A Kalman-filtering derivation of simultaneous input and state estimation ⋆

Robert R. Bitmead [a], Morten Hovd [b], Mohammad Ali Abooshahab [b]

[a] *Department of Mechanical & Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, MS-0411, La Jolla CA 92093-0411, USA.*

[b] *Department of Engineering Cybernetics, Norwegian University of Science & Technology, 7491 Trondheim,Norway.*

Abstract

Simultaneous input and state estimation algorithms are studied as particular limits of Kalman filtering problems. This admits interpretation of the algorithm properties and critical analysis of their claims to being partly model-free and to providing unbiased estimates. A disturbance model, white noise of unbounded variance, is provided and the bias feature is shown to be a geometric projection property rather than probabilistic in nature. As a consequence of this analysis, the algorithm is connected, in the stationary case, to Algebraic Riccati equation computations for the gains, estimate covariances and filter frequency response.

Key words: state estimation, input estimation, Kalman filters, estimation algorithms

1 Introduction

The SISE algorithm has been the subject of considerable research interest since its inception [1–4] as an input reconstruction method suited to signal recovery in environmental and geophysical linear array analysis. Kitanidis [3] is generally credited with the formulation which seeks also to generate reliable state estimates. More recent works [5–8] have developed the algorithm *per se* for systems with direct feedthrough and for nonlinear problems, again with the emphasis on environmental estimation when an application is developed. The genesis of the algorithm is clearly based on least-squares linear estimation but invokes a number of properties to motivate and guide its derivation. These focus on the absence of two features: any statistical signal model for the input signal and any 'bias' in the state or input estimates. Part of our aim in this paper is re-derive and then extend the SISE algorithm by providing a specific input signal model (curiously suggested and then abandoned by both [1] and [3]) and applying standard Kalman filtering ideas.

⋆ The authors would like to thank for financial support from CINELDI – the Centre for Intelligent Electricity Distribution, Research Council of Norway project no. 257626/E20. This paper was not presented at any IFAC meeting. Corresponding author R.R. Bitmead. Tel. +1 858 822 3477.

Email addresses: `rbitmead@ucsd.edu` (Robert R. Bitmead), `morten.hovd@ntnu.no` (Morten Hovd), `mohammad.ali.abooshahab@ntnu.no` (Mohammad Ali Abooshahab).

The contribution of the paper is fourfold.

(i) demonstrating that SISE is a standard Kalman filtering algorithm with a specific disturbance model,
(ii) providing a Riccati equation approach to design SISE,
(iii) decomposing SISE as projections,
(iv) critiquing the model of *absence of model.*

2 The SISE algorithm

Consider the linear time-invariant system without direct feedthrough and with zero known control input,

$$x_{t+1} = Ax_t + Gd_t + w_t, \tag{1}$$
$$y_t = Cx_t + v_t, \tag{2}$$

[We take the time-invariant and zero-control system (1-2) solely for clarity in exposition. The time-varying and control-inclusive versions are direct and available in the cited references.] Make the following assumptions.

Assumption 1 *(i)* $x_t, w_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^q$, $d_t \in \mathbb{R}^m$, $v_t, y_t \in \mathbb{R}^p$.
(ii) these signals are mutually independent Gaussian white noises, $w_t \sim \mathcal{N}(0, Q)$, $v_t \sim \mathcal{N}(0, R)$ *and initial condition* $x_0 \sim \mathcal{N}(\hat{x}_{0|0}, P_0)$,
(iii) $R_t > 0$,
(iv) $\mathit{rank}\, CG = \mathit{rank}\, G = m$.

Then the simultaneous input and state estimation (SISE) algorithm, developed by [3] and summarized, refined and analyzed by [5], is as follows. At time t with current state estimate $\hat{x}^{\text{SISE}}_{t|t}$ with covariance P_t and measurement y_{t+1},

$$
\begin{aligned}
X_{t+1} &= AP_tA^T + Q, && (3)\\
K_{t+1} &= X_{t+1}C^T(CX_{t+1}C^T + R)^{-1}, && (4)\\
M_{t+1} &= [G^TC^T(CX_{t+1}C^T + R)^{-1}CG]^{-1}\\
&\quad \times G^TC^T(CX_{t+1}C^T + R)^{-1}, && (5)\\
P_{t+1} &= (I - K_{t+1}C)\left[(I - GM_{t+1}C)X_{t+1}\right.\\
&\quad \left.\times(I - GM_{t+1}C)^T + GM_{t+1}RM_{t+1}^TG^T\right]\\
&\quad + K_{t+1}RM_{t+1}^TG^T, && (6)\\
\hat{d}^{\text{SISE}}_{t|t+1} &= M_{t+1}(y_{t+1} - CA\hat{x}^{\text{SISE}}_{t|t}), && (7)\\
\hat{x}^{\text{SISE}}_{t+1|t+1} &= A\hat{x}^{\text{SISE}}_{t|t} + G\hat{d}^{\text{SISE}}_{t|t+1} + K_{t+1}\\
&\quad \times(y_{t+1} - CA\hat{x}^{\text{SISE}}_{t|t} - CG\hat{d}^{\text{SISE}}_{t|t+1}). && (8)\\
\operatorname{cov}(x_{t+1}|\mathbf{Y}^{t+1}) &= P_{t+1}, && (9)
\end{aligned}
$$

The purpose of this algorithm is to take the measurement sequence, $\mathbf{Y}^{t+1} \triangleq \{y_{t+1}, y_t, \dots, y_1\}$, and the current state estimate, $\hat{x}^{\text{SISE}}_{t|t}$, and covariance, P_t, and to produce estimates, $\hat{d}^{\text{SISE}}_{t|t+1}$ and $\hat{x}^{\text{SISE}}_{t+1|t+1}$ respectively, of the input and the state signals. The properties claimed of these estimates are as follows.

(1) No model whatsoever is provided for the evolution of the disturbance sequence $\{d_t\}$, including presumably that it might depend on $x_{t+\tau}$ or anything else.
(2) The estimates $\hat{d}^{\text{SISE}}_{t-1|t}$ and $\hat{x}^{\text{SISE}}_{t|t}$ are 'unbiased,' viz. $\mathrm{E}(\hat{d}^{\text{SISE}}_{t-1|t}|\mathbf{Y}^t) = d_{t-1}$ and $\mathrm{E}(\hat{x}^{\text{SISE}}_{t|t}|\mathbf{Y}^t) = x_t$, regardless of the values taken by $\{d_t\}$ and with expectations taken over $\sigma\{x_0, w_k, v_k : k = 0, 1, \dots, t\}$.
(3) Subject to possession of the above properties, the estimates are least mean squares [9], minimizing the criterion

$$J = \text{trace cov}\left(x_t|\mathbf{Y}^t\right). \qquad (10)$$

Our aim is to demonstrate that the SISE algorithm can be derived from a standard Kalman filtering problem and the non-properties of *no model* and *unbiasedness* can be linked to assumed signal properties. To achieve this, we provide a model for the $\{d_t\}$ sequence and for its relationship with the $\{x_t\}$ sequence; we assume that d_t is a Gaussian white noise sequence independent from other signals, with finite mean, $\mathfrak{d}$, but variance, D, tending to infinity. By doing so, we are able to provide a genealogy for the SISE algorithm and to show: that the algorithm's properties of convergence and stability in the time-invariant case, established by [7], follow naturally; that there are aspects of the algorithm preserved for finite D; and that the algorithm might be derived in a standard way by selecting a specific augmenting disturbance model.

3 Kalman filtering formulation

Make the following assumptions regarding the signals

Assumption 2 *The disturbance signal $d_t \sim \mathcal{N}(\mathfrak{d}, D)$ and is white and independent from x_0, w_τ, v_τ for all t and τ.*

Theorem 1 *Given system (1-2) subject to Assumptions 1 and 2, the Kalman filtering solution to input and state estimation is given as follows, from $\hat{x}_{t|t}$ and P_t.*

$$
\begin{aligned}
\mathcal{X}_{t+1} &= AP_tA^T + GDG^T + Q, && (11)\\
\mathcal{K}_{t+1} &= \mathcal{X}_{t+1}C^T(C\mathcal{X}_{t+1}C^T + R)^{-1}, && (12)\\
\mathcal{M}_{t+1} &= DG^TC^T(C\mathcal{X}_{t+1}C^T + R)^{-1}. && (13)\\
\mathcal{P}_{t+1} &= \mathcal{X}_{t+1} - \mathcal{X}_{t+1}C^T(C\mathcal{X}_{t+1}C^T + R)^{-1}C\mathcal{X}_{t+1},\\
&= (I - \mathcal{K}_{t+1}C)\mathcal{X}_{t+1}, && (14)\\
\hat{x}_{t+1|t+1} &= \mathrm{E}\left[x_{t+1}|\mathbf{Y}^{t+1}\right]\\
&= A\hat{x}_{t|t} + G\mathfrak{d}\\
&\quad + \mathcal{K}_{t+1}(y_{t+1} - CA\hat{x}_{t|t} - CG\mathfrak{d}), && (15)\\
\hat{d}_{t|t+1} &= \mathrm{E}\left[d_t|\mathbf{Y}^{t+1}\right]\\
&= \mathfrak{d} + \mathcal{M}_{t+1}(y_{t+1} - CA\hat{x}_{t|t} - CG\mathfrak{d}). && (16)
\end{aligned}
$$

The criterion minimized is altered from (10)*, which deals with $d_t - \hat{d}_{t|t+1}$ via the 'unbiasedness' condition, to*

$$J = \text{trace cov}\left(d_{t-1}|\mathbf{Y}^t\right) + \text{trace cov}\left(x_t|\mathbf{Y}^t\right), \qquad (17)$$

and these covariances are given by

$$
\begin{aligned}
&\operatorname{cov}\left(x_{t+1}|\mathbf{Y}^{t+1}\right) = \mathcal{P}_{t+1},\\
&\operatorname{cov}\left(d_t|\mathbf{Y}^{t+1}\right) = D - DG^TC^T(C\mathcal{X}_{t+1}C^T + R)^{-1}CGD, && (18)\\
&= (I - \mathcal{M}_{t+1}CG)D \triangleq \mathcal{D}_t. && (19)
\end{aligned}
$$

The proof, included in the Appendix, differs from those sketched by [1] and alluded to by [3]. Part of our aim is to establish, in Theorem 2 below, that as $D^{-1} \to 0$ the two algorithms coincide. This is more algebraic in nature than probabilistic. We achieve this via eight linking identities.

4 Identities for finite D

From the earlier definitions of matrices: X_{t+1}, $\mathcal{M}_{t+1}$, $\mathcal{X}_{t+1}$, $\mathcal{K}_{t+1}$, K_{t+1}, for finite values of D, we have the following set of sequential identities linking quantities in the Kalman filtering formulation to SISE.

Identity 1 (divisors)

$$CX_{t+1}C^T + R = (I_p - CG\mathcal{M}_{t+1})(C\mathcal{X}_{t+1}C^T + R).$$

Identity 2 (innovations)

$$y_{t+1} - CA\hat{x}_{t|t} - CG\hat{d}_{t|t+1} \\ = (I_p - CG\mathcal{M}_{t+1})(y_{t+1} - CA\hat{x}_{t|t} - CG\mathfrak{d}).$$

Identity 3 (state update gains)

$$\mathcal{K}_{t+1} = G\mathcal{M}_{t+1} + K_{t+1}(I_p - CG\mathcal{M}_{t+1}), \\ I - \mathcal{K}_{t+1}C = (I - K_{t+1}C)(I - G\mathcal{M}_{t+1}C).$$

Identity 4 (state updates)

$$\begin{aligned}\hat{x}_{t+1|t+1} &= A\hat{x}_{t|t} + G\mathfrak{d} + \mathcal{K}_{t+1}(y_{t+1} - CA\hat{x}_{t|t} - CG\mathfrak{d}), \\ &= A\hat{x}_{t|t} + G\hat{d}_{t|t+1} \\ &\quad + K_{t+1}(y_{t+1} - CA\hat{x}_{t|t} - G\hat{d}_{t|t+1}).\end{aligned}$$

Identity 4 establishes that the filtered state estimate updates for the finite-D Kalman filter and for SISE starting from the same values of $\hat{x}_{t|t}$ and P_t coincide when $\hat{d}_{t|t+1}$ is the same. Since the matrices $\mathcal{M}_{t+1}$ and M_{t+1} are not identical for finite D, the algorithms will differ in the $\hat{d}_{t-1|t}$ update, which is addressed by the next finite-D identity.

Identity 5 (disturbance update gain)

$$\mathcal{M}_{t+1} = \left[D^{-1} + G^TC^T(CX_{t+1}C^T + R)^{-1}CG\right]^{-1} \\ \times G^TC^T(CX_{t+1}C^T + R)^{-1}.$$

Identity 6 (disturbance update)

$$\hat{d}_{t|t+1} = (I_m - \mathcal{M}_{t+1}CG)\mathfrak{d} + \mathcal{M}_{t+1}(y_{t+1} - CA\hat{x}_{t|t}).$$

Identity 7 (disturbance estimation error covariance) *The covariance of $\hat{d}_{t|t+1}$, $\mathcal{D}_t$, satisfies*

$$\mathcal{D}_t = \left[D^{-1} + G^TC^T(CX_{t+1}C^T + R)^{-1}CG\right]^{-1}.$$

Identity 8 (covariances)

$$\mathcal{P}_{t+1} = (I - K_{t+1}C)\left\{(I - G\mathcal{M}_{t+1}C)X_{t+1} + G\mathcal{D}_tG^T\right\}.$$

5 Properties when $D^{-1} \to 0$: KF→SISE

Identities 1-8 lead to 9, following, which in turn permits the identification of SISE as the limit of a Kalman filter.

Identity 9 (disturbance and filter update gains)

As $D^{-1} \to 0$,

$$\mathcal{M}_t \to M_t, \;\; \mathcal{K}_t \to GM_t + K_t(I_p - CGM_t).$$

Lemma 1 *For M_{t+1} given by* (5)*, $M_{t+1}CG = I_m$. Whence, the matrices*

$$M_{t+1}CG \in \mathbb{R}^{m\times m}, \; CGM_{t+1} \in \mathbb{R}^{p\times p}, \; GM_{t+1}C \in \mathbb{R}^{n\times n},$$

are rank m projections on $\mathbb{R}^m$, $\mathbb{R}^p$, $\mathbb{R}^n$ respectively. The range spaces are given by

$$\mathcal{R}a\,(M_{t+1}CG) = \mathbb{R}^m, \\ \mathcal{R}a\,(CGM_{t+1}) = \mathcal{R}a\,(CG) \subseteq \mathbb{R}^p, \\ \mathcal{R}a\,(GM_{t+1}C) = \mathcal{R}a(G) \subseteq \mathbb{R}^n.$$

Theorem 2 *In the limit that $D^{-1} \to 0$, the Kalman filtering algorithm (11-16) coincides with the SISE algorithm (4-8).*

$$\hat{d}_{t|t+1} = \hat{d}^{\text{SISE}}_{t|t+1}, \quad \hat{x}_{t+1|t+1} = \hat{x}^{\text{SISE}}_{t+1|t+1},$$

with $\operatorname{cov}\left(x_{t+1}|\mathbf{Y}^{t+1}\right) = P_{t+1} = \mathcal{P}_{t+1}$ *and*

$$\operatorname{cov}\left(d_t|\mathbf{Y}^{t+1}\right) = \\ \left[G^TC^T(CAP_{t+1}A^TC^T + CQC^T + R)^{-1}CG\right]^{-1}.$$

We may next combine: Theorem 2, Lemma 1, (14) and Identity 8, to yield an interpretation of the SISE algorithm.

Corollary 1 *For the SISE algorithm, define the signals*

$$\hat{x}^{\text{SISE}}_{t+1|t} \triangleq (I - GM_{t+1}C)\left(A\hat{x}^{\text{SISE}}_{t|t} + G\mathfrak{d}\right), \\ \hat{x}^{\text{SISE}}_{t+1|t+\frac{1}{2}} \triangleq \hat{x}^{\text{SISE}}_{t+1|t} + GM_{t+1}y_{t+1}.$$

Then

$$\hat{x}^{\text{SISE}}_{t+1|t+1} = (I - K_{t+1}C)\,\hat{x}^{\text{SISE}}_{t+1|t+\frac{1}{2}} + K_{t+1}y_{t+1}.$$

Corollary 1 decomposes SISE into three steps.

(i) A time update projected onto the null space of G.
(ii) An update in the range space of G.
(iii) A Kalman-filter-like measurement update.

This sequence decodes the SISE covariance formula (6). Further, but consistent, reinterpretation of SISE unbiasedness as prioritizing the input signal estimate over state estimation is examined in Section 7 following.

6 Riccati-based steady-state SISE gains, performance and design

An evident and troubling absence from SISE is the Riccati difference equation associated with recursive linear least-squares optimal estimation. While, for the stationary case, the existence of and convergence to stationary values for the SISE gains and covariances has been established by [7] and others, the computation of these values is problematic without an algebraic Riccati equation connection. The algorithm performance, and hence design in terms of Q, R and perhaps D, is evaluated from these error covariances. By the same token, appreciation of the noise amplification properties of the algorithm is wanting in earlier works. A numerical example from a three-bus power system was computed but omitted due to imposed space restrictions. The steady-state SISE gain was computed by iterating for 500 steps and seen to equal the ARE solution.

For the steady-state Kalman filter version of SISE, KF-SISE, we solve the following ARE in MATLAB.

```
Sig = dare(A',C',Q+G*D*G',R);
KF = Sig*C'/(C*Sig*C'+R);
MKF = D*G'*C'/(C*Sig*C'+R);
P = (eye(n)-KF*C)*Sig;
Dd = (eye(m)-MKF*C*G)*D;
```

The ARE solution, `Sig` here, is the steady-state prediction error covariance $\mathcal{X}_\infty$ from (11). Variables `KF`, `MKF`, `P`, `Dd` are the Kalman filter gain $\mathcal{K}_\infty$ from (12), the disturbance gain $\mathcal{M}_\infty$ from (13), the filtered state error covariance $\mathcal{P}_\infty$ from (14) and the smoothed disturbance error covariance $\mathcal{D}_\infty$ from the proof of Identity 7, which follows, in turn, from the proof of Theorem 1. The systems from $d_t \to y_t$ and from $y_t \to \hat{d}_{t-1|t}$ may be computed as follows.

```
fwdsysd = ss(A,G,C,0,1)
deconsys = ...
    ss((eye(n)-KF*C)*A,KF,-MKF*C*A,MKF,1)
```

Thus, their frequency responses can be plotted to reveal that SISE implements a system inversion. This yields accessibility of standard linear systems design tools for the KF-SISE algorithm via the application of the ARE to derive the steady-state gain and covariance values.

7 Non-models, unbiasedness and input estimation

It is usually attributed to John von Neumann or to Stanislaw Ulam that the study of non-equilibrium thermodynamics in Physics is akin to the study of non-elephants in Zoology. By the same token, the study of *model-free* estimation is an unhelpful even meaningless description in this domain. Theorem 2 establishes that the SISE algorithm does indeed correspond to a particular model for the disturbance input process $\{d_t\}$ and thereby admits access to standard tools of linear least-squares estimation. The SISE concept that estimates are independent is replaced by the sounder hypothesis that in the signal model the disturbance input is independent from early values of the state.

Unbiasedness of the estimates, used in a probabilistically non-standard (but at least consistent) fashion in SISE since [3], refers to the property that, no matter the specific value taken by the disturbance, d_t, the conditional expected value, $\hat{x}_{t|t} = \mathrm{E}(x_t|\mathbf{Y}^t) = x_t$. This is not so much a statistical property as a geometric one captured by the projection operations of Corollary 1. The juxtaposition of probabilistic signal properties with non-models and absence of assumptions concerning the disturbance leads to fundamental questions regarding the nature of filtrations over which one is meant to take the expected values. By assuming a model, albeit a singular one, we are able to clarify these statements and to prove that they disguise a deterministic projection property. This might better be interpreted as a prioritization of the estimation of the input signal over that of the state, with the constraint $\mathrm{E}(\hat{d}_{t|t+1}) = d_t$, trumping the subsequent optimization of (10). The formulation developed in this paper can then be seen as a penalty function approach to this same constrained optimization.

8 Conclusion

We have derived from an algebraic perspective the SISE algorithms as Kalman filters of a specific type, suggested by Mendel [4] in his seismic deconvolution work. The input signal model is white noise, which, if its variance tends to infinity, yields a Kalman filter coinciding with SISE. As we mention, this was hinted at earlier but not carried through. Equipped with a fuller understanding of the connections, we were able to present new interpretations and to connect the approach to algebraic Riccati equation computational methods. Further, we were able to clarify – the uncharitable might say *debunk* – the ideas of model-free state estimation and estimate unbiasedness, showing that the methods necessarily involve projections induced by the large variances.

Appendix – proofs

Proof of Theorem 1

From Assumption 2, (1-2) and $\hat{x}_{t|t}$, P_t, write the joint conditional density

$$\mathrm{pdf}\left(\begin{bmatrix} x_{t+1} \\ d_t \\ y_{t+1} \end{bmatrix} \middle| \mathbf{Y}^t\right) = \mathcal{N}\left(\begin{bmatrix} A\hat{x}_{t|t} + G\mathfrak{d} \\ \mathfrak{d} \\ CA\hat{x}_{t|t} + CG\mathfrak{d} \end{bmatrix}, \begin{bmatrix} \mathcal{X}_{t+1} & GD & \mathcal{X}_{t+1}C^T \\ DG^T & D & DG^TC^T \\ C\mathcal{X}_{t+1} & CGD & C\mathcal{X}_{t+1}C^T + R \end{bmatrix}\right).$$

Then appeal to the standard Gaussian conditional density calculation to yield

$$\mathrm{E}\left(\begin{bmatrix}x_{t+1}\\ d_t\end{bmatrix}\middle|\mathbf{Y}^{t+1}\right)=\begin{bmatrix}A\hat{x}_{t|t}+G\mathfrak{d}\\ \mathfrak{d}\end{bmatrix}+\begin{bmatrix}\mathcal{X}_{t+1}C^T\\ DG^TC^T\end{bmatrix}\times(C\mathcal{X}_{t+1}C^T+R)^{-1}(y_{t+1}-CA\hat{x}_{t|t}-CG\mathfrak{d}),$$

or,

$$\begin{aligned}\hat{x}_{t+1|t+1}&=\mathrm{E}\left(x_{t+1}|\mathbf{Y}^{t+1}\right),\\ &=A\hat{x}_{t|t}+G\mathfrak{d}+\mathcal{X}_{t+1}C^T(C\mathcal{X}_{t+1}C^T+R)^{-1}\\ &\qquad\times(y_{t+1}-CA\hat{x}_{t|t}-CG\mathfrak{d}),\end{aligned}$$

and

$$\begin{aligned}\hat{d}_{t|t+1}&=\mathrm{E}\left(d_t|\mathbf{Y}^{t+1}\right),\\ &=\mathfrak{d}+DG^TC^T(C\mathcal{X}_{t+1}C^T+R)^{-1}\\ &\qquad\times(y_{t+1}-CA\hat{x}_{t|t}-CG\mathfrak{d}).\end{aligned}$$

These are (15) and (16), respectively in Theorem 1.

The covariance calculation follows similarly.

$$\begin{aligned}&\mathrm{cov}\left(\begin{bmatrix}x_{t+1}\\ d_t\end{bmatrix}\middle|\mathbf{Y}^{t+1}\right)=\\ &=\begin{bmatrix}\mathcal{X}_{t+1}&GD\\ DG^T&D\end{bmatrix}\\ &\quad-\begin{bmatrix}\mathcal{X}_{t+1}C^T\\ DG^TC^T\end{bmatrix}(C\mathcal{X}_{t+1}C^T+R)^{-1}\begin{bmatrix}C\mathcal{X}_{t+1}&CGD\end{bmatrix}.\end{aligned}$$

The (1,1)-block-element gives (14) for $\mathcal{P}_{t+1}$ and the (2,2)-block-element gives (18) for $\mathcal{D}_t$.. □

Proof of Identity 5

Using Identity 1 and denoting $Y_{t+1}=CX_{t+1}C^T+R$,

$$\begin{aligned}\mathcal{M}_{t+1}&=DG^TC^T(C\mathcal{X}_{t+1}C^T+R)^{-1},\\ &=DG^TC^T(CX_{t+1}C^T+R)^{-1}(I-CG\mathcal{M}_{t+1}),\\ &=DG^TC^TY_{t+1}^{-1}-DG^TC^TY_{t+1}^{-1}CG\mathcal{M}_{t+1},\end{aligned}$$

$$\begin{aligned}\left[I+DG^TC^TY_{t+1}^{-1}CG\right]\mathcal{M}_{t+1}&=DG^TC^TY_{t+1}^{-1},\\ \left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]\mathcal{M}_{t+1}&=G^TC^TY_{t+1}^{-1},\end{aligned}$$

$$\mathcal{M}_{t+1}=\left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]^{-1}G^TC^TY_{t+1}^{-1}.$$ □

Proof of Identity 7

From (18),

$$\begin{aligned}\mathcal{D}_t&=D-DG^TC^T(C\mathcal{X}_{t+1}C^T+R)^{-1}CGD,\\ &=(I-\mathcal{M}_{t+1}CG)D,\end{aligned}$$

Now, using Identity 5 and continuing the notation $Y_{t+1}=CX_{t+1}C^T+R$,

$$\begin{aligned}\mathcal{M}_{t+1}CG&=\left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]^{-1}G^TC^TY_{t+1}^{-1}CG,\\ &=\left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]^{-1}\\ &\qquad\times\left[-D^{-1}+D^{-1}+G^TC^TY_{t+1}^{-1}CG\right],\\ &=I-\left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]^{-1}D^{-1}.\end{aligned}$$

So,

$$\begin{aligned}\mathcal{D}_t&=\left[D^{-1}+G^TC^TY_{t+1}^{-1}CG\right]^{-1},\\ &=\left[D^{-1}+G^TC^T(CX_{t+1}C^T+R)^{-1}CG\right]^{-1}.\end{aligned}$$

□

Proof of Identity 8

Substitute for $\mathcal{K}_{t+1}$ from Identity 3 into (14), drop the time indices, and pay attention to the dimensions and typefaces,

$$\begin{aligned}\mathcal{P}_{t+1}&=(I_p-\mathcal{K}_{t+1}C)\,\mathcal{X}_{t+1},\\ &=(I-KC)(I-G\mathcal{M}C)\mathcal{X},\\ &=(I-KC)(I-G\mathcal{M}C)(X+GDG^T),\\ &=(I-KC)\left[(I-G\mathcal{M}C)X+(I-G\mathcal{M}C)GDG^T\right],\\ &=(I-KC)\left[(I-G\mathcal{M}C)X+G(I-\mathcal{M}CG)DG^T\right].\end{aligned}$$

Now, denoting (as above) $Y_{t+1}=CX_{t+1}C^T+R$, and using Identity 5,

$$\begin{aligned}I-\mathcal{M}CG&=(D^{-1}+G^TC^TY^{-1}CG)^{-1}\\ &\qquad\times\left\{D^{-1}+G^TC^TY^{-1}CG-G^TC^TY^{-1}CG\right\},\\ &=(D^{-1}+G^TC^TY^{-1}CG)^{-1}D^{-1},\\ &=\mathcal{D}_tD^{-1}.\end{aligned}$$

Whence, substituting this above,

$$\mathcal{P}_{t+1}=(I-K_{t+1}C)\left\{(I-G\mathcal{M}C)X_{t+1}+G\mathcal{D}_tG^T\right\}.$$ □

Proof of Theorem 2

Identity 9 establishes the convergence of $\mathcal{M}_{t+1}$ and $\mathcal{K}_{t+1}$. With the convergence of $\mathcal{M}_{t+1}$ and $M_{t+1}CG=I$, from

Lemma 1, we see that (7) and (16) are identical updates regardless of the value of $\mathfrak{d}$. Substituting the limiting value for $\mathcal{K}_{t+1}$ in (15) shows that this update and (8) also are identical. Thus, SISE and KF-SISE yield identical estimates at this t from the same starting data and, hence, the estimates remain identical. Since $\mathcal{P}_{t+1}$ is the conditional error covariance of x_{t+1} for KF-SISE and P_{t+1} is shown in [5] to be the conditional covariance of the SISE estimate, these covariances must also be identical. $\square$

Proof of Corollary 1

Applying Identity 4 to (8) and then Identities 3 and 9,

$$\begin{aligned}
\hat{x}^{\text{SISE}}_{t+1|t+1} &= A\hat{x}^{\text{SISE}}_{t|t} + G\mathfrak{d} + \mathcal{K}_{t+1}(y_{t+1} - CA\hat{x}^{\text{SISE}}_{t|t} - CG\mathfrak{d}),\\
&= (I - \mathcal{K}_{t+1}C)\, A\hat{x}^{\text{SISE}}_{t|t} + (I - \mathcal{K}_{t+1}C)\, G\mathfrak{d} + \mathcal{K}_{t+1}y_{t+1},\\
&= (I - K_{t+1}C)\,(I - GM_{t+1}C)\left(A\hat{x}^{\text{SISE}}_{t|t} + G\mathfrak{d}\right) + \mathcal{K}_{t+1}y_{t+1},\\
&= (I - K_{t+1}C)\,\hat{x}^{\text{SISE}}_{t+1|t} + \mathcal{K}_{t+1}y_{t+1},\\
&= (I - K_{t+1}C)\,\hat{x}^{\text{SISE}}_{t+1|t} + GM_{t+1}y_{t+1}\\
&\qquad\qquad + K_{t+1}(I - CGM_{t+1})y_{t+1},\\
&= (I - K_{t+1}C)\,\hat{x}^{\text{SISE}}_{t+1|t}\\
&\qquad + (I - K_{t+1}C)GM_{t+1}y_{t+1} + K_{t+1}y_{t+1},\\
&= (I - K_{t+1}C)\left(\hat{x}^{\text{SISE}}_{t+1|t} + GM_{t+1}y_{t+1}\right) + K_{t+1}y_{t+1},\\
&= (I - K_{t+1}C)\,\hat{x}^{\text{SISE}}_{t+1|t+\frac{1}{2}} + K_{t+1}y_{t+1}. \qquad \square
\end{aligned}$$

References

[1] J.D. Glover. The linear estimation of completely unknown signals. *IEEE Transactions on Automatic Control*, AC-14(6):766–769, 1969.

[2] P. Sanyal and C.N. Shen. Bayes' decision rule for rapid detection and adaptive estimation scheme with space applications. *IEEE Transactions on Automatic Control*, AC-19(3):228–231, 1974.

[3] P.K. Kitanidis. Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778, 1987.

[4] J.M. Mendel. White-noise estimators for seismic data processing in oil exploration. *IEEE Transactions on Automatic Control*, AC-22(5):694–706, October 1977.

[5] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time sysems. *Automatica*, 43:111–116, 2007.

[6] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time sysems with direct feedthrough. *Automatica*, 43:934–937, 2007.

[7] Huazhen Fang, R.A. de Callafon, and J. Cortés. Simultaneous input and state estimation for nonlinear systems with applications to flow field estimation. *Automatica*, 49:2805–2812, 2013.

[8] Sze Zhong Yong, Minghui Zhu, and Emilio Frazzoli. A unified filter for simultaneous input and state estiamtin of linear discrete-time stochastic systems. *Automatica*, 63:321–329, 2016.

[9] W. S. Kerwin and J. L. Prince. On the optimality of recursive unbiased state estimation with unknown inputs. *Automatica*, 36(9):1381–1383, September 2000.