Kernel-Based Simultaneous Parameter-State Estimation for Continuous-Time Systems

Peng Li, Francesca Boem, Gilberto Pin, Thomas Parisini

Abstract—In this note, the problem of jointly estimating the state and the parameters of continuous-time systems is addressed. Making use of suitably designed Volterra integral operators, the proposed estimator does not need the availability of time-derivatives of the measurable signals and the dependence on the unknown initial conditions is removed. As a result, the estimates converge to the true values in arbitrarily short time in noise-free scenario. In the presence of bounded measurement and process disturbances, the estimation error is shown to be bounded. The numerical implementation aspects are dealt with and extensive simulation results are provides showing the effectiveness of the estimator.

I. Introduction

In many practical engineering problems, due to incomplete knowledge of the system, both the parameters and the state variables are usually uncertain and need to be estimated. Examples can be seen in power systems [8], mechanical engineering [9], [13] and architecture [10].

One of the early approaches to solve the problem of jointly estimating the parameters and the state of linear systems consisted of using augmented observers, in which the unknown parameters are treated as additional state variables that enter nonlinearly in the system's dynamics [11]. However, the augmented nonlinear dynamics result in unnecessary increase of the complexity and computational burden. Alternatively, a hierarchical methodology has been explored to separate the problem into two interconnected parts: parameter estimation and state estimation. Iterative corrections between the estimators are implemented using Kalman filters [12], adaptive methods [9], [10] and so on. Besides, in [25], parameters and states are estimated through an EM or Rauch-Tung-Striebel process with iterative correction. However, the aforementioned methods can only achieve asymptotic convergence while in many time-critical applications, such as medical diagnosing systems [13] and power systems [14], it is often desirable to obtain accurate and simultaneous state-parameters estimates

P. Li is with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK (`lipenglp1991@126.com`).

F. Boem is with Dept. of Electrical and Electronic Engineering at the University College London, UK (`f.boem@ucl.ac.uk`).

G. Pin is with Electrolux Italia S.p.A, Italy, e-mail: (`gilbertopin@alice.it`).

T. Parisini is with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK, with the KIOS Research and Innovation Centre of Excellence, University of Cyprus, and also with the Dept. of Engineering and Architecture at the University of Trieste, Italy (`t.parisini@gmail.com`).

in finite or fixed time. The finite-time estimation has been gaining research attention in the recent decades. Sliding mode methods [9], [15] rely on discontinuous high-gain injection to achieve the finite-time convergence, which makes this class of estimators quite sensitive to the measurement noise. Alternatively, algebraic and modulating function (MF) methods are proposed to deal with the unmeasurable signal derivatives and to remove the effects of the initial conditions, thus achieving the finite or fixed-time convergence (see [16], [17]). It can also be seen in [5] that the modulating theorem is a key tool to avoid explicit time-differentiation and achieve the identification of a continuous-time stochastic state-space model. More importantly, MF methods are proven to be effective in practical applications, such as in smart grids [14] and for heating ventilation and air conditioning systems [1]. It is worth noting that, for algebraic and some univariate MF-based approaches, integrals with expanding window suffer from internal instability that may cause error accumulation as shown in [2]. To address this issue, moving-horizon integrals are adopted as in [3], [18]. Alternatively, periodic resettings, also known as periodic deflation, can be useful as shown in [14], [17] to avoid error-accumulation.

Furthermore, a methodology based on Volterra operators has been developed to solve the specific tasks of parameter estimation [19] and state estimation [20], separately. With properly shaped kernel functions, the Volterra operators avoid the explicit time-differentiation, annihilating, at the same time, the effects of the initial conditions, thus attaining deadbeat convergence. Moreover, thanks to the specific class of kernel functions defined therein, the Volterra operators are inherently endowed with BIBO stability and can be realized by internally stable LTV systems, not requiring neither periodic resetting neither moving window implementation.

In this note (see [22] for some preliminary results), a joint parameter-state estimator based on Volterra operators is illustrated, providing simultaneous estimates of parameters and state variables with fixed-time convergence. Differently from classical hierarchical methods, the proposed methodology is non-iterative. Furthermore, the persistency of excitation (PE) conditions, the robustness against bounded measurement noise, process disturbance and digital implementation are characterized analytically. Extensive simulation results are also provided to show the effectivess of the proposed estimation methodology.

II. Problem statement

Consider a Multi-Input-Single-Output (MISO) system in the input-output form:

$$y^{(n)}(t) = \sum_{i=0}^{n-1} a_i y^{(i)}(t) + \sum_{k=0}^{p-1}\sum_{j=0}^{m_k-1} b_{k,j} u_k^{(j)}(t) \qquad (1)$$

with unknown initial conditions $y^{(i)}(0)$, $i \in \{0,\ldots,n-1\}$ and $u_k^{(j)}(0)$, $j \in \{0,\ldots,m_k-1\}$, $k \in \{0,\ldots,p-1\}$ where n, $m_k \in \mathbb{Z}_{>0}$, $m_k \leq n$, $\forall k$. The parameters $\{a_i\}$, $\{b_{k,j}\}$ are assumed to be constant but unknown. Only the input and output signals, i.e. $u_k(t), k \in \{0,...,p-1\}$ and $y(t)$, are assumed to be available.

Consider now the observer canonical form for system (1):

$$\begin{cases} z^{(1)}(t) &= Az(t) + Bu(t) \\ y(t) &= c^\top z(t) \end{cases} \qquad (2)$$

where $z(t) \triangleq [z_0(t) z_1(t) \ldots z_{n-1}(t)]^\top \in \mathbb{R}^n$ is the state vector and the input vector is defined as $u(t) \triangleq [u_0(t), u_1(t), \ldots, u_{p-1}(t)]$. $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$ and $c \in \mathbb{R}^n$ are given by:

$$A = \begin{bmatrix} a_{n-1} & 1 & 0 & \cdots & 0 \\ a_{n-2} & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ a_1 & 0 & \ldots & 0 & 1 \\ a_0 & 0 & \cdots & 0 & 0 \end{bmatrix}, c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$B = [b_0, b_1, \ldots, b_{p-1}]$, $b_k = [0,\ldots,0, b_{k,m_k-1},\ldots,b_{k,0}]^\top$, $\forall k \in \{0,...,p-1\}$. The state dynamics of (2) can be written as

$$z_r(t) = y^{(r)}(t) - \sum_{j=0}^{r-1} a_{n-r+j} y^{(j)}(t) - \sum_{k=0}^{p-1}\sum_{j=0}^{r-1+p_k} b_{k,n-r+j} u_k^{(j)}(t)$$

where $p_k \triangleq n - m_k$ and the convention $\sum_{j=0}^{k}\{\cdot\} = 0$, for $k < 0$ is used. The proposed approach aims at simultaneously estimating the state vector $z(t)$ and the unknown parameters $\{a_i\}$ and $\{b_{k,j}\}$.

III. Fixed-time parameter-state joint estimation

The key tool used by the proposed algorithm to achieve non-asymptotic convergence is the Volterra integral operator V_K. It maps the signal $x(t)$ to the transformed one: $[V_K x](t) \triangleq \int_0^t K(t,\tau)x(\tau)d\tau$, $t \in \mathbb{R}_{\geq 0}$, with $K(t,\tau)$ being an nth order non-asymptotic kernel satisfying the non-asymptotic condition $K^{(i)}(t,0) = 0$, $\forall i \in \{0,\ldots,n-1\}$. For further details about the algebra of the Volterra integral operator, readers are suggested to refer to [19], [20] and the references therein.

The kernel adopted in this paper is the Bivariate Feedthrough Non-asymptotic Kernel (BF-NK proposed in [20]) function $K_h(t,\tau)$ having the form

$$K_h(t,\tau) = e^{-\omega_h(t-\tau)}\left(1 - e^{-\bar{\omega}t}\right)^N, \qquad (3)$$

where $N \in \mathbb{Z}_{\geq n}$, $\omega_h, \bar{\omega} \in \mathbb{R}_{>0}$ are the design tuning parameters. Remarkably, the kernel (3) satisfies $K^{(i)}(t,0) = 0, \forall t \in \mathbb{R}_{\geq 0}, i \in \{0,\ldots,N-1\}$, that allows the Volterra transformation of the signal derivative $x^{(i)}(t)$ to become

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

where $K^{(i)}(t,\tau)$ denotes the ith partial derivative of $K(t,\tau)$ with respect to the second argument. As a result, the estimator can get rid of explicit time-differentiation without the effects of the unknown initial conditions. For more details regarding the non-asymptotic property of the BF-NK, readers can refer to [20], [21] and the reference therein. It is worth noting that the tuning parameter ω_h acts as a forgetting factor that reduces the error accumulation in the noisy scenario.

Applying the BF-NK to (1) and after some quite lengthy algebra, we get

$$(-1)^{n-1}[V_{K_h^{(n)}}y](t) = -\sum_{i=0}^{n-1} a_i(-1)^i[V_{K_h^{(i)}}y](t)$$
$$-\sum_{k=0}^{p-1}\sum_{i=0}^{m_k-1} b_{k,i}(-1)^i[V_{K_h^{(i)}}u_k](t) + \sum_{r=0}^{n-1}\gamma_h(t)z_r(t),$$
$$(4)$$

where $\gamma_{h,r}(t) = (-1)^{n-r-1}K_h^{(n-r-1)}(t,t)$. In turn, (4) can be rearranged as a vector equation

$$(-1)^{n-1}[V_{K_h^{(n)}}y](t) = \nu_h(t)^\top\theta(t), \qquad (5)$$

where $\theta(t) \triangleq [a_0,\ldots,a_{n-1}, b_{0,0},\ldots,b_{0,m_0-1},\ldots, b_{p-1,m_{p-1}-1}, z_0(t),\ldots,z_{n-1}(t)]$ and $\nu_h(t) \triangleq [-[V_{K_h}y](t), \ldots, (-1)^n[V_{K_h^{(n-1)}}](t), -[V_{K_h}u_0](t),\ldots, (-1)^{m_{p-1}}[V_{K_h^{(m_{p-1}-1)}}u_{p-1}](t), \gamma_{h,n-1}(t),\ldots,\gamma_{h,0}(t)]^\top$.

Note that the vector $\nu_h(t)$ consists of the derivatives of the kernel $K_h(t,\tau)$ and the Volterra images $\xi_{h,i,\star}(t) \triangleq [V_{K_h^{(i)}}\star](t)$ which can be calculated from the I/O by a dynamic system

$$\xi_{h,i,\star}^{(1)}(t) = -\omega_h\xi_{h,i,\star}(t) + K_h(t,t)\star(t), \qquad (6)$$

where $\star$ represents signals $y(t)$ and u_k, $\forall k \in \{0,\ldots,p-1\}$.

In order to solve for $n_\delta \triangleq 2n + \sum_{k=0}^{p-1} m_k$ unknowns, n of which are time-varying, an augmentation of (5) is necessary. Therefore, we use n_δ BF-NKs tuned by a common $\bar{\omega}$ but different ω_h, $h \in \{0,\ldots,n_\delta-1\}$. As a result, the problem consists in solving the following matrix equation for θ:

$$\kappa(t) = \Gamma(t)\theta, \qquad (7)$$

where $\kappa(t) = [(-1)^{n-1}[V_{K_0^{(n)}}y](t),\ldots,(-1)^{n-1}[V_{K_{n_\delta-1}^{(n)}}y](t)]^\top$ and $\Gamma(t) = [\nu_0(t),\nu_1(t),\ldots,\nu_{n_\delta-1}(t)]^\top$. Moreover, the following definition and assumption is needed to guarantee the invertibility of the matrix $\Gamma(t)$ in (7).

Definition III.1. *(Persistency of Excitation (PE))[23]: A bounded locally-integrable vector function $\varphi(t) : \mathbb{R} \to \mathbb{R}^n$ is said to be persistently exciting if there exist a constant s_0 and positive constants T_0 and α such that*
$$\frac{1}{T_0}\int_s^{s+T_0} \varphi(t)\varphi(t)^\top dt \geq \alpha I > 0 \ \forall s \geq s_0.$$

Assumption 1. *Given the I/O measurement $y(t)$ and $u_k(t)$ and the BF-NKs in (3), there exist some $\alpha_h \in \mathbb{R}_{>0}$ and $t_{\alpha,h} \in \mathbb{R}_{>0}, \forall h \in \{0,\ldots,n_\delta-1\}$ such that*

$$\int_{t-t_{\alpha,h}}^t \nu_h(\tau)\nu_h(\tau)^\top \geq \alpha_h I, \ \forall t > t_{\alpha,h}, \qquad (8)$$

following the Definition III.1.

Proposition III.1. *(Implication) The PE condition* (8) *is satisfied if the following conditions are verified*

- *The inputs* $u_k(t)$ *are sufficiently rich (SR) of at least order* m_k, $\forall k \in \{0, \ldots, p-1\}$, $\forall t > 0$;
- *The output* $y(t)$ *is SR of at least order* n, $\forall t > 0$;
- *The I/O vector* $\zeta(t) \triangleq [y(t), y^{(1)}(t), \ldots, y^{(n-1)}(t), u_0(t), \ldots, u_0^{(m_0-1)}(t), \ldots u_{p-1}^{(m_{p-1}-1)}(t)]$ *is PE.*

Proof. Let us split the vector $\nu_h(t)$ into $p+2$ vectors $\nu_{h,y}(t)$, $\nu_{h,u_k}(t)$, $k \in \{0, \ldots, p-1\}$ and $\nu_{h,\gamma}(t)$ where $\nu_{h,y}(t) = \big[[V_{K_h}y](t), [V_{K_h^{(1)}}y](t), \ldots, (-1)^n[V_{K_h^{(n-1)}}y](t)\big]$, $\nu_{h,u_k}(t) = \big[[V_{K_h}u_k](t), [V_{K_h^{(1)}}u_k](t), \ldots, (-1)^{m_k}[V_{K_h^{(m_k-1)}}u_k](t)\big]$, $k \in \{0, \ldots, p-1\}$, $\nu_{h,\gamma}(t) = \big[\gamma_{h,n-1}(t), \gamma_{h,n-2}(t), \ldots, \gamma_{h,0}(t)\big]$.

Recall the shape of the kernel (3) and its derivatives, which can be expressed in the following forms

$$K_h^{(i)}(t, \tau) = e^{-\omega_h t} \sum_{q=0}^{N} (-1)^q \binom{N}{q} (\omega_h - q\bar{\omega})^i e^{(\omega_h - q\bar{\omega})\tau}. \quad (9)$$

Substituting τ in (9) with t, it is obvious that with the designed set of kernel functions, the vector $\nu_{h,\gamma}(t)$ is PE.

By analysing the transformed signals in the Laplace domain, one gets $\mathscr{L}\{\nu_{h,y}(t)\} = [-\chi_{h,0,y}(s), \chi_{h,1,y}(s), \ldots, (-1)^n \chi_{h,n-1,y}(s)]$, $\mathscr{L}\{\nu_{h,u_k}(t)\} = [-\chi_{h,0,u_k}(s), \chi_{h,1,u_k}(s), \ldots, (-1)^{m_k} \chi_{h,m_k-1,u_k}(s)]$, for all $k \in \{0, \ldots, p-1\}$ and $\chi_{h,i,\star}(s) = \sum_{q=0}^{N} (-1)^q \binom{N}{q} (\omega_h - q\bar{\omega})^i \frac{1}{s+\omega_h} \star (s + q\bar{\omega})$, where $\star(s)$ represents the I/O signals in the Laplace domain. In this regard, one gets

$$\mathscr{L}\{\nu_{h,y}(t)^\top\} = \Psi_{h,y}(s)[Y(s), Y(s+\bar{\omega}), \ldots, Y(s+N\bar{\omega})]^\top,$$

where

$$\Psi_{h,y}(s) =$$
$$\begin{bmatrix} -\binom{N}{0}\frac{1}{s+\omega_h} & \binom{N}{1}\frac{1}{s+\omega_h} & \cdots & (-1)^{N+1}\binom{N}{N}\frac{1}{s+\omega_h} \\ \binom{N}{0}\frac{\omega_h}{s+\omega_h} & -\binom{N}{1}\frac{\omega_h-\bar{\omega}}{s+\omega_h} & \cdots & (-1)^N\binom{N}{N}\frac{\omega_h-n\bar{\omega}}{s+\omega_h} \\ -\binom{N}{0}\frac{\omega_h^2}{s+\omega_h} & \binom{N}{1}\frac{(\omega_h-\bar{\omega})^2}{s+\omega_h} & \cdots & (-1)^{N+1}\binom{N}{N}\frac{(\omega_h-n\bar{\omega})^2}{s+\omega_h} \\ \vdots & \vdots & & \vdots \\ -\binom{N}{0}\frac{(-\omega_h)^{n-1}}{s+\omega_h} & \binom{N}{1}\frac{(\bar{\omega}-\omega_h)^{n-1}}{s+\omega_h} & \cdots & (-1)^{n+N}\binom{N}{N}\frac{(\omega_h-n\bar{\omega})^{n-1}}{s+\omega_h} \end{bmatrix}. \quad (10)$$

Therefore, $\mathscr{L}\{\nu_{h,y}(t)^\top\}$ can be regarded as the output of an MIMO system with transfer function matrix $\Psi_{h,y}(s)$ fed by an input vector $[Y(s), Y(s+\bar{\omega}), \ldots, Y(s+N\bar{\omega})]$. Referring to [24], this input vector is at least sufficiently rich of order $(n, n, \ldots, n/1)$. Note that the system (10) is output reachable, i.e. there exists no $\beta \neq 0, \beta \in \mathbb{R}^n$ such that $\beta^\top \Psi_{h,y}(s) = 0$, $\forall s$. Hence, in the light of Theorem 4.2 of [24], one can conclude that in the time-domain the vector $\nu_{h,y}(t)$ is PE. In the same line of reasoning, vectors $\nu_{h,u_k}(t)$ are PE for all $k \in \{0, \ldots, p-1\}$. Owing to the third condition in Proposition III.1, i.e. the I/O vector $\zeta(t)$ is PE, it turns out that the vector $[\nu_{h,y}(t), \nu_{h,u_0}(t), \ldots, \nu_{h,u_{p-1}}(t)]$ is also PE. Moreover, considering the Laplace transformation of the kernel derivatives (9) and the fact that $y(t)$ and $u_k(t)$ are SR of at least order n and m_k, $\forall k \in \{0, \ldots, p-1\}$ respectively, it is straightforward to prove that the combined vector $\nu_h(t)$ is PE. $\qquad\square$

It is worth noting that conditions in Proposition III.1 can be examined off-line. Therefore, under Assumption 1, stacking n_δ vectors $\nu_h(t)$ with $K_h(t, \tau)$, $h \in \{0, \ldots, n_\delta - 1\}$ ensures that the matrix $\Gamma(t)$ is invertible $\forall t > 0$. In turn, the unknown vector θ can be estimated by

$$\hat{\theta}(t) = \Gamma(t)^{-1}\kappa(t), \forall t > 0.$$

IV. ROBUSTNESS ANALYSIS

In this section, we characterize the robustness of the proposed estimator against measurement noise and process disturbance. Consider the following model affected by disturbances:

$$\begin{cases} z_d^{(1)}(t) &= Az_d(t) + Bu(t) + \eta(t) \\ y_d(t) &= c^\top z_d(t) + d_y(t), \end{cases} \quad (11)$$

where $d_y(t) \in \mathbb{R}$ denotes the measurement noise and $\eta(t) = [\eta_0, \eta_1, \ldots, \eta_{n-1}]^\top \in \mathbb{R}^n$ represents the process disturbance, both are assumed to be bounded, *i.e.* $|d_y(t)| \leq \bar{d}_y$ and $|\eta(t)| \leq \bar{\eta}$, $\forall t \geq 0$.

Under the action of $d_y(t)$ and $\eta(t)$, thanks to the linearity of the Volterra operator, (5) becomes

$$(-1)^{n-1}[V_{K_h^{(n)}}y](t) + (-1)^{n-1}[V_{K_h^{(n)}}d_y](t) \\ + \sum_{i=0}^{n-1}(-1)^i[V_{K_h^{(i)}}\eta_{n-1-i}](t) = \hat{\nu}_h(t)^\top\hat{\theta}(t), \quad (12)$$

where $\hat{\nu}_h(t) =$

$$\big[-[V_{K_h}y_d](t), \ldots, (-1)^n[V_{K_h^{(n-1)}}y_d](t), -[V_{K_h}u_0](t),$$
$$\ldots, (-1)^{m_{p-1}}[V_{K_h^{(m_{p-1}-1)}}u_{p-1}](t), \gamma_{h,n-1}(t), \ldots, \gamma_{h,0}(t)\big]^\top.$$

In turn, by defining $\epsilon_{\nu_h}(t) \triangleq \hat{\nu}_h(t) - \nu_h(t) = \big[-[V_{K_h}d_y](t), \ldots, (-1)^n[V_{K_h^{(n-1)}}d_y](t), 0, \ldots, 0\big]^\top$ and $\epsilon_\theta(t) \triangleq \hat{\theta}(t) - \theta(t)$, by subtracting (5) from (12), we have $\epsilon_{\kappa_h} = \hat{\nu}_h(t)^\top\epsilon_\theta(t)$. with $\epsilon_{\kappa_h} = (-1)^{n-1}[V_{K_h^{(n)}}d_y](t) + \sum_{i=0}^{n-1}(-1)^i[V_{K_h^{(i)}}\eta_{n-1-i}](t) - \epsilon_{\nu_h}(t)^\top\theta$.

We now introduce the PE conditions in the noisy scenario.

Assumption 2. *Given the noisy output measurement $y_d(t)$, the input $u_k(t)$ and the kernel* (3)*, there exists some $\alpha_{d,h} \in \mathbb{R}_{>0}$ and $t_{d,h} \in \mathbb{R}_{>0}$ satisfying*

$$\int_{t-t_{d,h}}^{t} \hat{\nu}_h(\tau)\hat{\nu}_h(\tau) \geq \alpha_{d,h}I, \ \forall t > t_{d,h},$$

for all $h \in \{0, \ldots, n_\delta - 1\}$.

Namely, in the presence of noise, the noisy $\Gamma(t)$, denoted as $\hat{\Gamma}(t)$, could be singular on some time instances. Therefore, in the noisy scenario, we resort to online calculating the determinant of $\hat{\Gamma}(t)$ and switching off the estimator on certain time instances when $\hat{\Gamma}(t)$ loses rank. Therefore, under Assumption 2, the estimation error can be written as

$$\epsilon_\theta(t) = \hat{\Gamma}(t)^{-1}\epsilon_{\kappa_h}(t), \quad \forall t \in \{\mathbb{R}_{>0} | det(\hat{\Gamma}(t)) > 0\}, \quad (13)$$

where $\hat{\Gamma}(t) \triangleq [\hat{\nu}_0(t), \hat{\nu}_1(t), \ldots, \hat{\nu}_{n_\delta-1}(t)]^\top$, $\epsilon_\kappa(t) \triangleq [\epsilon_{\kappa_0}(t), \ldots, \epsilon_{\kappa_{n_\delta-1}}(t)]^\top$. Remarkably, elements in $\epsilon_\kappa(t)$ are linear combinations of the Volterra images of the measurement noise $[V_{K_h^{(i)}} d_y](t) = \epsilon_{\xi_d,i,h}(t)$, and of the process noise $[V_{K_h^{(i)}} \eta](t) = \epsilon_{\xi_\eta,i,h}(t)$, $i \in \{0, \ldots, n\}$, $h \in \{0, \ldots, n_\delta - 1\}$. $\epsilon_{\xi_d,i,h}(t)$ has the dynamics

$$\dot{\epsilon}_{\xi_d,i,h}((t) = -\omega_h \epsilon_{\xi_d,i,h}(t) + K_h^{(i)}(t,t) d_y(t).$$

Therefore, after applying the Volterra mapping, the transformed measurement disturbance satisfies:

$$|\epsilon_{\xi_d,i,h}(t)| \leq \left| \frac{1}{\omega_h} \bar{d}_y \sup_{0 < \tau \leq t} K_h^{(i)}(\tau, \tau) \right| \triangleq \bar{\epsilon}_{\xi_d,i,h}(t). \quad (14)$$

Following the same line of reasoning, the Volterra images of $\eta_i(t)$, $\forall i \in \{0, \ldots, n-1\}$ have their upper bounds $\bar{\epsilon}_{\xi_\eta,i,h}(t) \triangleq |\frac{1}{\omega_h} \bar{\eta} \sup_{0 < \tau \leq t} K_h^{(i)}(\tau, \tau)|$. In turn, $|\epsilon_{\nu_h}(t)| \leq [\bar{\epsilon}_{\xi_d,0,h}, \bar{\epsilon}_{\xi_d,1,h}, \ldots, \bar{\epsilon}_{\xi_d,n-1,h}, 0, \ldots, 0] \triangleq \bar{\epsilon}_{\nu_h}(t)$. Hence, the upper bound $\bar{\epsilon}_{\kappa_h}(t)$ can be defined as

$$|\epsilon_{\kappa_h}(t)| \leq \bar{\epsilon}_{\xi_d,n,h}(t) + \sum_{i=0}^{n-1} \bar{\epsilon}_{\xi_\eta,i,h} + \bar{\epsilon}_{\nu_h}(t)^\top |\theta| \triangleq \bar{\epsilon}_{\kappa_h}(t)$$

for all $h \in \{0, 1, \ldots, n_\delta - 1\}$. Thus, the upper bound on $\epsilon_\kappa(t)$ is $\bar{\epsilon}_\kappa(t) \triangleq [\bar{\epsilon}_{\kappa_0}(t), \ldots, \bar{\epsilon}_{\kappa_{n_\delta-1}}(t)]$. Under Assumption 2, the estimation error is bounded as long as the measurement noise $d_y(t)$ and the process noise $\eta(t)$ are bounded, *i.e.*

$$|\epsilon_\theta(t)| \leq [\hat{\Gamma}(t)^{-1}] \bar{\epsilon}_\kappa(t), \quad \forall t \in \{\mathbb{R}_{>0} | det(\hat{\Gamma}(t)) > 0\}.$$

V. NUMERICAL ISSUES CAUSED BY DISCRETIZATION

When deploying the proposed algorithm in a digital computing platform, the impact of time discretization has to be carefully analyzed. Specifically, the discretization error is unavoidable and causes some deterioration of performance compared to the theoretical characterization obtained so far in the continuous-time case.

Let us consider the discrete-time implementation of (6). The discretized Volterra transformation admits the dynamics

$$\breve{\xi}_{h,\star}(k+1) = \breve{\xi}_{h,\star}(k) + \Upsilon_\star(k), \quad (15)$$

with $\breve{\xi}_{h,\star}(1) = 0$ and $\Upsilon_\star(k)$ is the integral approximation which, depending on the discretization methods, could be function of $\star(t_k), \star(t_{k+1}), K(t_k, t_k), K(t_{k+1}, t_{k+1})$ and ω_h, with $t_k = (k-1)Ts$. The global discretization error $\epsilon_{\breve{V}_\star,i,h}(k) \triangleq [V_{K_h^{(i)}} \star](t_k) - [\breve{V}_{K_h^{(i)}} \star](k)$, varies between different discretization methods due to different $\Upsilon_\star$, e.g. $\mathcal{O}(T_s^2)$ for midpoint method (2nd order) and $\mathcal{O}(T_s^4)$ for classic order Runge-Kutta method (4th order). Needless to say, the discretization error decreases with increasing complexity of the discretization methods [26]. Similarly as the deduction from (12) to (13), the estimation error introduced by discretization in the noise-free scenario can be expressed as

$$\epsilon_{\breve{\theta}}(k) \triangleq \breve{\theta}(k) - \theta(t_k) = \breve{\Gamma}(t)^{-1} \epsilon_{\breve{\kappa}}(k), \quad \forall t > 0, \quad (16)$$

where $\breve{\Gamma}(k) \triangleq [\breve{\nu}_0(k), \breve{\nu}_1(k), \ldots, \breve{\nu}_{n_\delta-1}(k)]^\top$, $\epsilon_{\breve{\kappa}}(k) \triangleq [(-1)^{n-1} \epsilon_{\breve{V}_y,n,h}(k) - \epsilon_{\breve{\nu}_0}(k)^\top \theta, \ldots, (-1)^{n-1} \epsilon_{\breve{V}_y,n,h}(k) -$

$\epsilon_{\breve{\nu}_{n_\delta-1}}(k)^\top \theta]^\top$, $\epsilon_{\breve{\nu}_h}(k) \triangleq \breve{\nu}_h(k) - \nu_h(t_k)$, $\forall h \in \{0, \ldots, n_\delta - 1\}$. Therefore, reducing the discretization error $\epsilon_{\breve{V}_\star,i,h}(k)$ minimizes the error in (16) by the term $\epsilon_{\breve{\kappa}}(k)$. On the other hand, the discretization error undergoes a major amplification for small t, when $\breve{\Gamma}(t)$ is close to singularity. To address this issue, we use an activation λ_a to hold the algorithm for a short time T_a (activation time) before activation until the $det(\breve{\Gamma}(t)) \geq \lambda_a$. Specific implementation and tuning are detailed in Section VI with examples.

VI. NUMERICAL EXAMPLES

In this section, numerical examples are presented to show the effectiveness of the proposed simultaneous parameter-state estimation method.

A. Example 1

We consider a double-input single-output LTI system

$$y^{(2)}(t) = a_1 y^{(1)}(t) + a_0 y(t) + b_{11} u_1^{(1)}(t) + b_{20} u_2(t).$$

The inputs are $u_1(t) = 10 \sin(10t)$ and $u_2(t) = \sin(2t) - 0.5$, where both $u_1(t)$ is sufficiently rich of order 2 and $u_2(t)$ is sufficiently rich of order 3. As a result, the output will be rich of order 5. Moreover, the elements in the I/O vector $[y(t), u_1(t), u_2(t)]$ are mutually independent, thus verifying the PE condition.

1) Discretization and parameter tuning: Theoretically, the proposed estimator achieves deadbeat convergence instantaneously as $t \neq 0$ in the noise-free scenario. However, in the numerical implementation, distortion may occur due to the discretization, significantly depending on the sampling time T_s and the discretization scheme. Evidence can be found in Table I, in which the *Root Mean Square Error (RMSE)* is defined as $RMSE(\varsigma)_{(t_1,t_2)} \triangleq \sqrt{(\sum_{n_{rmse}} e_\varsigma^\top e_\varsigma)/n_{rmse}}$, where ς represents the parameter vector $\varsigma \triangleq [a_0, a_1, b_{11}, b_{20}]^\top$ and (t_1, t_2) represents the time interval we consider to calculate the RMSE and n_{rmse} is the number of samples in this interval.

Moreover, Table I also shows the effectiveness of the activation threshold λ_a which minimizes the influence of discretization in the initial phase comparing to the no-threshold ($\lambda_a = 0$) mechanism. Notably, after the initial phase (e.g. $t \in (2,3]$ in Table I), effects of the discretization becomes so tiny that can be ignored with Runge-Kutta solver.

TABLE I: RMSE with different T_s and discretization methods

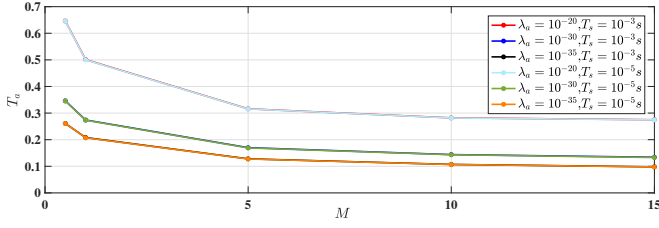| Methods	λ_a	T_s	$RMSE(\varsigma)_{(0,1]}$	$RMSE(\varsigma)_{(2,3]}$
Euler (1st order)	0	10^{-3}	3.43×10^{20} p.u.	34.3 p.u.
		10^{-4}	1.64×10^{14} p.u.	3.40 p.u.
	10^{-20}	10^{-3}	5.87×10^3 p.u.	34.3 p.u.
		10^{-4}	498 p.u.	3.40 p.u.
Runge-Kutta (4th order)	0	10^{-3}	5.95×10^3 p.u.	5.11×10^{-4} p.u.
		10^{-4}	1.88×10^4 p.u.	3.62×10^{-7} p.u.
	10^{-20}	10^{-3}	1 p.u.	5.11×10^{-4} p.u.
		10^{-4}	7.11×10^{-4} p.u.	3.62×10^{-7} p.u.

Fig. 1: Influence of activation threshold λ_a, kernel parameter M and sampling interval T_s on activation time $T_a(s)$.

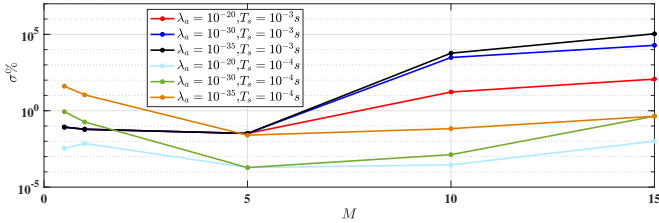Fig. 2: Influence of activation threshold λ_a, kernel parameter M and sampling interval T_s on distortion overshoot $\sigma\%(dB)$.

Therefore, to implement the proposed estimator, the activation threshold λ_a, the sampling interval T_s and the kernel parameters $[\omega_0, \ldots, \omega_5], \bar{\omega}$ must be determined off-line. In order to guarantee an accurate approximation of the continuous-time algorithms, the parameter tuning are conducted to show how the effects of the discretization, in terms of the the activation time T_a and the overshoot $\sigma\% \triangleq \frac{\text{maximal estimate} - \text{true value}}{\text{true value}} \times 100\%$, are influenced by λ_a, Ts. Differently parametrized sets of BF-NKs with $[\omega_0, \ldots, \omega_5] = [1\ 2\ 3\ 4\ 5\ 6] \times M$, $M \in \{0.5, 1, 5, 10, 15\}$ and $\bar{\omega} = 2.5$ are utilized obtaining results in Fig. 1 and 2.

In Fig. 1, it is indicated that smaller activation thresholds will shorten the activation times. Meanwhile, with fixed threshold λ_a, the activation times tend to be shorter with larger kernel parameters M. However, the influence of the thresholds on the activation time is more significant than that of the kernel parameters. Fig. 1 also shows that the sampling time T_s does not have obvious effects on the activation time. As depicted in Fig. 2, smaller activation thresholds or larger kernel parameters result in larger overshoots. Conclusion can be drawn intuitively that shorter activation times correspond to larger overshoot. Therefore, the designer should consider the trade off between the overshoot and the activation time in tuning λ_a. Needless to say, the overshoot, caused by discretization, will be reduced significantly in more precise discretization (*i.e.*, more complex scheme and smaller sampling time). Moreover, there is no explicit relation between the variation of the kernel parameters and the overshoot which reaches a minimum at $M = 5$. Taking the accuracy and complexity into consideration, we adopt Runge-Kutta solver with sampling time $T_s = 10^{-3}$.

As for the kernel parameters $[\omega_0, \ldots, \omega_5]$, they play a more significant role in noise immunity than discretization. In the noisy scenario, we consider a measurement noise within $[-0.2, 0.2]$ additive to the output $y(t)$ and a process noise within $[-0.5, 0.5]$. The RMSE of the parameter-state estimates $\hat{\varsigma}(t)$ and $\hat{x}(t)$ with different parameter settings are listed in

Table II. Recall the fact that the threshold is also used to detect the loss of excitation in the noisy scenario, its variation is also considered. For the purpose of analyzing only the effects of the parameters on the robustness, we exclude the effect of the overshoot by using the samples from $0.2s$ after activation time until the end of the simulation to calculate RMSEs.

TABLE II: RMSEs with different parameters choice in the noisy scenario

T_s	λ_a	M	T_a	$RMSE(\varsigma)$	$RMSE(x)$
10^{-3}	10^{-7}	0.5	3.667	1 p.u.	1 p.u.
		1	2.715	4.14 p.u.	1.22 p.u.
		2	1.371	31.12 p.u.	2.14 p.u.
		3	1.272	230.44 p.u.	5.33 p.u.
		5	1.374	1.19×10^3 p.u.	28.39 p.u.
	10^{-6}	0.5	4.398	0.42 p.u.	4.16 p.u.
		1	2.717	1.56 p.u.	2.24 p.u.
		2	1.821	10.37 p.u.	0.91 p.u.
		3	1.680	68.69 p.u.	2.85 p.u.
		5	2.109	446.90 p.u.	11.47 p.u.

According to Table. II, the robustness of the estimator degrades with increasing kernel parameters verifying the analysis in Section IV. Furthermore, the threshold λ_a is also able to detect the singularity of $\hat{\Gamma}(t)$. In our simulation, when the loss of excitation is detected, we switch off the estimation and all the estimates are frozen to the last values until the measurements become excited again. Therefore, the increase of λ_a helps to enhance the noise attenuation which has been verified in Table II. However, this may not be applicable in the case of state estimation for small kernel parameters ($M \le 1$), since smaller values of the parameters result in smaller determinants of $\hat{\Gamma}(t)$, and therefore higher activation thresholds make the estimator to be switched off more often. On the other hand, although being theoretically feasible, extremely large M may be computationally challenging thus risking to increase the sensitivity to the noise, and therefore should be avoided. In this regard, designer should judge comprehensively the activation time, robustness to noise and PE conditions, in tuning M and λ_a.

2) Noise-free scenario: According to Table I and Fig. 1 2, to verify the effectiveness of the parameter-state joint estimator in the noise-free scenario, we use 6 BF-NK functions with $\bar{\omega} = 2.5$, $[\omega_0, \ldots, \omega_5] = [1, 2, 3, 4, 5, 6] \times 5$ and $N = 4$. The activation threshold is chosen as $\lambda_a = 10^{-20}$.

Fig. 3 shows the instantaneous convergence of the estimator where both the estimates of the parameters and state variables jump to the corresponding true values.

3) Noisy scenario: We examine the estimator performance in a scenario where the measurement $y(t)$ is corrupted by a uniformly distributed random disturbance within $[-0.2, 0.2]$ and the system dynamics are perturbed by a signal $\eta(t) = [0.5 \sin(50t), 0.5 \sin(50t)]^\top$ as in (11).

In this example, we are adopting the parameter settings $M = 0.5$ and $\lambda_a = 10^{-7}$ obtaining the estimation results as in Fig. 4. It is worth noting that, at the beginning, the
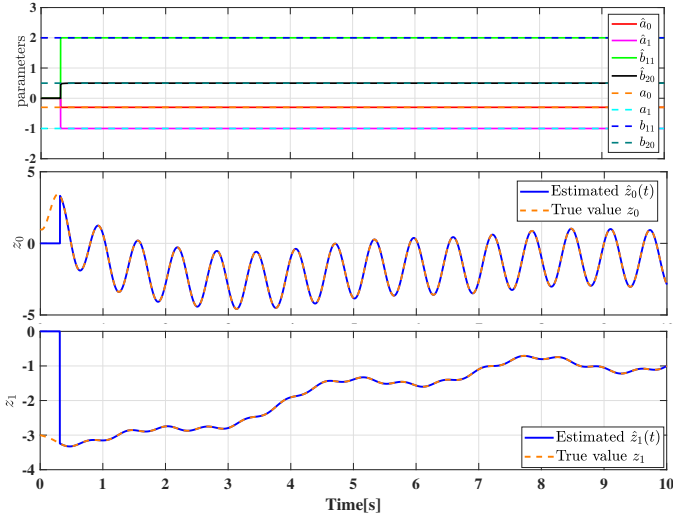
Fig. 3: Parameter-state joint estimates in noise-free scenario.

Fig. 4: Estimation results in noisy scenario with $T_s = 10^{-3}s$.

Fig. 5: Estimation results in noisy scenario with $T_s = 10^{-2}s$.

discretization error tends to amplify the effect of noise; therefore, a relatively large threshold λ_a is needed to pause the estimator until the elements in $\hat{\Gamma}(t)$ are large enough to avoid the distortion. Namely, it is not the number of data samples collected before activation, but the values of the elements in $\hat{\Gamma}(t)$, that influence the performance. This is shown by the parameter-state joint estimates in Fig. 5, where the sampling interval is set to be $T_s = 10^{-2}s$. With the proposed activation mechanism, 342 data points are collected before activation while the proposed method still provides good estimation. However, a longer sampling interval is prone to enlarge the discretization error increasing the sensitivity to the noise. Therefore, the peak of the spikes tends to be amplified while $T_s = 10^{-2}s$. All in all, the proposed estimator is able to deal with the measurement and process noise. All the estimates converge to the close neighbour of the true values albeit with some spikes which are likely to happen periodically. This phenomena is due to the effect of the measurement noise, making the matrix $\hat{\Gamma}(t)$ close to its singularity. Moreover, since the output signal is sinusoidal whose derivatives will suffer from the maximum effects of noise at their positive and negative peaks thus give rise to the periodicity of the spikes.

B. Example 2

In this example, inspired by a recent finite-time parameter-state joint estimator in [18], our methodology is extended to deal with a certain class of nonlinear system in which the parameters appear linearly in a nonlinear input-output differential equation. Herein, we are considering the parameter-state estimation of the standard form of the unforced Van der Pol oscillator $y^{(2)}(t) = \mu(1 - y^2)y^{(1)}(t) - y(t)$, which has its state-space form

$$
\begin{aligned}
x_1^{(1)}(t) &= \mu x_1(t) + x_2(t) - \frac{\mu}{3}x_1^3(t), \\
x_2^{(1)}(t) &= -x_1(t).
\end{aligned}
\tag{17}
$$

The estimation results of the proposed method will be compared with the finite-time modulating function (MF)-based estimator with $T_s = 10^{-3}s$. For the MF method, we refer to the two cascading estimator structure in [18] with the
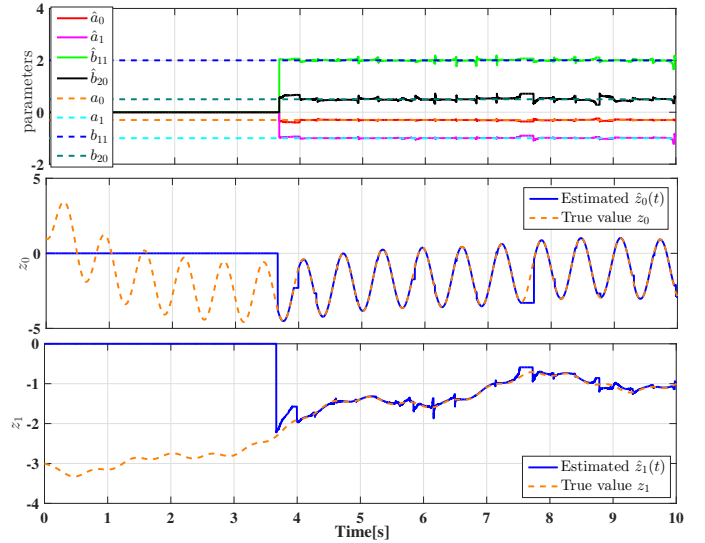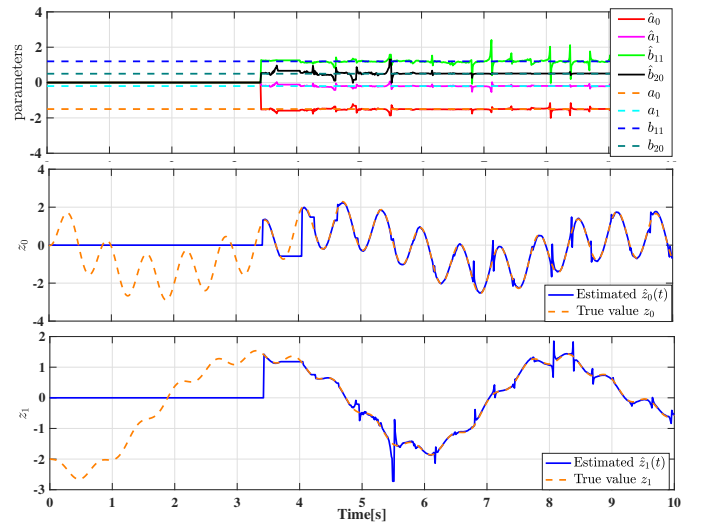
left and total modulating functions used therein, i.e. $\phi_k(t) = (T - t)^k t^k$, $k = 4, 5, 6, 7$ and $\varphi_1 = t^i e^{-t}$, $i = 2, 3$. Moreover, integrals with expanding window are used for $t < 2s$ and for $t \geq 2s$, receding horizon are applied with $T = 2s$. All the kernel parameters remain the same with Example 1.

In noise-free scenario, the estimation results are depicted as in Fig. 6. The figure indicates that both the kernel-based and the MF-based estimator are able to achieve finite-time convergence in parameter-state estimation. However, thanks to the usage of the activation threshold, the kernel-based parameter estimation appears to be less vulnerable to the discretization error with the same discretization technique. Owing to the fact that only a single estimator is used, there is no cascading error affecting the state estimation.

In the noisy scenario, we consider only measurement noise on $y(t)$ with SNR=28. The RMSE of the parameters and the state variables for $t \in (2, 10]s$ of both methods are compared in Table III, where the proposed kernel-based estimator appears to be robust against measurement noise.
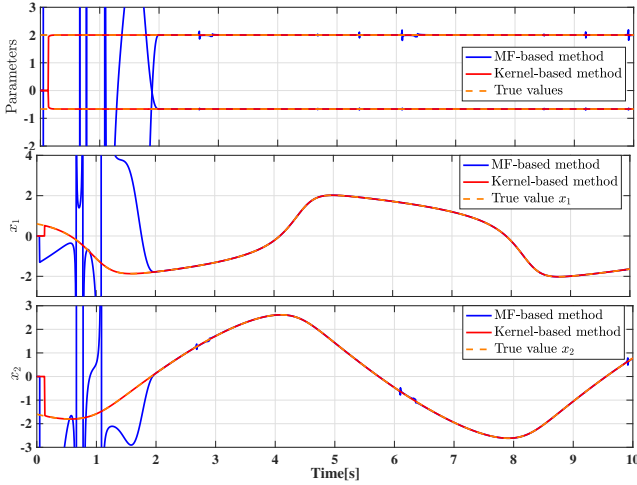
Fig. 6: State estimates comparison in noise-free scenario.

TABLE III: state estimation RMSE of both methods

Methods	Kernel-based method	MF-based method
$RMSE_{(2,10](\mu)}$	1.38	2.89
$RMSE_{(2,10](x)}$	0.15	0.81

VII. CONCLUDING REMARKS

In this work, a simultaneous parameter-state estimator for MISO systems has been developed resorting to Volterra integral operators. Thanks to the usage of the non-asymptotic kernels, the exact estimates can be achieved in an arbitrarily small fixed time in the noise-free scenario. In the noisy scenario, the estimator still guarantees bounded estimation error against bounded noises. The numerical analysis also provides implementation hints. The effectiveness of the proposed estimator has been verified by extensive simulations.

REFERENCES

[1] A. Ionesi and J. Jouffroy, "On-line parameter estimation of an Air Handling Unit model: experimental results using the modulating function method," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 81-86, 2018.

[2] J. Reger, P. Mai, and H. Sira-Ramirez, "Robust algebraic state estimation of chaotic systems," *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications*, pp. 326-331, 2006.

[3] J. Reger and J. Jouffroy, "On algebraic time-derivative estimation and deadbeat state reconstruction," *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pp. 1740-1745, 2009.

[4] A. Concha, L. A.-Icaza and R. Garrido, "Simultaneous parameter and state estimation of shear buildings," *Mechanical Systems and Signal Processing*, vol. 70, pp. 788- 810, 2016.

[5] A. Ohsumi, K. Kameyama, and K. Yamaguchi, "Subspace identification for continuous-time stochastic systems via distribution-based approach," *Automatica*, vol. 38, no. 1 pp. 63-79, 2002.

[6] P. Moireau, D. Chapelle, and P. L. Tallec, "Joint state and parameter estimation for distributed mechanical systems," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 6-8, pp. 659-677, 2008.

[7] A. B. Proca and A. Keyhani, "Sliding-mode flux observer with online rotor parameter estimation for induction motors," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 2, pp. 716-723, 2007.

[8] G. L. Plett, "Sigma-point kalman filtering for battery management systems of lipb-based hev battery packs: Part 2: Simultaneous state and parameter estimation," *Journal of Power Sources*, vol. 161, no. 2, pp. 1369-1384, 2006.

[9] A. B. Proca and A. Keyhani, "Sliding-mode flux observer with online rotor parameter estimation for induction motors," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 2, pp. 716-723, 2007.

[10] A. Concha, L. Alvarez-Icaza, and R. Garrido, "Simultaneous parameter and state estimation of shear buildings," *Mechanical Systems and Signal Processing*, vol. 70, pp. 788-810, 2016.

[11] R. E. Kopp and R. J. Orford, "Linear regression applied to system identification for adaptive control systems," *Aiaa Journal*, vol. 1, no. 10, pp. 2300-2306, 1963.

[12] L. Nelson and E. Stear, "The simultaneous on-line estimation of parameters and states in linear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 94-98, 1976.

[13] P. Moireau, D. Chapelle, and P. Le Tallec, "Joint state and parameter estimation for distributed mechanical systems," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 6-8, pp. 659–677, 2008.

[14] B. Chen, G. Pin, W. M. Ng, T. Parisini, and S.-Y. R. Hui, "A fast-convergent modulation integral observer for online detection of the fundamental and harmonics in grid-connected power electronics systems," *IEEE Transactions on Power Electronics*, vol. 32, no. 4, pp. 2596-2607, 2017.

[15] S. K. Spurgeon, "Sliding mode observers: a survey," *International Journal of Systems Science*, vol. 39, no. 8, pp. 751-764, 2008.

[16] M. Mboup, "Parameter estimation for signals described by differential equations," *Applicable Analysis*, vol. 88, no. 1, pp. 29-52, 2009.

[17] G. Pin, B. Chen, and T. Parisini, " Robust deadbeat continuous-time observer design based on modulation integrals, " *Automatica*, vol. 107, pp. 95-102, 2019.

[18] J. Jouffroy and J. Reger, "Finite-time simultaneous parameter and state estimation using modulating functions," in *Proc. of IEEE Conference on Control Applications (CCA)*, pp. 394-399, 2015.

[19] G. Pin, A. Assalone, M. Lovera, and T. Parisini, "Non-asymptotic kernel-based parametric estimation of continuous-time linear systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 360–373, 2016.

[20] G. Pin, M. Lovera, A. Assalone, and T. Parisini, "Kernel-based non-asymptotic state estimation for linear continuous-time system," in *Proc. of IEEE American Control Conference (ACC)*, pp. 3123-3128, 2013.

[21] P. Li, "Finite-time system identification, estimation and fault detection," *Imperial College London*, pp.1-137, 2019.

[22] P. Li, F. Boem, G. Pin, and T. Parisini, "Deadbeat simultaneous parameter-state estimation for linear continuous-time systems: a kernel-based approach," in *Proc. of IEEE European Control Conference (ECC)*, pp. 2493-2498, 2018.

[23] N. Shimkin and A. Feuer, "Persistency of excitation in continuous-time systems," *Systems & Control Letters*, vol. 9, no. 3, pp. 225-233, 1987.

[24] I. Mareels and M. Gevers, "Persistency of excitation criteria for linear, multivariable, time-varying systems," *Mathematics of Control, Signals and Systems*, vol. 1, no. 3, pp. 203-226, 1988.

[25] G. Prando, M. Zorzi, A. Bertoldo, and A. Chiuso, "Estimating effective connectivity in linear brain network models," *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5931-5936, 2017.

[26] A. Iserles, "On the global error of discretization methods for highly-oscillatory ordinary differential equations," *BIT Numerical Mathematics*, vol. 42, no. 3, pp. 561-599, 2002.