2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Variational Approach to Joint Linear Model and State Estimation*

Abhishek Pandey[1], Debarshi Patanjali Ghoshal[2] and Hannah Michalska

Abstract— **An approach similar to variational data assimilation is presented for simultaneous model and state estimation of linear systems from output measurement perturbed by large noise of unknown characteristic. Estimation can be carried out over time windows of arbitrary length and is based on a special kernel representation of a hypothetical linear model. A functional criterion is employed to assess the distance of a model to a data cloud. The reproducing kernel representation of the model is particularly advantageous as it permits a Hilbert subspace representation of the model sought. The approach is similar to variational data assimilation as it relies on the adjoint method for calculation of the gradient of the cost functional involved.**

I. Introduction

Model estimation, adaptive filtering, and joint parameter and state estimation are topics of paramount importance to applications in almost all disciplines of science. The literature is overwhelmingly vast and diverse, ranging from general estimation methods, [1], [2], [3], through recursive stochastic filtering approaches [4], [5] to system modeling via data assimilation [6], [7]. Recursive stochastic approaches prevail due to their elegance, simplicity, and capability of efficient noise attenuation. Most filtering techniques, maybe with the exception of the Wiener filters [8], assume at least partial information about the initial state of the system and measurement noise characteristics. Convergence of such classical methods is conditioned by the validity of the assumptions made.

New trends in system estimation and modelling are rapidly developing and include: algebraic dead-beat observers based on the concept of differential flatness, algebraic parameter estimation [9], [10], [11], and invariant observers [12], [13]. The algebraic dead-beat methods are increasingly important as many applications, such as target tracking, call for finite time reliable estimation.

In this context, the present paper endeavours to extend previous work on algebraic approaches to state and parameter estimation of linear systems [14], [15], [16], [17]. The state estimation approach presented in [14] introduces a kernel system representation which allows to calculate higher output derivatives despite considerable levels of output measurement noise. The algorithm has shown to enjoy strong noise attenuation properties so the output derivatives can be used to reconstruct state estimates over arbitrary short observation intervals. In [15] a kernel representation of linear systems is used for the purpose of parameter estimation. The state estimation algorithm in [14] is further generalized to apply to linear time-varying systems in [16]. Applications include a target tracking algorithm developed in [17]. The kernel representation of linear systems is based on the concept of differential invariance which roots from the validity of the Cayley-Hamilton theorem. Differential invariants are particularly useful in state and parameter estimation as they embody the "deterministic signature" of a system that holds for all times during its evolution notwithstanding any measurement noise.

A more radical, non-asymptotic approach to simultaneous model and state estimation is adopted in this paper. Specifically, given a cloud of measurement points presumed to represent a dynamical system output trajectory, a linear model is constructed to fit it best. The approach is somewhat similar to variational data assimilation in that it employs a cost functional as a measure of model fitness and computes its gradient by solution of an adjoint sensitivity problem. In distinction to the methods used in [14], [15], a kernel system model to be constructed is viewed as a linear finite dimensional subspace of a reproducing Hilbert space. The subspace is linearly parametrized by the unknown system constants whose values determine the subspace "orientation" vis á vis the cloud of measurement points. The system constants are found simultaneously with reconstructing the system state. A geometrical interpretation of this process is particularly appealing as it can be visualized as subspace re-orientation by way of minimizing the residual of the modelling error (i.e. the cost functional in this case). The idea of rotation of subspaces for the purpose of parameter estimation has been used before, see [18], albeit in a completely different context.

The method does not employ numerical differentiation techniques. The advantages of the described approach are transparent. The joint model and state estimation is non-asymptotic, does not require any information about boundary conditions of the presumed dynamical system nor the measurement noise characteristic. Additionally, as the optimization process searches for the best estimates of the system parameters the "empirical" statistical distribution of the modelling residual error can be tested for congruence with a priori knowledge about the measurement noise, if such knowledge is available.

Examples confirm high accuracy of estimation that can be carried out over time windows of arbitrary length. The estimation process is, however, computationally expensive as simultaneous denoisification and parameter estimation employs optimization by iterative search.

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

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

[2]Corresponding author debarshi.ghoshal@mail.mcgill.ca

The nonparametric Kolmogorov probability density test is used to assess congruence of the modelling residual errors with the (in this case assumed) statistical density of the measurement noise.

II. Model and State Estimation

A. Joint Estimation Problem

Given a data cloud of time-tagged measurements, a linear model can be sought to best explain the local output behaviour of a presumed dynamical system. For simplicity of exposition, consider single output, zero-input, LTI systems as candidates of such dynamical system models. The order of the model being fitted is not of critical importance as it can also be determined during the estimation process; see [17].

The goal of simultaneous model parameter and state estimation is then stated as follows:

Linear Simultaneous State and Parameter Estimation
Assuming an LTI SISO system model :$\dot{x} = Ax; y = c^T x$ of order n, identify the values of the parameters $a_i, i = 0, .., n-1$, in its characteristic equation

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{1}$$

from a noisy system output measurement $\tilde{y}(t)$ over a finite, but arbitrary, interval of time $t \in [0, T]$, $T > 0$. No assumption is made about the noise characteristic. Also, provide the estimates of the noise-free output $y(t)$ and all output derivatives : $y^{(i)}(t), i = 1, \cdots, n-1$, for $t \in [0, T]$.

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. The left hand side of (1) is clearly a differential invariant for the system as it results from the validity of the Cayley-Hamilton's theorem, a quantity that remains invariant under the action of the flow of the system.

B. A Kernel Model Representation

The following kernel representation of the system model of Section II-A was introduced and derived in [15]:

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

$$y(t) = \int_a^b K_{DS}(t, \tau) y(\tau) \, \mathrm{d}\tau \ ; \ \forall t \in [a, b] \tag{2}$$

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 $t \in [a, b]$:

$$y^{(i)}(t) = \sum_{k=0}^{i-1} b_k(t) y^{(k)}(t) + \int_a^b K_{DS}^i(t, \tau) y(\tau) \, \mathrm{d}\tau \tag{3}$$

where $y^{(0)} \equiv y$ and $b_k(\cdot)$ are rational functions of t. Hilbert-Schmidt kernels are square integrable functions on $L^2[a, b] \times L^2[a, b]$.
The kernel in (2) induces a reproducing kernel Hilbert space (RKHS) uniquely corresponding to the symmetric, positive-type kernel function
$K(t_1, t_2) \triangleq \langle K_{DS}(t_1, \cdot) | K_{DS}(t_2, \cdot) \rangle_2$
for all $t_1, t_2 \in [a, b]$ 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 (2) with a K-weighted norm as defined in [19]. The reproducing equality (2) has then yet another useful interpretation - that of a linear subspace of $\mathcal{H}_K$:

$$\mathcal{S}_I \triangleq \{ y \in \mathcal{H}_K \mid y \text{ satisfies (2) } \} \tag{4}$$

More importantly, by construction of the kernel, the behavioural system model of Theorem 1 is equivalent to the differential model as described by the invariance equation (1). This fact is stated as follows:

Corollary 1: An output function $y : [a, b] \to \mathbb{R}$ satisfies the invariance equation (1) on the interval $[a, b]$ if an only if it is reproduced by the evaluation functional in (2).
The proof of this fact is redundant as the multiple iterated integration in the derivation of the reproducing kernel can be reversed by multiple differentiation to retrieve the original invariance equation; see [14]. Clearly, at this point, the initial conditions of the original system play no role as the behaviour of the system is fully characterized in terms of its "trajectory behaviour" over $[a, b]$. The filtering problem for the system output and its derivatives of any order then amounts to a "output trajectory reconstruction" that best fits the noisy measurement while preserving the system invariant.

C. Parameter Estimation and Practical Identifiability

Model parameter estimation, as described briefly below; see also [15], is obviously a component of the joint estimation approach as proposed here.

The proof of Theorem 1 ; see [15], shows that the 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

$$y(t) = \int_a^b K_{DS}(t, \tau) y(\tau) \, \mathrm{d}\tau = \sum_{i=1}^{n} \tilde{a}_i g_i(t, y) \tag{5}$$

$$\text{where } g_i(t, y) \stackrel{def}{=} \int_a^b K_{DS}^{(i)}(t, \tau) y(\tau) \, \mathrm{d}\tau; \ t \in [a, b] \tag{6}$$

where the $K_{DS}^{(i)}; i = 1, n$ are "component kernels" of K_{DS} and $\tilde{a}_i = a_{i-1}$ for notational convenience; see also the explicit expressions for model kernels used in the Example section of this paper.

Given distinct time instants $t_1, \cdots, t_m \in (a, b]$, $m \geq n$, equation (5) is now re-written point-wise in the form of a linear algebraic system of equations

$$q(y) = P(y) a \ ; \ \text{mapping trajectories } y : [0, t] \to \mathbb{R} \tag{7}$$

$$q(y) \stackrel{def}{=} \begin{bmatrix} y(t_1) \\ \vdots \\ y(t_m) \end{bmatrix}; a \stackrel{def}{=} \begin{bmatrix} \tilde{a}_1 \\ \vdots \\ \tilde{a}_n \end{bmatrix};$$

$$P(y) \stackrel{def}{=} \begin{bmatrix} g_1(t_1, y) \cdots g_n(t_1, y) \\ \ddots \\ g_1(t_m, y) \cdots g_n(t_m, y) \end{bmatrix} \tag{8}$$

As no assumptions are made about the noise which may determine the existence of solutions to the linear system (7), we give the following practical definition of linear identifiability.

D. Practical Linear Identifiability

Definition 1. The homogeneous system (1) 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 $t_1, \cdots, t_m \in (a, b]$ such that rank$P(y) = n$. By analogy with the nomenclature used in [9] output trajectories which render rank$P(y) = n$ will be called *persistent*.

In practical applications the m distinct time instants needed to generate (8) 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 (8) is best solved in terms of a pseudo-inverse $P^\dagger$ of P:

$$a = P^\dagger(y)q(y) \tag{9}$$

Finally, note that the output reproducing property, as written in (5), implies that the model representing subspace (4) has explicit linear expression in terms of the system model parameters. This fact is important in futher development of the joint estimation approach.

III. Simultaneous Denoisification and Model Shaping

With essentially no information about the system model nor measurement noise the task of output differentiation requires a special treatment.

While the first step of the approach will use raw data to deliver the initial estimate of the system parameters as outlined in II-C, the latter estimate must be further refined during an exhaustive search for an optimal orientation of the model representing subspace $\mathcal{H}$ of (4) to best fit the data cloud. Such a refinement process will be referred to as "simultaneous denoisification and model shaping". To pose this problem in a mathematical form, it is convenient to re-write (1) in state space form:

$$\dot{x} = A_K x \; ; \; y = Cx; \quad x \in \mathbb{R}^n \tag{10}$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \stackrel{def}{=} \begin{bmatrix} y \\ \vdots \\ y^{(n-1)} \end{bmatrix}; \; A_K \stackrel{def}{=} \begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ & & & \\ 0 & 0 & 0 & \ddots \\ -a_0 & -a_1 & \cdots & -a_{n-1} \end{bmatrix}$$

$$K \stackrel{def}{=} \begin{bmatrix} k_1 & k_2 & \cdots & k_n \end{bmatrix} \stackrel{def}{=} \begin{bmatrix} -a_0 & -a_1 & \cdots & -a_{n-1} \end{bmatrix}$$

$$C \stackrel{def}{=} \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix};$$

with an unknown initial condition $x(0) = x_0$. The simultaneous denoisification and parameter estimation problem is now cast as an optimization problem

$$\min\{Q(K, x_0) \mid \dot{x} = A_K x \; ; x(0) = x_0; y = Cx\} \tag{11}$$

$$Q(K, x_0) \stackrel{def}{=} \frac{1}{2} \int_0^T (y - \tilde{y})^2 dt \tag{12}$$

where the cost Q represents the model residual error. As the initial condition is unknown the problem is equivalently reformulated in terms of fundamental solutions of (10) as follows. Let, $x_i^e, i = 1, \cdots, n$ denote the solutions of the system equation $\dot{x} = A_K x$ corresponding to the initial conditions $x(0) = e_i$, respectively, where e_i are unit canonical basis vectors in $\mathbb{R}^n$. All solutions of $\dot{x} = A_K x$ then form an n-dimensional subspace $\mathcal{S}_x \stackrel{def}{=} \text{span}\{x_i^e; i = 1, \cdots, n\}$ i.e. any solution of the system equation (10) is a linear combination

$$x = \sum_{i=1}^n c_i x_i^e; \; c \stackrel{def}{=} \begin{bmatrix} c_1, \cdots, c_n \end{bmatrix} \tag{13}$$

and (11) - (12) takes a form of a finite dimensional optimization problem with equality constraints which is, however, free of any initial conditions

$$\min\{J(K, c) \mid \dot{x}_i^e = A_K x_i^e; \; i = 1, \cdots, n\} \tag{14}$$

$$J(K, c) \stackrel{def}{=} \frac{1}{2} \int_0^T (C \sum_{i=1}^n c_i x_i^e - \tilde{y})^2 dt \tag{15}$$

$$J : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

It is now easy to see that *for any fixed value* of the vector K the minimum of J is attained for a vector $\hat{c} \in \mathbb{R}^n$ which renders the orthogonal projection

$$\hat{y} = \sum_{i=1}^n \hat{c}_i y_i^e; \quad y_i^e \stackrel{def}{=} C x_i^e \tag{16}$$

of the output measurement function $\tilde{y}(\cdot)$ onto the n-dimensional subspace

$$\mathcal{S}_y \stackrel{def}{=} \text{span}\{y_i^e; i = 1, \cdots, n\} \subset L^2[0, T]; \tag{17}$$

so that

$$(\hat{y} - \tilde{y}) \perp \mathcal{S}_y \tag{18}$$

$$\text{i.e.} \quad (\sum_{i=1}^n \hat{c}_i y_i^e - \tilde{y}) \perp y_k^e \quad \text{for all } k = 1, \cdots, n \tag{19}$$

$$\Longrightarrow \; \hat{c} = G(y_i^e)^{-1} h(\tilde{y}, y_i^e); \tag{20}$$

$$G(y_i^e) \stackrel{def}{=} \begin{bmatrix} (y_1^e|y_1^e) & \cdots & (y_n^e|y_1^e) \\ (y_1^e|y_2^e) & \cdots & \cdots \\ \cdots & \ddots & \cdots \\ (y_1^e|y_n^e) & \cdots & (y_n^e|y_n^e) \end{bmatrix}; h(\tilde{y}, y_i^e) \stackrel{def}{=} \begin{bmatrix} (\tilde{y}|y_1^e) \\ \vdots \\ (\tilde{y}|y_n^e) \end{bmatrix}$$

where the Gram matrix $G(y_i^e)$ is invertible as the fundamental set spanning $\mathcal{S}_y$ is linearly independent.

Remark 1: Note that if the system parameters in vector K are known then equations (16) and (20) deliver a denoisified measurement corresponding to $\tilde{y}$ in the form of $y(t) = \hat{y}(t), t \in [0, T]$.

Importantly, the optimization problem (14) - (15) has infinitely many solutions if the noisy measurement is identically zero (i.e. when $\tilde{y} \equiv 0$). This is because then the orthogonal projection (16) is also identically zero, i.e. $\hat{y} \equiv 0$. Consequently, $\hat{y}^{(i)} \equiv 0$ for $i = 1, \cdots, n-1$, so the system homogeneous equation (1) is satisfied for *any* set of parameters $a_i; i = 0, \cdots, n-1$. The system (1) is then *non-identifiable*. As such a situation is unlikely to occur the minima of (14) - (15) can be sought by iterative search direction optimization methods. Computation of such search directions requires evaluation of the gradient of the cost index J at any given parameter pair (K, c). This is a complex task as the computation of $\frac{\partial J}{\partial K}(\bar{K}, \bar{c})$ at any given point $(\bar{K}, \bar{c})$ requires calculation of the sensitivities of the fundamental solutions of $\dot{x}_i^e = A_{\bar{K}} x_i^e$ with respect to variation of the parameters in $\bar{K}$. The solutions x_i^e depend on the parameters K nonlinearly and a direct *forward* calculation of sensitivities (via the solution of the respective variational equations; [20]) is not helpful especially that the sensitivities enter not only the cost index J but also the constraints. The adjoint method is used instead as explained next.

A. Necessary Conditions of Optimality and Gradient Computation

The adjoint gradient method for dynamical systems was provided rigorously in [21] and less formally in [22]. It will be explained here with strict reference to our particular problem.

The gradient of the cost index in the optimization problem (14) - (15) is efficiently computed in a dual approach by incorporating a Lagrangian function:

$$\mathcal{L}(K, c) \stackrel{def}{=} \frac{1}{2} \int_0^T (C \sum_{i=1}^n c_i x_i^e - \tilde{y})^2 dt + \int_0^T \sum_{i=1}^n c_i \lambda_i^T (\dot{x}_i^e - A_K x_i^e) dt \tag{21}$$

where the Lagrangian multipliers are *linearly independent vector functions* $\lambda^T : [0, T] \to \mathbb{R}^n$. The Lagrange multipliers constrain the dynamical system to variations around the path of $\dot{x}_i^e = A_K x_i^e$. Clearly, once the constraint equation $\dot{x}_i^e = A_K x_i^e$ is satisfied the Lagrange multiplier term will disappear leaving

$$\frac{\partial \mathcal{L}(K, c)}{\partial k_j} = \frac{\partial J(K, c)}{\partial k_j}; \; j = 1, \cdots, n \tag{22}$$

Therefore, in the dual representation of the constrained problem the contribution of the derivative $\frac{dx_i^e}{dK}$ to the gradient $\frac{d\mathcal{L}}{dK}$ can be "annihilated" by choosing $\lambda^T(\cdot)$ in a specific way (in effect circumventing the need for explicit forward calculation of $\frac{dx_i^e}{dK}$). Such adjoint procedure is explained as follows. Calculating the derivatives with respect to the components of K yields

$$\frac{\partial \mathcal{L}(K, c)}{\partial k_j} = \int_0^T (C \sum_{l=1}^n c_l x_l^e - \tilde{y})^T C \sum_{i=1}^n c_i \frac{dx_i^e}{dk_j} dt + \int_0^T \sum_{i=1}^n c_i \lambda_i^T (\frac{d\dot{x}_i^e}{dk_j} - A_K \frac{dx_i^e}{dk_j} - \frac{dA_K}{dk_j} x_i^e) dt \tag{23}$$

for all $k_j, j = 1, \cdots, n$, where explicit evaluation gives the sensitivity of the matrix A_K with respect to parameter k_j

$$\frac{dA_K}{dk_j} = I_K; \; I_K \stackrel{def}{=} \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{24}$$

where $I_K \in \mathbb{R}^n \times \mathbb{R}^n$ is a matrix with a single non-zero element in the last row and k-th column : $I_K^{(k,n)} = 1$. Using integration by parts allows to write

$$\int_0^T \lambda_i^T \frac{d\dot{x}_i^e}{dk_j} dt = \lambda_i^T \frac{dx_i^e}{dk_j}\Big|_0^T - \int_0^T \frac{d\lambda_i^T}{dt} \frac{dx_i^e}{dk_j} dt \tag{25}$$

Since the Lagrangian multiplier functions have no effect on the value of the Lagrangian when the constraint is satisfied; see (22), and the initial conditions needed to compute the fundamental solutions $x_i^e; i = 1, \cdots n$ do not depend on the parameters k_j then we can set, without the loss of generality

$$\lambda_i^T(T) = 0; \quad \frac{dx_i^e}{dk_j}(0) = 0; \quad i, j = 1, \cdots, n \tag{26}$$

for an immediate simplification

$$\int_0^T \lambda_i^T \frac{d\dot{x}_i^e}{dk_j} dt = -\int_0^T \frac{d\lambda_i^T}{dt} \frac{dx_i^e}{dk_j} dt \tag{27}$$

Substituting (24) and (27) into (23) yields

$$\frac{\partial \mathcal{L}(K, c)}{\partial k_j} = \int_0^T \sum_{i=1}^n c_i \left[(C \sum_{l=1}^n c_l x_l^e - \tilde{y})^T C \frac{dx_i^e}{dk_j} \right. \tag{28}$$

$$\left. - \dot{\lambda}_i^T \frac{dx_i^e}{dk_j} - \lambda_i^T A_K \frac{dx_i^e}{dk_j} \right] dt - \int_0^T \sum_{i=1}^n c_i \lambda_i^T I_j x_i^e dt \tag{29}$$

It is now time to select a "trajectory" for the Lagrange multipliers that best serves the goal of removing excessive computational burden by imposing that

$$\begin{aligned} \dot{\lambda}_i^T &= -\lambda_i^T A_K + (C \sum_{l=1}^n c_l x_l^e - \tilde{y})^T C \\ \lambda_i^T(T) &= 0; \quad i = 1, \cdots, n \end{aligned} \tag{30}$$

The remaining term of (29) then constitutes the essential part of the gradient with respect to the parameters in K:

$$\frac{\partial \mathcal{L}(K, c)}{\partial k_j} = -\int_0^T \sum_{i=1}^n c_i \lambda_i^T I_j x_i^e dt; j = 1, \cdots, n \tag{31}$$

It is also easy to see that the gradient of J with respect to the parameters c_i that replace the influence of the initial conditions is given by

$$\frac{\partial J(K, c)}{\partial c_i} = \int_0^T (C \sum_{l=1}^n c_l x_l^e - \tilde{y})^T C x_i^e dt; \; i = 1, \cdots, n \tag{32}$$

It is now possible to state:

Theorem 2: The necessary conditions for optimality in the simultaneous state and parameter estimation optimization problem (14) - (15) are given by

$$\frac{\partial J(K,c)}{\partial k_j} = -\int_0^T \sum_{i=1}^{n} c_i \lambda_i^T I_j x_i^e dt = 0; \tag{33}$$

$$\frac{\partial J(K,c)}{\partial c_i} = \int_0^T (C\sum_{l=1}^{n} c_l x_l^e - \tilde{y})^T C x_i^e dt = 0; \tag{34}$$

$$\dot{\lambda}_i^T = -\lambda_i^T A_K + (C\sum_{l=1}^{n} c_l x_l^e - \tilde{y})^T C; \tag{35}$$

$$\dot{x}_i^e = A_K x_i^e; \tag{36}$$

$$\lambda_i^T(T) = 0; \quad x_i^e(0) = e_i; \tag{37}$$

$$\text{for all} \quad i = 1, \cdots, n; \quad j = 1, \cdots, n$$

The above conditions can readily be used in a gradient-search direction based iterative optimization procedure which has been employed here. Once satisfactory precision is achieved during the iterative search for the minimum in (14) - (15) the algorithm is exited. The latest values of the iterated parameters $\hat{c}$ and $\hat{K}$ are then used in the next stage of the estimation process in which flatness-based estimates of the system states requires prior estimation of all the output derivatives. Their calculation is described next.

B. Estimation of Output Derivatives and System States

Estimation of the output $y(t)$ is immediately obtained in terms of the optimized values $\hat{c}$, $\hat{K}$, as found at the exit from the optimization algorithm of the previous section ,i.e.

$$\hat{y}(t) = \sum_{i-1}^{n} \hat{c}_i \hat{x}_i^e(t); \;\; t \in [0, T] \tag{38}$$

where $\hat{x}_i^e; i = 1, \cdot, n$ are is the fundamental set of solutions corresponding to the optimal parameter $\hat{K}$. The estimate of the state of the system is then calculated in terms of all the derivatives of the output $y^{(1)}, \cdots y^{(n-1)}$ explicitly using the integral transform formulae as delivered by Theorem 1 of II-B; see (3) .

The advantage of using integral transforms to compute derivatives in place of conventional numerical differentiation methods should be clear as such an approach provides additional robustness with respect to computational noise.

IV. Example

A third order LTI system was considered as follows:

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -10 & 0 \end{bmatrix} x \;; y = x_1 \;; \; x(0) = [1, 1, 0] \tag{39}$$

with its corresponding characteristic equation

$$y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)} + a_0 y(t) = 0 \tag{40}$$

All the parameters a_0, a_1, a_2 were assumed to be unknown, but the "nominal" reference values used for comparison are however given in Table 1 as True values.

The system output was quantized and further perturbed by white noise with SNR=30dB to represent "raw measurement data" which was used to compute initial estimates of the parameters a_0, a_1, a_2 as in section II-C . The result is shown in Table I.

	a_0	a_1	a_2
True values	-1	10	0
Initial estimates	-1.0601	10.5965	0.0541

TABLE I

Initial estimates for parameter values (prior to denoisification and model shaping).

The estimates of the output signal y and its time-derivatives $y^{(1)}$, $y^{(2)}$ for any given values of parameter estimates are derived as in section III-B , while using the kernels corresponding to the example system (39) via Theorem 1 (for derivation see [15]); the expressions for the kernels are omitted here for reason of brevity.

Comparing the estimated parameter values to their true counterparts could be very misleading (the absolute differences seem small as seen in Table I). However, when these estimates were subsequently used in the kernel transforms of Theorem 1 they failed to reproduce the system output correctly as proved by Figure 1. The need for further refinement of the initial estimates was then transparent.

Simultaneous denoisification and model shaping was subsequently applied Figures 2, 3 and 4 show the y, $y^{(1)}$, $y^{(2)}$ estimates after convergence was achieved in denoisification and model shaping. These estimates are the same in both cases: with and without prior parameter estimation from raw data.

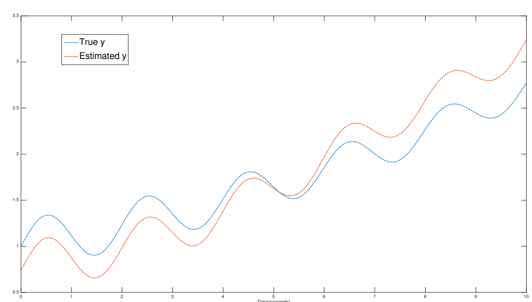

Fig. 1. Comparison of the true and estimated outputs prior to denoisification and model shaping.

A. Kolmogorov Smirnov test

The Kolmogorov Smirnov test was used to compare the two distributions of empirical samples from $\tilde{y} - y$ and $\tilde{y} - \hat{y}$, where $\tilde{y}$ is the noise signal, y is the true system trajectory, and $\hat{y}$ is the estimated trajectory. The KS test indicated a good fit of the residuals, confirming successful model shaping at a 5% significance level.

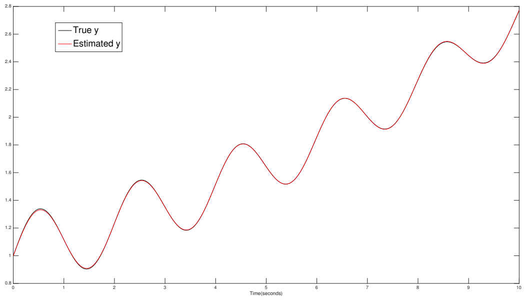

Fig. 2. Estimated system output y after convergence is achieved.

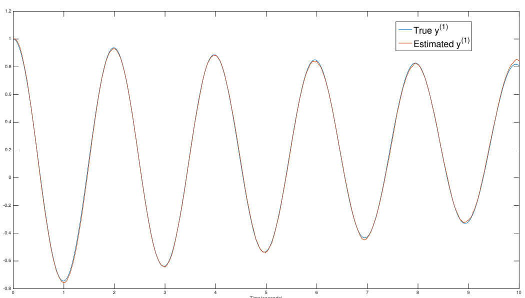

Fig. 3. Estimated first derivative of the output $y^{(1)}$ after convergence is achieved.

V. Conclusions

The paper addresses the problem of simultaneous state and parameter estimation in linear systems from the measurement of its output over a finite interval of time. Additionally, the output measurement is subject to noise of an unknown characteristics. The state is reconstructed as a function of the derivatives of the system output which are obtained by way of integral transforms. It is somewhat surprising that such a problem can be solved successfully as it is widely acknowledged that high order, high accuracy differentiation of noisy signals is impossible.

As noise is habitually represented by stochastic processes that also posess kernel representations, it is not hard to see the importance of the above approach. The two kernel representations: that of the hypothetical model and that of the stochastic process noise give rise to two objective functions: the value of the model residual error and the distance measure between two statistical distributions, that of the presumed noise and that of the observed residual. The need of a trade-off between model fitness to data and fitness of residual to noise statistics is apparent. Such and other probabilistic approaches to the problem will be explored elsewhere.

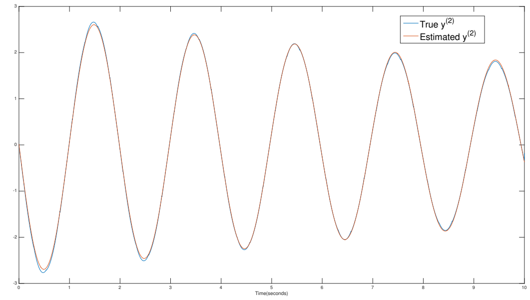

Fig. 4. Estimated second derivative of the output $y^{(2)}$ after convergence is achieved.

References

[1] L. Ljung, "System identification: Theory for the user, ptr prentice hall information and system sciences series," *ed: Prentice Hall, New Jersey*, 1999.

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

[3] G. Goodwin and K. Sin, "Adaptive filtering prediction and control. 1984," *Englewood Clifs: Prentice Ha II*.

[4] R. E. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[5] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.

[6] E. Kalnay, *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.

[7] J. L. Anderson, "An ensemble adjustment kalman filter for data assimilation," *Monthly weather review*, vol. 129, no. 12, pp. 2884–2903, 2001.

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

[9] M. Fliess and H. Sira-Ramírez, "An algebraic framework for linear identification," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 9, pp. 151–168, 2003.

[10] M. Fliess and H. Sira-Ramirez, "Closed-loop parametric identification for continuous-time linear systems via new algebraic techniques," in *Identification of Continuous-time Models from sampled Data*. Springer, 2008, pp. 363–391.

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

[12] S. Bhattacharyya, "Parameter invariant observers," *International Journal of Control*, vol. 32, no. 6, pp. 1127–1132, 1980.

[13] P. Martin and E. Salaun, "Invariant observers for attitude and heading estimation from low-cost inertial and magnetic sensors," in *Decision and Control, 2007 46th IEEE Conference on*. IEEE, 2007, pp. 1039–1045.

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

[15] ——, "Algebraic parameter estimation using kernel representation of linear systems," in *The 20th World Congress of the International Federation of Automatic Control*, 2017.

[16] ——, "Double-sided kernel observer for linear time-varying systems," in *IEEE Conference on Control Technology and Applications*, 2017.

[17] ——, "Kernel-based adaptive multiple model target tracking," in *IEEE Conference on Control Technology and Applications*, 2017.

[18] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.

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

[20] M. Eslami, *Theory of sensitivity in dynamic systems: an introduction*. Springer Science & Business Media, 2013.

[21] Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution," *SIAM Journal on Scientific Computing*, vol. 24, no. 3, pp. 1076–1089, 2003.

[22] A. C. Hindmarsh and R. Serban, "User documentation for cvodes, an ode solver with sensitivity analysis capabilities," *Livermore (California): Lawrence Livermore National Laboratory*, vol. 189, 2002.