Int J Geomath manuscript No.
(will be inserted by the editor)

A recursive linear MMSE filter for dynamic systems with unknown state vector means

Amir Khodabandeh · **Peter J.G. Teunissen**

Received: date / Accepted: date

Abstract In this contribution we extend Kalman-filter theory by introducing a new recursive linear Minimum Mean Squared Error (MMSE) filter for dynamic systems with unknown state-vector means. The recursive filter enables the joint MMSE prediction and estimation of the random state vectors and their unknown means, respectively. We show how the new filter reduces to the Kalman-filter in case the state-vector means are known and we discuss the fundamentally different roles played by the intitialization of the two filters.

Keywords Minimum mean squared error (MMSE), Best linear unbiased estimation (BLUE), Best linear unbiased prediction (BLUP), Kalman filter, BLUE-BLUP recursion

1 Introduction

The minimum mean squared error (MMSE) criterion is a popular criterion for determining estimators and predictors. Depending on the class of functions considered, different MMSE predictors exist. The conditional mean achieves the smallest MSE and is therefore the best predictor (BP) of all. Within the class of linear functions however, it is the best linear predictor (BLP) that achieves the smallest MSE.

The Kalman filter is a recursive MMSE filter, which has found a widespread usage in various Earth science disciplines (Grafarend, 1976; Sanso, 1980; Bertino et al, 2002; Marx and Potthast, 2012). It is used, for example, in deformation and

A. Khodabandeh
GNSS Research Centre, Department of Spatial Sciences, Curtin University of Technology, Perth, Australia
E-mail: amir.khodabandeh@curtin.edu.au

P.J.G. Teunissen
GNSS Research Centre, Department of Spatial Sciences, Curtin University of Technology, Perth, Australia; Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands
E-mail: p.teunissen@curtin.edu.au

earth-orientation studies (Gross et al, 1998; Ince and Sahin, 2000), in physical and space geodesy (Grafarend and Rapp, 1984; Sanso, 1986; Herring et al, 1990), and in hydrology and atmospheric studies (Ferraresi et al, 1996; Cao et al, 2006; Acharya et al, 2011).

In the literature, the recursive Kalman-filter is derived as either a BP or a BLP, see e.g., Kalman (1960); Gelb (1974); Kailath (1981); Candy (1986); Brammer and Siffling (1989); Jazwinski (1991); Gibbs (2011). Both these predictors however, require the mean of the to-be-predicted random vector to be known. This is why in the derivation of the Kalman filter the mean of the random initial state-vector is assumed known, see e.g., Sorenson (1966, p. 222), Kailath (1974, p. 148), Maybeck (1979, p. 204), Anderson and Moore (1979, p. 15), Stark and Woods (1986, p. 393), Bar-Shalom and Li (1993, p. 209), Kailath et al (2000, p. 311), Christensen (2001, p. 261), Simon (2006, p. 125), Grewal and Andrews (2008, p. 138). Hence, the BP, the BLP, nor the Kalman filter, are applicable in case the mean of the random state vector is unknown.

As shown in Teunissen and Khodabandeh (2013), one can do away with this need to have the means known. In this contribution we build on that fact and develop from first principles the recursive linear MMSE filter for dynamic systems with unknown state vector means. This filter generalizes standard Kalman filter theory and it enables the joint recursive prediction and estimation of the random state vector and its unknown mean, respectively. In the standard Kalman filter set-up, with known state-vector means, this difference between estimation and prediction does not occur since one is then only left with predicting the outcomes of the random state vectors. The generalized filter links BLUE-BLUP with BLP and shows how the outcomes of the BLUE-BLUP recursions can be directly used in tandem to obtain those of the standard Kalman filter as special case.

This contribution is organized as follows. In Sect. 2, we briefly review the necessary ingredients of prediction and estimation for linear models. We use the misclosure vector of the linear model as an ancillary statistic to give a useful joint representation for the best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP). This representation is used in Sect. 3 to derive our recursive linear MMSE filter for dynamic models with unknown state vector means. In Sect. 4 we show how this generalized filter specializes to that of the Kalman filter in case the state-vector means are known. It demonstrates how the different recursions are related and interacting, and in what way their quality descriptions differ. We discuss the role of system noise and that of the error-covariance matrices in the generalized filter. Hereby we also discuss the fundamentally different role played by the initialization of the two filters.

Throughout this contribution, the estimator and the predictor are distinguished by the $\hat{.}$-symbol and $\check{.}$-symbol, respectively, while the joint estimator-predictor is denoted by using the $\tilde{.}$-symbol. Random variates are indicated by an underscore. Thus $\underline{x}$ is random, while x is not. $\mathsf{E}(.)$, $\mathsf{C}(.,.)$ and $\mathsf{D}(.)$ denote the expectation, covariance and dispersion operators, respectively. Thus $\mathsf{E}([\underline{x}-\mathsf{E}(\underline{x})][\underline{x}-\mathsf{E}(\underline{x})]^T) = \mathsf{C}(\underline{x},\underline{x}) = \mathsf{D}(\underline{x})$. The norm of a vector is denoted as $||.||$. Thus $||.||^2 = (.)^T(.)$.

2 Estimation and Prediction in linear models

2.1 Linear unbiased statistics

Consider the linear model

$$\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix} = \begin{bmatrix} A \\ A_z \end{bmatrix} x + \begin{bmatrix} \underline{e} \\ \underline{e}_z \end{bmatrix} \tag{1}$$

with known matrices $A \in \mathbb{R}^{m\times n}$, $A_z \in \mathbb{R}^{k\times n}$, zero-mean $\mathsf{E}([\underline{e}^T \underline{e}_z^T]^T) = 0$, and known dispersion

$$\mathsf{D}(\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}) = \begin{bmatrix} Q_{yy} & Q_{yz} \\ Q_{zy} & Q_{zz} \end{bmatrix} \tag{2}$$

It is assumed that rank$A = n$, Q_{yy} is positive definite and the nonrandom vector $x \in \mathbb{R}^n$ is unknown.

It is our aim to use a linear unbiased statistic of $\underline{y}$ to *estimate* the unknown mean $\bar{z} = A_z x$ and to *predict* the outcome of $\underline{z} = \bar{z} + \underline{e}_z$. In order to perform the estimation and prediction jointly, we define the target vector $\underline{\mathscr{Z}} = [\bar{z}^T, \underline{z}^T]^T$.

Let $\mathscr{G}(\underline{y}) = F\underline{y} + f$ and $\mathscr{G}_J(\underline{y}) = F_J\underline{y} + f_J$ be two arbitrary linear unbiased statistics for $\underline{\mathscr{Z}}$. Then it follows from the condition of unbiasedness that the expectation of their difference satisfies $\mathsf{E}(\mathscr{G}_J(\underline{y}) - \mathscr{G}(\underline{y})) = (F_J - F)Ax + (f_J - f) = 0$ for all x. Hence,

$$F_J = F + JB^T, \quad \text{and} \quad f_J = f \tag{3}$$

for some matrix $J \in \mathbb{R}^{2k\times(m-n)}$, where B is an $m\times(m-n)$ basis matrix of the orthogonal complement of the range space of A, $B^TA = 0$, or equivalently, B is a basis matrix of the null space of A^T. Using the above representation, we arrive at the following lemma.

Lemma 1 *Let the misclosure of $\underline{y}$ be given as $\underline{v} = B^T\underline{y}$, with B a basis matrix of the null space of A^T. Then any two linear unbiased statistics $\mathscr{G}_J(\underline{y})$ and $\mathscr{G}(\underline{y})$ for $\underline{\mathscr{Z}}$, are related as*

$$\mathscr{G}_J(\underline{y}) = \mathscr{G}(\underline{y}) + J\underline{v}, \quad \textit{for some } J \in \mathbb{R}^{2k\times(m-n)} \tag{4}$$

This lemma shows that any two linear unbiased statistics differ only by a linear function of the random misclosure vector $\underline{v}$.

2.2 MMSE-Estimator and Predictor

We now use representation (4) to establish the connection between any arbitrary linear unbiased statistic and the one achieving the minimum mean squared error (MMSE). The error vector $\underline{\varepsilon}_J = \underline{\mathscr{Z}} - \mathscr{G}_J(\underline{y})$, of which the squared norm is to be minimized, consists of the *estimation error* as well as the *prediction error*. One may then, through the choice of matrix $J \in \mathbb{R}^{2k\times r}$, minimize the mean squared norm of the error vector $\underline{\varepsilon}_J$ to obtain the joint MMSE estimator/predictor $\underline{\tilde{\mathscr{Z}}} = [\underline{\hat{z}}^T, \underline{\check{z}}^T]^T$. Recall that its two components, $\underline{\hat{z}}$ and $\underline{\check{z}}$, respectively, are referred to as the best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP), see e.g.,Goldberger (1962); Anderson and Moore (1979); Stark and Woods (1986); Simon (2006); Teunissen (2007). The idea is finalized in the following theorem.

Theorem 1 *Let $\mathscr{G}(\underline{y})$ be an arbitrary linear unbiased statistic for $\underline{\mathscr{Z}}$. Then the joint BLUE-BLUP of $\underline{\mathscr{Z}}$ can be computed as*

$$\underline{\tilde{\mathscr{Z}}} = \mathscr{G}(\underline{y}) + Q_{\varepsilon v} Q_{vv}^{-1} \underline{v} \tag{5}$$

with $\underline{\varepsilon} = \underline{\mathscr{Z}} - \mathscr{G}(\underline{y})$.

Proof Since $\mathscr{G}_J(\underline{y}) = \mathscr{G}(\underline{y}) + J\underline{v}$ for some $J \in \mathbb{R}^{2k\times(m-n)}$, then $\mathsf{E}||\underline{\varepsilon}_J||^2 = \mathsf{E}||\underline{\varepsilon} - J\underline{v}||^2$. This can be further decomposed as

$$\begin{aligned}\mathsf{E}||\underline{\varepsilon} - J\underline{v}||^2 &= \mathsf{E}||\underline{\varepsilon} - Q_{\varepsilon v}Q_{vv}^{-1}\underline{v} - (J - Q_{\varepsilon v}Q_{vv}^{-1})\underline{v}||^2 \\ &= \mathsf{E}||\underline{\varepsilon} - Q_{\varepsilon v}Q_{vv}^{-1}\underline{v}||^2 + \mathsf{E}||(J - Q_{\varepsilon v}Q_{vv}^{-1})\underline{v}||^2\end{aligned}$$

since $\underline{\varepsilon} - Q_{\varepsilon v}Q_{vv}^{-1}\underline{v}$ is uncorrelated with $\underline{v}$. By setting $J - Q_{\varepsilon v}Q_{vv}^{-1} = 0$, $\mathsf{E}||\underline{\varepsilon}_J||^2$ attains its minimum, which proves the claim. □

Theorem 1 implies that the joint BLUE-BLUP error vector $\underline{\tilde{\varepsilon}} = \underline{\mathscr{Z}} - \underline{\tilde{\mathscr{Z}}}$ is *uncorrelated* with the misclosure vector $\underline{v}$, i.e. $\mathsf{C}(\underline{\tilde{\varepsilon}}, \underline{v}) = 0$.

3 BLUE-BLUP recursion

In this section the recursive formulation of Theorem 1 is presented. It is based on the measurement- and dynamic model that forms the basis of the Kalman-filter. However, instead of the standard assumption of known state-vector means, we assume the means to be *unknown*.

3.1 Model assumptions

First we state the assumptions concerning the measurement- and dynamic model. Accordingly, the observational vector $\underline{y}$ is generalized to a time series of vectorial observables, $\underline{y}_1, \ldots, \underline{y}_t$. Here the role of the to-be-predicted vector $\underline{z}$ is taken by the state-vector $\underline{x}_t$. Hence, it is our aim to estimate the unknown state vector mean $x_t = \mathsf{E}(\underline{x}_t)$ and to predict the outcome of the random state-vector $\underline{x}_t$. It will be shown how such joint estimation/prediction can be performed recursively.

The dynamic model: The linear dynamic model, describing the time-evolution of the random state-vector $\underline{x}_i$, is given as

$$\underline{x}_i = \Phi_{i,i-1}\underline{x}_{i-1} + \underline{d}_i,\ i = 1, 2, \ldots, t \tag{6}$$

with

$$\mathsf{E}(\underline{x}_0) = x_0 \text{ (unknown)}, \quad \mathsf{D}(\underline{x}_0) = Q_{x_0 x_0} \tag{7}$$

and

$$\mathsf{E}(\underline{d}_i) = 0,\ \mathsf{C}(\underline{d}_i, \underline{d}_j) = S_i \delta_{i,j},\ \mathsf{C}(\underline{d}_i, \underline{x}_0) = 0 \tag{8}$$

for $i, j = 1, 2, \ldots, t$, with $\delta_{i,j}$ being the Kronecker delta, and where the $n \times n$ nonsingular matrix $\Phi_{i,i-1}$ denotes the transition matrix and the random vector $\underline{d}_i$ is the system noise. The system noise $\underline{d}_i$ is thus assumed to have a zero mean, to be uncorrelated in

time and to be uncorrelated with the initial state-vector $\underline{x}_0$. The transition matrix from epoch j to i is denoted as $\Phi_{i,j}$. Thus $\Phi_{i,j}^{-1} = \Phi_{j,i}$ and $\Phi_{i,i} = I_n$, the identity matrix of size n.
The measurement model: The link between the random vector of observables $\underline{y}_i \in \mathbb{R}^{m_i}$ and the random state-vector $\underline{x}_i \in \mathbb{R}^n$ is assumed given as

$$\underline{y}_i = A_i \underline{x}_i + \underline{n}_i,\ i = 1, 2, \ldots, t, \tag{9}$$

with

$$\mathsf{E}(\underline{n}_i) = 0,\ \mathsf{C}(\underline{n}_i, \underline{n}_j) = R_i \delta_{i,j} \tag{10}$$

and

$$\mathsf{C}(\underline{n}_i, \underline{x}_0) = 0,\ \mathsf{C}(\underline{n}_i, \underline{d}_j) = 0 \tag{11}$$

for $i, j = 1, 2, \ldots, t$. Thus the zero-mean measurement noise $\underline{n}_i$ is assumed to be uncorrelated in time and to be uncorrelated with the initial state-vector $\underline{x}_0$ and the system noise $\underline{d}_i$. Matrix A_1 of (9) is assumed to be of full column rank.

3.2 The three-step recursion

In the following, to show on which set of observables estimation/prediction are based, we use the notation $\underline{\tilde{x}}_{t|[\tau]} = [\underline{\hat{x}}_{t|[\tau]}^T,\ \underline{\check{x}}_{t|[\tau]}^T]^T$ when based on $\underline{y}_{[\tau]} = [\underline{y}_1^T, \ldots, \underline{y}_\tau^T]^T$. The variance matrix of the joint estimation-prediction error

$$\underline{\tilde{\varepsilon}}_{t|[\tau]} = [(x_t - \underline{\hat{x}}_{t|[\tau]})^T,\ (\underline{x}_t - \underline{\check{x}}_{t|[\tau]})^T]^T,$$

will be denoted by $\tilde{P}_{t|[\tau]}$.

Before forming the recursive counterpart of Theorem 1, an appropriate representation of the random misclosure vector $\underline{v}$, defined in lemma 1, must be formulated.

Lemma 2 *Let the linear model $\mathsf{E}(\underline{y}_{[t]}) = A_{[t],\tau} x_\tau$, $t = 1, 2, \ldots$, be structured by those given in (6) and (9). That is, $\underline{y}_{[t]} = [\underline{y}_{[t-1]}^T,\ \underline{y}_t^T]^T$, $A_{[t],\tau} = [A_{[t-1],\tau}^T,\ A_{t,\tau}^T]^T$ with $A_{i,\tau} = A_i \Phi_{i,\tau}$. Then there exists a representation of $\underline{v}_{[t]} = B_{[t]}^T \underline{y}_{[t]}$ as*

$$\underline{v}_{[t]} = \begin{bmatrix} \underline{v}_{[t-1]} \\ \underline{v}_t \end{bmatrix} = \begin{bmatrix} B_{[t-1]}^T \underline{y}_{[t-1]} \\ \underline{y}_t - A_t \underline{\check{x}}_{t|[t-1]} \end{bmatrix} \tag{12}$$

with $B_{[t-1]}$ and $B_{[t]}$ being basis matrices of the null spaces of $A_{[t-1],t}^T$ and $A_{[t],t}^T$, respectively.

Proof Matrix $B_{[t]}^T$ can be represented as

$$B_{[t]}^T = \begin{bmatrix} B_{[t-1]}^T & 0 \\ -A_t A_{[t-1],t}^- & I \end{bmatrix}, \quad t = 2, 3, \ldots \tag{13}$$

where $A^-_{[t-1],t}$ denotes an arbitrary left-inverse of $A_{[t-1],t}$, i.e. $A^-_{[t-1],t}A_{[t-1],t} = I_n$. Hence,

$$B^T_{[t]}\underline{y}_{[t]} = \left[[B^T_{[t-1]}\underline{y}_{[t-1]}]^T, [\underline{y}_t - A_t A^-_{[t-1],t}\underline{y}_{[t-1]}]^T \right]^T \tag{14}$$

The lemma is proven if $A^-_{[t-1],t}$ can be chosen such that $A^-_{[t-1],t}\underline{y}_{[t-1]}$ is the BLUP of $\underline{x}_t$ based on $\underline{y}_{[t-1]}$. Let $A^-_{[t-1],t}$ therefore be of the form

$$A^-_{[t-1],t} = A^+_{[t-1],t} + HB^T_{[t-1]} \tag{15}$$

for some H and where $A^+_{[t-1],t}$ is another left-inverse of $A_{[t-1],t}$. Then, since $A^+_{[t-1],t}\underline{y}_{[t-1]}$ is a linear unbiased statistic for x_t based on $\underline{y}_{[t-1]}$, it follows from Theorem 1 that matrix H can always be chosen such that $A^-_{[t-1],t}\underline{y}_{[t-1]} = \check{\underline{x}}_{t|[t-1]}$. □

We are now in a position to present the three-step procedure of the BLUE-BLUP recursion. In each step, use is made of Theorem 1, i.e. the MMSE-estimator/predictor is obtained from the sum of an unbiased linear statistic $\mathscr{G}$ and a linear function of $\underline{v}_{[t]}$ in (12).

Initialization ($t = 1$): We start with $\underline{y}_1 = A_1\underline{x}_1 + \underline{n}_1$. Since the random vector $\underline{v}_1 = B^T_1\underline{y}_1 = B^T_1\underline{n}_1$ is uncorrelated with the state-vector $\underline{x}_1$, we choose the following linear unbiased statistic

$$\mathscr{G}(\underline{y}_1) \mapsto U(A^T_1 R^{-1}_1 A_1)^{-1}A^T_1 R^{-1}_1 \underline{y}_1 \tag{16}$$

with $U = [I_n, I_n]^T$.

Using the identity $B^T_1 A_1 = 0$, the zero-covariance property $\mathsf{C}(A^T_1 R^{-1}_1 \underline{y}_1, \underline{v}_1) = 0$ follows as well. Thus the joint estimation-prediction error $[x^T_1, \underline{x}^T_1]^T - \mathscr{G}(\underline{y}_1)$ is uncorrelated with $\underline{v}_1$, meaning that the proposed statistic $\mathscr{G}(\underline{y}_1)$ itself is the joint BLUE-BLUP $\tilde{\underline{x}}_{1|1} = [\hat{\underline{x}}^T_{1|1}, \check{\underline{x}}^T_{1|1}]^T$. The error variance matrix $\tilde{P}_{1|1}$ also follows by an application of the variance propagation law to

$$\tilde{\underline{\varepsilon}}_{1|1} = U(\underline{x}_1 - \hat{\underline{x}}_{1|1}) + [I_n, 0]^T(x_1 - \underline{x}_1)$$

This, together with $Q_{x_1x_1} = \Phi_{1,0}Q_{x_0x_0}\Phi^T_{1,0} + S_1$, results in

$$\tilde{P}_{1|1} = U(A^T_1 R^{-1}_1 A_1)^{-1}U^T + \text{blockdiag}(Q_{x_1x_1}, 0), \tag{17}$$

since $\mathsf{D}(\underline{x}_1 - \hat{\underline{x}}_{1|1}) = (A^T_1 R^{-1}_1 A_1)^{-1}$ and $\mathsf{C}(\underline{x}_1, \underline{x}_1 - \hat{\underline{x}}_{1|1}) = 0$

Time update: In case of the time update step, we set $\mathscr{G}$ as

$$\mathscr{G}(\underline{y}_{[t-1]}) \mapsto \tilde{\Phi}_{t,t-1}\tilde{\underline{x}}_{t-1|[t-1]} \tag{18}$$

with $\tilde{\Phi}_{t,t-1} = \text{blockdiag}(\Phi_{t,t-1}, \Phi_{t,t-1})$. The corresponding joint estimation-prediction error can be expressed as

$$[x^T_t, \underline{x}^T_t]^T - \mathscr{G}(\underline{y}_{[t-1]}) = \tilde{\Phi}_{t,t-1}\tilde{\underline{\varepsilon}}_{t-1|[t-1]} + [0, I_n]^T\underline{d}_t \tag{19}$$

The estimation-prediction error $\underline{\tilde{\varepsilon}}_{t-1|[t-1]}$ is uncorrelated with $\underline{v}_{[t-1]}$ of (12) (cf. Theorem 1). Given the assumptions (8), (10) and (11), the system noise $\underline{d}_t$ is also uncorrelated with the previous observables, thus with any linear functions thereof, i.e. $\mathsf{C}(\underline{d}_t, \underline{v}_{[t-1]}) = 0$. This confirms the zero-covariance property between the estimation-prediction error (19) and $\underline{v}_{[t-1]}$. The BLUE-BLUP time-update is thus nothing else but the statistic given in (18).

With $\mathsf{C}(\underline{\tilde{\varepsilon}}_{t-1|[t-1]}, \underline{d}_t) = 0$, the error variance matrix $\tilde{P}_{t|[t-1]}$ is obtained by applying the variance propagation law to the representation of $\underline{\tilde{\varepsilon}}_{t|[t-1]}$ given in (19). This yields

$$\tilde{P}_{t|[t-1]} = \tilde{\Phi}_{t,t-1}\tilde{P}_{t-1|[t-1]}\tilde{\Phi}^T_{t,t-1} + \text{blockdiag}(0, S_t) \tag{20}$$

Measurement update: For the measurement-update, the BLUE-BLUP based on the data vector $\underline{y}_{[t-1]}$ is taken as the linear unbiased statistic of the data vector $\underline{y}_{[t]}$, that is

$$\mathscr{G}(\underline{y}_{[t]}) \mapsto \underline{\tilde{x}}_{t|[t-1]} \tag{21}$$

Now we make use of the representation of (12) by which $\underline{v}_t$ can also be re-written as

$$\underline{v}_t = \tilde{A}_t \underline{\tilde{\varepsilon}}_{t|[t-1]} + \underline{n}_t, \quad \text{with} \quad \tilde{A}_t = A_t[0, I_n] \tag{22}$$

Given the assumptions (8), (10) and (11), the measurement noise $\underline{n}_t$ is uncorrelated with the previous observables and the state-vectors. This, together with (22), yields $\mathsf{C}(\underline{\tilde{\varepsilon}}_{t|[t-1]}, \underline{v}_t) = \tilde{P}_{t|[t-1]}\tilde{A}^T_t$. Combining the results with $\mathsf{C}(\underline{\tilde{\varepsilon}}_{t|[t-1]}, \underline{v}_{[t-1]}) = 0$, an application of Theorem 1 gives finally

$$\underline{\tilde{x}}_{t|[t]} = \mathscr{G}(\underline{y}_{[t]}) + \tilde{P}_{t|[t-1]}\tilde{A}^T_t Q^{-1}_{v_t v_t} \underline{v}_t \tag{23}$$

With $\mathsf{C}(\underline{\tilde{\varepsilon}}_{t|[t-1]}, \underline{n}_t) = 0$, an application of the variance propagation law to (22) provides the following expression of the variance matrix $Q_{v_t v_t}$

$$Q_{v_t v_t} = R_t + \tilde{A}_t \tilde{P}_{t|[t-1]}\tilde{A}^T_t \tag{24}$$

Using the identity $\underline{\tilde{\varepsilon}}_{t|[t]} = \underline{\tilde{\varepsilon}}_{t|[t-1]} - \tilde{P}_{t|[t-1]}\tilde{A}^T_t Q^{-1}_{v_t v_t}\underline{v}_t$, the error variance matrix $\tilde{P}_{t|[t]}$ reads

$$\tilde{P}_{t|[t]} = \tilde{P}_{t|[t-1]} - \tilde{P}_{t|[t-1]}\tilde{A}^T_t Q^{-1}_{v_t v_t}\tilde{A}_t \tilde{P}_{t|[t-1]} \tag{25}$$

since $\mathsf{C}(\underline{\tilde{\varepsilon}}_{t|[t-1]}, \underline{v}_t) = \tilde{P}_{t|[t-1]}\tilde{A}^T_t$.
The structure of the above recursive procedure has been summarized in Theorem 2.

Theorem 2 (Recursive BLUE-BLUP) *The three steps of the BLUE-BLUP recursion are given as follows.*

Initialization:

$$\begin{aligned} \underline{\tilde{x}}_{1|1} &= U\underline{\hat{x}}_{1|1}, \\ \tilde{P}_{1|1} &= UP_{1|1}U^T + \tilde{Q}_{x_1 x_1} \end{aligned} \tag{26}$$

with $U = [I_n, I_n]^T$, $\underline{\hat{x}}_{1|1} = (A^T_1 R^{-1}_1 A_1)^{-1} A^T_1 R^{-1}_1 \underline{y}_1$, $P_{1|1} = (A^T_1 R^{-1}_1 A_1)^{-1}$, *and* $\tilde{Q}_{x_1 x_1} = \text{blockdiag}(Q_{x_1 x_1}, 0)$.

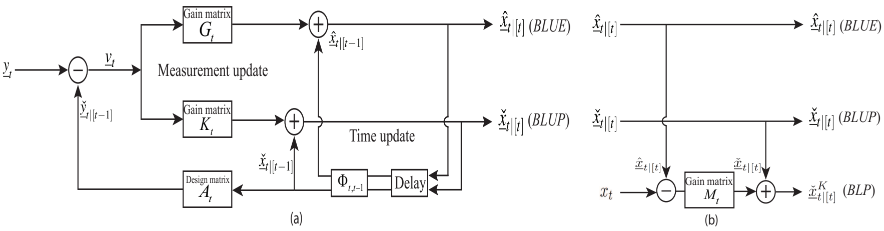

Fig. 1 (a) BLUE-BLUP recursion: measurement- and time-update, with gain matrices $G_t = C_{t|[t-1]}A_t^T Q_{v_t v_t}^{-1}$ and $K_t = P_{t|[t-1]}A_t^T Q_{v_t v_t}^{-1}$; (b) BLP from the BLUE and BLUP, with gain matrix $M_t = C_{t|[t]} Q_{t|[t]}^{-1}$.

Time-update:

$$\begin{aligned} \underline{\tilde{x}}_{t|[t-1]} &= \tilde{\Phi}_{t,t-1}\underline{\tilde{x}}_{t-1|[t-1]}, \\ \tilde{P}_{t|[t-1]} &= \tilde{\Phi}_{t,t-1}\tilde{P}_{t-1|[t-1]}\tilde{\Phi}_{t,t-1}^T + \tilde{S}_t \end{aligned} \tag{27}$$

with transition matrix $\tilde{\Phi}_{t,t-1} = \text{blockdiag}(\Phi_{t,t-1}, \Phi_{t,t-1})$ *and system noise variance matrix* $\tilde{S}_t = \text{blockdiag}(0, S_t)$.

Measurement-update:

$$\begin{aligned} \underline{\tilde{x}}_{t|[t]} &= \underline{\tilde{x}}_{t|[t-1]} + \tilde{K}_t \underline{v}_t, \\ \tilde{P}_{t|[t]} &= (I_{2n} - \tilde{K}_t \tilde{A}_t)\tilde{P}_{t|[t-1]} \end{aligned} \tag{28}$$

with $\underline{v}_t = \underline{y}_t - \tilde{A}_t \underline{\tilde{x}}_{t|[t-1]}$, $\tilde{A}_t = A_t[0, I_n]$, $Q_{v_t v_t} = R_t + A_t P_{t|[t-1]} A_t^T$, *and gain matrix* $\tilde{K}_t = \tilde{P}_{t|[t-1]}\tilde{A}_t^T Q_{v_t v_t}^{-1}$. $\diamondsuit$

With the use of the partitioning $\underline{\tilde{x}}_{t|[t]} = (\underline{\hat{x}}_{t|[t]}^T, \underline{\check{x}}_{t|[t]}^T)^T$ and $\tilde{K}_t = (G_t^T, K_t^T)^T$, the mechanism of the BLUE-BLUP recursion is illustrated with the block diagram given in Fig. 1 (a). It shows that, in contrast to the BLUP recursion, the BLUE-part of the BLUE-BLUP recursion cannot stand on its own. It requires $\underline{v}_t = \underline{y}_t - A_t \underline{\check{x}}_{t|[t-1]}$, and therefore the BLUP $\underline{\check{x}}_{t|[t-1]}$.

3.3 Role of the system noise

As the expressions of (26) and (27) show, the BLUE and BLUP both have the same initialization ($\underline{\hat{x}}_{1|1} = \underline{\check{x}}_{1|1}$) and the same time-update structure ($\tilde{\Phi}_{t,t-1} = \text{blockdiag}(\Phi_{t,t-1}, \Phi_{t,t-1})$). They differ however in their error variance matrices, which in turn makes their measurement-updates different. As the structure of $\tilde{P}_{t|[t-1]}$ in (27) shows, the difference between the BLUE and the BLUP is only driven by the system noise. This difference starts to be felt in the measurement-update of the time instance $t = 2$, where the corresponding BLUE/BLUP components of the partitioned gain matrix $\tilde{K}_t = [G_t^T,\ K_t^T]^T$ start deviating from each other (i.e. $G_t \neq K_t$). Would the system noises be absent (i.e. $S_t = 0\ \forall t$), then $G_t = K_t,\ \forall t$, thus making the outcomes of the recursive BLUE identical to that of the recursive BLUP.

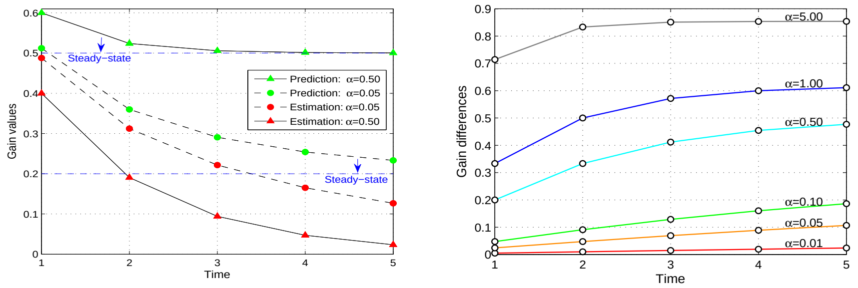

Fig. 2 *Left-panel*: Estimation gain values (in red) versus their prediction counterparts (in green) for two different values of $\alpha = 0.50$ (triangles) and $\alpha = 0.05$ (circles) over time. *Right-panel*: The difference in the gain values (i.e. $K_t - G_t$) for different values of α over time.

Example As illustration of the stated system noise role, we consider as example an observed time-series of random-walk noise with unknown trend. The underlying model follows from (9) and (6) by setting ($n = 1$)

$$A_t = 1, \quad \Phi_{t,t-1} = 1, \quad \forall\, t \tag{29}$$

We further assume the variance of the system noise to be related by that of the measurement noise, say σ^2, via the nonnegative scalar α as

$$R_t = \sigma^2, \quad S_t = \alpha\,\sigma^2, \quad \forall\, t \tag{30}$$

Employing the BLUE-BLUP recursion, the estimation and prediction gain values can be shown to read

$$G_t = \frac{1}{\sum_{i=1}^{t} w_i}, \quad K_t = \frac{w_t}{\sum_{i=1}^{t} w_i} \tag{31}$$

where the nonnegative weights w_t, $t = 1, 2, \ldots$, as polynomials of α, are computed as

$$w_1 = 1 + \alpha, \quad w_t = w_{t-1} + \alpha \sum_{i=1}^{t-1} w_i, \quad t = 2, 3, \ldots \tag{32}$$

Fig. 2 shows the gain values (and their difference) for different values of α. As shown, the difference between the two gain values is insignificant for small values of α, while the gain values deviate from each other by increasing α (right-panel).

The identities in (31) show that the estimation gain values, in this example, get smaller faster than their prediction counterparts, that is $G_t \leq K_t$ (see also Fig. 2, left-panel).This can be explained as follows. As stated, the estimation target vector is the unknown mean x_t which, in this example, does not change over time (i.e. $\Phi_{t,t-1} = 1$). Therefore, as the information content in the data vectors $\underline{y}_t$ is accumulated, the gain in improving the estimator due to the upcoming data gets less. In case of prediction however, the target vector is an outcome of the state-vector $\underline{x}_t$. Thus the gain in improving the predictor does generally rely on the observables in time. In the extreme case when the time instance tends to infinity, the *steady-state* gain values follow, namely

$$\lim_{t\to\infty} G_t = 0, \quad \lim_{t\to\infty} K_t = \frac{1}{2}(\sqrt{\alpha^2+4\alpha}-\alpha) \tag{33}$$

According to (33), as the filter converges to its steady-state form, the BLUE $\hat{\underline{x}}_{t|[t]}$ does not improve any more by accumulating further data. In case of prediction however, the constant gain values are generally different from zero meaning that the BLUP $\check{\underline{x}}_{t|[t]}$ still benefits from the further data. The steady-state error variance matrix of the joint BLUE-BLUP reads similarly

$$\lim_{t\to\infty} \tilde{P}_{t|[t]} = \begin{bmatrix} \sigma^2_{x_1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2}(\sqrt{\alpha^2+4\alpha}-\alpha)\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{34}$$

with $\sigma^2_{x_1} = \sigma^2_{x_0} + \alpha\sigma^2$ being the variance of the state-vector $\underline{x}_1$.

3.4 Role of the estimation-error variance matrix

To appreciate the contribution to the BLUE-BLUP recursion of the entries of the joint estimation-prediction error variance matrices, we partition $\tilde{P}_{t|[t]}$ as

$$\tilde{P}_{t|[t]} = \begin{bmatrix} Q_{t|[t]} & C_{t|[t]} \\ C^T_{t|[t]} & P_{t|[t]} \end{bmatrix}, \quad t = 1, 2, \ldots \tag{35}$$

with $Q_{t|[t]} = \mathsf{D}(x_t - \hat{\underline{x}}_{t|[t]})$ and $P_{t|[t]} = \mathsf{D}(\underline{x}_t - \check{\underline{x}}_{t|[t]})$, the error-variance matrices of estimation and prediction, and $C_{t|[t]} = \mathsf{C}(x_t - \hat{\underline{x}}_{t|[t]}, \underline{x}_t - \check{\underline{x}}_{t|[t]})$ their error cross-covariance. A similar partitioning is used for $\tilde{P}_{t|[t-1]}$. From Theorem 2 follows then:

Initialization:

$$\begin{aligned} P_{1|1} &= (A_1^T R_1^{-1} A_1)^{-1}, \\ C^T_{1|1} &= P_{1|1}, \\ Q_{1|1} &= P_{1|1} + Q_{x_1 x_1} \end{aligned} \tag{36}$$

Time-update:

$$\begin{aligned} P_{t|[t-1]} &= \Phi_{t,t-1} P_{t-1|[t-1]} \Phi^T_{t,t-1} + S_t, \\ C^T_{t|[t-1]} &= \Phi_{t,t-1} C^T_{t-1|[t-1]} \Phi^T_{t,t-1}, \\ Q_{t|[t-1]} &= \Phi_{t,t-1} Q_{t-1|[t-1]} \Phi^T_{t,t-1} \end{aligned} \tag{37}$$

Measurement-update:

$$\begin{aligned} P_{t|[t]} &= (I_n - K_t A_t) P_{t|[t-1]}, \\ C^T_{t|[t]} &= (I_n - K_t A_t) C^T_{t|[t-1]}, \\ Q_{t|[t]} &= Q_{t|[t-1]} - G_t A_t C^T_{t|[t-1]} \end{aligned} \tag{38}$$

with the gain matrices $G_t = C_{t|[t-1]} A_t^T Q^{-1}_{v_t v_t}$ and $K_t = P_{t|[t-1]} A_t^T Q^{-1}_{v_t v_t}$. This shows that the P- and C-matrices are not impacted by the error-estimation variance matrices Q. In particular note, that neither the estimation gain matrix G_t, nor the prediction gain matrix K_t, depend on the initial uncertainty $\mathsf{D}(\underline{x}_1) = Q_{x_1 x_1}$. This implies that the numerical sampling outcome of the BLUE-BLUP recursion is invariant for changes in $Q_{x_1 x_1}$. This variance matrix, and therefore also $Q_{x_0 x_0}$ and S_1, are thus not needed for computing the BLUE-BLUP outcomes $\hat{x}_{t|[t]}$ and $\check{x}_{t|[t]}$. The only role played by $Q_{x_1x_1}$ lies in describing how the uncertainty of $\underline{x}_1$ contributes to the uncertainty of the estimators at various time instances.

4 Relation to the Kalman filter

In this section we show how the BLUE-BLUP recursion specializes to that of the Kalman-filter in case the state-vector means are known.

4.1 From BLUE-BLUP to the BLP recursion

Let the mean $\mathsf{E}(\underline{x}_0) = x_0$ (cf. (7)) be known. Then $\mathsf{E}(\underline{x}_t) = x_t$ is known for all times, since $x_i = \Phi_{i,i-1}x_{i-1},\ i = 1,2,\ldots,t$. With all state-vector means known, the need for estimation disappears and the mean squared error of prediction can be improved. Hence, the BLP can now take over from the BLUE-BLUP. The BLP of $\underline{x}_t$, when based on $\underline{y}_1,\ldots,\underline{y}_\tau$, is denoted as $\underline{\check{x}}^K_{t|[\tau]}$ and its error variance matrix is denoted as $P^K_{t|[\tau]}$.

Lemma 3 (BLUE-BLUP and BLP) *In the presence of data, the BLP $\underline{\check{x}}^K_{t|[\tau]}$ and its error variance matrix $P^K_{t|[\tau]}$ can be expressed in the BLUE $\underline{\hat{x}}_{t|\tau}$ and BLUP $\underline{\check{x}}_{t|\tau}$, and their error variance matrices $P_{t|\tau}$ and $Q_{t|\tau}$, as*

$$\begin{array}{l} (i)\ \underline{\check{x}}^K_{t|[\tau]} = \underline{\check{x}}_{t|[\tau]} + C^T_{t|[\tau]}Q^{-1}_{t|[\tau]}(x_t - \underline{\hat{x}}_{t|[\tau]}) \\ (ii)\ P^K_{t|[\tau]} = P_{t|[\tau]} - C^T_{t|[\tau]}Q^{-1}_{t|[\tau]}C_{t|[\tau]} \end{array} \quad (39)$$

In the absence of data, the BLP of $\underline{x}_t$ is given as $\underline{\check{x}}^K_{t|[0]} = x_t$, with error variance matrix $P^K_{t|[0]} = Q_{x_tx_t}$.

Proof We first prove (i). With the mean $\mathsf{E}(\underline{x}_t) = x_t$ known, the misclosure vector $\underline{v}_{[\tau]}$ extends to $\underline{v}'_{[\tau]} = [\underline{v}^T_{[\tau]}, (x_t - \underline{\hat{x}}_{t|[\tau]})^T]^T$. Note, since $\mathsf{C}(\underline{v}_{[\tau]}, \underline{\hat{x}}_{t|[\tau]}) = 0$, that the variance matrix of $\underline{v}'_{[\tau]}$ is blockdiagional. To determine the MMSE-predictor $\underline{\check{x}}^K_{t|[\tau]}$, we apply Theorem 1. Accordingly, using $\mathscr{G}(\underline{y}_{[\tau]}) \mapsto \underline{\check{x}}_{t|[\tau]}$ as the linear unbiased statistic, we get

$$\begin{aligned} \underline{\check{x}}^K_{t|[\tau]} &= \underline{\check{x}}_{t|[\tau]} + \mathsf{C}(\underline{x}_t - \underline{\check{x}}_{t|[\tau]}, \underline{v}'_{[\tau]})Q^{-1}_{v'_{[\tau]}v'_{[\tau]}}\underline{v}'_{[\tau]} \\ &= \underline{\check{x}}_{t|[\tau]} + \mathsf{C}(\underline{x}_t - \underline{\check{x}}_{t|[\tau]}, x_t - \underline{\hat{x}}_{t|[\tau]})Q^{-1}_{t|[\tau]}(x_t - \underline{\hat{x}}_{t|[\tau]}) \end{aligned} \quad (40)$$

since $Q_{v'_{[\tau]}v'_{[\tau]}}$ is blockdiagonal and $\mathsf{C}(\underline{x}_t - \underline{\check{x}}_{t|[\tau]}, \underline{v}_{[\tau]}) = 0$. The result (i) now follows, since $C_{t|[\tau]} = \mathsf{C}(x_t - \underline{\hat{x}}_{t|[\tau]}, \underline{x}_t - \underline{\check{x}}_{t|[\tau]})$ by definition.

To prove (ii), recall that the MMSE prediction error is uncorrelated with the misclosure vector (cf. Theorem 1). Hence, the prediction error of $\underline{\check{x}}^K_{t|[\tau]}$ is uncorrelated with $\underline{v}'_{[\tau]}$ and thus also with $x_t - \underline{\hat{x}}_{t|[\tau]}$. With $\mathsf{C}(\underline{x}_t - \underline{\check{x}}^K_{t|[\tau]}, x_t - \underline{\hat{x}}_{t|[\tau]}) = 0$ and (i), the variance matrix of $\underline{x}_t - \underline{\check{x}}^K_{t|[\tau]}$ follows as given in (ii). □

This lemma shows how the BLP can be obtained from the BLUE, the BLUP and the known state-vector mean x_t. This is illustrated in the block diagram given in Fig. 1 (b). As the BLP makes use of the known mean x_t, it is a better predictor than the BLUP, i.e. $P^K_{t|[\tau]} \leq P_{t|[\tau]}$ (cf. (39)). Also note that the BLP prediction error is uncorrelated with the BLUE estimation error, i.e. $\mathsf{C}(\underline{x}_t - \underline{\check{x}}^K_{t|[\tau]}, x_t - \underline{\hat{x}}_{t|[\tau]}) = 0$.

We now use the above lemma to determine the recursive form of the BLP $\underline{\check{x}}^K_{t|[\tau]}$, thus giving the Kalman filter. This will also show how the Kalman gain matrix K^K_t is formed from the gain matrices K_t, G_t and M_t (cf. Fig. 1).

Lemma 4 (The Kalman Filter) *The three steps of the BLP recursion are given as follows.*

Initialization:

$$\begin{aligned} \check{x}^K_{0|0} &= \mathsf{E}(\underline{x}_0) = x_0, \\ P^K_{0|0} &= \mathsf{D}(\underline{x}_0 - \check{x}^K_{0|0}) = Q_{x_0 x_0} \end{aligned} \tag{41}$$

Time-update:

$$\begin{aligned} \underline{\check{x}}^K_{t|[t-1]} &= \Phi_{t,t-1}\underline{\check{x}}^K_{t-1|[t-1]} \\ P^K_{t|[t-1]} &= \Phi_{t,t-1}P^K_{t-1|[t-1]}\Phi^T_{t,t-1} + S_t \end{aligned} \tag{42}$$

Measurement-update:

$$\begin{aligned} \underline{\check{x}}^K_{t|[t]} &= \underline{\check{x}}^K_{t|[t-1]} + K^K_t \underline{v}^K_t \\ P^K_{t|[t]} &= (I_n - K^K_t A_t)P^K_{t|[t-1]} \end{aligned} \tag{43}$$

with $\underline{v}^K_t = \underline{y}_t - A_t\underline{\check{x}}^K_{t|[t-1]}$, $Q_{v^K_t v^K_t} = R_t + A_t P^K_{t|[t-1]} A^T_t$*, and Kalman gain matrix*

$$\begin{aligned} K^K_t &= K_t - M_t G_t \\ &= P^K_{t|[t-1]} A^T_t Q^{-1}_{v^K_t v^K_t} \end{aligned} \tag{44}$$

Proof As the mean x_0 is known, the best predictor of $\underline{x}_0$ in the absence of data is the mean. Hence, the initialization is given as in (41). To prove the time-update (42), first note that

$$\begin{aligned} \underline{\check{x}}_{t|[t-1]} &= \Phi_{t,t-1}\underline{\check{x}}_{t-1|[t-1]} \\ (x_t - \underline{\hat{x}}_{t|[t-1]}) &= \Phi_{t,t-1}(x_{t-1} - \underline{\hat{x}}_{t-1|[t-1]}) \\ C^T_{t|[t-1]}Q^{-1}_{t|[t-1]} &= \Phi_{t,t-1}(C^T_{t-1|[t-1]}Q^{-1}_{t-1|[t-1]})\Phi^{-1}_{t,t-1} \end{aligned} \tag{45}$$

where the last equation follows from (37). Substitution of (45) into the expression of (39) for $\tau = t-1$, gives $\underline{\check{x}}^K_{t,|[t-1]} = \Phi_{t,t-1}[\underline{\check{x}}_{t-1|[t-1]} + C^T_{t-1|[t-1]}Q^{-1}_{t-1|[t-1]}(x_{t-1} - \underline{\hat{x}}_{t-1|[t-1]})] = \Phi_{t,t-1}\underline{\check{x}}^K_{t-1|[t-1]}$, and thus the time-update (42). To prove (43), we first substitute $\underline{\check{x}}_{t|[t]} = \underline{\check{x}}_{t|[t-1]} + K_t\underline{v}_t$, $\underline{\hat{x}}_{t|[t]} = \underline{\hat{x}}_{t|[t-1]} + G_t\underline{v}_t$, and $M_t = C^T_{t|[t]}Q^{-1}_{t|[t]}$ into $\underline{\check{x}}^K_{t|[t]} = \underline{\check{x}}_{t|[t]} + C^T_{t|[t]}Q^{-1}_{t|[t]}(x_t - \underline{\hat{x}}_{t|[t]})$ (cf. (39) for $\tau = t$). This gives

$$\underline{\check{x}}^K_{t|[t]} = \underline{\check{x}}_{t|[t-1]} + (K_t - M_tG_t)\underline{v}_t + C^T_{t|[t]}Q^{-1}_{t|[t]}(x_t - \underline{\hat{x}}_{t|[t-1]}) \tag{46}$$

From the last two expressions of (38) follows

$$C^T_{t|[t]}Q^{-1}_{t|[t]} = [I_n - (K_t - M_tG_t)A_t]C^T_{t|[t-1]}Q^{-1}_{t|[t-1]} \tag{47}$$

Substitution into (46) gives, with (39) for $\tau = t-1$,

$$\check{\underline{x}}^K_{t|[t]} = \check{\underline{x}}^K_{t|[t-1]} + (K_t - M_t G_t)[\underline{y}_t - A_t(\check{\underline{x}}^K_{t|[t-1]} - \check{\underline{x}}_{t|[t-1]})] \tag{48}$$

from which the measurement update (43), with gain matrix (44), follows. □

Apart from the initialization, the recursive structure of the Kalman filter is the same as that of the BLUE-BLUP recursion. The initialization is different as the Kalman filter assumes the state-vector means known. The estimation of the mean $\mathsf{E}(\underline{x}_t) = x_t$ is therefore not needed and the initialization can start with the known mean $\mathsf{E}(\underline{x}_0) = x_0$. As a consequence, the initial uncertainty needs to be specified through $Q_{x_0x_0}$ (cf. (41)), which takes the role of the error variance matrix $P^K_{0|0}$. The BLUE-BLUP initialization however, does not require this variance matrix. As shown earlier, the BLUE-BLUP outcomes, $\hat{x}_{t|[t]}$ and $\check{x}_{t|[t]}$, do not depend on $Q_{x_0x_0}$. Hence, with the BLUE-BLUP recursion, the same results are obtained, irrespective of the choice made for this variance matrix. This is in marked contrast to the Kalman filter where the results are affected by $P^K_{0|0} = Q_{x_0x_0}$.

5 Conclusion

In this contribution we introduced a new recursive filter that does away with the need to have the state vector means of a dynamic system known. The recursive filter enables the joint linear MMSE prediction and estimation of the random state vectors and their unknown means, respectively (cf. Fig. 1). We discussed the role of the system noise and of the estimation-error variance matrix in the joint prediction and estimation of the filter. We showed how the filter specialize to the Kalman-filter in case the state-vector means are known and determined the relation between their respective error variance matrices and gain matrices. We also discussed the fundamentally different roles played by the intitialization of the two filters. In particular, it was shown that for the new filter the initial variance-matrix $Q_{x_0x_0}$ need not be known, this in contrast to the Kalman-filter.

References

Acharya R, Roy B, Sivaraman M, Dasgupta A (2011) Estimation of equatorial electrojet from total electron content at geomagnetic equator using Kalman filter. Advances in Space Research 47(6):938–944

Anderson BDO, Moore JB (1979) Optimal filtering, vol 11. Prentice-hall Englewood Cliffs, New Jersey

Bar-Shalom Y, Li X (1993) Estimation and tracking- principles, techniques, and software. Norwood, MA: Artech House, Inc, 1993

Bertino L, Evensen G, Wackernagel H (2002) Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. Inverse problems 18(1):1–23

Brammer K, Siffling G (1989) Kalman-Bucy Filters. Artech House

Candy J (1986) Signal Processing: Model Based Approach. McGraw-Hill, Inc.

Cao Y, Chen Y, Li P (2006) Wet refractivity tomography with an improved Kalman-filter method. Advances in Atmospheric Sciences 23:693–699

Christensen R (2001) Advanced linear modeling: multivariate, time series, and spatial data; nonparametric regression and response surface maximization, 2nd edn. Springer

Ferraresi M, Todini E, Vignoli R (1996) A solution to the inverse problem in groundwater hydrology based on Kalman filtering. Journal of hydrology 175(1):567–581

Gelb A (1974) Applied optimal estimation. MIT Press

Gibbs B (2011) Advanced Kalman Filtering, Least-squares and Modeling: A Practical Handbook. Wiley

Goldberger A (1962) Best linear unbiased prediction in the generalized linear regression model. Journal of the American Statistical Association 57(298):369–375

Grafarend EW (1976) Geodetic applications of stochastic processes. Physics of the Earth and Planetary Interiors 12(2):151–179

Grafarend EW, Rapp RH (1984) Advances in geodesy: selected papers from reviews of geophysics and space physics. Washington, DC: American Geophysical Union

Grewal MS, Andrews AP (2008) Kalman Filtering; Theory and Practice Using MATLAB, 3rd edn. John Wiley and Sons

Gross RS, Eubanks TM, Steppe JA, Freedman AP, Dickey JO, Runge TF (1998) A Kalman-filter-based approach to combining independent Earth-orientation series. J Geod 72(4):215–235

Herring TA, Davis JL, Shapiro II (1990) Geodesy by radio interferometry: The application of Kalman filtering to the analysis of very long baseline interferometry data. J Geophys Res (1978–2012) 95(B8):12,561–12,581

Ince CD, Sahin M (2000) Real-time deformation monitoring with GPS and Kalman Filter. Earth Planets and Space 52(10):837–840

Jazwinski A (1991) Stochastic processes and filtering theory. Dover Publications

Kailath T (1974) A view of three decades of linear filtering theory. IEEE Trans Inf Theory 20(2):146–181

Kailath T (1981) Lectures on Wiener and Kalman filtering. 140, Springer

Kailath T, Sayed AH, Hassibi B (2000) Linear Estimation. Prentice-Hall

Kalman RE (1960) A new approach to linear filtering and prediction problems. Journal of basic Engineering 82(1):35–45

Marx BA, Potthast RWE (2012) On instabilities in data assimilation algorithms. GEM-International Journal on Geomathematics 3(2):253–278

Maybeck P (1979) Stochastic Models, Estimation, and Control, vol 1. Academic Press, republished 1994

Sanso F (1980) The minimum mean square estimation error principle in physical geodesy (stochastic and non-stochastic interpretation). Boll Geod Sci Affi 39(2):112–129

Sanso F (1986) Statistical methods in physical geodesy. In: Sunkel H (ed) Mathematical and Numerical Techniques in Physical Geodesy, Lecture Notes in Earth Sciences, vol 7, Springer Berlin Heidelberg, pp 49–155

Simon D (2006) Optimal state estimation: Kalman, H [infinity] and nonlinear approaches. John Wiley and Sons

Sorenson HW (1966) Kalman filtering techniques. In Advances in Control Systems Theory and Applications, Ed CT Leondes 3:219–292

Stark H, Woods J (1986) Probability, random processes, and estimation theory for engineers. Prentice-Hall Englewood Cliffs, New Jersey

Teunissen PJG (2007) Best prediction in linear models with mixed integer/real unknowns: theory and application. J Geod 81(12):759–780

Teunissen PJG, Khodabandeh A (2013) BLUE, BLUP and the Kalman filter: some new results. J Geod 87(5):461–473