Linear dynamic filtering with noisy input and output [1]

Ivan Markovsky and Bart De Moor[2]

20 November 2002

Published in the proc. of
13th IFAC Symposium on System Identification (SYSID-2003)
pages 1749–1754

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/markovsky/reports/02-191.ps.gz*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD (SISTA), Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium,
Tel. 32/16/32 17 09, Fax 32/16/32 19 70,
WWW: *http://www.esat.kuleuven.ac.be/sista*.
E-mail: *ivan.markovsky@esat.kuleuven.ac.be*. **Research Council KUL:** GOA–Mefisto 666, several PhD/postdoc & fellow grants; **Flemish Government:** - **FWO:** PhD/postdoc grants, projects, G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), research communities (ICCoS, ANMMM); - **AWI:** Bil. Int. Collaboration Hungary/Poland; - **IWT:** PhD Grants, Soft4s (softsensors), **Belgian Federal Government:** DWTC (IUAP IV–02 (1996–2001) and IUAP V–22 (2002–2006), PODO–II (CP/40: TMS and Sustainibility); **EU:** CAGE; ERNSI; Eureka 2063–IMPACT; Eureka 2419–FliTE; **Contract Research/agreements:** Data4s, Electrabel, Elia, LMS, IPCOS, VIB;

Abstract

We establish the equivalence between the optimal least-squares state estimator for a linear time-invariant dynamic system with noise corrupted input and output, and an appropriately modified Kalman filter. The approach used is algebraic and the result shows that the noisy input/output filtering problem is not fundamentally different from the classical Kalman filtering problem. The result is illustrated with a simulation example.

LINEAR DYNAMIC FILTERING WITH NOISY INPUT AND OUTPUT

Ivan Markovsky * and **Bart De Moor** *

* *ESAT, SCD-SISTA, K.U.Leuven,*
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
{Ivan.Markovsky,Bart.DeMoor}@esat.kuleuven.ac.be
http://www.esat.kuleuven.ac.be/sista-cosic-docarch
Tel: +32–16–32 17 09, Fax: +32–16–32 19 70

Abstract: We establish the equivalence between the optimal least-squares state estimator for a linear time-invariant dynamic system with noise corrupted input and output, and an appropriately modified Kalman filter. The approach used is algebraic and the result shows that the noisy input/output filtering problem is not fundamentally different from the classical Kalman filtering problem. The result is illustrated with a simulation example.

Keywords: dynamic errors-in-variables model, Kalman filtering, optimal smoothing, total least squares.

1. INTRODUCTION

Optimal least-squares state estimation for linear dynamical systems is a well developed topic with many practical applications. A central result is the Kalman filter. In the discrete-time case, one considers the model

$$\begin{aligned} x(t+1) &= Ax(t)+Bu(t)+w(t),\\ y(t) &= Cx(t)+Du(t)+v(t), \end{aligned} \tag{1}$$

with $x(0)=x_0$ and $t=0,1,\ldots$ Here $u(t)\in\mathbb{R}^m$, $y(t)\in\mathbb{R}^l$, and $x(t)\in\mathbb{R}^n$ are the input, output, and state vectors at time instant t. When the model matrices A, B, C, and D are known, the state x depends linearly on the input u and the output y, so that it can be estimated from the input/output (I/O) data via the least-squares method. The unknown *process noise* w and the *measurement noise* v play the role of *equation errors* in (1). Even though v and w are unknown, we assume that they are zero mean, white, Gaussian noises, with covariance matrix

$$\mathbf{E}\begin{bmatrix} w(t)\\ v(t)\end{bmatrix}\begin{bmatrix} w^T(t+\tau) & v^T(t+\tau)\end{bmatrix}=\begin{bmatrix} Q(t) & S(t)\\ S^T(t) & R(t)\end{bmatrix}\delta(\tau).$$

The importance of the Kalman filter is that it solves the least-squares problem recursively and in real time. Model (1), however, is in a certain sense asymmetric. The process noise w acts as an unobserved input while the measurement noise v represents measurement error on y. However, u is assumed to be noiseless! In the paper, we pose and answer the question:

> How should we modify the Kalman filter when *both* the input and the output of the system are noisy?

The paradigm of treating the input and the output on an equal footing leads to the behavioral approach (Polderman and Willems, 1998). A derivation of the Kalman filter in the behavioral context is given in (Fagnani and Willems, 1997).

In this paper, we consider the deterministic discrete-time LTI state-space system

$$\begin{aligned} x(t+1) &= Ax(t)+Bu(t), \qquad x(0)=x_0,\\ y(t) &= Cx(t)+Du(t), \qquad t=0,1,\ldots \end{aligned} \tag{2}$$

together with the *measurement errors model*

$$u_{\mathrm{d}}(t)=u(t)+\tilde{u}(t), \qquad y_{\mathrm{d}}(t)=y(t)+\tilde{y}(t). \tag{3}$$

and refer to the model (2) together with the measurement errors model (3) as the *noisy I/O model* (see Figure 1). We note that (2–3) is an *errors-in-variables* (EIV) model in the sense of, *e.g.*, (Söderström *et*

al., 2002; Zheng, 2002), when the problem would be to estimate A, B, C, and D. But since the purpose of this paper is to estimate x, knowing A, B, C, and D, we call it a noisy I/O model.

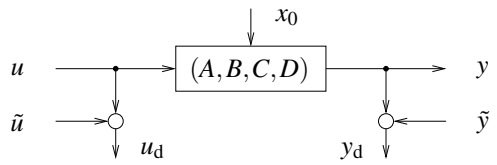


Fig. 1. Block scheme of the noisy I/O model.

A full analysis of a static EIV estimation problem, also known as the *total least squares* (TLS) problem, is given in the monograph (Van Huffel and Vandewalle, 1991). The dynamic equivalent of the TLS problem, see (Aoki and Yue, 1970; De Moor and Roorda, 1994), is a system identification problem: given noisy I/O measurements, find the model (2). The dynamic TLS problem can be expressed as a static TLS problem with Toeplitz (or Hankel) structured data matrices. The *structured total least squares* problem is treated in (De Moor, 1993; Lemmerling, 1999).

Different type of estimation problem, in the EIV context, occurs when the model is exactly known and the state vector has to be estimated from the noisy I/O measurements. We refer to this latter problem as the *noisy I/O state estimation* problem. In (Guidorzi *et al.*, 2003; Diversi *et al.*, 2003), the noisy I/O state estimation problem has been considered, using the language of transfer functions. The authors there claim that it is fundamentally different problem from the Kalman filtering problem and describe new recursive algorithms for its solution.

In this paper, we prove that the noisy I/O state estimation problem is *not* fundamentally different from the Kalman filtering problem. Its solution boils down to the solution of a sequence of linear least-squares problems with a special structure coming from the state space equation (1). Our approach is linear algebraic. We represent the model over a finite time horizon as a set of linear equations and apply standard linear algebra techniques for its analysis. The optimal solution is shown to be a Kalman filter with correlated process and measurement noises. The continuous-time version of the noisy I/O state estimation problem is treated in (Markovsky *et al.*, 2002), where a completion of squares approach is used and the solution is also shown to be a Kalman filter-type system.

In Section 2, we define the smoothing and filtering noisy I/O state estimation problems. Two explicit block solutions of the smoothing problem are derived in Section 3. They are weighted least-squares problems. In Section 4, we transform the noisy I/O model in the form (1) and define the Kalman filter for the resulting system as the modified Kalman filter. In Section 5, we prove the deterministic equivalence of the estimates of the noisy I/O filter and the modified Kalman filter. Section 6 further establishes the optimal estimates of the true input/output signals. In Section 7, we confirm the results on an example and in Section 8, give conclusions.

2. PROBLEM FORMULATION

An on-line, recursive estimation procedure can be realized by a causal dynamical system called a *filter*. Such a filter operates on previous and current measurements and produces an estimate of the to-be-estimated signal for the current moment of time. The problem of a filter synthesis is referred to as a *filtering problem*. A *smoothing problem* is an estimation problem, in which on the basis of the available measurements, an estimate for the to-be-estimated signal is produced for the whole (past) period of observation.

We introduce some notation used in the rest of the paper. A signal variable, without time index, denotes the vector obtained by stacking one over another the signal samples for the consecutive time instances. For example, over the time horizon $0,1,\ldots,t_f-1$, the vector of the consecutive input samples u is defined as

$$u := [u^T(0) \ \cdots \ u^T(t_f-1)]^T$$

and the vector of the consecutive state samples is

$$x := [x^T(0) \ \cdots \ x^T(t_f-1) \ \ x^T(t_f)]^T.$$

For a time indexed matrix sequence $\{V(t)\}_{t=0}^{t_f-1}$, we denote by V, without time argument, the block matrix

$$V := \operatorname{blk}\operatorname{diag}\big(V(0),\ldots,V(t_f-1)\big).$$

Definition 1. (Optimal noisy I/O smoothing problem). Assume that the measurement errors $\tilde{u}$ and $\tilde{y}$ are random, centered, normal, uncorrelated, and white with known covariance matrices

$$\operatorname{cov}\big(\tilde{u}(t)\big) =: V_{\tilde{u}}(t), \qquad \operatorname{cov}\big(\tilde{y}(t)\big) =: V_{\tilde{y}}(t), \tag{4}$$

and that the initial condition x_0 is unknown. Then, given the matrices A, B, C, D, the optimal noisy I/O smoothing problem is defined as

$$\begin{aligned} \min_{\hat{u},\hat{y},\hat{x}} \ & \left\| \begin{bmatrix} V_{\tilde{u}} & \\ & V_{\tilde{y}} \end{bmatrix}^{-1/2} \begin{bmatrix} \hat{u}-u_\mathrm{d} \\ \hat{y}-y_\mathrm{d} \end{bmatrix} \right\|^2 \\ \text{s.t.} \ & \begin{aligned} \hat{x}(t+1) &= A\hat{x}(t)+B\hat{u}(t) \\ \hat{y}(t) &= C\hat{x}(t)+D\hat{u}(t) \end{aligned} \end{aligned} \tag{5}$$

for $t = 0,1,\ldots,t_f-1$. The *optimal smoothed state estimate* $\hat{x}(\cdot,t_f)$ is the solution of (5).

Under the normality assumption for the noises, $\hat{x}(\cdot,t_f)$ is the minimum variance estimate, the maximum likelihood estimate, and the conditional expectation estimate of the state x. The equivalence is well known in the Kalman filter case (Willems, 2002; Bryson and Ho, 1975; Anderson and Moore, 1979), and is shown in the noisy I/O case in (Guidorzi *et al.*, 2003).

Definition 2. (Optimal noisy I/O filtering problem). Given the model (2–3), satisfying the assumptions of Definition 1, the optimal noisy I/O filtering problem is to find a dynamical system,

$$\begin{aligned} z(t+1) &= A_{\mathrm{f}}(t)z(t) + B_{\mathrm{f}}(t)\begin{bmatrix} u_{\mathrm{d}}(t) \\ y_{\mathrm{d}}(t) \end{bmatrix}, \\ \hat{x}(t) &= C_{\mathrm{f}}(t)z(t) + D_{\mathrm{f}}(t)\begin{bmatrix} u_{\mathrm{d}}(t) \\ y_{\mathrm{d}}(t) \end{bmatrix}, \end{aligned} \tag{6}$$

such that $\hat{x}(t) = \hat{x}(t,t+1)$, where $\hat{x}(\cdot)$ is the solution of (6), *i.e.*, the *optimal filtered state estimate*, and $\hat{x}(\cdot,t+1)$ is the optimal smoothed state estimate with a time horizon $t+1$.

3. SMOOTHING BY BLOCK PROCESSING

In this section, we write the optimal noisy I/O smoothing problem (5) as a weighted least-squares (WLS) problem. This representation is used as a conceptual tool for the analysis and not as a means to carry out the actual computations needed for the estimation.

We represent the I/O dynamics of the system (2), over the time horizon $0,\ldots,t_f-1$, explicitly as

$$y = \Gamma x_0 + Tu, \tag{7}$$

where

$$\Gamma := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{t_f-2} \end{bmatrix}, \quad \text{and} \quad T := \begin{bmatrix} H(0) & 0 & \cdots & 0 \\ H(1) & H(0) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ H(t_f-1) & H(t_f-2) & \cdots & H(0) \end{bmatrix}.$$

The matrix Γ is an extended observability matrix and T is a Toeplitz matrix formed from the Markov parameters

$$H(0) = D, \quad H(t) = CA^{t-1}B,\ t = 1,\ldots,t_f-1.$$

Using (7), we see that the optimal noisy I/O smoothing problem (5) is a weighted least-squares problem

$$\min_{\hat{x}_0,\hat{u}} \left\| \begin{bmatrix} V_{\tilde{u}} & \\ & V_{\tilde{y}} \end{bmatrix}^{-\frac{1}{2}} \left(\begin{bmatrix} u_{\mathrm{d}} \\ y_{\mathrm{d}} \end{bmatrix} - \begin{bmatrix} 0 & I \\ \Gamma & T \end{bmatrix} \begin{bmatrix} \hat{x}_0 \\ \hat{u} \end{bmatrix} \right) \right\|^2. \tag{8}$$

Alternatively, we represent the input/state/output dynamics of the system, over the time horizon $0,\ldots,t_f-1$, as

$$\begin{bmatrix} y(0) \\ 0 \\ y(1) \\ 0 \\ \vdots \\ y(t_f-1) \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} C & 0 & & & & \\ A & -I & & & & \\ & C & 0 & & & \\ & A & -I & & & \\ & & \ddots & \ddots & & \\ & & & & C & 0 \\ & & & & A & -I \end{bmatrix}}_{\mathscr{A}} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(t_f) \end{bmatrix} + \underbrace{\begin{bmatrix} D & & & \\ B & & & \\ & D & & \\ & B & & \\ & & \ddots & \\ & & & D \\ & & & B \end{bmatrix}}_{-\mathscr{B}} \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(t_f-1) \end{bmatrix}.$$

Substituting $y_{\mathrm{d}} - \tilde{y}$ for y and $u_{\mathrm{d}} - \tilde{u}$ for u (see (3)), we have

$$\begin{aligned} &\underbrace{\begin{bmatrix} y_{\mathrm{d}}(0) \\ 0 \\ y_{\mathrm{d}}(1) \\ 0 \\ \vdots \\ y_{\mathrm{d}}(t_f-1) \\ 0 \end{bmatrix}}_{\bar{y}_{\mathrm{d}}} + \mathscr{B} \begin{bmatrix} u_{\mathrm{d}}(0) \\ u_{\mathrm{d}}(1) \\ \vdots \\ u_{\mathrm{d}}(t_f-1) \end{bmatrix} = \\ &= \mathscr{A}x + \mathscr{B} \begin{bmatrix} \tilde{u}(0) \\ \tilde{u}(1) \\ \vdots \\ \tilde{u}(t_f-1) \end{bmatrix} + \underbrace{\begin{bmatrix} I & & & \\ 0 & & & \\ & I & & \\ & 0 & & \\ & & \ddots & \\ & & & I \\ & & & 0 \end{bmatrix}}_{\mathscr{C}} \begin{bmatrix} \tilde{y}(0) \\ \tilde{y}(1) \\ \vdots \\ \tilde{y}(t_f-1) \end{bmatrix}, \end{aligned}$$

or with the definition of the new variables

$$\bar{y}_{\mathrm{d}} + \mathscr{B}u_{\mathrm{d}} = \mathscr{A}x + \mathscr{B}\tilde{u} + \mathscr{C}\tilde{y}. \tag{9}$$

Using (9), the optimal noisy I/O smoothing problem is equivalent to the following problem

$$\begin{aligned} &\min_{\hat{x},\Delta u,\Delta y} \left\| \begin{bmatrix} V_{\tilde{u}} & \\ & V_{\tilde{y}} \end{bmatrix}^{-1/2} \begin{bmatrix} \Delta u \\ \Delta y \end{bmatrix} \right\|^2 \\ &\text{s.t.} \quad \bar{y}_{\mathrm{d}} + \mathscr{B}u_{\mathrm{d}} = \mathscr{A}\hat{x} + \begin{bmatrix} \mathscr{B} & \mathscr{C} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta y \end{bmatrix}, \end{aligned} \tag{10}$$

which solution is given in Section 5.

The solutions (8) and (10) of the noisy I/O smoothing problem are not recursive and thus not practical for large data sets. See (Markovsky *et al.*, 2002) for recursive solution of an equivalent continuous-time problem. In the discrete-time case, the recursive solution is given by two time-varying filters; one running backward in time and one running forward in time. The forward recursion is defined by a time-varying Riccati equation. The backward filter produces the optimal smoothed state estimate.

4. THE MODIFIED KALMAN FILTER

We convert the noisy I/O model (2–3) in the form (1) by substituting $u_{\mathrm{d}}(t) - \tilde{u}$ for $u(t)$ and $y_{\mathrm{d}}(t) - \tilde{y}$ for $y(t)$ (see (3)) in (2)

$$\begin{aligned} x(t+1) &= Ax(t) + Bu_{\mathrm{d}}(t) - B\tilde{u}(t) \\ y_{\mathrm{d}}(t) &= Cx(t) + Du_{\mathrm{d}}(t) - D\tilde{u}(t) + \tilde{y}(t) \end{aligned}$$

and define (fake) process noise w and measurement noise v by

$$w := -B\tilde{u} \quad \text{and} \quad v := -D\tilde{u} + \tilde{y}.$$

The resulting system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu_{\mathrm{d}}(t) + w(t) \\ y_{\mathrm{d}}(t) &= Cx(t) + Du_{\mathrm{d}}(t) + v(t) \end{aligned} \tag{11}$$

is in the form (1), where

$$\begin{bmatrix} Q(t) & S(t) \\ S^T(t) & R(t) \end{bmatrix} = \begin{bmatrix} -B & 0 \\ -D & I \end{bmatrix} \begin{bmatrix} V_u(t) & \\ & V_y(t) \end{bmatrix} \begin{bmatrix} -B & 0 \\ -D & I \end{bmatrix}^T.$$

We call the Kalman filter of the modified system (11), *i.e.*, the system

$$\begin{aligned} z(t+1) &= A_{\mathrm{KF}}(t)z(t) + B_{\mathrm{KF}}(t)\begin{bmatrix} u_{\mathrm{d}}(t) \\ y_{\mathrm{d}}(t) \end{bmatrix}, \\ \hat{x}(t) &= C_{\mathrm{KF}}(t)z(t) + D_{\mathrm{KF}}(t)\begin{bmatrix} u_{\mathrm{d}}(t) \\ y_{\mathrm{d}}(t) \end{bmatrix}, \end{aligned} \tag{12}$$

where

$$\begin{aligned} A_{\mathrm{KF}}(t) &= \big(A - K(t)C\big), \quad B_{\mathrm{KF}}(t) = \big[B - K(t)D, \;\; K(t)\big], \\ C_{\mathrm{KF}}(t) &= I - P(t)C^T\big(CP(t)C^T + R(t)\big)^{-1}C, \\ D_{\mathrm{KF}}(t) &= P(t)C^T\big(CP(t)C^T + R(t)\big)^{-1}\big[-D \;\; I\big], \\ K(t) &= \big(AP(t)C^T + S(t)\big)\big(CP(t)C^T + R(t)\big)^{-1}, \end{aligned}$$

and

$$\begin{aligned} P(t+1) &= AP(t)A^T - \big(AP(t)C^T + S(t)\big)\times \\ &\times \big(CP(t)C^T + R(t)\big)^{-1}\big(AP(t)C^T + S(t)\big)^T + Q(t), \end{aligned}$$

the *modified Kalman filter*. It recursively solves (9) for the last block entry of the unknown x. The solution is in the sense of the weighted least-squares problem

$$\min_{\hat{x},\hat{e}} ||V_e^{-1/2}\hat{e}||^2 \quad \text{s.t.} \quad \bar{y}_{\mathrm{d}} + \mathscr{B}u_{\mathrm{d}} = \mathscr{A}\hat{x} + \hat{e}. \tag{13}$$

The variable $\hat{e}$ accounts for the cumulative noise

$$e := [\mathscr{B} \;\; \mathscr{C}]\begin{bmatrix} \tilde{u} \\ \tilde{y} \end{bmatrix}$$

added to the equation and the covariance matrix of e is $V_e = \mathscr{B}V_{\tilde{u}}\mathscr{B}^T + \mathscr{C}V_{\tilde{y}}\mathscr{C}^T$.

When the measurement noise covariances $V_{\tilde{u}}(t)$ and $V_{\tilde{u}}(t)$ does not depend on t, one can replace the time-varying Kalman filter with the (suboptimal) time-invariant filter, obtained by replacing $P(t)$ in (12) with the steady-state solution $\bar{P}$ of the algebraic Riccati equation

$$\bar{P} = A\bar{P}A^T - \big(A\bar{P}C^T + S\big)\big(C\bar{P}C^T + R\big)^{-1}\big(A\bar{P}C^T + S\big)^T + Q.$$

In the following section, we investigate the relation between the modified Kalman filter (12) state estimate and the noisy I/O filter state estimate.

5. EQUIVALENCE OF THE MODIFIED KALMAN FILTER AND THE NOISY I/O FILTER

Consider the linear system of equations (9) and the two solution methods (10) and (13). Denote

$$z := \bar{y}_{\mathrm{d}} + \mathscr{B}u_{\mathrm{d}}, \quad \delta := \begin{bmatrix} \tilde{u} \\ \tilde{y} \end{bmatrix}, \quad V_\delta := \begin{bmatrix} V_{\tilde{u}} & \\ & V_{\tilde{y}} \end{bmatrix},$$
$$\hat{\delta} := \begin{bmatrix} \Delta u \\ \Delta y \end{bmatrix}, \quad \text{and} \quad \mathscr{D} := [\mathscr{B} \;\; \mathscr{C}].$$

We want to find a relation between the solutions of the following problems

$$\min_{\hat{x},\hat{e}} ||V_e^{-1/2}\hat{e}||_2^2 \quad \text{s.t.} \quad z = \mathscr{A}\hat{x} + \hat{e}, \tag{14}$$

where $V_e = \mathscr{D}V_\delta\mathscr{D}^T$ and

$$\min_{\hat{x},\hat{\delta}} ||V_\delta^{-1/2}\hat{\delta}||_2^2 \quad \text{s.t.} \quad z = \mathscr{A}\hat{x} + \mathscr{D}\hat{\delta}. \tag{15}$$

The first problem is a weighted least-squares problem and its solution is

$$\hat{x}_{\mathrm{KF}} = (\mathscr{A}^T V_e^{-1}\mathscr{A})^{-1}\mathscr{A}^T V_e^{-1} z. \tag{16}$$

The noisy I/O estimation problem (15) is a minimum-norm type problem and its solution is (see Lemma 3 in the Appendix)

$$\hat{x} = [0 \;\; I][V_e\mathscr{A}^\perp \;\; \mathscr{A}]^{-1}z, \tag{17}$$

where $\mathscr{A}^\perp$ is a matrix which columns form a basis for the orthogonal complement of the range space of $\mathscr{A}$.

We transform (16) and (17) by the change of variables

$$\bar{\mathscr{A}} := V_e^{-1/2}\mathscr{A} \qquad \text{and} \qquad \bar{z} := V_e^{-1/2}z.$$

Then

$$\hat{x}_{\mathrm{KF}} = (\bar{\mathscr{A}}^T\bar{\mathscr{A}})^{-1}\bar{\mathscr{A}}^T\bar{z}, \tag{18}$$

and

$$\hat{x} = [0 \;\; I][\bar{\mathscr{A}}^\perp \;\; \bar{\mathscr{A}}]^{-1}\bar{z}. \tag{19}$$

((19) follows from the identity $\bar{\mathscr{A}}^\perp = V_e^{1/2}\mathscr{A}^\perp$.) Now the question of the solutions' equivalence is answered by Theorem 4, see the Appendix, which states that $\hat{x}_{\mathrm{KF}} = \hat{x}$. Thus the two solutions are *deterministically* equal and the noisy I/O filtering problem is solved by the modified Kalman filter (12), *i.e.* $A_{\mathrm{f}} = A_{\mathrm{KF}}$, $B_{\mathrm{f}} = B_{\mathrm{KF}}$, $C_{\mathrm{f}} = C_{\mathrm{KF}}$, and $D_{\mathrm{f}} = D_{\mathrm{KF}}$.

6. OPTIMAL ESTIMATION OF THE TRUE INPUT/OUTPUT SIGNALS

Up to now we were interested in the optimal filtering in the sense of state estimation. In this section, we show how the optimal estimates of the input and the output can be derived from the modified Kalman filter.

The solutions $\hat{e}$ and $\hat{\delta}$ of (14) and (15), respectively, satisfy the following relation

$$\begin{aligned} \hat{e} &= V_e\mathscr{A}^\perp[V_e\mathscr{A}^\perp \;\; \mathscr{A}]^{-1}z \\ &= \mathscr{D}V_\delta\mathscr{D}^T\mathscr{A}^\perp[V_e\mathscr{A}^\perp \;\; \mathscr{A}]^{-1}z = \mathscr{D}\hat{\delta}. \end{aligned}$$

This implies that the state estimate $\hat{x}$, the one-step-ahead prediction $z(t+1)$, and the optimal input estimate $\hat{u}$ satisfy the equation

$$z(t+1) = A\hat{x}(t) + B\hat{u}(t). \tag{20}$$

Then we can find $\hat{u}$ *exactly* from $\hat{x}$ and $z(t+1)$, obtained from the modified Kalman filter (12). In fact, (20) and the Kalman filter equations imply that

$$\hat{u}(t) = E(t)z(t) + F(t)\begin{bmatrix} u_{\mathrm{d}}(t) \\ y_{\mathrm{d}}(t) \end{bmatrix},$$

where

$$E(t) := -V_{\tilde{u}}D^T\big(CP(t)C^T + R(t)\big)^{-1}C \quad \text{and}$$
$$F(t) := \Big[I - V_{\tilde{u}}D^T\big(CP(t)C^T + R(t)\big)^{-1}D, \;\; V_{\tilde{u}}D^T\big(CP(t)C^T + R(t)\big)^{-1}\Big].$$

The optimal output estimate is $\hat{y}(t) = C\hat{x}(t) + D\hat{u}(t)$.

7. NUMERICAL EXAMPLE

In this section, we verify numerically the equivalence of the solutions established in Section 5. The particular system, we use, is

$$A = \begin{bmatrix} 0.6 & -0.45 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$
$$C = [0.48429 \quad -0.45739], \quad \text{and} \quad D = 0.5381.$$

The time horizon is $t_f = 100$, the initial state is $x_0 = 0$, and the input signal is a normal white noise sequence with unit variance. The input and the output of the system are corrupted by independent, centered, normal, white noises with variances $\mathrm{var}(\tilde{u}(t)) = 0.4$ and $\mathrm{var}(\tilde{y}(t)) = 0.4$ for all t.

The estimate of the noisy I/O filter is computed directly from the definition, *i.e.*, we solve a sequence of optimal smoothing problems with increasing time-horizon. Every smoothing problem is a weighted least-squares problem that is solved explicitly according to (8). The last block entries of the obtained sequence of solutions form the noisy I/O filter state estimate.

We compare the noisy I/O filter estimate with the estimate of the modified Kalman filter (12). The experiment is carried out in MATLAB. The state estimate $\hat{x}_{\mathrm{KF}}$ obtained by the modified Kalman filter is up to the numerical errors equal to the state estimate $\hat{x}_{\mathrm{f}}$ obtained by the noisy I/O filter, $||\hat{x}_{\mathrm{KF}} - \hat{x}_{\mathrm{f}}|| = 5.7723e-15$. This is the desired numerical verification of the theoretical result of the paper. The absolute errors of estimation $||\hat{x}-x||^2$, $||\hat{u}-u||^2$, $||\hat{y}-y||^2$ for all estimation methods, discussed in the paper is given in Table 1.

Table 1. Comparison of the absolute errors of the state, input, and output estimates for all methods and the noisy data. (MKF — modified Kalman filter)

Method	$\|\|\hat{x}-x\|\|^2$	$\|\|\hat{u}-u\|\|^2$	$\|\|\hat{y}-y\|\|^2$
optimal smoothing	75.3981	29.2195	15.5409
optimal filtering	75.7711	35.5604	16.4571
time-varying MKF	75.7711	35.5604	16.4571
time-invariant MKF	76.1835	35.7687	16.5675
noisy data	116.3374	42.4711	41.2419

8. CONCLUSIONS

We considered optimal noisy I/O estimation problems for discrete-time LTI systems. The filtering problem is solved via a modified Kalman filter. The equivalence between the optimal noisy I/O filter and the modified Kalman filter is proven algebraically using explicit state-space representation of the system.

ACKNOWLEDGEMENTS

Ivan Markovsky is a research assistant and Dr. Bart De Moor is a full professor at the Katholieke Universiteit Leuven, Belgium. Our research is supported by **Research Council KUL:** GOA–Mefisto 666, several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), research communities (ICCoS, ANMMM); **AWI:** Bil. Int. Collaboration Hungary/Poland; **IWT:** PhD Grants, Soft4s (softsensors), **Belgian Federal Government:** DWTC (IUAP IV–02 (1996–2001) and IUAP V–22 (2002–2006), PODO–II (CP/40: TMS and Sustainibility); **EU:** CAGE; ERNSI; Eureka 2063–IMPACT; Eureka 2419–FliTE; **Contract Research/agreements:** Data4s, Electrabel, Elia, LMS, IPCOS, VIB.

9. REFERENCES

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Prentice Hall.

Aoki, M. and P. C. Yue (1970). On certain convergence questions in system identification. *SIAM J. Control* **8**(2), 239–256.

Bryson, A. and Y. Ho (1975). *Applied Optimal Control*. Hemisphere, Washington, D.C.

De Moor, B. (1993). Structured total least squares and L_2 approximation problems. *Lin. Alg. and Its Appl.* **188–189**, 163–207.

De Moor, B. and B. Roorda (1994). L_2-optimal linear system identification structured total least squares for SISO systems. In: *In the proceedings of the CDC*. pp. 2874–2879.

Diversi, R., R. Guidorzi and U. Soverini (2003). Algorithms for optimal errors-in-variables filtering. *Systems & Control Letters* **48**, 1–13.

Fagnani, F. and J. C. Willems (1997). Deterministic Kalman filtering in a behavioral framework. *Systems & Control Letters* **32**, 301–312.

Guidorzi, R., R. Diversi and U. Soverini (2003). Optimal errors-in-variables filtering. *Automatica* **39**, 281–289.

Lemmerling, P. (1999). Structured total least squares: analysis, algorithms and applications. PhD thesis. ESAT/SISTA, K.U. Leuven.

Markovsky, I., J. C. Willems and B. De Moor (2002). Continuous-time errors-in-variables filtering. In: *Proc. of the Conference on Decision and Control*. pp. 2576–2581.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM.

Polderman, J. W. and J. C. Willems (1998). *Introduction to mathematical systems theory*. Springer-Verlag.

Söderström, T., U. Soverini and K. Mahata (2002). Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing* **82**, 1139–1154.

Van Huffel, S. and J. Vandewalle (1991). *The total least squares problem: Computational aspects and analysis*. SIAM, Philadelphia.

Willems, J. C. (2002). Deterministic Kalman filtering. *J. of Econometrics*, to appear.

Zheng, W. X. (2002). A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Trans. on Aut. Control* **47**(7), 1142–1147.

Appendix A. SOLUTION OF THE OPTIMIZATION PROBLEM (15) AND PROOF OF THE STATE ESTIMATES' EQUIVALENCE

Lemma 3. Assuming that $\mathscr{A}$ is full rank and V_δ is positive definite. The solution of the minimum-norm type problem (15) is

$$\begin{bmatrix}\hat{\delta}\\ \hat{x}\end{bmatrix} = \begin{bmatrix}V_\delta \mathscr{D}^T \mathscr{A}^\perp & 0\\ 0 & I\end{bmatrix} [V_e \mathscr{A}^\perp \;\; \mathscr{A}]^{-1} z.$$

PROOF. The Lagrangian of (15) is

$$L(\hat{x},\hat{\delta},\lambda) = \hat{\delta}^T V_\delta^{-1}\hat{\delta} + \lambda^T(\mathscr{A}\hat{x} + \mathscr{D}\delta - z).$$

The first order optimality condition

$$\frac{\partial L}{\partial \delta} = 2V_\delta^{-1}\hat{\delta} + \mathscr{D}^T\lambda = 0 \quad \Rightarrow \quad \hat{\delta} = -\frac{1}{2}V_\delta \mathscr{D}^T\lambda, \qquad \text{(A.1)}$$

$$\frac{\partial L}{\partial \hat{x}} = \mathscr{A}^T\lambda = 0 \quad \Rightarrow \quad \lambda = -2\mathscr{A}^\perp\bar{\lambda}, \qquad \text{(A.2)}$$

$$\frac{\partial L}{\partial \lambda} = \mathscr{A}\hat{x} + \mathscr{D}\hat{\delta} - z = 0 \quad \Rightarrow \quad z = \mathscr{A}\hat{x} + \mathscr{D}\hat{\delta} \qquad \text{(A.3)}$$

is a necessary and sufficient condition for a global minimum. The matrix $\mathscr{A}^\perp$ is any matrix which columns form a basis for the range space of $\mathscr{A}^T$. Substituting (A.1) and (A.2) in (A.3), we have

$$[\mathscr{D}V_\delta\mathscr{D}^T\mathscr{A}^\perp \;\; \mathscr{A}]\begin{bmatrix}\bar{\lambda}\\ \hat{x}\end{bmatrix} = z.$$

Using $V_e = \mathscr{D}V_\delta\mathscr{D}^T$ and the assumption that $\mathscr{A}$ is full rank, the result follows. □

Theorem 4. For a full rank matrix $A \in \mathbb{R}^{m\times n}$, with m greater then n,

$$(A^TA)^{-1}A^T = [0_{n\times(m-n)} \;\; I_n]\begin{bmatrix}A^\perp & A\end{bmatrix}^{-1}, \qquad \text{(A.4)}$$

where $A^\perp$ is a matrix which columns form a basis for the orthogonal complement of the range space of A, *i.e.*, $\text{Range}(A^\perp) = \text{Null}(A^T)$, and $\text{rank}(A^\perp) = m-n$.

PROOF. In the proof, we use the SVD of the matrix A

$$A = U\Sigma V^T = [U_1 \; U_2]\begin{bmatrix}\Sigma_1\\ 0\end{bmatrix}V^T = U_1\Sigma_1 V^T.$$

Partition U as follows

$$U = [U_1 \; U_2] = \begin{matrix}\\ m-n\\ n\end{matrix}\overset{\begin{matrix}n & m-n\end{matrix}}{\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}}.$$

The matrix U_2 satisfies $\text{Range}(U_2) = \text{Null}(A^T)$ and $\text{rank}(U_2) = m-n$, so it serves as a particular $A^\perp$.

The left-hand side of the desired identity (A.4) is

$$(A^TA)^{-1}A^T = V\Sigma_1^{-1}U_1^T = V\Sigma_1^{-1}\begin{bmatrix}U_{11}^T & U_{21}^T\end{bmatrix},$$

and the right-hand side is

$$[0 \;\; I_n]\begin{bmatrix}A^\perp & A\end{bmatrix}^{-1} = [0 \;\; I_n]\begin{bmatrix}U_2 & U_1\Sigma_1V^T\end{bmatrix}^{-1} = [0 \;\; I_n]\begin{bmatrix}U_{12} & U_{11}\Sigma_1V^T\\ U_{22} & U_{21}\Sigma_1V^T\end{bmatrix}^{-1}. \qquad \text{(A.5)}$$

To find explicitly the inverse matrix in (A.5), we use the formula for inverse of a block matrix (Meyer, 2000, p.123)

$$\begin{bmatrix}B & C\\ D & E\end{bmatrix}^{-1} = \begin{bmatrix}B^{-1}+B^{-1}CS^{-1}DB^{-1} & -B^{-1}CS^{-1}\\ -S^{-1}DB^{-1} & S^{-1}\end{bmatrix},$$

where

$$S = E - DB^{-1}C$$

is the Schur complement of B in the matrix $\left[\begin{smallmatrix}B & C\\ D & E\end{smallmatrix}\right]$. For the block matrix in (A.5), the Schur complement of U_{12} is

$$\begin{aligned}S &= U_{21}\Sigma_1V^T - U_{22}U_{12}^{-1}U_{11}\Sigma_1V^T\\ &= (U_{21} - U_{22}U_{12}^{-1}U_{11})\Sigma_1V^T.\end{aligned}$$

Then

$$[0 \;\; I_n]\begin{bmatrix}A^\perp & A\end{bmatrix}^{-1} = V\Sigma_1^{-1}(U_{21} - U_{22}U_{12}^{-1}U_{11})^{-1}[U_{22}U_{12}^{-1} \;\; I]. \qquad \text{(A.6)}$$

Because of the orthogonality of U,

$$\begin{aligned}U_1^TU_2 = U_{11}^TU_{12} + U_{21}^TU_{22} &= 0\\ \Rightarrow \quad -U_{22}U_{12}^{-1} &= U_{21}^{-T}U_{11}^T. \qquad \text{(A.7)}\end{aligned}$$

Then

$$\begin{aligned}(U_{21} - U_{22}U_{12}^{-1}U_{11})^{-1} &= (U_{21} + U_{21}^{-T}U_{11}^TU_{11})^{-1}\\ &= \big(U_{21}^{-T}(\underbrace{U_{21}^TU_{21} + U_{11}^TU_{11}}_{U_1^TU_1 = I})\big)^{-1} = U_{21}^T. \qquad \text{(A.8)}\end{aligned}$$

Substitution of the expressions of (A.7) and (A.8) into (A.6) establishes the identity. □