A continuous-time framework for least squares parameter estimation

Samer Riachy

HAL Id: hal-01080376
https://inria.hal.science/hal-01080376

Submitted on 5 Nov 2014

A continuous-time framework
for least squares parameter estimation

Samer RIACHY[1,2]

[1]Ecole National Supérieure de l'Electronique et de ses Applications, ECS-Lab/EA-3649,
6, avenue du Ponceau, 95014, Cergy, France

[2] Non-A team, INRIA Lille-Nord-Europe, Lille, France. *riachy@ensea.fr*

Abstract

This paper proposes a continuous-time framework for the least-squares parameter estimation method through evolution equations. Nonlinear systems in the standard state space representation that are linear in the unknown, constant parameters are investigated. Two estimators are studied. The first one consists of a linear evolution equation while the second one consists of an impulsive linear evolution equation. The paper discusses some theoretical aspects related to the proposed estimators: uniqueness of a solution and an attractive equilibrium point which solves for the unknown parameters. A deterministic framework for the estimation under noisy measurements is proposed using a Sobolev space with negative index to model the noise. The noise can be of large magnitude. Concrete signals issued from an electronic device are used to discuss numerical aspects.

Keywords: Evolution equations, parameter estimation, least squares, Sobolev spaces, estimation under noise.

1 Introduction

Least squares (LS) is by far the most popular method for parameter estimation. It has been developed under different guises. Discrete and continuous-time approaches were proposed for discrete and continuous-time, linear and nonlinear systems [7, 8, 9, 10, 11, 15, 16]. The present work is concerned with continuous-time frameworks for LS. It seems that continuous-time LS estimators have been mainly developed in the context of adaptive control [7, 8, 9, 10, 15, 16]. In general, such estimators consist of a set of ordinary differential equations fed by the system input and output data. Major advantages of such estimators reside in the real time implementability and suitability to an adaptive control loop. Nonetheless, during its convergence, the estimator should be continuously fed by the system data. Consequently, depending on the convergence rate, the system data can be needed on a big interval of time[1] which may not be possible in some applications[2]. Typical examples are unstable plants or plants with restrictions on the state vector (linear actuators for example) when a certainty equivalence controller or a primary stabilizing controller are not available. Therefore, it is quite interesting to develop a continuous-time theoretical basis to account for LS estimation when the data is available on a bounded interval of time. The present work proposes such a continuous time framework. The approach relies on evolution equations (infinite dimension) and thus cannot be integrated in an adaptive control scheme. In order to highlight the practical utility of the approach, a discrete implementation is given. The computation cost is evaluated and shown to be carried out by low-cost real-time microcontrollers in order to accomplish fast parameter estimation for online plants.

An underlying problem to the design of a continuous-time estimator resides in the unavailability of the derivative of the state vector[3]. The standard solution [7, 8, 15] uses a filtered version of the system equations. The basics of the approach can be seen on the scalar example $y'(\tau) = \theta y(\tau)$ where θ denotes the unknown parameter. With $\check{y}(\lambda)$ to denote the Laplace transform of $y(\tau)$, one has $\frac{1}{\lambda+1}(\lambda \check{y} - y_0) = \theta \frac{1}{\lambda+1} \check{y}$. Then the algebraic equation $Y_1 - e^{-\tau} y_0 = Y_2 \theta$ is used in the LS estimator with $\check{Y}_1 = \frac{\lambda}{\lambda+1} \check{y}$ and $\check{Y}_2 = \frac{1}{\lambda+1} \check{y}$. A typical estimator is given by $\dot{\hat{\theta}} = \frac{Y_2(Y_1 - e^{-\tau} y_0 - Y_2 \hat{\theta})}{\sqrt{\kappa + Y_2^2}}$ where κ is a positive constant and the normalization $(\sqrt{\kappa + Y_2^2})^{-1}$ enhances the convergence rate of the estimator. A quite interesting approach that do not rely on estimating the state derivative is developed in [9]. An ingenious method [4, 5] based on the algebraic derivative concept permits to obtain an estimator that do not depend neither on the state derivative nor the initial condition. Let us illustrate the basics of this method on our scalar example. Differentiate

---

[1]This time interval is the whole positive real line in theory due to asymptotic convergence while in practice it can be considerably reduced by an efficient tuning of the estimator.

[2]The same reasoning applies also to recursive discrete-time least squares [11] which are not addressed in this work.

[3]Assuming that the state vector is measured.

$\lambda \check{y} - y_0 = \theta \check{y}$ once with respect to λ gives $\lambda^{-2} \left[\check{y} + \lambda \frac{d}{d\lambda} \check{y} - \theta \frac{d\check{y}}{d\lambda} \right] = 0$. Notice that y_0 has disappeared. In addition, the multiplication by λ^{-2} permit to obtain integrals in the time domain $\int_0^\tau (\tau - 2\tau_1) y d\tau_1 = \theta \int_0^\tau (\tau - \tau_1)(-\tau_1) y d\tau_1$. The integrals offer a low-pass filtering effect. Thus one can solve for θ whenever the right hand side is different from zero. The approach of [4, 5] has been specialized to noisy signal derivation in [12] and shown to admit a least squares interpretation. The method offers a systematic approach to annihilate initial conditions as well as structured perturbations. In order to deal with the state derivative in the present work, the computation technique summarized in the scalar example ([4, 5]) is applied prior to the estimator design.

The contribution of this work is twofold. First, LS estimators are developed which are suitable when the systems data are available on a bounded interval of time. Second, the proposed framework provides a simple deterministic description for the LS estimation under noisy measurements where the noise can be of large magnitude. Note that bounded unknown functions are usually used to model the noise in a deterministic context.

The paper is organized as follows. Two LS estimators, the first one based on a linear evolution equation, the second one based on an impulsive evolution equation are proposed and discussed in section 3. The deterministic framework for the estimation under noisy measurements is proposed in section 4 for linear input-output systems. Section 5 proposes a numerical implementation of the estimators. Section 6 is dedicated to an experimental validation in order to discuss the numerical aspect. Let us start in section 2 with a brief recall about evolution equations in order to clarify subsequent developments.

2 Elementary notions about evolution equations

The content of this section can be found in any introductory textbook to the theory of evolution equations (see for example [3, 13, 14, 17]). Let $\{S(t); t \geqslant 0\}$ be a family of linear operators defined on a Banach space $\mathcal{B}$. $S(t)$ is said to be a linear semigroup on $\mathcal{B}$ if $S(0) = I$ (I denotes the identity operator on $\mathcal{B}$) and $S(t_1 + t_2) = S(t_2)S(t_1) = S(t_1)S(t_2)$. It is said to be a semigroup of contractions if, moreover, $\|S(t)\| \leqslant 1$ where $\| \cdot \|$ is an operator norm on $\mathcal{B}$. In addition, the semigroup is strongly continuous if $\lim_{h \to +0} S(h)\xi = \xi$. The operator A defined by $A := \lim_{h \to +0} \frac{S(h)-I}{h}$ is called the infinitesimal generator of $S(t)$. The Hille-Yosida theorem provides necessary and sufficient conditions an operator A should satisfy in order to be a generator of a semigroup of contractions. Let us denote by $\mathcal{D}(A) \subset \mathcal{B}$ the domain of definition of an operator A.

Theorem 2.1 (Hille-Yosida [17]) *The linear operator $A : \mathcal{D}(A) \subset \mathcal{B} \mapsto \mathcal{B}$ is the infinitesimal generator of a linear semigroup of contractions if, and only if,*

1. *A is a densely defined ($\mathcal{D}(A)$ is dense in $\mathcal{B}$) and closed operator in $\mathcal{B}$,*

2. *$\forall \lambda > 0$, $\lambda I - A$ is a one-to-one and onto mapping,*

3. *$\|(\lambda I - A)^{-1}\| \leqslant \frac{1}{\lambda}$.*

A linear homogeneous evolution equation is a system given by ($\dot{\tilde{\Theta}} := \frac{d\tilde{\Theta}}{dt}$):

$$\dot{\tilde{\Theta}} = A\tilde{\Theta}; \quad \tilde{\Theta}_{t=0} = \tilde{\xi}, \ \tilde{\xi} \in \mathcal{D}(A) \tag{1}$$

where A is the infinitesimal generator of a strongly continuous semigroup. A (Banach space valued) function $\tilde{\Theta}(t) : [0, +\infty) \mapsto \mathcal{B}$ is said to be a solution of (1) if $\tilde{\Theta}(t) \in \mathfrak{C}^1([0, \infty); \mathcal{D}(A))$ such that (1) is satisfied. $\mathfrak{C}^1$ represents the space of continuously differentiable functions defined on $\mathcal{B}$. The existence of a unique solution for (1) is ensured by the following theorem[4].

Theorem 2.2 (Existence and uniqueness) *If $A : \mathcal{D}(A) \subset \mathcal{B} \mapsto \mathcal{B}$ generates a strongly continuous semigroup of contractions $S(t)$ on $\mathcal{B}$ then $\forall \tilde{\xi} \in \mathcal{D}(A)$, (1) admits a unique solution given by $\tilde{\Theta}(t) = S(t)\tilde{\xi}$.*

3 The estimators

3.1 Problem description and basic assumptions

Consider the nonlinear system which is linear in the parameter Θ:

$$x'(\tau) = \phi(x, u) + \varphi(x, u)\Theta \tag{2}$$

where x and u are scalar real variables while $\Theta = [\theta_1, \cdots, \theta_p]^T \in \mathbb{R}^p$ is the vector of unknown constant parameters. The scalar field $\phi(x, u) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ and the vector field $\varphi = [\varphi_1, \cdots, \varphi_p]$ with $\varphi_i(x, u) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$, $i = 1, \cdots, p$,

---

[4]Theorem 2.2 corresponds to proposition 6.2 page 145 of [13].

are known and satisfy standard assumptions about the existence of a unique classical solution for (2) for a given initial condition. A scalar system is considered in order to simplify the presentation, the extension to multidimensional systems can be done straightforwardly if the state vector is accessible for measurements. Let $\mathbb{T}$ be a bounded and connected subset of $\mathbb{R}$. Given a solution of (2) on $\mathbb{T}$, it is legitimate to identify $\phi(x(\tau), u(\tau))$ and $\varphi(x(\tau), u(\tau))$ with functions depending on τ only. Thus the notations $\phi(\tau) := \phi(x(\tau), u(\tau))$ and $\varphi_i(\tau) := \varphi_i(x(\tau), u(\tau))$, $\tau \in \mathbb{T}$ are adopted. Moreover, the following is assumed.

Assumption 3.1 $u(\tau)$, $\phi(\tau)$ and $\varphi_i(\tau) \in \mathcal{L}^2(\mathbb{T})$. $\mathcal{L}^2(\mathbb{T})$ is the space of square summable functions on $\mathbb{T}$.

Let ϵ be a positive constant and denote by $\mathbb{D}$ the set $\mathbb{D} = [-\epsilon, \epsilon]$. In order to deal with the derivative x', let $v(s)$, $s \in \mathbb{D}$, be a continuously differentiable function, supported on $\mathbb{D}$ such that $v(-\epsilon) = v(\epsilon) = 0$. Introduce $\mathbb{T}_\epsilon$, an ϵ-neighborhood of $\mathbb{T}$ such that $\mathbb{T} \subseteq \mathbb{T}_\epsilon \subset \mathbb{R}$ and $T_\epsilon = T_1 + 2\epsilon$. T_ϵ, T_1 and 2ϵ are the Lebesgue measures of $\mathbb{T}_\epsilon$, $\mathbb{T}$ and $\mathbb{D}$ respectively. Define the convolutions $x'_\epsilon = v \star x'$, $\phi_\epsilon = v \star \phi$ and $\varphi_{\epsilon,i} = v \star \varphi_i$. They are supported on $\mathbb{T}_\epsilon$ since x', ϕ and φ are supported on $\mathbb{T}$. Moreover, notice that $-\frac{dv}{ds} \star x = v \star x'$. Such a $v(s)$ can be given by

$$v(s) = (\epsilon^2 - s^2)\chi_{\mathbb{D}} \tag{3}$$

where $\chi_{\mathbb{D}}$ is the indicator function of $\mathbb{D}$. Convolving (2) with (3) leads to:

$$x'_\epsilon(\tau) = \phi_\epsilon(\tau) + \varphi_\epsilon(\tau)\Theta, \quad \tau \in \mathbb{T}_\epsilon. \tag{4}$$

Remark 3.1 *Equality $v \star x' = -\frac{dv}{ds} \star x$ can be linked to the notion of weak derivative (see [1, 3]). Moreover, x'_ϵ, ϕ_ϵ and $\varphi_{\epsilon,i}$ can be seen as mollified versions of x', ϕ and φ_i. Then, $v(s)$ can be called a mollifier (For more details, see [1] pages 36 and 66 and [3] page 629). In addition, $v(\epsilon s)$ given by (3) appears in Jacobi orthogonal polynomials and $v(2\epsilon s - \epsilon)$ appears in [12].*

In the sequel we use $\mathbb{T}_\epsilon = [0, T_\epsilon]$ and $\mathbb{T} = [\epsilon, T_\epsilon - \epsilon]$. The estimators are developed based on (4) which is supposed to satisfy the following assumption.

Assumption 3.2 (Persistence of excitation) *The sets*

- $\{\tau; \ \varphi_\epsilon(\tau)(\Psi - \Theta) = 0, \Psi = (\psi_1, \cdots, \psi_p)^T \in \mathbb{R}^p,$
  $\Psi \neq \Theta, \tau \in \mathbb{T}_\epsilon\}$

- $\{\tau; \varphi_{\epsilon,i}(\tau) = 0, i = 1 \cdots p, \tau \in \mathbb{T}_\epsilon\},$

have a zero Lebesgue measure ($\psi_1, \cdots, \psi_p$ are constants).

The first item of assumption 3.2 ensures the uniqueness of Θ up to sets of zero Lebesgue measure. An estimator based on a linear homogeneous evolution equation is first presented. Then an estimator based on an impulsive evolution equation is proposed in section 3.3.

3.2 LS estimator based on a linear evolution equation

Let

$$A = A_1 + A_2, \tag{5}$$

where $A_1 = -\varphi_\epsilon^T \varphi_\epsilon$, $A_2 = cI\frac{\partial^2}{\partial \tau^2}$, $B = -\varphi_\epsilon^T(\phi_\epsilon - x'_\epsilon)$, c denotes a strictly positive constant and I denotes the identity matrix. The domain of definition of A is given by:

$$\mathcal{D}(A) = \{\hat{\Theta} \in \left[\mathcal{C}^2(\mathbb{T}_\epsilon)\right]^p; \hat{\Theta}'(0) = \hat{\Theta}'(T_\epsilon) = 0, \tau \in \mathbb{T}_\epsilon\}. \tag{6}$$

Introduce the system

$$\dot{\hat{\Theta}} = A\hat{\Theta} + B, \quad \hat{\Theta}_{t=0} = \hat{\xi} \tag{7}$$

where $\dot{\hat{\Theta}} := \frac{d\hat{\Theta}}{dt}$. With $\tilde{\Theta} = \hat{\Theta} - \Theta$, equation (7) transforms to

$$\dot{\tilde{\Theta}} = A\tilde{\Theta}, \quad \tilde{\Theta}_{t=0} = \tilde{\xi}. \tag{8}$$

Remark 3.1 *The right-hand-side of (7) corresponds to the minimum of the quadratic cost $\int_{\mathbb{T}_\epsilon}[(\varphi_\epsilon\hat{\Theta} + \phi_\epsilon - x'_\epsilon)^2 + c\hat{\Theta}'^2]d\tau$ according to Euler-Lagrange equations.*

Remark 3.2 *A similar cost, involving a Taylor expansion, is frequently used in image processing for the optical flow estimation problem [2, 6]. In contrast to our context, the unknown parameters are not constant (and can be discontinuous is some cases), however, the applicability of the method is motivated by heuristics and do not rely on a theoretical basis.*

The following theorem describes the estimator (8).

Theorem 3.1 *The operator A defined in (5), (6) generates a strongly continuous semigroup of contractions. Consequently, (8) admits a unique solution for a given initial condition. Moreover, under assumptions 3.2, the equilibrium $\tilde{\Theta} = 0$ is unique and attractive.*

proof 3.1 *Notice first that $\mathcal{D}(A)$ given by (6) is dense in $\mathcal{L}^2(T_\epsilon)$. Rewrite (8) using the Laplace transform with the Laplace variable $\lambda \in \mathbb{C}$:*

$$\lambda \check{\Theta} - \tilde{\xi} = A_1 \check{\Theta} + c\check{\Theta}'', \quad \tilde{\Theta}'(0) = \tilde{\Theta}'(T_\epsilon) = 0. \tag{9}$$

For each $\lambda > 0$ (a positive real number), equation (9) is a linear boundary value problem (with $\check{\Theta}$ and τ the dependent and independent variables respectively) admitting a unique solution whenever $\varphi_{\epsilon,i}$ are summable (the summability is insured by assumption 3.1). Consequently, the operator $\lambda I - A$ is invertible with $\check{\Theta} = (\lambda I - A)^{-1}\tilde{\xi}$.

Multiply (9) with $\check{\Theta}^T$ and integrate over $\mathbb{T}_\epsilon$ leads to $\int_0^{T_\epsilon} \check{\Theta}^T(\lambda I - A_1)\check{\Theta} d\tau - c\int_0^{T_\epsilon} \check{\Theta}^T\check{\Theta}'' d\tau = \int_0^{T_\epsilon} \check{\Theta}^T\tilde{\xi} d\tau$. Note that the spectrum of A_1 is given by $\{0, -\sum_{i=1}^{p} \varphi_{\epsilon,i}^2\}$. Next we use the identity $e_1|\check{\Theta}| < \check{\Theta}^T(-A_1)\check{\Theta} \leqslant e_2|\check{\Theta}|$ where e_1 and e_2 are the smallest and largest eigenvalues of $-A_1$ and $|\cdot|$ denotes the Euclidean norm. Integrating by parts on the left hand side and using Cauchy-Schwartz inequality on the right hand side leads to $\lambda\|\check{\Theta}\|^2 + c\|\check{\Theta}'\|^2 \leqslant \|\check{\Theta}\|\|\tilde{\xi}\|$, consequently, $\|\check{\Theta}\| \left(\lambda \|\check{\Theta}\| - \|\tilde{\xi}\|\right) \leqslant -c\|\check{\Theta}'\| \leqslant 0$ then $\lambda \|\check{\Theta}\| \leqslant \|\tilde{\xi}\|$. Using $\check{\Theta} = (\lambda I - A)^{-1}\tilde{\xi}$ one obtains $\lambda \|(\lambda I - A)^{-1}\tilde{\xi}\| \leqslant \|\tilde{\xi}\|$ from which one deduces that $\|(\lambda I - A)^{-1}\| \leqslant \frac{1}{\lambda}$. Then, by theorem 2.1, A generates a strongly continuous semigroup of contractions and, by theorem 2.2, (8) admits a unique solution for a given $\tilde{\xi} \in \mathcal{D}(A)$.

Consider the functional $V = \frac{1}{2}\int_0^{T_\epsilon} \tilde{\Theta}^T\tilde{\Theta} d\tau > 0$. Its derivative along the trajectories of (8) gives $\dot{V} = -\int_0^{T_\epsilon} \tilde{\Theta}^T\varphi_\epsilon^T\varphi_\epsilon\tilde{\Theta} d\tau + c\int_0^{T_\epsilon} \tilde{\Theta}^T\tilde{\Theta}'' d\tau$. After integration by parts one obtains $\dot{V} = -\int_0^{T_\epsilon}(\varphi_\epsilon\tilde{\Theta})^2 d\tau - c\int_0^{T_\epsilon} \|\tilde{\Theta}'\|^2 d\tau$. $\dot{V}$ is negative definite whenever $\tilde{\Theta}$ is different from a constant. Then by assumption 3.2, $\tilde{\Theta} = 0$ is the only equilibrium of (7) and it is attractive. This ends the proof. □

3.3 LS estimator based on an impulsive evolution equation

Let δ be a strictly positive constant and introduce the following impulsive system

$$\begin{cases} \dot{\tilde{\Theta}} & = & A_1\tilde{\Theta}, \quad n\delta < t < (n+1)\delta \\ \tilde{\Theta}^+ & = & \frac{1}{a(\tau) + b(\tau)} \int_{-a(\tau)}^{b(\tau)} \tilde{\Theta}(t, \tau + s)ds, \, t = n\delta \end{cases} \tag{10}$$

where $n \in \mathbb{N}$, $a(\tau) = \min(\bar{a}, \tau)$, $b(\tau) = \min(\bar{a}, T_\epsilon - \tau)$ and $\bar{a} \leqslant T_\epsilon$ is a strictly positive constant. $A_1 : [\mathcal{L}^2(\mathbb{T}_\epsilon)]^p \mapsto [\mathcal{L}^2(\mathbb{T}_\epsilon)]^p$ is a closed operator defined on the whole of $[\mathcal{L}^2(\mathbb{T}_\epsilon)]^p$. Its spectrum is given by $\{0, -\sum_{i=1}^{p} \varphi_{\epsilon,i}^2\}$. It generates thus a strongly continuous semigroup according to theorem 2.1 . System (10) admits a unique solution (in forward time) satisfying $\tilde{\Theta}_{t=0} = \tilde{\xi}$ since, on $n\delta < t < (n+1)\delta$, the first equation of (10) admits, by theorem 2.2, a unique solution. Moreover, on $t = n\delta$, the second equation of (10) permits to uniquely determine $\tilde{\Theta}^+$ from $\tilde{\Theta}$. The following theorem describes the estimator (10).

Theorem 3.1 *Under assumption 3.2, $\tilde{\Theta} = 0$ is the unique equilibrium of (10) and it is attractive.*

proof 3.2 *Consider the functional $V = \frac{1}{2}\tilde{\Theta}^T\tilde{\Theta}$. Its derivative along the trajectories of (10) gives $\dot{V} = -(\varphi_\epsilon\tilde{\Theta})^2$. $\dot{V}$ is negative definite whenever $\varphi_\epsilon\tilde{\Theta} \neq 0$. On the other hand, notice that if $\tilde{\Theta}$ is not constant, the strict norm inequality $\|\tilde{\Theta}^+\| < \|\Theta\|$ is satisfied at $t = n\delta$. Consequently, at $t = n\delta$, the quadratic functional $V^+ = \frac{1}{2}\tilde{\Theta}^{+T}\tilde{\Theta}^+$ satisfies $V^+ < V$. On the other hand, assumption 3.2 states that $\tilde{\Theta}$ being constant is possible for $\tilde{\Theta} = 0$ only. This ends the proof.*

Remark 3.3 *The diffusion term $\tilde{\Theta}''$ in (8) and the second equation of (10) both offer a smoothing effect. The smoothing is more pronounced with (10) especially for $\bar{a} = T_\epsilon$. This fact has a direct influence on the convergence rate of the estimator as it is confirmed in the experimentation section 5.*

4 Linear input-output systems under noise

Setting and identifying the parameters of a linearized model (around an operating point) of a nonlinear system often arises in practice. We will be interested by the parameter estimation in linear systems since these systems offer a simple framework to introduce a deterministic representation of *big magnitude noise* affecting the input and output. A noise with big magnitude may be encountered in power electronics, telecommunications etc. With the aim of setting a deterministic representation of such a noise, we introduce the Sobolev space with negative index $\mathcal{W}^{-1,2}(\mathbb{T})$ which is the dual

space of $\mathcal{W}^{1,2}(\mathbb{T})$, defined by $\mathcal{W}^{1,2}(\mathbb{T}) = \{\omega \in \mathcal{L}^2(\mathbb{T}), D\omega \in \mathcal{L}^2(\mathbb{T}); \|\omega\| + \|D\omega\| < +\infty\}$ where $D\omega$ is the weak derivative of ω. $\mathcal{W}^{-1,2}(\mathbb{T})$ is a Banach space under the norm (see [1]) $\|\varpi\|_{\mathcal{W}^{-1,2}} = \sup_{\omega \in \mathcal{W}^{1,2}(\mathbb{T}), \|\omega\|_{\mathcal{W}^{1,2}} \leqslant 1} |\langle \omega, \varpi \rangle|$ where $\|\omega\|_{\mathcal{W}^{1,2}} = (\|w\|^2 + \|Dw\|^2)^{\frac{1}{2}}$. Consider $\mathcal{V}^{-1,2}(\mathbb{T})$, a subset of $\mathcal{W}^{-1,2}(\mathbb{T})$, given by $\mathcal{V}^{-1,2}(\mathbb{T}) = \{\varpi \in \mathcal{W}^{-1,2}(\mathbb{T}); \|\varpi\|_{\mathcal{W}^{-1,2}} < N\}$ where N is a positive constant. Roughly speaking, $\mathcal{V}^{-1,2}$ contains (but is not restricted to) large magnitude, rapidly oscillating functions such that their integral is less than N. Consider now the linear input output system

$$x^{(q)} = \sum_{i=0}^{q-1} \theta_i x^{(i)} + \theta_q u \tag{11}$$

which can be written in the compact form $X\Theta = x^{(q)}$; $X = [x, \cdots, x^{(q-1)}, u]$, $\Theta = [\theta_0, \cdots, \theta_q]^T$ and $|\theta_i| > 0$. Let y be a noisy measurement of x such that $y = x + \varpi$ where $\varpi \in \mathcal{V}^{-1,2}(\mathbb{T})$. With

$$\rho(s) = (\epsilon^2 - s^2)^q \chi_{\mathbb{D}}, \tag{12}$$

define[5] $y_{\epsilon,i} = \rho \star D^{(i)} y$ and notice that $y_{\epsilon,i} = (-1)^i \rho^{(i)} \star y$ and that $y_{\epsilon,i}$, $i = 0, \cdots, q$, are supported on $\mathbb{T}_\epsilon$. Since (11) is linear and for $u \in \mathcal{L}^2(\mathbb{T})$, its solution x is square summable (it belongs in fact to the Sobolev space $\mathcal{W}^{q,2}(\mathbb{T})$ since $\|x\|_{\mathcal{W}^{q,2}(\mathbb{T})} = (\sum_{i=0}^q \|D^{(i)} x\|^2)^{\frac{1}{2}} < +\infty)$. Let Y_ϵ be given by $Y_\epsilon = [y_\epsilon, \cdots, y_{\epsilon,(q-1)}, u_\epsilon]$, $u_\epsilon = \rho \star u$, $Y_\epsilon = X_\epsilon + W_\epsilon$, $W_\epsilon = [\cdots, (-1)^i \rho^{(i)} \star \varpi, \cdots, 0]$ and $X_\epsilon = [x_\epsilon, \cdots, u_\epsilon]$.

For simplicity sake, take $\bar{a} = T_\epsilon$ (see Remark 3.3) and set the impulsive evolution equation (similar to (10)):

$$\begin{cases} \dot{\hat{\Theta}} & = & -Y_\epsilon^T (Y_\epsilon \hat{\Theta} - y_\epsilon^{(q)}), \ n\delta < t < (n+1)\delta \\ \hat{\Theta}^+ & = & \frac{1}{T_\epsilon} \int_0^{T_\epsilon} \hat{\Theta}(t, \tau + s) ds, \ t = n\delta. \end{cases} \tag{13}$$

System (13) can be transformed to

$$\begin{cases} \dot{\tilde{\Theta}} & = & -Y_\epsilon^T (Y_\epsilon \tilde{\Theta} + G_\epsilon), \ n\delta < t < (n+1)\delta \\ \tilde{\Theta}^+ & = & \frac{1}{T_\epsilon} \int_0^{T_\epsilon} \tilde{\Theta}(t, \tau + s) ds, \ t = n\delta \end{cases} \tag{14}$$

with $G_\epsilon = W_\epsilon \Theta - \varpi_{\epsilon,q}$. Notice that $\varpi_{\epsilon,i} = \int_{-\epsilon}^{\epsilon} \frac{d^i \rho}{ds^i} \varpi ds$ is bounded since $\int \varpi$ is bounded by N and $\frac{d^i \rho}{ds^i}$ is polynomial. Thus, there exist two constants k_1 and k_2 such that $|G_\epsilon| < KN$ with $K = k_1|\Theta| + k_2$. The following assumption is made.

Assumption 4.1 (Persistence of excitation) *With $\bar{x}_{\epsilon,i} = \frac{1}{T_\epsilon} \int_0^{T_\epsilon} x_{\epsilon,i} d\tau$ and $\eta_{\epsilon,i} = x_{\epsilon,i} - \bar{x}_{\epsilon,i}$, the positive definite matrix*

$$I = \frac{1}{T_\epsilon} \int_0^{T_\epsilon} \begin{bmatrix} \eta_{\epsilon,i}^2 & \eta_\epsilon \eta_{\epsilon,1} & \cdots & \eta_\epsilon \eta_{\epsilon,p} \\ \eta_\epsilon \eta_{\epsilon,1} & \eta_{\epsilon,2}^2 & \cdots & \eta_{\epsilon,1} \eta_{\epsilon,p} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{\epsilon,p} \eta_\epsilon & \eta_{\epsilon,p} \eta_{\epsilon,1} & \cdots & \eta_{\epsilon,p}^2 \end{bmatrix} d\tau$$

has its smaller eigenvalue bigger than the positive constant L.

Assumption 4.1 ensures, in particular, that $|X_\epsilon \tilde{\Theta}^+|^2 \geqslant L|\tilde{\Theta}^+|^2$ at each application of the second equation of (14).

Theorem 4.1 *Under assumption 4.1 with $L > (KN)^2$, the ball centered at zero, with radius $\frac{kN}{\sqrt{L}-kN}$ is attractive for the system (14).*

proof 4.1 *Consider the functional $V = \frac{1}{2}\tilde{\Theta}^T \tilde{\Theta}$. Its derivative is given by $\dot{V} = -(Y_\epsilon \tilde{\Theta})^2 - G_\epsilon Y_\epsilon \tilde{\Theta} \leqslant -|Y_\epsilon \tilde{\Theta}|(|Y_\epsilon \tilde{\Theta}| - KN)$. According to assumption 4.1 one has at $n = t\delta$ (i.e at each application of the second equation of (14)) that $|X_\epsilon \tilde{\Theta}^+| \geqslant \sqrt{L}|\tilde{\Theta}^+|$. Consequently, $|Y_\epsilon \tilde{\Theta}| \geqslant (\sqrt{L} - KN)|\tilde{\Theta}^+|$. Then $\dot{V}^+ \leqslant -(\sqrt{L} - KN)|\tilde{\Theta}^+|[(\sqrt{L} - KN)|\tilde{\Theta}^+| - KN]$ which ends the proof.*

5 Numerical implementation

In practice, the systems data are collected in a discrete form. Denote by d the sampling time such that $T_1 = md$, $m \in \mathbb{N}$. Let $\vec{y}$ and $\vec{u}$ be discrete versions of y and u ($\vec{y}[j+1] = y(jd)$, $\vec{u}[j+1] = u(jd)$, $j = 0, \cdots, m$). Moreover, let $\vec{\rho}_i$ be a discrete version of $\rho^{(i)}$ ($\vec{\rho}_i[j] = \rho^{(i)}(jd)$, $j = -l, \cdots, l$, $l \in \mathbb{N}$, $ld = \epsilon$). The convolutions are then approximated by $\rho^{(i)} \star y \approx \vec{\rho}_i \ast \vec{y}$ where $\ast$ denotes the discrete convolution. Prior to the application of the estimators, $\vec{y}_{\epsilon,i}$, $i = 0 \cdots q - 1$ and u_ϵ are generated by $\vec{y}_{\epsilon,i} = \vec{\rho}_i \ast \vec{y}$ and $\vec{u}_\epsilon = \vec{\rho}_0 \ast \vec{u}$. Notice that $\vec{y}_{\epsilon,i}$ and $\vec{u}_\epsilon$ consist of $m + 2l - 1$ elements.

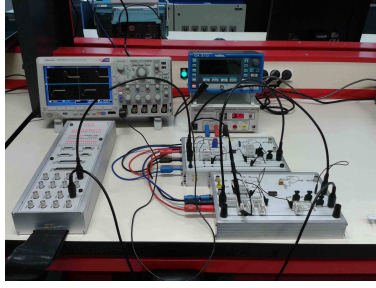[5]The weak derivative $D^{(i)} y$ is used since ϖ can be weakly differentiable only.

Figure 1: The experimental system

5.1 Implementation of estimator (7)

It can be implemented using standard finite elements or finite volume methods. We use here the basic forward Euler method. It leads to the recursive system

$$\vec{\tilde{\Theta}}_{k+1} = [I + \delta_1(\vec{A}_1 + \frac{c}{d^2}C)]\vec{\tilde{\Theta}}_k + \delta_1\vec{B} \tag{15}$$

where C is a square matrix which permit to approximate $\tilde{\Theta}''$. $C(i,i) = 2$, $C(i, i-1) = -1$, $C(i, i+1) = -1$, for $i = 2, \cdots, m-1$, $C(1,1) = C(m,m) = 1$ and $C(m, m-1) = -1$. All remaining entries of C are equal to zero. δ_1 is a positive constant, chosen such that the matrix $I + \delta_1(\vec{A}_1 + \frac{c}{d^2}C)$ has its eigenvalues inside the unit circle.

5.2 Implementation of estimator (13)

We consider only the case where $\bar{a} = T_\epsilon$ (The implementation of estimator (10) is quite similar). The estimator (13) is implemented by

$$\vec{\tilde{\Theta}}_{k+1} = \begin{cases} \vec{\tilde{\Theta}}_k - \delta_1\vec{Y}_\epsilon^T(\vec{Y}_\epsilon\vec{\tilde{\Theta}}_k + \vec{y}_{\epsilon,q}), & \text{if } \mathrm{mod}(k,r) \neq 0 \\ \mathrm{mean}(\vec{\tilde{\Theta}}_k), & \text{if } \mathrm{mod}(k,r) = 0, \end{cases} \tag{16}$$

with $r \in \mathbb{N}$, $r > 1$ and δ_1 is chosen in order to insure the asymptotic stability of the first equation of (16). That is the eigenvalues of $I - \delta_1\vec{Y}_\epsilon^T\vec{Y}_\epsilon$ lies within the unit circle.

6 An experimental validation

The linear system depicted, in figure 1, consists of four stages of first order low pass RC-circuits. With state variables taken as the capacitors voltage, the state space representation is given by:

$$x_1' = -2\omega_1 x_1 + \omega_1 x_2 + \omega_1 u; \ x_2' = \omega_1 x_1 - 2\omega_1 x_2 + \omega_1 x_3$$
$$x_3' = \omega_3 x_2 - 2\omega_3 x_3 + \omega_3 x_4; \ x_4' = \iota x_3 - \iota x_4 \tag{17}$$

where $\omega_1 = \omega_2 = \omega_3 = \dfrac{1}{RC} = \omega$, $\iota = \dfrac{1}{RC_2}$, $R = 100\mathrm{K}\Omega$, $C = 430\mathrm{nF}$ and $C_2 = 100\mathrm{nF}$. The input-state data are collected with a sampling time $d = 0.01$ second.

Let us use a standard LS algorithm from the literature [7, 8, 15] and estimate ω from the third equation of (17). The objective is to illustrate its operating principle. The estimator is given by

$$\hat{\omega}' = \frac{-(\breve{x}_2 - 2\breve{x}_3 + \breve{x}_4)[(\breve{x}_2 - 2\breve{x}_3 + \breve{x}_4)\hat{\omega} + \breve{x}_3']}{\sqrt{\kappa + (\breve{x}_2 - 2\breve{x}_3 + \breve{x}_4)^2}} \tag{18}$$

where $\breve{x}_2$, $\breve{x}_3$, $\breve{x}_4$ and $\breve{x}_3'$ are are low-pass filtered versions of x_2, x_3, x_4 and x_3' respectively. The experimental result is depicted on figure 2. It appears that the estimator converges towards the true parameter value ($\omega = 23.26$) as time goes on. Next we test our estimators, the estimators are normalized in order to increase the convergence rate.

Scalar estimators for ω and ι: The parameters ω and ι are estimated from the third and fourth equations of (17) respectively. Let us start by the estimation of ω. With $\hat{\Theta} = \hat{\omega}$, $A_1 = (x_{2,\epsilon} - 2x_{3,\epsilon} + x_{4,\epsilon})^2$, $B = (x_{2,\epsilon} - 2x_{3,\epsilon} + x_{4,\epsilon})x_{3,\epsilon}'$ where $x_{2,\epsilon}$, $x_{3,\epsilon}$, $x_{4,\epsilon}$ and $x_{3,\epsilon}'$ result from the convolution of x_2, x_3, x_4 and x_3' with $v(s)$ respectively. The estimators (7) and (14) are implemented. $v(s)$ is discretized using 31 samples while 30 samples of the system data are used corresponding to a step response. This number of sample corresponds to $30 \times d = 0.3$ second (compare with figure 2). Thus $30 + 2 \times 31 - 1 = 91$ samples are used in the iterative systems (15) and (16). A reliable estimate ($\hat{\omega} \approx 24$) is
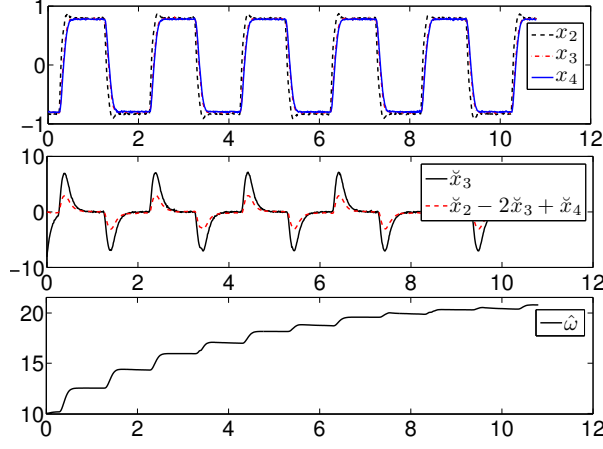
Figure 2: Estimation of ω from (18) versus time

obtained from (15) (resp. (16)) after 10 (resp. 4) iterations. The same scenario is used in the estimation of ι from the fourth equation of (17). Roughly, 10^4 multiplication operations are needed in the estimation procedure.

Two dimensional estimator: Let us slightly increase the complexity by using a second order estimator. The estimator is constituted with $\varphi_{1,\epsilon} = x_{1,\epsilon} - 4x_{2,\epsilon} + 5x_{3,\epsilon} - 2x_{4,\epsilon}$, $\varphi_{2,\epsilon} = x_{3,\epsilon} - x_{4,\epsilon}$, $\varphi_\epsilon = [\varphi_{1,\epsilon} \ \varphi_{2,\epsilon}]$, $A_1 = \varphi_\epsilon^T \varphi_\epsilon$, $B = \varphi_\epsilon x_{3,\epsilon}''$ and $\hat{\Theta} = [\hat{\theta}_1, \ \hat{\theta}_2]^T$ with $\theta_1 = \omega^2$ and $\theta_2 = \omega\iota$. The components of φ_ϵ are obtained by convolution with $\rho(s)$ ((12) with $p = 2$). 50 samples are used for the discretization of $\rho(s)$ and 50 samples of the systems data are used (0.5 second) corresponding to a step response. 40 iterations are needed in order to obtain a reliable estimate from (16) ($\hat{\theta}_1 \approx 510$, $\hat{\theta}_2 \approx 2150$). The estimation needs 90×10^4 multiplication operations.

discussions: The computation needed for both estimators can be realized in less than 0.1 second on a FPGA Xilinx Spartan-6 which is a low-cost microcontroller. The computation time is less than the time needed to collect the data (0.3 second and 0.5 second for previous tests). Consequently, the proposed estimator can be used without difficulty with "slow" systems which is the case of many industrial systems.

7   Conclusion and future works

A continuous-time approach for LS estimation is proposed through evolution equations. The developed framework provides estimators for systems whose data are available on a bounded interval of time. In addition, the measurements can be corrupted by a large magnitude noise.

Extending this approach to the reconstruction of time varying (maybe discontinuous) parameters is an interesting future research direction.

References

[1] R.A. Adams and J.F. Fournier, "Sobolev spaces" 2nd ed. *Academic press*, 2003.

[2] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black and R Szeliski. "A Database and Evaluation Methodology for Optical Flow." *International Journal of Computer Vision* No. 92, pp. 1-31, 2011.

[3] L.C. Evans (1997) "Partial Differential Equations" *AMS, Graduate studies in mathematics.* (chapter 7)

[4] M. Fliess, and H. Sira-Ramirez, (2003) "An algebraic framework for linear identification". *ESAIM: Contr. opt. calc. variat.*, Vol. 9, pp. 151-168.

[5] M. Fliess, C. Join, and H. Sira-Ramirez, (2008) "Nonlinear estimation is easy". *International Journal of Modelling, Identification and Control*, Vol. 4, No. 1, pp. 12-27.

[6] B.K.P. Horn and B.G. Schunck. "Determining optical flow." *Artificial Intelligence* vol. 17, pp. 185–203, 1981.

[7] P. Ioannou, and B. Fidan (2006) "Adaptive control tutorial" *SIAM.* (chapter 2)

[8] M. Krstic, I. Kanellakopoulos, and P. Kokotović, (1995) "Nonlinear and adaptive control design" *New York: Wiley.* (chapters 5 and 6)

[9] M. Krstic, (2009) "On using least-squares updates without regressor filtering in identification and adaptive control of nonlinear systems". *Automatica*, Vol. 45, pp. 731-735.

[10] M. Krstic, and P. Kokotović, (1995) "Adaptive nonlinear design with controller-identifier separation and swapping". *IEEE TAC*, Vol. 40, No. 3, pp. 426-440.

[11] L. Ljung, "System identification theory for the user", 2nd edition, *Prentice Hall*, 1999.

[12] M. Mboup, C. Join, and M. Fliess, (2009) "Numerical differentiation with annihilators in noisy environment". *Numerical algorithms*, Vol. 50, pp. 439-467.

[13] K-J Engel and R. Nagel, "One-parameter semigroups for linear evolution equations" *Springer-Verlag*, 2000.

[14] A. Pazy, "Semigroups of Linear Operators and Applications to Partial Differential Equations" *Springer-Verlag*, 1983.

[15] S. Sastry, and M. Bodson, (1989) "Adaptive control: stability, convergence and robustness" *Prentice Hall*. (chapter 3)

[16] S. Sastry and A. Isidori, (1989) "Adaptive control of linearizable systems". *IEEE TAC*, Vol. 34, No. 11, pp. 1123-1131.

[17] S. Zheng, "Nonlinear Evolution Equations" *Chapman & Hall/CRC*, 2004.