



**POLITECNICO**  
MILANO 1863

**[RE.PUBLIC@POLIMI](mailto:RE.PUBLIC@POLIMI)**

Research Publications at Politecnico di Milano

## Post-Print

This is the accepted version of:

G. Pin, A. Assalone, M. Lovera, T. Parisini

*Non-Asymptotic Kernel-Based Parametric Estimation of Continuous-Time Linear Systems*

IEEE Transactions On Automatic Control, Vol. 61, N. 2, 2016, p. 360-373

doi:10.1109/TAC.2015.2434075

The final publication is available at <https://doi.org/10.1109/TAC.2015.2434075>

Access to the published version may require subscription.

**When citing this work, cite the original published paper.**

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Permanent link to this version

<http://hdl.handle.net/11311/1007777>

# Non-Asymptotic Kernel-based Parametric Estimation of Continuous-time Linear Systems

Gilberto Pin, Andrea Assalone, Marco Lovera, and Thomas Parisini

**Abstract**—In this paper, a novel framework to address the problem of parametric estimation for continuous-time linear time-invariant dynamic systems is dealt with. The proposed methodology entails the design of suitable kernels of non-anticipative linear integral operators thus obtaining estimators showing, in the ideal case, “non-asymptotic” (*i.e.*, “finite-time”) convergence. The analysis of the properties of the kernels guaranteeing such a convergence behaviour is addressed and a novel class of admissible kernel functions is introduced. The operators induced by the proposed kernels admit implementable (*i.e.*, finite-dimensional and internally stable) state-space realizations. Extensive numerical results are reported to show the effectiveness of the proposed methodology. Comparisons with some existing continuous-time estimators are addressed as well and insights on the possible bias affecting the estimates are provided.

## I. INTRODUCTION

In many engineering applications the direct estimation of the parameters of a Continuous-Time (CT) model from sampled input-output data is an important problem for which several methods and tools have been developed so far.

Among the various techniques proposed in the literature for CT parameter identification of linear dynamical systems (see, for example, [1], [2], [3], [4], [5], [6], the contributed volume [7] and the recent special issue [8]), two main families of methods can be identified depending on the approach used to overcome the unavailability of the time-derivatives of the input-output signals: *i*) State Variable Filtering (SVF) and *ii*) integral methods (IMs).

The SVF approach - not dealt with in this paper - consists in filtering the system’s inputs and outputs in order to obtain *prefiltered* time-derivatives in the bandwidth of interest that may be exploited, in place of the unavailable derivatives of the signals, to estimate the model parameters. Instead, IMs are closely related to the proposed methodology and they have quite a long history in the field of continuous-time systems identification. Among integral techniques, we recall *i*) the Modulating Function (MF) method, which relies on the repeated integration of input-output signals over finite-length intervals to minimize the effect of unknown initial conditions on the estimates; *ii*) the linear integral filter method, in which the initial conditions must be considered explicitly, by augmenting the dimension of the decision space with the

unknown initialization variables; *iii*) the re-initialized partial moments method, that consists in the input-output signals over finite-length time windows, in sampling the integrals, and finally in performing the regression over a discrete time-series, making the overall estimator an inherently hybrid dynamical system.

In the context of CT identification, asymptotic convergence properties can be proved and several algorithms have been devised to provide good performance in terms of transient behaviour of the estimates (see, for example, the important work [3] and the references cited therein). However, in order to achieve estimates characterized by fast convergence properties, it is usually necessary to augment the vector of decision variables with the unknown initial conditions of the unmeasured states. The main drawback of this technique is related to numerical issues in estimating the initial hidden states as time goes on.

In the present paper, it is shown that the design of an *internally stable* dynamic estimator – characterized by very fast convergence properties – can be carried out by suitably shaping the kernel of a non-anticipative linear integral operator, thus yielding a stable linear dynamic system implementation. Namely, by transforming the measurable input-output signals of an unknown linear system through appropriately-designed Volterra integral operators, it is possible to obtain auxiliary signals that can be used in place of the unmeasurable derivatives to obtain the rapid convergence in terms of the data samples of the estimates of the system’s parameters. Such “surrogate” signal derivatives can be made independent from initial conditions by exploiting the so-called Non-asymptotic Kernel (NK) functions. The use of Volterra operators induced by NKs, together with a suitable augmentation strategy, allows to form a linear algebraic system that can be solved for the unknown parameters under suitable excitation conditions on the input-output signals. In this connection, a significant contribution of the paper consists in the definition and the characterization of the class of *Bivariate Causal Non-Asymptotic Kernels* (BC-NKs). The operators induced by the BC-NKs yield a “non-asymptotic” estimator that admits a finite-dimensional time-varying linear state-space realization with guaranteed internal stability.

The paper is organized as follows. In the next section the CT identification problem dealt with is stated and some preliminary concepts and tools on integral operators are given. In Section III, the main properties of non-anticipative and non-asymptotic Volterra operators are illustrated, whereas in Section IV, the use of these operators is specialized to the systems identification scenario. Section V provides the characterization of the BC-NK mentioned above, and Section VI addresses

G. Pin is with Electrolux Professional S.p.A, Pordenone, Italy (gilbertopin@alice.it); A. Assalone is with the Institute for Design and Control of Mechatronical Systems at the Johannes Kepler University, Linz, Austria (Andrea.Assalone@jku.at); M. Lovera is with the Dip. di Scienze e Tecnologie Aerospaziali, Politecnico di Milano, Italy (marco.lovera@polimi.it); T. Parisini is with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK and also with the Dept. of Engineering and Architecture at the University of Trieste, Italy (t.parisini@gmail.com).

the asymptotic analysis of the identification technique. Finally, Section VII reports extensive simulation results and comparisons showing the effectiveness of the proposed technique.

## II. PROBLEM STATEMENT AND PRELIMINARIES

Consider the following SISO CT system  $\mathcal{S}_{u \rightarrow y}$ :

$$y^{(n)}(t) = \sum_{i=0}^{n-1} a_i y^{(i)}(t) + \sum_{k=0}^{m-1} b_k u^{(k)}(t), \quad \forall t \in \mathbb{R}_{\geq 0} \quad (1)$$

where  $y^{(i)}(0) = y_0^{(i)}$ ,  $i \in \{0, \dots, n-1\}$  and  $u^{(k)}(0) = u_0^{(k)}$ ,  $k \in \{0, \dots, m-1\}$ , with  $m \in \mathbb{Z}_{>0}$ ,  $n \in \mathbb{Z}_{>0}$ ,  $m \leq n$ . The values of the constant parameters  $a_i \in \mathbb{R}$ ,  $i \in \{0, \dots, n-1\}$  and  $b_k \in \mathbb{R}$ ,  $k \in \{0, \dots, m-1\}$  are unknown. The only measurable signals are  $y(t)$  and  $u(t)$ , while their time-derivatives are not assumed to be available. Our objective consists in estimating the system's parameters  $a_i$  and  $b_k$  by suitably processing the input and output signals  $u(t)$  and  $y(t)$ .

In the following, for the reader's convenience, some basic concepts of linear integral operators' algebra (see [9] and the references therein) and realization theory (for example, refer to [10] and [11]) are recalled. More specifically, we consider transformations acting on the Hilbert space  $\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0})$  of locally square-integrable functions with domain  $\mathbb{R}_{\geq 0}$  and range  $\mathbb{R}$  (i.e.,  $x(\cdot) \in \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}) \Leftrightarrow (x(\cdot) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}) \wedge (\int_B |x(t)|^2 dt < \infty, \forall \text{ compact } B \subset \mathbb{R}_{\geq 0})$ ). The notation  $\mathcal{B}(\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}), \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}))$  will be used to denote the set of all the *bounded linear operators*  $T : \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}) \rightarrow \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0})$ . Given a function  $u(\cdot) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ , with  $u \in \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0})$ , the image function through a linear operator  $T \in \mathcal{B}(\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}), \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}))$  is denoted as  $Tu$ , and its value at time  $t \in \mathbb{R}_{\geq 0}$  is denoted as  $[Tu](t)$ .

In the paper, we resort to Volterra linear integral operators  $V_K \in \mathcal{B}(\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}), \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}))$ , defined as

$$[V_K u](t) \triangleq \int_0^t K(t, \tau) u(\tau) d\tau, \quad t \in \mathbb{R}_{\geq 0},$$

where  $u(\cdot) \in \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0})$  and the function  $K(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is a Hilbert-Schmidt ( $\mathcal{HS}$ ) *Kernel Function*. Volterra operators are a special case of the more general Fredholm operators, defined as

$$[F_{\tilde{K}} v](t) \triangleq \int_a^b \tilde{K}(t, \tau) v(\tau) d\tau, \quad t \in \mathbb{R},$$

with  $v(\cdot) \in \mathcal{L}_{loc}(\mathbb{R})$ ,  $a, b \in \mathbb{R}$  ( $a < b$ ) and  $\tilde{K}(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ . It is easy to show that the Fredholm operator specializes to the Volterra operator when we consider  $t \geq 0$ ,  $a \leq 0$ ,  $b \geq t$ , and the kernel  $\tilde{K}(t, \tau)$  verifies the condition

$$\tilde{K}(t, \tau) = 0, \quad \forall \tau \in \mathbb{R} \setminus [0, t], \forall t \in \mathbb{R}.$$

Indeed, for any  $K(t, \tau) \in \mathcal{HS}$ , we can write

$$\tilde{K}(t, \tau) = K(t, \tau)H(\tau)H(t - \tau), \quad \forall t \in \mathbb{R}_{\geq 0}, \forall \tau \in [a, b],$$

where  $H(\cdot)$  is the right-continuous Heaviside step-function. The kernel  $\tilde{K}(t, \tau)$  is called *Fredholm extension* of the Volterra kernel  $K(t, \tau)$ . The Fredholm kernel extension can be exploited to specialize to Volterra operators the properties and

the results conceived within the Fredholm operator theory. In particular, we will use the Fredholm kernel extension to characterize the kernel of composed Volterra operators (see the Appendix).

A *signal* is defined as a generic function of time  $u(t) : t \mapsto u_t$ ,  $u_t \in \mathbb{R}$ , such that  $u(\cdot) \in \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0})$ . Furthermore, given two scalars  $a, b \in \mathbb{R}_{\geq 0}$ , with  $a < b$ , let us denote by  $u_{[a,b]}(\cdot)$  and  $u_{(a,b]}(\cdot)$  the *restriction* of a signal  $u(\cdot)$  to the closed interval  $[a, b]$  and to the left-open interval  $(a, b]$ , respectively. Then, we have the following:

**Definition 2.1 (Weak (generalized) Derivative):** Let  $u(\cdot) \in \mathcal{L}_{loc}^1(\mathbb{R}_{\geq 0})$ . We say that  $u^{(1)}(\cdot)$  is a *weak derivative* of  $u(\cdot)$  if

$$\int_0^t u(\tau) \left( \frac{d}{d\tau} \phi(\tau) \right) d\tau = - \int_0^t u^{(1)}(\tau) \phi(\tau) d\tau, \quad \forall t \in \mathbb{R}_{\geq 0}$$

for all  $\phi \in C^\infty$ , with  $\phi(0) = \phi(t) = 0$ .

We denote the  $i$ -th order generalized derivative as  $u^{(i)}(\cdot)$ ,  $i \in \mathbb{Z}_{\geq 0}$ . Moreover, given a kernel function  $K(\cdot, \cdot) \in \mathcal{HS}$  in two variables, the  $i$ -th order weak derivative of  $K$  with respect to the second argument will be denoted as  $K^{(i)}$ ,  $i \in \mathbb{Z}_{\geq 0}$ . Finally, the notion of BIBO stability for an integral operator is introduced.

**Definition 2.2 (BIBO Stability):** A bounded linear operator  $T \in \mathcal{B}(\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}), \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}))$  is said *BIBO-stable* if:

$$\begin{aligned} & |[Tx](t)| < \infty, \quad \forall t \in \mathbb{R}_{\geq 0}, \\ & \forall x(\cdot) \in \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}) : \{ |x(\tau)| < \infty, \forall \tau \in \mathbb{R}_{\geq 0} \}. \end{aligned}$$

In the case of a Volterra operator  $V_K$ , BIBO stability is equivalent to the following property of the kernel:

$$\sup_{t \in \mathbb{R}_{>0}} \left\{ \int_0^t |K(t, \tau)| d\tau \right\} < \infty. \quad (2)$$

Condition (2) will be used in the sequel to assess the stability of the operators in our setting. A kernel fulfilling (2) will be called a *BIBO stable kernel*. In this respect, it is worth noting that BIBO stability *per se* is not sufficient to establish the existence of a finite-dimensional state-space realization for an operator, that is, its implementability. The order of the realization can be determined only when an analytical expression for the kernel is available.

## III. NON-ANTICIPATIVE AND NON-ASYMPTOTIC VOLTERRA OPERATORS

In this section, the concepts of causality and non-anticipativity are addressed as they play a key role in characterizing the implementability (existence of a stable finite-dimensional state-space realization for the integral operators) of the proposed methodology.

### A. Non-anticipativity of the Volterra operator

In qualitative terms, an operator  $T \in \mathcal{B}(\mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}), \mathcal{L}_{loc}^2(\mathbb{R}_{\geq 0}))$  is said to be *causal* (non-anticipative) if at any time  $t > 0$  (respectively,  $t \geq 0$ ) the image of a signal  $x(\cdot)$  at time  $t$ ,  $[Tx](t)$ , depends only on the

restriction  $x_{[0,t]}(\cdot)$  (respectively,  $x_{[0,t]}(\cdot)$ ). Being the Volterra operator inherently non-anticipative, the signal  $[V_K x](t)$ , for  $t > 0$ , can be obtained as the output of a dynamic system described by the following scalar integro-differential equation

$$\begin{cases} \xi^{(1)}(t) = K(t, t)x(t) + \int_0^t \left( \frac{\partial}{\partial t} K(t, \tau) \right) x(\tau) d\tau \\ [V_K x](t) = \xi(t) \end{cases},$$

where  $\xi^{(1)}(0) = 0$  and  $\xi(0) = \xi_0 = \int_0^0 K(0, \tau)x(\tau)d\tau$ . Clearly,  $\xi^{(1)}(t) = \frac{d}{dt}[V_K x](t)$  is obtained by applying the Leibnitz rule in deriving the integral.

The following result is useful in dealing with the application of Volterra operators to the derivatives of a signal.

*Lemma 3.1 (Proof in Appendix):* For a given  $i \geq 0$ , consider a signal  $x(\cdot) \in \mathcal{L}^2(\mathbb{R}_{\geq 0})$  that admits a  $i$ -th weak derivative in  $\mathbb{R}_{\geq 0}$  and a kernel function  $K(\cdot, \cdot) \in \mathcal{HS}$  that admits the  $i$ -th derivative (in the conventional sense) with respect to the second argument,  $\forall t \in \mathbb{R}_{\geq 0}$ . Then:

$$\begin{aligned} [V_K x^{(i)}](t) &= \sum_{j=0}^{i-1} (-1)^{i-j-1} x^{(j)}(t) K^{(i-j-1)}(t, t) \\ &+ \sum_{j=0}^{i-1} (-1)^{i-j} x^{(j)}(0) K^{(i-j-1)}(t, 0) \\ &+ (-1)^i [V_{K^{(i)}} x](t), \quad \forall t \in \mathbb{R}_{\geq 0} \end{aligned} \quad (3)$$

that is, the function  $[V_{K_h} x^{(i)}](\cdot)$  is non-anticipative with respect to the lower-order derivatives  $x(\cdot), \dots, x^{(i-1)}(\cdot)$ .  $\square$

### B. Non-asymptoticity conditions

By exploiting (3), we now characterize the kernels for which the transformed signal  $[V_K x^{(i)}](\cdot)$  is independent from the initial values of the signal and of its derivatives. The following definition characterizes the kernels yielding non-asymptotic Volterra operators.

*Definition 3.1 (i-th Order Non-Asymptotic Kernel):*

Consider a kernel  $K(\cdot, \cdot)$  satisfying the assumptions posed in the statement of Lemma 3.1; if, for a given  $i \geq 1$ , the kernel verifies the supplementary condition

$$K^{(j)}(t, 0) = 0, \quad \forall t \in \mathbb{R}_{\geq 0}, \forall j \in \{0, \dots, i-1\},$$

then, it is called an  $i$ -th order non-asymptotic kernel.

*Lemma 3.2 (Non-asymptoticity):* If a kernel  $K(\cdot, \cdot)$  is at least  $i$ -th order non-asymptotic, then the image function of  $x^{(i)}(\cdot)$  at time  $t$ ,  $[V_K x^{(i)}](t)$ , depends only on the instantaneous values of the lower-order derivatives  $(x(t), x^{(1)}(t), \dots, x^{(i-1)}(t))$  and on the restriction  $x_{[0,t]}(\cdot)$ , but not on the initial states  $x(0), x^{(1)}(0), \dots, x^{(i-1)}(0)$ .  $\square$

The proof of Lemma 3.2 follows immediately from Lemma 3.1 and is therefore omitted.

Up to now, a candidate class of kernels has been characterized which allows to remove the influence of the unknown initial derivatives from the transformed signal  $[V_K x^{(i)}](t)$ . However, beyond depending on the current value  $x(t)$  and its past time-behaviour, such a signal depends also on the unmeasurable instantaneous values of the lower-order derivatives

$x^{(j)}(t)$ , with  $j \in \{1, \dots, n-1\}$ . To address this issue, we need to introduce the concept of composed (or nested) Volterra operators and to discuss some relevant properties.

Let us denote by  $[V_{K_N \circ \dots \circ K_1} x^{(i)}](\cdot)$ , the image function obtained by the composition of  $N$  Volterra integral operators applied to  $x^{(i)}(\cdot)$ :

$$[V_{K_N \circ \dots \circ K_1} x^{(i)}] = [V_{K_N} \cdots [V_{K_2} [V_{K_1} x^{(i)}]]]. \quad (4)$$

In view of the composition property of Volterra operators (see (39) and (40) in the Appendix), it holds that the composed operator is in turn a Volterra operator with kernel  $K_N \circ K_{N-1} \circ \dots \circ K_i \circ \dots \circ K_2 \circ K_1$ , where  $\cdot \circ \cdot$  denotes the kernel-composition integral (see (40) in the Appendix). The following result can now be proved.

*Theorem 3.1 (Non-asymptotic Derivative Image):* Let  $x^{(i)}(\cdot)$  be the  $i$ -th derivative of the signal  $x(\cdot)$  and let  $N \geq i$  be an arbitrary integer. Given  $N$  kernel functions  $K_1(\cdot, \cdot), \dots, K_N(\cdot, \cdot)$ , such that  $K_1$  is  $d$ -th order non-asymptotic, with  $d \geq i-1$  and  $K_j \in \mathcal{HS}, \forall j \in \{1, \dots, N\}$ , consider the composed operator  $V_{P_N}$ , with kernel

$$P_N \triangleq K_N \circ \dots \circ K_2 \circ K_1.$$

The image of the derivative signal  $x^{(i)}(\cdot)$  through  $V_{P_N}$ , i.e.,  $[V_{P_N} x^{(i)}](\cdot)$ , can be obtained as the image of the restriction  $x_{[0,t]}(\cdot)$  through a non-anticipative operator. Indeed, there exists an operator  $V_{R_{N,i}}$ , induced by the kernel

$$R_{N,i} \triangleq -K_N \circ K_{N-1} \circ \dots \circ K_{i+1} \circ T_i,$$

with  $T_i(\cdot, \cdot)$  defined recursively by

$$\begin{aligned} T_j(t, \tau) &\triangleq - \left( K_j \circ T_{j-1}^{(1)} \right) (t, \tau) + K_j(t, \tau) T_{j-1}(\tau, \tau) \\ &\forall j \in \{2, \dots, i\}, \forall (t, \tau) \in \mathbb{R}^2 \end{aligned} \quad (5)$$

with the initialization  $T_1 \triangleq K_1$ , such that

$$\begin{aligned} [V_{P_N} x^{(i)}](t) &= R_{N,i}(t, t)x(t) - R_{N,i}(t, 0)x(0) - [V_{R_{N,i}^{(1)}} x](t). \end{aligned} \quad (6)$$

*Proof:* First, by integrating by parts, the innermost operator in the right-hand side of (4) can be decomposed as

$$[V_{P_N} x^{(i)}](t) = [V_{K_N} \cdots [V_{K_2} (\tilde{x}_1 - [V_{K_1^{(1)}} x^{(i-1)}])]](t) \quad (7)$$

where  $\tilde{x}_1(t) \triangleq K_1(t, t)x^{(i-1)}(t) - K_1(t, 0)x^{(i-1)}(0)$ . Now, consider the composed kernel

$$T_2(t, \tau) = -(K_2 \circ K_1^{(1)})(t, \tau) + K_2(t, \tau)K_1(\tau, \tau)$$

obtained by (5) (recall that  $K_2 \circ K_1^{(1)} = \int_{\tau}^t K_2(t, \sigma)K_1^{(1)}(\sigma, \tau)d\sigma$ , see also the Appendix). By the non-asymptoticity property of  $K_1$ :  $K_1(t, 0) = 0, \forall t \in \mathbb{R}_{\geq 0}$ , and in view of (5) and (7) we get

$$[V_{P_N} x^{(i)}] = [V_{K_N} \cdots [V_{K_3} [V_{T_2} (-x^{(i-1)})]]](t). \quad (8)$$

Integrating by parts, the innermost operator in (8) can be decomposed as

$$[V_{P_N} x^{(i)}] = \left[ V_{K_N} \cdots \left[ V_{K_3} (\tilde{x}_2 + [V_{T_2^{(1)}} x^{(i-2)}]) \right] \right] (t),$$

where  $\tilde{x}_2(t) \triangleq -T_2(t, t)x^{(i-2)}(t) + T_2(t, 0)x^{(i-2)}(0)$ . Since the  $i$ -th order non-asymptoticity condition is implicitly in place when considering the kernel  $K_1(\cdot, \cdot)$ , then, by (5), we have  $T_2(t, 0) = 0, \forall t \in \mathbb{R}_{\geq 0}$ , i.e., also  $T_2(\cdot, \cdot)$  is non-asymptotic. Thus, we can write

$$\left[ V_{P_N} x^{(i)} \right] (t) = \left[ V_{K_N} \cdots \left[ V_{K_4} [V_{T_3} (-x^{(i-2)})] \right] \right] (t).$$

Integrating again by parts, the innermost operator can be decomposed as

$$\left[ V_{P_N} x^{(i)} \right] (t) = \left[ V_{K_N} \cdots \left[ V_{K_4} (\tilde{x}_3 + [V_{T_3^{(1)}} x^{(i-3)}]) \right] \right],$$

where  $\tilde{x}_3(t) \triangleq -T_3(t, t)x^{(i-3)}(t) + T_3(t, 0)x^{(i-3)}(0)$ . Due to the fact that the kernel  $K_1(\cdot, \cdot)$  is at least  $i-1$ -th order non-asymptotic, then  $T_j(t, 0) = 0, \forall j \in 1, \dots, i-1$ . By iterating this line of reasoning, we finally get

$$\begin{aligned} [V_{P_N} x^{(i)}] (t) &= \left[ V_{K_N} \cdots \left[ V_{K_{i+1}} [V_{T_i} (-x^{(1)})] \right] \right] (t) \\ &= R_{N,i}(t, t)x(t) - R_{N,i}(t, 0)x(0) - [V_{R_{N,i}^{(1)}} x] (t), \end{aligned}$$

thus ending the proof.  $\blacksquare$

*Remark 3.1 (Implications):* In Theorem 3.1, the existence of a composed Volterra integral operator has been shown, namely  $V_{P_N} = V_{K_N \circ \dots \circ K_1}$ , that, fed by the  $i$ -th derivative  $x^{(i)}(\cdot)$  of a signal, produces an image signal, say  $[V_{P_N} x^{(i)}] (\cdot)$ , which, in turn, can be expressed, in the most general case, in terms of the sole restriction  $x_{[0,t]}(\cdot)$  and that, in any case, does not depend on the initial conditions of the hidden derivatives. Assume now that  $x^{(i)}(\cdot)$  is not measurable while  $x(\cdot)$  is available; then, thanks to (6), the signal  $[V_{P_N} x^{(i)}] (\cdot)$  can be obtained by applying a non-anticipative operator (see (6)) to  $x(\cdot)$ .

#### IV. NON-ASYMPTOTIC KERNELS FOR PARAMETER ESTIMATION

In the context of the parameter estimation problem formulated in Section II, consider (1) which relates the unknown parameters with the time-derivatives of the signals  $u(\cdot)$  and  $y(\cdot)$ . In the sequel, the results presented in the previous section are exploited to overcome the unavailability of signal derivatives (hidden internal states of the system) in (1), thus obtaining *non-asymptotic estimates* of the unknown parameters by means of *causal filtering*.

Let us select an integer  $N \geq n$  and apply the Volterra operator  $V_{P_N} = V_{K_N \circ \dots \circ K_1}$  (with kernels taken as in Theorem 3.1) to both sides of (1) thus obtaining

$$[V_{P_N} y^{(n)}] (\cdot) = \sum_{i=0}^{n-1} a_i [V_{P_N} y^{(i)}] (\cdot) + \sum_{k=0}^{m-1} b_k [V_{P_N} u^{(k)}] (\cdot). \quad (9)$$

In view of (6), (9) can be rewritten as

$$r_{y,n}(t) = \sum_{i=0}^{n-1} a_i r_{y,i}(t) + \sum_{k=0}^{m-1} b_k r_{u,k}(t), \quad \forall t \in \mathbb{R}_{\geq 0}, \quad (10)$$

where the auxiliary signals in (10) can be obtained as the image of measurable signals  $y(\cdot)$  and  $u(\cdot)$  through non-anticipative operators, that is, for  $j \in \{0, \dots, n\}$ ,

$$r_{y,j}(t) = R_{N,j}(t, t)y(t) - R_{N,j}(t, 0)y(0) - [V_{R_{N,j}^{(1)}} y] (t),$$

and, for  $j \in \{0, \dots, m-1\}$ ,

$$r_{u,j}(t) = R_{N,j}(t, t)u(t) - R_{N,j}(t, 0)u(0) - [V_{R_{N,j}^{(1)}} u] (t).$$

Finally, by introducing the true parameter vector

$$\theta^* \triangleq [a_0, \dots, a_{n-1}, b_0, \dots, b_{m-1}]^\top,$$

and the vector of auxiliary signals

$$\mathbf{z}(t) \triangleq [r_y(t), \dots, r_{y,n-1}(t), r_u(t), \dots, r_{u,m-1}(t)]^\top,$$

equation (10) can be rewritten in a compact form as

$$\mathbf{z}^\top(t)\theta^* = r_{y,n}(t), \quad t \in \mathbb{R}_{\geq 0}. \quad (11)$$

Now, assuming that all the operators in our formulation admit a stable realization, we need to collect a suitable number of equations like (11) in order to form a well-posed algebraic system, to be solved in the unknown parameters. Several approaches can be used to obtain the needed set of constraints. As usual in the CT identification framework, a suitable number of constraints can be formed by sampling the signals in (10)-(11) at different time-instants and in collecting the data for batch regression.

In order to emphasize the generality of the proposed methodology, we still have not assigned explicit analytic expressions to the kernels  $K_1, \dots, K_N$  and to  $R_{N,j}$ ,  $j \in 0, \dots, n$ , which are needed to compute the auxiliary signals. The problem of selecting a class of non-asymptotic kernels yielding stable finite-dimensional state-space realizations is addressed in the following.

#### V. BIVARIATE CAUSAL NON-ASYMPTOTIC KERNELS

In this section, the main contribution of this paper is presented by introducing a novel class of non-asymptotic causal kernels with guaranteed stability properties that allow the application of Theorem 3.1 to construct auxiliary signals yielding the unavailable input-output time-derivatives. To this end, let us first introduce the following definition.

*Definition 5.1 (i-th Order BC-NK):* A kernel  $K(\cdot, \cdot) \in \mathcal{HS}$  that satisfies the assumptions given in Lemma 3.1 and that, for a given  $i \geq 1$ , also verifies the conditions

$$\begin{aligned} K^{(j)}(t, 0) &= 0, \quad K^{(j)}(t, t) = 0, \\ \forall t \in \mathbb{R}_{\geq 0}, \forall j &\in \{0, \dots, i-1\} \end{aligned} \quad (12)$$

is called *i-th Order Bivariate (strict) Causal Non-Asymptotic (BC-NK) kernel*.

It is worth noting that, owing to Lemma 3.1, if a kernel  $P_N$  verifies both the conditions stated in (12), then the image of a derivative signal  $x^{(i)}$  can be expressed as

$$\left[ V_{P_N} x^{(i)} \right] (t) = (-1)^i \left[ V_{P_N^{(i)}} x \right] (t).$$

On the other hand, a kernel  $P_N$  verifying (12) is not easy to obtain by means of the kernel composition integral, as outlined in the statement of Theorem 3.1.

Therefore, in the paper we devise the following analytical form for a kernel  $P_N = C_{\omega,N}(t, \tau)$ . Specifically, we propose the following bivariate function as a BC-NK that we exploit later on to design a stable non-asymptotic continuous-time parametric estimator:

$$C_{\omega,N}(t, \tau) \triangleq e^{-\omega(t-\tau)} (1 - e^{-\omega\tau})^N \left[ 1 - e^{-\omega(t-\tau)} \right]^N, \quad (13)$$

where  $\omega \in \mathbb{R}_{>0}$  is a tuning parameter (see Sections VI and VII). The non-asymptoticity, causality and stability characteristics of the BC-NK given by (13) are provided in the following lemma.

*Lemma 5.1 (Kernel Characterization:  $C_{\omega,N}(t, \tau)$ ):* The bivariate kernel  $C_{\omega,N}(t, \tau)$  is BIBO-stable and  $N$ -th order BC-NK. Moreover, all the kernel derivatives  $C_{\omega,N}^{(i)}(t, \tau)$ , with  $i \in \{0, \dots, N-1\}$ , are BIBO-stable.  $\square$

*Proof:* First, we prove that the kernel  $C_{\omega,N}(t, \tau)$  is a  $N$ -th order BC-NK. Indeed, all the non-anticipativity conditions up to the  $N$ -th order are met by the factor  $(1 - e^{-\omega\tau})^N$ . The causality conditions up to the  $N$ -th order are met by the third factor  $(1 - e^{-\omega(t-\tau)})^N$ . The BIBO-stability of  $C_{\omega,N}^{(i)}(t, \tau)$  is implied by the fact that each  $(e^{-\omega(t-\tau)})^{(i)}$ , with  $i \in \{0, \dots, N-1\}$ , is BIBO-stable and the following terms are bounded:  $|(1 - e^{-\omega(t-\tau)})^N| < 1$ ,  $\forall \tau : 0 \leq \tau \leq t$  and their derivatives up to the  $(N-1)$ -th order are bounded too.  $\blacksquare$

Now, we describe how the image of the derivative  $x^{(i)}(\cdot)$  through the operator  $V_{C_{\omega,N}}$ , i.e.,  $[V_{C_{\omega,N}} x^{(i)}] = (-1)^i [V_{C_{\omega,N}^{(i)}} x]$  can be obtained as the output of a BIBO-stable finite-dimensional time-varying linear system. First, the  $i$ -th derivative of the BC-NK (13) with respect to the second argument can be expressed as:

$$C_{\omega,N}^{(i)}(t, \tau) = \sum_{j=1}^{N+1} e^{-\omega j t} f_{\omega,N|i,j}(\tau), \quad (14)$$

where  $f_{\omega,N|i,j}(\cdot)$  are univariate functions of  $\tau$ . Let  $C_{\omega,N|i,j}(t, \tau) \triangleq (-1)^i e^{-\omega j t} f_{\omega,N|i,j}(\tau)$ ; then, by the linearity of the Volterra operator, it follows that

$$[V_{C_{\omega,N}} x^{(i)}](t) = (-1)^i [V_{C_{\omega,N}^{(i)}} x](t) = \sum_{j=1}^{N+1} [V_{C_{\omega,N|i,j}} x](t). \quad (15)$$

Moreover, letting  $\xi_{i,j}(t) \triangleq [V_{C_{\omega,N|i,j}} x](t)$ , with  $i \in \{0, \dots, n\}$ ,  $j \in \{1, \dots, N+1\}$ , and taking into account that,  $\forall t \in \mathbb{R}_{\geq 0}$ , we have

$$\begin{aligned} C_{\omega,N|i,j}(t, 0) &= 0 \\ \frac{\partial}{\partial t} C_{\omega,N|i,j}(t, \tau) &= -\omega j e^{-\omega j t} f_{\omega,N|i,j}(\tau) \end{aligned} ,$$

then,  $[V_{C_{\omega,N}} x^{(i)}]$  admits the following  $(N+1)$ -th dimensional state-space realization:

$$\begin{cases} \xi_{i,j}^{(1)}(t) = C_{\omega,N|i,j}(t, t)x(t) - \omega j \xi_{i,j}(t), & j = 1, \dots, N+1 \\ [V_{C_{\omega,N}} x^{(i)}](t) = \sum_{j=1}^{N+1} \xi_{i,j}(t), \end{cases} \quad (16)$$

with  $\xi_{i,1}(0) = 0, \dots, \xi_{i,N+1}(0) = 0$ . As  $|e^{-\omega j t} f_{\omega,N|i,j}(t)| < \infty$ ,  $\forall j \in \{1, \dots, N+1\}$ , (i.e., all the time-varying terms affine to the  $x(t)$ -injection are bounded), and since the system is diagonal with  $\omega > 0$ , then (16) is a BIBO-stable time-varying linear system.

Moreover, there exist finite scalars  $\beta_{i,j} \in \mathbb{R}_{>0}$  such that  $(e^{-\omega j t} f_{\omega,N|i,j}(t)) \rightarrow \beta_{i,j}$  for  $t \rightarrow \infty$ . This implies that the time-varying system (16), for  $t \rightarrow \infty$ , tends to a stable linear time-invariant system in which the  $x(t)$ -injection is never suppressed. Thanks to (16), the extended auxiliary signal vector  $\mathbf{z}_e(t)$ , which embeds both the signals  $z(t)$  and  $r_{y,n}$  needed to form the constraint (11):

$$\mathbf{z}_e(t) = [r_{y,0}(t), \dots, r_{y,n}(t), r_{u,0}(t), \dots, r_{u,m-1}(t)],$$

with

$$\begin{aligned} r_{y,i} &= [V_{C_{\omega,N}} y^{(i)}], & i \in \{0, \dots, n\}, \\ r_{u,i} &= [V_{C_{\omega,N}} u^{(i)}], & i \in \{0, \dots, m-1\}, \end{aligned}$$

can be obtained as the output of an overall  $n_\xi = (n + m + 1)(N+1)$ -dimensional linear time-varying dynamical system, described as:

$$\mathcal{G}_{u,y \rightarrow \mathbf{z}_e} : \begin{cases} \xi^{(1)}(t) = \mathbf{G}_\xi \xi(t) + \mathbf{E}_y(t)y(t) + \mathbf{E}_u(t)u(t), \\ \mathbf{z}_e(t) = \mathbf{H}_\xi \xi(t), \end{cases} \quad (17)$$

with  $\xi(0) = \mathbf{0}$  and where  $\xi \in \mathbb{R}^{n_\xi}$  is the overall state-vector; moreover, the other quantities in (17) are

$$\mathbf{G}_\xi = \text{blockdiag}(\mathbf{G}, \dots, \mathbf{G}) \in \mathbb{R}^{n_\xi \times n_\xi},$$

$$\mathbf{G} = \text{diag}(-\omega, \dots, -\omega(N+1)) \in \mathbb{R}^{(N+1) \times (N+1)},$$

$$\mathbf{E}_y(t) = [\mathbf{E}_0(t) \ \dots \ \mathbf{E}_n(t) \ | \ 0 \ \dots \ 0]^\top \in \mathbb{R}^{n_\xi},$$

$$\mathbf{E}_u(t) = [0 \ \dots \ 0 \ | \ \mathbf{E}_0(t) \ \dots \ \mathbf{E}_{m-1}(t)]^\top \in \mathbb{R}^{n_\xi},$$

$$\mathbf{E}_i(t) = \begin{bmatrix} C_{\omega,N|i,1}(t, t) \\ \vdots \\ C_{\omega,N|i,N+1}(t, t) \end{bmatrix} \in \mathbb{R}^{N+1}, \quad (18)$$

and

$$\mathbf{H}_\xi = \text{blockdiag}(\mathbf{1}^\top, \dots, \mathbf{1}^\top) \in \mathbb{R}^{n_\xi \times n_\xi},$$

where  $\mathbf{1}^\top$  denotes a row vector of ones with  $(N+1)$  elements.

## VI. ASYMPTOTIC ANALYSIS OF THE KERNEL-BASED CONTINUOUS-TIME MODEL IDENTIFICATION ALGORITHM

In the following, the analysis of the asymptotic properties of the proposed identification algorithm is addressed. To this purpose, the so-called hybrid framework of continuous-time model identification (see [7]) is assumed, i.e., the system generating the data is assumed to be a continuous-time, linear time-invariant system, while the noise model is defined in discrete-time.

Consider the continuous-time linear time-invariant input-output system

$$x^{(n)}(t) = \sum_{i=0}^{n-1} a_i x^{(i)}(t) + \sum_{i=0}^{m-1} b_i u^{(i)}(t), \quad (19)$$

corresponding to (1), and introduce the following assumptions:

*Assumption 1:* the system (19) is asymptotically stable.

*Assumption 2:* The input  $u$  is a quasi stationary, piece-wise constant, deterministic sequence.

Consider now a dataset consisting of  $K$  input-output measurements associated with the sampling instants  $t_k = t_0 + kT$ ,  $k = 0, \dots, K-1$  (uniform sampling is assumed, for the sake of simplicity), defined as follows

$$y(k) = x(t_k) + e(k), \quad u(k) = u(t_k),$$

where  $e(k)$  represents (output) measurement noise. Then, the following further assumptions are in place.

*Assumption 3:* the process  $e$  is a stationary zero mean white process noise with second moments given by

$$\mathbb{E}[e(i)e^\top(j)] = R_e \delta_{ij}$$

where  $\delta_{ij}$  is the Kronecker delta.

*Assumption 4:* The input  $u$  is uncorrelated with noise  $e$ .

*Assumption 5:* Instantaneous sampling, in the sense of [12], is assumed, *i.e.*, sampling is assumed to be "fast" with respect to the dynamics of interest.

To deal with deterministic and stochastic signals in a compact manner, the following operator is defined

$$\bar{\mathbb{E}}[\cdot] = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=1}^K \mathbb{E}[\cdot]$$

where  $\mathbb{E}[\cdot]$  is the expectation operator. For two signals  $a(t)$  and  $b(t)$ , the cross-covariance matrix is denoted as  $\mathbf{R}_{ab} = \bar{\mathbb{E}}[a(t)b(t)^\top]$ . Then, the proposed identification algorithm aims at estimating the parameter vector

$$\boldsymbol{\theta} = [a_0, a_1, \dots, a_{n-1}, b_0, b_1, \dots, b_{m-1}]^\top = [\boldsymbol{\theta}_y^\top \boldsymbol{\theta}_u^\top]^\top$$

on the basis of the available data by solving the linear regression problem

$$r_{y,n} = \mathbf{z}^\top \boldsymbol{\theta} \quad (20)$$

where

$$\mathbf{z} = [r_{y,0}, \dots, r_{y,n-1}, r_{u,0}, \dots, r_{u,m-1}]^\top = [\mathbf{z}_y^\top, \mathbf{z}_u^\top]^\top \quad (21)$$

and  $r_{y,i}, r_{u,i}$  are given by the outputs of a suitably discretised version of the filter bank in equation (16), to be defined in the following. Note that the state space representation of the filter bank in (16) can be broken down to a set of  $n+1$  filters for the output  $y$

$$\dot{\boldsymbol{\xi}}_{y,i}(t) = \mathbf{G}\boldsymbol{\xi}_{y,i}(t) + \mathbf{E}_i(t)y(t) \quad (22)$$

$$r_{y,i}(t) = \mathbf{H}\boldsymbol{\xi}_{y,i}(t), \quad (23)$$

$i = 0, \dots, n$ , and  $m$  filters for the input  $u$

$$\dot{\boldsymbol{\xi}}_{u,i}(t) = \mathbf{G}\boldsymbol{\xi}_{u,i}(t) + \mathbf{E}_i(t)u(t) \quad (24)$$

$$r_{u,i}(t) = \mathbf{H}\boldsymbol{\xi}_{u,i}(t), \quad (25)$$

$i = 0, \dots, m-1$ , where

$$\mathbf{G} = \text{diag}[-\omega, \dots, -\omega(N+1)] \quad \text{and} \quad \mathbf{H} = [1, \dots, 1].$$

For the purpose of an asymptotic analysis, the time-varying matrices  $\mathbf{E}_i(t)$  can be replaced with their asymptotic values, so in the following the constant matrices

$$\bar{\mathbf{E}}_i = \lim_{t \rightarrow \infty} \mathbf{E}_i(t) \quad (26)$$

are considered in the definition of the filters in equations (23) and (25). Note that, in view of the definitions of the kernels giving rise to the time-varying vectors  $\mathbf{E}_i(t)$ , the limits in (26) are well-defined.

In the following, we denote by  $F_i(s)$  the transfer function associated with the state space quadruple  $(\mathbf{G}, \bar{\mathbf{E}}_i, \mathbf{H}, 0)$ . For the purpose of the following analysis it is interesting to point out and exploit the particular structure of the filters  $F_i(s)$ : indeed, as  $\mathbf{G}$  is diagonal and the definition of the output matrix  $\mathbf{H}$  implies that the output of each filter is simply given by the sum of its states, one can conclude that

$$F_i(s) = \sum_{j=1}^{N+1} F_{ij}(s), \quad (27)$$

with

$$F_{ij}(s) = \frac{\bar{E}_{ij}}{s - G_j}, \quad j = 1, \dots, N+1, \quad (28)$$

where  $G_j = -j\omega$ ,  $j = 1, \dots, N+1$  and where  $\bar{E}_{ij}$  denotes the  $j$ th element of vector  $\bar{\mathbf{E}}_i$ .

Finally, as the hybrid framework of continuous-time identification has been assumed, for the sake of implementation, a discretised version of the above defined continuous-time filters has to be derived. By using, *e.g.*, the backward Euler transformation, the discrete-time counterparts of the  $F_{ij}(s)$  filters can be derived as

$$F_{ij}(z) = F_{ij}(s)|_{s=(z-1)/(Tz)} = \frac{z\bar{E}_{d,ij}}{z - G_{d,j}}, \quad j = 1, \dots, N+1$$

where

$$\bar{E}_{d,ij} = \frac{\bar{E}_{ij}T}{1 - G_jT}, \quad G_{d,j} = \frac{1}{1 - G_jT}, \quad (29)$$

so that

$$F_i(z) = \sum_{j=1}^{N+1} F_{ij}(z).$$

Therefore, in discrete-time, the variables appearing in the regressions (20) and in (21) can be defined as

$$r_{y,i}(k) = F_i(z)y(k), \quad i = 0, \dots, n,$$

$$r_{u,i}(k) = F_i(z)u(k), \quad i = 0, \dots, m-1.$$

Finally, in the definition of  $r_{y,i}$  it is convenient to highlight the deterministic part, resulting from the filtering of  $x(k)$  and the stochastic part, resulting from the filtering of  $e(k)$ , as follows:

$$r_{y,i}(k) = r_{x,i}(k) + e_i(k), \quad i = 0, \dots, n,$$

where  $r_{x,i}(k) = F_i(z)x(k)$ ,  $i = 0, \dots, n$  and  $e_i(k) = F_i(z)e(k)$ ,  $i = 0, \dots, n$ . Similarly,  $\mathbf{z}_y$  in (21) can be expressed as  $\mathbf{z}_y = \mathbf{z}_x + \mathbf{z}_e$ , with obvious definitions of  $\mathbf{z}_x$  and  $\mathbf{z}_e$ .

On the basis of the above definitions, our aim is to establish an expression for the bias of the estimate of  $\boldsymbol{\theta}$  computed by solving the discrete-time regression. In this respect, for the purpose of parameter bias analysis, it is convenient to express the linear regression (20) as

$$\begin{aligned} r_{y,n} &= \mathbf{z}^\top \boldsymbol{\theta} \\ &= [\mathbf{z}_x^\top + \mathbf{z}_e^\top \quad \mathbf{z}_u^\top] \begin{bmatrix} \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_u \end{bmatrix} = [\mathbf{z}_x^\top \quad \mathbf{z}_u^\top] \begin{bmatrix} \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_u \end{bmatrix} + \mathbf{z}_e^\top \boldsymbol{\theta}_y. \end{aligned} \quad (30)$$

Left-multiplying by  $\mathbf{z}$  equation (30) and letting  $\mathbf{z}_{x,u} = [\mathbf{z}_x \quad \mathbf{z}_u]$ , we obtain

$$\begin{aligned} \mathbf{z}_{x,u} r_{y,n} &= \mathbf{z}_{x,u} \mathbf{z}_{x,u}^\top \boldsymbol{\theta} - \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} r_{x,n} - \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} e_n \\ &+ \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} [\mathbf{z}_x^\top \quad \mathbf{z}_u^\top] + \begin{bmatrix} \mathbf{z}_x \mathbf{z}_e^\top + \mathbf{z}_e \mathbf{z}_e^\top & 0 \\ \mathbf{z}_u \mathbf{z}_e^\top & 0 \end{bmatrix} \boldsymbol{\theta}. \end{aligned} \quad (31)$$

Letting  $\mathbf{R}_{\mathbf{z}_{x,u} r_{x,n}} = \bar{\mathbb{E}}[\mathbf{z}_{x,u} r_{x,n}]$  and  $\mathbf{R}_{\mathbf{z}_{x,u}} = \bar{\mathbb{E}}[\mathbf{z}_{x,u} \mathbf{z}_{x,u}^\top]$ , and applying the  $\bar{\mathbb{E}}[\cdot]$  operator to both sides of (31), after some algebra, we get

$$(\mathbf{R}_{\mathbf{z}_{x,u} r_{y,n}} - \mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} e_n) = (\mathbf{R}_{\mathbf{z}_{x,u}} + \mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix}) \boldsymbol{\theta}. \quad (32)$$

Therefore, from (32), it follows immediately that, in the absence of measurement noise, the regression reduces to

$$\mathbf{R}_{\mathbf{z}_{x,u} r_{y,n}} = \mathbf{R}_{\mathbf{z}_{x,u}} \boldsymbol{\theta}. \quad (33)$$

When noise is taken into account, the resulting estimate is necessarily affected by bias if a least squares solution of the linear regression problem in (32) is considered. While this is a known fact in the continuous-time identification literature, which has led to the development of sophisticated instrumental variable algorithms for bias elimination (see, *e.g.*, [7] and the references therein), it is interesting to pursue the above analysis further, exploiting the above derived expressions for the discrete-time counterparts of the filters (23) and (25). More precisely, letting  $\boldsymbol{\theta}^*$  the true value of the unknown parameter vector (corresponding to the solution of the noise-free regression (33)) and denoting with  $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$  the bias in the computed noisy estimate, from (32) and (33) we get

$$\mathbb{E}[\Delta\boldsymbol{\theta}] = [\mathbf{R}_{\mathbf{z}_{x,u}} + \mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix}]^{-1} [-\mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} e_n - \mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} \boldsymbol{\theta}^o]. \quad (34)$$

The noise-dependent covariance functions in (34) can be further analysed by noting that

$$\mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\mathbf{z}_e \mathbf{z}_e^\top] & 0 \\ 0 & 0 \end{bmatrix},$$

where, in turn,

$$\mathbb{E}[\mathbf{z}_e \mathbf{z}_e^\top] = \mathbb{E} \begin{bmatrix} e_0^2 & e_0 e_1 & \dots & e_0 e_{n-1} \\ \vdots & \vdots & & \vdots \\ e_0 e_{n-1} & e_1 e_{n-1} & \dots & e_{n-1}^2 \end{bmatrix},$$

and

$$\mathbf{R} \begin{bmatrix} \mathbf{z}_e \\ 0 \end{bmatrix} e_n = \mathbb{E}[e_0 e_n \quad e_1 e_n \cdots e_{n-1} e_n \quad 0 \cdots 0].$$

Hence, to evaluate the bias on the estimated parameters, we have to compute  $R_{e_0 e_q} = \mathbb{E}[e_0 e_q]$ ,  $q = 0, \dots, n$ . To this purpose, note that, by definition  $\mathbb{E}[e_0 e_q] = \mathbb{E}[(F_0(z)e)(F_q(z)e)]$ , which, in view of (27), becomes

$$\mathbb{E}[(F_0(z)e)(F_q(z)e)] = \mathbb{E} \left[ \left( \sum_{j=1}^{N+1} e_{0j} \right) \left( \sum_{l=1}^{N+1} e_{ql} \right) \right],$$

where (see (28))  $e_{0j}(k) \triangleq F_{0j}(z)e(k) = [\bar{E}_{d,0j}/(1-G_{d,j}z^{-1})]e(k)$ , or, equivalently,  $e_{0j}(k) = G_{d,j}e_{0j}(k-1) + \bar{E}_{d,0j}e(k-1)$ , and similarly for  $e_{ql}$ .

It follows that the covariance  $R_{e_{0j} e_{ql}} = \mathbb{E}[e_{0j}(k)e_{ql}(k)]$  corresponds to the covariance between two first order AR processes forced by the same white noise input, so that

$$R_{e_{0j} e_{ql}} = \frac{\bar{E}_{d,0j} \bar{E}_{d,ql}}{1 - G_{d,j} G_{d,l}} R_e, \quad (35)$$

and, in turn,

$$R_{e_0 e_q} = \sum_{j=1}^{N+1} \sum_{l=1}^{N+1} \frac{\bar{E}_{d,0j} \bar{E}_{d,ql}}{1 - G_{d,j} G_{d,l}} R_e,$$

which can be used in (34) to quantify the bias in the estimate of  $\boldsymbol{\theta}^*$ . These results on the asymptotic expression of the bias can be used to quantify the performance of the proposed identification algorithm in many respects. In particular, one expects the asymptotic bias to depend on the sampling interval  $T$  and on the filter parameter  $\omega$ . In this connection, the influence of  $T$  and  $\omega$  on the quality of the computed estimates is analyzed in the following, for given values of  $N$  in the kernel (13). Focusing initially on the dependence on  $T$ , in view of the expressions in (29) for the parameters of the discrete-time filters  $F_{ij}(z)$ , it follows that

$$\bar{E}_{d,ij} = \frac{\bar{E}_{ij} T}{1 - G_j T} \simeq \frac{T}{1 - T}, \quad G_{d,j} = \frac{1}{1 - G_j T} \simeq \frac{1}{1 - T},$$

so that, again in terms of the  $T$ -dependence only, we obtain

$$\bar{E}_{d,0j} \bar{E}_{d,ql} \simeq \frac{T^2}{(1 - T)^2}, \quad G_{d,j} G_{d,l} \simeq \frac{1}{(1 - T)^2}.$$

Substituting the above relationships for  $\bar{E}_{d,0j} \bar{E}_{d,ql}$  and  $G_{d,j} G_{d,l}$  in (35) one can see that the scalar covariances have the following dependence on  $T$ :

$$\begin{aligned} R_{e_{0j} e_{ql}} &= \frac{\bar{E}_{d,0j} \bar{E}_{d,ql}}{1 - G_{d,j} G_{d,l}} R_e \simeq \frac{T^2}{(1 - T)^2} \frac{1}{1 - 1/(1 - T)^2} \\ &= \frac{T}{T - 2}, \end{aligned} \quad (36)$$

which clearly becomes smaller and smaller for decreasing values of  $T$ . Also, in view of the structure of the bias expression in (34), we conclude that  $\mathbb{E}[\Delta\boldsymbol{\theta}]$  decreases as  $T$  decreases (recall that  $\omega$  is given, as mentioned above). Note that this conclusion is in agreement with the simulation results

presented in Section VII, from which this effect is apparent. For the sake of generality, a complete expression for the covariance  $R_{e_{0j}e_{ql}}$  is reported, in which the dependence on  $\omega$  is made explicit, by substituting expressions for  $\bar{E}_{ij}$  and  $G_j$  in (29) (the derivation is omitted for brevity):

$$\mathbf{R}_{z_e e_{n+p}} = -\frac{1}{4}T\omega^2 \frac{(-11\omega T + 18\omega^2 T^2 + 6)}{(1 + 3\omega T)^2(3 + 2\omega T)(2 + \omega T)(1 + \omega T)}.$$

Concerning  $\mathbf{R}_{z_e e_{n+p}}$ , it is worth noting that:

- for a given value of  $\omega$ , the covariance vanishes (linearly) for decreasing values of  $T$  (in agreement with (36) and with the numerical evaluation shown in Fig. 1).
- For a given value of  $T$ , the covariance is a function of  $\omega$ , again as depicted in Fig. 1.

Finally, as a simple example of application of the above analysis, the covariances  $\mathbf{R}_{z_e}$  and  $\mathbf{R}_{z_e e_n}$ , which appear in the numerator of the bias expression in (34) are computed, for the case of  $n = 1$  and  $N = 2$ . The kernel (13) is used considering increasing values of  $\omega$  ranging from 0.1 to 10 and three choices for the sampling period  $T$ , namely  $10^{-3}s, 10^{-4}s$  and  $10^{-5}s$ . The results are summarised in Fig. 1, in which the dependence of the two covariances (normalised to the noise variance  $R_e$ ) on  $\omega$  and  $T$  is depicted. As can be clearly seen from the figure, the analysis confirms the numerical results in Section VII as far as the effect of  $T$  is concerned: both  $\mathbf{R}_{z_e}$  and  $\mathbf{R}_{z_e e_n}$  become negligible with respect to  $R_e$  for decreasing  $T$ ; in particular, for very small values of  $T$ , the result becomes almost insensitive to the value of  $\omega$ . Furthermore, as far as the dependence on  $\omega$  is concerned, the results in the figure indicate that smaller values of  $\omega$  appear to be more suitable to minimize the bias induced by the time-discretization.

*Remark 6.1:* It is worth noting that the choice of  $\omega$  and  $T$  is critical both in terms of bias (as well as on the bias/variance tradeoff) and in terms of transient behavior. In particular, it is important to underline that the tuning of the algorithm cannot be based on the sole bias analysis but has to take into account the impact of such choices on the variance of the estimates. Indeed, during transient operation the minimisation of the mean square error could be a more relevant requirement than unbiasedness. A detailed investigation on this important aspect, however, is out of the scope of the present paper. Nevertheless, in the next section, for given values of  $\omega$  and  $N$ , the effects of the choice of the sampling period  $T$  will be evaluated also in the context of comparative analysis with other techniques and the numerical results will confirm the key conclusion of this section, *i.e.*, that bias becomes negligible for fast sampling applications.

## VII. SIMULATION RESULTS

Consider the benchmark proposed by Rao and Garnier in [13] (see also [14] and [15])

$$\begin{cases} x^{(4)}(t) = a_1 x^{(3)}(t) + a_2 x^{(2)}(t) + a_3 x^{(1)}(t) + a_4 x(t) \\ \quad + b_1 u^{(1)}(t) + b_2 u(t), \\ y(t) = x(t) + \eta_y(t), \end{cases} \quad (37)$$

where  $x(0) = x_0$ ,  $x^{(1)}(0) = x_0^{(1)}$ ,  $x^{(2)}(0) = x_0^{(2)}$ ,  $x^{(3)}(0) = x_0^{(3)}$  and  $b_1 = -6400$ ,  $b_2 = 1600$ ,  $a_1 = -5$ ,

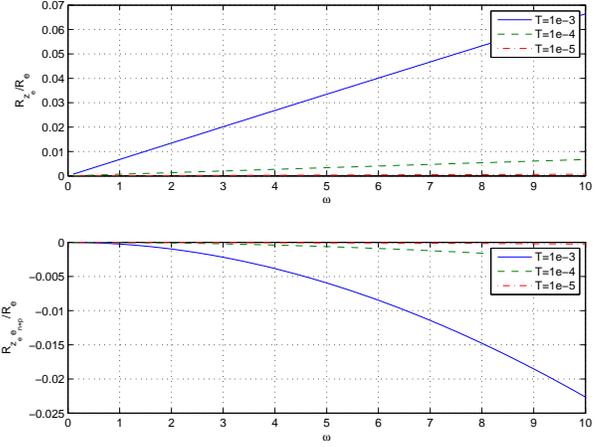


Fig. 1.  $\mathbf{R}_{z_e}$  and  $\mathbf{R}_{z_e e_{n+p}}$  (normalised with respect to  $R_e$ ) as functions of the kernel parameter  $\omega$  and of the sampling period  $T$ .

$a_2 = -408$ ,  $a_3 = -416$  and  $a_4 = -1600$ .  $y(\cdot)$  is the output of the system affected by an output measurement noise  $\eta_y(\cdot)$ , while  $u(\cdot)$  is the measured forcing input signal.

To carry out a simulation in the BC-NK framework, the value of  $N = 5$  is used in the implementation of the BC-NK kernel (13). The kernel parameter  $\omega$  is set to  $\omega = 1$ . As mentioned in Remark 6.1 an analysis concerning the best choice of this tuning parameter is out of the scope of the paper. However, in the following examples, different values of  $T$  are considered.

The procedure for constructing the auxiliary signals generation system by BC-NK kernels (depicted in Fig. 2) consists in taking the derivatives  $C_{\omega,N}^{(i)}(t,\tau)$ ,  $i \in \{1, 2, 3, 4, 5\}$  of the BC-NK (13), then in identifying the terms  $C_{\omega,N|i,j}$ , with  $j \in \{1, 2, 3, 4, 5, 6\}$  (see (14) and (15)), and finally in computing  $C_{\omega,N|i,j}(t,t)$  to form the  $\mathbf{E}_i(t)$  matrices (see (18)) needed for the implementation of the auxiliary signal generation system  $\mathcal{G}_{u,y \rightarrow z_e}$  (see (17)).

The initial conditions for system (37) have been set to  $x(0) = 1$ ,  $x^{(1)}(0) = 10$  and  $x^{(2)}(0) = x^{(3)}(0) = 0$ , while the forcing input has been chosen as a combination of sinusoidal signals:

$$u(t) = 10 \sin(5t) + 6 \sin(20t) + 3 \sin(8t) + \sin(2t) + 7 \sin(4t) + 9 \sin(12t).$$

The sampling period is set as  $T = 1 \cdot 10^{-4}s$  and the additive output measurement noise  $\eta_y(\cdot)$  is simulated as a uniformly distributed random signal taking values in the interval  $[-0.5, 0.5]$ , corresponding to a signal-to-noise ratio (SNR)  $SNR = 42.4152$ , where the SNR is defined as

$$SNR = 10 \log_{10} \left( \frac{P_y}{P_v} \right).$$

$P_v$  denotes the average power of the additive noise on the system output (*e.g.*, the variance), while  $P_y$  denotes the average power of the noise-free output fluctuations. As can be seen in Fig. 3, the BC-NK estimator shows very fast convergence and good robustness against the output measurement noise. Clearly, in a noisy scenario, the bias on the estimates is

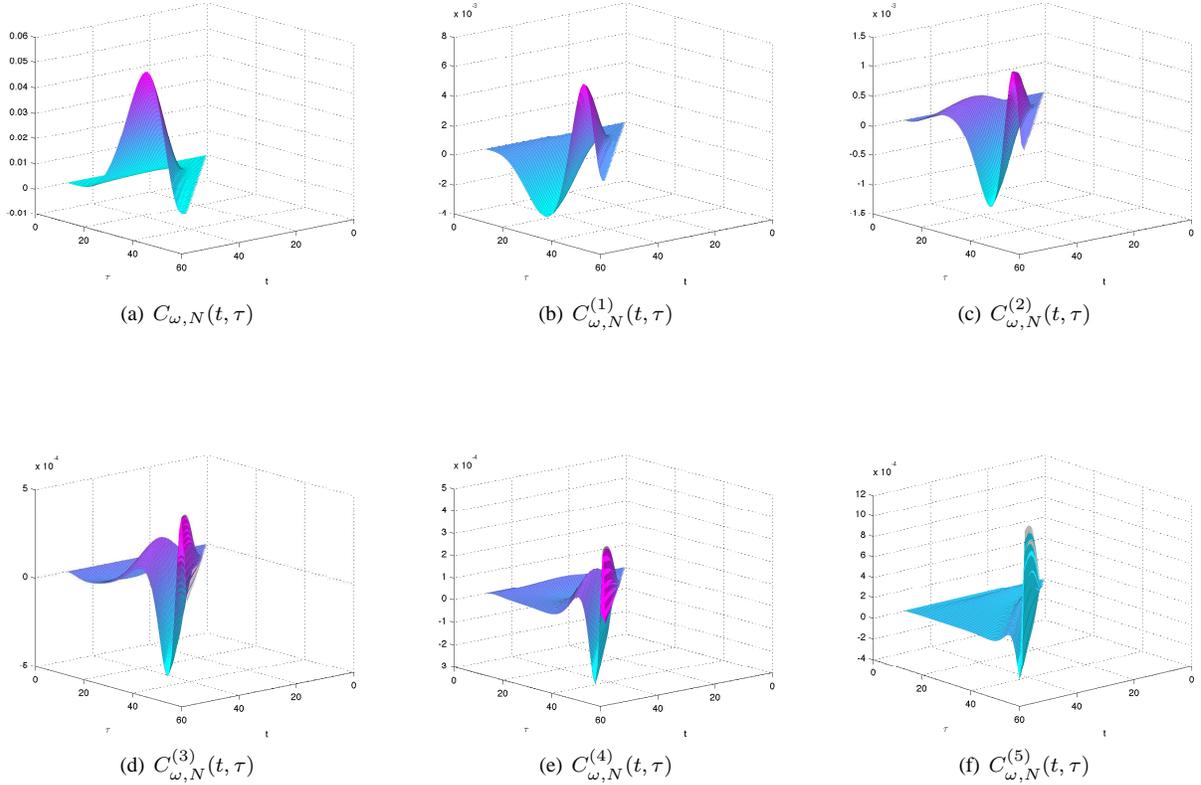


Fig. 2. Plots of the Bivariate Causal Non-Asymptotic Kernel (13) and its derivatives (see (14)), for  $\omega = 0.1$  and  $N = 5$ . The value of  $\omega$  is different from the one used for the simulations due to mere graphical rendering reasons.

unavoidable as illustrated in Section VI.

#### A. Comparative Analysis

Now, the proposed BC-NK parameter estimation method is compared with two well-known techniques from the literature, namely, the State Variable Filtering (SVF) method and the Integral Method (IM). More specifically, the BC-NK algorithm is compared with the Hartley Modulating Function method (HMF, see *e.g.*, [7]) and with the Simplified Refined Instrumental Variable one (SRIVC, see *e.g.*, [16]).

*Comparison with the Hartley Modulating Function method:* An important advantage of using Hartley-based modulating functions is that the system identification problem can be equivalently posed entirely in the frequency domain which makes it possible to exploit efficient DFT/FFT techniques. This method is well suited for digital implementation and has been included in the CONTSID toolbox ([17], [18]).

A Monte Carlo comparative analysis is carried out and the results (averaged over 200 runs) are presented in Table I for different values of the SNR. Note that in the table, each row provides an estimate of the unknown parameters for different levels of noise and sampling intervals  $T$ .

The table shows that the proposed technique yields accurate estimates of the parameters in several sampling and noise conditions, showing a better behavior with smaller values

of the sampling interval  $T$ , which is in agreement with the analysis provided in Section VI. Compared with the HMF technique, the BC-NK shows good robustness at decreasing values of the SNR and of the sampling interval.

*Comparison with the Refined Instrumental Variable Method:* The SRIVC method is a very powerful tool for the identification and estimation of continuous-time transfer function models. It was first suggested and implemented in [16], while the full RIVC has been implemented more recently (see [14]) and the reader is referred to this work for more details.

The comparison with the SRIVC method has been carried out under two different scenarios. In the first one, the asymptotic properties of the three methods (BC-NK, HMF and SRIVC) have been compared, as reported in Table I. The second scenario refers to the transient mode of behavior. More specifically, a simulation trial characterized by a  $SNR = 30$  and a sampling interval of  $T = 1 \cdot 10^{-4} s$  has been performed. The results concerning the parameter  $a_1$  are shown in Fig. 4 on a time-window of  $15s$  (the results are obtained computing the ergodic mean values over a number of 100 runs). Similar results are obtained concerning parameters  $a_2$ ,  $a_3$ ,  $a_4$ ,  $b_1$ , and  $b_2$  and they are not shown for the sake of brevity.

As can be noticed, the BC-NK technique shows, in the transient phase, very good results compared with SRIVC, while SRIVC shows slightly better performance asymptotically. In this regard, the good transient behavior is a relevant feature

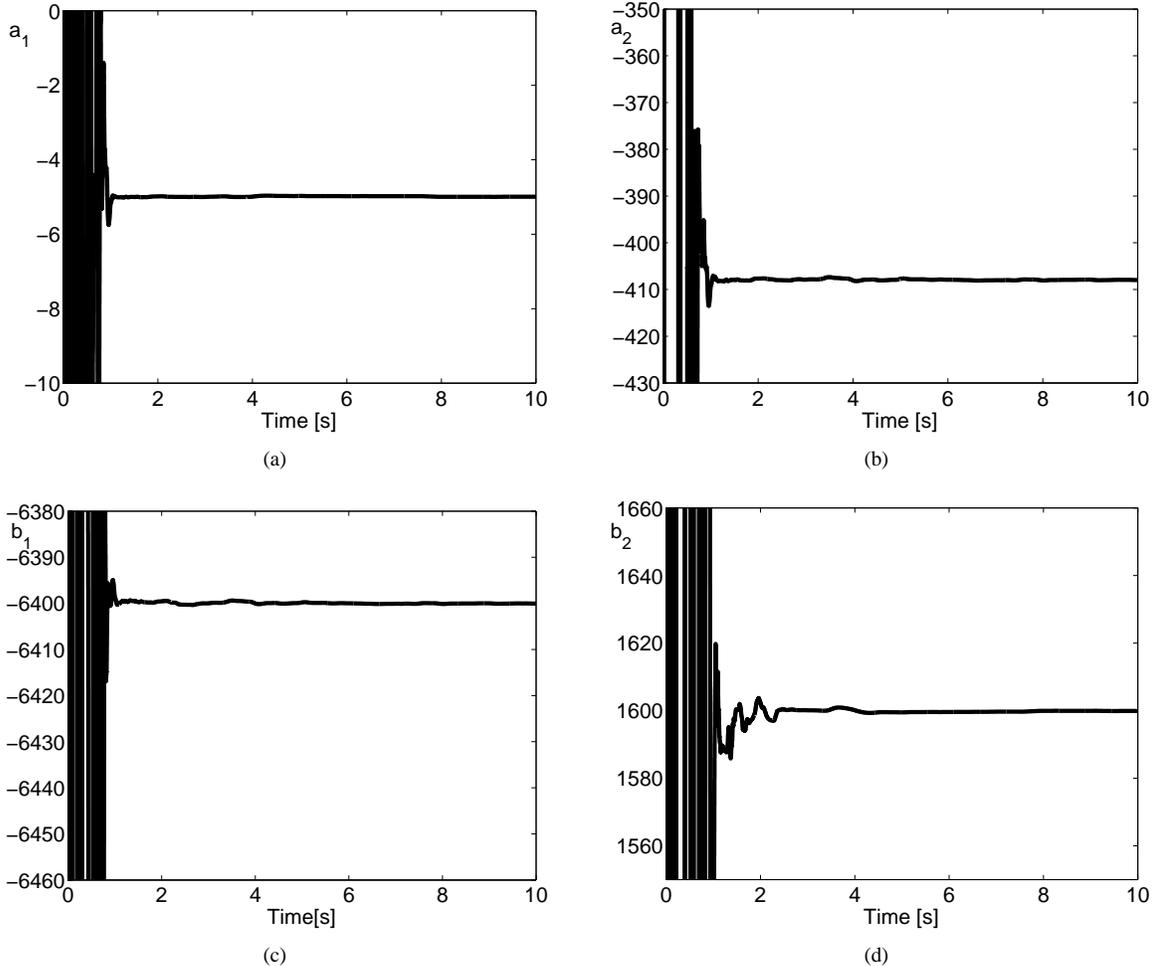


Fig. 3. (a)-(d): Time-behaviors of the estimates of parameters  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  provided by the BC-NK estimator in noisy conditions (the time-behaviors of the estimates of parameters  $a_3$  and  $a_4$  are similar).

of the proposed technique and enhances in simulation the effectiveness of the guaranteed internal stability properties of the estimator. For the sake of completeness, in Figs. 5 and 6 the behaviors of the standard deviations on the whole time-period and in the transient phase are shown with reference to the estimation of parameter  $a_1$  by the proposed method and by the SRIVC technique, respectively. Similar behaviors take place for the other parameters.

### VIII. CONCLUDING REMARKS

In this work, a novel framework arising from the algebra of linear integral operators has been established for the design of non-asymptotic parametric estimators for continuous-time linear dynamical systems. In particular, a new class of bivariate kernels has been devised allowing to get rid of the influence of the unknown initial conditions while, at the same time, guaranteeing the internal stability of the estimator. As a consequence, very fast convergence of the estimates can be achieved. The proposed estimation method has been thoroughly evaluated and compared with other techniques in simulation on a benchmark problem available in the literature. A parameter-bias analysis has also been provided.

Future research efforts will be devoted to a deeper analysis of the delicate issue of good (even optimal) selection of the tuning parameters and to an extensive robustness analysis. This will lead to the evaluation of the proposed technique in real-world scenarios. Moreover, future work will also aim at exploring the bias/variance tradeoff, as well as the behaviour of the algorithm in the presence of under-modelling, both for the deterministic and the stochastic parts of the model class.

### APPENDIX

#### A. Proof of Lemma 3.1

Integrating by parts, we have:

$$\begin{aligned} [V_K x^{(i)}](t) &= \int_0^t K(t, \tau) x^{(i)}(\tau) d\tau = x^{(i-1)}(t) K(t, t) \\ &\quad - x^{(i-1)}(0) K(t, 0) \int_0^t K^{(1)}(t, \tau) x^{(i-1)}(\tau) d\tau. \end{aligned} \quad (38)$$

Sampling Interval	Method	SNR	Param. value & st. dev.	Parameters					
				$a_1 = -5$	$a_2 = -408$	$a_3 = -416$	$a_4 = -1600$	$b_1 = -6400$	$b_2 = 1600$
$1 \cdot 10^{-3}$	BC-NK	20	$\hat{\theta}^*$	-4.685	-405.135	-413.231	-1568.656	-6250.399	1493.692
			$\sigma_{\theta^*}$	0.0005	0.1923	0.3285	3.6495	57.1571	34.9348
		30	$\hat{\theta}^*$	-4.919	-407.179	-414.543	-1597.616	-6384.814	1585.849
			$\sigma_{\theta^*}$	0.0000	0.0041	0.079	0.790	3.2375	5.0161
		35	$\hat{\theta}^*$	-4.957	-407.759	-415.413	-1599.539	-6394.814	1585.910
			$\sigma_{\theta^*}$	0.0000	0.0027	0.0048	0.0483	0.796	0.6702
$1 \cdot 10^{-3}$	HMF	20	$\hat{\theta}^*$	-5.079	-408.883	-450.807	-1666.169	-6389.525	848.298
			$\sigma_{\theta^*}$	0.0005	0.5547	173.8672	577.4829	375.3087	546.8073
		30	$\hat{\theta}^*$	-5.008	-408.113	-419.262	-1604.249	-6398.651	1519.295
			$\sigma_{\theta^*}$	0.0000	0.0193	4.3941	13.1635	12.1270	246.4390
		35	$\hat{\theta}^*$	-5.002	-408.036	-416.442	-1599.270	-6399.602	1582.447
			$\sigma_{\theta^*}$	0.0000	0.0134	2.8191	8.3934	7.8879	157.4980
$1 \cdot 10^{-3}$	SRIVC	20	$\hat{\theta}^*$	-5.001	-407.811	-415.741	-1591.133	-6380.437	1561.847
			$\sigma_{\theta^*}$	0.0001	0.0761	0.27529	2.1745	16.0868	29.7428
		30	$\hat{\theta}^*$	-5.001	-407.665	-415.549	-1602.287	-6388.085	1589.601
			$\sigma_{\theta^*}$	0.0000	0.0006	0.0039	0.0246	0.3530	0.4893
		35	$\hat{\theta}^*$	-5.000	-407.968	-415.953	-1600.158	-6397.152	1598.641
			$\sigma_{\theta^*}$	0.0000	0.0000	0.0020	0.0157	0.9927	0.2435
$1 \cdot 10^{-4}$	BC-NK	20	$\hat{\theta}^*$	-4.919	-406.793	-414.791	-1594.948	-6387.959	1579.771
			$\sigma_{\theta^*}$	0.0004	0.0155	0.2131	0.6797	10.0588	12.3100
		30	$\hat{\theta}^*$	-4.976	-407.453	-415.750	-1598.907	-6398.317	1594.118
			$\sigma_{\theta^*}$	0.0000	0.0006	0.0012	0.0009	0.1722	0.3065
		35	$\hat{\theta}^*$	-4.986	-407.871	-415.875	-1599.443	-6399.297	1596.752
			$\sigma_{\theta^*}$	0.0000	0.0018	0.0013	0.0309	0.4501	0.1142
$1 \cdot 10^{-4}$	HMF	20	$\hat{\theta}^*$	-5.005	-408.180	-418.937	-1617.120	-6401.134	1440.808
			$\sigma_{\theta^*}$	0.0000	0.02826	12.0170	54.754	32.2305	182.5871
		30	$\hat{\theta}^*$	-4.995	-408.001	-413.943	-1598.309	-6401.138	1608.519
			$\sigma_{\theta^*}$	0.0000	0.0018	0.6144	2.7075	1.7241	82.8194
		35	$\hat{\theta}^*$	-4.993	-407.967	-412.914	-1594.777	-6401.000	1641.983
			$\sigma_{\theta^*}$	0.0000	0.0147	0.404	3.853	2.135	68.173
$1 \cdot 10^{-4}$	SRIVC	20	$\hat{\theta}^*$	4.969	407.418	416.319	1597.030	6408.718	1578.066
			$\sigma_{\theta^*}$	0.001	0.069	3.820	9.750	8.391	54.424
		30	$\hat{\theta}^*$	4.987	407.762	416.191	1599.706	6401.018	1597.049
			$\sigma_{\theta^*}$	0.000	0.001	0.059	0.474	0.130	1.921
		35	$\hat{\theta}^*$	4.999	407.972	416.182	1599.816	6400.090	1597.676
			$\sigma_{\theta^*}$	0.000	0.001	0.072	0.058	0.202	1.218
$5 \cdot 10^{-5}$	BC-NK	20	$\hat{\theta}^*$	-5.009	-407.515	-415.742	-1598.016	-6395.546	1589.910
			$\sigma_{\theta^*}$	0.000	0.021	0.007	0.186	2.697	6.828
		30	$\hat{\theta}^*$	-5.003	-407.892	-415.966	-1599.553	-6399.711	1596.998
			$\sigma_{\theta^*}$	0.000	0.018	0.002	0.048	1.114	2.609
		35	$\hat{\theta}^*$	-5.002	-407.944	-415.986	-1599.769	-6399.721	1598.333
			$\sigma_{\theta^*}$	0.000	0.002	0.001	0.031	0.224	0.450
$5 \cdot 10^{-5}$	HMF	20	$\hat{\theta}^*$	-5.007	-408.154	-415.087	-1599.959	-6399.602	1568.693
			$\sigma_{\theta^*}$	0.000	0.0322	1.363	10.116	21.364	61.407
		30	$\hat{\theta}^*$	-4.996	-407.999	-412.984	-1593.468	-6402.867	1643.441
			$\sigma_{\theta^*}$	0.000	0.013	0.651	2.629	5.542	46.590
		35	$\hat{\theta}^*$	-4.993	-407.967	-412.404	-1592.121	-6401.965	1660.997
			$\sigma_{\theta^*}$	0.000	0.005	0.228	1.712	3.605	58.242
$5 \cdot 10^{-5}$	SRIVC	20	$\hat{\theta}^*$	-4.989	-407.776	-413.612	-1592.580	-6404.072	1643.590
			$\sigma_{\theta^*}$	0.002	0.277	2.846	7.706	4.557	31.112
		30	$\hat{\theta}^*$	-5.002	-407.981	-415.761	-1599.715	-6400.839	1598.274
			$\sigma_{\theta^*}$	0.000	0.068	0.567	1.425	1.651	2.902
		35	$\hat{\theta}^*$	-5.002	-408.051	-416.177	-1600.253	-6400.695	1598.189
			$\sigma_{\theta^*}$	0.000	0.002	0.039	0.114	0.680	0.842

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE ESTIMATED PARAMETERS CALCULATED USING BC-NK, HMF AND SRIVC APPROACHES, FOR DIFFERENT SAMPLING INTERVALS AND SNRS.

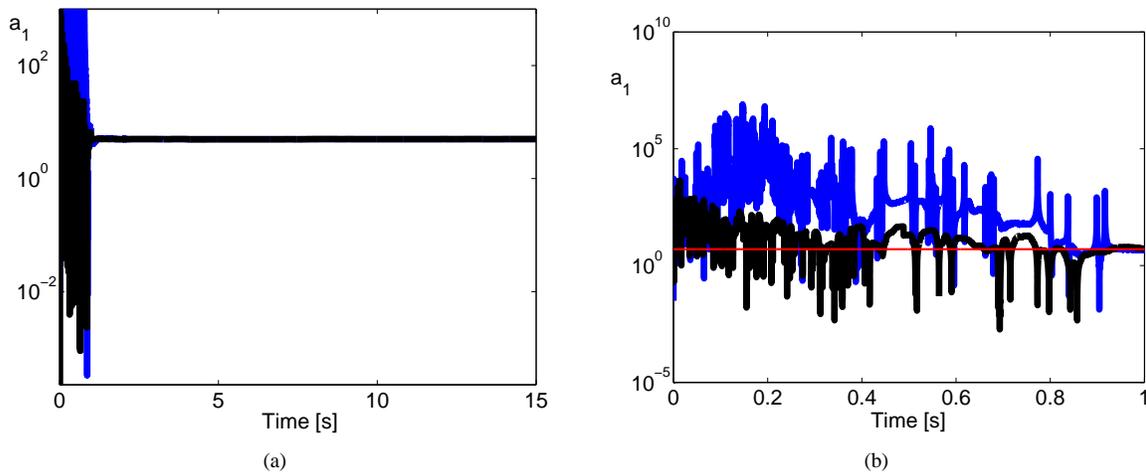


Fig. 4. (a) Parameter  $a_1 = -5$  estimated by the SRIVC method (blue) and by the BC-NK estimator (black); the horizontal line (red) denotes the true parameter value. (b) Enhancement of the transient modes of behavior.

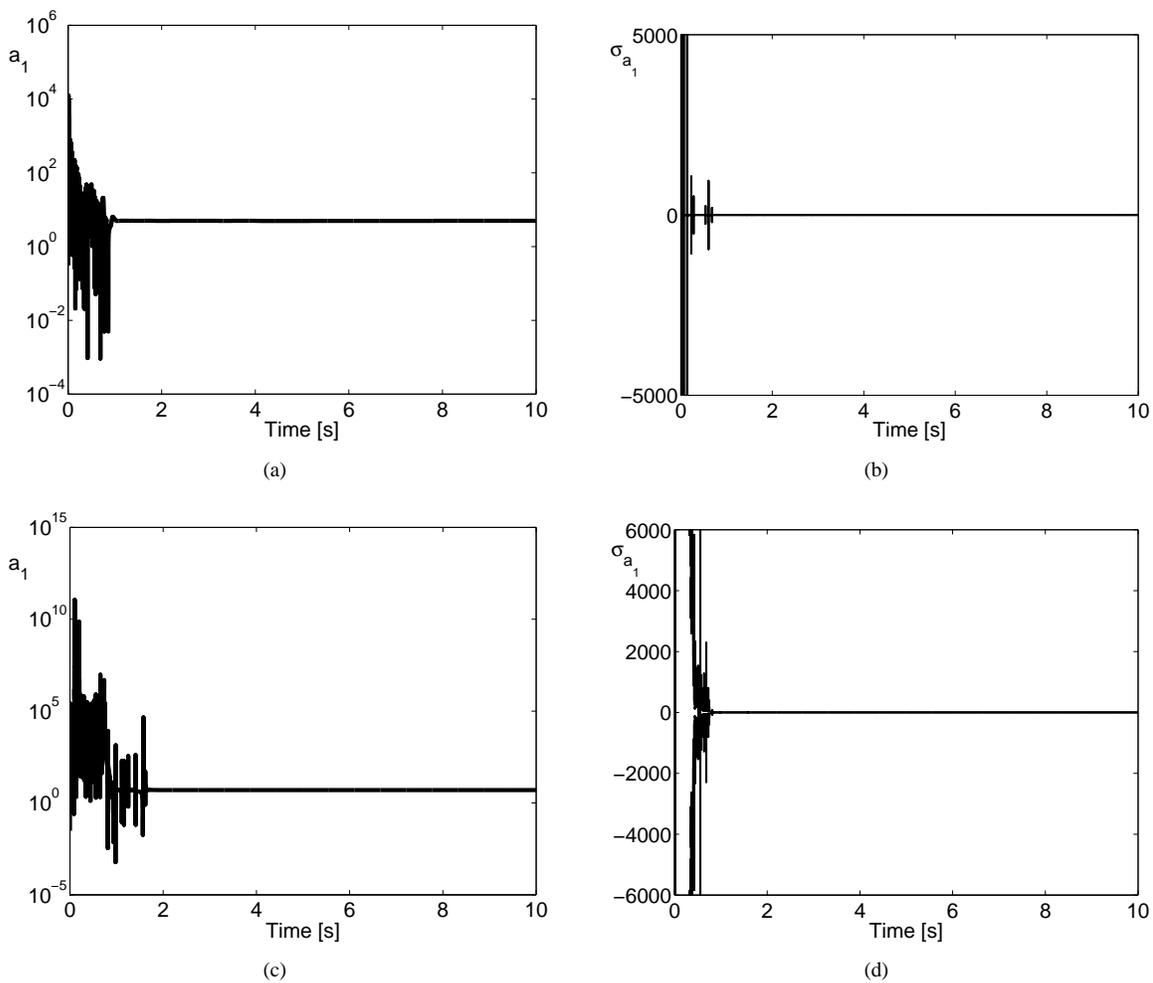


Fig. 5. (a) Behavior of the estimate of  $a_1$  by the BC-NK estimator and (b) respective standard-deviation. (c) Behavior of the estimate of  $a_1$  by the SRIVC estimator and (d) respective standard-deviation.

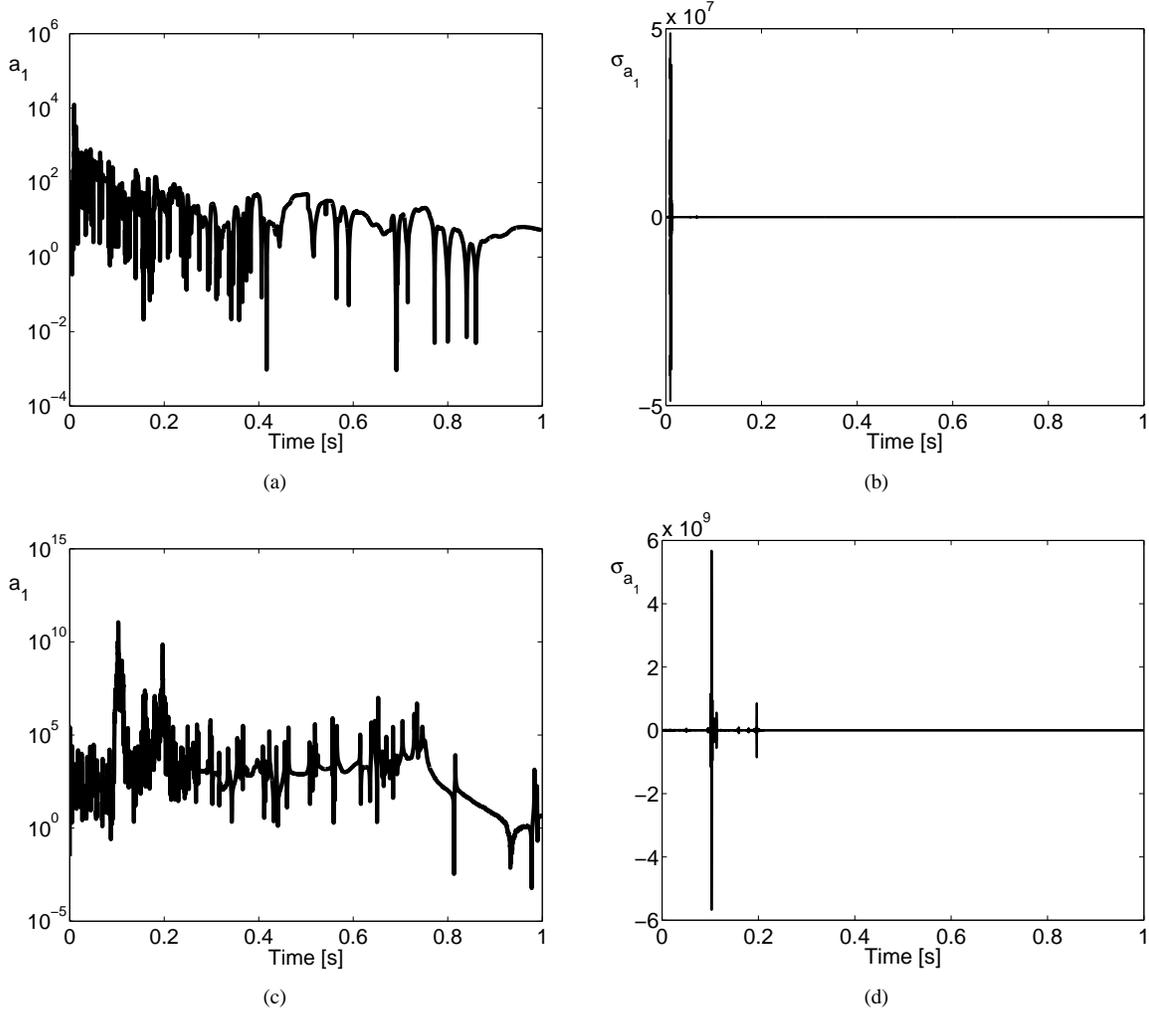


Fig. 6. (a) Transient behavior of the estimate of  $a_1$  by the BC-NK estimator and (b) respective standard-deviation. (c) Transient behavior of the estimate of  $a_1$  by the SRIVC estimator and (d) respective standard-deviation.

The integral operator on the right-hand side of (38) can be further split by parts:

$$\begin{aligned}
 - \int_0^t K^{(1)}(t, \tau) x^{(i-1)}(\tau) d\tau &= -x^{(i-2)}(t) K^{(1)}(t, t) \\
 &+ x^{(i-2)}(0) K^{(1)}(t, 0) + \int_0^t K^{(2)}(t, \tau) x^{(i-2)}(\tau) d\tau.
 \end{aligned}$$

Proceeding by induction we obtain

$$\begin{aligned}
 \int_0^t K(t, \tau) x^{(i)}(\tau) d\tau &= \sum_{j=1}^i (-1)^{j+1} x^{(i-j)}(t) K^{(j-1)}(t, t) + \\
 &\sum_{j=1}^i (-1)^j x^{(i-j)}(0) K^{(j-1)}(t, 0) + (-1)^i \int_0^t K^{(i)}(t, \tau) x(\tau) d\tau
 \end{aligned}$$

that is, the function obtained by applying the Volterra operator to the  $i$ -th derivative is non-anticipative with respect to lower-order derivatives. The proof is concluded by rearranging the indexing of the summation in the above expression.

## B. Composition of Volterra operators

The composition of two integral Volterra operators yields, in turn, a Volterra integral operator (see [9]). The derivation is non-trivial due to the necessity of considering explicitly the integration over finite domains. Suppose that  $V_{K_h}$  and  $V_{K_g}$  are two Volterra operators induced by the  $\mathcal{HS}$  kernels  $K_h$  and  $K_g$  respectively. By introducing the Fredholm extension of the Volterra kernel  $\tilde{K}_g(\sigma, \tau) \triangleq K_g(\sigma, \tau)H(\tau)H(\sigma - \tau)$ , we get:

$$\begin{aligned}
 [V_{K_h} [V_{K_g} x]](t) &= \int_0^t K_h(t, \sigma) \left( \int_0^\sigma K_g(\sigma, \tau) x(\tau) d\tau \right) d\sigma \\
 &= \int_0^t K_h(t, \sigma) \int_0^\sigma \tilde{K}_g(\sigma, \tau) x(\tau) d\tau d\sigma \\
 &= \int_0^t (K_h \circ K_g)(t, \tau) x(\tau) d\tau, \quad (39)
 \end{aligned}$$

where the kernel of the composed integral operator can be thus obtained by the *kernel composition integral*  $(\cdot \circ \cdot)$ , defined as:

$$(K_h \circ K_g)(t, \tau) \triangleq \int_\tau^t K_h(t, \sigma) K_g(\sigma, \tau) d\sigma. \quad (40)$$

From the causality of the Volterra operator, we have  $(K_h \circ K_g)(t, t) = 0, \forall t \in \mathbb{R}_{\geq 0}$ . Moreover, if for some  $i \in \mathbb{N}$ ,  $K_g^{(i)}(t, 0) = 0, \forall t \in \mathbb{R}_{\geq 0}$ , then  $(K_h \circ K_g)^{(i)}(t, 0) = 0, \forall t$ .

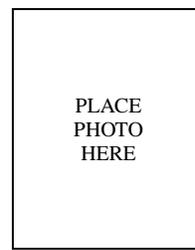
#### ACKNOWLEDGMENT

The authors wish to thank Prof. Peter Young for useful feedback and suggestions on an early draft of the paper.

#### REFERENCES

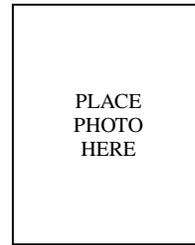
- [1] L. Ljung, *System Identification - Theory for the user*. Englewood Cliffs: Prentice-Hall, 1987.
- [2] T. Söderström and P. Stoica, *System Identification*. New York: Prentice-Hall, 1989.
- [3] P. Young, *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Berlin: Springer-Verlag, 2011.
- [4] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Prentice-Hall, 1994.
- [5] P. Young, "The refined instrumental variable method: unified estimation of discrete and continuous-time transfer function models," *Journal Europ'een des Syst'emes Automatis'es*, vol. 42, no. 1, pp. 149–179, 2008.
- [6] M. Bergamasco and M. Lovera, "Continuous-time predictor-based subspace identification using Laguerre filters," *IET Control Theory and Applications*, vol. 5, no. 7, pp. 856–867, 2011.
- [7] H. Garnier and L. Wang, Eds., *Identification of Continuous-time Models from Sampled Data*. Advances in Industrial Control: Springer, 2008.
- [8] H. Garnier, T. Soderstrom, and J. Yuz, "Editorial: Continuous-time model identification," *IET Control Theory and Applications*, vol. 5, no. 7, 2011.
- [9] T. Burton, *Volterra Integral and Differential Equations*. Elsevier, 2005.
- [10] A. Feintuch and R. Saeks, *System Theory. A Hilbert Space Approach*. New York: Academic Press, 1982.
- [11] R. Brockett, *Finite Dimensional Linear Systems*. New York: Wiley, 1970.
- [12] L. Ljung and A. Wills, "Issues in sampling and estimating continuous-time models with stochastic disturbances," *Automatica*, vol. 46, no. 5, pp. 925–931, 2010.
- [13] G. P. Rao and H. Garnier, "Numerical illustrations of the relevance of direct continuous-time model identification," in *15th Triennial IFAC World Congress on Automatic Control*, Barcelona, Spain, 2002.
- [14] V. Laurain, H. Garnier, M. Gilson, and P. Young, "Refined instrumental variable method for identification of Hammerstein continuous-time box-jenkins models," in *Proc. of IEEE Conference on Decision and Control*, Cancun, 2008, pp. 1386–1391.
- [15] L. Ljung, "Initialisation aspects for subspace and output-error identification methods," in *Proc. European Control Conference, ECC 03*, J. Maciejowski, Ed., Cambridge, UK, Sept. 2003.
- [16] P. Young and A. Jakeman, "Refined instrumental variable methods of time-series analysis: part III, extensions," *International Journal of Control*, vol. 31, pp. 741–764, 1980.
- [17] H. Garnier, M. Gilson, and E. Huesestein, "Developments for the matlab CONTSID toolbox," in *Proc. IFAC Symp. System Identification American Control Conference*, 2003.
- [18] H. Garnier and M. Mensler, "The CONTSID toolbox: A Matlab toolbox for continuous-time system identification," in *Proc. 12th IFAC Symp. System Identification*, 2000.

PLACE  
PHOTO  
HERE



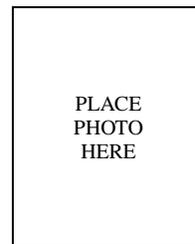
**Gilberto Pin** received the Laurea (M.Sc.) degree in Electrical Engineering (with honors) and the Ph.D. in Information Engineering from the University of Trieste, Italy, in 2005 and 2009, respectively. From 2009 to 2012 he was an Automation Engineer at Danieli Automation S.p.A., Italy. Since 2013, he has been a Control Systems Engineer at the R&D Dept. of Electrolux Professional S.p.A., Italy. He is author of several papers published in international conferences and scientific journals concerning systems theory, signal processing methods and control applications. He is a co-recipient of the IFAC Best Application Paper Prize of the Journal of Process Control, Elsevier, for the three-year period 2011-2013. His current research interests include networked control, model predictive control, advanced system's identification methods and deadbeat observers design. His activity is also devoted to the industrial application of advanced control techniques

PLACE  
PHOTO  
HERE



**Andrea Assalone** received the Laurea (M.Sc.) degree in Automation Engineering and the Ph.D in Information Engineering from the University of Trieste, Italy, in 2010 and 2014, respectively. He is currently with the Institute for Design and Control at the Johannes Kepler University, Linz, Austria. His research interests include system's identification methods of linear systems and data-based modeling and control of biomedical systems.

PLACE  
PHOTO  
HERE



**Marco Lovera** is a Professor of Automatic Controls at the Politecnico di Milano. He received the MSc in Electronic Engineering from the Politecnico di Milano in 1993 and the PhD from the same University in 1998. After a one-year period in industry he joined in 1999 the Dipartimento di Elettronica, Informazione e Bioingegneria of the Politecnico di Milano. Presently (from January 2015), he is with the Dipartimento di Scienze e Tecnologie Aerospaziali of the Politecnico di Milano, where he leads the Aerospace Systems and Control Laboratory (ASCL).

Main research interests: model identification and aerospace applications of advanced systems and control concepts, with specific reference to spacecraft attitude determination and control, rotorcraft modelling and control, small scale autonomous vehicles. He is currently Associate editor of *Automatica*, member of the Editorial Board of *IET Control Theory and Applications*, Corresponding Editor of *IEEE Control Systems* and arXiv moderator for the categories "Control and Optimisation" and "Systems and Control". He is an elected member of the Board of Governors of the IEEE CSS.

PLACE  
PHOTO  
HERE

**Thomas Parisini** received the Ph.D. degree in Electronic Engineering and Computer Science in 1993 from the University of Genoa. He was with Politecnico di Milano and since 2010 he holds the Chair of Industrial Control at Imperial College London. Since 2001 he is also Danieli Endowed Chair of Automation Engineering with University of Trieste. In 2009-2012 he was Deputy Rector of University of Trieste. He authored or co-authored more than 250 research papers in archival journals, book chapters, and international conference proceedings. His

research interests include neural-network approximations for optimal control problems, fault diagnosis for nonlinear and distributed systems, nonlinear model predictive control systems and nonlinear estimation. He is a co-recipient of the IFAC Best Application Paper Prize of the Journal of Process Control, Elsevier, for the three-year period 2011-2013 and of the 2004 Outstanding Paper Award of the IEEE Trans. on Neural Networks. He is also a recipient of the 2007 IEEE Distinguished Member Award. He is involved as Project Leader in several projects funded by the European Union, by the Italian Ministry for Research, and he is currently leading consultancy projects with some major process control companies (Danieli, Duferco, Electrolux, among others). In 2012 he was awarded a prestigious ABB Research Grant dealing with energy-autonomous sensor networks for self-monitoring industrial environments. Thomas Parisini is the Editor-in-Chief of the IEEE Trans. on Control Systems Technology. He is also the Chair of the IFAC Technical Committee on Fault Detection, Supervision & Safety of Technical Processes - SAFEPROCESS. He was the Chair of the IEEE Control Systems Society Conference Editorial Board and a Distinguished Lecturer of the IEEE Control Systems Society . He was an elected member of the Board of Governors of the IEEE Control Systems Society and of the European Control Association (EUCA) and a member of the board of evaluators of the 7th Framework ICT Research Program of the European Union. Thomas Parisini is currently serving as an Associate Editor of the Int. J. of Control and served as Associate Editor of the IEEE Trans. on Automatic Control, of the IEEE Trans. on Neural Networks, of Automatica, and of the Int. J. of Robust and Nonlinear Control. Among other activities, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and a General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a Fellow of the IEEE.