An interval-valued recursive estimation framework for linearly parameterized systems

Laurent Bako, Seydi Ndiaye, Eric Blanco

An interval-valued recursive estimation framework for linearly parameterized systems

Laurent Bako, Seydi Ndiaye, Eric Blanco[a]

[a]*Univ Lyon, Ecole Centrale Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CNRS, Ampère, UMR 5005, 69130 Ecully, France*

Abstract

This paper proposes a recursive interval-valued estimation framework for identifying the parameters of linearly parameterized systems which may be slowly time-varying. It is assumed that the model error (which may consist in measurement noise or model mismatch or both) is unknown but lies at each time instant in a known interval. In this context, the proposed method relies on bounding the error generated by a given reference point-valued recursive estimator, for example, the well-known recursive least squares algorithm. We discuss the trade-off between computational complexity and tightness of the estimated parametric interval.

Keywords: adaptive estimation, interval-valued estimator

1. Introduction

In system identification, the estimation problem refers to the task of finding the parameters of a given parametrized model family in such a way that the resulting model matches (in some sense) a set of data. The main challenge with this task is most presumably how to deal with the uncertainty affecting the data with regards to the assumed model structure (e.g., in the form of model error or measurement noise). To hope for good estimates when the uncertainty is not negligible, it is important to model somehow the uncertainty. Probabilistic distributions are probably the most common models for describing uncertainties in many engineering fields. Such a modelling, when accurate, can lead to the design of better estimation schemes. A problem however is that a fine probabilistic modeling of the uncertainty may require a strong prior knowledge of the process being modelled. While such a reliable knowledge is rarely available, probabilistic models of the uncertainty may be severely wrong hence damaging the performance of the estimator. An alternative approach to the probabilistic one is to assume that the uncertain variables (e.g., the noise component) of the model, although unknown, live in bounded and predefined sets. This corresponds to the so-called set-membership representation of the uncertainty. In this latter setting the underlying idea of the parameter estimator design is to characterize the entire set of parameters which, through the induced models, are consistent with the data samples and the uncertainty sets. The literature of system identification abounds in such set-memberships approaches, see e.g., [11, 13, 8, 12, 5, 18, 19]. Various predefined geometrical forms can be considered for representing the parameter sets but we restrict our attention here to the case of intervals [8, 14]. Assuming that the model error (uncertainty) takes values in a known sequence of intervals, one can estimate intervals containing all the parameters which are consistent with the data, the model and its uncertainty. Among the existing methods which have tackled this question we can cite [10, 17] for batch mode of estimation and [15, 7] for recursive (online) mode of learning.

In this paper, we consider the problem of deriving a recursive interval-valued estimator for linearly parameterized models subject to a bounded uncertainty. The data model is assumed to be linear with respect to the parameters (although the input-output map may indeed be nonlinear). Then, under the assumption that the model error sequence is only known to lie in some interval bounds, we first construct a tight interval-valued estimator based on the error generated by the recursive least squares (RLS) algorithm [6, 16, 4]. However, this (tight) interval-valued estimator suffers from a level of computational complexity which is not affordable in practice when the estimation horizon grows towards infinity. We therefore turn to a family of approximate implementations whose complexity can be calibrated in function of the desired level of performance (measured here in term of tightness of the interval-valued estimate). That is, the proposed family of estimators offer the user the possibility to tie the size of the desired interval-valued estimate to the available computational resources. The proposed estimation framework applies to both stationary systems and slowly time-varying ones.

Closely related works to the current paper are the ones reported in [15] and [7]. The former solves at each time a linear program on a sliding window of constant length. The latter bounds the error generated by a bank of RLS identifiers. The current paper proposes a complementary development of this latter idea by considering a more general framework with regards to the model error representation. Moreover, our design method appears to be more systematic as it leads to a family of parametrizable interval-valued estimators. Also, it applies to both time-invariant and time-varying models with bounded change rate. In particular, it is shown in the time-invariant case that by applying an appropriate intersection operation, the size of the es-

Email address: laurent.bako@ec-lyon.fr,eric.blanco@ec-lyon.fr (Laurent Bako, Seydi Ndiaye, Eric Blanco)

timate is guaranteed to decrease monotonically.

Outline. Section 2 states the problem of recursive interval-valued estimation and outlines some necessary preliminaries on interval set-membership representation of uncertainty. Our method for designing interval-valued estimators relies on the error generated by a reference point-valued recursive estimator. Hence, Section 3 discusses one possible candidate for such a reference point-valued estimator, the recursive least squares (RLS) estimator with exponential forgetting factor. In Section 4, we present the main estimator and its variants. In Section 5, we consider the more general estimation setting where the to-be-estimated model is no longer constant but slowly time-varying. Section 6 reports some simulation results to illustrate and analyze the performance of the proposed estimation framework. Finally, Section 7 presents some conclusions.

Notation. $\mathbb{R}$ (resp. $\mathbb{R}_+$) is the set of real (resp. nonnegative real) numbers; $\mathbb{Z}$ (resp. $\mathbb{Z}_+$) is the set of (resp. nonnegative) integers. For a real number x, $|x|$ will refer to the absolute value of x. For $x = [x_1 \ \cdots \ x_n]^\top \in \mathbb{R}^n$, $\|x\|_p$ will denote the p-norm of x defined by $\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$, for $p \geq 1$. In particular for $p = \infty$, $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$. For a matrix $A \in \mathbb{R}^{n\times m}$, $\|A\|_F$ is the Frobenius norm of A defined by $\|A\|_F = \mathrm{tr}(A^\top A)^{1/2}$ (with tr referring to the trace of a matrix).

If $A = [a_{ij}]$ and $B = [b_{ij}]$ are real matrices of the same dimensions, the notation $A \leq B$ will be understood as an elementwise inequality on the entries, i.e., $a_{ij} \leq b_{ij}$ for all (i, j). $|A|$ corresponds to the matrix $[|a_{ij}|]$ obtained by taking the absolute value of each entry of A. In case A and B are real square matrices, $A \succeq B$ (resp. $A \succ B$) means that $A - B$ is positive semi-definite (resp. positive definite). I_n will denote the identity matrix of dimension n.

2. Problem statement

We consider a linearly parameterized[1] discrete-time dynamic system defined by

$$y(t) = x(t)^\top \theta^\circ + v(t), \tag{1}$$

where $y(t) \in \mathbb{R}$ is the measured output at the discrete time $t \in \mathbb{Z}_+$, $x(t) \in \mathbb{R}^n$ is the (known) regressor and $v(t) \in \mathbb{R}$ denotes an (unknown) noise component or a modeling error. The regressor $x(t)$ may, among other possibilities, assume a structure of the form

$$x(t) = \left[y(t-1) \ \cdots \ y(t-n_a) \, u(t)^\top \, u(t-1)^\top \ \cdots \ u(t-n_b)^\top\right]^\top$$

where $u(t) \in \mathbb{R}^{n_u}$ is the input of the system and the integers n_a and n_b are the model orders. $\theta^\circ \in \mathbb{R}^n$ in (1) is an unknown constant parameter vector which is to be estimated from data. The problem we consider in this paper is the following: given data points $\{(y(k), x(k))\}_{k=1}^t$ generated by the system (1) up to time t, we want to infer an estimate of the parameter vector θ°. However, since the sequence $\{v(t)\}$ is unknown here we can hardly hope for an exact recovery of θ°. Hence we consider the scenario where $v(t)$ is componentwise bounded for all $t \in \mathbb{Z}_+$ with known bounds and setup as our objective to characterize a set-valued estimate which is guaranteed to contain θ° while being consistent with the observed data.

Assumption 1. There exist (known) bounded sequences $\{(\underline{v}(t), \overline{v}(t))\}$ such that the noise sequence $\{v(t)\}$ in (1) satisfies $\underline{v}(t) \leq v(t) \leq \overline{v}(t)$ for all $t \in \mathbb{Z}_+$.

2.1. Some preliminaries on interval representation

Consider two vectors $\underline{x}$ and $\overline{x}$ in $\mathbb{R}^n$ such that $\underline{x} \leq \overline{x}$ with the inequality holding componentwise. An interval $[\underline{x}, \overline{x}]$ of $\mathbb{R}^n$ is the subset of $\mathbb{R}^n$ defined by

$$[\underline{x}, \overline{x}] = \{x \in \mathbb{R}^n : \underline{x} \leq x \leq \overline{x}\}. \tag{2}$$

An interval $[\underline{x}, \overline{x}]$ of $\mathbb{R}^n$ can be equivalently represented by

$$\mathscr{I}(c_x, r_x) \triangleq \{c_x + \mathrm{diag}\,(r_x)\alpha : \alpha \in \mathbb{R}^n, \ \|\alpha\|_\infty \leq 1\} \tag{3}$$

where

$$c_x = \frac{\overline{x} + \underline{x}}{2}, \quad r_x = \frac{\overline{x} - \underline{x}}{2} \tag{4}$$

The notation $\mathrm{diag}(v)$ for a vector $v \in \mathbb{R}^n$ refers to the diagonal matrix whose diagonal elements are the entries of v. We will call the so-defined c_x the *center or mid-point* of the interval $[\underline{x}, \overline{x}]$ and r_x its *radius* (a half of the width). To sum up, the interval set can be equivalently represented by the pairs $(\underline{x}, \overline{x}) \in \mathbb{R}^n \times \mathbb{R}^n$ and $(c_x, r_x) \in \mathbb{R}^n \times \mathbb{R}^n_+$ so that $[\underline{x}, \overline{x}] = \mathscr{I}(c_x, r_x)$. Finally, it will be useful to keep in mind for the rest of the paper that $\underline{x} = c_x - r_x$ and $\overline{x} = c_x + r_x$.

Definition 1 (Parametric interval estimator)**.**
Consider the system (1) under Assumption 1 and let $V^t = ((\underline{v}(0), \overline{v}(0)), \ldots, (\underline{v}(t), \overline{v}(t)))$ and $Y^t = (y(1), \ldots, y(t))$.
Consider a dynamical system defined by

$$\begin{aligned} \underline{\theta}(t) &= F_t(V^t, Y^t, \underline{\theta}(0), \overline{\theta}(0)) \\ \overline{\theta}(t) &= G_t(V^t, Y^t, \underline{\theta}(0), \overline{\theta}(0)) \end{aligned} \tag{5}$$

where F_t and G_t are some functions indexed by time, $(\underline{\theta}(t), \overline{\theta}(t))$ denote the output (or the state) of the system for any $t \in \mathbb{Z}_+$. The system (5) is called a (parametric) *interval-valued estimator* for the parameter vector θ° of system (1) if:

- **(a)** $\underline{\theta}(t) \leq \theta^\circ \leq \overline{\theta}(t)$ for all $t \in \mathbb{Z}_+$, whenever $\underline{\theta}(0) \leq \theta^\circ \leq \overline{\theta}(0)$
- **(b)** (5) is Bounded Input-Bounded Output (BIBO) stable i.e., if the signals v and y and the initial state $(\underline{\theta}(0), \overline{\theta}(0))$ are all bounded then so is $(\underline{\theta}, \overline{\theta})$.

Now we recall from [2, 3] a lemma that will play a central role in the design of interval-valued estimators.

Lemma 1. *Let $M \in \mathbb{R}^{n\times m}$ and $(\underline{z}, \overline{z}) \in \mathbb{R}^m \times \mathbb{R}^m$ such that $\underline{z} \leq \overline{z}$. Consider the set $\mathcal{I}$ defined by $\mathcal{I} = \{Mz : \underline{z} \leq z \leq \overline{z}\}$. Define the*

[1] Note that the considered system may indeed be nonlinear in term of input-output relation. For example, $x(t)$ may be of the form $x(t) = \phi(z(t))$ with ϕ being a known nonlinear map and $z(t)$ is formed from measurements.

vectors (c, r) *by*

$$\begin{aligned} c &= Mc_z \\ r &= |M|\, r_z, \end{aligned} \tag{6}$$

with $c_z = (\overline{z} + \underline{z})/2$ *and* $r_z = (\overline{z} - \underline{z})/2$.
Then $[c - r, c + r]$ *is the tightest interval containing* $\mathcal{I}$.

We now state formally the estimation problem.

Problem. Given the data $(y(k), x(k))_{1\le k\le t}$ generated by system (1) up to an arbitrary time $t \in \mathbb{Z}_+$, the uncertainty bounds $\left\{(\underline{v}(t), \overline{v}(t))\right\}_{1\le k\le t}$ on the noise sequence as defined in Assumption 1 and a prior (initial) interval set $\mathcal{I}(c_\theta(0), r_\theta(0))$ containing the true parameter vector θ° from (1), we are interested in finding an interval-valued estimate of the form $\mathcal{I}(c_\theta(t), r_\theta(t)) \subset \mathbb{R}^n$ (in the sense of Definition 1), of the parameter vector θ° in (1) which is consistent with data. Moreover, it is desirable that the estimate $(c_\theta(t), r_\theta(t))$ at time t be obtained by a simple update mechanism from the measurements $(x(t), y(t), \underline{v}(t), \overline{v}(t))$ at time t and a finite number m of past estimates $(c_\theta(t-i), r_\theta(t-i))$, $i = 1, \dots, m$.

We will describe in Section 4 a framework for deriving a solution to this problem. Our method for constructing a recursive set-valued estimator requires three ingredients: (a) a reference adaptive point-valued identifier; (b) a characterization of the stability of the associated error dynamics ; (c) an appropriate mechanism for deducing the set-valued estimate from the point-valued one. Many recursive identifiers may be suitable for the role (a) mentioned above. Here however we choose to discuss only the RLS algorithm.

3. A reference adaptive identifier

For the purpose of designing the recursive interval-valued estimator as stated above, we first study a reference adaptive point-valued identifier.

3.1. Recursive least squares (RLS)

A candidate adaptive identifier for point (a) above is the exponentially weighted recursive least squares (RLS) algorithm which returns a point-valued estimate $\theta(t)$ of θ°, selected at each time t to be the minimizing point of an objective function $\theta \mapsto V_t(\theta)$,

$$\theta(t) = \arg\min_{\theta\in\mathbb{R}^n} V_t(\theta), \tag{7}$$

with $V_t(\theta)$ defined by

$$V_t(\theta) = \frac{1}{2}\sum_{k=1}^{t} \lambda^{t-k}(y(t) - x(t)^\top\theta)^2 + \frac{\lambda^t}{2}(\theta-\theta_0)^\top P_0^{-1}(\theta-\theta_0). \tag{8}$$

In Eq. (8), θ_0 refers to a prior guess for the parameter vector, $P_0 \succ 0$ is a symmetric positive-definite weighting matrix reflecting the uncertainty related to the guess θ_0, and $\lambda \in]0, 1[$ is a forgetting factor which intends to downweight the information contained in the oldest data with respect to time t.
Note that the objective function in (8) is continuous, coercive and strictly convex, hence implying that the minimizer in (7) exists and is unique. It can be shown that there exists a sequence of symmetric matrices[2] $\{P(t)\}$ such that the solution $\theta(t)$ to the optimization problem (7) can be recursively expressed as [6]:

$$\theta(t) = \theta(t-1) + q(t)(y(t) - x(t)^\top\theta(t-1)) \tag{9}$$

$$q(t) = \frac{P(t-1)x(t)}{\lambda + x(t)^\top P(t-1)x(t)} \tag{10}$$

$$P(t) = \frac{1}{\lambda}(P(t-1) - q(t)x(t)^\top P(t-1)) \tag{11}$$

where $\theta(0) = \theta_0$, $P(0) = P_0$. Eqs (9)-(11) define the well-known recursive least squares (point-valued) identifier with exponential forgetting factor [9].
For the purpose of the analysis to be presented in the sequel, define the parametric error $\tilde{\theta}(t) = \theta(t) - \theta^\circ$. Then it follows from the system equation (1) and the θ-update equation (9) that the error has the following dynamics

$$\tilde{\theta}(t) = A(t)\tilde{\theta}(t-1) + q(t)v(t), \tag{12}$$

with $A(t) = I_n - q(t)x(t)^\top$. Eq. (12) together with (10)-(11) represents a dynamic system with input $\{v(t)\}$ and state $\{\tilde{\theta}(t)\}$. For future use in the paper, we can further express $\tilde{\theta}(t)$ in function of the initial error $\tilde{\theta}(0)$ and the noise sequence $\{v(k)\}_{1\le k\le t}$ up to time t,

$$\tilde{\theta}(t) = \Phi(t, 0)\tilde{\theta}(0) + \sum_{j=1}^{t} \Phi(t, j)q(j)v(j), \tag{13}$$

where Φ is the state transition matrix defined by

$$\Phi(t, t_0) = \begin{cases} I_n & t = t_0 \\ A(t)\cdots A(t_0+1) & t > t_0 \end{cases} \tag{14}$$

An interesting property of the state transition matrix is that for any triplet (t, t_1, t_0) of nonnegative integers satisfying $t \ge t_1 \ge t_0$,

$$\Phi(t, t_0) = \Phi(t, t_1)\Phi(t_1, t_0). \tag{15}$$

Now we recall the stability concept which is of interest in the following developments. For this purpose, consider the homogenous part of system (12) (i.e., the one obtained when the input v satisfies $v \equiv 0$), which we may generically describe by

$$\xi(t) = A(t)\xi(t-1), \quad \xi(0) = \xi_0 \tag{16}$$

where $A : \mathbb{Z}_+ \to \mathbb{R}^{n\times n}$ is a matrix-valued function and $\xi(t) \in \mathbb{R}^n$ is the state of the system (16) at time $t \in \mathbb{Z}_+$. For any $(t, t_0) \in \mathbb{Z}_+$ with $t \ge t_0$, $\xi(t)$ can be related to $\xi(t_0)$ by $\xi(t) = \Phi(t, t_0)\xi(t_0)$. Using the generic LTV system (16), we now define the notion of exponential stability.

Definition 2. The LTV system (16) is said to be exponentially stable if there exist some constants $\gamma > 0$ and $\rho \in [0, 1[$ such that

$$\|\xi(t)\|_2 \le \gamma\rho^{t-t_0}\,\|\xi(t_0)\|_2 \tag{17}$$

for all $(t, t_0) \in \mathbb{Z}_+$ such that $t \ge t_0$. Indeed (17) is equivalent to

[2] Indeed we have $P(t) = \left[\sum_{k=1}^{t} \lambda^{t-k}x(t)x(t)^\top + \lambda^t P_0^{-1}\right]^{-1}$ so that $P^{-1}(t) = \lambda P^{-1}(t-1) + x(t)x(t)^\top$.

$\|\Phi(t,t_0)\|_2 \leq \gamma\rho^{t-t_0}$.

Finally, note that the non-homogenous system (12) is stable in the BIBO sense if (16) is (exponentially) stable and the gain sequence $\{q(t)\}$ is bounded. We will see in the next section that such a property is guaranteed for the error system (12) provided that the regressor $\{x(t)\}$ from the system (1) is bounded and enjoys some richness condition.

3.2. A stability property for the RLS

We first recall a definition of the concept of (uniform) persistence of excitation [9].

Definition 3. A vector-valued sequence $\{x(t)\} \subset \mathbb{R}^n$ is said to be persistently exciting (PE) if there exist some strictly positive constants α and β (called excitation levels) and a time horizon T such that

$$\alpha I_n \leq \sum_{k=t+1}^{t+T} x(k)x(k)^\top \leq \beta I_n \quad \forall t \in \mathbb{Z}_+ \tag{18}$$

The lower bound of (18) requires that the matrix of regressor $X_t \triangleq [x(t+1) \quad \cdots \quad x(t+T)]$ be full rank on any time horizon of length T. Additionally, the smallest eigenvalue of $X_t X_t^\top$ must be larger than a minimum level $\alpha > 0$. The upper bound in (18) expresses uniform boundedness of the sequence $\{x(t)\}$.

Lemma 2. *Consider the RLS algorithm* (9)-(11) *under the assumptions that* $P(0) > 0$ *and* $\lambda \in]0,1[$. *Then the matrices* $P(t)$ *defined by* (11) *are invertible for all* $t \in \mathbb{Z}_+$ *and the sequence* $\left\{P^{-1}(t)\right\}$ *of their inverses satisfy*

$$P^{-1}(t) = \lambda P^{-1}(t-1) + x(t)x(t)^\top. \tag{19}$$

Moreover, if $\{x(t)\}$ *is PE in the sense of Definition 3 with horizon* T *and excitation levels* (α,β), *then* $\left\{P^{-1}(t)\right\}$ *is uniformly bounded as follows*

$$\gamma_1 I_n \leq P^{-1}(t) \leq \gamma_2 I_n \quad \forall t \geq 0 \tag{20}$$

with

$$\gamma_1 = \min\left(\delta_1, \alpha\lambda^{2T-1}\right) \tag{21}$$

$$\gamma_2 = \max\left(\delta_2, \lambda^T \sigma_{\max}(P^{-1}(0)) + \beta\frac{2-\lambda}{1-\lambda}\right) \tag{22}$$

and $\delta_1 = \min_{t=0,\ldots,T-1} \sigma_{\min}[P^{-1}(t)]$, $\delta_2 = \max_{t=0,\ldots,T-1} \sigma_{\max}[P^{-1}(t)]$, $\sigma_{\min}[\cdot]$ *and* $\sigma_{\max}[\cdot]$ *standing for the minimum and maximum eigenvalues respectively.*

A proof of this lemma can be found in [1].
Next we derive an input-to-state-stability (ISS) property for the error dynamics (12) subject to (10)-(11).

Theorem 1. *Consider the RLS algorithm applied to the data generated by system* (1). *If the regressor sequence* $\{x(t)\}$ *is PE, then*

$$\left\|\tilde{\theta}(t)\right\|_2^2 \leq \frac{1}{\gamma_1}\left[\lambda^t \sigma_{\max}[P^{-1}(0)]\left\|\tilde{\theta}(0)\right\|_2^2 + \sum_{k=1}^{t} \lambda^{t-k} v(k)^2\right] \tag{23}$$

where $\tilde{\theta}(t) = \theta(t) - \theta^\circ$ *is the parametric estimation error at time* t *and* γ_1 *is any positive number satisfying* (20).

Proof. Let $V(t) = \tilde{\theta}(t)^\top P^{-1}(t)\tilde{\theta}(t)$. By subtracting the true parameter vector θ° from each side of (9) and invoking the equation of the data-generating system (1), it is easy to see that $\tilde{\theta}(t) = \tilde{\theta}(t-1) + q(t)\varepsilon(t)$, where $\varepsilon(t) = y(t) - x(t)^\top\hat{\theta}(t-1) = v(t) - x(t)^\top\tilde{\theta}(t-1)$. On the other hand, we know from Lemma 2 that $\{P^{-1}(t)\}$ obeys the recursive relation (19). Now by direct algebraic calculations it can be seen that

$$\begin{aligned} V(t) &= (\tilde{\theta}(t-1) + q(t)\varepsilon(t))^\top(\lambda P^{-1}(t-1) + x(t)x(t)^\top) \times \ldots \\ &\qquad \ldots \times (\tilde{\theta}(t-1) + q(t)\varepsilon(t)) \\ &= \lambda V(t-1) + 2\lambda\tilde{\theta}(t-1)^\top P^{-1}(t-1)q(t)\varepsilon(t) \\ &\quad + 2(x(t)^\top\tilde{\theta}(t-1))(x(t)^\top q(t))\varepsilon(t) + (x(t)^\top\tilde{\theta}(t-1))^2 \\ &\quad + \lambda q(t)^\top P^{-1}(t-1)q(t)\varepsilon(t)^2 + (x(t)^\top q(t))^2\varepsilon(t)^2 \end{aligned}$$

Note now that by posing $s(t) = \lambda + x(t)^\top P(t-1)x(t)$, we have

$$\begin{aligned} q(t)^\top P^{-1}(t-1)q(t) &= \frac{1}{s(t)} - \frac{\lambda}{s(t)^2} \\ x(t)^\top q(t) &= 1 - \frac{\lambda}{s(t)} \\ P^{-1}(t-1)q(t) &= \frac{x(t)}{s(t)} \end{aligned}$$

Substituting these formulas in the above expression of $V(t)$ gives

$$V(t) = \lambda V(t-1) - \frac{\lambda}{s(t)}\varepsilon(t)^2 + v(t)^2.$$

It follows that $V(t) \leq \lambda V(t-1) + v(t)^2$. Iterating this last equation and invoking the property (20) of $\{P^{-1}(t)\}$ yields

$$\gamma_1\left\|\tilde{\theta}(t)\right\|_2^2 \leq V(t) \leq \lambda^t V(0) + \sum_{k=1}^{t} \lambda^{t-k} v(k)^2$$

which, by using the fact that $P^{-1}(0) \leq \sigma_{\max}[P^{-1}(0)]I_n$, implies that

$$\gamma_1\left\|\tilde{\theta}(t)\right\|_2^2 \leq \lambda^t \sigma_{\max}[P^{-1}(0)]\left\|\tilde{\theta}(0)\right\|_2^2 + \sum_{k=1}^{t} \lambda^{t-k} v(k)^2.$$

Hence the claim of the theorem is established. □

Corollary 1. *Under the conditions of Theorem 1, if the noise* v *of model* (1) *is identically equal to zero, then*

$$\left\|\tilde{\theta}(t)\right\|_2 \leq \left(\frac{\lambda^t \sigma_{\max}[P^{-1}(0)]}{\gamma_1}\right)^{1/2}\left\|\tilde{\theta}(0)\right\|_2 \tag{24}$$

that is, the sequence $\{\theta(t)\}$ *generated by the RLS algorithm converges to* θ° *exponentially fast regardless of the initial point* $\theta(0)$.

Lemma 3. *Consider the state transition matrix-valued function* Φ *defined in* (14) *from the RLS error system* (12). *If the regressor sequence* $\{x(t)\} \subset \mathbb{R}^n$ *is PE, then there exist constant*

real positive numbers γ_1 and γ_2 such that for all (t, t_0) obeying $t \geq t_0$,

$$\|\Phi(t, t_0)\|_F \leq c\rho^{t-t_0} \tag{25}$$

where $c = (n\gamma_2\gamma_1^{-1})^{1/2}$ and $\rho = \lambda^{1/2}$, λ being the forgetting factor of the RLS algorithm.

Proof. Consider the error system (12) under the assumption that the noise sequence $\{v(t)\}$ is equal to zero. Then for any (t, t_0) such that $t \geq t_0$, we have $\tilde{\theta}(t) = \Phi(t, t_0)\tilde{\theta}(t_0)$. Moreover, Corollary 1 can be applied by replacing the time origin for an arbitrary $t_0 \in \mathbb{Z}_+$ such that $t \geq t_0 \geq 0$. This gives

$$\left\|\tilde{\theta}(t)\right\|_2 \leq \left(\frac{\lambda^{t-t_0}\sigma_{\max}[P^{-1}(t_0)]}{\gamma_1}\right)^{1/2} \left\|\tilde{\theta}(t_0)\right\|_2$$

for any value of $\tilde{\theta}(t_0) \in \mathbb{R}^n$. Since the PE condition holds here for $\{x(t)\}$, we know by Lemma 2 that there exists a constant number $\gamma_2 > 0$ such that $\sigma_{\max}[P^{-1}(t_0)] \leq \gamma_2$ (see Eq. (20)). We can hence write

$$\left\|\tilde{\theta}(t)\right\|_2 \leq \left(\frac{\gamma_2}{\gamma_1}\right)^{1/2} \rho^{t-t_0} \left\|\tilde{\theta}(t_0)\right\|_2.$$

This implies that

$$\|\Phi(t, t_0)\|_2 = \sup_{\tilde{\theta}(t_0)\neq 0} \frac{\left\|\Phi(t, t_0)\tilde{\theta}(t_0)\right\|_2}{\left\|\tilde{\theta}(t_0)\right\|_2} \leq \left(\frac{\gamma_2}{\gamma_1}\right)^{1/2} \rho^{t-t_0}.$$

Finally, the result follows by recalling that $\|\Phi(t, t_0)\|_F \leq \sqrt{n}\,\|\Phi(t, t_0)\|_2$. □

4. Interval-valued estimator

In this section we present the main contributions of the paper concerning the development of an adaptive interval-valued parametric estimator. As explained at the end of Section 2, our method relies on the error sequence generated by a point-value adaptive estimator. Considering the special case of the RLS, we obtain the error dynamics expressed in (13) which is directly related to the noise. Applying Lemma 1 to this equation gives an interval estimate of the error $\tilde{\theta}(t) = \hat{\theta}(t) - \theta^\circ$, which can then be modified to get an estimate of θ°.

4.1. Derivation of an interval-valued estimator

Assume now that we are given a recursive point-valued estimator (say the RLS algorithm dicussed earlier) generating a sequence of estimates $\{\theta(t)\}$ for θ° in (1). To derive an interval-valued estimator for θ°, we first find an interval-valued estimate for the error $\tilde{\theta}(t)$ defined in (12). We do so by applying Lemma 1 to (13) which, for convenience, can be rewritten as $\tilde{\theta}(t) = M(t)z(t)$ with

$$M(t) = \begin{bmatrix}\Phi(t,0) & \Phi(t,1)q(1) & \cdots & \Phi(t,t)q(t)\end{bmatrix}$$
$$z(t) = \begin{bmatrix}\tilde{\theta}(0)^\top & v(1) & \cdots & v(t)\end{bmatrix}^\top.$$

We hence obtain immediately from Lemma 1 that the smallest interval set containing the parametric error $\tilde{\theta}(t)$ can be expressed in term of its center-radius pair $(c_{\tilde{\theta}}, r_{\tilde{\theta}})$ given by

$$c_{\tilde{\theta}}(t) = \Phi(t, 0)c_{\tilde{\theta}}(0) + \sum_{j=1}^{t} \Phi(t, j)q(j)c_v(j) \tag{26}$$

$$r_{\tilde{\theta}}(t) = |\Phi(t, 0)|r_{\tilde{\theta}}(0) + \sum_{j=1}^{t} |\Phi(t, j)q(j)|r_v(j) \tag{27}$$

where (c_v, r_v) is the pair of signals defining the intervals of the noise sequence $\{v(t)\}$ and $c_{\tilde{\theta}}(0) = c_\theta(0) - \theta^\circ$ and $r_{\tilde{\theta}}(0) = r_\theta(0)$. Recalling now that $\theta^\circ = \theta(t) - \tilde{\theta}(t)$, an interval-valued estimate of the θ° can be obtained as proposed in the following proposition.

Proposition 1. *Consider the system* (1) *and assume that the regressor sequence $\{x(t)\}$ is PE in the sense of Definition 3 and that the noise $\{v(t)\}$ is bounded and admits an interval representation $(c_v(t), r_v(t))$. Then the pair $(c_\theta(t), r_\theta(t))$ given by*

$$\begin{cases} c_\theta(t) = \theta(t) - c_{\tilde{\theta}}(t) \\ r_\theta(t) = r_{\tilde{\theta}}(t) \end{cases} \tag{28}$$

with $(c_{\tilde{\theta}}(t), r_{\tilde{\theta}}(t))$ as in (26)-(27)*, defines an interval estimator for the parameter vector θ°.*

Proof. To begin with, let us recall that $\theta^\circ = \theta(t) - \tilde{\theta}(t)$ and (26)-(27) is such that $\tilde{\theta}(t) \in [c_{\tilde{\theta}}(t) - r_{\tilde{\theta}}(t), c_{\tilde{\theta}}(t) + r_{\tilde{\theta}}(t)]$ for all $t \in \mathbb{Z}_+$. It follows, as an immediate consequence, that $\theta^\circ \in [c_\theta(t) - r_\theta(t), c_\theta(t) + r_\theta(t)]$ with (c_θ, r_θ) defined as in (28). To reach the conclusion of the proposition, it remains to prove that the dynamical systems (operators) $(c_\theta(0), c_v) \mapsto c_\theta$ and $(r_\theta(0), r_v) \mapsto r_\theta$ are BIBO stable. By relying on (26)-(27), it is immediate to see that, under the PE condition, both properties follow indeed from (25) which in turn is a consequence of Theorem 1. □

4.2. Computational aspects

Implementing numerically the estimator (28) requires computing $(c_{\tilde{\theta}}(t), r_{\tilde{\theta}}(t))$ defined in (26)-(27) for any time t. A problem however is that these convolutional formulas become infeasible in practice when t grows towards infinity. Therefore it is desirable to find an efficient implementation of this estimator for example, in the form of a one-step-ahead state-space recursive realization. In this perspective, note that $c_\theta(t)$ can be computed recursively through the following equation

$$c_\theta(t) = A(t)c_\theta(t-1) + q(t)(y(t) - c_v(t)). \tag{29}$$

Unfortunately, there is, in general, no simple recursive implementation for the interval radius r_θ as defined (27)-(28). Hence, a strategy would be to search for a more pessimistic estimate r_θ but which would be implementable. That is, the computational constraint introduces a dose of pessimism in the estimation, resulting in a less tight interval-valued estimate of θ°. A possible solution is to replace r_θ with a truncated version $\hat{r}_{\theta,m}$ defined,

for a given integer $m > 0$, by

$$\hat{r}_{\theta,m}(t) = \begin{cases} |\Phi(t,0)|\, r_\theta(0) + \sum_{k=1}^{t} |\Phi(t,k)q(k)| r_v(k) \\ \quad \text{if } t = 0,\dots,m \\ |\Phi(t,t-m)|\, \hat{r}_{\theta,m}(t-m) + \sum_{k=t-m+1}^{t} |\Phi(t,k)q(k)| r_v(k), \\ \quad \text{if } t > m \end{cases} \tag{30}$$

Intuitively $\hat{r}_{\theta,m}$ is all the smaller as m is large. On the other hand the computational complexity grows with m. Note that in the extreme case where $m = t$, we get $\hat{r}_{\theta,m}(t) = r_\theta(t)$ for all $t \in \mathbb{Z}_+$. The simplest version (but also the most pessimistic) of the family (30) of estimates is obtained for $m = 1$,

$$\hat{r}_{\theta,1}(t) = |A(t)|\, \hat{r}_{\theta,1}(t-1) + |q(t)| r_v(t) \tag{31}$$

However, as we will see shortly, such an estimate is unlikely to satisfy the BIBO condition of Definition 1. As a consequence it will not qualify in general as an interval-valued estimator

The result below formally shows that for any time t, the interval $\mathscr{I}(c_\theta(t, r_\theta(t))$ is included in $\mathscr{I}(c_\theta(t), \hat{r}_{\theta,m}(t))$.

Lemma 4. *Consider the interval radii $r_\theta(t)$ and $\hat{r}_{\theta,m}(t)$ defined in* (27) *and* (30) *respectively. For any fixed integer m, it holds that $r_\theta(t) \le \hat{r}_{\theta,m}(t)$ for all $t \in \mathbb{Z}_+$.*

Proof. We start by observing that $\hat{r}_{\theta,m}(t) = r_\theta(t)$ for $t = 0,\dots,m$. Hence the inequality is true for $t = 0,\dots,m$. If $t > m$, write $t = \alpha(t)m + \beta(t)$ for some positive integers $(\alpha(t), \beta(t))$ such that $0 \le \beta(t) < m$. By iterating the second equation of (30), we ultimately get

$$\begin{aligned} \hat{r}_{\theta,m}(t) =& \Big(\prod_{\ell=1}^{\alpha(t)} \Big|\Phi(t-(\ell-1)m, t-\ell m)\Big|\Big)\hat{r}_{\theta,m}(\beta(t)) \\ &+ \sum_{j=1}^{\alpha(t)} \sum_{k=t-jm+1}^{t-(j-1)m} \Big(\prod_{\ell=1}^{j-1} \Big|\Phi(t-(\ell-1)m, t-\ell m)\Big|\Big) \times \dots \\ &\qquad \dots \times \Big|\Phi(t-(j-1)m, k)q(k)\Big| r_v(k) \end{aligned}$$

Moreover, since $\beta(t) = t - \alpha(t)m$ satisfies $0 \le \beta(t) < m$, by the definition (30) of $\hat{r}_{\theta,m}(t)$ we know that

$$\hat{r}_{\theta,m}(\beta(t)) = |\Phi(t-\alpha(t)m, 0)|\, r_\theta(0) + \sum_{k=1}^{t-\alpha(t)m} |\Phi(t-\alpha(t)m, k)q(k)| r_v(k).$$

Plugging this in the above expression yields

$$\begin{aligned} \hat{r}_{\theta,m}(t) =& \Big(\prod_{\ell=1}^{\alpha(t)+1} \Big|\Phi(\kappa(t,\ell-1), \kappa(t,\ell))\Big|\Big) r_\theta(0) \\ &+ \sum_{j=1}^{\alpha(t)+1} \sum_{\kappa(t,j)+1}^{\kappa(t,j-1)} \Big(\prod_{\ell=1}^{j-1} \Big|\Phi(\kappa(t,\ell-1), \kappa(t,\ell))\Big|\Big) \times \dots \\ &\qquad \dots \times \Big|\Phi(\kappa(t,j-1), k)q(k)\Big| r_v(k) \end{aligned} \tag{32}$$

where $\kappa(t, j) = \max(t - jm, 0)$. Using the property $|A||B| \ge |AB|$, we observe, for example, that the matrix in the first term of (32) can be bounded as follows

$$\begin{aligned} &\prod_{\ell=1}^{\alpha(t)+1} \Big|\Phi(\kappa(t,\ell-1), \kappa(t,\ell))\Big| \\ &= \Big(\prod_{i=1}^{\alpha(t)} \Big|\Phi(t-(i-1)m, t-im)\Big|\Big) |\Phi(t-\alpha(t)m, 0)| \\ &\ge \Big|\Phi(t,t-m)\Phi(t-m,t-2m)\cdots\Phi(t-\alpha(t)m,0)\Big| \\ &= \Big|\Phi(t,0)\Big|, \end{aligned}$$

where the last equality follows from the property (15) of the state transition matrix. Applying this property to the whole expression of $\hat{r}_{\theta,m}(t)$ above leads to

$$\begin{aligned} \hat{r}_{\theta,m}(t) &\ge \Big|\Phi(t,0)\Big| r_\theta(0) + \sum_{j=1}^{\alpha(t)} \sum_{k=t-jm+1}^{t-(j-1)m} \Big|\Phi(t,k)q(k)\Big| r_v(k) \\ &\qquad + \sum_{k=1}^{t-\alpha(t)m} |\Phi(t,k)q(k)| r_v(k) \\ &= \Big|\Phi(t,0)\Big| r_\theta(0) + \sum_{k=1}^{t} \Big|\Phi(t,k)q(k)\Big| r_v(k) \\ &= r_\theta(t) \end{aligned}$$

This concludes the proof. □

Theorem 2. *Consider the system* (1) *and assume that the regressor sequence $\{x(t)\}$ is PE in the sense of Definition 3 and that the noise $\{v(t)\}$ is bounded and admits an interval representation $(c_v(t), r_v(t))$. Then there exists a number $m^\star > 0$ such that for all $m \ge m^\star$, $(c_\theta(t), \hat{r}_{\theta,m}(t))$ defined in* (28) *and* (30) *constitutes an interval estimator for the parameter vector θ°.*

Proof. By Lemma 4, we know that $r_\theta(t) \le \hat{r}_{\theta,m}(t)$. Hence it is immediate by Proposition 1 that $\theta^\circ \in [c_\theta(t) - \hat{r}_{\theta,m}(t), c_\theta(t) + \hat{r}_{\theta,m}(t)]\ \forall t \in \mathbb{Z}_+$. We just need to establish the second condition of Definition 1. More precisely, we need to show that the sequence of intervals $[c_\theta(t) - \hat{r}_{\theta,m}(t), c_\theta(t) + \hat{r}_{\theta,m}(t)]$ is bounded. For this purpose we will show that the systems $(c_\theta(0), c_v) \mapsto c_\theta$ and $(r_\theta(0), r_v) \mapsto \hat{r}_{\theta,m}$ are stable.

Boundedness of $\{c_\theta(t)\}$. Recall that a state-space realization of c_θ is given in (29). Relying on this equation, it is clear that $(c_\theta(0), c_v) \mapsto c_\theta$ is BIBO state if and only if $(\tilde{\theta}(0), v) \mapsto \tilde{\theta}$ defined in (12) is BIBO stable. As already shown in the proof of Theorem 1, the PE condition suffices to guarantee that $\tilde{\theta}$ is bounded whenever $(\tilde{\theta}(0), v)$ is bounded. We hence conclude that $(c_\theta(0), c_v) \mapsto c_\theta$ is BIBO.

Boundedness of $\{\hat{r}_{\theta,m}(t)\}$. We will rely on formula (32). Taking the Frobenius norm of $\hat{r}_{\theta,m}(t)$ and applying the submultiplicativity property of the Frobenius norm and the fact that $|||A|||_F = \|A\|_F$ (i.e., Frobenius norm of A equals Frobenius norm

of $|A|$) for any matrix A, we can write

$$\left\|\hat{r}_{\theta,m}(t)\right\|_2 \leq \Big(\prod_{i=1}^{\alpha(t)+1}\left\|\Phi(\kappa(t,i-1),\kappa(t,i))\right\|_F\Big)\|r_\theta(0)\|_2 + \sum_{j=1}^{\alpha(t)+1}\sum_{k=\kappa(t,j)+1}^{\kappa(t,j-1)}\Big(\prod_{\ell=1}^{j-1}\left\|\Phi(\kappa(t,\ell-1),\kappa(t,\ell))\right\|_F\Big)\times\ldots \ldots\times\left\|\Phi(\kappa(t,j-1),k)\right\|_F\|q(k)\|_2\,\|r_v(k)\|_2$$

Here, we have used the fact that $\|x\|_F = \|x\|_2$ for any vector x. Since the PE condition holds, it follows from the analysis of Section 3 (See Eq. (25)) that the transition matrix Φ satisfies $\|\Phi(t,t_0)\|_F \leq c\rho^{t-t_0}$ with the constants $c>0$ and ρ being defined as in (25). Applying this in the above inequality gives

$$\begin{aligned}\left\|\hat{r}_{\theta,m}(t)\right\|_2 &\leq (c\rho^m)^{\alpha(t)}(c\rho^{t-\alpha(t)m})\,\|r_\theta(0)\|_2 \\ &+\sum_{j=1}^{\alpha(t)}\sum_{k=t-jm+1}^{t-(j-1)m}(c\rho^m)^{j-1}(c\rho^{t-(j-1)m-k})\,\|q(k)\|_2\,\|r_v(k)\|_2 \\ &+\sum_{k=1}^{t-\alpha(t)m}(c\rho^m)^{\alpha(t)}(c\rho^{t-\alpha(t)m-k})\,\|q(k)\|_2\,\|r_v(k)\|_2\end{aligned}$$

Under the PE condition of $\{x(t)\}$, we know by Lemma 2 that $\{P(t)\}$ is uniformly bounded as $1/\gamma_2 I_n \preceq P(t) \preceq 1/\gamma_1 I_n$ for all t. From this, it is easy to see that the vector $q(t)$ defined in (10) satisfies

$$\|q(k)\|_2 \leq \frac{1/\gamma_1\,\sup_t\|x(t)\|_2}{\lambda + 1/\gamma_2\,\inf_t\|x(t)\|_2^2}. \tag{33}$$

This implies that $\{q(t)\}$ is upper-bounded. On the other hand, r_v is bounded by assumption. Let therefore consider the bounds $\eta_q = \sup_{k\in\mathbb{Z}_+}\|q(k)\|_2$ and $\eta_v = \sup_{k\in\mathbb{Z}_+}\|r_v(k)\|_2$. Using these notations and proceeding from above gives

$$\left\|\hat{r}_{\theta,m}(t)\right\|_2 \leq (c\rho^m)^{\alpha(t)}c\,\|r_\theta(0)\|_2 + c\eta_q\eta_v\frac{1-\rho^m}{1-\rho}\frac{1-(c\rho^m)^{\alpha(t)}}{1-c\rho^m} + c\eta_q\eta_v(c\rho^m)^{\alpha(t)}\frac{1-\rho^{t-\alpha(t)m}}{1-\rho}$$

This inequality can be refined as

$$\left\|\hat{r}_{\theta,m}(t)\right\|_2 \leq (c\rho^m)^{\alpha(t)}c\,\|r_\theta(0)\|_2 + \frac{c\eta_q\eta_v}{1-\rho}\left[\frac{1-\rho^m}{1-c\rho^m} + (c\rho^m)^{\alpha(t)}(1-\rho^{m-1})\right]. \tag{34}$$

Note that $\alpha(t) = \lfloor t/m\rfloor \to +\infty$ as $t\to+\infty$. The notation $\lfloor\cdot\rfloor$ refers here to the floor function (integer part). Hence, if $c\rho^m < 1$, that is, if $m > m^\star \triangleq -\dfrac{\ln(c)}{\ln(r)}$, then the sequence $\{\hat{r}_{\theta,m}(t)\}$ is bounded. □

A candidate for the constant c is the one expressed in (25). By making use of it, a full expression of $m^\star$ can be obtained as

$$m^\star = -\frac{\ln\left(n\gamma_2\gamma_1^{-1}\right)}{\ln(\lambda)} \tag{35}$$

This suggests that the richer the regressor sequence $\{x(t)\}$ (that is, the smaller the ratio γ_1/γ_2), the smaller the threshold $m^\star$ will be. Note indeed that γ_1 and γ_2 depend not only on the data sequence $\{x(t)\}$ but also on the forgetting factor λ and the initial weighting matrix $P^{-1}(0)$. A few further comments can be made concerning the behavior of $\hat{r}_{\theta,m}$. First, note that an asymptotic bound on the estimated interval radius can be derived as follows

$$\limsup_{t\to+\infty}\left\|\hat{r}_{\theta,m}(t)\right\|_2 \leq \frac{c\eta_q\eta_v}{1-\rho}\frac{1-\rho^m}{1-c\rho^m}. \tag{36}$$

Then we see that as the truncation order m grows, the asymptotic bound on $\hat{r}_{\theta,m}(t)$ gets closer to $b_\infty^\star \triangleq c\eta_q\eta_v/(1-\rho)$. By invoking Eq. (33) it is immediate to see that if we let $h_{\min} = \inf_t\|x(t)\|_2$ and $h_{\max} = \sup_t\|x(t)\|_2$, then $\eta_q \leq \frac{\gamma_2}{\gamma_1}\frac{h_{\max}}{h_{\min}^2+\lambda\gamma_2}$ which, by using the expressions of c and ρ given in (25), implies that

$$b_\infty^\star \leq \frac{\eta_v n^{1/2}}{1-\lambda^{1/2}}\left(\frac{\gamma_2}{\gamma_1}\right)^{3/2}\frac{h_{\max}}{h_{\min}^2+\lambda\gamma_2}. \tag{37}$$

What this shows is that the influencing parameters of the bound $b_\infty^\star$ originates from three sources: (i) the parameters measuring richness of the learning data: $\gamma_1,\gamma_2,h_{\min},h_{\max}$; (ii) the design parameters of the estimator: λ, $P^{-1}(0)$; (iii) the magnitude η_v of the uncertainty associated with the mathematical representation (1) of the data.

4.3. Further improvements

Due to the presence of noise in the data, the size of the interval estimates (c_θ, r_θ) or $(c_\theta, \hat{r}_{\theta,m})$ discussed above may oscillate over time instead of decreasing monotonically (See Figure 1 for a visual illustration of this phenomenon). This behavior is undesirable in practice and should be mitigated as much as possible. For this purpose, we discuss here a simple recursive intersection operation for removing such possible non monotonic trend of the interval-valued estimate for the estimators proposed in the previous sections. To this end, consider a pair $(\underline{\xi},\overline{\xi}) : \mathbb{Z}_+ \to \mathbb{R}^n\times\mathbb{R}^n$ such that the to-be-estimated parameter vector θ° lies in $[\underline{\xi}(t),\overline{\xi}(t)]$ for all t. Define the pair of vector-valued functions $(\underline{\theta},\overline{\theta}) : \mathbb{Z}_+ \to \mathbb{R}^n\times\mathbb{R}^n$ such that $\theta^\circ \in [\underline{\theta}(0),\overline{\theta}(0)]$ and for all $t\geq 1$,

$$\underline{\theta}(t) = \max\left(\underline{\theta}(t-1),\underline{\xi}(t)\right) \tag{38}$$

$$\overline{\theta}(t) = \min\left(\overline{\theta}(t-1),\overline{\xi}(t)\right), \tag{39}$$

where the minimum/maximum operators apply componentwise, i.e., when x and y are vectors of the same dimension, $\min(x,y)$ refers to the vector whose entries are given by $\min(x_i,y_i)$. We will call $(\underline{\xi},\overline{\xi})$ the input of the dynamic system (38)-(39) and $(\underline{\theta},\overline{\theta})$ its state. In fact (38)-(39) is equivalent to $[\underline{\theta}(t),\overline{\theta}(t)] = [\underline{\theta}(t-1),\overline{\theta}(t-1)]\cap[\underline{\xi}(t),\overline{\xi}(t)]$.
We now state some basic properties of the estimator (38)-(39).

Lemma 5. *Assume $\underline{\theta}(0)\leq\overline{\theta}(0)$ and $\underline{\xi}(t)\leq\overline{\xi}(t)$ for all t. Then the following facts are true:*

1. *Boundedness:* $\underline{\theta}(0)\leq\underline{\theta}(t)\leq\overline{\theta}(t)\leq\overline{\theta}(0)\ \forall t\geq 0$

2. *Monotonically decreasing widths:* $[\underline{\theta}(t), \overline{\theta}(t)] \subset [\underline{\theta}(k), \overline{\theta}(k)]$ $\forall (k,t)$ *such that* $k \leq t$.
3. *Convergence: The sequences* $\left\{\underline{\theta}(t)\right\}$ *and* $\left\{\overline{\theta}(t)\right\}$ *converge to* $\underline{\theta}^*$ *and* $\overline{\theta}^*$ *respectively with*

$$\underline{\theta}^* \triangleq \max\left(\underline{\theta}(0), \max_t \underline{\xi}(t)\right)$$
$$\overline{\theta}^* \triangleq \min\left(\overline{\theta}(0), \min_t \overline{\xi}(t)\right)$$

If $\max_t \underline{\xi}(t) \leq \underline{\theta}(0)$ *and* $\min_t \overline{\xi}(t) \leq \overline{\theta}(0)$*, then the input sequence* $\left\{\underline{\xi}(t), \overline{\xi}(t)\right\}$ *does not bring any information since in this case* $\underline{\theta}(t) = \underline{\theta}(0)$ *and* $\overline{\theta}(t) = \overline{\theta}(0)$ *for all* t.
4. *If* $\theta^\circ \in [\underline{\theta}(0), \overline{\theta}(0)] \cap [\underline{\xi}(t), \overline{\xi}(t)]$ *for all* t*, then* $\theta^\circ \in \bigcap_{t=0}^{\infty} [\underline{\theta}(t), \overline{\theta}(t)]$.

Proof. The facts 1, 2 and 4 are quite immediate. To see why fact 3 holds, note that $\left\{\underline{\theta}(t)\right\}$ is (componentwise) nondecreasing and upper-bounded while $\left\{\overline{\theta}(t)\right\}$ is nonincreasing and lower-bounded. Hence by the monotone convergence theorem, both sequences are convergent and their limits are the maximal element $\underline{\theta}^*$ and minimal element $\overline{\theta}^*$ of the respective sequences as expressed above. □

Remark 1. *In virtue of the properties stated in Lemma 5, the estimator in* (38)-(39) *is naturally robust to potential outliers in the sequence* $\{(\underline{v}(t), \overline{v}(t))\}$ *of bounds on the equation errors in* (1).

5. Application to a time-varying system

We now consider the case where the true parameter vector θ° in (1) is no longer constant but may be time varying with a limited rate of change. Let us pose

$$\theta^\circ(t) = \theta^\circ(t-1) + \delta(t), \tag{40}$$

where $\{\delta(t)\}$ is unknown but assumed to be bounded in an interval. More precisely, we assume that we know a sequence $\{\mathscr{I}(c_\delta(t), r_\delta(t))\}$ of intervals such that $\delta(t) \in \mathscr{I}(c_\delta(t), r_\delta(t))$ for all $t \in \mathbb{Z}_+$. Let us still use the notation $\tilde{\theta}(t)$ to refer to the parametric error now defined by $\tilde{\theta}(t) = \theta(t) - \theta^\circ(t)$ with $\theta(t)$ generated as in (9) from the data. It can then be shown that the error dynamics take the form

$$\tilde{\theta}(t) = A(t)\tilde{\theta}(t-1) + B(t)\overline{v}(t) \tag{41}$$

with $A(t) = I_n - q(t)x(t)^\top$ as in (12) and

$$\begin{aligned} B(t) &= [q(t) \quad -A(t)] \\ \overline{v}(t) &= [v(t) \quad \delta(t)^\top]^\top \end{aligned} \tag{42}$$

Note in passing that one recovers the error dynamics (12) from (41) when $\delta(t) = 0$ for all t, that is, when $\theta^\circ(t)$ is assumed constant. Now an interval representation of $\overline{v}(t)$ in (41) is given by

$$\begin{cases} c_{\overline{v}}(t) = [c_v(t) \quad c_\delta(t)]^\top \\ r_{\overline{v}}(t) = [r_v(t) \quad r_\delta(t)]^\top. \end{cases} \tag{43}$$

The relation (41) is key for deriving an interval-valued estimator. In effect, by relying on it and following the preceding discussions, it is easy to obtain, under the PE condition, an interval-valued estimator for the vector-valued sequence $\{\theta^\circ(t)\}$. More precisely, the complete form of the estimator is $\mathscr{I}(c'_\theta(t), r'_\theta(t)) = [c'_\theta(t) - r'_\theta(t), c'_\theta(t) + r'_\theta(t)]$, with center c'_θ defined by the state-space equation

$$c'_\theta(t) = A(t)c'_\theta(t-1) + q(t)(y(t) - c_v(t)) + A(t)c_\delta(t), \tag{44}$$

$c'_\theta(0) = c_\theta(0)$, and radius r'_θ given in convolution form by

$$r'_\theta(t) = |\Phi(t,0)|r_\theta(0) + \sum_{j=1}^{t} |\Phi(t,j)B(j)|r_{\overline{v}}(j). \tag{45}$$

Recall that in (44), $\{\theta(t)\}$ still refers to the sequence generated by the point-valued RLS identifier (9)-(11). Likewise, Φ is the RLS transition matrix expressed in (14). As to the truncated form of the estimator, it now admits the expression $\mathscr{I}(c'_\theta(t), \hat{r}'_{\theta,m}(t)) = [c'_\theta(t) - \hat{r}'_{\theta,m}(t), c'_\theta(t) + \hat{r}'_{\theta,m}(t)]$ with c'_θ as in (44) and $\hat{r}_{\theta,m}(t)$ defined by

$$\hat{r}'_{\theta,m}(t) = \begin{cases} |\Phi(t,0)|\, r_\theta(0) + \displaystyle\sum_{k=1}^{t} |\Phi(t,k)B(k)|r_{\overline{v}}(k), \\ \text{if } t = 0, \ldots, m \\ |\Phi(t,t-m)|\, \hat{r}'_{\theta,m}(t-m) + \displaystyle\sum_{k=t-m+1}^{t} |\Phi(t,k)B(k)|r_{\overline{v}}(k), \\ \text{if } t > m \end{cases} \tag{46}$$

Finally, let us remark that it is possible, similarly as in Section 4.3, to derive improved versions of the above interval-valued estimators for the case of time-varying systems. For this purpose, consider any pair $(\underline{\xi}, \overline{\xi})$ of functions such that $\underline{\xi}(t) \leq \overline{\xi}(t)$ and $\theta^\circ(t) \in [\underline{\xi}(t), \overline{\xi}(t)]$ for all $t \geq 0$. Then by letting $(\underline{p}, \overline{p})$ be defined by

$$\underline{p}(t) = \max\left(\underline{p}(t-1) + \underline{\delta}(t), \underline{\xi}(t)\right) \tag{47}$$
$$\overline{p}(t) = \min\left(\overline{p}(t-1) + \overline{\delta}(t), \overline{\xi}(t)\right), \tag{48}$$

with $\underline{\delta}(t) = c_\delta(t) - r_\delta(t)$ and $\overline{\delta}(t) = c_\delta(t) + r_\delta(t)$, it holds that $\theta^\circ(t) \in [\underline{p}(t), \overline{p}(t)]$ for all $t \geq 0$ provided that $\theta^\circ(0) \in [\underline{p}(0), \overline{p}(0)]$. Moreover, $(\underline{p}, \overline{p})$ is bounded provided that $(\underline{\xi}, \overline{\xi})$ is bounded. Of course the inputs $(\underline{\xi}, \overline{\xi})$ of (47)-(48) can be taken to be any of the estimates $(c'_\theta - r'_\theta, c'_\theta - r'_\theta)$ in (44)-(45) or $(c'_\theta - \hat{r}'_{\theta,m}, c'_\theta - \hat{r}'_{\theta,m})$ with $\hat{r}'_{\theta,m}$ defined in (46).

6. Some simulation results

6.1. Linear Time Invariant system

To illustrate the performance of the proposed estimators, we first consider a dynamical LTI system described by a model of the form (1) where $\theta^\circ = [-1.40 \quad 0.75 \quad 0.60 \quad -0.10]^\top \in \mathbb{R}^4$ and $x(t) = [-y(t-1) \quad -y(t-2) \quad u(t-1) \quad u(t-2)]^\top \in \mathbb{R}^4$ with the input $\{u(t)\}$ being generated as the realization of a zero-mean white Gaussian noise with unit variance. As to the noise

sequence $\{v(t)\}$, it is uniformly sampled from an interval of the form $[-a, a]$ with $a = 0.2$. In these conditions, we consider an estimation horizon of length $N = 200$ data points and compute the interval-valued parameter estimates described in (28). The initial parameter set $\mathscr{I}(c_\theta(0), r_\theta(0))$ is selected such that $c_\theta(0) = 0$ and $r_\theta(0) = \alpha_0 \mathbf{1}_n$ with $\alpha_0 = 4$ and $n = 4$ here and $\mathbf{1}_n$ being a n-dimensional vector of ones. The reference RLS algorithm (9)-(11) is run with initial value $\theta(0) = 0$, covariance matrix $P(0) = 10^3 I_4$ and forgetting factor $\lambda = 0.99$.

Evaluations of the preliminary estimators. Considering the truncated estimates (29)-(30), we start by recalling that, as established by Theorem 2, there is a minimum value of the horizon m beyond which boundedness of the estimate can be hoped for. With the experimental setting described above, a minimal such value is empirically found to be about 10 for most realizations of the input-output data.

Figure 1 below presents the interval-valued parameter estimates for this example when applying the estimators described (29)-(30) for $m \in \{20, 50, N\}$. Note that the case $m = N$ with N being the entire estimation horizon generated an interval radius $\hat{r}_{\theta,m}$ such that $\hat{r}_{\theta,m} = r_\theta$ (See Eq. (28)). The results confirm that the estimate $\mathscr{I}(c_\theta(t), r_\theta(t))$ defined in (28) is tighter than the truncated forms $\mathscr{I}(c_\theta(t), \hat{r}_{\theta,m}(t))$ for $m < N$. Moreover, the larger the truncation horizon m, the tighter $\mathscr{I}(c_\theta(t), \hat{r}_{\theta,m}(t))$.

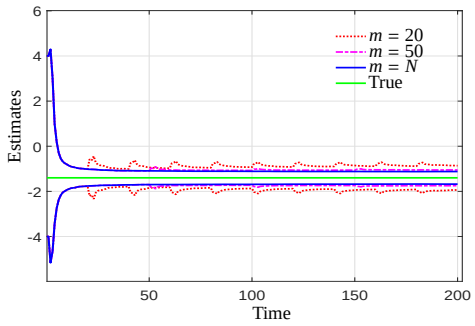

Figure 1: Interval-valued parameter estimates for the first entry of θ° (averaged over 100 independent runs). Truncated estimates $(c_\theta, \hat{r}_{\theta,m})$ with $\hat{r}_{\theta,m}$ as in (30) for $m = 20$ (dotted red), $m = 50$ (dashed magenta), $m = N$ (solid blue) and the true (constant) parameter θ° (solid green).

Monotonically improved estimators. To illustrate the benefit of the monotonic operators proposed in Section 4.3, we apply them to the three previous estimators, that is, the estimates are now computed by using $(c_\theta, \hat{r}_{\theta,m})$ as input in Eqs (38)-(39), with $m \in \{10, N\}$. Again recall that $\hat{r}_{\theta,N} = r_\theta$. The associated results are plotted in Figure 2. As argued before, we can see that the obtained estimates are smoother and tighter compared to those of Figure 1. Moreover, they effectively generate intervals with monotonically decreasing (nonincreasing) widths. For comparison purpose, we have also represented estimates[3] obtained by the method described in [7]. As it turns out, our estimator gives tighter estimates. We will see in the next paragraph that more tightness can be gained by using a smaller forgetting factor λ.

[3]Note that no line is visible in the time interval $[0, 20]$ because the first 20 samples are used here to initialize the algorithm.

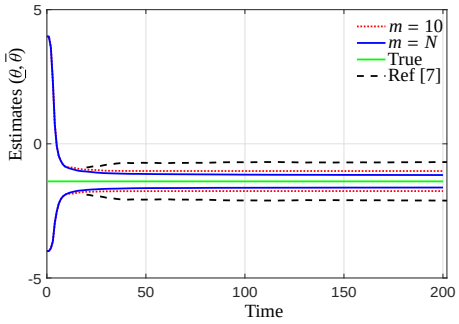

Figure 2: Interval-valued estimates (averaged over 100 independent runs) given by the estimator (38)-(39) with $\bar{\xi} = c_\theta + \hat{r}_{\theta,m}$, $\underline{\xi} = c_\theta - \hat{r}_{\theta,m}$ and $\hat{r}_{\theta,m}$ as in (30) for $m = 10$ (dotted red), $m = N$ (solid blue) : true (constant) parameter θ° (solid green) and estimates given by the method of [7] (dashed black).

Influence of the RLS forgetting factor. As can be intuitively guessed, the performance of the proposed interval-valued estimators depends on the properties of the RLS reference identifier which in turn are determined by the richness of the data and the user-defined parameters such as $P(0)$ and λ. In particular, it is interesting to study the impact of the forgetting factor $\lambda \in]0, 1]$. In general, for point-valued estimation, such a parameter is selected, to be close to 1 in order to smooth the trajectories of θ in (9). In contrast, the recursive interval-valued estimator (38)-(39) tends to perform better when λ is small. To see this consider Figure 3 where we have plotted the final interval width $\bar{\theta}(N) - \underline{\theta}(N)$ achieved by the estimator (38)-(39). Again only the estimates related to the first component of the parameter vector are represented. We consider the estimator (38)-(39) with $\underline{\xi} = c_\theta - \hat{r}_{\theta,m}$ and $\bar{\xi} = c_\theta + \hat{r}_{\theta,m}$ as defined in (28) and (30) for $m \in \{20, 50, N\}$. The results are indeed averages over 100 independent simulations. What this reveals is that the estimator's asymptotic performance depends on the forgetting factor in the sense that the width of the estimated interval is all the smaller as the forgetting factor λ is small. This behavior can be explained by the fact that a small λ in the RLS may cause the estimates $(c_\theta, \hat{r}_{\theta,m})$ to fluctuate substantially hence favoring the event that the associated interval jumps occasionally to a small value. We can further observe that for small values of the forgetting factor (e.g., $\lambda \leq 0.6$ in Figure 3), all truncation orders m tend to perform equally well. This suggests an important feature of the proposed estimation framework for practical implementation: provided the exciting input $\{u(t)\}$ is sufficiently rich and λ is then taken small enough, the computational complexity of the estimators can be reduced to the minimum by selecting a small truncation horizon m.

6.2. Linear Time Varying system

We now consider a model of the form (1) where the parameter vector θ° is time-varying with dynamics defined as in (40) where it is assumed that $\delta(t)$ belongs to an interval given by $c_\delta(t) = 0$ and $r_\delta(t) = [0.10 \quad 0.05 \quad 0.04 \quad 0.01]^\top$ for all t. For the simulation, we generate a sequence $\{\delta(t)\}$ in this interval such that $\delta(t) = r_\delta(t) \sin(2\pi t/30)$. The other settings remain the same as previously defined in the beginning of Section 6 except

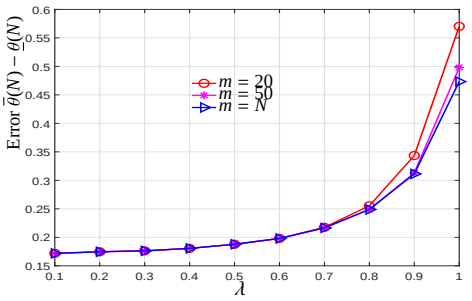

Figure 3: Widths $\bar{\theta}(N) - \underline{\theta}(N)$ (averaged over 100 independent runs) of the estimated parameter intervals versus forgetting factor λ. Only the first components of the parametric (vector-valued) error are represented for truncated estimators of the form (30) for $m = 20$ (dashed red), $m = 50$ (dashed magenta), $m = N$ (solid blue) and the true (constant) parameter θ° (solid green).

the forgetting factor which is now set to 0.1 (recall that as discussed earlier, the estimate is tighter when λ is small). Consider applying the estimator (47)-(48) with inputs $\underline{\xi} = c_\theta - \hat{r}_{\theta,m}$ and $\bar{\xi} = c_\theta + \hat{r}_{\theta,m}$ as defined in (28) and (30) for $m \in \{5, N\}$. The outcome of this experiment is depicted in Figure 4. For a value of λ as small as 0.1, the estimated interval appears to be very tight. Moreover, all values of the truncation horizon m give almost the same performance in this case.

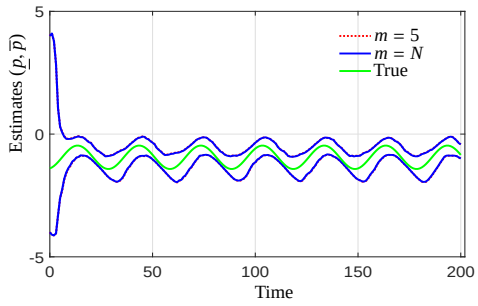

Figure 4: Interval-valued parameter estimates (averaged over 100 independent runs) given by the estimator (47)-(48) on the time-varying example with $\bar{\xi} = c'_\theta + \hat{r}'_{\theta,m}$, $\underline{\xi} = c'_\theta - \hat{r}'_{\theta,m}$ and $\hat{r}'_{\theta,m}$ as in (46) for $m = 5$ (dashed red) and $m = N$ (solid blue). The true time-varying parameter θ° is in solid green.

7. Conclusion

In this paper, we have presented a recursive interval-valued estimation framework for the identification of linearly parametrized models. The main idea of the method is to carefully bound the error generated by a certain reference adaptive algorithm, for example the recursive least squares. However the smallest interval-valued estimator we discussed turns out to be computationally costly to implement in an online scenario. We therefore turn to an alternative family of (over)-estimators which exhibits a trade-off between the achievable performance of and the price to pay for it in computational load. Two cases have been studied: one where the to-be-estimated parameter vector is constant and a more general situation where it is possibly time-varying. In the first case, we further show that the estimated interval size can be made monotonically decreasing. In the second, this monotonic property cannot be systematically achieved (as this depends on the change rate of the parameters) but the width of the estimated interval can be made very small by an appropriate design of the reference point-value identifier. For example, we have observed in simulation that when the reference identifier is the RLS algorithm, the performance of the estimator improves if the forgetting factor is small.

Future work may concern the extension of the proposed interval-valued estimation framework to systems whose models are nonlinear in the parameters.

References

[1] L. Bako. Adaptive identification of linear systems subject to gross errors. *Automatica*, 67:192–199, 2016.

[2] L. Bako and V. Andrieu. Interval-valued estimation for discrete-time linear systems: application to switched systems. https://arxiv.org/abs/1912.10770, 2018.

[3] L. Bako and V. Andrieu. Interval-valued state estimation for linear systems: the tightest estimator and its relaxations. *Automatica*, 106:168–177, 2019.

[4] S. Brüggemann and R. R. Bitmead. Exponential convergence of recursive least squares with forgetting factor for multiple-output systems. *Automatica*, 124:109389, 2021.

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

[6] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Dover Publications Inc., Mineola, New York, 2009.

[7] P.-O. Gutman. Online parameter interval estimation using recursive least squares. *International Journal of Adaptive Control and Signal Processing*, 8:61–72, 1994.

[8] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied interval analysis*. Springer, London, 2001.

[9] R. M. Johnstone, C. R. Johnson Jr, R. R. Bitmead, and B. D. Anderson. Exponential convergence of recursive least squares with exponential forgetting factor. *Systems & Control Letters*, 2:77–82, 1982.

[10] M. Kieffer and E. Walter. Interval analysis for guaranteed non-linear parameter and state estimation. *Mathematical and Computer Modelling of Dynamical Systems*, 11:171–181, 2005.

[11] M. Lauricella and L. Fagiano. Set membership identification of linear systems with guaranteed simulation accuracy. *IEEE Transactions on Automatic Control*, 65:5189–5204, 2020.

[12] M. Milanese, J. Norton, H. Piet-Lahanier, and E. W. (Eds). *Bounding approaches to system identification*. Springer Science & Business Media, 2013.

[13] M. Milanese and C. Novara. Unified set membership theory for identification, prediction and filtering of nonlinear systems. *Automatica*, 47:2141–2151, 2011.

[14] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. SIAM, 2009.

[15] X. F. Sun, H. Y. Zhang, and Y. Z. Fan. Recursive dual-linear-programming approach for parameter-uncertainty-interval estimation., 150(3), 303-310. *IEE Proceedings-Control Theory and Applications*, 150:303–310, 2003.

[16] G. Tao. *Adaptive control design and analysis*. John Wiley & Sons, 2003.

[17] E. van Kampen, Q. Chu, and J. Mulder. Interval analysis as a system identification tool. In *Advances in Aerospace Guidance, Navigation and Control, Springer, Berlin, Heidelberg*, pages 333–343, 2011.

[18] 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.

[19] E. Walter and H. Piet-Lahanier. Recursive robust minimax estimation for models linear in their parameters. *IFAC Proceedings Volumes*, 25:215–220, 1992.