ELSEVIER

Systems & Control Letters 32 (1997) 301–312

SYSTEMS & CONTROL LETTERS

Deterministic Kalman filtering in a behavioral framework

Fabio Fagnani[a,*], Jan C.Willems[b]

[a] *Scuola Normale Superiore, Piazza dei Cavalieri 7, 56100 Pisa, Italy*
[b] *Institute for Mathematics and Computer Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands*

Received 6 August 1996; received in revised form 23 January 1997; accepted 15 February 1997

Abstract

The purpose of this paper is to obtain a deterministic version of the Kalman filtering equations. We will use a behavioral description of the plant, specifically, an image representation. The resulting algorithm requires a matrix spectral factorization. We also show that the filter can be implemented recursively. © 1997 Elsevier Science B.V.

Keywords: Behavioral approach; Kalman filtering; Least-squares estimation; Observers; Spectral factorization

1. Introduction

A frequently encountered problem in control and signal processing is to estimate the value of one variable on the basis of another related, observed one. When the to-be-estimated and the observed variables are time signals then we can view such problems as filtering questions. Both the Wiener and the Kalman filter approach this area by casting it in a stochastic setting: it is assumed that the signals involved are realizations of stochastic processes with known statistics. The filter then produces the conditional expectation of the to-be-estimated signal given the observations.

What makes the Wiener and the Kalman filtering problem difficult, is the *real-time* aspect, the problem of having to obtain an estimate of the present value of the to-be-estimated signal from only the past of the observations. The algorithm underpinning the Kalman filter has, moreover, the important advantage that this real-time feature is implemented in a *recursive* fashion, in the sense that, by obtaining a state space representation of the filter, the estimates can be automatically updated as new measurements become available.

The purpose of the present paper is to treat the filtering problem, as we have just defined it, in a deterministic setting. There is, of course, a well-known deterministic interpretation of the filtering problem, being that the optimal filter minimizes the H_2-norm of the transfer function from the driving variables to the estimation error over all non-anticipating linear filters. Our approach is a different one and will be more akin to the formulation of the filtering problem given above. Other works which, though oriented towards a different goal, discuss deterministic Kalman filtering are [2, 3].

* Corresponding author.

PII S0167-6911(97)00086-8

2. Problem formulation

Our formulation of deterministic Kalman filter is rather subtle. It goes as follows. Assume that we observe a vector time signal $y:\mathbb{R}\to\mathbb{R}^q$. We would like to deduce from it a to-be-estimated vector time signal z. The estimate at time t, $\hat{z}(t)$, is allowed to depend only on the past of y. Assume that we have a procedure to find the related vector signal $\hat{z}$ which optimally explains observed signal y up to time t. This yields a map from the past of y to $\hat{z}$. In particular, it yields a map from the past of y at time t to $\hat{z}(t)$. This way we will obtain a mapping $\mathscr{E}$ (for '*estimate*'), $\mathscr{E}: y \mapsto \hat{z}$, having the property that $\hat{z}(t) = \mathscr{E}(y)(t)$. It is this mapping $\mathscr{E}$ that we are looking for.

In order to set up a rationale for constructing $\mathscr{E}$ we have to assume that the observed signal y and the to-be-estimated signal z are somehow related. We do this by postulating that this relationship is formalized by a linear time-invariant differential system with behavior $\mathscr{B}$. Now, consider two situations. First, assume that the observations y are compatible with $\mathscr{B}$, i.e., that $\mathscr{B}$ could actually explain the observations exactly. We will refer to this case as *exact observations* (granted, a bit of a misnomer). In this case our problem reduces to *observer design*: we have to derive, on the basis of the specifications of $\mathscr{B}$, the differential equation relating y and z.

However, the problem which we like to solve is the situation in which there is simply no element in $\mathscr{B}$ which could conceivably explain the observations. This case will be referred to (misnomed as) *approximate observations*. In this case we will replace the observation y by that element $\hat{y}$ that can be explained by the behavior and which approximates y optimally in the least-squares sense. Combining this with the observer then leads to the desired estimate of z, from y, via $\hat{y}$ and the observer. However, this optimal $\hat{y}$ will in principle depend on the whole past and future of y and so this procedure does not deal with the real-time aspect of filtering. This real-time element is brought in as follows. We introduce $\mathscr{B}^t_-$, the behavior restricted to the half-line $(-\infty,t]$ and look for the element $\hat{y}_t$ in $\mathscr{B}^t_-$ which approximates $y_{|(-\infty,t]}$ optimally in the least-squares sense. The observer requires a number of derivatives of y. The filter is now obtained by replacing, at time t, these derivatives by those of $\hat{y}_t$ at time t.

3. Differential systems

We will use the behavioral approach to the theory of dynamical systems as put forward, for example, in [5]. Let $\mathscr{L}^q$ denote the family of linear time-invariant differential dynamical systems in q real variables. Thus, each element of $\mathscr{L}^q$ is a dynamical system $\Sigma=(\mathbb{R},\mathbb{R}^q,\mathscr{B})$ with $\mathscr{B}$ the solution set of a system of linear constant coefficient differential equations, say

$$R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)w=0 \tag{1}$$

with $R\in\mathbb{R}^{\bullet\times q}[\xi]$. For simplicity we will consider, throughout this paper, only C^∞ solutions. Hence, we assume that $\mathscr{B}\subseteq C^\infty(\mathbb{R},\mathbb{R}^q)$. The system of differential equations, Eq. (1), is said to be a *kernel representation* of Σ. There exist many other possible representations of Σ. In particular, if Σ is *controllable* [5], then there will exist an $M(\xi)\in\mathbb{R}^{q\times\bullet}[\xi]$ such that

$$w=M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)l \tag{2}$$

is an *image representation* of Σ, where l is some other signal variable which is usually called *latent*. Equivalently, we can write

$$\mathscr{B}=M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)C^\infty(\mathbb{R},\mathbb{R}^m), \tag{3}$$

where m denotes the number of columns of M.

The notion of *observability* refers to a situation with two types of variables, say

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{4}$$

with $w_1 \in \mathbb{R}^{q_1}$ and $w_2 \in \mathbb{R}^{q_2}$, yielding the dynamical system $(\mathbb{R}, \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}, \mathscr{B})$. We will say that w_1 is *observable* from w_2 if $(w_1', w_2) \in \mathscr{B}$ and $(w_1'', w_2) \in \mathscr{B}$ together imply $w_1' = w_1''$. It can be shown that observability is equivalent to the existence of a kernel representation of the special form

$$w_1 = R_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w_2, \tag{5}$$

$$R_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w_2 = 0. \tag{6}$$

However, when discussing systems with latent variables, as Eq. (2), we will assume that observability refers to observability of the latent variable ℓ from the manifest variable w. It can be shown, in fact, that in Eq. (2), M can always be chosen to be an observable image representation, equivalently, such that the matrix $M(\lambda) \in \mathbb{C}^{q\times\bullet}$ is of full column rank for all $\lambda \in \mathbb{C}$.

4. The model

We will assume that, as explained in Section (2), the to-be-estimated signal z and the optimal approximation $\hat{y}$ of the observation y obey the laws of a given differential system. In particular, in the case of exact observations, we will assume that (z, y) belongs to a given behavior. In the case of approximate observations, we want to find a $\hat{y}$ such that $(z, \hat{y})$ belongs to this given behavior. Assume that in this behavior z is observable from $\hat{y}$, and that the behavior is controllable. Whence the system can be represented as

$$\hat{y} = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \ell, \tag{7}$$

$$z = P\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \hat{y}, \tag{8}$$

with Eq. (7) an observable image representation. We will take Eqs. (7) and (8) as the model from which we start our analysis. Note that Eq. (8) shows how to obtain the estimate z once $\hat{y}$ is known. We will, therefore, concentrate in this paper on the question of how to compute $\hat{y}$ from the observation y.

We will first briefly consider estimating $\hat{y}$ from both the past and the future of y (referred to as *smoothing*) and then consider the case that $\hat{y}$ can only depend on the past (referred to as *filtering*).

5. Smoothing

In this section we take a look at the question how to obtain an estimate $\hat{y}$ using both the past and the future of y. Assume, for simplicity, that $y \in \mathscr{D}^q$ (:= the elements of $C^\infty(\mathbb{R}, \mathbb{R}^q)$ with compact support). Then if $\hat{y} = M(\mathrm{d}/\mathrm{d}t)\ell^*$ is least-squares optimal, the associated ℓ^* will have to satisfy

$$\int_{-\infty}^{+\infty} \left(\left\| y - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(\ell^* + \Delta) \right\|^2 - \left\| y - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell^* \right\|^2 \right) \mathrm{d}t \geqslant 0 \tag{9}$$

for all $\Delta \in \mathscr{D}^m$ (m equals again the column dimension of M). It is easy to prove that if ℓ^* obeys Eq. (9), then

$$M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} y = M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \ell^*, \tag{10}$$

where T denotes transposition. Consider Eq. (10) as an equation for $\ell^*:\mathbb{R}\to\mathbb{R}^m$ with $y\in\mathscr{D}^q$ given. It can be shown that, among the finite-dimensional linear variety of solutions to Eq. (10), there exists *exactly one* such that

$$\int_{-\infty}^{+\infty}\left(\left\|y-M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell^*\right\|^2\right)\mathrm{d}t<\infty. \tag{11}$$

This is the solution we are looking for. Actually, this solution is the unique solution to Eq. (10) which is in $\mathscr{L}_2(\mathbb{R},\mathbb{R}^m)$. Hence, the optimal $\hat{y}$ is specified by

$$M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell^*=M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}y, \tag{12}$$

$$\ell^*\in\mathscr{L}_2(\mathbb{R},\mathbb{R}^m), \tag{13}$$

$$\hat{y}=M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell^*. \tag{14}$$

Various algorithms for computing ℓ^* and $\hat{y}$ can be constructed. Some of these use the polynomial matrix factorizations which will be discussed later.

6. Filtering

It can be seen that the solution ℓ^* and $\hat{y}$ obtained from Eqs. (12)–(14) will be such that $\ell^*(t)$ and $\hat{y}(t)$ will depend on the future as well on the past of y. Hence, Eqs. (12)–(14) is an *off-line* algorithm. We will now formulate a real-time version of the problem at hand. Let $y:\mathbb{R}\to\mathbb{R}^q$ be given and assume, again for simplicity, that $y\in\mathscr{D}^q$. We will now compute, with $t\in\mathbb{R}$ a fixed element, the element $\hat{y}_t\in\mathscr{B}$ (with $\mathscr{B}$ still given in image representation by Eq. (7)) such that

$$\int_{-\infty}^{t}\|y-\hat{y}_t\|^2\,\mathrm{d}t \tag{15}$$

is minimized. Now define $y^*:\mathbb{R}\to\mathbb{R}^q$ by

$$y^*(t):=\hat{y}_t(t). \tag{16}$$

Our problem is to find the map $\mathscr{E}:y\mapsto y^*$. Obviously by its construction, $y^*(t)$ will depend on the past $y|_{(-\infty,t]}$ only. We view the map $\mathscr{E}$ as the *deterministic Kalman filter*.

7. Quadratic differential forms

Let $\Phi\in\mathbb{R}^{n_1\times n_2}[\zeta,\eta]$, i.e., Φ is a two-variable polynomial matrix, say $\Phi(\zeta,\eta)=\sum_{k,l}\Phi_{kl}\zeta^k\eta^l$ with $\Phi_{kl}\in\mathbb{R}^{n_1\times n_2}$. Associated with Φ there is the bilinear differential form

$$L_\Phi: C^\infty(\mathbb{R},\mathbb{R}^{n_1})\times C^\infty(\mathbb{R},\mathbb{R}^{n_2})\to C^\infty(\mathbb{R},\mathbb{R}) \tag{17}$$

defined by

$$L_\Phi(w_1,w_2):=\sum_{k,l}\left(\frac{\mathrm{d}^k w_1}{\mathrm{d}t^k}\right)^{\mathrm{T}}\Phi_{kl}\left(\frac{\mathrm{d}^l w_2}{\mathrm{d}t^l}\right). \tag{18}$$

If $n_1=n_2$, we also define the quadratic differential form

$$Q_\Phi: C^\infty(\mathbb{R},\mathbb{R}^{n_1})\to C^\infty(\mathbb{R},\mathbb{R}) \tag{19}$$

by

$$Q_\Phi(w) := L_\Phi(w, w). \tag{20}$$

We will make use of special quadratic differential forms which can be obtained by differentiating another quadratic differential form. It is easy to see that if $\Phi, \Psi \in \mathbb{R}^{n_1 \times n_2}[\zeta, \eta]$, then

$$Q_\Phi = \frac{\mathrm{d}}{\mathrm{d}t} Q_\Psi \tag{21}$$

if and only if

$$\Phi(\zeta, \eta) = (\zeta + \eta)\Psi(\zeta, \eta). \tag{22}$$

This shows that Φ is such that Q_Φ is the derivative of another quadratic differential form if and only if

$$\Phi(-\xi, \xi) = 0. \tag{23}$$

This is equivalent to requiring that

$$\frac{\Phi(\zeta, \eta)}{\zeta + \eta} \tag{24}$$

is again a two-variable polynomial matrix. More details on this sort of use of two-variable polynomial matrices may be found in [6]. We recall the following result proved in [6, 7].

Lemma 1. *Let $\Sigma = (\mathbb{R}, \mathbb{R}^q, \mathscr{B}) \in \mathscr{L}^q$ and let $\langle , \rangle$ be a bilinear form on $\mathbb{R}^q$. Let $\Phi \in \mathbb{R}^{q \times q}[\zeta, \eta]$. Assume that*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_\Phi(w) \leqslant \langle w, w \rangle \quad \forall w \in \mathscr{B}. \tag{25}$$

Let now $X(\mathrm{d}/\mathrm{d}t)w$ be a minimal state map [6] for Σ. Then there exists a symmetric matrix $S \in \mathbb{R}^{q \times q}$ such that

$$Q_\Phi(w) = \left(X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w\right)^{\mathrm{T}} S X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w. \tag{26}$$

8. Polynomial factorization

Let $M \in \mathbb{R}^{q \times m}[\xi]$ be such that $M(\lambda)$ is full column rank for all $\lambda \in \mathbb{C}$ (observability of Eq. (7)). Assume also that M has the property that it can be partitioned as

$$M = \begin{bmatrix} N \\ D \end{bmatrix}, \tag{27}$$

with D square, ND^{-1} a matrix of proper rational functions, and D^{-1} a matrix of strictly proper rational functions. It is easy to see that, up to permutation of its rows, any M can be partitioned such that ND^{-1} is proper. The assumption that D^{-1} is strictly proper is a restriction. We will not dwell on the significance or the reason of this assumption.

Consider the polynomial matrix $\Phi \in \mathbb{R}^{m \times m}[\xi]$ defined by $\Phi(\xi) := M(-\xi)^{\mathrm{T}} M(\xi)$. Then, of course, $\Phi(\mathrm{i}\omega) > 0$ for all $\omega \in \mathbb{R}$. This implies that the matrix factorization equation

$$\Phi(\xi) = F(\xi) F(-\xi)^{\mathrm{T}} \tag{28}$$

has a solution $F \in \mathbb{R}^{m\times m}[\xi]$. In fact, it has many solutions. Among these, there will be one such that F is a Hurwitz polynomial matrix (meaning that $\det F$ has its roots in the left half of the complex plane). We will denote this solution by H. Hence, $H \in \mathbb{R}^{m\times m}[\xi]$ is a Hurwitz polynomial matrix that satisfies

$$M(-\xi)^{\mathrm{T}}M(\xi) = H(\xi)H(-\xi)^{\mathrm{T}}. \tag{29}$$

Using the results of Section (7), it follows that

$$\Psi(\zeta,\eta) := \frac{M(\zeta)^{\mathrm{T}}M(\eta) - H(-\zeta)H(-\eta)^{\mathrm{T}}}{\zeta+\eta} \tag{30}$$

is a two-variable polynomial matrix. There holds

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_\Psi(\ell) = \left\| M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\|^2 - \left\| H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell\right\|^2. \tag{31}$$

We now want to study in more detail the structure of this quadratic differential form Q_Ψ. First we need a preliminary result:

Lemma 2. *$M(\xi)H(-\xi)^{-\mathrm{T}}$ and $H(-\xi)^{-\mathrm{T}}$ are both proper rational matrices ($^{-\mathrm{T}}$ means inverse of the transpose).*

Proof. The fact that $M(\xi)H(-\xi)^{-\mathrm{T}}$ is proper is an immediate consequence of Eq. (29). Consider now the partition, Eq. (27), and put $G(\xi) := N(\xi)D(\xi)^{-1}$. It follows from Eq. (29) that

$$H(-\xi)^{-\mathrm{T}}H(\xi)^{-1} = D(\xi)^{-1}[I + G(-\xi)^{\mathrm{T}}G(\xi)]^{-1}D(-\xi)^{-\mathrm{T}}. \tag{32}$$

Since G is proper, for $\xi \to +\infty$, $G(\xi)$ will converge to $G(\infty)$. Since $I + G(\infty)^{\mathrm{T}}G(\infty)$ is obviously invertible, this shows that $[I + G(-\xi)^{\mathrm{T}}G(\xi)]^{-1}$ is proper. Since $D(\xi)^{-1}$ is also proper, we have the result. □

Before stating the next lemma, we need to introduce few more concepts which have been introduced in [1]. First define the so-called 'shift and cut' operator

$$\sigma_+ : \mathbb{R}[\xi] \to \mathbb{R}[\xi], \tag{33}$$

$$\sigma_+ p := \xi^{-1}(p(\xi) - p(0)). \tag{34}$$

Of course, we can extend the action of σ_+ to vectors of polynomials by making it act on each component. Consider

$$M(\xi) := \begin{bmatrix} M_1(\xi) \\ \vdots \\ M_g(\xi) \end{bmatrix}. \tag{35}$$

We define $\Xi_{M(\xi)}$ as the vector space spanned by $\{\sigma_+^k M_i \mid k \in \mathbb{N},\ i = 1,\dots,g\}$. It is proved in [1] that if $X(\xi)$ is a matrix whose rows form a basis of $\Xi_{M(\xi)}$, then $X(\mathrm{d}/\mathrm{d}t)\ell$ is a minimal state map for the system

$$y = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \tag{36}$$

Lemma 3.

$$\Xi_{M(\xi)} = \Xi_{H(-\xi)^{\mathrm{T}}}. \tag{37}$$

Proof. Consider the partition, Eq. (27). Since $Y(\xi)D(\xi)^{-1}$ and $D(\xi)^{-1}$ are both proper, it follows that

$$\Xi_{M(\xi)} = \Xi_{D(\xi)} = \{r(\xi) \in \mathbb{R}[\xi]^{1\times p} \mid r(\xi)D(\xi)^{-1} \text{ is strictly proper}\}. \tag{38}$$

Similarly, since $H(-\xi)^{-\mathrm{T}}$ is proper,

$$\Xi_{H(-\xi)^{\mathrm{T}}} = \{r(\xi) \in \mathbb{R}[\xi]^{1\times p} \mid r(\xi)H(-\xi)^{-\mathrm{T}} \text{ is strictly proper}\}. \tag{39}$$

Now, since $M(\xi)H(-\xi)^{-\mathrm{T}}$ is proper, it follows that $D(\xi)H(-\xi)^{-\mathrm{T}}$ is proper. This immediately yields $\Xi_{M(\xi)} \subseteq \Xi_{H(-\xi)^{\mathrm{T}}}$. On the other hand, from Eq. (32) we easily see that $H(-\xi)^{\mathrm{T}}D(\xi)^{-1}$ is also proper. Hence, $\Xi_{M(\xi)} \supseteq \Xi_{H(-\xi)^{\mathrm{T}}}$. Therefore, they must be equal and the result is proved. □

Consider the behavior $\mathscr{B}$ described by the relation

$$y = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell. \tag{40}$$

We have the following:

Proposition 4. *There exists a minimal state map $X(\mathrm{d}/\mathrm{d}t)\ell$ for $\mathscr{B}$ such that*

$$\Psi(\zeta,\eta) = X(\zeta)^{\mathrm{T}}X(\eta). \tag{41}$$

Proof. It follows from Eq. (31) that Q_Ψ is a storage function [6] for $\mathscr{B}$. Fix a minimal state map $X(\mathrm{d}/\mathrm{d}t)\ell$ for $\mathscr{B}$. Hence, by Lemma 1 there exists a symmetric matrix S such that

$$\Psi(\zeta,\eta) = X(\zeta)^{\mathrm{T}}SX(\eta). \tag{42}$$

We now prove that $S > 0$. Eq. (41) then follows by taking as new state map $\sqrt{S}X(\mathrm{d}/\mathrm{d}t)\ell$.

Consider the behavior $\mathscr{B}'$ described by the equations

$$y = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell, \qquad f = H^{\mathrm{T}}\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell. \tag{43}$$

It follows from Lemma 3 that $X(\mathrm{d}/\mathrm{d}t)\ell$ is also a minimal state map for $\mathscr{B}'$. Notice now that f is the input in $\mathscr{B}'$. Indeed,

$$\begin{pmatrix} I & M(\xi) \\ 0 & H(-\xi)^{\mathrm{T}} \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ I \end{pmatrix} = \begin{pmatrix} I & -M(\xi)H(-\xi)^{\mathrm{T}} \\ 0 & H(-\xi)^{-\mathrm{T}} \end{pmatrix}\begin{pmatrix} 0 \\ I \end{pmatrix} = \begin{pmatrix} -M(\xi)H(-\xi)^{\mathrm{T}} \\ H(-\xi)^{-\mathrm{T}} \end{pmatrix}, \tag{44}$$

which is proper by Lemma 2. Fix now $\ell \in C^\infty(\mathbb{R},\mathbb{R}^m)$ and let $x_0 := X(\mathrm{d}/\mathrm{d}t)\ell(0)$. It follows that there must exist an $\ell_1 \in C^\infty(\mathbb{R},\mathbb{R}^m)$ such that

$$H^{\mathrm{T}}\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell_1 = 0, \qquad X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell_1(0) = x_0. \tag{45}$$

Since $H(-\xi)^{\mathrm{T}}$ is anti-Hurwitz, it follows that $\ell_{1|(-\infty,0]}$ and all its derivatives are in $\mathscr{L}_2$. Hence,

$$Q_\Psi(\ell)(0) = Q_\Psi(\ell_1)(0) = \int_{-\infty}^{0}\left\|M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\|^2 \mathrm{d}t \geqslant 0. \tag{46}$$

Notice, moreover, that

$$Q_\Psi(\ell)(0) = 0 \Leftrightarrow \ell_1 = 0 \Leftrightarrow X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell = 0. \tag{47}$$

This clearly yields $S > 0$. The result is thus proved. □

Next, consider two further two-variable polynomial matrices:

$$\Theta(\zeta,\eta) := \frac{M(-\zeta) - M(\eta)}{\zeta + \eta}, \tag{48}$$

$$\Phi(\zeta,\eta) := \frac{-H^{\mathrm{T}}(\zeta) + H^{\mathrm{T}}(-\eta)}{\zeta + \eta}. \tag{49}$$

Of course, just as X, both Θ and Φ are immediately related to M and H. This has the following consequence:

Proposition 5. *There exist polynomial matrices $Y \in \mathbb{R}^{\bullet \times q}[\xi]$ and $F \in \mathbb{R}^{\bullet \times m}[\xi]$ such that*

$$\Theta(\zeta,\eta) = Y(\zeta)^{\mathrm{T}} X(\eta) + X(\zeta)^{\mathrm{T}} Y(\eta), \tag{50}$$

$$\Phi(\zeta,\eta) = F(\zeta)^{\mathrm{T}} X(\eta) + X(\zeta)^{\mathrm{T}} F(\eta). \tag{51}$$

Proof. It follows from the definition of Θ, Eq. (48), that

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_{\Theta}\begin{pmatrix} \ell \\ \tilde{\ell} \end{pmatrix} = \left\langle \ell, M^{\mathrm{T}}\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)\tilde{\ell} \right\rangle - \left\langle \tilde{\ell}, M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\rangle. \tag{52}$$

Denote by $\tilde{\mathscr{B}}$ the behavior described by the image representation

$$\tilde{y} = M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} \tilde{\ell} \tag{53}$$

and let $\tilde{X}(\mathrm{d}/\mathrm{d}t)\tilde{\ell}$ be a minimal state map for $\tilde{\mathscr{B}}$. Hence,

$$\begin{bmatrix} X(\mathrm{d}/\mathrm{d}t)\ell \\ \tilde{X}(\mathrm{d}/\mathrm{d}t)\tilde{\ell} \end{bmatrix} \tag{54}$$

is a minimal state map for the direct sum system $\mathscr{B} \oplus \tilde{\mathscr{B}}$. It follows from the definition of Θ, Eq. (48), that

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_{\Theta}\begin{pmatrix} \ell \\ \tilde{\ell} \end{pmatrix} = \left\langle \ell, M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\tilde{\ell} \right\rangle - \left\langle \tilde{\ell}, M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\rangle. \tag{55}$$

By Lemma 1 that there exist a real matrix

$$S := \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^{\mathrm{T}} & S_{22} \end{pmatrix} \tag{56}$$

with $S_{11}^{\mathrm{T}} = S_{11}$ and $S_{22}^{\mathrm{T}} = S_{22}$, such that

$$\Theta(\zeta,\eta) = \begin{bmatrix} X(\zeta) \\ \tilde{X}(\zeta) \end{bmatrix}^{\mathrm{T}} \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^{\mathrm{T}} & S_{22} \end{pmatrix} \begin{bmatrix} X(\eta) \\ \tilde{X}(\eta) \end{bmatrix}. \tag{57}$$

We now show that S_{11} and S_{22} must both be equal to 0. Indeed, assume that ℓ and $\tilde{\ell}$ have compact support. By integrating Eq. (55), we obtain

$$Q_{\Theta}\begin{pmatrix} \ell \\ \tilde{\ell} \end{pmatrix}(0) = \int_{-\infty}^{t} \left\langle \ell, M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\tilde{\ell} \right\rangle(s)\,\mathrm{d}s - \int_{-\infty}^{t} \left\langle \tilde{\ell}, M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\rangle(s)\,\mathrm{d}s, \tag{58}$$

which shows that

$$\left(X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell(0)\right)^{\mathrm{T}} S_{11}X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell(0) = Q_{\Theta}\begin{pmatrix}\ell\\0\end{pmatrix}(0) = 0. \tag{59}$$

On the other hand, since $\mathscr{B}$ is controllable, for every x_0 there exists $\ell \in C^{\infty}$ with compact support such that $X(\mathrm{d}/\mathrm{d}t)\ell(0) = x_0$. Together with Eq. (59) this implies that $S_{11} = 0$. Analogously we can prove that $S_{22} = 0$. This yields

$$\Theta(\zeta,\eta) = \tilde{X}(\zeta)^{\mathrm{T}} S_{12}^{\mathrm{T}} X(\eta) + X(\zeta)^{\mathrm{T}} S_{12}\tilde{X}(\eta). \tag{60}$$

Hence, Eq. (50) holds with $Y(\zeta) := S_{12}\tilde{X}(\zeta)$. Eq. (51) can be proved in exactly the same way, using Lemma 3. □

Relations of Eqs. (41), (50) and (51) imply

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\|X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\|^2 = \left\|M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\|^2 - \left\|H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell\right\|^2, \tag{61}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y, X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\rangle = \left\langle M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}y, \ell\right\rangle - \left\langle y, M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\rangle, \tag{62}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f, X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\rangle = -\left\langle H\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f, \ell\right\rangle + \left\langle f, H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell\right\rangle. \tag{63}$$

9. Basic identity

We will now derive the basic identity which will let us obtain our deterministic version of the Kalman filter.

The observation $y \in \mathscr{D}^q$ is assumed to be given. Let H be the polynomial matrix obtained in Eq. (29). Next, define $f \in C^{\infty}(\mathbb{R}, \mathbb{R}^m)$ as the unique $\mathscr{L}_2(\mathbb{R}, \mathbb{R}^m)$ solution of

$$H\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f = M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} y. \tag{64}$$

Note that f is well defined because H is assumed to be Hurwitz and that it will have left compact support. Henceforth, we can thus assume that both y and f are given.

Denote by Λ_t the subset of $C^{\infty}((-\infty, t], \mathbb{R}^m)$ consisting of the functions which are of class $\mathscr{L}_2$ together with all their derivatives.

Theorem 6.

$$\int_{-\infty}^{t}\left\|y - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\|^2 \mathrm{d}t \tag{65}$$

assumes its minimum value over Λ_t in the unique point ℓ_t which satisfies the following conditions:

$$f = H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell_t \quad on\ (-\infty, t], \tag{66}$$

$$\left(X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell_t\right)(t) + \left(Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y\right)(t) + \left(F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f\right)(t) = 0. \tag{67}$$

Proof. A straightforward calculation yields the following identity:

$$\left\| y - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\|^2 = \left\| f - H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell \right\|^2 + \frac{\mathrm{d}}{\mathrm{d}t}\left\| X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell + Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f \right\|^2$$
$$+(\|y\|^2 - \|f\|^2) - \frac{\mathrm{d}}{\mathrm{d}t}\left\| Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f \right\|^2. \tag{68}$$

If $\ell \in \Lambda_t$, integrating Eq. (68) from $-\infty$ to t, yields

$$\int_{-\infty}^{t} \left\| y - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\|^2 \mathrm{d}t$$
$$= \int_{-\infty}^{t} \left\| f - H^{\mathrm{T}}\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \right\|^2 \mathrm{d}t + \left\| X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell + Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f \right\|^2(t)$$
$$+ \int_{-\infty}^{t} (\|y\|^2 - \|f\|^2)\,\mathrm{d}t - \left\| Y\left(\frac{\mathrm{d}}{-\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f \right\|^2(t). \tag{69}$$

We are trying to find an ℓ which minimizes the left-hand side of the above expression with y given. Recall that we may also consider that f is given by Eq. (64). Now, the third and fourth term on the right-hand side of Eq. (69) depend on y and f only. Therefore, they cannot be changed by choosing ℓ. Hence, if there exists an ℓ_t satisfying conditions Eqs. (66) and (67), it will clearly minimize Eq. (65). The fact that there exists a unique ℓ_t which satisfies conditions Eqs. (66) and (67) is a straightforward consequence of the fact, that $X(\mathrm{d}/\mathrm{d}t)\ell$ is a minimal state for the dynamical system

$$f = H^{\mathrm{T}}\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \tag{70}$$

with input f (see Lemma 3). □

Remark. It may seem not natural to minimize Eq. (65) over Λ_t since ℓ is only latent variable, while the external variable is $M(\mathrm{d}/\mathrm{d}t)\ell$. It would be thus more natural to impose that only $M(\mathrm{d}/\mathrm{d}t)\ell$ is in $\mathscr{L}_2$. However, in [4] it is proven that $M(\mathrm{d}/\mathrm{d}t)\ell \in \mathscr{L}_2$ automatically implies that ℓ and $X(\mathrm{d}/\mathrm{d}t)\ell$ are also in $\mathscr{L}_2$. Thus, Eq. (69) will remain valid if only $M(\mathrm{d}/\mathrm{d}t)\ell$ is assumed to be in $\mathscr{L}_2$. From this, straightforward considerations would show that the unique minimum is again given by the ℓ_t found above.

The optimal estimation of y at time t is thus given by

$$y^*(t) = \hat{y}_t(t) = \left(M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell_t\right)(t). \tag{71}$$

10. Recursive implementation

Similarly as in the stochastic Kalman filter, it is possible to implement the deterministic Kalman filter in a recursive way. The key fact is the following:

Proposition 7. *Let Y and F be polynomial matrices satisfying Eqs.* (50) *and* (51). *Then*

$$-Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y - F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f \tag{72}$$

is a minimal state map for the system, Eq. (64).

Proof. Consider

$$\Omega(\zeta,\eta) := X(\zeta)^{\mathrm{T}}[Y(\eta),F(\eta)] + [Y(\zeta),F(\zeta)]^{\mathrm{T}}X(\eta). \tag{73}$$

It follows from the definition of Y and F that

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_\Omega\begin{pmatrix}\ell\\ y\\ f\end{pmatrix} = \left\langle \ell, M\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} y\right\rangle - \left\langle y, M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\rangle + \left\langle f, H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell\right\rangle - \left\langle \ell, H\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f\right\rangle. \tag{74}$$

If $\tilde{X}(\mathrm{d}/\mathrm{d}t)(y,f)^{\mathrm{T}}$ is a minimal state map for the system, Eq. (64), it follows from Lemma 1 that there exists a symmetric matrix

$$S = \begin{pmatrix} S_{11} & S_{12}\\ S_{12}^{\mathrm{T}} & S_{22}\end{pmatrix} \tag{75}$$

such that

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_\Omega\begin{pmatrix}\ell\\ y\\ f\end{pmatrix} = \begin{pmatrix} (X(\mathrm{d}/\mathrm{d}t)\ell\\ \tilde{X}(\mathrm{d}/\mathrm{d}t)(y,f)^{\mathrm{T}})\end{pmatrix}^{\mathrm{T}} \begin{pmatrix} S_{11} & S_{12}\\ S_{12}^{\mathrm{T}} & S_{22}\end{pmatrix} \begin{pmatrix} (X(\mathrm{d}/\mathrm{d}t)\ell\\ \tilde{X}(\mathrm{d}/\mathrm{d}t)(y,f)^{\mathrm{T}})\end{pmatrix}. \tag{76}$$

By integrating along compact support trajectories, similar considerations as in the proof of Proposition 5, show that S_{11} and S_{22} must be equal to 0 and, moreover, that

$$Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f = S_{12}\tilde{X}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(y,f)^{\mathrm{T}}. \tag{77}$$

We now complete the proof by showing that S_{12} is square and invertible. It follows from Lemma 2 that the Mc-Millan degree of system, Eq. (64), is equal to the degree of $\det H(\xi)$ which is obviously equal to the degree of $\det H(-\xi)$. This last is, again by Lemma 2, equal to the Mc-Millan degree of the system

$$f = H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}}\ell. \tag{78}$$

By Lemma 3, $X(\mathrm{d}/\mathrm{d}t)\ell$ is a minimal state map for this system. It follows that S_{12} must indeed be a square matrix. Assume now that

$$\left(Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y + F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f\right)(0) = 0 \tag{79}$$

and let

$$\tilde{x}_0 := \tilde{X}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(y,f)^{\mathrm{T}}. \tag{80}$$

Since y acts as an input in system, Eq. (64), it follows that there exists f such that

$$H\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f = 0, \tag{81}$$

$$\tilde{X}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(0,f)^{\mathrm{T}}(0) = \tilde{x}_0. \tag{82}$$

Take now ℓ compact support and integrate Eq. (74). We obtain

$$\int_{-\infty}^{t} \left\langle f, H\left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{\mathrm{T}} \ell \right\rangle \mathrm{d}s = 0. \tag{83}$$

Since as ℓ varies among the maps with compact support the image $H(-\mathrm{d}/\mathrm{d}t)^{\mathrm{T}}\ell$ covers all the compact support maps, it follows from Eq. (83) that $f = 0$. Hence, $\tilde{x}_0 = 0$ and this proves that S_{12} is non-singular. This proves the result. □

Proposition 7, combined with the fact that y acts as input to Eq. (64), implies that there exist matrices A and B of suitable dimensions such that

$$\hat{x}(t) := \left(X\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell_t\right)(t) = \left(-Y\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)y - F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f\right)(t) \tag{84}$$

is governed by

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{x} = A\hat{x} + By. \tag{85}$$

Moreover, there exist matrices G_1 and L_1 such that

$$f = G_1\hat{x} + L_1 y. \tag{86}$$

On the other hand, since $X(\mathrm{d}/\mathrm{d}t)\ell$ is a state map for Eq. (70) and $M(\xi)H(-\xi)^{-\mathrm{T}}$ is a matrix of proper rational functions (see Eq. (29)), there exist matrices G_2 and L_2 such that

$$y^* = G_2\hat{x} + L_2 f. \tag{87}$$

Combined with Eq. (86) this yields the existence of matrices G and L such that

$$y^* = G\hat{x} + Ly. \tag{88}$$

Eqs. (85) and (88) yields a recursive implementation of the deterministic Kalman filter.

11. Summary

Let us close by summarizing our deterministic Kalman filtering algorithm.

1. The model is specified by the polynomial matrix M (see Eq. (7)).
2. The observations consist of a vector time signal y.
3. The factorization equation, Eq. (29), yields a Hurwitz polynomial matrix H.
4. Equations (41), (50) and (51) deliver the polynomial matrices X, Y, F.
5. The filter is determined by Eqs. (64), (66), (67) and (71).
6. The filter can be implemented recursively through Eqs. (85) and (88).

References

[1] P. Rapisarda, J.C. Willems, State maps for linear systems, SIAM J. Control Opt. 35 (1997) 254–264.
[2] B. Roorda, Algorithms for global total least-squares modelling of finite multivariable time series, Automatica 31 (3) (1995) 391–404.
[3] B. Roorda, C. Heij, Global total least-squares modeling of multivariable time series, IEEE Trans. Automat. Control 40 (1995) 50–63.
[4] S. Weiland, Ph.D. Dissertation, University of Groningen, 1991.
[5] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, IEEE Trans. Automat. Control 36 (1991) 259–294.
[6] J.C. Willems, H.L. Trentelman, On quadratic differential forms, SIAM J. Control Opt., to appear.
[7] H.L. Trentelman, J.C. Willems, Every storage function is a state function, System Control Lett. 32 (1997) 249–259, this issue.