2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Finite-interval kernel-based identification and state estimation for LTI systems with noisy output data*

Debarshi Patanjali Ghoshal[1], Hannah Michalska[2]

Abstract— **This note extends previous results pertaining to algebraic state and parameter estimation of linear systems based on a special construction of kernel system representations that incorporate system differential invariants. Main results include explicit expressions for the kernel functions for single-input, single-output LTI systems of arbitrary order. A recursive regression type algorithm is also proposed for the purpose of joint system identification and finite interval filtering. As compared with previous results the proposed non-asymptotic estimation method proves remarkably robust to Gaussian noise in output measurements. The approach has been shown to extend to linear time-varying and linear parameter-varying systems in a multivariate setting. The idea of system-related kernels can further be employed to enhance convergence properties of moving-window and minimum energy nonlinear filtering methods.**

I. INTRODUCTION

The body of work on system identification and filtering is extremely vast, due to its countless important applications; to cite merely a few leading texts in the area [1], [2], [3]. Recursive approaches prevail, as primarily rooted in Kalman's seminal work [4], as they are computationally fast and avoid handling of large data. Recently, however, there is a growing interest in non-asymptotic estimation methods which is justified by modern developments in systems with rapidly switching dynamics [5], advanced nonlinear control methods based on differential flatness [6], and need for powerful target tracking algorithms of superior performance in speed and precision [7].

The idea of constructing kernel representations of differential systems, as first presented in [8], initially sprung from an attempt to improve on the algebraic signal differentiation approach proposed in [9], [10], as the principal tool required in the implementation of controllers for nonlinear systems that are differentially flat. The original algebraic differentiator was based on truncated Taylor series signal approximation and attributed its properties to the introduction of an algebraic annihilator of the Taylor series coefficients up to any desired order, [9], [11]. After iterated integration it then delivered the values of the selected coefficients that escaped annihilation, effectively the values of the desired higher order derivatives. The annihilator served yet a different purpose: that of shaping the noise rejection characteristic of the resulting "filter", see [12]. As was admitted, however, the method required frequent re-initialization when used forward in time, and its noise rejection properties were characterized as non-standard [13].

As quickly became clear, improvements could be achieved if more information about the signal were available; e.g. if it were known that it represents an output from a dynamical system exhibiting a certain differential invariance retaining a deterministic footprint of the system. The obvious fact the Cayley-Hamilton theorem delivers one such invariance as expressed by the characteristic equation of an LTI system led to the construction of the kernel representation in [8] with its immediate application to state estimation and system identification [14] that is quite competitive with the deterministic version of subspace identification [15], [16]. It is noted that there are other kernel-based identification methods reported in the literature, notably [17], [18], [1]. The Gaussian kernels introduced there are used to identify finite-impulse responses for stable LTI systems. To the best of our knowledge the approach of [17] has been employed only to stable LTI systems.

The kernel system representation of Ghoshal has been extended to linear time-varying systems in [19] where it is used as a deterministic system observer and also in [20] where it is combined with B-spline function approximations to identify time varying system parameters and even unknown system inputs. Recently the kernel representation was also adapted to serve the identification of linear parameter-varying systems [21]. In the last contribution the kernel representation was employed in a moving-window "minimum energy" nonlinear observer to deliver state estimates for a strongly nonlinear system for which the extended Kalman filter is known to fail. The filtering converged very fast due to the presence of a nonlinear differential invariance that was easy to isolate. In the contributions [19]-[21] the identification process is separate from state estimation that is performed after the parameters are already identified. The construction of the kernels was best explained in [19] by way of a low-order example of a linear time-varying system. The calculation of the kernel system representation is conceptually simple as it is related to the derivation of a double-sided Green's function for a linear system with general boundary conditions, see [22], but requires handling long algebraic calculations if the order of the system is high.

In the context of the previous developments as summarized above, the contributions of the present note are summarized as follows:

*This work was supported by The National Science & Engineering Research Council of Canada

[1] Authors are with the Department of Electrical, Computer & Software Engineering, McGill University, 3480 University Street, Montreal H3A 2A7, Quebec, Canada.

[2] Corresponding author hannah.michalska@mcgill.ca

- The kernels of the system representation introduced in [8] are derived for general SISO LTI systems of arbitrary order. Delivering explicit formulae for the kernel representation of high order systems is hoped to encourage further diverse applications of the kernels once the hurdle associated with their calculation is removed.
- A kernel-based recursive least-squares regression algorithm is introduced for simultaneous parameter identification and state estimation from observation of a noisy system output over any finite interval of time.
- The joint state and parameter estimation approach is shown to be strongly robust with respect to additive white Gaussian output measurement noise. While the estimation method assumes the order of the LTI system it does not need any information regarding the initial conditions of the system or knowledge of the variance of the noise.

As the analytic expressions for all kernels are rendered available, the order of the LTI system that best fits the output data can also become subject to estimation. Attempts in this direction were already made in [23]. This task is, however, best carried out in a fully stochastic setting of a kernel-based approach which is currently developed in a reproducing kernel Hilbert space setting.

II. A DOUBLE-SIDED KERNEL REPRESENTATION OF A SISO LTI SYSTEM OF ARBITRARY ORDER

A. General n-th order SISO LTI system and assumptions

Consider a general n-th order, strictly proper, and minimal SISO LTI system in state space form evolving on a given finite time interval $[a,b] \subset \mathbb{R}$:

$$\dot{x} = Ax + bu \ ; \quad y = c^T x \ ; \ x \in \mathbb{R}^n \tag{1}$$

with matching dimensions of the system matrices and characteristic equation

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0 \tag{2}$$

The input-output equation for system (1) becomes

$$\begin{aligned} y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) \\ = -b_{n-1}u^{(n-1)}(t) - \cdots - b_0 u(t) \end{aligned} \tag{3}$$

where $-b_i, i = 0, n-1$ are the coefficients of the polynomial in the numerator of the rational transfer function for (1).

In the light of the assumptions made, the system is differentially flat, so that the entire state vector $x(t)$, for $t \in [a,b]$, can be instantaneously recovered from the knowledge of the input and output functions $u(t), y(t), t \in [a,b]$ and their derivatives $u^{(i)}(t), y^{(i)}(t); i = 1, ..., n-1, t \in [a,b]$; for the exact formulae see e.g. [20].

B. A Differential invariant and its controlled version

The left and right hand sides of (3) involve differential operators F and U acting on the system output and input in suitably defined spaces of differentiable functions (such as the Sobolev space $H_n^2[a,b]$ of functions whose n-th derivatives are absolutely integrable), so that (3) can be compactly written as

$$F(y)(t) = U(u)(t); t \in [a,b] \tag{4}$$

If the system input is constant (or zero) then $F(y)(t) = const.$ for all times, thus F represents a differential invariant whose value remains constant under the action of the flow of such a system. When the input u varies in time the equality (4) delivers what is referred to as a *controlled invariance*.

C. A Kernel representation of the n-order SISO LTI system

Employing the controlled invariance of (4), which, in this particular case, happens to coincide with the input-output equation, it is easy to construct an initial-condition free integral representation of the system (1). Before stating the result a definition is helpful.

Definition 1:
A pair of smooth (class $\mathcal{C}^\infty$) functions (α_a, α_b), $\alpha_s : [a,b] \to \mathbb{R}, s = a$ or b, is an annihilator of the boundary conditions for system (1) if the functions α_s are non-negative, monotonic, vanish with their derivatives up to order $n-1$ at the respective ends of the interval $[a,b]$; i.e.

$$\alpha_s^{(i)}(s) = 0 \quad i = 0, \cdots, n-1; \quad s = a, b; \ \alpha_s^{(0)} \equiv \alpha_s \tag{5}$$

such that their sum is strictly positive, i.e. that for some constant $c > 0$

$$\alpha_{ab}(t) := \alpha_a(t) + \alpha_b(t) > c \quad t \in [a,b] \tag{6}$$

A simplest example of such an annihilator for (1) is the pair

$$\begin{aligned} &\alpha_a(t) := (t-a)^n, \quad \alpha_b(t) := (b-t)^n \quad t \in [a,b] \\ &\alpha_{ab}(t) := \alpha_a(t) + \alpha_b(t) > 0 \quad t \in [a,b] \\ &\alpha_{ab}(s) = (b-a)^n, \quad s = a, b \end{aligned} \tag{7}$$

Indeed, it is easy to see that (7) holds for all $n \geq 1$ because $\alpha_{ab}(a) = \alpha_{ab}(b) = (b-a)^n > 0$ and, for $n \geq 2$, α_{ab} has a unique stationary point $t^* = 0.5(a+b) \in [a,b]$ at which $\alpha_{ab}(t^*) = (0.5)^{n-1}(b-a)^n$.

Employing this particular annihilator the integral representation for system (1) is rendered by the following:

Theorem 1: There exist Hilbert-Schmidt kernels $K_{DS,y}$, $K_{DS,u}$, such that the input and output functions u and y of (1) satisfy

$$\begin{aligned} y(t) = \alpha_{ab}^{-1}(t) \Bigg[\int_a^b K_{DS,y}(n,t,\tau) y(\tau) d\tau \\ + \int_a^b K_{DS,u}(n,t,\tau) u(\tau) d\tau \Bigg] \end{aligned}$$

with

$$\alpha_{ab}^{-1}(t) := \frac{1}{(t-a)^n + (b-t)^n} \tag{8}$$

Hilbert-Schmidt double-sided kernels of (8) are square integrable functions on $L^2[a,b] \times L^2[a,b]$ and are expressed in terms of the "forward" and "backward" kernels for the respective integral representations of the operators $\alpha_a F$ and $\alpha_b F^*$, with the operator F^* denoting the formal adjoint of F:

$$K_{DS,y}(n,t,\tau) \triangleq \begin{cases} K_{F,y}(n,t,\tau), & \text{for } \tau \le t \\ K_{B,y}(n,t,\tau), & \text{for } \tau > t \end{cases}$$

$$K_{DS,u}(n,t,\tau) \triangleq \begin{cases} K_{F,u}(n,t,\tau), & \text{for } \tau \le t \\ K_{B,u}(n,t,\tau), & \text{for } \tau > t \end{cases}$$

The kernel functions $K_{DS,y}, K_{DS,u}$ are $n-1$ times differentiable as functions of t.

□

The easiest proof of the representation theorem is conducted by mathematical induction on n, delivering explicit formulae for both kernels, as stated next.
Full proof can be found in the forthcoming publication [22]. The kernels of Theorem 1 have the following expressions.

$$K_{F,y}(n,t,\tau) = \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}\frac{n!(t-\tau)^{j-1}(\tau-a)^{n-j}}{(n-j)!(j-1)!} + \sum_{i=0}^{n-1} a_i \sum_{j=0}^{i}(-1)^{j+1}\binom{i}{j}\frac{n!(t-\tau)^{n-i+j-1}(\tau-a)^{n-j}}{(n-j)!(n-i+j-1)!}$$

$$K_{B,y}(n,t,\tau) = \sum_{j=1}^{n}\binom{n}{j}\frac{n!(t-\tau)^{j-1}(b-\tau)^{n-j}}{(n-j)!(j-1)!} + \sum_{i=0}^{n-1} a_i \sum_{j=0}^{i}\binom{i}{j}\frac{n!(t-\tau)^{n-i+j-1}(b-\tau)^{n-j}}{(n-j)!(n-i+j-1)!}$$

$$K_{F,u}(n,t,\tau) = \sum_{i=0}^{n-1} b_i \sum_{j=0}^{i}(-1)^{j+1}\binom{i}{j}\frac{n!(t-\tau)^{n-i+j-1}(\tau-a)^{n-j}}{(n-j)!(n-i+j-1)!}$$

$$K_{B,u}(n,t,\tau) = \sum_{i=0}^{n-1} b_i \sum_{j=0}^{i}\binom{i}{j}\frac{n!(t-\tau)^{n-i+j-1}(b-\tau)^{n-j}}{(n-j)!(n-i+j-1)!}$$

Explicit kernel expressions for the derivatives of the output function

Due to the regularity properties of the kernel functions in Theorem 1, it is straightforward to obtain the corresponding recursive formulae for the time derivatives of the system output $y^{(i)}, i = 1, \cdots, n-1$.

Theorem 2: There exist Hilbert-Schmidt kernels $K_{F,k,y}$, $K_{F,k,u}$, $K_{B,k,y}$, $K_{B,k,u}$, $k = 1, \cdots, n-1$ such that the derivatives of the output function in (1) can be computed rercursively as follows:

$$\begin{aligned} y^{(k)}(t) = \frac{1}{(t-a)^n + (b-t)^n}\Bigg[& \sum_{i=1}^{k}(-1)^{i+1}\binom{p+i-1}{i}\frac{n!(t-a)^{n-i}y^{(k-i)}(t)}{(n-i)!} \\ & + \sum_{i=p}^{n-1} a_i \sum_{j=0}^{i-p}(-1)^{j+1}\binom{p+j-1}{j}\frac{n!(t-a)^{n-j}y^{(i-j-p)}(t)}{(n-j)!} \\ & + \int_a^t K_{F,k,y}(n,p,t,\tau)y(\tau)d\tau \\ & + \sum_{i=p}^{n-1} b_i \sum_{j=0}^{i-p}(-1)^{j+1}\binom{p+j-1}{j}\frac{n!(t-a)^{n-j}u^{(i-j-p)}(t)}{(n-j)!} \\ & + \int_a^t K_{F,k,u}(n,p,t,\tau)u(\tau)d\tau \\ & - \sum_{i=1}^{k}\binom{p+i-1}{i}\frac{n!(b-t)^{n-i}y^{(k-i)}(t)}{(n-i)!} \\ & - \sum_{i=p}^{n-1} a_i \sum_{j=0}^{i-p}\binom{p+j-1}{j}\frac{n!(b-t)^{n-j}y^{(i-j-p)}(t)}{(n-j)!} \\ & + \int_t^b K_{B,k,y}(n,p,t,\tau)y(\tau)d\tau \\ & - \sum_{i=p}^{n-1} b_i \sum_{j=0}^{i-p}\binom{p+j-1}{j}\frac{n!(b-t)^{n-j}u^{(i-j-p)}(t)}{(n-j)!} \\ & + \int_t^b K_{B,k,u}(n,p,t,\tau)u(\tau)d\tau\Bigg] \end{aligned}$$

where

$$p = n-k$$

$$\begin{aligned} K_{F,k,y}(n,p,t,\tau) = & \sum_{j=1}^{p}(-1)^{j+n-p+1}\binom{n}{n-p+j}\frac{n!(t-\tau)^{j-1}(\tau-a)^{p-j}}{(p-j)!(j-1)!} \\ & + \sum_{i=0}^{p-1} a_i \sum_{j=0}^{i}(-1)^{j+1}\binom{i}{j}\frac{n!(t-\tau)^{p-i+j-1}(\tau-a)^{n-j}}{(n-j)!(p-i+j-1)!} \\ & + \sum_{i=p}^{n-1} a_i \sum_{j=1}^{p}(-1)^{j+i-p+1}\binom{i}{i-p+j} \\ & \frac{n!(t-\tau)^{j-1}(\tau-a)^{n-i+p-j}}{(n-i+p-j)!(j-1)!} \end{aligned}$$

$$\begin{aligned} K_{F,k,u}(n,p,t,\tau) = & \sum_{i=0}^{p-1} b_i \sum_{j=0}^{i}(-1)^{j+1}\binom{i}{j}\frac{n!(t-\tau)^{p-i+j-1}(\tau-a)^{n-j}}{(n-j)!(p-i+j-1)!} \\ & + \sum_{i=p}^{n-1} b_i \sum_{j=1}^{p}(-1)^{j+i-p+1}\binom{i}{i-p+j} \\ & \frac{n!(t-\tau)^{j-1}(\tau-a)^{n-i+p-j}}{(n-i+p-j)!(j-1)!} \end{aligned}$$

$$K_{B,k,y}(n,p,t,\tau)=\sum_{j=1}^{p}\binom{n}{n-p+j}\frac{n!(t-\tau)^{j-1}(b-\tau)^{p-j}}{(p-j)!(j-1)!}+\sum_{i=0}^{p-1}a_i\sum_{j=0}^{i}\binom{i}{j}\frac{n!(t-\tau)^{p-i+j-1}(b-\tau)^{n-j}}{(n-j)!(p-i+j-1)!}+\sum_{i=p}^{n-1}a_i\sum_{j=1}^{p}\binom{i}{i-p+j}\frac{n!(t-\tau)^{j-1}(b-\tau)^{n-i+p-j}}{(n-i+p-j)!(j-1)!}$$

$$K_{B,k,u}(n,p,t,\tau)=\sum_{i=0}^{p-1}b_i\sum_{j=0}^{i}\binom{i}{j}\frac{n!(t-\tau)^{p-i+j-1}(b-\tau)^{n-j}}{(n-j)!(p-i+j-1)!}+\sum_{i=p}^{n-1}b_i\sum_{j=1}^{p}\binom{i}{i-p+j}\frac{n!(t-\tau)^{j-1}(b-\tau)^{n-i+p-j}}{(n-i+p-j)!(j-1)!}$$

□

Full proof can be found in the forthcoming publication [22].

D. Properties of the kernel system representation

The kernel representation for an input-free system follows immediately with some interesting consequences that can be exploited in filtering, smoothing, and identification.

Corollary 1:
The kernel system representation of a linear homogeneous differential equation (3) (with input $u \equiv 0$) takes the form of a continuous linear evaluation functional

$$y(t)=\alpha_{ab}^{-1}(t)\int_a^b K_{DS,y}(n,t,\tau)y(\tau)d\tau \quad t\in[a,b] \tag{9}$$

The output derivative kernels are then obtained from these of Theorem 2 by substituting $b_0=b_1=\cdots=b_{n-1}=0$.

By virtue of the continuity of the evaluation functional (9) on the space $L^2[a,b]$, the symmetric, positive-type kernel defined by $K(t_1,t_2)\triangleq\langle K_{DS}(t_1,\cdot)|K_{DS}(t_2,\cdot)\rangle_2$ for all $t_1,t_2\in[a,b]$ where $\langle\cdot|\cdot\rangle_2$ denotes the scalar product in $L^2[a,b]$ induces a reproducing kernel Hilbert space, here denoted by $\mathcal{H}_K$, that is isomorphic with the image of the space $L^2[a,b]$ under the integral transform defined by the double-sided kernel $K_{DS,y}$ of (9) with a K-weighted norm as defined in [24].

□

The proof of the representation Theorem 1 is essentially conducted by construction, involving the induction argument only at its final stage. The construction of the kernels for $n=2$ can be followed step-by step in [19] and is seen to involve only reversible mathematical operations such as pre-multiplication by annihilator functions, iterated integrations, and integration by parts. The only loss of information in the passage from the input-output equation of the system to the integral kernel representation of Theorem 1 is that of any pre-existing boundary conditions in (3) as the latter are annihilated during every integration operation by the presence of the annihilating factors α_a and α_b. The following conjecture is then quite obvious.

Corollary 2:
For any given input function u, the output function $y:[a,b]\to\mathbb{R}$ satisfies the system input-output equation (3) on the interval $[a,b]$ if an only if it satisfies the integral equation (8) regardless of any boundary conditions that may be imposed. For the homogeneous system with zero input and given coefficients of its characteristic equation, the smallest linear space containing all functions that satisfy the reproducing property (9) is finite dimensional as it is spanned by all fundamental solutions of the homogeneous system.

□

III. Simultaneous finite-interval state and parameter estimation

For a given input function u, the estimation problem for system (3) reduces to that of a homogeneous system as the influence of the input can be factored out from the output measurement prior to the estimation procedure. The double-sided kernel $K_{DS,y}$ are clearly linear with respect to the system parameters $a_i, i=0,\cdots,n-1$, so

$$y(t)=\int_a^b K_{DS,y}(n,t,\tau)y(\tau)\,d\tau \tag{10}$$

$$=\int_a^b\sum_{i=0}^{n}\tilde{a}_iK_{DS(i),y}(n,t,\tau)y(\tau)d\tau \tag{11}$$

where the $K_{DS(i),y}; i=0,\cdots,n$ are some "component kernels" of $K_{DS,y}$ with $\tilde{a}_i=a_i; i=0,\cdots,n-1$, and $\tilde{a}_n=1$.

Commuting integration with summation then permits to write

$$y(t)-g_n(n,t,y)=\sum_{i=0}^{n-1}a_ig_i(n,t,y) \tag{12}$$

with

$$g_i(n,t,y):=\int_a^b K_{DS(i),y}(n,t,\tau)y(\tau)d\tau \quad i=0,\dots n \tag{13}$$

Given distinct time instants $t_1,\cdots,t_N\in(a,b]$, here referred to as *knots*, the last equation is re-written point-wise in the form of a matrix equation

$$Q(y)=P(y)\bar{a} \tag{14}$$

$$Q\overset{def}{=}\begin{bmatrix}q(t_1)\\ \vdots\\ q(t_N)\end{bmatrix};\quad \bar{a}\overset{def}{=}\begin{bmatrix}\tilde{a}_0\\ \vdots\\ \tilde{a}_{n-1}\end{bmatrix}; \tag{15}$$

$$P\overset{def}{=}\begin{bmatrix}p_0(t_1)\cdots p_{n-1}(t_1)\\ \ddots\\ p_0(t_N)\cdots p_{n-1}(t_N)\end{bmatrix}$$

$$q(t_i)=y(t_i)-g_n(n,t_i,y);\quad p_k(t_i)=g_k(n,t_i,y) \tag{16}$$

Equation (14) is the parametric estimation equation to be solved by least squares error minimization.

The above system of linear algebraic equations can be solved to find the value of the system parameters $\bar{a}$ provided that an identifiability condition is satisfied. In the presence of noise the latter is best stated in practical terms as

Definition 2: Practical linear identifiability
The homogeneous system (3) with $u \equiv 0$ is practically linearly identifiable on $[a, b]$ with respect to a particular noisy realization of the output measurement process, $y(t), t \in [a, b]$, if and only if there exist distinct knots $t_1, \cdots, t_N \in (a, b]$ which render $rank P(y) = n$. Any such output realization is then called *persistent* as it yields an estimate $\bar{a} := P(y)^{\dagger} Q(y)$, where the pseudo-inverse $P^{\dagger}$ is the left inverse of $P(y)$.

Let a given noisy realization of the observed system output $y_M(t), t \in [a, b]$, be persistent and the observation noise be white Gaussian and additive. In search for the knots for which identifiability holds and to compensate for the noise, the search for the system parameters $\bar{a}$ is best solved by way of a recursive least-square regression algorithm (RLS); see e.g. [25], [26], [27], [28] for standard assumptions and statistical properties.

For simplicity of the example computations presented, the RLS employed here will use batches of N distinct knots generated randomly using a uniform distribution over $[a, b]$.

The parameter and state estimation will be carried out simultaneously using RLS by first obtaining the estimates of the system output and its derivatives $y_E^{(i)}(t), t \in [a, b], i = 0, 1, \cdots, n-1$, by evaluating the right hand sides of the equations of Theorems 1 and 2 at the observed output realization $y_M(t), t \in [a, b]$. In this calculation, the kernel functions will use the estimates of the system parameters $\bar{a}$ as returned from RLS. The estimated system state $x_E(t), t \in [a, b]$ will finally be recovered as a function of $y_E^{(i)}(t), t \in [a, b]$ as explained in II-A, with exact expressions cited in [20].

A. Recursive least squares algorithm [28]

The estimation model which coincides with the estimation equation (14) at iteration k is assumed to be

$$Q_k = P_k \bar{a}_k + e_k \quad , \qquad k = 0, 1, \ldots \tag{17}$$

where $Q_k \in \mathbb{R}^{N \times 1}$, and $P_k \in \mathbb{R}^{N \times n}$, are evaluated at the k-th batch of knots sampled from $y_M(t), t \in [a, b]$; with $\bar{a}_k \in \mathbb{R}^n$, $e_k \in \mathbb{R}^N$ for all k , and where the vectors $\bar{a}$ and e_k represent the parameter estimate and observation noise, respectively.

At each iteration k the parameter estimate is sought as

$$\hat{a}_k = \arg\min_{\bar{a}} \left(\sum_{i=1}^{k} (Q_i - P_i \bar{a})'(Q_i - P_i \bar{a}) \right) \tag{18}$$

Introducing aggregated matrices

$$\bar{Q}_{k+1} = \begin{bmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_{k+1} \end{bmatrix}; \quad \bar{P}_{k+1} = \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_{k+1} \end{bmatrix}; \quad \bar{e}_{k+1} = \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{k+1} \end{bmatrix} \tag{19}$$

the algorithm strives to minimize

$$\begin{aligned} & \min(\bar{e}'_{k+1} \bar{e}_{k+1}) \\ \text{subject to:} \quad & \bar{Q}_{k+1} = \bar{P}_{k+1} \bar{a}_{k+1} + \bar{e}_{k+1} \end{aligned}$$

to compute the next estimate $\hat{a}_{k+1}$. The solution of the above least-squares problem is written as

$$(\bar{P}'_{k+1} \bar{P}_{k+1}) \hat{a}_{k+1} = \bar{P}'_{k+1} \bar{Q}_{k+1} \tag{20}$$

or in summation form as

$$\left(\sum_{i=0}^{k+1} P'_i P_i \right) \hat{a}_{k+1} = \sum_{i=0}^{k+1} P'_i Q_i \tag{21}$$

Defining

$$M_{k+1} = \sum_{i=0}^{k+1} P'_i P_i \tag{22}$$

the recursion for M_{k+1} is:

$$M_{k+1} = M_k + P'_{k+1} P_{k+1} \tag{23}$$

Rearranging (21) gives

$$\begin{aligned} \hat{a}_{k+1} &= M_{k+1}^{-1} \left[\left(\sum_{i=0}^{k} P'_i P_i \right) \hat{a}_k + P'_{k+1} Q_{k+1} \right] \\ &= M_{k+1}^{-1} \left[M_k \hat{a}_k + P'_{k+1} Q_{k+1} \right] \end{aligned} \tag{24}$$

Another form of (24) is delivered by the recursion (23) and reads

$$\begin{aligned} \hat{a}_{k+1} &= \hat{a}_k - M_{k+1}^{-1} (P'_{k+1} P_{k+1} \hat{a}_k - P'_{k+1} Q_{k+1}) \\ &= \hat{a}_k + M_{k+1}^{-1} P'_{k+1} (Q_{k+1} - P_{k+1} \hat{a}_k) \end{aligned} \tag{25}$$

A recursion for M_{k+1}^{-1} is obtained by using the identity to the recursion in (23)

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B (DA^{-1} B + C^{-1})^{-1} DA^{-1} \tag{26}$$

which yields

$$M_{k+1}^{-1} = M_k^{-1} - M_k^{-1} P'_{k+1} (P_{k+1} M_k^{-1} P'_{k+1})^{-1} P_{k+1} M_k^{-1}$$

Defining $R_{k+1} = M_{k+1}^{-1}$ the latter becomes

$$R_{k+1} = R_k - R_k P'_{k+1} (P_{k+1} R_k P'_{k+1})^{-1} P_{k+1} R_k \tag{27}$$

Equations (25) and (27) constitute the RLS. The stopping criterion employed here is:

$$\|\hat{a}_{k+1} - \hat{a}_k\| < \epsilon \text{ for some } \epsilon > 0 \tag{28}$$

Further modifications and adaptation of the RLS is presented in forthcoming publication [22].

IV. Results

Example system

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -150 & -125 & -31 & -5 \end{bmatrix} x \quad ; \quad y = x_1 \quad ; \tag{29}$$

$$x(0) = [1, 1, 1, 1] \tag{30}$$

The characteristic equation is

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{31}$$

with unknown, nominal values of parameters

$$a_0 = 150 \quad , \quad a_1 = 125 \quad , \quad a_2 = 31 \quad , \quad a_3 = 5 \tag{32}$$

A. Low noise

White Gaussian noise with 0.1 standard deviation was added to the nominal y trajectory to simulate a noisy observation y_M. Figure 1 shows y vs. y_M. The RLS uses batches of $N = 100$ knots sampled from a uniform distribution over $[a, b]$ where a, b are 0 and 5 second, respectively. The threshold value for stopping criterion is $\epsilon = 0.05$. The final parameter estimates are:

$$a_0 = 116.155, a_1 = 104.250, a_2 = 30.062, a_3 = 4.194 \tag{33}$$

Figures 2, 3, 4, 5 show the estimations y_E, $y_E^{(1)}$, $y_E^{(2)}$, $y_E^{(3)}$, respectively, calculated as explained in section III.

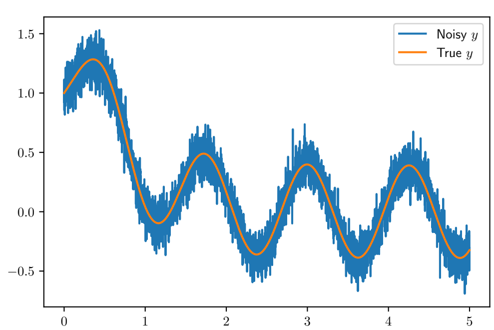

Fig. 1. Noise of 0.1 SD added to y

B. High noise

Additive white Gaussian noise with 1.0 standard deviation is used to simulate y_M. The parameters of the RLS and the estimation method used are the same as in the case of the low observation noise. The final parameter estimates are:

$$a_0 = 45.376, a_1 = 51.867, a_2 = 23.366, a_3 = 2.218 \tag{34}$$

Figure 6 shows y vs. y_M. Figures 7, 8, 9, 10 show the estimated y_E, $y_E^{(1)}$, $y_E^{(2)}$, $y_E^{(3)}$ calculated employing y_M in the right hand sides of the equations of Theorems 1 and 2.

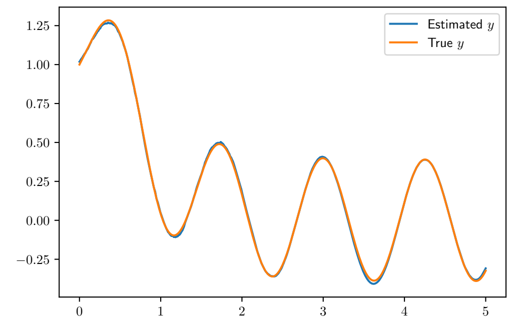

Fig. 2. Estimation of y from low noise

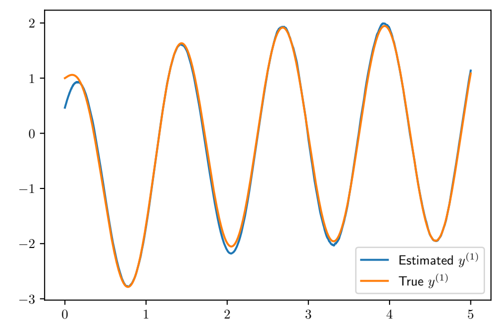

Fig. 3. Estimation of $y^{(1)}$ from low noise

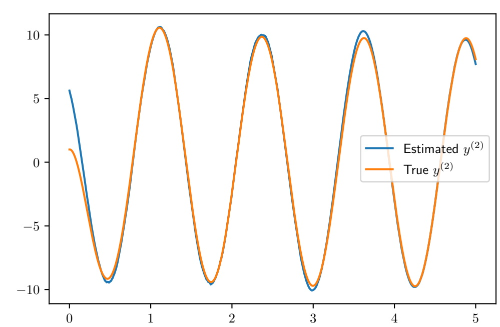

Fig. 4. Estimation of $y^{(2)}$ from low noise

C. Discussion of results

Generally, the estimated trajectories exhibit large errors close to the ends of the interval $[a, b]$. This is easily explained by first noting that the kernel-based estimation presented here is closely related to L-spline function approximation; see [29], and [30]. The L-spline resulting from the development in this note is, however, rather special as the associated output approximation does not require using projections onto the direct sum of subspaces of which the first retains the information about the initial conditions for the L-spline and the second determines its shape and regularity. Due to the action of the annihilator functions

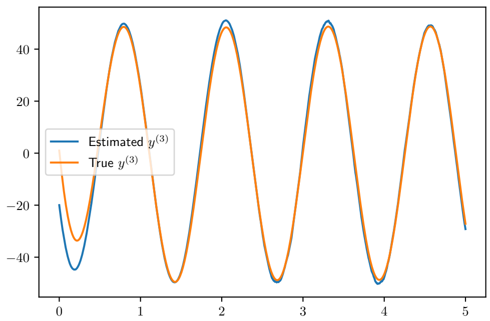

Fig. 5. Estimation of $y^{(3)}$ from low noise

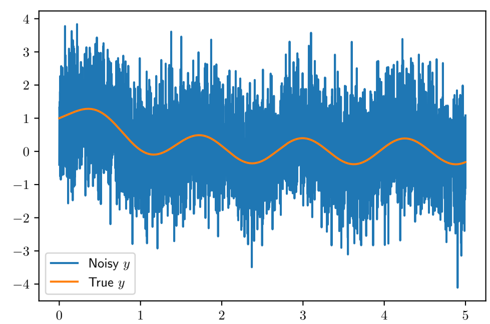

Fig. 6. Noise of 1 SD added to y

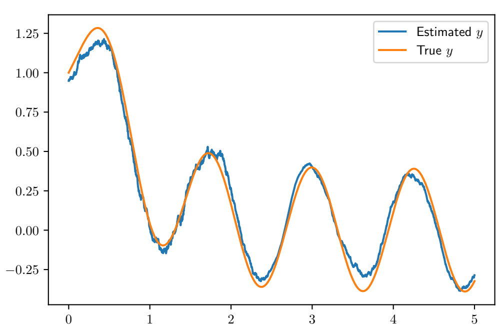

Fig. 7. Estimation of y from high noise

in Theorem 1 the initial conditions are irrelevant as they result from the approximated function by virtue of the derivative formulae of Theorem 2. This fact is alluded to in the statement of Theorem 1, where it is mentioned that the kernels are essentially generating a space of L-splines in which the differential operator $L := \alpha_a F + \alpha_b B$ with $B := F^*$, the formal adjoint to F (here F and B correspond to the "forward" and "backward" versions of equation (3) with input $u \equiv 0$). The exciting correspondence between spline approximation and Bayesian filtering can be contemplated by referring to [31].

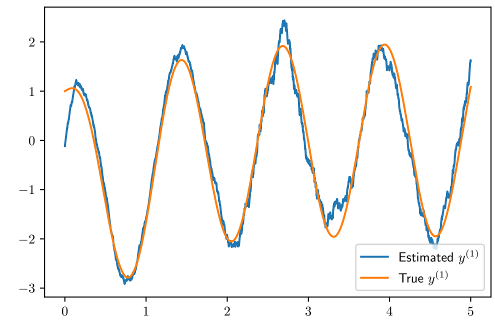

Fig. 8. Estimation of $y^{(1)}$ from high noise

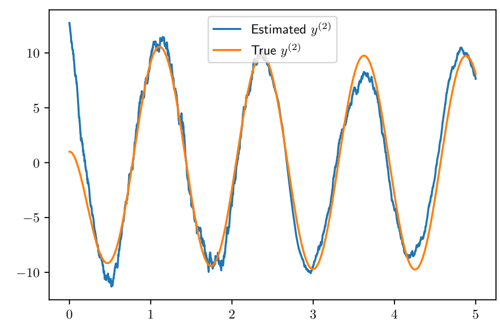

Fig. 9. Estimation of $y^{(2)}$ from high noise

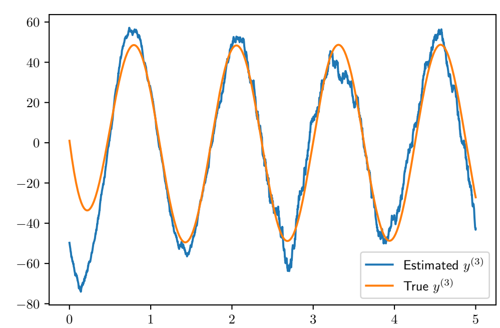

Fig. 10. Estimation of $y^{(3)}$ from high noise

Next, as is widely known, any spline approximation on a closed interval will give worse estimation error at the extremities of the interval unless the *spline knots* are chosen optimally, which usually corresponds to a denser population of knots at these extremities. Translating this fact into the case of the adapted RLS employed here, better results are highly likely if a non-uniform density distribution were used to generate the knot batches. Such modifications of the RLS is explored in forthcoming publication [22].

The estimation approach exhibits, overall, a promising performance. By comparing the low and high noise cases, its robustness to noise is impressive, given that no prior information is available about neither: system parameters, initial or boundary conditions, and variance of the observation noise. Needless to say, a variety of methods can be applied to further smooth the higher derivative estimates.

V. Conclusions

Now, that the explicit formulae for the special forward-backward kernels are delivered, various more advanced applications come to mind such as those in the areas of identification of linear parameter-varying systems, and nonlinear filtering; see [21] for first attempts in such directions.

Somewhat surprisingly, the computational burden associated with the kernel method does not change much with the increasing system order. The increase in computational complexity is by and large limited to the algebraic complexity of the kernel expressions which now are readily available in closed form and need a single finite-interval integration.

The kernel estimation approach embedded in the RKHS spaces of stochastic processes; see section II-D and Corollaries 1 and 2, is expected to yield further improvements and is the current research effort. Such stochastic setting will permit to incorporate any available characteristic of the observation noise. The approach may then be considered a viable alternative to the Wiener filter.

References

[1] L. Ljung, *System identification: theory for the user*. Prentice-hall, 1987.

[2] T. Kailath, "Lectures on wiener and kalman filtering," in *Lectures on Wiener and Kalman Filtering*. Springer, 1981, pp. 1–143.

[3] K. J. Åström and P. Eykhoff, "System identificationa survey," *Automatica*, vol. 7, no. 2, pp. 123–162, 1971.

[4] C. K. Chui, G. Chen *et al.*, *Kalman filtering*. Springer, 2017.

[5] R. Alur, T. A. Henzinger, and E. D. Sontag, *Hybrid systems III: verification and control*. Springer Science & Business Media, 1996, vol. 3.

[6] M. Fliess and H. Sira-Ramirez, "Control via state estimations of some nonlinear systems," in *IFAC Symposium on Nonlinear Control Systems (NOLCOS 2004)*, 2004.

[7] Y. Bar-Shalom, "Multitarget-multisensor tracking: Applications and advances. volume iii," *Norwood, MA, Artech House, Inc., 2000*, 2000.

[8] D. P. Ghoshal, K. Gopalakrishnan, and H. Michalska, "Using invariance to extract signal from noise," in *American Control Conference*, 2017.

[9] M. Fliess, C. Join, and H. Sira-Ramirez, "Non-linear estimation is easy," *International Journal of Modelling, Identification and Control*, vol. 4, no. 1, pp. 12–27, 2008.

[10] H. Sira-Ramírez, C. G. Rodríguez, J. C. Romero, and A. L. Juárez, *Algebraic identification and estimation methods in feedback control systems*. John Wiley & Sons, 2014.

[11] M. Mboup, C. Join, and M. Fliess, "A revised look at numerical differentiation with an application to nonlinear feedback control," in *The 15th Mediterrean Conference on Control and Automation-MED'2007*, 2007.

[12] ——, "Numerical differentiation with annihilators in noisy environment," *Numerical algorithms*, vol. 50, no. 4, pp. 439–467, 2009.

[13] M. Fliess, "Analyse non standard du bruit," *Comptes Rendus Mathematique*, vol. 342, no. 10, pp. 797–802, 2006.

[14] D. P. Ghoshal, K. Gopalakrishnan, and H. Michalska, "Algebraic parameter estimation using kernel representation of linear systems," in *The 20th World Congress of the International Federation of Automatic Control*, 2017.

[15] R. Mohd-Mokhtar and L. Wang, "Continuous time system identification using subspace methods," *ANZIAM Journal*, vol. 47, pp. 712–732, 2007.

[16] G. Mercère, R. Ouvrard, M. Gilson, and H. Garnier, "Subspace based methods for continuous-time model identification of mimo systems from filtered sampled data," in *Control Conference (ECC), 2007 European*. IEEE, 2007, pp. 4057–4064.

[17] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.

[18] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[19] D. P. Ghoshal and H. Michalska, "Double-sided kernel observer for linear time-varying systems," in *IEEE Conference on Control Technology and Applications*, 2017.

[20] D. Sridhar, D. P. Ghoshal, and H. Michalska, "B-splines in joint parameter and state estimation in linear time-varying systems," in *American Control Conference*, 2018.

[21] D. Sridhar and H. Michalska, "Non-asymptotic state and input estimation for smooth linear parameter varying systems," in *IEEE Conference on Decision and Control*, 2018.

[22] D. P. Ghoshal and H. Michalska, "Forward-backward kernel-based state and parameter estimation for linear systems of arbitrary order," *(submitted) International Journal of Control*.

[23] D. P. Ghoshal, K. Gopalakrishnan, and H. Michalska, "Kernel-based adaptive multiple model target tracking," in *IEEE Conference on Control Technology and Applications*, 2017.

[24] S. Saitoh, *Theory of reproducing kernels and its applications*. Longman, 1988, vol. 189.

[25] C. S. Leung, K. W. Wong, P. F. Sum, and L. W. Chan, "On-line training and pruning for recursive least square algorithms," *Electronics Letters*, vol. 32, no. 23, pp. 2152–2153, 1996.

[26] C. S. Leung, G. H. Young, J. Sum, and W.-K. Kan, "On the regularization of forgetting recursive least square," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1482–1486, 1999.

[27] G. Wang, T. Li, G. Zhang, X. Gui, and D. Xu, "Position estimation error reduction using recursive-least-square adaptive filter for model-based sensorless interior permanent-magnet synchronous motor drives," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 9, pp. 5115–5125, 2014.

[28] M. Dahleh, M. A. Dahleh, and G. Verghese, "Lectures on dynamic systems and control," *A+ A*, vol. 4, no. 100, pp. 1–100, 2004.

[29] L. Schumaker, *Spline functions: basic theory*. Cambridge University Press, 2007.

[30] G. Wahba, *Spline models for observational data*. Siam, 1990, vol. 59.

[31] G. S. Kimeldorf and G. Wahba, "A correspondence between bayesian estimation on stochastic processes and smoothing by splines," *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 495–502, 1970.