2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Finite Interval Estimation of LTI Systems Using Differential Invariance, Instrumental Variables, and Covariance Weighting*

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

Abstract— **It is shown how the kernel approach to joint parameter and state estimation can be improved to handle large measurement noise. High accuracy of estimation results from combining the powers of the kernel representation of the differential invariance in the system, a feasible recursive version of the generalized least squares with covariance weighting to eliminate regression dilution and suitable choices of instrumental variables to compensate for the error-in-the-variable in the stochastic regression formulation.**

I. INTRODUCTION

As rich as is the field of system identification and filtering, [1], [2], [3], there is a growing interest in non-asymptotic estimation methods which is justified by modern advances in nonlinear and hybrid systems, [4], modeling of complex dynamical systems and need for ever faster and more precise target tracking algorithms [5].

The idea of constructing kernel representations of differential systems, as first presented in [6], initially sprung from an attempt to improve on the algebraic signal differentiation approach proposed in [7], [8], as the principal tool required in the implementation of controllers for nonlinear systems that are differentially flat. It has since developed into a more general approach of finite interval system identification and orbit reconstruction for dynamical systems, [9], [10], from a single realization of an output measurement process subject to high noise. Indeed, robustness to noise proved to be a weak point of the original algebraic estimation idea despite attempts to justify it by arguments lent by non-standard analysis, [11], and design of the annihilator components of the estimators towards improved noise rejection properties, [12].

It was noticed early that the special double-sided kernel system representation introduced in [6] had an obvious advantage of alleviating the singularity problem at the extremities of the estimation interval and eliminating the need of re-initialization to compensate for accumulation of numerical error in the original estimation method of [13]. More importantly, the idea of kernel representation of differential invariants of LTI systems arrived with a bonus in the form of a reproducing kernel Hilbert space induced by the kernel themselves where the kernel functions could serve as a basis for projections and representations of bounded linear functionals. The kernel expressions were hence explicitly derived for systems of arbitrary order [9] and showed effective in joint parameter and state estimation of LTI, LTV, or even LPV systems, [10] [14], under small measurement noise.

The next step towards improvement then became quite obvious: to compensate for higher levels of measurement noise and fully explore the potential of kernel estimation based on differential invariance, the approach needed some rigour of statistical and stochastic analysis. This is the topic of the present contribution, which combines the advantages of both the deterministic and stochastic aspects and analysis of the kernel estimation problem:

- While the existence of a differential invariance is still assumed as a deterministic backbone of the kernel-based LTI estimation problem considered here, the resulting least squares regression formulation is shown to fall short of the satisfaction of the assumptions of the Gauss-Markov Theorem;
- Remedies under which consistency is restored in the estimation problem considered comprise: (a) the introduction of kernel-based instrumental variables to eliminate bias caused by what is known as the "error-in-the-variable" (a random regressor variable correlated with the regression error); (b) the introduction of a recursive version of what is known as a "feasible generalized least squares" to alleviate any problems associated with heteroskedastic regression error that cannot be avoided in reproducing kernel approach to estimation.

The estimation approach proposed is applied to an example problem which involves an unstable homogeneous LTI system of order four. It is shown that high fidelity reconstruction of the nominal system output and its first three time derivatives is possible from a single realization of a measurement process with large level of additive Gaussian noise.

II. A FINITE INTERVAL ESTIMATION PROBLEM FOR AN LTI SYSTEM

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 the 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).

The estimation problem is stated as follows. Given an arbitrary finite interval of time $[a,b]$, suppose that:

(i) The dimension of the state vector of the LTI system is known a priori alongside with the system input function $u(t)$ with its derivatives $u^{(i)}(t), i = 1, \cdots, n-1$ for $t \in [a,b]$ (implicit is the statement that the u is $n-1$ times differentiable);

(ii) The output of the system is observed as a single realization of a "continuous" measurement process $y_M(t) := y(t) + \eta(t)$, $t \in [a,b]$ in which η denotes additive white Gaussian noise with unknown intensity (variance) σ^2.

Although some of the derivations in this paper will indeed employ continuous representation of the measurement

*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

978-1-5386-8266-1/$31.00 ©2020 AACC

process, an implementable version of assumption (ii) simply requires availability of an unrestricted number of output measurements over the observation horizon $[a, b]$.

Under these conditions it is required that:

(a) Identifiability of the system input-output parameters $a_i, b_i, i = 0, \cdots, n-1$ from a single realization of the measurement process $y_M(t), t \in [a, b]$ is determined for the given observation horizon $[a, b]$;

(b) Under identifiability condition, a parameter estimator for $a_i, b_i, i = 0, \cdots, n-1$ be proposed that is statistically consistent;

(c) The true system output $y(t)$ and its $n-1$ time derivatives $y^{(i)}(t), i = 1, \cdot, n-1$ be reconstructed from the noisy observation $y_M(t)$ over $t \in [a, b]$;

(d) The reconstruction error in the output function y and its derivatives $y^{(i)}, i = 1, \cdot, n-1$ converges to zero uniformly as the number of measurement samples of y_M increases freely.

It is worth pointing out that assumption (i) requesting complete knowledge of the system input and its derivatives is not absolutely necessary. The approach can be potentially extended to the nonlinear case when the input function is also identified using B-splines or other non-parametric representations.

For simplicity of exposition, the estimation approach is presented here with respect to the more interesting case of homogeneous systems in which persistent excitation cannot be rendered by way of the system input u. Without the loss of generality, the order of the system considered in an example demonstration of the validity and properties of the approach was taken to be $n = 4$.

III. A KERNEL REPRESENTATION OF A SYSTEM DIFFERENTIAL INVARIANCE

The cornerstone of the finite interval estimation approach presented here is the integral representation of the controlled differential invariance of the system (3). For the homogeneous system the latter is given by a mapping J:

$$\begin{aligned}&J(y, y^{(1)}, \cdots, y^{(n)}) := \\ &y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1 y^{(1)}(t) + a_0 y(t) \\ &\quad \equiv const. = 0; \quad t \in [a, b]\end{aligned} \tag{4}$$

that remains constant under the action of the flow of the system. The validity of (4) is seen to deliver additional measurement-noise independent information about the behaviour of the system beyond mere observation of the noisy y_M. To be useful, however, the "zero-input response" characterization (4) has to be put in a form, which does not depend on the initial or boundary conditions of the system, and that does not involve any time derivatives of the output as the last are not available through direct measurement.

Such a representation of (4) has been presented for SISO LTI systems of arbitrary order in [9] as rigorously proved in a forthcoming publication [15]. It relies on the construction of an integral kernel as recalled in the result below.

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

$$\begin{aligned}y(t) = \alpha_{ab}^{-1}(t)\Bigg[&\int_a^b K_{DS,y}(n,t,\tau)y(\tau)d\tau \\ &+ \int_a^b K_{DS,u}(n,t,\tau)u(\tau)d\tau\Bigg]\end{aligned} \tag{5}$$

where the inverse of the annihilator of the boundary conditions of the system has been chosen as

$$\alpha_{ab}^{-1}(t) := \frac{1}{(t-a)^n + (b-t)^n} \tag{6}$$

Hilbert-Schmidt double-sided kernels of (5) 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:

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

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

The kernel functions $K_{DS,y}, K_{DS,u}$ are linear in the system parameters $a_i, b_i, i = 0, \cdots, n-1$ and are $n-1$ times differentiable as functions of the argument t.

□

It is important to note that for $u \equiv 0$, the invariance representation (5) is in fact a continuous evaluation functional for the system output functions and hence induces a unique reproducing kernel Hilbert space (RKHS) with kernel

$$K_y(t_1, t_2) := \alpha_{ab}^{-2}(t)\int_a^b K_{DS,y}(t_1,\tau)K_{DS,y}(t_2,\tau)d\tau \tag{7}$$

$t_1, t_2 \in [a, b]$ where the dependence of the kernels on the system order n has been suppressed for brevity. It then follows immediately that any bounded linear functional $L : y \to L(y)(t)$ on that space has a representer with the kernel $L(K_y(t, \cdot))$. This fact implies that the kernel representation of the time derivatives of the output functions involves only the derivatives of the $K_{DS}(\cdot, \tau)$, i.e. for the homogeneous case

$$y^{(i)}(t) = \int_a^b \frac{\partial}{\partial t}\left[\alpha_{ab}^{-1}(t)K_{DS,y}(t,\tau)\right] y(\tau)d\tau \quad t \in (a, b) \tag{8}$$

for $i = 1, \cdot, n-1$.

Finally, it should be understood, see also [9], that the kernel representation of the system invariance provides a unique criterion whose reproducing property unambiguously characterizes all zero input solutions of the SISO LTI system. In particular all the fundamental solutions of the LTI system share the reproducing property (5) as they span a subspace of the RKHS of dimension n.

Given the kernel representation of the system differential invariance, the estimation problem will be solved in two separate stages: (a) parametric estimation of the coefficients of the characteristic equation $a_i, i = 0, n-1$; (b) non-parametric estimation/reconstruction of the system output trajectory and its time derivatives over the interval $[a, b]$.

IV. PARAMETRIC ESTIMATION AS A LEAST SQUARES PROBLEM

As the kernels of Theorem 1 are linear in the unknown system coefficients, the reproducing property is first re-written to bring out this fact while omitting the obvious dependence of the kernels on n. [9].

$$y(t) = \int_a^b K_{DS,y}(t,\tau)y(\tau)\ d\tau \tag{9}$$

$$= \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t,\tau)y(\tau)d\tau \tag{10}$$

where the $K_{DS(i),y}; i = 0, \cdots, n$ are "component kernels" of $K_{DS,y}$ that post-multiply the coefficients $\beta_i = a_i; i = 0, \cdots, n-1$, with $\beta_n = 1$ for convenience of notation. In a noise-free deterministic setting, the output variable y becomes the measured output coinciding with the nominal output trajectory y_T, so the regression equation for the constant parameters $a_i, i = 0, \ldots, n-1$, (10), can be written in a partitioned form as

$$\begin{aligned} y_T(t) &= [K^{\bar{a}}, K^1](t; y_T)\beta^T \\ \bar{a} &:= [a_0, \cdots, a_{n-1}]; \quad \beta^T := [\bar{a}; \beta_n] \end{aligned} \tag{11}$$

where $K^{\bar{a}}(t; y_T)$ is a row vector with integral components

$$K^{\bar{a}}(t; y_T)_k := \int_a^b K_{DS(k),y}(t, \tau) y_T(\tau) d\tau \; ; \; k = 0, \ldots n-1 \tag{12}$$

while $K^1(t; y_T)$ is a scalar

$$K^1(t; y_T) := \int_a^b K_{DS(n),y}(t, \tau) y_T(\tau) d\tau \tag{13}$$

corresponding to $\beta_n = 1$. Given distinct time instants $t_1, \cdots, t_N \in (a, b]$, here referred to as *knots*, the regression equation is re-written point-wise in the form of a matrix equation

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

$$Q := \begin{bmatrix} q(t_1) \\ \vdots \\ q(t_N) \end{bmatrix} \; ; P := \begin{bmatrix} p_0(t_1) \cdots p_{n-1}(t_1) \\ \ddots \\ p_0(t_N) \cdots p_{n-1}(t_N) \end{bmatrix}$$

$$\begin{aligned} q(t_i) &= y_T(t_i) - K^1(t_i, y_T); \\ p_k(t_i) &= K^{\bar{a}}(t_i; y_T)_k \end{aligned} \tag{15}$$

that can be solved using least squares error minimization provided adequate identifiability assumptions are met and the output is measured without error.

Identifiability of homogeneous LTI systems from a single realization of a measured output

Referring to Definition and Theorem from [16], a homogeneous LTI system such as

$$\dot{x}(t) = Ax(t); \quad y = Cx; \quad x(0) = b; \quad x \in \mathbb{R}^n \tag{16}$$

is identifiable from a single noise-free realization of its output trajectory y under precise conditions, which admittedly are difficult to verify computationally.

However, the *practical version of identifiability* is sufficient for the present estimation purpose as defined below.

Definition 2: Practical linear identifiability
The homogeneous system (3) is practically linearly identifiable on $[a, b]$ with respect to a particular noisy discrete realization of the output measurement process, $y_M(t), t \in [a, b]$, 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*.

In the presence of large measurement noise, here assumed to be AWGN - white Gaussian and additive, the regression equation (10) is no longer valid as the reproducing property fails to hold along an inexact output trajectory. It must thus be suitably replaced leading to a stochastic regression problem. First, the stochastic output measurement process on a suitable space such as $L^2(\Omega, \mathcal{F}, \mathbb{P})$ where y_M is assumed to be adapted to the natural filtration of the standard Wiener process W on $[a, b]$ is

$$y_M(t, \omega) = y_T(t) + \sigma\dot{W}(t, \omega) \; ; \quad t \in [a, b] \tag{17}$$

where y_T is the true system output and $\dot{W}$ signifies the generalized derivative of the standard Wiener process; see e.g. [17], i.e. $\sigma\dot{W}$ is identified with the white noise process with constant variance σ^2. Now, without adhering to any particular realization of the measurement process (17) the kernel expression

$$\begin{aligned} &\int_a^b K_{DS,y}(t, \tau) y_M(\tau) \; d\tau \\ &= \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t, \tau) y_M(\tau) d\tau \\ &= \int_a^b K_{DS,y}(t, \tau) y_T(\tau) \; d\tau + \int_a^b K_{DS,y}(t, \tau) \; \sigma\dot{W}(\tau) d\tau \end{aligned} \tag{18}$$

is a random variable and the following equality holds

$$y_M(t) = \int_a^b K_{DS,y}(t, \tau) y_M(\tau) \; d\tau + e(t) \tag{19}$$

$$\text{with} \quad e(t) := \sigma\dot{W}(t) - \int_a^b K_{DS,y}(t, \tau) \; \sigma\dot{W}(\tau) d\tau \tag{20}$$

since y_T satisfies the reproducing property in the deterministic regression equation (10). It is noted that the random error variable e is dependent on the unknown system parameters $a_i, i = 0, \cdots, n-1$ while the stochastic regression equation

$$y_M(t) = \sum_{i=0}^{n} \beta_i \int_a^b K_{DS(i),y}(t, \tau) y_M(\tau) d\tau + e(t) \tag{21}$$

has the random regressor vector

$$\left[\int_a^b K_{DS(0),y}(t, \tau) y_M(\tau) d\tau, \cdots, \int_a^b K_{DS(n),y}(t, \tau) y_M(\tau) d\tau \right]^T \tag{22}$$

It is easily verified (see later developments) that the assumptions of the Gauss-Markov Theorem are violated in the linear regression problem (21) because the random regressor is correlated with a regression error, which additionally fails to be homoskedastic. The above regression is thus a typical "error-in-the-variable" problem with heteroskedastic noise. These complications are addressed next.

A. Heteroskedasticity

Heteroskedasticity has serious consequences for the OLS estimator. Despite that the OLS estimator remains unbiased, the estimated regression error is wrong while confidence intervals cannot be relied on. A standard quite powerful way to deal with *unknown* heteroskedasticity is to resort to generalized least squares (GLS), which can be shown to be BLUE (Best Linear Unbiased Estimator); see [18]. The GLS employs inverse covariance weighting in the regression error minimization problem associated with (21). In similarity with the notation used in the deterministic OLS (14) let $Q(y_M)$ and $P(y_M)$ be the matrices corresponding to N samples of the measurement process realization y_M at a batch of knots $t_1, \cdots, t_N$. Then the stochastic regression error vector is given by

$$e := [e(t_i), \cdots, e(t_N)]^T = Q(y_M) - P(y_M)\bar{a} \tag{23}$$

where $e(t_i)$ are as in (20). The standard GLS regression error minimization for estimation of the parameter vector $\bar{a}$ is

$$\min_{\bar{a}} \left([Q(y_M) - P(y_M)\bar{a}]^T S[Q(y_M) - P(y_M)\bar{a})] \right) \quad (24)$$

$$\text{with} \quad S := [\text{Cov}(e)]^{-1} \quad (25)$$

Applying the expectation operator to equations (17) and (20) and using standard properties of white noise, the covariance calculation yields (full derivation presented in forthcoming publication [15]):

$$\text{Cov}[e(t_i), e(t_j)] = \sigma^2 \delta(t_i - t_j) - \sigma^2 K_{DS}(t_i, t_j)$$
$$\sigma^2 K_{DS}(t_j, t_i) + \sigma^2 \int_a^b K_{DS}(t_i, \tau) K_{DS}(t_j, \tau) d\tau$$

At this point it should be clear that the standard GLS as in (24) cannot be applied directly as the covariance matrix depends on the unknown variance σ^2 and also on the unknown parameter vector $\bar{a}$ in the K_{DS} kernels. Hence a *feasible* version of the GLS must be employed here in which the covariance matrix is estimated progressively as more information about the regression residuals becomes available. This is typically performed as part of a recursive scheme in which consecutive batches of samples are drawn from the realization of y_M. Letting $Q_i - P_i\bar{a}$ denote the regression error e_i in batch i, the recursive GLS algorithm computes

$$\hat{a}_k = \arg\min_{\bar{a}} \left(\sum_{i=1}^{k} (Q_i - P_i\bar{a})^T S_i (Q_i - P_i\bar{a}) \right) \quad (26)$$

where $\hat{a}_k$ is the parameter estimate update at iteration k of the algorithm. Each weighting matrix S_{k+1}, is calculated as the inverse of the covariance matrix based on the parameter estimate $\hat{a}_k$ and an estimate of the variance σ^2 obtained from the residual trajectory $y_M(t) - y_E(t)$ in previous iteration k, where y_E signifies the estimated/reconstructed output.

B. Errors-in-variables

It is well known that the presence of errors-in-variables induces an asymptotic bias in OLS regression estimates which is proportional to the signal-to-noise ratio in the observed regressand. In such situations the leading way to eliminate estimation bias is to use *Instrumental Variables* (IV); see [19], in the normal equations that deliver the optimal estimates. The IV method has a long history and multiple applications; refer to [20], [21], [22], [23], [24], [25].

To render statistical consistency for the estimation problem at hand the IV is constructed by way of the backward reproducing kernel as described below.

Referring to the exposition of the basic regression problem in section IV it follows from Theorem 1 that the "double-sided" regression equation (11), can be cloned as two statistically independent regression equations corresponding to the forward and backward kernels $K_{F,y}$ and $K_{B,y}$ as follows:

$$(t-a)^n y_T(t) = \int_a^t K_{F,y}(t,\tau)\ y_T(\tau) d\tau$$
$$:= [K_F^{\bar{a}}, K_F^1](t; y_T)\ \beta\ ; t \in [a,b] \quad (27)$$

$$(b-t)^n y_T(t) = \int_t^b K_{B,y}(t,\tau)\ y_T(\tau) d\tau$$
$$:= [K_B^{\bar{a}}, K_B^1](t; y_T)\ \beta\ ; t \in [a,b] \quad (28)$$

Given a set of knots $[t_1, \cdots, t_N]; N > n; t_1 >> a; t_N << b$, the latter are written in discrete time as N copies of (27) and (28) in matrix-vector form

$$Y_T = K_F(y_T)\beta \quad (29)$$
$$Y_T = K_B(y_T)\beta \quad (30)$$

The equation (29) delivers OLS estimator for the parameter vector β corresponding to normal equation:

$$K_F(y_T)^T K_F(y_T)\beta = K_F(y_T)^T Y_T$$

namely:

$$\hat{\beta}_F := \left[K_F(y_T)^T K_F(y_T)\right]^{-1} K_F(y_T)^T Y_T \quad (31)$$

provided that the inverted matrix is nonsingular.

Of course, nothing stands in the way of pre-multiplying the forward estimation equation (29) by the backward matrix $K_B(y_T)$. If the matrix satisfies:

$$\det\left[K_B(y_T)^T K_F(y_T)\right] \neq 0 \quad (32)$$

then it delivers one more estimator which, in the noiseless case, will be competitive with that in (31):

$$\hat{\beta}_{IVF} := \left[K_B(y_T)^T K_F(y_T)\right]^{-1} K_B(y_T)^T Y_T \quad (33)$$

Clearly, as long as the nonsingularity condition (32) is satisfied then, in the noiseless case, both the estimators are bound to produce the same value of the estimated parameter vector, i.e. $\beta_F = \beta_{IVF}$.

However, it is easy to see that the forward and the backward kernel integrals (27) - (28) are defined on the respective intervals $[a,t]$ and $[t,b]$, which have t as their only point in common. This fact indicates possible selection of "instrumental variable": with K_B as the IV for the forward equation (33). Then (33) could be regarded as an "instrumental variable estimator".

Detailed justification of such choice of IV is discussed in forthcoming publication [15].

The statistical properties of the IV instrument as discussed above restore consistency of the modified GLS estimator employing such IV instrument; see [19] for a proof in the OLS case. However, the IV-GLS recursions are now modified to

$$\hat{\beta}_{k+1} = \hat{\beta}_k + R_{k+1}\tilde{P}_{k+1}^T S_{k+1}(\tilde{Q}_{k+1} - \tilde{P}_{k+1}\hat{\beta}_k) \quad (34)$$

$$R_{k+1} = R_k - R_k\tilde{P}_{k+1}^T(S_{k+1}^{-1} + \tilde{P}_{k+1}R_k\tilde{P}_{k+1}^T)^{-1}\tilde{P}_{k+1}R_k \quad (35)$$

where the matrices P and Q are replaced by $\tilde{P}$ and $\tilde{Q}$

$$\tilde{P} = K_B(y_M)^T K_F(y_M)\ ;\ \tilde{Q} = K_B(y_M)^T Y_M \quad (36)$$

and $M_{k+1} = \sum_{i=0}^{k+1} \tilde{P}_i^T S_i \tilde{P}_i$ and $R_{k+1} = M_{k+1}^{-1}$.

It is worth noticing that a stopping criterion of the recursive scheme is elegantly delivered by the fact that the final estimate should satisfy $\beta_n = 1$ where $\beta := [a_0, \cdots, a_{n-1}, \beta_n]$, lending a criterion

$$|\hat{\beta}_n - 1| < \epsilon \text{ for some } \epsilon > 0 \quad (37)$$

A recursive scheme for the estimation of $\hat{\beta}_{IVB}$ follows from a symmetric development.

When using the feasible recursive IV-GLS algorithm, the covariance of the error terms should be calculated by replacing K_{DS} by K_F or K_B appropriately.

V. Trajectory and Derivative Reconstruction

Once the parameters of the system are estimated successfully, the system output can be reconstructed by projection onto the finite dimensional subspace of the RKHS spanned by the fundamental solutions, here denoted by $\xi_1, \cdots, \xi_n$. The fundamental solutions are found by integrating the characteristic equations for n sets of initial conditions:

$$Y(0)_k := [y(0), y^{(1)}, \cdots, y^{(n-1)}] = B_k; \quad k = 1, \cdots, n \tag{38}$$

where B_k are the canonical basis vectors in $\mathbb{R}^n$. For efficiency of computations it makes sense to orthonormalize the set $\xi_k, k = 1, \cdots, n$ into $\zeta_k, k = 1, \cdots, n$ using the Gram-Schmidt orthonormalization procedure in L^2. As the orthonormalizing transformation is linear

$$\text{span}\{\xi_k, k = 1, \cdots, n\} = \text{span}\{\zeta_k, k = 1, \cdots, n\} := \Xi$$
$$< \zeta_i | \zeta_j >_2 = 0, i \neq j; \quad < \zeta_i | \zeta_i >_2 = 1$$

with $\langle \cdot | \cdot \rangle_2$ as the inner products in L^2. Since the noise-free output to be estimated is a linear combination of fundamental solutions

$$y_T = \sum_{i=1}^{n} c_i \zeta_i; \quad \text{with} \quad c_i = \langle y_T | \zeta_i \rangle_2, \quad i = 1, \cdots, n \tag{39}$$

it is natural to consider a similar form for the estimator $\hat{y}_T$ with linear estimators $\hat{c}_i$ for the coefficients c_i in the form

$$\hat{c}_i := \langle y_M | \zeta \rangle_2 = \int_a^b y_M(\tau) \zeta_i(\tau) d\tau, \quad i = 1, \cdots, n \tag{40}$$

The resulting estimator $\hat{y}_T$ using the coefficients of (40) is also BLUE because it can be viewed as an OLS in which a set of measurements y_M is projected onto a given subspace Ξ where the measurement errors are Gaussian i.i.d. hence satisfying the strict Gauss-Markov assumptions.

To conclude, given a measurement process realization $\bar{y}_M$ on $[a, b]$, the reconstructed output trajectory is obtained as

$$y_E(t) = \sum_{i=1}^{n} \langle \bar{y}_M | \zeta_i \rangle_2 \zeta_i(t); \quad t \in [a, b] \tag{41}$$

The estimates of the derivatives $y_E^{(i)}, i = 1, \cdots, n-1$ are obtained from y_E via the integral expressions

$$y^{(i)}(t) = \int_a^b K_{DS,y}^{(i)}(t, \tau) y_E(\tau) d\tau$$
$$K_{DS,y}^{(i)}(t, \tau) := \left(\frac{\partial}{\partial t} \right)^i K_{DS,y}(t, \tau); \quad i = 1, \cdots, n-1 \tag{42}$$

involving i-fold differentiations of the kernel $K_{DS,y}$.

VI. Results

The 4-th order system considered for demonstration of results is stated in controllable canonical form with characteristic equation

$$y^{(4)}(t) + a_3 y^{(3)}(t) + a_2 y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = 0 \tag{43}$$

and nominal values of parameters which makes it unstable

$$a_0 = 1, \quad a_1 = 5, \quad a_2 = 5, \quad a_3 = 0 \tag{44}$$

The following initial values are used:

$$y(0) = 0, \quad y^{(1)}(0) = 0, \quad y^{(2)} = 0, \quad y^{(3)} = 1 \tag{45}$$

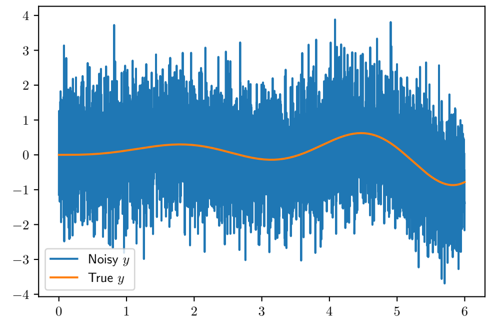

Fig. 1. Noisy y_M vs. nominal y_T

.

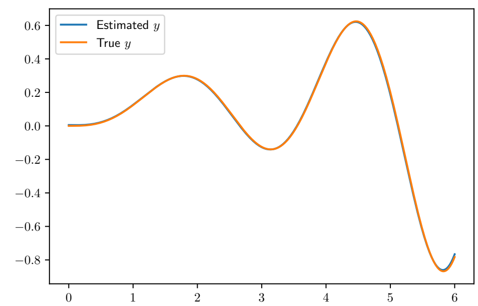

Fig. 2. Reconstructed nominal solution y_E

.

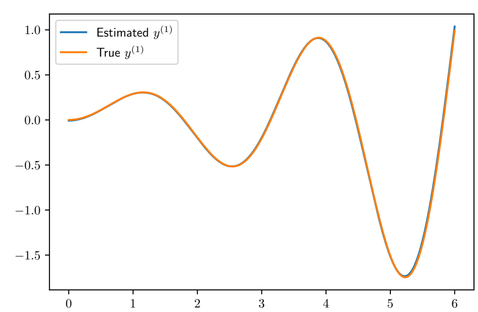

Fig. 3. Reconstructed $y_E^{(1)}$

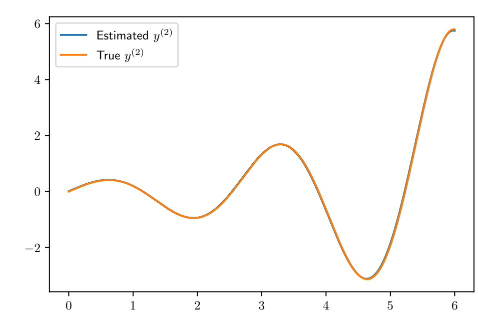

Fig. 4. Reconstructed $y_E^{(2)}$

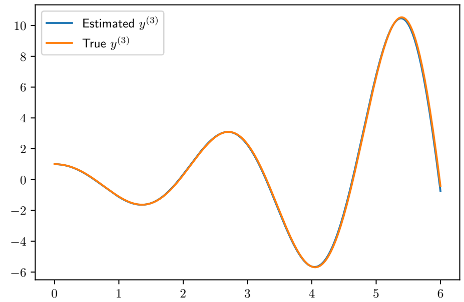

Fig. 5. Reconstructed $y_E^{(3)}$

TABLE I
ESTIMATED PARAMETER VALUES

	a_0	a_1	a_2	a_3	RMSD
True	1	5	5	0	
Estimated ($N = 500$)	0.9464	4.8137	5.1652	0.1293	0.0095
Estimated ($N = 1000$)	0.9806	5.0187	4.9978	0.0026	0.0060
Estimated ($N = 2000$)	0.9851	4.9987	5.0066	0.0018	0.0058

The measured realization of the output y_M is obtained as $y_M = y_T + \sigma\dot{W}$ with $\sigma = 1$ and sampled as needed. Figure 1 shows y_M vs. y_T.

As the quality of the feasible GLR is known to rely on large sample theory, the recursive feasible GLR with the IV as explained in section IV is employed in batches of $N = 500$, $N = 1000$, $N = 2000$ knots sampled from a uniform distribution over $[a, b] = [0, 6]$. The separation between the K_F and K_B used in IV is $\varepsilon = 0.01$sec. The threshold value for stopping the algorithm is $\epsilon = 0.01$.

The final parameter estimates are shown in Table I. The standard Root Mean Square Deviation (RMSD) is used to assess proximity of the nominal and estimated/reconstructed trajectories y_T and y_E. The parameter estimates did not improve further beyond the number of samples $N = 2000$, most probably due to accumulation of numerical errors.

Figures 2, 3, 4, 5 show the estimated/reconstructed output trajectories and their derivatives: y_E, $y_E^{(1)}$, $y_E^{(2)}$, $y_E^{(3)}$, respectively, calculated as explained in section V, using the estimated parameter values for $N = 2000$.

It is seen that the reconstructed output and derivative trajectories are virtually indistinguishable from their nominal counterpart, hence confirming the power of the estimation approach.

As the system is differentially flat [8], the state vector x_E, for $t \in [a, b]$, can be faithfully recovered from estimated y_E and its derivatives.

VII. CONCLUSIONS

The results presented here confirm that the knowledge of any differential invariance of a system is a powerful tool that can be used in estimation of system trajectories from very noisy measurements. Although there is an obvious price to pay: accessibility of a very large measurement data set and computational burden of processing it, the results are surprising in their precision. The approach presented is useful in several ways: (a) it provides a construction of $L - splines$ that are ideally fitted to the dynamic properties of the underlying invariance thus enabling parameter and trajectory estimation without the need of any regularization, (b) it delivers a vehicle of high fidelity system estimation and high order trajectory differentiation. In a way, the approach is a counterpart of the continuous-time Wiener filter, but necessitates fewer assumptions and is computationally more tractable. Finally, it should be mentioned that trajectory reconstruction can be performed in the basis provided by the kernel functions in place of fundamental solutions; this will be discussed elsewhere.

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] M. Fliess and H. Sira-Ramirez, "Control via state estimations of some nonlinear systems," in *IFAC Symposium on Nonlinear Control Systems (NOLCOS 2004)*, 2004.
[5] Y. Bar-Shalom, "Multitarget-multisensor tracking: Applications and advances. volume iii," *Norwood, MA, Artech House, Inc., 2000*, 2000.
[6] D. P. Ghoshal, K. Gopalakrishnan, and H. Michalska, "Using invariance to extract signal from noise," in *American Control Conference (ACC), 2017*. IEEE, 2017, pp. 2588–2593.
[7] 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.
[8] 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.
[9] D. P. Ghoshal and H. Michalska, "Finite-interval kernel-based identification and state estimation for LTI systems with noisy output data," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4982–4989.
[10] D. Sridhar, D. P. Ghoshal, and H. Michalska, "B-splines in joint parameter and state estimation in linear time-varying systems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 3508–3513.
[11] M. Fliess, "Analyse non standard du bruit," *Comptes Rendus Mathematique*, vol. 342, no. 10, pp. 797–802, 2006.
[12] M. Mboup, C. Join, and M. Fliess, "Numerical differentiation with annihilators in noisy environment," *Numerical algorithms*, vol. 50, no. 4, pp. 439–467, 2009.
[13] M. Fliess, M. Mboup, H. Mounier, and H. Sira-Ramirez, "Questioning some paradigms of signal processing via concrete examples," in *Algebraic Methods in Flatness, Signal Processing and State Estimation*. Editorial Lagares, 2003, pp. pp–1.
[14] 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.
[15] D. P. Ghoshal and H. Michalska, "Forward-backward kernel-based state and parameter estimation for linear systems of arbitrary order," *(submitted to) Numerical Algorithms*.
[16] S. Stanhope, J. Rubin, and D. Swigon, "Identifiability of linear and linear-in-parameters dynamical systems from a single trajectory," *SIAM Journal on Applied Dynamical Systems*, vol. 13, no. 4, pp. 1792–1815, 2014.
[17] S. Schäffler, *Generalized Stochastic Processes: Modelling and Applications of Noise Processes*. Springer, 2018.
[18] A. C. Aitken, "Iv.on least squares and linear combination of observations," *Proceedings of the Royal Society of Edinburgh*, vol. 55, pp. 42–48, 1936.
[19] P. C. Young, *Recursive estimation and time-series analysis: An introduction for the student and practitioner*. Springer Science & Business Media, 2011.
[20] D. Mayne, "A method for estimating discrete time transfer functions," in *Advances in Computer Control, Second UKAC Control Convention*. The University of Bristol, 1967.
[21] K. Wong and E. Polak, "Identification of linear discrete time systems using the instrumental variable method," *IEEE Transactions on Automatic Control*, vol. 12, no. 6, pp. 707–718, 1967.
[22] P. C. Young, "The use of linear regression and related procedures for the identification of dynamic processes," in *Seventh Symposium on Adaptive Processes*. IEEE, 1968, pp. 53–53.
[23] P. Caines, "On the asymptotic normality of instrumental variable and least squares estimators," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 598–600, 1976.
[24] B. Friedlander, "The overdetermined recursive instrumental variable method," *IEEE Transactions on Automatic Control*, vol. 29, no. 4, pp. 353–356, 1984.
[25] P. E. Caines, *Linear stochastic systems*. SIAM, 2018, vol. 77.