Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 50-1 (2017) 12898–12904

Algebraic Parameter Estimation Using Kernel Representation of Linear Systems ⋆

Debarshi Patanjali Ghoshal * **Kumar Gopalakrishnan** ** **Hannah Michalska** ***

* *Authors are with the Department of Electrical, Computer and Software Engineering, McGill University, 3480 University Street, Montreal H3A 2A7, Quebec, Canada (e-mail: debarshi.ghoshal@mail.mcgill.ca)*
** *(e-mail: kumar.gopalakrishnan@mail.mcgill.ca)*
*** *(e-mail: hannah.michalska@mcgill.ca)*

Abstract: This work makes a contribution to algebraic parameter estimation as it proposes a simple alternative to the derivation of the algebraic estimation equations. The idea is based on a system representation in the form of an evaluation functional which does not exhibit any singularities in the neighbourhood of zero. Implied is the fact that algebraic estimation of parameters as well as system states can then truly be performed in arbitrary time and with uniform accuracy over the entire estimation interval. Additionally, the result offers a geometric representation of a linear system as a finite dimensional subspace of a Hilbert space, that readily suggests powerful noise rejection methods in which invariance plays a central role.

Keywords: Algebraic Parameter and State Estimation, Dead-Beat Observers.

1. INTRODUCTION

The work presented here was inspired by the results of Fliess and Sira-Ramirez (2003), Fliess and Sira-Ramirez (2008), who proposed a framework for an algebraic approach to linear parameter estimation. Although the literature on the subject of system identification and state estimation is extremely rich; see an excellent compendium in Gevers (2006), the importance of the algebraic approach should be strongly emphasized due to the following superior attributes: (i) in principle (in the absence of noise), algebraic estimation is not only "dead-beat", but is able to deliver parameter and state estimates in *arbitrary short time* provided that a system output trajectory is measured continuously in time; (ii) even if measurement error is present, the estimators exhibit natural noise rejection properties as they employ multiple integration to convert differential systems into integral systems; (iii) the initial conditions of the system are rendered irrelevant as algebraic estimation operates on output trajectories rather than data points. In practice, continuous measurement of the system output is replaced by high-frequency discrete time measurement; the sampling rate is then adjusted to the estimation horizon. The noise attenuation properties of iterated integrals are adequately dicussed in Fliess (2006).

Given the above attributes the pursuit of improved algebraic estimation methods has a clear goal: the development of a next generation exact estimators for hybrid and nonlinear systems. Fast switching requires dead-beat observers and the theory of flat nonlinear systems warrants the existence of differential system invariants that can be viewed as nonlinear versions of linear system invariants as delivered by the Cayley Hamilton Theorem.

In the light of the above the following research objectives are of priority: (a) improving the estimation accuracy over arbitrary estimation horizons; (b) developing noise "annihilation" methods that are capable of system model reconstruction from output measurements subject to noise of arbitrary characteristics. We use the word "annihilation" to stress the fact that to deliver faithful estimates of high order time derivatives of a measured noisy output signal the derivatives must be obtained from a "reconstructed" trajectory of the model that best "fits the noise".

In this paper we address the first objective i.e. (a). Backward integration and the Cauchy formula is employed to convert a linear high order differential equation, that embodies a system invariant, into an integral form with no singularities at the boundaries of the observation window. For simplicity and clarity of exposition, but without any loss of generality we focus on homogeneous system equations (pure *torsion* systems; see Fliess and Sira-Ramirez (2003)); the generalization to SISO systems is immediate. The system invariance then assumes a form of an output reproducing property that unambiguously characterizes the system trajectories on any given observation window. An equivalent system representation takes the form of a subspace of a Hilbert space such as $L^2[a, b]$; the subspace is itself a reproducing kernel Hilbert space. Such geometric system representation is seen to be useful in pursuing the denoisification objective as stated in (b).

Not surprisingly, the alternative integral system representation results in improved accuracy of estimation. It also renders the technique truly valid on estimation windows of

⋆ This work was supported by The National Science and Engineering Research Council of Canada

10.1016/j.ifacol.2017.08.1943

arbitrary length without compromising accuracy. Importantly, the new system representation lends itself directly to all existing improvements and developments of the algebraic estimation theory. The approach extends naturally to LTV systems; details will be presented elsewhere due to limited length of this article.

2. THE ALGEBRAIC PARAMETER ESTIMATION PROBLEM

Although algebraic estimation methods can be extended to multi-input systems, for reasons of simplicity of exposition, we consider only single-input single-output case. A standard algebraic (non-asymptotic) parameter estimation problem for a linear time invariant system is then:

Algebraic Parameter Estimation for LTI Systems

Identify the values of the parameters $a_i, i = 0, .., n-1$, and $b_i, i = 0, ..., m$ in a given input-output system structure:

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) = b_m u^{(m)}(t) + \cdots + b_0 u(t) \tag{1}$$

using noisy observations of the system's input and output $u(t)$ and $y(t)$ over a finite, but arbitrary, interval of time $t \in [0, T]$, $T > 0$.

What makes this estimation problem different is the lack of any assumptions about the initial conditions of the system as well as its non-asymptotic nature - the observation interval is not only finite but, in principle, can be arbitrarily short.

An elegant algebraic framework for the solution of such a problem was first laid out in Fliess and Sira-Ramirez (2003). While the general conditions for parameter identifiability are stated in terms of free modules over the rings of linear differential operators their practical interpretation is actually quite simple and leads to explicit analytic solution to the problem. This is best illustrated in terms of a simple example; see Sira-Ramirez et al. (2014), p. 337, which is also useful to elucidate the contribution of the results presented here.

Example: Consider the system with unknown parameters a and b and unknown initial conditions:

$$y^{(2)}(t) = ay(t) + bu(t) \tag{2}$$

The effect of the initial conditions vanishes if (2) is first multiplied by t^2

$$t^2 y^{(2)}(t) = at^2 y(t) + bt^2 u(t) \tag{3}$$

and then integrated by parts two times over the intervals $[0, \sigma]$ and $[0, t]$ leading to:

$$a\int_0^t\int_0^\sigma \theta^2 y(\theta)\ \mathrm{d}\theta\mathrm{d}\sigma + b\int_0^t\int_0^\sigma \theta^2 u(\theta)\ \mathrm{d}\theta\mathrm{d}\sigma = t^2 y(t) - 4\int_0^t \sigma y(\sigma)\ \mathrm{d}\sigma + 2\int_0^t\int_0^\sigma y(\theta)\ \mathrm{d}\theta\mathrm{d}\sigma \tag{4}$$

Integrating the above once more leads to

$$a\int_0^t\int_0^\sigma\int_0^\theta \lambda^2 y(\lambda)\ \mathrm{d}\lambda\mathrm{d}\theta\mathrm{d}\sigma + b\int_0^t\int_0^\sigma\int_0^\theta \lambda^2 u(\lambda)\ \mathrm{d}\lambda\mathrm{d}\theta\mathrm{d}\sigma = \int_0^t \sigma^2 y(\sigma)\ \mathrm{d}\sigma - 4\int_0^t\int_0^\sigma \theta y(\theta)\ \mathrm{d}\theta\mathrm{d}\sigma + 2\int_0^t\int_0^\sigma\int_0^\theta y(\lambda)\ \mathrm{d}\lambda\mathrm{d}\theta\mathrm{d}\sigma \tag{5}$$

Equations (4) and (5) can now be written in a matrix-vector form

$$P(t; y, u)\begin{bmatrix} a \\ b \end{bmatrix} = Q(t; y) \tag{6}$$

Identifiability of parameters is then insured as long as $\det(P(t; y, u)) \neq 0$ which holds whenever t is away from zero and when the trajectories u and y are *persistent*; see Sira-Ramirez et al. (2014). Needless to mention that identifiability pertains to the *nominal system* as the matrix $P(t : y, u)$ is normally computed based on noisy data $y(\sigma), \sigma \in [0, t]$. Identification equation for the nominal system can thus be solved for $t > \epsilon > 0$ and for $u \neq 0$ and delivers estimates of parameters a, b at any such time instant t.

The singularity at $t = 0$ is more than a mere inconvenience as the choice of the value of $\epsilon > 0$ is not obvious to warrant good conditioning of the solution and also because the observation horizon T is required to be arbitrarily short. Since the trajectories y, u are not persistent when $u \equiv 0$ rendering a singular $P(t; y, 0)$, the case of the estimation of a homogeneous equation (1) deserves separate attention, if only from the point of view of identifiability - a property that obviously fails to hold for $y \equiv 0$.

An alternative representation of the single output system will be introduced in the next section that addresses these problems. The parameter estimation of a homogeneous system can be viewed as the identification of a differential invariant $\mathcal{I}$ ($\mathcal{I} \equiv 0$) under the action of the flow of some closed loop system (such as its characteristic equation):

$$\mathcal{I}(t, y(t), y^{(1)}(t), \cdots, y^{(n)}(t)) = y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_0 y(t)\ ;\ t \geq 0 \tag{7}$$

The estimation of such an invariant might be of independent interest in answering the following question: "can the stability properties of a closed loop system be determined from its output trajectory $y(t), t \in [0, T]$, on an arbitrary small interval of time?"

3. KERNEL REPRESENTATION OF A LINEAR SYSTEM

The problem with singularity at $t = 0$ in the integral system representation that is the cornerstone of "initial-condition independent" system representation can easily be removed by employing two-sided forward-backward integration. This ultimately leads to an alternative integral system representation of (7) in a reproducing kernel Hilbert space (RKHS). Such representation is also advantageous from the point of view of noise attenuation that can be further achieved by regularization as is used in standard spline approximation.

We note the following result (see authors' earlier work Ghoshal et al. (2017) for details):

Theorem 1. There exist Hilbert-Schmidt kernels K_{DS}, K^i_{DS}, $i = 1, \cdots n-1$, such that the output function y of (7) is reproduced on any given interval $[a, b]$ in accordance with the action of the evaluation functional

$$y(t) = \int_a^b K_{DS}(t, \tau) y(\tau)\, \mathrm{d}\tau \ ; \ \forall t \in [a, b] \tag{8}$$

and the derivatives of the output $y^{(1)}, \cdots y^{(n-1)}$ can be computed *recursively* by way of output integration, so that for $i = 1, \cdots n-1$ and for all $t \in [a, b]$:

$$y^{(i)}(t) = \sum_{k=0}^{i-1} b_k(t) y^{(k)}(t) + \int_a^b K^i_{DS}(t, \tau) y(\tau)\, \mathrm{d}\tau \tag{9}$$

where $y^{(0)} \equiv y$ and $b_k(\cdot)$ are rational functions of t. Hilbert-Schmidt kernels are square integrable functions on $L^2[a, b] \times L^2[a, b]$.

While the analytical formulae for the kernels for a general n-dimensional system maybe somewhat cumbersome their derivation is surprisingly straightforward. It is to mention that the derivation of such system representation can equivalently be conducted in the Laplace domain or else by direct embedding of (7) into a Sobolev space $H^2_n[a, b]$ of functions whose n-th derivatives are absolutely integrable. For clarity and simplicity we will only demonstrate the derivation directly in the time domain with reference to a 3-dimensional example. To this end, let $n = 3$, and consider two equations obtained from (7) by premultiplication by the respective factors $(\xi - a)^3$ and $(b - \zeta)^3$:

$$(\xi-a)^3 y^{(3)} + a_2(\xi-a)^3 y^{(2)} + a_1(\xi-a)^3 y^{(1)} + a_0(\xi-a)^3 y = 0 \tag{10}$$

$$(b-\zeta)^3 y^{(3)} + a_2(b-\zeta)^3 y^{(2)} + a_1(b-\zeta)^3 y^{(1)} + a_0(b-\zeta)^3 y = 0 \tag{11}$$

Each of the above are then integrated three times, on the respective intervals $[a, a+\tau]$ and $[b-\sigma, b]$ while assuming that τ and σ are related by $a + \tau = b - \sigma$. Integration by parts is used whenever it allows to lower the degree of the derivatives appearing under the integrals and the result is then simplified algebraically before proceeding to the next integration. To illustrate this process, integrating the first term of (10) yields

$$\int_a^{a+\tau} (\xi - a)^3 y^{(3)}(\xi)\, \mathrm{d}\xi = \tau^3 y^{(2)}(a+\tau) - 3\tau^2 y^{(1)}(a+\tau) + 6\tau y(a+\tau) - \int_a^{a+\tau} 6y(\xi)\, \mathrm{d}\xi \tag{12}$$

Integrating again gives

$$\begin{aligned}&\int_a^{a+\tau}\int_a^{\xi'} (\xi - a)^3 y^{(3)}(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' = \tau^3 y^{(1)}(a+\tau) - 6\tau^2 y(a+\tau) \\ &+ \int_a^{a+\tau} 18(\xi' - a) y(\xi')\, \mathrm{d}\xi' - \int_a^{a+\tau}\int_a^{\xi'} 6y(\xi)\, \mathrm{d}\xi \mathrm{d}\xi'\end{aligned} \tag{13}$$

Integrating the third time yields

$$\begin{aligned}&\int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} (\xi - a)^3 y^{(3)}(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi'' \\ &= \tau^3 y(a+\tau) - \int_a^{a+\tau} 9(\xi'' - a)^2 y(\xi'')\, \mathrm{d}\xi'' \\ &+ \int_a^{a+\tau}\int_a^{\xi''} 18(\xi' - a) y(\xi')\, \mathrm{d}\xi' \mathrm{d}\xi'' - \int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} 6y(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi''\end{aligned} \tag{14}$$

To abbreviate derivations, we give the end results after triple consecutive integration of the remaining terms in (10).

$$\begin{aligned}&\int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} a_2(\xi - a)^3 y^{(2)}(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi'' \\ &= \int_a^{a+\tau} a_2(\xi'' - a)^3 y(\xi'')\, \mathrm{d}\xi'' \\ &- \int_a^{a+\tau}\int_a^{\xi''} 6a_2(\xi' - a)^2 y(\xi')\, \mathrm{d}\xi' \mathrm{d}\xi'' \\ &+ \int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} 6a_2(\xi - a) y(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi''\end{aligned} \tag{15}$$

The term containing the first derivative yields

$$\begin{aligned}&\int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} a_1(\xi - a)^3 y^{(1)}(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi'' \\ &= \int_a^{a+\tau}\int_a^{\xi''} a_1(\xi' - a)^3 y(\xi')\, \mathrm{d}\xi' \\ &- \int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} 3a_1(\xi - a)^2 y(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi''\end{aligned} \tag{16}$$

Finally, the last term is

$$\int_a^{a+\tau}\int_a^{\xi''}\int_a^{\xi'} a_0(\xi - a)^3 y(\xi)\, \mathrm{d}\xi \mathrm{d}\xi' \mathrm{d}\xi'' \tag{17}$$

Collecting the terms in (12) - (17) yields

$$
\begin{aligned}
-\tau^3 y(a+\tau) = & \int\limits_a^{a+\tau} \Big[-9(\xi''-a)^2 + a_2(\xi''-a)^3\Big] y(\xi'')\,\mathrm{d}\xi'' \\
& + \int\limits_a^{a+\tau}\int\limits_a^{\xi''} \Big[18(\xi'-a) - 6a_2(\xi'-a)^2 \\
& + a_1(\xi'-a)^3\Big] y(\xi')\,\mathrm{d}\xi'\mathrm{d}\xi'' \\
& + \int\limits_a^{a+\tau}\int\limits_a^{\xi''}\int\limits_a^{\xi'} \Big[-6 + 6a_2(\xi-a) \\
& - 3a_1(\xi-a)^2 + a_0(\xi-a)^3\Big] y(\xi)\,\mathrm{d}\xi\mathrm{d}\xi'\mathrm{d}\xi''
\end{aligned} \tag{18}
$$

Applying the Cauchy formula for repeated integration while letting $a + \tau = t$ gives

$$(t-a)^3 y(t) \triangleq \int\limits_a^t K_F(t,\tau) y(\tau)\,\mathrm{d}\tau \tag{19}$$

with $K_F(t,\tau)$ defined as

$$
\begin{aligned}
K_F(t,\tau) \triangleq & \Big[9(\tau-a)^2 - a_2(\tau-a)^3\Big] \\
& + (t-\tau)\Big[-18(\tau-a) + 6a_2(\tau-a)^2 - a_1(\tau-a)^3\Big] \\
& + \frac{(t-\tau)^2}{2}\Big[6 - 6a_2(\tau-a) + 3a_1(\tau-a)^2 - a_0(\tau-a)^3\Big]
\end{aligned} \tag{20}
$$

Equation (11) is processed similarly, with the end result obtained after collecting terms and applying the Cauchy formula while letting $b - \sigma = t$

$$(b-t)^3 y(t) \triangleq \int\limits_t^b K_B(t,\tau) y(\tau)\mathrm{d}\tau \tag{21}$$

with $K_B(t,\tau)$ given by

$$
\begin{aligned}
K_B(t,\tau) \triangleq & \Big[9(b-\tau)^2 + a_2(b-\tau)^3\Big] \\
& + (t-\tau)\Big[18(b-\tau) + 6a_2(b-\tau)^2 + a_1(b-\tau)^3\Big] \\
& + \frac{(t-\tau)^2}{2}\Big[6 + 6a_2(b-\tau) + 3a_1(b-\tau)^2 + a_0(b-\tau)^3\Big]
\end{aligned} \tag{22}
$$

Adding (19) and (21) side by side while dividing both sides by $[(t-a)^3 + (b-t)^3]$ yields

$$y(t) = \int\limits_a^b K_{DS}(t,\tau) y(\tau)\,\mathrm{d}\tau \tag{23}$$

with

$$K_{DS}(t,\tau) \triangleq \frac{1}{[(t-a)^3 + (b-t)^3]}\begin{cases} K_F(t,\tau) & : \tau \le t \\ K_B(t,\tau) & : \tau > t \end{cases} \tag{24}$$

This delivers the formula for the "double-sided" kernel that effectively combines the operations of forward integration on $[a,t]$ and backward integration on $[b,t]$.

The recursive expressions for the derivatives of the output (9) can be derived by proceeding similarly as when deriving the K_{DS}. To obtain the exression for $y^{(1)}$ the equations (10) & (11) need to be integrated two times :

$$
\begin{aligned}
(t-a)^3 y^{(1)}(t) = & \,6(t-a)^2 y(t) - a_2(t-a)^3 y(t) \\
& + \int_a^t \big[-18(\tau-a) + 6a_2(\tau-a)^2 - a_1(\tau-a)^3\big] y(\tau)\,\mathrm{d}\tau \\
& + \int_a^t (t-\tau)\big[6 - 6a_2(\tau-a) + 3a_1(\tau-a)^2 \\
& - a_0(\tau-a)^3\big] y(\tau)\,\mathrm{d}\tau
\end{aligned} \tag{25}
$$

and

$$
\begin{aligned}
(b-t)^3 y^{(1)}(t) = & -6(b-t)^2 y(t) - a_2(b-t)^3 y(t) \\
& + \int_t^b \big[18(b-\tau) + 6a_2(b-\tau)^2 + a_1(b-\tau)^3\big] y(\tau)\,\mathrm{d}\tau \\
& + \int_t^b (t-\tau)\big[6 + 6a_2(b-\tau) + 3a_1(b-\tau)^2 \\
& + a_0(b-\tau)^3\big] y(\tau)\,\mathrm{d}\tau
\end{aligned} \tag{26}
$$

The final expression for $y^{(1)}$ is obtained by adding the results of (25) and (26) while dividing by $[(t-a)^3 + (b-t)^3]$.

To obtain a formula for $y^{(2)}$, equations (10) & (11) need to be integrated only once :

$$
\begin{aligned}
(t-a)^3 y^{(2)}(t) = & \,3(t-a)^2 y^{(1)}(t) - a_2(t-a)^3 y^{(1)}(t) \\
& - 6(t-a)y(t) + 3a_2(t-a)^2 y(t) - a_1(t-a)^3 y(t) \\
& + \int_a^t \big[6 - 6a_2(\tau-a) + 3a_1(\tau-a)^2 - a_0(\tau-a)^3\big] y(\tau)\,\mathrm{d}\tau
\end{aligned} \tag{27}
$$

and

$$
\begin{aligned}
(b-t)^3 y^{(2)}(t) = & -3(b-t)^2 y^{(1)}(t) - a_2(b-t)^3 y^{(1)}(t) \\
& - 6(b-t)y(t) - 3a_2(b-t)^2 y(t) - a_1(b-t)^3 y(t) \\
& + \int_t^b \big[6 + 6a_2(b-\tau) + 3a_1(b-\tau)^2 + a_0(b-\tau)^3\big] y(\tau)\,\mathrm{d}\tau
\end{aligned} \tag{28}
$$

The expression for $y^{(2)}$ is obtained by adding (27) and (28) while dividing by the factor $[(t-a)^3 + (b-t)^3]$.

Fact 2. It follows from the construction that the resulting kernel K_{DS} does not exhibit any singularities at the extremities of the integration interval $[a,b]$ because $\min_t[(t-a)^3 + (b-t)^3] = 0.25(b-a)^3 > 0$. An important consequence of the double sided integration is that the estimates of the derivatives of the output are equally accurate over the entire estimation interval $[a,b]$.

Fact 3. The system representation above immediately extends to the full input-output case (1) as all the algebraic operations applied to the left-hand side of (1) can concurrently be applied to its right-hand side which would result in a system representation with inputs of the type

$$y(t) = \int_a^b K_{DS}(t,\tau) y(\tau)\ \mathrm{d}\tau + \int_a^b K_{DS}^u(t,\tau) u(\tau)\ \mathrm{d}\tau$$

Fact 4. It is also worth noting that the reproducing kernel induces a reproducing kernel Hilbert space (RKHS) uniquely corresponding to the symmetric, positive-type kernel function

$$K(t_1,t_2) \triangleq \langle K_{DS}(t_1,\cdot)|K_{DS}(t_2,\cdot)\rangle_2, \forall t_1,t_2 \in [a,b]$$

where $\langle\cdot|\cdot\rangle_2$ denotes the scalar product in $L^2[a,b]$. The RKHS, here denoted by $\mathcal{H}_K$, is then simply defined as the image of the space $L^2[a,b]$ under the integral transform defined by the double-sided kernel K_{DS} of (8) with a K-weighted norm as defined in Saitoh (1988). The reproducing equality (8) has then yet another useful interpretation - that of a linear subspace of $\mathcal{H}_K$:

$$\mathcal{S}_I \triangleq \{y \in \mathcal{H}_K \mid y \text{ satisfies } \mathcal{I} \equiv 0\ \} \tag{29}$$

More importantly, since the operations leading to the system representation are reversible (by direct differentiation) the homogeneous system $\mathcal{I} \equiv 0$ has an equivalent geometric characterization (29); precisely

Corollary 5. An output function $y : [a,b] \to \mathbb{R}$ satisfies the homogeneous equation $\mathcal{I} \equiv 0$ on the interval $[a,b]$ if an only if it is reproduced by the evaluation functional in (8).

Fact 6. The subspace $\mathcal{S}_I$ is in fact n-dimensional as it is spanned by the fundamental solutions of the invariance equation $\mathcal{I} = 0$ hence guaranteeing the existence of an orthogonal projection onto $\mathcal{S}_I \subset L^2[a,b]$. The parameter estimation procedure then has the obvious geometric interpretation as finding the "orientation" of $\mathcal{S}_I$ that best fits the measurement output data. This line of thought will be pursued in another paper as a way of achieving complete denoisification of a measured output trajectory.

4. SIMULTANEOUS STATE AND PARAMETER ESTIMATION

The derivation of the result in Theorem 1, see equations (12) - (24), clearly shows that the double-sided kernel K_{DS} is in fact linear with respect to the system parameters $a_i, i = 0,\cdots,n-1$, i.e. one can write

$$y(t) = \int_a^b K_{DS}(t,\tau)y(\tau)\ \mathrm{d}\tau \tag{30}$$

$$= \int_a^b \sum_{i=0}^{n} \tilde{a}_i K_{DS(i)}(t,\tau)y(\tau)\ \mathrm{d}\tau \tag{31}$$

where the $K_{DS(i)}; i = 0,\cdots,n$ are "component kernels" of K_{DS} (not to be confused with derivative kernels) with $\tilde{a}_i = a_i; i = 0,\cdots,n-1$, and $\tilde{a}_n = 1$.

Commuting integration with summation then permits to write

$$y(t) - g_n(t,y) = \sum_{i=0}^{n-1} a_i g_i(t,y) \tag{32}$$

with

$$g_i(t,y) := \int_a^b K_{DS(i)}(t,\tau)y(\tau)\ \mathrm{d}\tau \quad i = 0,\dots n \tag{33}$$

Integrating (32) once more gives

$$\int_a^t [y(\sigma) - g_n(\sigma,y)]\ \mathrm{d}\sigma = \sum_{i=0}^{n-1} a_i \int_a^t g_i(\sigma,y)\ \mathrm{d}\sigma\ ; \quad t \in [a,b] \tag{34}$$

Given distinct time instants $t_1,\cdots,t_n \in (a,b]$, the last equation is now re-written point-wise in the form of a matrix equation

$$Q(y) = P(y)a\ ;\ \text{mapping trajectories } y : [0,t] \to \mathbb{R}$$

$$Q \stackrel{def}{=} \begin{bmatrix} q(t_1) \\ \vdots \\ q(t_n) \end{bmatrix}; a \stackrel{def}{=} \begin{bmatrix} \tilde{a}_0 \\ \vdots \\ \tilde{a}_{n-1} \end{bmatrix}; P \stackrel{def}{=} \begin{bmatrix} p_1(t_1)\cdots p_n(t_1) \\ \ddots \\ p_1(t_n)\cdots p_n(t_n) \end{bmatrix}$$

$$q(t_i) = \int_a^{t_i} [y(\sigma) - g_n(\sigma,y)]\ \mathrm{d}\sigma;$$

$$p_k(t_i) = \int_a^{t_i} g_k(\sigma,y)\ \mathrm{d}\sigma \tag{35}$$

As no assumptions are made about the noise which may determine the invertibility of the matrix P, we give the following practical definition of linear identifiability

4.1 Practical Linear Identifiability

Definition 1. The homogeneous system $\mathcal{I} = 0$ is practically linearly identifiable on $[a,b]$ with respect to a particular realization of the output measurement, $y(t), t \in [a,b]$, if and only if there exist distinct time instants $t_1,\cdots,t_n \in (a,b]$ which render the matrix $P(y)$ non-singular. By analogy with the nomenclature used in Fliess and Sira-Ramirez (2003) output trajectories which render $\det P \neq 0$ will be called *persistent*.

Fact 7. Clearly, a zero trajectory ($y \equiv 0$) is not identifiable in any terms, but nominal non-zero (noise free) system trajectories are. This fact is easy to see by considering a system output in the form of a generic fundamental solution spanning (29).

An alternative way to set up the estimation equations is to integrate equation (32) consecutively $n-1$ times generating $n-1$ integral equations by application of the Cauchy formula for multiple iterated integrals. These equations would clearly appear in the form

$$\frac{1}{(k-1)!}\int_a^t (t-\sigma)^{k-1} y(\sigma)\ \mathrm{d}\sigma$$
$$= \sum_{i=0}^{n} a_i \frac{1}{(k-1)!}\int_a^t (t-\sigma)^{k-1} g_i(\sigma,y)\ \mathrm{d}\sigma; k = 1,\cdots n-1 \tag{36}$$

Repeated iterated integration has the advantage that the nodal time instants t_i are not needed and the alternative estimation equation is $Q'(y) = P'a$ in which the k-th row is directly corresponding to the k-th iterated integral of (36). Additionally it delivers additional smoothing by the low pass nature of integration. As before, linear identifiability is guaranteed for non-zero nominal output trajectories.

4.2 Practical Parameter Estimation Procedures

In practical applications the n distinct time instants needed to generate (35) can be placed equidistantly over the interval $(a,b]$ or else generated randomly. Since no assumptions are made about system perturbations or measurement noise, the estimation equation (35) is best solved in terms of a pseudo-inverse $P^\dagger$ of P:

$$a = P^\dagger(y)Q(y) \tag{37}$$

while simultaneously introducing an efficient assessment criterion of the accuracy of estimation eg. in terms of a redundant "sentinel" parameter as described next.

4.3 Invariance Based Estimation Improvements

It is rather remarkable how differential system invariants can be employed to achieve multiple improvements in algebraic estimation procedures. It is immediate to see that the basic system invariant is conserved if the invariant equation is transformed in such a way as to preserve its validity. This point is best illustrated in the Laplace domain where the original homogeneous system is written as

$$(s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0)y(s) = 0 \qquad (38)$$

and is multiplied by a real linear factor in s

$$[(s^m + b_{m-1}s^{m-1} + \cdots + 1)\cdot \\ \cdot (s^n + a_{n-1}s^{n-1} + \cdots + a_0)]y(s) = 0 \qquad (39)$$

which preserves the validity of $\mathcal{I} \equiv 0$ in the time domain (the output of the linear pre-filter is identically zero only for identically zero inputs). This can be used to develop a criterion to assess estimation accuracy in noisy environments. For $m = 1$, we write in place of (39)

$$[(s+1)s^n + (s+1)a_{n-1}s^{n-1} + \cdots + (s+1)a_0 + (s+1) \\ - (s + b_x)]y(s) = 0 \qquad (40)$$

where b_x is called the *sentinel* parameter and is estimated alongside with $a_0, \cdots, a_{n-1}$ after representing (40) in the time domain using Theorem 1. Clearly the estimated value of b_x is an indication of the accuracy of estimation as $b_x = 1$ is the only guarantee of the validity of $\mathcal{I} \equiv 0$.

Other pre-filtering factors can be employed in (39) that were shown to annihilate given structured output perturbations; see Fliess and Sira-Ramirez (2003). Optimization-based selection of pre-filter parameters $b_i, i = 1, \cdots, m-1$ are possible to further enhance denoisification of the measured output; see Mboup et al. (2009) and Hu and Mao (2014) .

Fact 8. Finally, it is worth noting that parameter estimation can be conducted simultaneously with state estimation. Under the assumption of system flatness, the system states are immediately recovered as functions of the time derivatives of the output. Following parametric estimation the output derivatives are easily computed using the recursive kernels in (9).

5. EXAMPLE SYSTEM IDENTIFICATION

The following third order LTI system was considered:

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -10 & 0 \end{bmatrix} x \; ; y = x_1 \; ; \; x(0) = [1, 1, 0] \qquad (41)$$

with the corresponding characteristic equation

$$y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)} + a_0 y(t) = 0 \qquad (42)$$

where the "nominal" values for the parameters are given in Table 1 (True values). In all the subsequent experiments the system parameters a_0, a_1, a_2 were considered unknown. In the first experiment, output "measurement noise" was introduced by simple sampling with frequency $10^3/sec$. The parameters a_0, a_1, a_2 were next estimated as in section $4. The sampled output was subsequently fed into the integral transforms of Theorem 1 using the estimated parameter values to produce the estimated output and its derivatives $y, \dot{y}, \ddot{y}$. Romberg integration was used to evaluate the integral transforms. Romberg integration employs repeated Richardson's extrapolation, effectively reducing the numerical error to near machine precision. Figures 1-3 show the estimated signals obtained by kernel filtering as they compare against their "true" counterparts (plotted 10 times more densely than the sampled signal used for estimation). It is seen that accuracy is uniform over the entire estimation window.

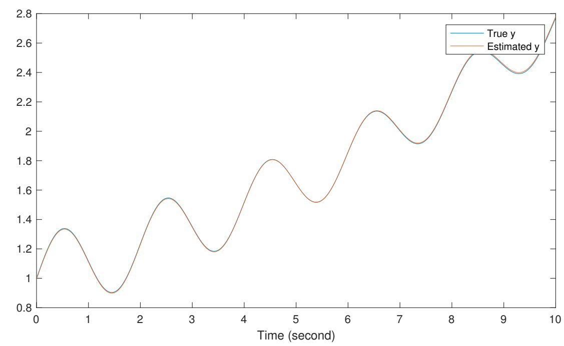

Fig. 1. Estimated signal y (Sampled case)

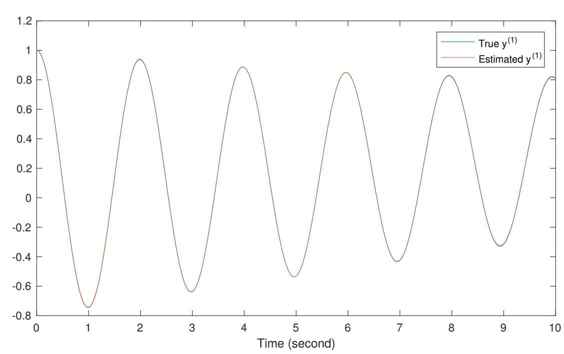

Fig. 2. Estimated first derivative of y (Sampled case)

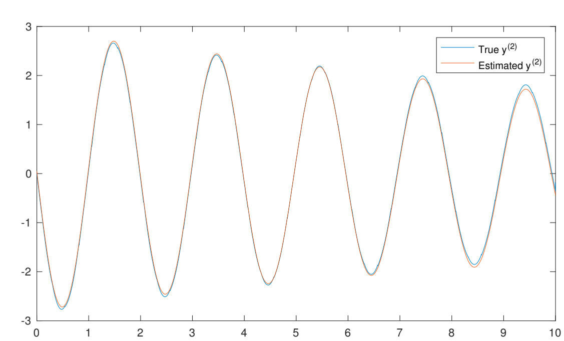

Fig. 3. Estimated second derivative of y (Sampled case)

The nominal system output y was then perturbed additively by noise of different characteristics prior to parameter estimation as in section $4. The following noise characteristics were considered:

Case I: Gaussian noise; 30 SNR

Case II: Gaussian noise; 20 SNR

Case III: Quantization with additive noise of 30 SNR

In Case III the nominal system output was quantized using Lloyds algorithm before adding Gaussian noise of 30 SNR.

The three noisy outputs are shown in Figures 4 - 6.

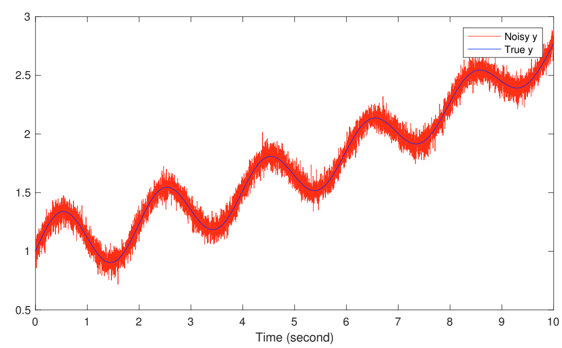

Fig. 4. Case I: Gaussian noise; 30 SNR

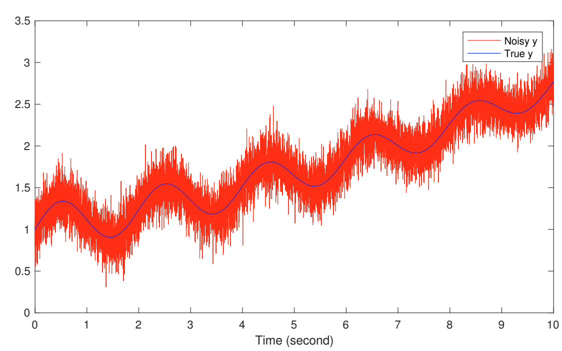

Fig. 5. Case II: Gaussian noise; 20 SNR

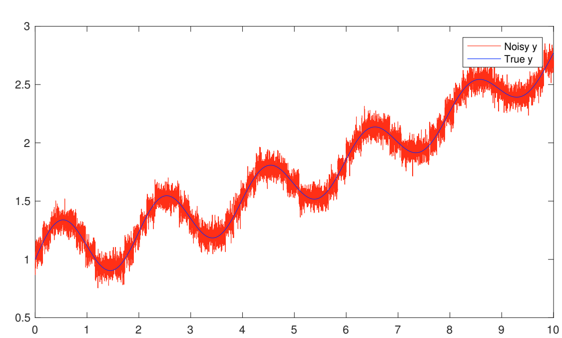

Fig. 6. Case III: Quantization with additional noise of 30 SNR

The estimated system parameter values a_0, a_1, a_2 for the sampled case as well as the three noisy cases are listed in Table 1.

	a_0	a_1	a_2
True values	-1	10	0
Estimated values (Sampled Case)	-1.0002	10.0021	-0.0002
Estimated values (Case I)	-0.9859	9.8691	0.0153
Estimated values (Case II)	-0.9785	9.7349	0.0510
Estimated values (Case III)	-1.0077	10.1244	0.0369

Table 1. Estimated parameter values for different measurement noises

6. CONCLUSIONS

The kernel transforms introduced deliver high accuracy in output derivative estimation which is seen to be uniform over estimation windows of arbitrary length. The double-sided kernel also performs well when employed as part of parameter estimation procedures. Adaptive window length should be used for best performance if the measured output variations are large, e.g. when the underlying system is highly unstable. Simultaneous denoisification and estimation is possible, and will be demonstrated in another paper.

REFERENCES

Fliess, M. (2006). Analyse non standard du bruit. *Comptes Rendus Mathematique*, 342(10), 797–802.

Fliess, M. and Sira-Ramirez, H. (2003). An algebraic framework for linear identification. *ESAIM: Control, Optimisation and Calculus of Variations*, 9, 151–168.

Fliess, M. and Sira-Ramirez, H. (2008). Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques. In *Identification of Continuous-time Models from sampled Data*, 363–391. Springer.

Gevers, M. (2006). A personal view of the development of system identification: A 30-year journey through an exciting field. *IEEE Control Systems*, 26(6), 93–105.

Ghoshal, D.P., Gopalakrishnan, K., and Michalska, H. (2017). Using invariance to extract signal from noise. In *American Control Conference*.

Hu, W. and Mao, J. (2014). Improved algebraic method for linear continuous-time model identification. *Journal of Control and Decision*, 1(2), 180–189.

Mboup, M., Join, C., and Fliess, M. (2009). Numerical differentiation with annihilators in noisy environment. *Numerical algorithms*, 50(4), 439–467.

Saitoh, S. (1988). *Theory of reproducing kernels and its applications*, volume 189. Longman.

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