International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Kalman filtering with constrained output injection

J. Chandrasekar [a]; D. S. Bernstein [a]; O. Barrero [b]; B. L. R. De Moor [b]
[a] Department of Aerospace Engineering, University of Michigan, Arun Arbor, USA
[b] ESAT-SCD (SISTA), Katholieke Univeriteit Leuven, Leuven, Belgium

First Published on: 05 September 2007
To cite this Article: Chandrasekar, J., Bernstein, D. S., Barrero, O. and De Moor, B. L. R. (2007) 'Kalman filtering with constrained output injection', International Journal of Control, 80:12, 1863 - 1879
To link to this article: DOI: 10.1080/00207170701373633
URL: http://dx.doi.org/10.1080/00207170701373633

International Journal of Control
Vol. 80, No. 12, December 2007, 1863–1879

Kalman filtering with constrained output injection

J. CHANDRASEKAR†, D. S. BERNSTEIN*†, O. BARRERO‡ and
B. L. R. DE MOOR‡

†Department of Aerospace Engineering, University of Michigan, Arun Arbor, USA
‡ESAT-SCD (SISTA), Katholieke Univeriteit Leuven, Leuven, Belgium

(*Received 1 May 2006; in final form 29 March 2007*)

In applications involving large scale systems such as discretized partial differential equations, it is often of interest to use data to estimate state variables associated with a subregion of the spatial domain. In this paper we derive an extension of the classical Kalman filter in which data injection is confined to a subspace of the system states.

1. Introduction

The classical Kalman filter provides optimal least-squares estimates of all of the states of a linear time-varying system under process and measurement noise. In many applications, however, optimal estimates are desired for a specified subset of the system states, rather than all of the system states. For example, for systems arising from discretized partial differential equations, the chosen subset of states can represent a subregion of the spatial domain. However, it is well known that the optimal state estimator for a subset of system states coincides with the classical Kalman filter (Gelb 1974, pp. 104–109).

For applications involving high-order systems, it is often difficult to implement the classical Kalman filter, and thus it is of interest to consider computationally simpler filters that yield suboptimal estimates of a specified subset of states. One approach to this problem is to consider reduced-order Kalman filters. These reduced-complexity filters provide state estimates that are suboptimal relative to the classical Kalman filter (Bernstein and Hyland 1985, Hippe and Wurmthaler 1990, Haddad and Bernstein 1990, Hsieh 2003). Alternative variants of the classical Kalman filter have been developed for computationally demanding applications such as weather forecasting (Farrell and Ioannou 2001, Heemink *et al*. 2001, Ballabrera *et al*. 2001, Fieguth *et al*. 2003), where the classical Kalman filter gain and covariance are modified so as to reduce the computational requirements.

The present paper is motivated by computationally demanding applications such as those discussed in Farrell and Ioannou (2001), Heemink *et al*. (2001), Ballabrera *et al*. (2001) and Fieguth *et al*. (2003). For such applications, a high-order simulation model is assumed to be available, but the derivation of a reduced-order filter in the sense of Bernstein and Hyland (1985), Hippe and Wurmthaler (1990), Haddad and Bernstein (1990), Hsieh (2003) is not feasible due to the high dimensionality of the analytic model. Instead, we use a full-order state estimator based directly on the simulation model. However, rather than implementing the classical Kalman filter, we derive an optimal spatially localized Kalman filter in which the structure of the filter gain is constrained to reflect the desire to estimate a specified subset of states. Our development is also more general than the classical treatment since the state dimension can be time varying, which is useful for variable-resolution discretizations of partial differential equations. Some of the results in this paper appeared in Barerro *et al*. (2005).

The use of a spatially localized Kalman filter in place of the classical Kalman filter is also motivated by computational architecture constraints arising from a multi-processor implementation of the Kalman filter (Lawrie *et al*. 1992) in which the Kalman filter operations can be confined to the subset of processors associated with the states whose estimates are desired.

*Corresponding author. Email: dsbaero@umich.edu

International Journal of Control
ISSN 0020–7179 print/ISSN 1366–5820 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207170701373633

2. Spatially localized Kalman filter

We consider the discrete-time dynamical system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad k \geqslant 0, \tag{1}$$

with output

$$y_k = C_k x_k + v_k, \tag{2}$$

where $x_k \in \mathbb{R}^{n_k}$, $u_k \in \mathbb{R}^{m_k}$, $y_k \in \mathbb{R}^{l_k}$, and A_k, B_k, C_k are known real matrices of appropriate size. The input u_k and output y_k are assumed to be measured, and $w_k \in \mathbb{R}^{n_{k+1}}$ and $v_k \in \mathbb{R}^{l_k}$ are zero-mean white noise processes with variances and correlation

$$\mathcal{E}\left[w_k w_j^{\mathrm{T}}\right] = Q_k \delta_{kj}, \quad \mathcal{E}\left[w_k v_j^{\mathrm{T}}\right] = S_k \delta_{kj},$$
$$\mathcal{E}\left[v_k v_j^{\mathrm{T}}\right]z = R_k \delta_{kj}, \tag{3}$$

where δ_{kj} is the Kronecker delta, and $\mathcal{E}[\cdot]$ denotes expected value. We assume that R_k is positive definite. The initial state x_0 is assumed to be uncorrelated with w_k and v_k. Note that the dimension n_k of the state x_k can be time varying, and thus $A_k \in \mathbb{R}^{n_{k+1} \times n_k}$ is not necessarily square.

For the system (1) and (2), we consider a state estimator of the form

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + \Gamma_k K_k (y_k - \hat{y}_k), \quad k \geqslant 0, \tag{4}$$

with output

$$\hat{y}_k = C_k \hat{x}_k, \tag{5}$$

where $\hat{x}_k \in \mathbb{R}^{n_k}$, $\hat{y}_k \in \mathbb{R}^{l_k}$, $\Gamma_k \in \mathbb{R}^{n_{k+1} \times p_k}$, and $K_k \in \mathbb{R}^{p_k \times l_k}$. The non-traditional feature of (4) is the presence of the term Γ_k, which, in the classical case is the identity matrix. Here, Γ_k constrains the state estimator so that only estimator states in the range of Γ_k are directly affected by the gain K_k. For example, Γ_k can have the form

$$\Gamma_k = \begin{bmatrix} 0 \\ I_{p_k} \\ 0 \end{bmatrix}, \tag{6}$$

where I_r denotes the $r \times r$ identity matrix. We assume that Γ_k has full column rank for all $k \geq 0$.

Next, define the state-estimation error state e_k by

$$e_k \triangleq x_k - \hat{x}_k, \tag{7}$$

which satisfies

$$e_{k+1} = \tilde{A}_k e_k + \tilde{w}_k, \quad k \geq 0, \tag{8}$$

where

$$\tilde{A}_k \triangleq A_k - \Gamma_k K_k C_k, \quad \tilde{w}_k \triangleq w_k - \Gamma_k K_k v_k. \tag{9}$$

Furthermore, we define the state-estimation error

$$J_k(K_k) \triangleq \mathcal{E}\left[(L_k e_{k+1})^{\mathrm{T}} L_k e_{k+1}\right], \tag{10}$$

where $L_k \in \mathbb{R}^{q_k \times n_{k+1}}$ determines the weighted error components. Then,

$$J_k(K_k) = \mathrm{tr}[P_{k+1} M_k], \tag{11}$$

where the error covariance $P_k \in \mathbb{R}^{n_k \times n_k}$ is defined by

$$P_k \triangleq \mathcal{E}\left[e_k e_k^{\mathrm{T}}\right] \tag{12}$$

and $M_k \triangleq L_k^{\mathrm{T}} L_k \in \mathbb{R}^{n_{k+1} \times n_{k+1}}$. We assume that M_k is positive definite for all $k \geqslant 0$. The following lemma will be useful.

Lemma 1: *The error* (7) *satisfies*

$$\mathcal{E}\left[e_k \tilde{w}_k^{\mathrm{T}}\right] = 0. \tag{13}$$

It thus follows from (8) and (13) that

$$\mathcal{E}\left[e_{k+1} e_{k+1}^{\mathrm{T}}\right] = \tilde{A}_k \mathcal{E}\left[e_k e_k^{\mathrm{T}}\right] \tilde{A}_k^{\mathrm{T}} + \mathcal{E}\left[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}\right]. \tag{14}$$

Note that (3) and (9) imply that

$$\mathcal{E}\left[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}\right] = \tilde{Q}_k, \tag{15}$$

where

$$\tilde{Q}_k \triangleq Q_k - \Gamma_k K_k S_k^{\mathrm{T}} - S_k K_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}} + \Gamma_k K_k R_k K_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}}. \tag{16}$$

It thus follows from (12), (14) and (15) that P_k satisfies

$$P_{k+1} = \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \tilde{Q}_k. \tag{17}$$

Therefore,

$$J_k(K_k) = \mathrm{tr}\left[(\tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \tilde{Q}_k) M_k\right]. \tag{18}$$

It follows from (9) and (16) that $J_k(K_k)$ can be expressed as

$$J_k(K_k) = \mathrm{tr}\Big[\Big((A_k - \Gamma_k K_k C_k)P_k(A_k - \Gamma_k K_k C_k)^{\mathrm{T}} + \tilde{Q}_k\Big)M_k\Big]. \quad (19)$$

3. Removing the noise correlation

In the classical case where $n_k = n$ and $\Gamma_k = I_n$ for all $k \geqslant 0$, the correlation S_k can be removed by introducing a linear combination of the measurements as deterministic inputs to the plant (Lewis 1986, pp. 181–183]. For the case $\Gamma_k \neq I_n$, we now state a condition under which we can derive an equivalent system with uncorrelated process and sensor noise.

Proposition 1: *Let $k \geqslant 0$ and suppose there exists $H_k \in \mathbb{R}^{p_k \times l_k}$ such that*

$$\Gamma_k H_k R_k = S_k. \quad (20)$$

Then

$$J_k(K_k) = \overline{J}_k(\overline{K}_k), \quad (21)$$

where

$$\overline{J}_k(\overline{K}_k) \triangleq \mathrm{tr}\Big[\Big((\overline{A}_k - \Gamma_k \overline{K}_k C_k)P_k(\overline{A}_k - \Gamma_k \overline{K}_k C_k)^{\mathrm{T}} + \overline{Q}_k + \Gamma_k \overline{K}_k R_k \overline{K}_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}})M_k\Big], \quad (22)$$

$$\overline{K}_k \triangleq K_k - H_k, \ \overline{A}_k \triangleq A_k - \Gamma_k H_k C_k, \quad (23)$$

and

$$\overline{Q}_k \triangleq Q_k - \Gamma_k H_k S_k^{\mathrm{T}} - S_k H_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}} + \Gamma_k H_k R_k H_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}}. \quad (24)$$

Proof: It follows from (24) that (18) can be expressed as

$$J_k(\overline{K}_k) = \mathrm{tr}\Big[\Big((\overline{A}_k - \Gamma_k \overline{K}_k C_k)P_k(\overline{A}_k - \Gamma_k \overline{K}_k C_k)^{\mathrm{T}} + \overline{Q}_k + \Gamma \overline{K}_k R_k \overline{K}_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}} - \Gamma_k \overline{K}_k S_k^{\mathrm{T}} - S_k \overline{K}_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}} + \Gamma_k \overline{K}_k R_k H_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}} + \Gamma_k H_k R_k \overline{K}_k^{\mathrm{T}} \Gamma_k^{\mathrm{T}}\Big)M_k\Big].$$

Using (20) yields (22). □

Note that replacing A_k, Q_k, and K_k in (18) by $\overline{A}_k$, $\overline{Q}_k$, and $\overline{K}_k$, respectively, and setting $S_k = 0$ in (18) yields (22). Hence, $\overline{J}_k(\overline{K}_k)$ is the cost of a system with uncorrelated process and sensor noise. It follows from (21) that $\overline{J}_k(\overline{K}_k)$ can be minimized with respect to $\overline{K}_k$, and K_k can be determined by using (23). If Γ_k is square and thus invertible by assumption, then $H_k = \Gamma_k^{-1} S_k R_k^{-1}$. In general, however, there may not exist a matrix H_k satisfying (20).

4 One-step spatially constrained Kalman filter

In this section we derive a one-step spatially constrained Kalman filter that minimizes the state-estimation error (18). For convenience, we define

$$\hat{S}_k \triangleq A_k P_k C_k^{\mathrm{T}} + S_k, \ \hat{R}_k \triangleq R_k + C_k P_k C_k^{\mathrm{T}}, \quad (25)$$

and $\pi_k \in \mathbb{R}^{n_{k+1} \times n_{k+1}}$ by

$$\pi_k \triangleq \Gamma_k (\Gamma_k^{\mathrm{T}} M_k \Gamma_k)^{-1} \Gamma_k^{\mathrm{T}} M_k. \quad (26)$$

Note that π_k is an oblique projector, that is, $\pi_k^2 = \pi_k$, but is not necessarily symmetric. Next, define the complementary oblique projector $\pi_{k\perp}$ by

$$\pi_{k\perp} \triangleq I_{n_{k+1}} - \pi_k. \quad (27)$$

Proposition 2: *The gain K_k that minimizes the cost $J_k(K_k)$ in (18) is given by*

$$K_k = \left(\Gamma_k^{\mathrm{T}} M_k \Gamma_k\right)^{-1} \Gamma_k^{\mathrm{T}} M_k \hat{S}_k \hat{R}_k^{-1}, \quad (28)$$

where the error covariance P_k is updated using

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} + \pi_{k\perp} \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}} \pi_{k\perp}^{\mathrm{T}} + Q_k - \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}}. \quad (29)$$

Proof: Setting $J_k'(K_k) = 0$ and using the fact that $\Gamma_k^{\mathrm{T}} M_k \Gamma_k$ is positive definite for all $k \geqslant 0$ yields (28). It follows from Bernstein (2005, p. 286) that, for all $0 < \alpha < 1$, all distinct $A_1, A_2 \in \mathbb{R}^{n \times m}$, and positive-definite $B \in \mathbb{R}^{m \times m}$, $\mathrm{tr}[\alpha(1-\alpha)(A_1 - A_2) \times B(A_1 - A_2)^{\mathrm{T}}] > 0$. Hence, the mapping $A \to \mathrm{tr}(ABA^{\mathrm{T}})$ is strictly convex. It thus follows that $J_k(K_k)$ is strictly convex, and hence K_k in (29) is the unique global minimizer of $J_k(K_k)$. To update the error covariance, we first note that

$$\Gamma_k K_k = \pi_k \hat{S}_k \hat{R}_k^{-1}, \quad (30)$$

where π_k is defined by (26). Now, using (30) with (17) yields (29). □

If either $M_k = I_{n_{k+1}}$ or $L_k = \Gamma_k^{\mathrm{T}}$, then π_k is the orthogonal projector

$$\pi_k = \Gamma_k(\Gamma_k^{\mathrm{T}}\Gamma_k)^{-1}\Gamma_k^{\mathrm{T}}, \tag{31}$$

and it follows from (28) that

$$K_k = \left(\Gamma_k^{\mathrm{T}}\Gamma_k\right)^{-1}\Gamma_k^{\mathrm{T}}\hat{S}_k\hat{R}_k^{-1}. \tag{32}$$

Alternatively, specializing to the case in which Γ_k is square yields $\pi_k = I_n$ and $\pi_{k\perp} = 0$, as well as the standard Riccati update equation

$$\begin{aligned} P_{k+1} = A_kP_kA_k^{\mathrm{T}} + Q_k - (A_kP_kC_k^{\mathrm{T}} + S_k) \\ \times (R_k + C_kP_kC_k^{\mathrm{T}})^{-1}(C_kP_kA_k^{\mathrm{T}} + S_k^{\mathrm{T}}). \end{aligned} \tag{33}$$

In this case the Kalman filter gain is given by

$$K_k = (A_kP_kC_k^{\mathrm{T}} + S_k)(R_k + C_kP_kC_k^{\mathrm{T}})^{-1} \tag{34}$$

and the estimator equation is

$$\hat{x}_{k+1} = A_k\hat{x}_k + B_ku_k + K_k(y_k - \hat{y}_k). \tag{35}$$

Furthermore, the one-step filter provides optimal estimates of all of the states, that is, the filter does not depend on the state-estimate error weighting L_k.

Next, we show that increasing the number of estimator states that are directly injected with the output improves the filter performance. Define $\hat{\pi}_k$ and $\hat{\pi}_{k\perp}$ by

$$\hat{\pi}_k \triangleq \hat{\Gamma}_k\left(\hat{\Gamma}_k^{\mathrm{T}}M_k\hat{\Gamma}_k\right)^{-1}\hat{\Gamma}_k^{\mathrm{T}}M_k, \quad \hat{\pi}_{k\perp} \triangleq I - \hat{\pi}_k. \tag{36}$$

where $\hat{\Gamma}_k$ has full column rank. Next, let $\hat{K}_k$ be the optimal gain given by (28) with Γ_k replaced by $\hat{\Gamma}_k$, that is,

$$\hat{K} \triangleq \left(\hat{\Gamma}_k^{\mathrm{T}}M_k\hat{\Gamma}_k\right)^{-1}\hat{\Gamma}_k^{\mathrm{T}}M_k\hat{S}_k\hat{R}_k^{-1}, \tag{37}$$

and let $\hat{P}_{k+1}$ be the corresponding error covariance when $\hat{K}_k$ is used, that is,

$$\begin{aligned} \hat{P}_{k+1} = A_kP_kA_k^{\mathrm{T}} + \hat{\pi}_{k\perp}\hat{S}_k\hat{R}_k^{-1}\hat{S}_k^{\mathrm{T}}\hat{\pi}_{k\perp}^{\mathrm{T}} \\ + Q_k - \hat{S}_k\hat{R}_k^{-1}\hat{S}_k^{\mathrm{T}}. \end{aligned} \tag{38}$$

Proposition 3: *Assume that $M_k = I$, let $\hat{\Gamma}_k = [\Gamma_k \ \ G_k]$, and assume $\hat{\Gamma}_k$ has full column rank. Then*

$$\mathrm{tr}(\hat{P}_k + 1) \leqslant \mathrm{tr}(P_k + 1). \tag{39}$$

Proof: Noting that π_k and $\hat{\pi}_k$ are symmetric, it follows from (36) that

$$\hat{\pi}_k = \pi_k + \pi_{k\perp}G_k\left(G_k^{\mathrm{T}}\pi_{k\perp}G_k\right)^{-1}G_k^{\mathrm{T}}\pi_{k\perp}. \tag{40}$$

Therefore,

$$\pi_{k\perp} = \hat{\pi}_{k\perp} + \pi_{k\perp}G_k\left(G_k^{\mathrm{T}}\pi_{k\perp}G_k\right)^{-1}G_k^{\mathrm{T}}\pi_{k\perp}. \tag{41}$$

Hence, subtracting (35) from (29) yields

$$\mathrm{tr}(P_{k+1} - \hat{P}_{k+1}) = \mathrm{tr}\left((\pi_{k\perp} - \hat{\pi}_{k\perp})\hat{S}_k\hat{R}_k^{-1}\hat{S}_k^{\mathrm{T}}\right) \geqslant 0. \quad \square$$

5. Two-step spatially constrained Kalman filter

In this section, we consider a two-step state estimator. The data assimilation step is given by

$$w_k^{\mathrm{da}} = \Upsilon_kK_{w,k}(y_k - y_k^{\mathrm{f}}), \quad k \geqslant 0, \tag{42}$$

and

$$x_k^{\mathrm{da}} = \mathrm{x}_k^{\mathrm{f}} + \Gamma_kK_{x,k}(y_k - \mathrm{y}_k^{\mathrm{f}}), \quad k \geqslant 0, \tag{43}$$

where $w_k^{\mathrm{da}} \in \mathbb{R}^{n_k}$ is the data assimilation estimate of w_k, $x_k^{\mathrm{da}} \in \mathbb{R}^{n_k}$ is the data assimilation estimate of x_k, and $x_k^{\mathrm{f}} \in \mathbb{R}^{n_k}$ is the forecast estimate of x_k. The forecast step or physics update is given by

$$x_{k+1}^{\mathrm{f}} = A_kx_k^{\mathrm{da}} + B_ku_k + w_k^{\mathrm{da}}, \quad k \geqslant 0, \tag{44}$$

$$y_k^{\mathrm{f}} = C_kx_k^{\mathrm{f}}. \tag{45}$$

Here, Υ_k is analogous to Γ_k in ensuring that only components of the process noise estimate in the range of Υ_k are directly affected by the gain $K_{w,k}$. We assume that Υ_k has full column rank for all $k \geq 0$. In traditional notation, x_k^{da} is denoted by $\hat{x}_{k|k}$ to indicate that $\hat{x}_{k|k}$ is the estimate of x_k obtained by using the measurements $y_0, \ldots, y_k$, while x_k^{f} is denoted by $\hat{x}_{k|k-1}$ to indicate that $\hat{x}_{k|k-1}$ is the estimate of x_k obtained by using the measurements $y_0, \ldots, y_{k-1}$. The notation x_k^{f} and x_k^{da} is motivated by the data assimilation literature (Scherliess *et al.* 2004).

Define the forecast state error e_k^{f} by

$$e_k^{\mathrm{f}} \triangleq x_k - x_k^{\mathrm{f}} \tag{46}$$

and the forecast error covariance $P_k^{\rm f}$ by

$$P_k^{\rm f} \triangleq \mathcal{E}\Big[e_k^{\rm f}\big(e_k^{\rm f}\big)^{\rm T}\Big]. \tag{47}$$

It follows from (1) and (44) that

$$e_{k+1}^{\rm f} = A_k e_k^{\rm da} + w_k - wa_k^{\rm da}, \quad k \geq 0, \tag{48}$$

where the data assimilation error state $e_k^{\rm da}$ is defined by

$$e_k^{\rm da} \triangleq x_k - x_k^{\rm da}. \tag{49}$$

Lemma 2: *The forecast error $e_k^{\rm f}$ satisfies*

$$\mathcal{E}\big[e_k^{\rm f} w_k^{\rm T}\big] = 0, \tag{50}$$

$$\mathcal{E}\big[e_k^{\rm f} v_k^{\rm T}\big] = 0. \tag{51}$$

Now, define the process noise estimation error

$$J_{w,k}(K_{w,k}) \triangleq \mathcal{E}\Big[\big(H_k\big(w_k - w_k^{\rm da}\big)\big)^{\rm T} H_k\big(w_k - w_k^{\rm da}\big)\Big], \tag{52}$$

where $H_k \in \mathbb{R}^{d_k \times n_{k+1}}$ determines the weighted error components. For convenience, define

$$N_k \triangleq H_k^{\rm T} H_k, \quad \chi_k \triangleq \Upsilon_k\big(\Upsilon_k^{\rm T} N_k \Upsilon_k\big)^{-1}\Upsilon_k^{\rm T} N_k,$$
$$\chi_{k\perp} \triangleq I_{n_{k+1}} - \chi_k. \tag{53}$$

Proposition 4: *The gain $K_{w,k}$ that minimizes the cost $J_{w,k}(K_{w,k})$ is given by*

$$K_{w,k} = \big(\Upsilon_k^{\rm T} N_k \Upsilon_k\big)^{-1}\Upsilon_k^{\rm T} N_k S_k\big(C_k P_k^{\rm f} C_k^{\rm T} + R_k\big)^{-1}. \tag{54}$$

Proof: Substituting (42) into (52), and using (3) and (50) in the resulting expression yields

$$J_w,k(K_w,k) = {\rm tr}\Big[\Big(Q_k - S_k K_{w,k}^{\rm T}\Upsilon_k^{\rm T} - \Upsilon_k K_{w,k} S_k^{\rm T} + \Upsilon_k K_{w,k}(C_k P_k^{\rm f} C_k + R_k)K_{w,k}^{\rm T}\Upsilon_k^{\rm T}\Big)N_k\Big]. \tag{55}$$

As in the proof of Proposition 2, $J_{w,k}(K_{w,k})$ is strictly convex. To obtain the optimal gain $K_{w,k}$, we set $J'_{w,k}(K_{w,k}) = 0$, which yields (54), the unique global minimizer of $J_{w,k}(K_{w,k})$. □

Next, define the state-estimation error

$$J_{x,k}(K_{x,k}) \triangleq \mathcal{E}\Big[\big(L_k e_k^{\rm da}\big)^{\rm T} L_k e_k^{\rm da}\Big] \tag{56}$$

so that

$$J_{x,k}(K_{x,k}) = {\rm tr}\big[P_k^{\rm da} M_k\big], \tag{57}$$

where the data assimilation error covariance $P_d^{\rm da} \in \mathbb{R}^{n_k \times n_k}$ is defined by

$$P_k^{\rm da} \triangleq \mathcal{E}\Big[e_k^{\rm da}\big(e_k^{\rm da}\big)^{\rm T}\Big]. \tag{58}$$

It follows from (43), (45), and (49) that

$$e_k^{\rm da} = \tilde{K}_{x,k} e_k^{\rm f} - \Gamma_k K_{x,k} v_k, \tag{59}$$

where

$$\tilde{K}_{x,k} \triangleq I - \Gamma_k K_{x,k} C_k. \tag{60}$$

Substituting (42) and (59) into (48) yields

$$e_{k+1}^{\rm f} = (A_k \tilde{K}_{x,k} - \Upsilon_k K_{w,k} C_k) e_k^{\rm f} + w_k - (A_k \Gamma_k K_{x,k} + \Upsilon_k K_{w,k}) v_k. \tag{61}$$

Next, define

$$R_k^{\rm f} \triangleq R_k + C_k P_k^{\rm f} C_k^{\rm T} \tag{62}$$

and

$$\begin{aligned} Q_k^{\rm f} \triangleq\, & Q_k - \big(A_k P_k^{\rm f} C_k^{\rm T} + S_k\big)\big(R_k^{\rm f}\big)^{-1}\big(A_k P_k^{\rm f} C_k^{\rm T} + S_k\big)^{\rm T} \\ & + A_k P_k^{\rm f} C_k^{\rm T}\big(R_k^{\rm f}\big)^{-1} C_k P_k^{\rm f} A_k^{\rm T} \\ & + \big(A_k \pi_{k\perp} P_k^{\rm f} C_k^{\rm T} + \chi_{k\perp} S_k\big)\big(R_k^{\rm f}\big)^{-1} \\ & \times \big(A_k \pi_{k\perp} P_k^{\rm f} C_k^{\rm T} + \chi_{k\perp} S_k\big)^{\rm T} \\ & - A_k \pi_{k\perp} P_k^{\rm f} C_k^{\rm T}\big(R_k^{\rm f}\big)^{-1} C_k P_k^{\rm f} \pi_{k\perp}^{\rm T} A_k^{\rm T}. \end{aligned} \tag{63}$$

Proposition 5: *The gain $K_{x,k}$ that minimizes the cost $J_{x,k}(K_{x,k})$ is given by*

$$K_{x,k} = \big(\Gamma_k^{\rm T} M_k \Gamma_k\big)^{-1}\Gamma_k^{\rm T} M_k P_k^{\rm f} C_k^{\rm T}\big(R_k^f\big)^{-1}, \tag{64}$$

where $P_k^{\rm da}$ and $P_k^{\rm f}$ are given by

$$P_k^{\rm da} = P_k^{\rm f} - P_k^{\rm f} C_k^{\rm T}\big(R_k^{\rm f}\big)^{-1} C_k P_k^{\rm f} + \pi_{k\perp} P_k^{\rm f} C_k^{\rm T}\big(R_k^{\rm f}\big)^{-1} C_k P_k^{\rm f} \pi_{k\perp}^{\rm T} \tag{65}$$

and

$$P_{k+1}^{\rm f} = A_k P_k^{\rm da} A_k^{\rm T} + Q_k^{\rm f}. \tag{66}$$

Proof: Using (58) and (59), $P_k^{\rm da}$ satisfies

$$\begin{aligned} {\rm P}_k^{\rm da} =\, & \tilde{K}_{x,k} P_k^{\rm f} \tilde{K}_{x,k}^{\rm T} - \tilde{K}_{x,k}\mathcal{E}\big[{\rm e}_k^{\rm f} v_k^{\rm T}\big] K_{x,k}^{\rm T}\Gamma_k^{\rm T} \\ & - \Gamma_k K_{x,k}\mathcal{E}\big[v_k\big({\rm e}_k^{\rm f}\big)^{\rm T}\big]\tilde{K}_{x,k}^{\rm T} + \Gamma_k K_{x,k} R_k K_{x,k}^{\rm T}\Gamma_k^{\rm T}. \end{aligned} \tag{67}$$

Substituting (51) into (67) and substituting the resulting equation into (57) yields

$$J_{x,k}(K_{x,k}) = {\rm tr}\Big[\Big(\tilde{K}_{x,k} P_k^{\rm f} \tilde{K}_{x,k}^{\rm T} + \Gamma_k K_{x,k} R_k K_{x,k}^{\rm T} \Gamma_k^{\rm T}\Big) M_k\Big]. \tag{68}$$

To obtain the optimal gain $K_{x,k}$, we set $J'_{x,k}(K_{x,k}) = 0$, which yields (64). As in the proof of Proposition 2, it can be shown that $J_{x,k}(K_{x,k})$ is strictly convex, and hence $K_{x,k}$ in (64) is the unique global minimizer of $J_{x,k}(K_{x,k})$. Substituting (50) and (64) into (67) yields (65).

To update the forecast error covariance, we substitute (42) into (48) so that

$$e^{\rm f}_{k+1} = A_k e^{\rm da}_k - \Upsilon_k K_{w,k} C_k e^{\rm f}_k + w_k - \Upsilon_k K_{w,k} v_k.$$

Hence,

$$\begin{aligned} P^{\rm f}_{k+1} = {} & A_k P^{\rm da}_k A^{\rm T}_k + Q_k + \Upsilon_k K_{w,k}\big(C_k P^{\rm f}_k C^{\rm T}_k + R_k\big)K^{\rm T}_{w,k}\Upsilon^{\rm T}_k \\ & + A_k\mathcal{E}\big[e^{\rm da}_k w^{\rm T}_k\big] + \mathcal{E}\big[w_k\big(e^{\rm da}_k\big)^{\rm T}\big]A^{\rm T}_k \\ & - A_k\mathcal{E}\big[e^{\rm da}_k\big(e^{\rm f}_k\big)^{\rm T}\big]C^{\rm T}_k K^{\rm T}_{w,k}\Upsilon^{\rm T}_k \\ & - \Upsilon_k K_{w,k} C_k\mathcal{E}\big[e^{\rm f}_k\big(e^{\rm da}_k\big)^{\rm T}\big]A^{\rm T}_k \\ & - A_k\mathcal{E}\big[e^{\rm da}_k v^{\rm T}_k\big]K^{\rm T}_{w,k}\Upsilon^{\rm T}_k - \Upsilon_k K_{w,k}\mathcal{E}\big[v_k\big(e^{\rm da}_k\big)^{\rm T}\big]A^{\rm T}_k \\ & - \mathcal{E}\big[w_k\big(e^{\rm f}_k\big)^{\rm T}\big]C^{\rm T}_k K^{\rm T}_{w,k}\Upsilon^{\rm T}_k - \Upsilon_k K_{w,k} C_k\mathcal{E}\big[e^{\rm f}_k w^{\rm T}_k\big] \\ & - \mathcal{E}\big[w_k v^{\rm T}_k\big]K^{\rm T}_{w,k}\Upsilon^{\rm T}_k - \Upsilon_k K_{w,k}\mathcal{E}\big[v_k w^{\rm T}_k\big] \\ & + \Upsilon_k K_{w,k}\big(C_k\mathcal{E}\big[e^{\rm f}_k v^{\rm T}_k\big] \\ & + \mathcal{E}\big[v_k\big(e^{\rm f}_k\big)^{\rm T}\big]C^{\rm T}_k\big)K^{\rm T}_{w,k}\Upsilon^{\rm T}_k. \end{aligned} \tag{69}$$

Substituting (59) into (69), and using (50) and (51) in the resulting expression yields (11). □

The two-step estimator can be summarized as follows

Data assimilation step:

$$w^{\rm da}_k = \Upsilon_k K_{w,k}\big(y_k - y^{\rm f}_k\big), \tag{70}$$

$$K_{w,k} = \big(\Upsilon^{\rm T}_k N_k \Upsilon_k\big)^{-1}\Upsilon^{\rm T}_k N_k S_k\big(R^{\rm f}_k\big)^{-1}, \tag{71}$$

$$x^{\rm da}_k = x^{\rm f}_k + \Gamma_k K_{x,k}\big(y_k - y^{\rm f}_k\big), \tag{72}$$

$$K_{x,k} = \big(\Gamma^{\rm T}_k M_k \Gamma_k\big)^{-1}\Gamma^{\rm T}_k M_k P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}, \tag{73}$$

$$\begin{aligned} P^{\rm da}_k = {} & P^{\rm f}_k - P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}C_k P^{\rm f}_k \\ & + \pi_{k\perp}P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}C_k P^{\rm f}_k \pi^{\rm T}_{k\perp}. \end{aligned} \tag{74}$$

Forecast step:

$$x^{\rm f}_{k+1} = A_k x^{\rm da}_k + B_k u_k + w^{\rm da}_k, \tag{75}$$

$$P^{\rm f}_{k+1} = A_k P^{\rm da}_k A^{\rm T}_k + Q^{\rm f}_k. \tag{76}$$

Assume that Γ_k and Υ_k are square for all $k \geqslant 0$. Substituting (70) and (72) into (75) yields the familiar one-step Kalman filter

$$\begin{aligned} x^{\rm f}_{k+1} = {} & A_k x^{\rm f}_k + B_k u_k \\ & + \big(A_k P^{\rm f}_k C^{\rm T}_k + S_k\big)\big(R_k + C_k P^{\rm f}_k C_k\big)^{-1}\big(y_k - y^{\rm f}_k\big). \end{aligned} \tag{77}$$

Furthermore, substituting (74) into (75) yields

$$\begin{aligned} P^{\rm f}_{k+1} = {} & A_k P^{\rm f}_k A^{\rm T}_k - \big(A_k P^{\rm f}_k C_k + S_k\big) \\ & \times \big(R_k + C_k P^{\rm f}_k C^{\rm T}_k\big)^{-1}\big(C_k P^{\rm f}_k A^{\rm T}_k + S^{\rm T}_k\big) + Q_k. \end{aligned} \tag{78}$$

Next, as in Proposition 3, we show that when additional estimator states are directly injected with the output data, the performance of the two-step filter improves. Define $\hat{K}_{x,k}$ by (64) with Γ_k replaced by $\hat{\Gamma}_k$, that is,

$$\hat{K}_{x,k} = \big(\hat{\Gamma}^{\rm T}_k M_k \hat{\Gamma}_k\big)^{-1}\hat{\Gamma}^{\rm T}_k M_k P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}. \tag{79}$$

Furthermore, let $\hat{P}^{\rm da}_k$ be the corresponding data assimilation error covariance when $\hat{K}_{x,k}$ is used instead of $K_{x,k}$, that is,

$$\begin{aligned} \hat{P}^{\rm da}_k \triangleq {} & P^{\rm f}_k - P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}C_k P^{\rm f}_k \\ & + \hat{\pi}_{k\perp}P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}C_k P^{\rm f}_k \hat{\pi}^{\rm T}_{k\perp}. \end{aligned} \tag{80}$$

Proposition 6: *Let $M_k = I$, $\hat{\Gamma}_k = [\Gamma_k \;\; G_k]$, and assume that $\hat{\Gamma}_k$ has full column rank. Then*

$$\mathrm{tr}\big(\hat{P}^{\rm da}_k\big) \leqslant \mathrm{tr}\big(P^{\rm da}_k\big). \tag{81}$$

Proof: Subtracting (80) from (65) and using the fact from (41) that $\pi_{k\perp} - \hat{\pi}_{k\perp}$ is positive semi-definite, it follows that

$$\mathrm{tr}\big(P^{\rm da}_k - \hat{P}^{\rm da}_k\big) = \mathrm{tr}\Big[\big(\pi_{k\perp} - \hat{\pi}_{k\perp}\big)P^{\rm f}_k C^{\rm T}_k\big(R^{\rm f}_k\big)^{-1}C_k P^{\rm f}_k\Big] \geqslant 0.$$

□

6. Comparison of the one-step and two-step filters

When Γ_k and Υ_k are square, comparing (33) with (78) and (35) with (75) shows that the two-step filter is equivalent to the one-step filter with $K_k = AK_{x,k} + K_{w,k}$, $\hat{x}_k = x^{\rm f}_k$ and $P_k = P^{\rm f}_k$. When Γ_k and Υ_k are not square, we obtain a sufficient condition under which the one-step and two-step spatially constrained Kalman filters are equivalent.

Proposition 7: *Suppose that $\hat{x}_0 = x^{\rm f}_0$ and $P_0 = P^{\rm f}_0$, and, for all $k \geqslant 0$,*

$$A_k \pi_{k\perp} P^{\rm f}_k C^{\rm T}_k + \chi_{k\perp} S_k = \pi_{k\perp}\big(A_k P^{\rm f}_k C^{\rm T}_k + S_k\big). \tag{82}$$

Then the one-step filter (28), (29) *and the two-step filter in* (70)–(76) *are equivalent, that is, for all $k > 0$, $\hat{x}_k = x^{\rm f}_k$ and $P_k = P^{\rm f}_k$.*

Proof: Substituting (63) and (74) into (76) yields

$$P^{\mathrm{f}}_{k+1} = A_k P^{\mathrm{f}}_k A_k^{\mathrm{T}} + \left(A_k \pi_{k\perp} P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_{k\perp} S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1} \times \left(A_k \pi_{k\perp} P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_{k\perp} S_k\right)^{\mathrm{T}} - \left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right) \times \left(R^{\mathrm{f}}_k\right)^{-1}\left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right)^{\mathrm{T}} + Q_k. \quad (83)$$

Substituting (88) into (83) yields

$$P^{\mathrm{f}}_{k+1} = A_k P^{\mathrm{f}}_k A_k^{\mathrm{T}} + \pi_{k\perp}\left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1} \times \left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right)^{\mathrm{T}} \pi_{k\perp}^{\mathrm{T}} + Q_k - \left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right) \times \left(R^{\mathrm{f}}_k\right)^{-1}\left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right)^{\mathrm{T}}. \quad (84)$$

Since $P^{\mathrm{f}}_0 = P_0$, it follows from (25), (29), and (62) that, for all $k > 0$, $P^{\mathrm{f}}_k = P_k$.

Next, substituting (42) and (72) into (35) yields

$$x^{\mathrm{f}}_{k+1} = A_k x^{\mathrm{f}}_k + B_k u_k + \left(A_k \pi_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_k S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1}\left(y_k - y^{\mathrm{f}}_k\right). \quad (85)$$

Now, (62) and (88) imply that

$$x^{\mathrm{f}}_{k+1} = A_k x^{\mathrm{f}}_k + B_k u_k + \pi_k\left(A_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + S_k\right) \times \left(C_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + R_k\right)^{-1}\left(y_k - C_k x^{\mathrm{f}}_k\right). \quad (86)$$

It follows from (4) and (28) that, for all $k \geqslant 0$,

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + \pi_k\left(A_k P_k C_k^{\mathrm{T}} + S_k\right) \times \left(C_k P_k C_k^{\mathrm{T}} + R_k\right)^{-1}\left(y_k - C_k \hat{x}_k\right). \quad (87)$$

Since $\hat{x}_0 = x^{\mathrm{f}}_0$ and $P^{\mathrm{f}}_k = P_k$ for all $k \geqslant 0$, (86) and (87) imply that $\hat{x}_k = x^{\mathrm{f}}_k$ for all $k \geqslant 0$. □

Note that, if Γ_k and Υ_k are square, then $\pi_{k\perp} = 0$ and $\chi_{k\perp} = 0$, and thus (88) is satisfied. Furthermore, if $S_k = 0$ or $\pi_k = \chi_k$, then Proposition 7 specializes to the following result.

Corollary 1: *Suppose that $\hat{x}_0 = x^{\mathrm{f}}_0$, $P_0 = P^{\mathrm{f}}_0$, and, for all $k \geqslant 0$, either $S_k = 0$ or $\pi_k = \chi_k$. If*

$$A_k \pi_{k\perp} = \pi_{k\perp} A_k, \quad (88)$$

for all $k \geqslant 0$, then the one-step filter (28), (29) and the two-step filter in (70)–(76) are equivalent, that is, for all $k > 0$, $\hat{x}_k = x^{\mathrm{f}}_k$ and $P_k = P^{\mathrm{f}}_k$.

Next, we present a converse of Proposition 7.

Proposition 8: *Assume that the one-step filter (28), (29) and the two-step filter in (70)–(76) are equivalent, that is, for all $k \geqslant 0$, $\hat{x}_k = x^{\mathrm{f}}_k$ and $P_k = P^{\mathrm{f}}_k$. Then, for all $k \geqslant 0$, there exists an orthogonal matrix $U_k \in \mathbb{R}^{l_k \times l_k}$ such that*

$$\left(A_k \pi_{k\perp} P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_{k\perp} S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1/2} U_k = \pi_{k\perp}\left(A_k P_k C_k^{\mathrm{T}} + S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1/2}. \quad (89)$$

Proof: Since $P_k = P^{\mathrm{f}}_k$ for all $k \geqslant 0$, subtracting (29) from (84) yields

$$\pi_{k\perp} \hat{S}_k\left(C_k P_k C_k^{\mathrm{T}} + R_k\right)^{-1} \hat{S}_k^{\mathrm{T}} \pi_{k\perp}^{\mathrm{T}} = \left(A_k \pi_{k\perp} P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_{k\perp} S_k\right)\left(R^{\mathrm{f}}_k\right)^{-1} \times \left(A_k \pi_{k\perp} P^{\mathrm{f}}_k C_k^{\mathrm{T}} + \chi_{k\perp} S_k\right)^{\mathrm{T}}. \quad (90)$$

Hence, (89) follows from (25) and Bernstein (2005 p. 193]. □

Neither the one-step nor the two-step filter performs consistently better than the other. However, there are special cases when the performance of one filter is better than the other.

Proposition 9: *Assume that $C_k = 0$ and $P_k = P^{\mathrm{f}}_k$. If Γ_k is square and Υ_k is not square, then*

$$P_{k+1} \leqslant P^{\mathrm{f}}_{k+1}. \quad (91)$$

Alternatively, if Γ_k is not square and Υ_k is square, then

$$P^{\mathrm{f}}_{k+1} \leqslant P_{k+1}. \quad (92)$$

Proof: Assume that Γ_k is square and Υ_k is not square. It then follows from (26), (27) and (53) that

$$\pi_{k\perp} = 0, \quad \chi_{k\perp} \neq 0.$$

Substituting (74) and (63) into (76), and using $C_k = 0$ and $\pi_{k\perp} = 0$ yields

$$P^{\mathrm{f}}_{k+1} = A_k P^{\mathrm{f}}_k A_k^{\mathrm{T}} + \chi_{k\perp} S_k\left(C_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + R_k\right)^{-1} \times S_k^{\mathrm{T}} \chi_{k\perp}^{\mathrm{T}} - S_k\left(C_k P^{\mathrm{f}}_k C_k^{\mathrm{T}} + R_k\right)^{-1} S_k^{\mathrm{T}} + Q_k. \quad (93)$$

Substituting $C_k = 0$ and $\pi_{k\perp} = 0$ into (29) yields

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} - S_k\left(C_k P_k C_k^{\mathrm{T}} + R_k\right)^{-1} S_k^{\mathrm{T}} + Q_k. \quad (94)$$

Subtracting (94) from (93) yields (91).

Alternatively, if Υ_k is square and Γ_k is not square, then

$$\pi_{k\perp} \neq 0, \quad \chi_{k\perp} = 0.$$

Substituting (74) and (63) into (76), and using $C_k = 0$ and $\chi_{k\perp} = 0$ yields

$$P^{\rm f}_{k+1} = A_k P^{\rm f}_k A_k^{\rm T} - S_k\left(C_k P^{\rm f}_k C_k^{\rm T} + R_k\right)^{-1} S_k^{\rm T} + Q_k. \quad (95)$$

Substituting $C_k = 0$ into (29) yields

$$P_{k+1} = A_k P_k A_k^{\rm T} + \pi_{k\perp} S_k\left(C_k P_k C_k^{\rm T} + R_k\right)^{-1} \times S_k^{\rm T}\pi_{k\perp}^{\rm T} - S_k\left(C_k P_k C_k^{\rm T} + R_k\right)^{-1} S_k^{\rm T} + Q_k. \quad (96)$$

Subtracting (95) from (96) yields (92). □

7. Comparison of the open-loop and closed-loop covariances

Next, we consider the zero-gain filter

$$\hat{x}_{{\rm ol},k+1} = A_k \hat{x}_{{\rm ol},k} + B_k u_k \quad (97)$$

with the zero-gain state-estimation error state

$$e_{{\rm ol},k} \triangleq x_k - \hat{x}_{{\rm ol},k}. \quad (98)$$

It follows from (1), (97) and (98) that

$$P_{{\rm ol},k+1} = A_k P_{{\rm ol},k} A_k^{\rm T} + Q_k, \quad (99)$$

where the zero-gain error covariance $P_{{\rm ol},k} \in \mathbb{R}^{n_k\times n_k}$ is defined by $P_{{\rm ol},k} \triangleq \mathcal{E}[e_{{\rm ol},k} e_{{\rm ol},k}^{\rm T}]$. First, we show that the performance of the Kalman filter is better than the performance of the zero-gain filter.

Proposition 10: *If* $\pi_k = I_{n_{k+1}}$ *and* $P_k \leqslant P_{{\rm ol},k}$, *then* $P_{k+1} \leqslant P_{{\rm ol},k+1}$.

Proof: Since $\pi_k = I_{n_{k+1}}$, it follows from (27) that $\pi_{k\perp} = 0$, and hence (29) implies that

$$P_{k+1} = A_k P_k A_k^{\rm T} + Q_k - \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T}. \quad (100)$$

Subtracting (100) from (99) yields

$$P_{{\rm ol},k+1} - P_{k+1} = A_k\left(P_{{\rm ol},k} - P_k\right)A_k^{\rm T} + \hat{S}_k \hat{R}_k^{-1} \hat{S}_k \geqslant 0. \quad □$$

If $\pi_k \neq I_{n_{k+1}}$, then $\pi_{k\perp} \neq 0$, and subtracting (29) from (99) yields

$$P_{{\rm ol},k+1} - P_{k+1} = A_k\left(P_{{\rm ol},k} - P_k\right)A_k^{\rm T} + \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} - \pi_{k\perp}\hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T}\pi_{k\perp}^{\rm T}, \quad (101)$$

which suggests the following negative result.

Proposition 11: *If* $\pi_k \neq I_{n_{k+1}}$ *and* $P_k = P_{{\rm ol},k}$, *then* $P_{k+1} \leqslant P_{{\rm ol},k+1}$ *is not always true.*

Proof: Let $k \geqslant 0$, $n_k = n_{k+1} = 2$, and

$$A_k = \begin{bmatrix} 0 & \alpha \\ 0 & 0.5 \end{bmatrix}, \quad C_k = [0\ 1],$$

where $24\alpha^2 + 2\alpha < 1$, and

$$Q_k = 0, \quad S_k = 0, \quad R_k = I, \quad L_k = I, \quad \Gamma_k = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Furthermore, let P_k and $P_{{\rm ol},k}$ have the scalar entries

$$P_k = \begin{bmatrix} p_{1,k} & p_{12,k} \\ p_{12,k} & p_{2,k} \end{bmatrix}, \quad P_{{\rm ol},k} = \begin{bmatrix} p_{{\rm ol},1,k} & p_{{\rm ol},12,k} \\ p_{{\rm ol},12,k} & p_{{\rm ol},2,k} \end{bmatrix}.$$

It follows from (29) and (99) that, if $P_k = P_{{\rm ol},k}$, then

$$p_{{\rm ol},1,k+1} - p_{1,k+1} = \left(\frac{24\alpha^2 + 2\alpha - 1}{25}\right)\frac{p_{2,k}^2}{1 + p_{2,k}}.$$

Hence, $p_{{\rm ol},1,k+1} < p_{1,k+1}$, and thus $P_{{\rm ol},k+1} - P_{k+1}$ is not positive semidefinite. □

The following result guarantees that the performance of the constrained filter is better than the performance of the zero-gain filter.

Proposition 12: *If* $P_k \leqslant P_{{\rm ol},k}$, *then*

$$\operatorname{tr}(P_{k+1}M_k) \leqslant \operatorname{tr}(P_{{\rm ol},k+1}M_k). \quad (102)$$

Proof: It follows from (27) and (101) that

$$\begin{aligned}\operatorname{tr}((P_{{\rm ol},k+1} - P_{k+1})M_k) = {} & \operatorname{tr}\left(A_k\left(P_{{\rm ol},k-P_k}\right)A_k^{\rm T} M_k\right) \\ & + \operatorname{tr}\left(\pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} M_k \right. \\ & + M_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} \pi_k^{\rm T} \\ & \left. - \pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} \pi_k^{\rm T} M_k\right). \quad (103)\end{aligned}$$

Since $\pi_k^{\rm T} M_k \pi_k = M_k \pi_k = \pi_k^{\rm T} M_k$, it follows that

$$\begin{aligned}&\operatorname{tr}((P_{{\rm ol},k+1} - P_{k+1})M_k) \\ &= \operatorname{tr}\left(A_k\left(P_{{\rm ol},k} - P_k\right)A_k^{\rm T} M_k\right) \\ &\quad + \operatorname{tr}\left(\pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} \pi_k^{\rm T} M_k\right) \\ &= \operatorname{tr}\left(L_k A_k\left(P_{{\rm ol},k} - P_k\right)A_k^{\rm T} L_k^{\rm T}\right) \\ &\quad + \operatorname{tr}\left(L_k \pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\rm T} \pi_k^{\rm T} L_k^{\rm T}\right) \\ &\geqslant 0\end{aligned}$$

□

In fact, in the example in Proposition 11, $M_k = I$ and

$$\mathrm{tr}(P_{\mathrm{ol},k+1}) - \mathrm{tr}(P_{k+1}) = \left[\frac{22}{25}\left(\alpha + \frac{3}{22}\right)^2 + \frac{5}{44}\right]\frac{p_{2,k}^2}{1+p_{2,k}} \geqslant 0. \tag{104}$$

Hence, $\mathrm{tr}(P_{k+1}) \leqslant \mathrm{tr}(P_{\mathrm{ol},k+1})$, and the one-step filter with constrained output injection performs better than the zero-gain filter. Although Proposition 12 guarantees that the performance of the one-step filter with constrained output injection is better than the zero-gain filter at time $k+1$, it follows from Proposition 11 that $P_{k+1} \leqslant P_{\mathrm{ol},k+1}$ may not be true. Hence, Proposition 12 does not guarantee that the performance of the one-step filter with constrained output injection is better than the zero-gain filter at time $k+2$, that is, $\mathrm{tr}(P_{k+2}) \leqslant \mathrm{tr}(P_{\mathrm{ol},k+2})$ may not be true.

The following result gives a condition under which the state estimates in the range of Γ_k are better than the corresponding estimates from the zero-gain filter.

Proposition 13: *If $P_k \leqslant P_{\mathrm{ol},k}$, then*

$$\Gamma_k^{\mathrm{T}} M_k P_{k+1} M_k \Gamma_k \leqslant \Gamma_k^{\mathrm{T}} M_k P_{\mathrm{ol},k+1} M_k \Gamma_k. \tag{105}$$

Proof: Note that

$$\begin{aligned}
&\Gamma_k^{\mathrm{T}} M_k (P_{k+1} - P_{\mathrm{ol},k+1}) M_k \Gamma_k \\
&\quad = \Gamma_k^{\mathrm{T}} M_k A_k (P_k - P_{\mathrm{ol},k}) A_k^{\mathrm{T}} M_k \Gamma_k \\
&\qquad - \Gamma_k^{\mathrm{T}} M_k \pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}} M_k \Gamma_k \\
&\qquad - \Gamma_k^{\mathrm{T}} M_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}} \pi_k^{\mathrm{T}} M_k \Gamma_k \\
&\qquad + \Gamma_k^{\mathrm{T}} M_k \pi_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}} \pi_k^{\mathrm{T}} M_k \Gamma_k.
\end{aligned} \tag{106}$$

It follows from (26) that

$$\Gamma_k^{\mathrm{T}} M_k \pi_k = \Gamma_k^{\mathrm{T}} M_k. \tag{107}$$

Substituting (107) into (106) yields

$$\begin{aligned}
&\Gamma_k^{\mathrm{T}} M_k (P_{k+1} - P_{\mathrm{ol},k+1}) M_k \Gamma_k \\
&\quad = \Gamma_k^{\mathrm{T}} M_k A_k (P_k - P_{\mathrm{ol},k}) A_k^{\mathrm{T}} M_k \Gamma_k \\
&\qquad - \Gamma_k^{\mathrm{T}} M_k \hat{S}_k \hat{R}_k^{-1} \hat{S}_k^{\mathrm{T}} M_k \Gamma_k \leqslant 0.
\end{aligned}$$ □

Assume that Γ_k has the form (6). Then, it follows from Proposition 13 that, if $M_k = I$, that is, all of the states are weighted, then the state estimate in the range of Γ_k obtained using the Kalman filter with constrained output injection are better than the state estimates obtained when data assimilation is not performed. However, state estimates that are not in the range of Γ_k may be worse than estimates obtained when no data assimilation is performed.

8. Steady-state filters for linear time-invariant systems

Next, we discuss the steady-state behaviour of the one-step spatially constrained Kalman filter for linear time-invariant systems. For all $k \geqslant 0$, let $A_k = A$, $B_k = B$, $C_k = C$, $\Gamma_k = \Gamma$, $L_k = L$, $Q_k = Q$, $S_k = 0$ and $R_k = R$. Assuming R is positive definite, it follows from Proposition 2 that the optimal gain K_k that minimizes J_k is given by

$$K_k = \left(\Gamma^{\mathrm{T}} M \Gamma\right)^{-1} \Gamma^{\mathrm{T}} M A P_k C^{\mathrm{T}} \hat{R}_k, \tag{108}$$

where

$$\hat{R}_k \triangleq C P_k C^{\mathrm{T}} + R, \quad M \triangleq L^{\mathrm{T}} L. \tag{109}$$

Furthermore, the covariance update is given by

$$\begin{aligned}
P_{k+1} &= A P_k A^{\mathrm{T}} + Q \\
&\quad + \pi_\perp A P_k C^{\mathrm{T}} \hat{R}_k^{-1} C P_k A^{\mathrm{T}} \pi_\perp^{\mathrm{T}} \\
&\quad - A P_k C^{\mathrm{T}} \hat{R}_k^{-1} C P_k A^{\mathrm{T}},
\end{aligned} \tag{110}$$

where

$$\pi \triangleq \Gamma\left(\Gamma^{\mathrm{T}} M \Gamma\right)^{-1} \Gamma^{\mathrm{T}} M, \quad \pi_\perp \triangleq I - \pi. \tag{111}$$

If $\lim_{k\to\infty} P_k$ exists, then the filtering process reaches statistical steady state. If Γ is square and thus by assumption non-singular, then $y_k - \hat{y}_k$ is directly injected into all of the estimator states. In this case, the following lemma guarantees the existence of $\lim_{k\to\infty} P_k$.

Lemma 3: *If Γ is square and (A, C) is detectable, then $P \triangleq \lim_{k\to\infty} P_k$ exists and is positive semidefinite. If, in addition, (A, Q) is stabilizable, then P is positive definite and $A - \Gamma K C$ is asymptotically stable, where $K \triangleq \Gamma^{-1} A P C^{\mathrm{T}} (C P C^{\mathrm{T}} + R)^{-1}$.*

Proof: Since Γ is square, it follows from (26) and (27) that $\pi = I$ and $\pi_\perp = 0$. Hence, it follows from (110) that

$$P_{k+1} = A P_k A^{\mathrm{T}} - A P_k C^{\mathrm{T}} \left(C P_k C^{\mathrm{T}} + R\right)^{-1} C P_k A^{\mathrm{T}} + Q. \tag{112}$$

Since (A, C) is detectable, it follows from Lewis (1986, pp. 100–101) that, if P_0 is positive semidefinite, then

$P \triangleq \lim_{k\to\infty} P_k$ exists and satisfies the algebraic Riccati equation

$$P = APA^{\mathrm{T}} - APC^{\mathrm{T}}\left(CPC^{\mathrm{T}} + R\right)^{-1}CPA^{\mathrm{T}} + Q. \quad (113)$$

If (A, C) is detectable and (A, Q) is stabilizable, it follows from Lewis (1986, pp. 101–103) that P is positive definite and $A - \Gamma KC$ is asymptotically stable. □

When Γ is not square, the existence of $\lim_{k\to\infty} P_k$ is not guaranteed. In fact, we have the following negative result when $\pi \neq I_n$.

Proposition 14: *Assume that $\pi \neq I_n$ and A is asymptotically stable. Then $\lim_{k\to\infty} P_k$ does not always exist.*

Proof: Consider the example in Proposition 11. It follows from (110) that

$$p_{2,k+1} = p_{2,k}\left(\frac{1}{4} + \frac{1}{100}\left[8(\alpha-1)^2 - 25\right]\frac{p_{2,k}}{1+p_{2,k}}\right). \quad (114)$$

Hence, if α satisfies

$$(\alpha-1)^2 > 25 \quad (115)$$

and

$$p_{2,0} > \frac{175}{8(\alpha-1)^2 - 200}, \quad (116)$$

then, for all $k > 0$, $p_{2,k+1} > 2p_{2,k}$, which implies that $\lim_{k\to\infty} p_{2,k} = \infty$. Hence, if $P_0 \in \mathbb{R}^{2\times 2}$ satisfies (116), then $\lim_{k\to\infty} P_k$ does not exist. □

Next, we present a converse result concerning the existence of $\lim_{k\to\infty} P_k$. For all $M \in \mathbb{R}^{n\times m}$, let $\mathcal{R}(M)$ denote the range of M.

Proposition 15: *Assume that (A, Γ) is stabilizable. If $P = \lim_{k\to\infty} P_k$ exists and $\mathcal{R}(\pi APC^{\mathrm{T}}) = \mathcal{R}(\Gamma)$, then (A, Γ, C) is output feedback stabilizable.*

Proof: Letting $k \to \infty$ in (110) yields

$$\begin{aligned} P = APA + Q + \pi_\perp APC^{\mathrm{T}}\hat{R}^{-1}CPA^{\mathrm{T}}\pi_\perp^{\mathrm{T}} \\ - APC^{\mathrm{T}}\hat{R}^{-1}CPA^{\mathrm{T}}, \end{aligned} \quad (117)$$

where $\hat{R} \triangleq CPC^{\mathrm{T}} + R$. We can rewrite (117) as

$$\begin{aligned} P = APA^{\mathrm{T}} + Q - \Gamma KCPA^{\mathrm{T}} \\ - APC^{\mathrm{T}}K^{\mathrm{T}}\Gamma^{\mathrm{T}} + \Gamma K\hat{R}K^{\mathrm{T}}\Gamma^{\mathrm{T}}, \end{aligned} \quad (118)$$

where

$$K \triangleq \left(\Gamma^{\mathrm{T}}M\Gamma\right)^{-1}\Gamma^{\mathrm{T}}MAPC^{\mathrm{T}}\hat{R}^{-1}. \quad (119)$$

Hence, (118) can be expressed as

$$P = (A - \Gamma KC)P(A - \Gamma KC)^{\mathrm{T}} + Q + \Gamma KRK^{\mathrm{T}}\Gamma^{\mathrm{T}}. \quad (120)$$

Next, define $\tilde{A}$ and $\tilde{\Gamma}$ by

$$\tilde{A} \triangleq A - \Gamma KC, \quad \tilde{\Gamma} \triangleq \Gamma KR^{1/2}. \quad (121)$$

Since (A, Γ) is stabilizable and $\mathcal{R}(\Gamma) = \mathcal{R}(\pi APC^{\mathrm{T}})$, it follows from Bernstein (2005, pp. 510 and 551) that $(\tilde{A}, \tilde{\Gamma})$ is also stabilizable. Let $\lambda \in \mathbb{C}$ be an eigenvalue of $\tilde{A}$. Then, there exists an eigenvector $x \in \mathbb{C}^n$ of $\tilde{A}$ such that

$$x^*\tilde{A} = \lambda x^*. \quad (122)$$

Furthermore, (120) implies that

$$x^*Px = x^*\tilde{A}P\tilde{A}^{\mathrm{T}}x + x^*\left(Q + \tilde{\Gamma}\tilde{\Gamma}^{\mathrm{T}}\right)x. \quad (123)$$

Substituting (122) into (123) yields

$$(1 - |\lambda|^2)x^*Px = x^*\left(Q + \tilde{\Gamma}\tilde{\Gamma}^{\mathrm{T}}\right)x. \quad (124)$$

If $|\lambda| \geqslant 1$, then (124) implies that

$$x^*\left(Q + \tilde{\Gamma}\tilde{\Gamma}^{\mathrm{T}}\right)x = 0 \quad (125)$$

and hence

$$x^*\tilde{\Gamma} = 0. \quad (126)$$

It follows from (122) and (126) that λ is an unstable and uncontrollable eigenvalue of $(\tilde{A}, \tilde{\Gamma})$, which contradicts the fact that $(\tilde{A}, \tilde{\Gamma})$ is stabilizable. Hence, $|\lambda| < 1$ and $\tilde{A}$ is asymptotically stable. Since K given by (119) renders $A - \Gamma KC$ asymptotically stable, (A, Γ, C) is output feedback stabilizable. □

The following result provides a sufficient condition for P_k to be bounded when C is square and non-singular.

Proposition 16: *Assume that C is square and non-singular. If*

$$\mathrm{sprad}(\pi_\perp A) < 1, \quad (127)$$

then P_k is bounded.

Proof: Since C is non-singular, (110) can be expressed as

$$\begin{aligned} P_{k+1} = AP_kA^{\mathrm{T}} + Q + \pi_\perp AP_k\left(P_k + C^{-1}RC^{-\mathrm{T}}\right)^{-1}P_kA^{\mathrm{T}}\pi_\perp^{\mathrm{T}} \\ - AP_k\left(P_k + C^{-1}RC^{-\mathrm{T}}\right)^{-1}P_kA^{\mathrm{T}}. \end{aligned} \quad (128)$$

Next, consider the Lyapunov equation

$$\tilde{P}_{k+1} = \left(A - \Gamma\tilde{K}\right)\tilde{P}_k\left(A - \Gamma\tilde{K}\right)^{\mathrm{T}} + Q + \Gamma\tilde{K}\tilde{K}^{\mathrm{T}}\Gamma^{\mathrm{T}} + A\tilde{R}A^{\mathrm{T}}, \tag{129}$$

where

$$\tilde{K} \triangleq \left(\Gamma^{\mathrm{T}}M\Gamma\right)^{-1}\Gamma MA \tag{130}$$

and

$$\tilde{R} \triangleq C^{-1}RC^{-\mathrm{T}}. \tag{131}$$

Using (130), we rewrite (129) as

$$\tilde{P}_{k+1} = \pi_\perp A\tilde{P}_kA^{\mathrm{T}}\pi_\perp^{\mathrm{T}} + Q + \pi AA^{\mathrm{T}}\pi^{\mathrm{T}} + A\tilde{R}A^{\mathrm{T}}. \tag{132}$$

Since $\pi_\perp A$ is asymptotically stable and $Q + \pi AA^{\mathrm{T}}\pi^{\mathrm{T}} + A\tilde{R}A^{\mathrm{T}}$ is positive semidefinite, $\tilde{P} = \lim_{k\to\infty}\tilde{P}_k$ exists for all positive-semidefinite $\tilde{P}_0$. Subtracting (128) from (132) yields

$$\begin{aligned}\tilde{P}_{k+1} - P_{k+1} = {} & A\tilde{R}\left(\tilde{R} + P_k\right)^{-1}\tilde{R}A^{\mathrm{T}} + \pi AA^{\mathrm{T}}\pi^{\mathrm{T}} \\ & + \pi_\perp AP_k\left(P_k + \tilde{R}\right)^{-1}\tilde{R}A^{\mathrm{T}}\pi_\perp^{\mathrm{T}} \\ & + \pi_\perp A\left(\tilde{P}_k - P_k\right)A^{\mathrm{T}}\pi_\perp^{\mathrm{T}}.\end{aligned} \tag{133}$$

It follows from (133) that, if $\tilde{P}_k \geqslant \tilde{P}_k$, then $\tilde{P}_{k+1} \geqslant P_{k+1}$. Hence, if $P_0 \leqslant \tilde{P}_0$, then $P_k \leqslant \tilde{P}_k$ for all $k > 0$. Furthermore, since $\tilde{P}_k$ converges to $\tilde{P}$ for every choice of $\tilde{P}_0$, it follows that P_k is bounded. □

Numerical results suggest that the following strengthening of Proposition 15 is true.

Conjecture 1: *Assume that C is square and non-singular. If*

$$\mathrm{sprad}(\pi_\perp A) < 1, \tag{134}$$

then $\lim_{k\to\infty} P_k$ *exists.*

Example 1: Let

$$A = \begin{bmatrix} 0 & 5 \\ 0 & 3 \end{bmatrix}, \quad C = I, \quad Q = 0, \quad R = I, \tag{135}$$

and choose

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \tag{136}$$

where $\gamma_1, \gamma_2 \in \mathbb{R}$ so that

$$\pi = \frac{1}{\gamma_1^2 + \gamma_2^2}\begin{bmatrix} \gamma_1^2 & \gamma_1\gamma_2 \\ \gamma_1\gamma_2 & \gamma_2^2 \end{bmatrix}, \quad \pi_\perp = \frac{1}{\gamma_1^2 + \gamma_2^2}\begin{bmatrix} \gamma_2^2 & -\gamma_1\gamma_2 \\ -\gamma_1\gamma_2 & \gamma_1^2 \end{bmatrix} \tag{137}$$

Note that

$$\pi_\perp A = \frac{1}{\gamma_1^2 + \gamma_2^2}\begin{bmatrix} 0 & 5\gamma_2^2 - 3\gamma_1\gamma_2 \\ 0 & 3\gamma_1^2 - 5\gamma_1\gamma_2 \end{bmatrix} \tag{138}$$

and hence

$$\mathrm{sprad}(\pi_\perp A) = \frac{1}{\gamma_1^2 + \gamma_2^2}\left|3\gamma_1^2 - 5\gamma_1\gamma_2\right|. \tag{139}$$

It follows from Conjecture 1 that, if

$$-\left(\gamma_1^2 + \gamma_2^2\right) < 3\gamma_1^2 - 5\gamma_1\gamma_2 < \gamma_1^2 + \gamma_2^2, \tag{140}$$

then $\lim_{k\to\infty} P_k$ exists. The shaded region in figure 1 indicates values of γ_1 and γ_2 that satisfy (140). Next, we choose various values of γ_1, γ_2 and numerically evaluate P_k as $k \to \infty$ using (110). The values of γ_1, γ_2 for which $\lim_{k\to\infty} P_k$ exists, are indicated by "•" and the values of γ_1, γ_2 for which $\lim_{k\to\infty} P_k$ does not exist are indicated by "×". The numerical results are consistent with Lemma 3.

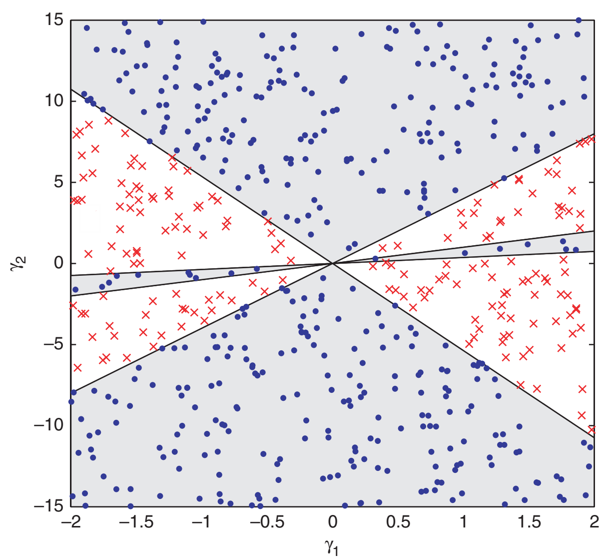

Figure 1. The shaded region indicates the values of γ_1, γ_2 that satisfy (140). The dots indicate the values of γ_1, γ_2 for which $\lim_{k\to\infty} P_k$ exists, whereas the values of γ_1, γ_2 for which $\lim_{k\to\infty} P_k$ does not exist are indicated by "×". These numerical results are consistent with Conjecture 1.

9. N-mass system example

Consider the N-mass system shown in figure 2 with stiffnesses $k_1, \ldots, k_{N+1} > 0$ and dashpots $c_1, \ldots, c_{N+1} > 0$. Let q_i denote the position of mass m_i. Define

$$q \triangleq [q_1 \cdots q_N]^{\mathrm{T}}, \quad M \triangleq \mathrm{diag}(m_1, \ldots, m_N). \tag{141}$$

$$K \triangleq \begin{bmatrix} k_1+k_2 & -k_2 & 0 & \cdots & 0 & 0 \\ -k_2 & k_2+k_3 & -k_3 & \cdots & 0 & 0 \\ 0 & -k_3 & k_3+k_4 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -k_N & k_N+k_{N+1} \end{bmatrix}, \tag{142}$$

$$C \triangleq \begin{bmatrix} c_1+c_2 & -c_2 & 0 & \cdots & 0 & 0 \\ -c_2 & c_2+c_3 & -c_3 & \cdots & 0 & 0 \\ 0 & -c_3 & c_3+c_4 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -c_N & c_N+c_{N+1} \end{bmatrix}. \tag{143}$$

We assume that d masses are disturbed by unknown force inputs $w \in \mathbb{R}^d$, which are zero-mean white noise with unit intensity, while p masses are actuated by known force inputs $u \in \mathbb{R}^p$. Let u and w have entries

$$u = \left[u_1 \cdots u_p\right]^{\mathrm{T}}, \quad w \triangleq [w_1 \cdots w_d]^{\mathrm{T}} \tag{144}$$

and let $\mathcal{B}_u$ and $\mathcal{D}_w$ have entries

$$\mathcal{B}_u = \left[\mathcal{B}_{u,1} \cdots \mathcal{B}_{u,p}\right], \quad \mathcal{D}_w = \left[\mathcal{D}_{w,1} \cdots \mathcal{D}_{w,d}\right], \tag{145}$$

where, for all $i = 1, \ldots, p$ and $j = 1, \ldots, d$, $\mathcal{B}_{u,i}$ and $\mathcal{D}_{w,j}$ are defined by

$$\mathcal{B}_{u,i} = \left[0_{1\times\hat{i}-1} \frac{1}{m_{\hat{i}}} 0_{1\times N-\hat{i}}\right]^{\mathrm{T}}, \quad \mathcal{D}_{w_j} = \left[0_{1\times\hat{j}-1} \frac{1}{m_{\hat{j}}} 0_{1\times N-\hat{j}}\right]^{\mathrm{T}} \tag{146}$$

and $\hat{i}$ and $\hat{j}$ correspond to the masses on which forces u_i and w_j act, respectively. The equations of motion can be written in first-order form as

$$\dot{x} = \mathcal{A}x + \mathcal{B}u + \mathcal{D}_1 w, \tag{147}$$

where $\mathcal{A} \in \mathbb{R}^{2N\times 2N}$, $\mathcal{B} \in \mathbb{R}^{2N\times m}$, $\mathcal{D}_1 \in \mathbb{R}^{2N\times d}$, and $x \in \mathbb{R}^{2N}$ are defined by

$$\mathcal{A} \triangleq \begin{bmatrix} 0_N & I_N \\ -M^{-1}K & -M^{-1}C \end{bmatrix}, \quad \mathcal{B} \triangleq \begin{bmatrix} 0_N \\ \mathcal{B}_u \end{bmatrix}, \quad \mathcal{D}_1 \triangleq \begin{bmatrix} 0_N \\ \mathcal{D}_w \end{bmatrix}, \quad x \triangleq [q_1 \cdots q_N \, \dot{q}_1 \cdots \dot{q}_N]^T. \tag{148}$$

Next, we assume that measurements of the positions of l masses are available so that the output $y \in \mathbb{R}^l$ can be expressed as

$$y = C_{\mathrm{pos}}x + v, \tag{149}$$

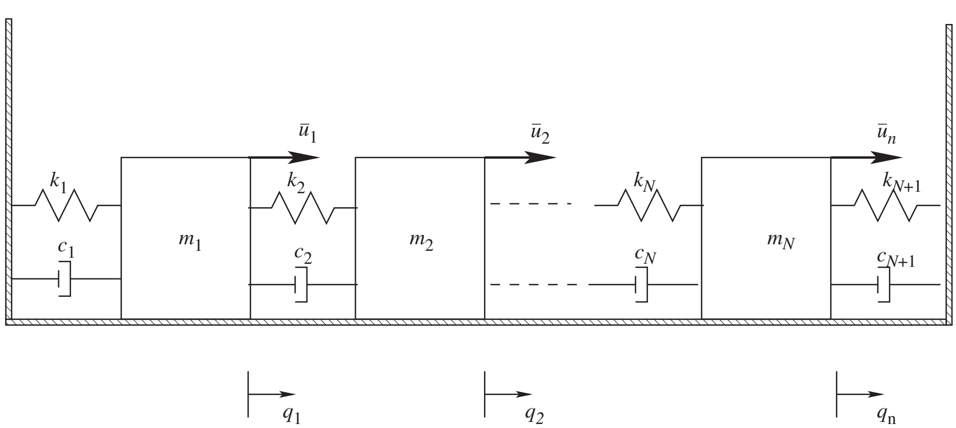

Figure 2. N-mass system.

where $C_{\text{pos}} \in \mathbb{R}^{l\times 2N}$ has entries

$$C_{\text{pos}} = \begin{bmatrix} C_{\text{pos}}^{[1]} \\ \vdots \\ C_{\text{pos}}^{[l]} \end{bmatrix} \tag{150}$$

and, for all $i = 1, \ldots, N$, $C_{\text{pos}}^{[i]} \in \mathbb{R}^{1\times 2N}$ is defined by

$$C_{\text{pos}}^{[i]} \triangleq \left[0_{1\times(\hat{i}-1)} \quad 1 \quad 0_{1\times(N-\hat{i})} \quad 0_{1\times N} \right], \tag{151}$$

where $\hat{i}$ corresponds to the index of the mass whose position measurements are available. With the sampling time $t = 0.1$ s, we obtain the zero-order-hold discrete-time model of (147) and (149) given by

$$x_{k+1} = Ax_k + Bu_k + D_1 w_k, \tag{152}$$

$$y_k = C_{\text{pos}} x_k + v_k. \tag{153}$$

Let $N = 20$, so that the (147) has order $n = 40$ with known inputs $u \in \mathbb{R}^3$ and unknown inputs $w \in \mathbb{R}^3$. We assume that w is zero-mean white Gaussian noise with unit covariance, and the known inputs $u \in \mathbb{R}^3$ are chosen to be sinusoids. The masses on which w and u act and the available measurements are given in table 1. We assume that the process noise and the measurement sensor noise are uncorrelated and hence $S_k = 0$. The values of the masses $m_1, \ldots, m_{20}$, damping coefficients $c_1, \ldots, c_{21}$, and spring constants $k_1, \ldots, k_{21}$ are $m_i = 10$ kg for $i = 1, \ldots, 20$, $c_i = 0.8$ N s/m and $k_i = 5$ N/m for $i = 1, \ldots, 21$. Finally, we assume that

Table 1. Forcing and measurement signals in the N-mass system.

Signal	Masses
Known force input u	m_1, m_5, m_{10}
Unknown force input w	m_4, m_{15}, m_{18}
Position measurement y	m_9, m_{12}

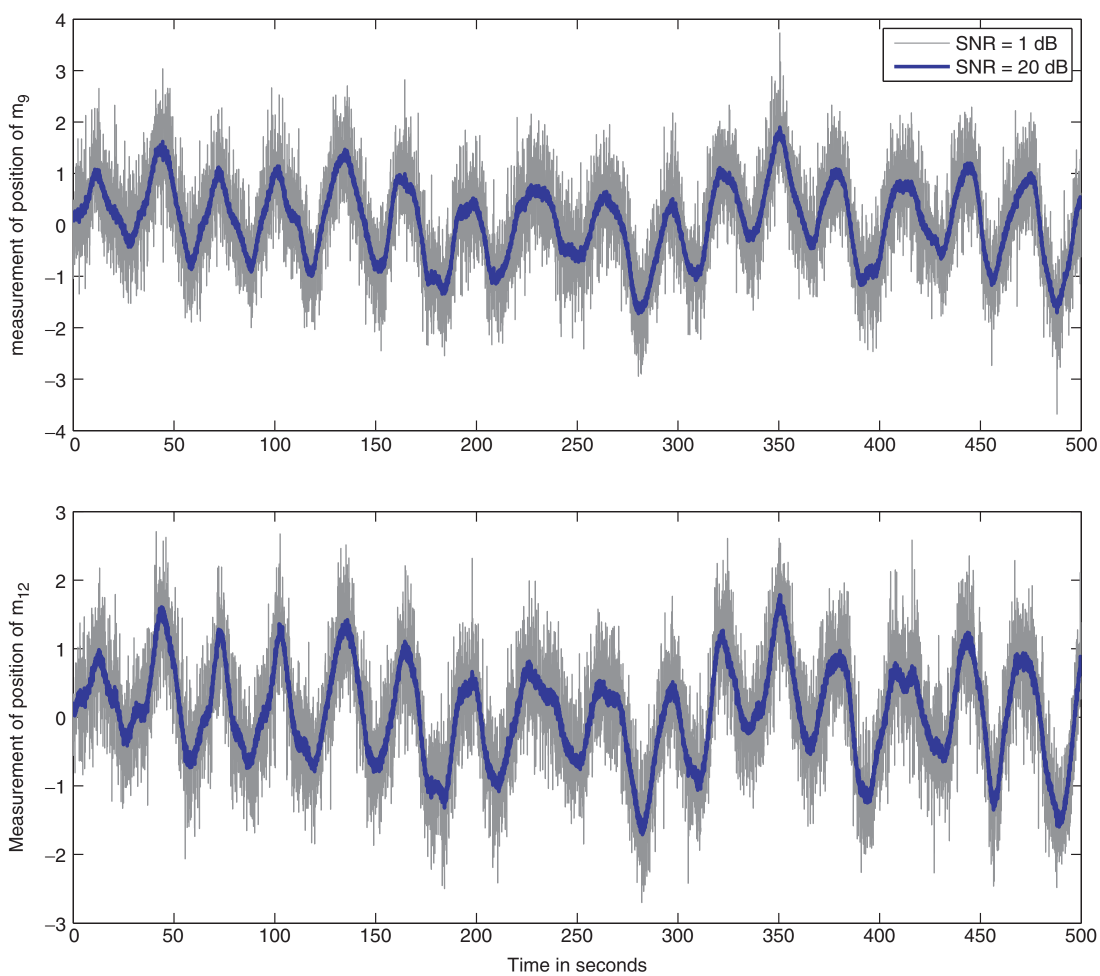

Figure 3. Noisy measurements of the positions of m_9 and m_{12} with SNR = 20 db and SNR = 1 dB. These measurements are used to estimate the positions and velocities of masses $m_1, \ldots, m_{20}$.

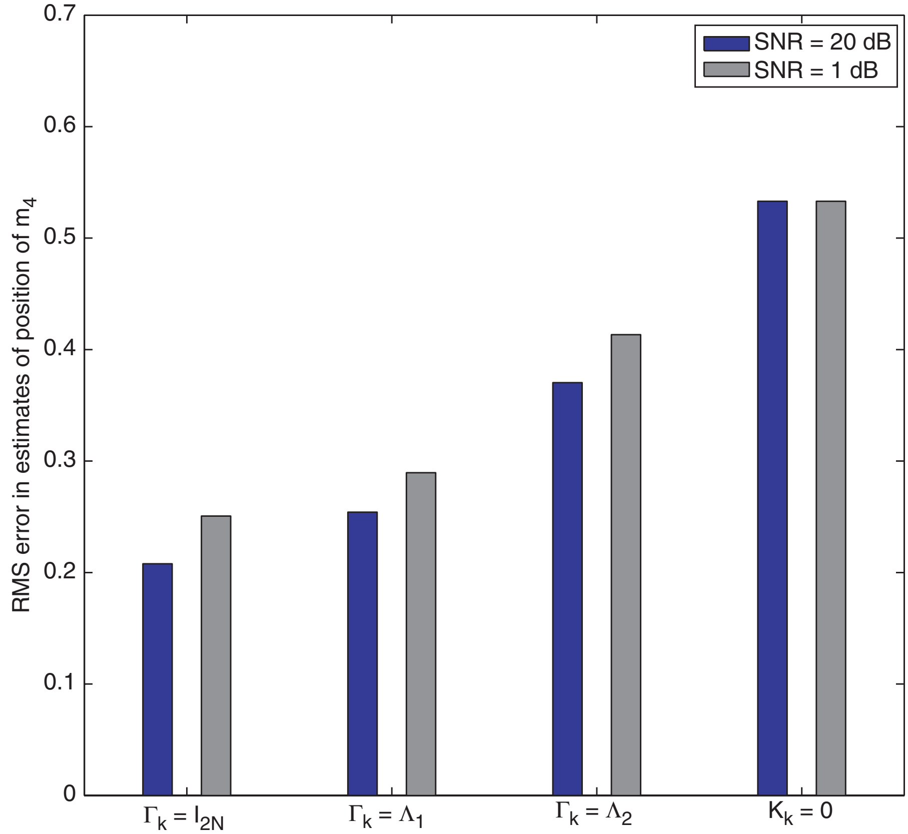

Figure 4. Root mean square value of the error in estimating the position of m4 obtained using the two-step filter with $\Gamma_k = I_{2N}$ (classical Kalman filter) and $\Gamma_k \neq I_{2N}$ using two different sets of measurements, one with SNR = 20 dB and another with SNR = 1 dB. When $\Gamma_k \neq \Lambda_1$, measurements are directly injected into the estimates of only the positions and velocities of masses $m_5, \ldots, m_{16}$, whereas when $\Gamma_k \neq \Lambda_2$, measurements are directly injected into estimates of only the positions and velocities of masses $m_g, \ldots, m_{12}$. As expected, the performance of the estimators with constrained output injection ($\Gamma_k \neq I$) is not as good as the estimator with $\Gamma_k = I_{2N}$. Since the zero-gain filter does not use the measurements, its performance does not depend on the value of the SNR of the measurement.

the process noise and sensor noise are uncorrelated, that is, $S_k = 0$ for all $k \geqslant 0$. Next, we obtain estimates of the position and velocity of $m_1, \ldots, m_{20}$ using two sets of measurements y, one with a signal to noise ratio (SNR) of 20 dB and another with a SNR of 1 dB. The measurements of position of m_9 and m_{12} with different signal to noise rations are shown in figure 3.

We first choose $\Gamma_k = I_{2N}$ and $L_k = I_{2N}$, that is, the available measurements are injected into all of the states of the estimator, and the errors between all of the states and the corresponding state estimates are weighted. In this case, the one-step and two-step Kalman filters are equivalent. The state estimates are obtained using the two-step filter (72)–(75). The root mean square (RMS) value of the error in the estimates of position of m_4 when measurements with a signal to noise ratio of 20 dB and 1 dB, respectively, are used is shown in figure 4. The RMS value of the errors in position and velocity estimates of $m_1, \ldots, m_{20}$ are plotted in figures 5 and 6, respectively.

Next, we obtain estimates by constraining the output injection into only some of the states of the estimator. First, we choose $\Gamma_k = \Lambda_1$ for all $k \geqslant 0$, where

$$\Lambda_1 \triangleq [0_{24\times 8} \quad I_{24} \quad 0_{24\times 8}]^{\mathrm{T}} \tag{154}$$

so that the measurements are injected into only the estimates of the positions and velocities of $m_5, \ldots, m_{16}$. Furthermore, we choose $L_k = I_{2N}$ so that the errors in all of the state estimates are weighted equally. The RMS value of the error in the position estimate of m_4 obtained when $\Gamma_k = \Lambda_1$ for all $k \geqslant 0$ is shown in figure 4. The RMS value of the errors in position and velocity estimates of $m_1, \ldots, m_{20}$, are shown in figures 5 and 6, respectively. Finally, we choose $\Gamma_k = \Lambda_2$ for all $k \geqslant 0$, where

$$\Lambda_2 \triangleq [0_{8\times 16} \quad I_8 \quad 0_{8\times 16}]^{\mathrm{T}} \tag{155}$$

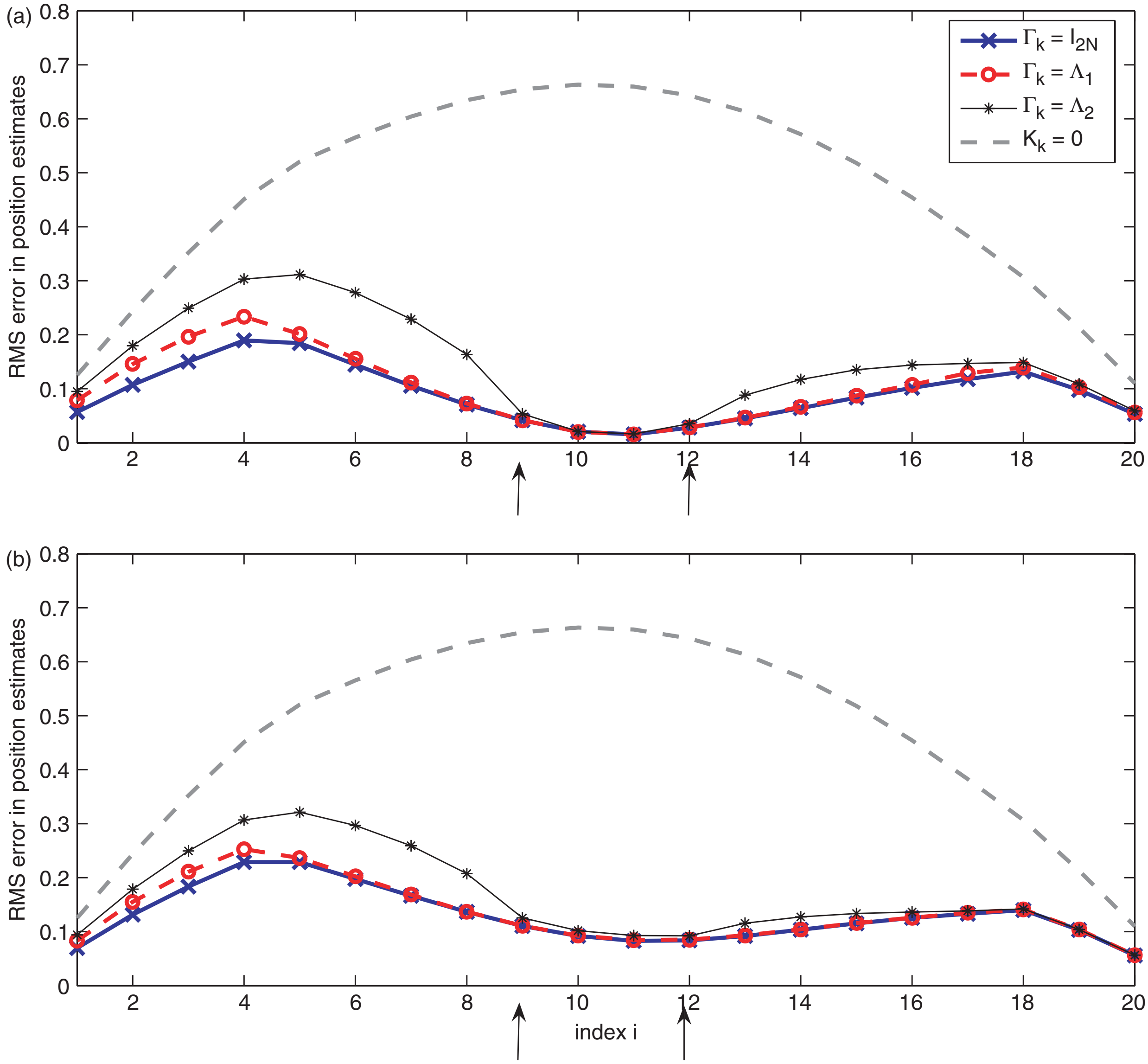

Figure 5. RMS value of the errors in the position estimates of all of the masses when measurements with (a) SNR = 20 dB and (b) SNR = 1 dB are injected into all of the state estimates ($\Gamma_k = I_{2N}$) and when measurements are injected into only the position and velocity estimates of some of the masses ($\Gamma_k \neq I_{2N}$). The performance of the zero-gain filter with $K_k \equiv 0$ is also shown for comparison. When measurements are injected into a larger number of the estimator states, the performance of the estimator improves. The arrows indicate the masses whose position measurements are available. As the SNR of the measurement increases, the difference in the performance of the filters with $\Gamma_k = I_{2N}$ and $\Gamma_k \neq I_{2N}$ decreases.

so that only the estimates of the positions and velocities of $m_9, \ldots, m_{12}$ are directly affected by the measurements y. Again, we choose $L_k = I_{2N}$ for all $k \geqslant 0$, and the performance of the estimator with $\Gamma_k = \Lambda_2$ for all $k \geqslant 0$ is shown in figures 4–6.

When $\Gamma_k = I_{2N}$, the measurements are injected directly into all of the states of the estimator, and figure 4 confirms the expected fact that the performance of the classical Kalman filter with $\Gamma_k = I_{2N}$ is better than the estimators with $\Gamma_k \neq I_{2N}$. Note that the number of states into which measurements are injected when $\Gamma_k = \Lambda_2$ is less than the number of states that are directly affected by measurements when $\Gamma_k = \Lambda_1$, and it follows from figure 4 that reducing the number of estimator states that are directly affected by measurements degrades the performance of the estimator. These observations are consistent with Proposition 6.

Although the errors in the position and velocity estimates of all of the masses are weighted in all three cases $\Gamma_k = I_{2N}$, $\Gamma_k = \Lambda_1$, and $\Gamma_k = \Lambda_2$, figures 5 and 6 demonstrate that the error in the position and velocity estimates of all of the masses is the least when $\Gamma_k = I_{2N}$ and the measurements are directly injected

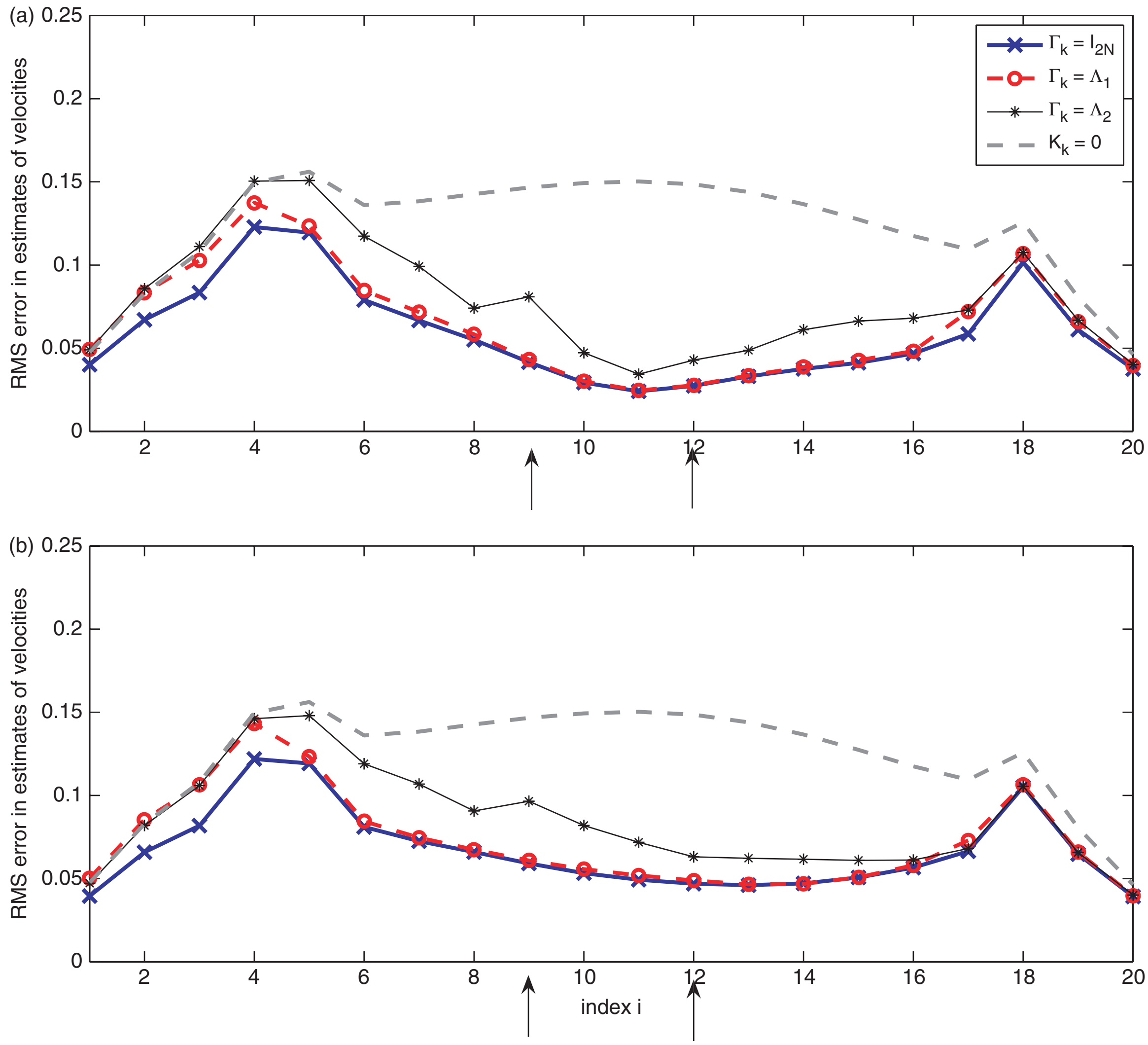

Figure 6. RMS value of the errors in the velocity estimates from the optimal filter with $\Gamma_k = I_{2N}$ and $\Gamma_k \neq I_{2N}$ when measurements with (a) SNR = 20 dB and (b) SNR = 1 dB are used. When $\Gamma_k \neq I_{2N}$, the one-step and two-step filters are not equivalent, and the results presented here are obtained using the two-step estimator. The performance of the estimators with $\Gamma_k \neq I_{2N}$ improves when additional states of the estimator are directly injected with measurements.

into all of the estimator states. Finally, it can be seen that when the measurements are injected into a subset of the estimator states, then the estimates of the states that are not directly affected by the measurements improve. The performance of the zero-gain filter with $K_k = 0$ for all $k \geqslant 0$ is also plotted in figures 4–6 for comparison.

10. Conclusions

This paper presents an extension of the Kalman filter that constrains data injection into only a specified subset of state estimates rather than the entire state estimate. This extension accounts for correlation between the process noise and the sensor noise. Conditions are given under which the one-step and two-step forms of the filter are equivalent. Future work will consider reduced-rank square root formulations of this filter to reduce the computational burden of propagating the covariance. More general conditions that guarantee the existence of a steady-state covariance for linear time-invariant dynamics are also of interest.

Acknowledgements

This research was supported by the National Science Foundation Information Technology Research

Initiative, through Grant ATM-0325332 to the University of Michigan, Ann Arbor, USA.

References

A. Gelb, *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, MA, 1974.

D.S. Bernstein and D.C. Hyland, "The optimal projection equations for reduced-order state estimation", *IEEE Trans. Autom. Contr.*, AC-30, pp. 583–585, 1985.

P. Hippe and C. Wurmthaler, "Optimal reduced-order estimators in the frequency domain: the discrete-time case", *Int. J. Contr.*, 52, pp. 1051–1064, 1990.

W.M. Haddad and D.S. Bernstein, "Optimal reduced-order observer-estimators", *AIAA J. Guid. Dyn. Contr.*, 13, pp. 1126–1135, 1990.

C.-S. Hsieh, "The unified structure of unbiased minimum-variance reduced-order filters", *Proc. Contr. Dec. Conf.*, pp. 4871–4876, Maui, HI, December 2003.

B.F. Farrell and P.J. Ioannou, "State estimation using a reduced-order Kalman filter", *J. Atmos. Sci.*, 58, pp. 3666–3680, 2001.

A.W. Heemink, M. Verlaan and A.J. Segers, "Variance reduced ensemble Kalman filtering", *Mon. Weather Rev.*, 129, pp. 1718–1728, 2001.

J. Ballabrera-Poy, A.J. Busalacchi and R. Murtugudde, "Application of a reduced-order Kalman filter to initialize a coupled atmosphere-ocean model: impact on the prediction of El Nino", *J. Climate*, 14, pp. 1720–1737, 2001.

P. Fieguth, D. Menemenlis and I. Fukumori, "Mapping and pseudo-inverse algorithms for ocean data assimilation", *IEEE Trans. Geoscience and Remote Sensing*, 41, pp. 43–51, 2003.

O. Barrero, D.S. Bernstein and B.L.R. De Moor, "Spatially localized Kalman filtering for data assimilation", *Proc. Amer. Contr. Conf.*, Portland, pp. 3468–3473, 2005.

D.I. Lawrie, P.J. Fleming, G.W. Irwin and S.R. Jones, "Kalman filtering: a survey of parallel processing alternatives", *Proc. IFAC Workshop on Algorithms and Architectures for Real-Time Control*, Pergamon, Seoul, South Korea, pp. 49–56, 1992.

F.L. Lewis, *Optimal Estimation*, John Wiley and Sons, New York, USA, 1986.

L. Scherliess, R.W. Schunk, J.J. Sojka and D.C. Thompson, "Development of a physics-based reduced state Kalman filter for the ionosphere", *Radio Science*, 39-RS1S04, pp. 1–22, 2004.

D.S. Bernstein, *Matrix Mathematics*, Princeton University Press, Princeton, USA, 2005.