

## Research Article

# Optimal Fusion Filtering in Multisensor Stochastic Systems with Missing Measurements and Correlated Noises

R. Caballero-Águila,<sup>1</sup> I. García-Garrido,<sup>2</sup> and J. Linares-Pérez<sup>2</sup>

<sup>1</sup>Departamento de Estadística, Universidad de Jaén, Paraje Las Lagunillas, 23071 Jaén, Spain

<sup>2</sup>Departamento de Estadística, Universidad de Granada, Avenida Fuentenueva, 18071 Granada, Spain

Correspondence should be addressed to R. Caballero-Águila; raguila@ujaen.es

Received 30 January 2013; Accepted 28 April 2013

Academic Editor: Weihai Zhang

Copyright © 2013 R. Caballero-Águila et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The optimal least-squares linear estimation problem is addressed for a class of discrete-time multisensor linear stochastic systems with missing measurements and autocorrelated and cross-correlated noises. The stochastic uncertainties in the measurements coming from each sensor (missing measurements) are described by scalar random variables with arbitrary discrete probability distribution over the interval  $[0, 1]$ ; hence, at each single sensor the information might be partially missed and the different sensors may have different missing probabilities. The noise correlation assumptions considered are (i) the process noise and all the sensor noises are one-step autocorrelated; (ii) different sensor noises are one-step cross-correlated; and (iii) the process noise and each sensor noise are two-step cross-correlated. Under these assumptions and by an innovation approach, recursive algorithms for the optimal linear filter are derived by using the two basic estimation fusion structures; more specifically, both centralized and distributed fusion estimation algorithms are proposed. The accuracy of these estimators is measured by their error covariance matrices, which allow us to compare their performance in a numerical simulation example that illustrates the feasibility of the proposed filtering algorithms and shows a comparison with other existing filters.

## 1. Introduction

For a long time, the least-squares (LS) estimation problem in linear stochastic systems from measurements perturbed by additive noises has received considerable attention in the scientific community due to its wide applicability in many practical situations (e.g., video and laser tracking systems, satellite navigation, radar and meteorological applications, etc. [1]). As it is well known, one of the major contributions made to solve this problem is the Kalman filter, which provides a recursive algorithm for the optimal LS estimator when the additive white noises and the initial state are Gaussian and mutually independent (or, equivalently, uncorrelated due to the Gaussianity assumption) and, therefore, the optimal LS estimator is the optimal LS linear estimator. From the publication of the Kalman filter [2] in 1960, numerous results and several solution methods have been reported in the literature to address the state estimation problem from noisy observations, which depend on models representing possible

relationships between the unknown state and the observable variables and also on the noise processes assumptions.

Specifically, during the past decades, there has been an increasing interest in the filtering problem in multisensor systems, where sensor networks are used to obtain the whole available information on the system state and its estimation must be carried out from the observations provided by all the sensors. A basic matter for this class of systems is how to fuse the measurement data from the different sensors to address the estimation problem. Commonly, two methods are used to process the measured data coming from multiple sensors: centralized and distributed fusion methods. In the centralized fusion method all the measured data from sensors are communicated to the fusion center for being processed; nevertheless, as is widely known, centralized estimators have many computational disadvantages, which motivate the research into other fusion methods. In the distributed fusion method, each sensor estimates the state based on its own single measurement data, and then it sends such

estimate to the fusion center for fusion according to a certain information fusion criterion. Although the use of sensor networks offers several advantages, the unreliable network characteristics usually cause problems during data transmission from sensors to the fusion center, such as missing measurements, random communication packet losses and/or delays. Taking into account these network uncertainties, the models representing the relationships between the state and measurements do not allow to apply the Kalman filter, and modifications of conventional estimation algorithms have been proposed (see e.g., [3–9] and references therein).

As in the Kalman filter, independent white noises are considered in all the mentioned papers; however, this assumption may not be realistic and can be a limitation in many real-world problems in which noise correlation may be present. This problem arises, for example, when a target is taking an electronic countermeasure, for example, noise jamming [10], or if the process noise and the sensor measurement noises are dependent on the system state, then there may be cross-correlation between different sensor noises and cross-correlation between process noise and sensor noises. Also, if all the sensors are observed in the same noisy environment, the measurement noises of different sensors are usually correlated.

For these reasons, the estimation problem in systems with correlated noises has received significant research interest in recent years. For example, the optimal Kalman filtering fusion problem in systems with cross-correlated sensor noises is addressed in [10], while [11, 12] study the same problem in systems with cross-correlated process noises and measurement noises; in these papers correlated noises at the same sampling time are considered. In general, the assumption of correlation and cross-correlation of the noise process and measurement noises in different sampling times makes difficult the identification of optimal estimators; this limitation has encouraged a wider research into suboptimal Kalman-type estimation problems. In [13], a Kalman-type recursive filter is presented for systems with finite-step correlated process noises, and the filtering problem with multistep correlated process and measurement noises is investigated in [14]. The optimal robust nonfragile Kalman-type recursive filtering problem is studied in [15] for a class of uncertain systems with finite-step autocorrelated measurement noises and multiple packet dropouts. The problem of distributed weighted robust Kalman filter fusion is studied in [16] for a class of uncertain systems with autocorrelated and cross-correlated noises. In [17], a stochastic singular system with correlated noises at the same sampling time is transformed into an equivalent nonsingular system with correlated noises at the same and neighboring sampling times. Also, in [18], an augmented parameterized system with correlated noises at the same and neighboring sampling times is used to describe the sensor delay, packet dropout, and uncertain observation phenomena.

On the other hand, as noted above, the use of communication networks for transmitting measured data motivates the need of considering stochastic uncertainties. Missing measurements have been widely treated due to its applicability to model a large class of real-world problems, such as fading

phenomena in propagation channels, target tracking or, in general, situations where there exist intermittent failures in the observation mechanism, accidental loss of some measurements, or inaccessibility of the data during certain times. The state estimation problem from missing measurement transmitted by multiple sensors has been studied based on the assumption that all the sensors are identical (see, e.g., [19–22]); however, this assumption can be unreasonable since some real systems usually involve multiple sensors with different characteristics. Recently, the filtering problem using missing measurements whose statistical properties are assumed not to be the same in all the sensors has been addressed by several authors under different approaches and hypotheses on the processes involved (see, e.g., [23–27]). In all the above papers, Bernoulli random variables are used to model the missing measurements phenomenon, and hence, it is assumed that the measurement signal is either completely lost (if the corresponding Bernoulli variable takes the value zero) or successfully transferred (when the Bernoulli variable is equal to one). Recently, this missing measurement model has been generalized considering any discrete distribution on the interval  $[0, 1]$ , which allows to cover some practical applications where only partial information is missing (see [28, 29] and references therein).

Motivated by the above considerations, our attention is focused on investigating the optimal LS linear centralized and distributed fusion estimation problems in multisensor systems with missing measurements and autocorrelated and cross-correlated noises. In each sensor, the missing measurement phenomenon is governed by a scalar random variable with arbitrary discrete probability distribution over the interval  $[0, 1]$ , and the different sensors may have different missing probabilities. Assume that the process noise and all the sensor noises are one-step autocorrelated; different sensor noises are one-step cross-correlated; and the process noise and each sensor noise are two-step cross-correlated. This paper makes a twofold substantial novel contribution: (1) unlike most previous results with correlated noises, in which suboptimal Kalman-type estimators are proposed, in this paper optimal LS linear estimators are obtained by using an innovation approach, which provides a simple derivation of the estimation algorithms due to the fact that the innovations constitute a white process; and (2) our missing measurement model considers at each sensor the possibility of observations containing only partial information about the state, or even only noise.

The paper is organized as follows. In Section 2 the system model with autocorrelated and cross-correlated noises and missing measurements coming from multiple sensors is described. Also, the suitable properties on the state and noise processes are specified and a brief description of the innovation approach to the optimal LS linear estimation problem is included. In Section 3 a recursive algorithm for the centralized optimal linear filter is presented for the considered model (the derivation has been deferred to Appendix 6). Next, in Section 4, the local LS linear filters and their corresponding error covariance matrices between any two local estimates are provided, and then

the distributed optimal weighted fusion estimators and their error covariance matrices are obtained by applying the optimal information fusion criterion weighted by matrices in the linear minimum variance sense. Finally, in Section 5, a numerical simulation example is presented to show the effectiveness of the estimation algorithms proposed in the current paper, and some conclusions are drawn in Section 6.

*Notation.* The notation used throughout the paper is standard. For any matrix  $A$ , the notation symbols  $A^T$  and  $A^{-1}$  represent its transpose and inverse, respectively;  $\mathbb{R}^n$  denotes the  $n$ -dimensional Euclidean space and  $\mathbb{R}^{m \times n}$  is the set of all real matrices of dimension  $m \times n$ . The shorthand  $\text{Diag}(a_1, \dots, a_r)$  denotes a diagonal matrix whose diagonal entries are  $a_1, \dots, a_r$ . If the dimensions of matrices are not explicitly stated, they are assumed to be compatible for algebraic operations.  $\delta_{k-s}$  is the Kronecker delta function, which is equal to one, if  $k = s$ , and zero otherwise. Moreover, for arbitrary random vectors  $\alpha$  and  $\beta$ , we will denote  $\text{Cov}[\alpha, \beta] = E[(\alpha - E[\alpha])(\beta - E[\beta])^T]$  and  $\text{Cov}[\alpha] = \text{Cov}[\alpha, \alpha]$ , where  $E[\cdot]$  stands for the mathematical expectation operator. Finally,  $\hat{\alpha}$  denotes the estimator of  $\alpha$  and  $\tilde{\alpha} = \alpha - \hat{\alpha}$  the estimation error.

## 2. Problem Formulation

Our aim is to obtain recursive algorithms for the optimal LS linear filtering problem in a class of discrete-time stochastic systems with missing measurements coming from multiple sensors, by using centralized and distributed fusion methods. In this section, firstly the system model and the assumptions about the state and noise processes are presented and, secondly, the optimal LS linear estimation problem is formulated using an innovation approach.

*2.1. Stochastic System Model.* Consider a discrete-time linear stochastic system with autocorrelated and cross-correlated noises and missing measurements coming from  $r$  sensors. The phenomenon of missing measurements occurs randomly and, for each sensor, a different sequence of scalar random variables with discrete distribution over the interval  $[0, 1]$  is used to model this phenomenon. Specifically, the following system is considered:

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \quad k \geq 1, \quad (1)$$

where  $x_k \in \mathbb{R}^n$  is the state,  $\{w_k; k \geq 0\}$  is the process noise, and  $F_k$ , for  $k \geq 0$ , are known matrices with compatible dimensions.

Consider  $r$  sensors which, at any time  $k$ , provide scalar measurements of the system state, perturbed by additive and multiplicative noises according to the following model:

$$y_k^i = \theta_k^i H_k^i x_k + v_k^i, \quad k \geq 1, \quad i = 1, 2, \dots, r, \quad (2)$$

where  $\{y_k^i; k \geq 1\}$  are the measured data;  $\{v_k^i; k \geq 1\}$  are measurement noises;  $\{\theta_k^i; k \geq 1\}$  are scalar random variables sequences;  $H_k^i$ , for  $k \geq 1$ , are known time-varying matrices

with compatible dimensions; superscript  $i$  denotes the  $i$ th sensor, and  $r$  is the number of sensors.

Next, the statistical properties assumed about the initial state and noise processes involved in (1) and (2) are specified.

- (i) The initial state  $x_0$  is a random vector with  $E[x_0] = \bar{x}_0$  and  $\text{Cov}[x_0] = P_0$ .
- (ii) The process noise,  $\{w_k; k \geq 0\}$ , and the measurement noises,  $\{v_k^i; k \geq 1\}$ ,  $i = 1, 2, \dots, r$ , are zero-mean sequences with covariances and cross-covariances:
 
$$\begin{aligned} \text{Cov}[w_k, w_s] &= Q_{k,k} \delta_{k-s} + Q_{k,s} \delta_{k-s+1} + Q_{k,s} \delta_{k-s-1}, \\ \text{Cov}[v_k^i, v_s^j] &= R_{k,k}^{ij} \delta_{k-s} + R_{k,s}^{ij} \delta_{k-s+1} + R_{k,s}^{ij} \delta_{k-s-1}, \\ \text{Cov}[w_k, v_s^j] &= S_{k,k}^i \delta_{k-s} + S_{k,s}^i \delta_{k-s+1} + S_{k,s}^i \delta_{k-s+2}. \end{aligned} \quad (3)$$
- (iii) The multiplicative noises  $\{\theta_k^i; k \geq 1\}$ ,  $i = 1, 2, \dots, r$ , are white sequences of scalar variables with discrete distribution over the interval  $[0, 1]$ , with  $E[\theta_k^i] = \bar{\theta}_k^i$  and  $\text{Var}[\theta_k^i] = V_k^{\theta^i}$ .
- (iv) The initial state  $x_0$  and the multiplicative noises  $\{\theta_k^i; k \geq 1\}$ , for  $i = 1, 2, \dots, r$ , are mutually independent, and they are independent of the additive noises  $\{w_k; k \geq 0\}$  and  $\{v_k^i; k \geq 1\}$ , for  $i = 1, 2, \dots, r$ .

*Remark 1.* From assumption (ii) the following correlation properties of the additive noises are easily deduced.

- (1) The noise vectors  $w_k$  and  $w_s$  are correlated at consecutive sampling times,  $|k - s| = 1$ , and independent otherwise; the covariance matrices of  $w_k$  with  $w_{k-1}$ , and  $w_{k+1}$  are  $Q_{k,k-1}$ , and  $Q_{k,k+1}$ , respectively.
- (2) For  $i, j = 1, 2, \dots, r$ , the measurement noises  $v_k^i$  and  $v_s^j$  are cross-correlated at the same sampling time and at consecutive sampling times,  $|k - s| = 0, 1$ , and independent otherwise; the cross-covariances of  $v_k^i$  with  $v_k^j$ ,  $v_{k-1}^j$  and  $v_{k+1}^j$  are  $R_{k,k}^{ij}$ ,  $R_{k,k-1}^{ij}$  and  $R_{k,k+1}^{ij}$ , respectively.
- (3) For  $i = 1, 2, \dots, r$ , the measurement noises  $v_k^i$  are correlated with the noise vectors  $w_s$ , for  $s = k, k-1, k-2$ , and independent otherwise; the cross-covariance matrices of  $v_k^i$  with  $w_k$ ,  $w_{k-1}$  and  $w_{k-2}$  are  $S_{k,k}^i$ ,  $S_{k-1,k}^i$  and  $S_{k-2,k}^i$ , respectively.

The correlation conditions of the process noise and the measurement noises considered in this paper are the same as those in [16]. Systems with only finite-step correlated process noises or multistep correlated process and measurement noises are considered in [13–15], among others. The current study can be extended to more general systems involving finite-step autocorrelated and cross-correlated noises with no difficulty, except for a greater complexity in the mathematical derivations.

*Remark 2.* From the state equation (1) and assumptions (ii) and (iv), it is easy to deduce that  $D_k = E[x_k x_k^T]$  is recursively calculated by

$$\begin{aligned} D_k &= F_{k-1} D_{k-1} F_{k-1}^T + Q_{k-1, k-1} + F_{k-1} Q_{k-2, k-1} \\ &\quad + Q_{k-1, k-2} F_{k-1}^T, \quad k \geq 2, \\ D_1 &= F_0 D_0 F_0^T + Q_{0,0}, \quad D_0 = P_0 + \bar{x}_0 \bar{x}_0^T. \end{aligned} \quad (4)$$

Also, it is easy to see that the state  $x_k$  is correlated with the measurement noises  $v_k^i$ , for  $i = 1, 2, \dots, r$ , and the expectations  $E_k^i = E[x_k v_k^i]$  satisfy

$$E_k^i = F_{k-1} S_{k-2, k}^i + S_{k-1, k}^i, \quad k \geq 2; \quad E_1^i = S_{0,1}^i. \quad (5)$$

*Remark 3.* According to assumption (iii), the scalar random variables  $\theta_k^i$  take values over the interval  $[0, 1]$  and they can satisfy any arbitrary discrete probability distribution over such interval, for instance, a Bernoulli distribution. Usually, Bernoulli random variables have been used to model the phenomenon of missing measurements (see, e.g., [25] and references therein), with  $\theta_k^i = 1$  meaning that the state  $x_k$  is present in the measurement  $y_k^i$  coming from the  $i$ th sensor at time  $k$ , while  $\theta_k^i = 0$  means that the state is missing in the measured data at time  $k$  or, equivalently, that such observation only contains additive noise  $v_k^i$ . However, in practice, the information transmitted at a sampling time can usually be neither completely missing nor completely successful, but only part of the information can go through; in such situations, only partial information is missing and the proportion of missed data at one moment is a *fraction* other than 0 or 1 (see, e.g., [28, 29] and references therein).

*2.2. Stacked Measurement Equation.* As noted above, our aim is to solve the optimal LS linear estimation problem of the state  $x_k$  based on the measurements  $\{y_1^i, y_2^i, \dots, y_k^i\}$ , for  $i = 1, 2, \dots, r$ , by using centralized and distributed fusion methods to process the measured sensor data. The centralized fusion method considers that all the measurement data coming from  $r$  sensors are transmitted to a fusion center for being processed; for this purpose and to simplify the notation, the measurement equation (2) is rewritten in a stacked form as follows:

$$y_k = \Theta_k H_k x_k + v_k, \quad k \geq 1, \quad (6)$$

where  $y_k = (y_k^1, \dots, y_k^r)^T$ ,  $v_k = (v_k^1, \dots, v_k^r)^T$ ,  $H_k = (H_k^{1T}, \dots, H_k^{rT})^T$ , and  $\Theta_k = \text{Diag}(\theta_k^1, \dots, \theta_k^r)$ .

The following properties of the noises in (6) are easily inferred from the model assumptions (ii)–(iv) previously stated.

- (i) The additive noise  $\{v_k; k \geq 1\}$  is a zero-mean process satisfying:

$$\text{Cov}[v_k, v_s] = R_{k,k} \delta_{k-s} + R_{k,s} \delta_{k-s+1} + R_{k,s} \delta_{k-s-1}, \quad (7)$$

$$\text{Cov}[w_k, v_s] = S_{k,k} \delta_{k-s} + S_{k,s} \delta_{k-s+1} + S_{k,s} \delta_{k-s+2},$$

where  $R_{k,s} = (R_{k,s}^{ij})_{i,j=1,2,\dots,r}$  and  $S_{k,s} = (S_{k,s}^1, \dots, S_{k,s}^r)$ .

- (ii) The state vector  $x_k$  and the measurement noise vector  $v_k$  are correlated with  $E_k = E[x_k v_k^T]$  satisfying

$$E_k = F_{k-1} S_{k-2, k} + S_{k-1, k}, \quad k \geq 2, \quad E_1 = S_{0,1}. \quad (8)$$

- (iii) The random matrices  $\{\Theta_k; k \geq 1\}$  satisfy  $E[\Theta_k] = \bar{\Theta}_k = \text{Diag}(\bar{\theta}_k^1, \dots, \bar{\theta}_k^r)$  and  $E[(\Theta_k - \bar{\Theta}_k)^2] = \text{Diag}(V_k^{\theta^1}, \dots, V_k^{\theta^r})$ ; also, denoting  $\theta_k = (\theta_k^1, \dots, \theta_k^r)^T$ , it is clear that  $\text{Cov}[\theta_k] = \text{Diag}(V_k^{\theta^1}, \dots, V_k^{\theta^r})$ . Moreover, for any random matrix  $G$  independent of  $\{\Theta_k; k \geq 1\}$ , it is easily deduced that

$$E[(\Theta_k - \bar{\Theta}_k) G (\Theta_k - \bar{\Theta}_k)] = \text{Cov}[\theta_k] \circ E[G], \quad (9)$$

where  $\circ$  denotes the Hadamard product [23].

- (iv) The initial state  $x_0$  and  $\{\Theta_k; k \geq 1\}$  are independent, and they are independent of  $\{w_k; k \geq 0\}$  and  $\{v_k; k \geq 1\}$ .

*2.3. Innovation Approach to the Optimal LS Linear Estimation Problem.* To address the optimal LS linear estimation problem of the state  $x_k$  based on the measurements  $\{y_1^i, y_2^i, \dots, y_k^i\}$ ,  $i = 1, 2, \dots, r$ , the centralized and distributed fusion methods will be used. In both cases, recursive algorithms for the LS linear estimators will be established using an innovation approach and the orthogonal projection Lemma (OPL); more specifically we have the following.

*Centralized Fusion Estimation Problem.* Our aim is to obtain the optimal LS linear filter,  $\hat{x}_{k/k}$ , of the state  $x_k$  based on the measurements  $\{y_1, y_2, \dots, y_k\}$ , given in (6), by recursive algorithms.

As known, the LS linear filter  $\hat{x}_{k/k}$  is the orthogonal projection of the state  $x_k$  over the linear space spanned by  $\{y_1, y_2, \dots, y_k\}$ . These observations are generally nonorthogonal vectors, but the Gram-Schmidt orthogonalization procedure allows us to substitute them by a set of orthogonal vectors, called *innovations*, defined as the difference between each observation and its one-stage predictor. Due to the orthogonality property of the innovations and since the innovation process is uniquely determined by the observations, the LS linear filter,  $\hat{x}_{k/k}$ , can be calculated as linear combination of the innovations; namely,

$$\hat{x}_{k/k} = \sum_{s=1}^k \mathcal{X}_{k,s} \Pi_{s,s}^{-1} \mu_s, \quad k \geq 1, \quad (10)$$

where  $\mu_s = y_s - \hat{y}_{s/s-1}$  are the innovation vectors, with  $\hat{y}_{s/s-1}$  the one-stage observation predictor,  $\Pi_{s,s} = E[\mu_s \mu_s^T]$ , and  $\mathcal{X}_{k,s} = E[x_k \mu_s^T]$ .

*Distributed Fusion Estimation Problem.* To address the distributed fusion estimation problem, firstly, recursive algorithms to obtain local LS linear filters,  $\hat{x}_{k/k}^i$ , for  $i = 1, 2, \dots, r$ , and the error cross-covariance matrices between any two local estimates, are derived. Secondly, the distributed fusion filter,  $\hat{x}_{k/k}^D$ , is established by applying the optimal information

fusion criterion weighted by matrices in the linear minimum variance sense [30].

Analogously to (10), denoting  $\mu_s^i = y_s^i - \hat{y}_{s/s-1}^i$ ,  $\Pi_{s,s}^{ii} = E[\mu_s^i \mu_s^i]$ , and  $\mathcal{X}_{k,s}^i = E[x_k \mu_s^i]$ , the local filter  $\hat{x}_{k/k}^i$  is expressed as

$$\hat{x}_{k/k}^i = \sum_{s=1}^k \mathcal{X}_{k,s}^i (\Pi_{s,s}^{ii})^{-1} \mu_s^i, \quad k \geq 1. \quad (11)$$

### 3. Optimal LS Linear Centralized Fusion Estimation

In this section a recursive algorithm for the centralized optimal (under the LS criterion) linear filter,  $\hat{x}_{k/k}$  is derived. Such algorithm is deduced using (10) and the OPL, and it is presented in Theorem 5. Firstly, in order to simplify the proof of Theorem 5, the following lemma is established.

**Lemma 4.** *Under assumptions (i)–(iv), the following results hold:*

$$\mathcal{W}_{k,k} = E[w_k \mu_k^T] = Q_{k,k-1} H_k^T \bar{\Theta}_k + S_{k,k}, \quad k \geq 1, \quad (12)$$

$$\mathcal{V}_{k,k-1} = E[v_k \mu_{k-1}^T] = S_{k-2,k}^T H_{k-1}^T \bar{\Theta}_{k-1} + R_{k,k-1}, \quad k \geq 2. \quad (13)$$

*Proof.* Since  $w_k$  is independent of  $y_1, \dots, y_{k-1}$ ,  $E[w_k \hat{y}_{k/k-1}^T] = 0$  and hence  $\mathcal{W}_{k,k} = E[w_k y_k^T]$ . Now, using (1) and (6),  $\mathcal{W}_{k,k}$  can be calculated as follows:

$$\begin{aligned} \mathcal{W}_{k,k} &= E[w_k (\Theta_k H_k x_k + v_k)^T] \\ &= E[w_k x_k^T] H_k^T \bar{\Theta}_k + S_{k,k} \\ &= E[w_k (F_{k-1} x_{k-1} + w_{k-1})^T] H_k^T \bar{\Theta}_k + S_{k,k} \\ &= Q_{k,k-1} H_k^T \bar{\Theta}_k + S_{k,k}. \end{aligned} \quad (14)$$

Taking into account that  $v_k$  is independent of  $y_1, \dots, y_{k-2}$ , the calculation of  $\mathcal{V}_{k,k-1}$  is similar to that of  $\mathcal{W}_{k,k}$ , and hence the proof is omitted.  $\square$

**Theorem 5.** *For the system model (1) and measurement model (6), under assumptions (i)–(iv), the optimal LS linear filter  $\hat{x}_{k/k}$  is obtained as*

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + \mathcal{X}_{k,k} \Pi_{k,k}^{-1} \mu_k, \quad k \geq 1, \quad \hat{x}_{0/0} = \bar{x}_0, \quad (15)$$

where the state predictor,  $\hat{x}_{k/k-1}$ , satisfies

$$\begin{aligned} \hat{x}_{k/k-1} &= F_{k-1} \hat{x}_{k-1/k-1} + \mathcal{W}_{k-1,k-1} \Pi_{k-1,k-1}^{-1} \mu_{k-1}, \quad k \geq 2, \\ \hat{x}_{1/0} &= F_0 \hat{x}_{0/0}. \end{aligned} \quad (16)$$

The innovation,  $\mu_k$ , is given by

$$\begin{aligned} \mu_k &= y_k - \bar{\Theta}_k H_k \hat{x}_{k/k-1} - \mathcal{V}_{k,k-1} \Pi_{k-1,k-1}^{-1} \mu_{k-1}, \quad k \geq 2, \\ \mu_1 &= y_1 - \bar{\Theta}_1 H_1 \hat{x}_{1/0}. \end{aligned} \quad (17)$$

The matrix  $\mathcal{X}_{k,k} = E[x_k \mu_k^T]$  is calculated by

$$\begin{aligned} \mathcal{X}_{k,k} &= P_{k/k-1} H_k^T \bar{\Theta}_k + E_k - \mathcal{X}_{k,k-1} \Pi_{k-1,k-1}^{-1} \mathcal{V}_{k,k-1}^T, \quad k \geq 2, \\ \mathcal{X}_{1,1} &= P_{1/0} H_1^T \bar{\Theta}_1 + E_1, \end{aligned} \quad (18)$$

where  $\mathcal{X}_{k,k-1} = E[x_k \mu_{k-1}^T]$  satisfies

$$\mathcal{X}_{k,k-1} = F_{k-1} \mathcal{X}_{k-1,k-1} + \mathcal{W}_{k-1,k-1}, \quad k \geq 2. \quad (19)$$

The prediction error covariance matrix,  $P_{k/k-1}$ , is obtained by

$$\begin{aligned} P_{k/k-1} &= F_{k-1} P_{k-1/k-1} F_{k-1}^T + Q_{k-1,k-1} + F_{k-1} \mathcal{F}_{k-1} + \mathcal{F}_{k-1}^T F_{k-1}^T \\ &\quad - \mathcal{W}_{k-1,k-1} \Pi_{k-1,k-1}^{-1} \mathcal{W}_{k-1,k-1}^T, \quad k \geq 2, \\ P_{1/0} &= F_0 P_{0/0} F_0^T + Q_{0,0}, \end{aligned} \quad (20)$$

where  $\mathcal{F}_k = E[\tilde{x}_{k/k} \mu_k^T]$  is calculated by

$$\mathcal{F}_k = Q_{k-1,k} - \mathcal{X}_{k,k} \Pi_{k,k}^{-1} \mathcal{W}_{k,k}^T, \quad k \geq 1. \quad (21)$$

The filtering error covariance matrix,  $P_{k/k}$ , is given by

$$P_{k/k} = P_{k/k-1} - \mathcal{X}_{k,k} \Pi_{k,k}^{-1} \mathcal{X}_{k,k}^T, \quad k \geq 1, \quad P_{0/0} = P_0. \quad (22)$$

The innovation covariance matrix,  $\Pi_{k,k}$ , satisfies

$$\begin{aligned} \Pi_{k,k} &= \text{Cov}(\theta_k) \circ (H_k D_k H_k^T) + R_{k,k} + \bar{\Theta}_k H_k \mathcal{X}_{k,k} \\ &\quad + \mathcal{X}_{k,k}^T H_k^T \bar{\Theta}_k - \bar{\Theta}_k H_k P_{k/k-1} H_k^T \bar{\Theta}_k \\ &\quad - \mathcal{V}_{k,k-1} \Pi_{k-1,k-1}^{-1} \mathcal{V}_{k,k-1}^T, \quad k \geq 2, \\ \Pi_{1,1} &= \text{Cov}(\theta_1) \circ (H_1 D_1 H_1^T) + R_{1,1} + \bar{\Theta}_1 H_1 \mathcal{X}_{1,1} \\ &\quad + \mathcal{X}_{1,1}^T H_1^T \bar{\Theta}_1 - \bar{\Theta}_1 H_1 P_{1/0} H_1^T \bar{\Theta}_1. \end{aligned} \quad (23)$$

The matrices  $D_k$ ,  $E_k$ ,  $\mathcal{W}_{k,k}$ , and  $\mathcal{V}_{k,k-1}$  are given in (4), (8), (12), and (13), respectively.

*Proof.* See Appendix 6.  $\square$

**Remark 6.** In conventional estimation problems in systems with missing measurements and uncorrelated additive white noises, the one-stage state and observation predictors are calculated as  $\hat{x}_{k/k-1} = F_{k-1} \hat{x}_{k-1/k-1}$  and  $\hat{y}_{k/k-1} = \bar{\Theta}_k H_k \hat{x}_{k/k-1}$ , respectively. However, this is not true for the problem at hand since, due to the correlation assumption (ii), the noise estimators  $\hat{w}_{k-1/k-1}$  and  $\hat{v}_{k/k-1}$  must be taken into account for the derivation of the predictors. Besides the fact of considering missing measurements, this is the main difference between the optimal estimators proposed in the current paper and the suboptimal Kalman-type ones proposed in [16], where the noise estimators are considered to be equal to zero.

#### 4. Distributed Fusion Estimation

One of the main disadvantages of the centralized fusion estimators derived in Section 3 is that they may have a high computational cost due to augmentation. Moreover, as is widely known, the centralized approach has several other drawbacks, such as fault detection, isolation, poor reliability, and so forth. To overcome these disadvantages, our aim in this section is to address the optimal distributed fusion estimation problem, in which each single sensor provides its local LS linear estimator and their estimation error covariance matrices, and then these local estimators along with the covariances and cross-covariance matrices of the estimation errors between any two sensors are sent to the fusion center for fusion based on the matrices-weighted fusion estimation criterion in the linear minimum variance sense [30].

**4.1. Local LS Linear Filtering Algorithms.** For each single sensor subsystem of systems (1) and (2), the following theorem provides recursive formulas for the local LS linear filters,  $\hat{x}_{k/k}^i$ , and their corresponding error covariance matrices,  $P_{k/k}^{ii}$ .

**Theorem 7.** For the  $i$ th sensor subsystem of systems (1) and (2) under assumptions (i)–(iv), the local LS linear filter,  $\hat{x}_{k/k}^i$ , is calculated by

$$\hat{x}_{k/k}^i = \hat{x}_{k/k-1}^i + \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \mu_k^i, \quad k \geq 1, \quad \hat{x}_{0/0}^i = \bar{x}_0, \quad (24)$$

where the local LS linear predictor,  $\hat{x}_{k/k-1}^i$ , satisfies

$$\begin{aligned} \hat{x}_{k/k-1}^i &= F_{k-1} \hat{x}_{k-1/k-1}^i + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mu_{k-1}^i, \quad k \geq 2, \\ \hat{x}_{1/0}^i &= F_0 \hat{x}_{0/0}^i, \end{aligned} \quad (25)$$

with  $\mathcal{W}_{k,k}^i = \bar{\theta}_k^i Q_{k,k-1} H_k^{iT} + S_{k,k}^i$ ,  $k \geq 1$ .

The innovation,  $\mu_k^i$ , is given by

$$\begin{aligned} \mu_k^i &= y_k^i - \bar{\theta}_k^i H_k^i \hat{x}_{k/k-1}^i - \mathcal{V}_{k,k-1}^{ii} (\Pi_{k-1,k-1}^{ii})^{-1} \mu_{k-1}^i, \quad k \geq 2, \\ \mu_1^i &= y_1^i - \bar{\theta}_1^i H_1^i \hat{x}_{1/0}^i, \end{aligned} \quad (26)$$

with  $\mathcal{V}_{k,k-1}^{ii} = \bar{\theta}_{k-1}^{ii} S_{k-2,k}^{iT} H_{k-1}^{iT} + R_{k,k-1}^{ii}$ ,  $k \geq 2$ .

The vector  $\mathcal{X}_{k,k}^i = E[x_k \mu_k^i]$  is calculated from the following expression

$$\begin{aligned} \mathcal{X}_{k,k}^i &= \bar{\theta}_k^i P_{k/k-1}^{ii} H_k^{iT} + E_k^i - \mathcal{X}_{k,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mathcal{V}_{k,k-1}^{ii}, \\ & \quad k \geq 2, \\ \mathcal{X}_{1,1}^i &= \bar{\theta}_1^i P_{1/0}^{ii} H_1^{iT} + E_1^i, \end{aligned} \quad (27)$$

where  $\mathcal{X}_{k,k-1}^i = F_{k-1} \mathcal{X}_{k-1,k-1}^i + \mathcal{W}_{k-1,k-1}^i$ ,  $k \geq 2$ .

The local prediction error covariance matrix,  $P_{k/k-1}^{ii}$ , is obtained by

$$\begin{aligned} P_{k/k-1}^{ii} &= F_{k-1} P_{k-1/k-1}^{ii} F_{k-1}^T + Q_{k-1,k-1} + F_{k-1} \mathcal{F}_{k-1}^i + \mathcal{F}_{k-1}^{iT} F_{k-1}^T \\ & \quad - \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mathcal{W}_{k-1,k-1}^{iT}, \quad k \geq 2, \\ P_{1/0}^{ii} &= F_0 P_{0/0}^{ii} F_0^T + Q_{0,0}, \end{aligned} \quad (28)$$

where  $\mathcal{F}_k^i = Q_{k-1,k} - \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \mathcal{W}_{k,k}^{iT}$ ,  $k \geq 1$ , and  $P_{k/k}^{ii}$  the filtering error covariance matrix, is given by

$$P_{k/k}^{ii} = P_{k/k-1}^{ii} - \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \mathcal{X}_{k,k}^{iT}, \quad k \geq 1, \quad P_{0/0}^{ii} = P_0. \quad (29)$$

The innovation variance,  $\Pi_{k,k}^{ii}$ , satisfies

$$\begin{aligned} \Pi_{k,k}^{ii} &= V_k^{\theta_i} H_k^i D_k H_k^{iT} + R_{k,k}^{ii} + \bar{\theta}_k^i H_k^i \mathcal{X}_{k,k}^i + \bar{\theta}_k^i \mathcal{X}_{k,k}^{iT} H_k^{iT} \\ & \quad - (\bar{\theta}_k^i)^2 H_k^i P_{k/k-1}^{ii} H_k^{iT} - (\mathcal{V}_{k,k-1}^{ii})^2 (\Pi_{k-1,k-1}^{ii})^{-1}, \\ & \quad k \geq 2, \\ \Pi_{1,1}^{ii} &= V_1^{\theta_i} H_1^i D_1 H_1^{iT} + R_{1,1}^{ii} + \bar{\theta}_1^i H_1^i \mathcal{X}_{1,1}^i \\ & \quad + \bar{\theta}_1^i \mathcal{X}_{1,1}^{iT} H_1^{iT} - (\bar{\theta}_1^i)^2 H_1^i P_{1/0}^{ii} H_1^{iT}. \end{aligned} \quad (30)$$

The matrix  $D_k$  and the vector  $E_k^i$  are given in (4) and (5), respectively.

*Proof.* The proof, based on the innovation approach and the OPL, is omitted for being analogous to that of Theorem 5. Nevertheless, it should be indicated that, in this proof, the Hadamard product is not used since, instead of the diagonal stochastic matrix  $\Theta_k$ , the scalar variable  $\bar{\theta}_k^i$  is now involved in the derivation of the estimators.  $\square$

**Remark 8.** As indicated in Remark 6 for the centralized estimators, it must be noted that, due to the correlation assumption (ii) of the additive noises  $\{w_k\}$  and  $\{v_k^j\}$ , the estimators  $\hat{w}_{k-1/k-1}^i = \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mu_{k-1}^i$  and  $\hat{v}_{k/k-1}^j = \mathcal{V}_{k,k-1}^{jj} (\Pi_{k-1,k-1}^{jj})^{-1} \mu_{k-1}^j$  are not equal to zero, and hence the optimal local state predictor,  $\hat{x}_{k/k-1}^i = F_{k-1} \hat{x}_{k-1/k-1}^i + \hat{w}_{k-1/k-1}^i$ , and the observation predictor,  $\hat{y}_{k/k-1}^j = \bar{\theta}_k^j H_k^j \hat{x}_{k/k-1}^i + \hat{v}_{k/k-1}^j$ , are quite different from conventional filtering algorithms with uncorrelated white noises. This issue, along with the consideration of missing measurements at each single sensor, constitutes the main difference between the current optimal local estimators and the suboptimal local estimators proposed in [16].

**4.2. Cross-Covariance Matrices of Local Estimation Errors.** To apply the optimal fusion criterion weighted by matrices in

the linear minimum variance sense, the filtering,  $P_{k/k}^{ij}$ , and prediction,  $P_{k/k-1}^{ij}$ , error cross-covariance matrices between local estimators of any two subsystems must be calculated.

For simplicity, besides the notation of Theorem 7, for  $i \neq j$ ,  $i, j = 1, 2, \dots, r$ , we introduce the following notation:

$$\begin{aligned} L_k^{ij} &= E[\hat{x}_{k/k-1}^i \mu_k^j], & \Pi_{k,s}^{ij} &= E[\mu_k^i \mu_s^j], \\ \mathcal{V}_{k,k-1}^{ij} &= E[v_k^i \mu_{k-1}^j]. \end{aligned} \quad (31)$$

Also, in order to simplify the calculation of the error cross-covariance matrices, the following lemmas are given.

**Lemma 9.** *Under assumptions (i)–(iv), the following results hold.*

(a) *The expectation  $E[\hat{x}_{k/k-1}^i \mu_{k-1}^j]$  satisfies*

$$\begin{aligned} E[\hat{x}_{k/k-1}^i \mu_{k-1}^j] &= F_{k-1} L_{k-1}^{ij} \\ &+ \mathcal{X}_{k,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij}, \quad k \geq 2. \end{aligned} \quad (32)$$

(b) *The expectation  $E[\hat{x}_{k/k-1}^i v_k^j]$  satisfies*

$$E[\hat{x}_{k/k-1}^i v_k^j] = \mathcal{X}_{k,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mathcal{V}_{k,k-1}^{ji}, \quad k \geq 2, \quad (33)$$

where  $\mathcal{V}_{k,k-1}^{ji} = \bar{\theta}_{k-1}^j H_{k-1}^i S_{k-2,k}^j + R_{k-1,k}^{ij}$ .

(c) *The expectation  $E[v_k^i \mu_k^j]$  satisfies*

$$\begin{aligned} E[v_k^i \mu_k^j] &= \bar{\theta}_k^j E_k^{iT} H_k^{jT} + R_{k,k}^{ij} \\ &- \mathcal{V}_{k,k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} (\bar{\theta}_k^j H_k^j \mathcal{X}_{k,k-1}^j + \mathcal{V}_{k,k-1}^{jj})^T, \\ &k \geq 2. \end{aligned} \quad (34)$$

*Proof.* (a) From (25) for  $\hat{x}_{k/k-1}^i$  and (24) for  $\hat{x}_{k-1/k-1}^j$ , we have

$$\begin{aligned} E[\hat{x}_{k/k-1}^i \mu_{k-1}^j] &= F_{k-1} E[\hat{x}_{k-1/k-1}^i \mu_{k-1}^j] + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &= F_{k-1} E[\hat{x}_{k-1/k-2}^i \mu_{k-1}^j] + F_{k-1} \mathcal{X}_{k-1,k-1}^i \\ &\quad \times (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &\quad + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &= F_{k-1} L_{k-1}^{ij} + (F_{k-1} \mathcal{X}_{k-1,k-1}^i + \mathcal{W}_{k-1,k-1}^i) \\ &\quad \times (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij}, \end{aligned} \quad (35)$$

and since  $\mathcal{X}_{k,k-1}^i = F_{k-1} \mathcal{X}_{k-1,k-1}^i + \mathcal{W}_{k-1,k-1}^i$ , expression (32) is proved.

(b) Analogously, taking into account that  $E[\hat{x}_{k-1/k-2}^i v_k^j] = 0$ , we have

$$\begin{aligned} E[\hat{x}_{k/k-1}^i v_k^j] &= F_{k-1} E[\hat{x}_{k-1/k-1}^i v_k^j] \\ &\quad + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} E[\mu_{k-1}^i v_k^j] \\ &= (F_{k-1} \mathcal{X}_{k-1,k-1}^i + \mathcal{W}_{k-1,k-1}^i) (\Pi_{k-1,k-1}^{ii})^{-1} \mathcal{V}_{k,k-1}^{ji}, \end{aligned} \quad (36)$$

and expression (33) is immediately obtained. Finally, the derivation of expression  $\mathcal{V}_{k,k-1}^{ji} = \bar{\theta}_{k-1}^j H_{k-1}^i S_{k-2,k}^j + R_{k-1,k}^{ij}$  is similar to that of (13) and hence it is omitted.

(c) Taking into account expression (26) for  $\mu_k^j$ , with (2) for  $y_k^j$ , we have

$$\begin{aligned} E[v_k^i \mu_k^j] &= \bar{\theta}_k^j E_k^{iT} H_k^{jT} + R_{k,k}^{ij} - \bar{\theta}_k^j E[v_k^i \hat{x}_{k/k-1}^{jT}] H_k^{jT} \\ &\quad - \mathcal{V}_{k,k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj}, \quad k \geq 2, \end{aligned} \quad (37)$$

and using (33) for  $E[v_k^i \hat{x}_{k/k-1}^{jT}]$ , expression (34) is obtained.  $\square$

**Lemma 10.** *Under assumptions (i)–(iv), for  $i \neq j$ ,  $i, j = 1, 2, \dots, r$ , the expectations  $L_k^{ij} = E[\hat{x}_{k/k-1}^i \mu_k^j]$  are recursively obtained by*

$$\begin{aligned} L_k^{ij} &= \bar{\theta}_k^j (P_{k/k-1}^{jj} - P_{k/k-1}^{ij}) H_k^{jT} - F_{k-1} L_{k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj} \\ &\quad + \mathcal{X}_{k,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \\ &\quad \times (\mathcal{V}_{k,k-1}^{ji} - \mathcal{V}_{k,k-1}^{jj} (\Pi_{k-1,k-1}^{jj})^{-1} \Pi_{k-1,k-1}^{ij}), \quad k \geq 2, \end{aligned} \quad (38)$$

with initial condition  $L_1^{ij} = 0$ .

*Proof.* Taking into account expression (26) for  $\mu_k^j$ , with (2) for  $y_k^j$ , we have

$$\begin{aligned} L_k^{ij} &= \bar{\theta}_k^j E[\hat{x}_{k/k-1}^i x_k^T] H_k^{jT} + E[\hat{x}_{k/k-1}^i v_k^j] \\ &\quad - \bar{\theta}_k^j E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{jT}] H_k^{jT} \\ &\quad - E[\hat{x}_{k/k-1}^i \mu_{k-1}^j] (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj}, \quad k \geq 2. \end{aligned} \quad (39)$$

From the OPL,  $E[\hat{x}_{k/k-1}^i x_k^T] = E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{iT}]$ ; then, taking into account (32) for  $E[\hat{x}_{k/k-1}^i \mu_{k-1}^j]$ , and (33) for  $E[\hat{x}_{k/k-1}^i v_k^j]$ , it is enough to prove that

$$E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{iT}] - E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{jT}] = P_{k/k-1}^{jj} - P_{k/k-1}^{ij}, \quad (40)$$

which is easily deduced since

$$\begin{aligned} E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{iT}] &= P_{k/k-1}^{ij} - D_k + E[\hat{x}_{k/k-1}^i \hat{x}_{k/k-1}^{iT}] + E[\hat{x}_{k/k-1}^j \hat{x}_{k/k-1}^{jT}], \\ E[\hat{x}_{k/k-1}^j \hat{x}_{k/k-1}^{jT}] &= D_k - P_{k/k-1}^{jj}. \end{aligned} \quad (41)$$

□

**Lemma 11.** Under assumptions (i)–(iv), for  $i \neq j$ ,  $i, j = 1, 2, \dots, r$ , the innovation cross-covariance  $\Pi_{k,k}^{ij} = E[\mu_k^i \mu_k^j]$  satisfies

$$\begin{aligned} \Pi_{k,k}^{ij} &= \bar{\theta}_k^i H_k^i (\mathcal{X}_{k,k}^j - L_k^{ij}) + \bar{\theta}_k^j E_k^{iT} H_k^{jT} + R_{k,k}^{ij} \\ &\quad - \mathcal{V}_{k,k-1}^{ii} (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k}^{ij} \\ &\quad - \mathcal{V}_{k,k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} (\bar{\theta}_k^j H_k^j \mathcal{X}_{k,k-1}^j + \mathcal{V}_{k,k-1}^{jj})^T, \\ &\quad k \geq 2, \\ \Pi_{1,1}^{ij} &= \bar{\theta}_1^i H_1^i \mathcal{X}_{1,1}^j + \bar{\theta}_1^j E_1^{iT} H_1^{jT} + R_{1,1}^{ij}, \end{aligned} \quad (42)$$

where  $\Pi_{k-1,k}^{ij} = E[\mu_{k-1}^i \mu_k^j]$  is given by

$$\begin{aligned} \Pi_{k-1,k}^{ij} &= \bar{\theta}_k^j (\mathcal{X}_{k,k-1}^i - F_{k-1} L_{k-1}^{ji}) \\ &\quad - \mathcal{X}_{k,k-1}^j (\Pi_{k-1,k-1}^{jj})^{-1} \Pi_{k-1,k-1}^{ji} H_k^{jT} \\ &\quad + \mathcal{V}_{k,k-1}^{ji} - \Pi_{k-1,k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj}, \quad k \geq 2. \end{aligned} \quad (43)$$

*Proof.* Taking into account expression (26) for  $\mu_k^i$ , with (2) for  $y_k^i$ , we have

$$\begin{aligned} \Pi_{k,k}^{ij} &= \bar{\theta}_k^i H_k^i E[x_k \mu_k^j] + E[v_k^i \mu_k^j] - \bar{\theta}_k^i H_k^i E[\hat{x}_{k/k-1}^i \mu_k^j] \\ &\quad - \mathcal{V}_{k,k-1}^{ii} (\Pi_{k-1,k-1}^{ii})^{-1} E[\mu_{k-1}^i \mu_k^j] \\ &= \bar{\theta}_k^i H_k^i (\mathcal{X}_{k,k}^j - L_k^{ij}) + E[v_k^i \mu_k^j] \\ &\quad - \mathcal{V}_{k,k-1}^{ii} (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k}^{ij}, \quad k \geq 2, \end{aligned} \quad (44)$$

and, from (34) for  $E[v_k^i \mu_k^j]$ , expression for  $\Pi_{k,k}^{ij}$  is clear.

Analogously, taking into account expression (26) for  $\mu_k^j$ , with (2) for  $y_k^j$ , we have

$$\begin{aligned} \Pi_{k-1,k}^{ij} &= \bar{\theta}_k^j E[\mu_{k-1}^i x_k^T] H_k^{jT} + E[\mu_{k-1}^i v_k^j] \\ &\quad - \bar{\theta}_k^j E[\mu_{k-1}^i \hat{x}_{k/k-1}^{iT}] H_k^{jT} \\ &\quad - E[\mu_{k-1}^i \mu_{k-1}^j] (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj}, \end{aligned}$$

$$\begin{aligned} &= \bar{\theta}_k^j (\mathcal{X}_{k,k-1}^i - E[\mu_{k-1}^i \hat{x}_{k/k-1}^{iT}]) H_k^{jT} + \mathcal{V}_{k,k-1}^{ji} \\ &\quad - \Pi_{k-1,k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{V}_{k,k-1}^{jj}, \quad k \geq 2, \end{aligned} \quad (45)$$

and, from (32) for  $E[\mu_{k-1}^i \hat{x}_{k/k-1}^{iT}]$ , expression for  $\Pi_{k-1,k}^{ij}$  is immediately derived. □

In the following theorem, recursive formulas to calculate the filtering and prediction error cross-covariance matrices,  $P_{k/k}^{ij}$  and  $P_{k/k-1}^{ij}$ , respectively, are derived.

**Theorem 12.** Under assumptions (i)–(iv), the cross-covariance matrices,  $P_{k/k}^{ij}$ , of the filtering errors between the  $i$ th and the  $j$ th sensor subsystems are recursively computed by

$$\begin{aligned} P_{k/k}^{ij} &= P_{k/k-1}^{ij} + \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \Pi_{k,k}^{ij} (\Pi_{k,k}^{jj})^{-1} \mathcal{X}_{k,k}^{jT} \\ &\quad - (\mathcal{X}_{k,k}^j - L_k^{ij}) (\Pi_{k,k}^{jj})^{-1} \mathcal{X}_{k,k}^{iT} \\ &\quad - \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} (\mathcal{X}_{k,k}^i - L_k^{ji})^T, \quad k \geq 1, \\ P_{0/0}^{ij} &= P_0, \end{aligned} \quad (46)$$

where  $P_{k/k-1}^{ij}$ , the cross-covariance matrix of the prediction error between the  $i$ th and the  $j$ th sensor subsystems, satisfies

$$\begin{aligned} P_{k/k-1}^{ij} &= F_{k-1} P_{k-1/k-1}^{ij} F_{k-1}^T + Q_{k-1,k-1} + F_{k-1} \mathcal{F}_{k-1}^i \\ &\quad + \mathcal{F}_{k-1}^{jT} F_{k-1}^T + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &\quad \times (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{W}_{k-1,k-1}^{jT} \\ &\quad - \mathcal{G}_{k-1}^{ij} (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{W}_{k-1,k-1}^{jT} \\ &\quad - \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \mathcal{G}_{k-1}^{jiT}, \quad k \geq 2, \\ P_{1/0}^{ij} &= F_0 P_{0/0}^{ij} F_0^T + Q_{0,0}, \end{aligned} \quad (47)$$

where  $\mathcal{F}_k^{ij} = \mathcal{W}_{k,k}^j + F_k (\mathcal{X}_{k,k}^j - L_k^{ij} - \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \Pi_{k,k}^{ij})$ ,  $k \geq 1$ . The vectors  $L_k^{ij}$  and the innovation cross-covariances  $\Pi_{k,k}^{ij}$  are given in Lemmas 10 and 11, respectively.

*Proof.* By using (24) for  $\hat{x}_{k/k}^i$  and  $\hat{x}_{k/k}^j$ , we have

$$\begin{aligned} P_{k/k}^{ij} &= P_{k/k-1}^{ij} + \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} \Pi_{k,k}^{ij} (\Pi_{k,k}^{jj})^{-1} \mathcal{X}_{k,k}^{jT} \\ &\quad - E[(x_k - \hat{x}_{k/k-1}^i) \mu_k^j] (\Pi_{k,k}^{jj})^{-1} \mathcal{X}_{k,k}^{jT} \\ &\quad - \mathcal{X}_{k,k}^i (\Pi_{k,k}^{ii})^{-1} E[\mu_k^i (x_k - \hat{x}_{k/k-1}^j)^T]. \end{aligned} \quad (48)$$

Taking into account that  $E[x_k \mu_k^j] = \mathcal{X}_{k,k}^j$  and  $E[\hat{x}_{k/k-1}^i \mu_k^j] = L_k^{ij}$ , the recursive expression for the cross-covariance matrices of the local filtering errors is immediately deduced.



Following an analogous reasoning, using now (25) and taking into account that  $E[(x_k - \hat{x}_{k/k}^i)w_k^T] = \mathcal{F}_k^i$  and  $E[\mu_k^i w_k^T] = \mathcal{W}_{k,k}^{iT}$ , it is easy to see that

$$\begin{aligned} P_{k/k-1}^{ij} &= F_{k-1} P_{k-1/k-1}^{ij} F_{k-1}^T + Q_{k-1,k-1} + F_{k-1} \mathcal{F}_{k-1}^i \\ &\quad + \mathcal{F}_{k-1}^{jT} F_{k-1}^T + \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &\quad \times (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{W}_{k-1,k-1}^{jT} \\ &\quad - (\mathcal{W}_{k-1,k-1}^j + F_{k-1} E[\tilde{x}_{k-1/k-1}^i \mu_{k-1}^j]) \\ &\quad \times (\Pi_{k-1,k-1}^{jj})^{-1} \mathcal{W}_{k-1,k-1}^{jT} \\ &\quad - \mathcal{W}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \\ &\quad \times (\mathcal{W}_{k-1,k-1}^i + F_{k-1} E[\tilde{x}_{k-1/k-1}^j \mu_{k-1}^i])^T. \end{aligned} \quad (49)$$

Finally, using again (24) for  $\tilde{x}_{k-1/k-1}^i$ , and since  $E[\tilde{x}_{k-1/k-2}^i \mu_{k-1}^j] = L_{k-1}^{ij}$ , we have

$$\begin{aligned} E[\tilde{x}_{k-1/k-1}^i \mu_{k-1}^j] &= \mathcal{X}_{k-1,k-1}^j - L_{k-1}^{ij} - \mathcal{X}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \\ &= \mathcal{X}_{k-1,k-1}^j - L_{k-1}^{ij} - \mathcal{X}_{k-1,k-1}^i (\Pi_{k-1,k-1}^{ii})^{-1} \Pi_{k-1,k-1}^{ij} \end{aligned} \quad (50)$$

and the expression for the cross-covariance matrices of the local prediction errors is easily obtained.  $\square$

**4.3. Distributed Fusion Filtering Estimators.** Once the local LS linear filtering estimators  $\hat{x}_{k/k}^i$  and their error covariance matrices  $P_{k/k}^{ii}$ , given in Theorem 7, along with the error cross-covariance matrices,  $P_{k/k}^{ij}$ , given in Theorem 12, are available, the distributed optimal weighted fusion estimators and their error covariance matrices are obtained by applying the optimal information fusion criterion weighted by matrices in the linear minimum variance sense [30].

**Theorem 13.** For the system model (1) and measurement model (2), under assumptions (i)–(iv), the distributed optimal fusion filter,  $\hat{x}_{k/k}^D$ , is given by

$$\hat{x}_{k/k}^D = A_k^1 \hat{x}_{k/k}^1 + \dots + A_k^r \hat{x}_{k/k}^r, \quad k \geq 0, \quad (51)$$

where the local estimators  $\hat{x}_{k/k}^i, k \geq 0$  ( $i = 1, 2, \dots, r$ ) are calculated by the recursive algorithm established in Theorem 7.

The optimal matrix weights  $A_k^i$  ( $i = 1, 2, \dots, r$ ) are computed by

$$A_k = \Sigma_{k/k}^{-1} e (e^T \Sigma_{k/k}^{-1} e)^{-1}, \quad (52)$$

where the matrices  $A_k = [A_k^1, \dots, A_k^r]^T$  and  $e = [I, \dots, I]^T$  are both  $nr \times n$  matrices, and

$$\begin{aligned} \Sigma_{k/k} &= E \left[ (\tilde{x}_{k/k}^1, \dots, \tilde{x}_{k/k}^r) (\tilde{x}_{k/k}^1, \dots, \tilde{x}_{k/k}^r)^T \right] \\ &= (P_{k/k}^{ij})_{i,j=1,2,\dots,r} \end{aligned} \quad (53)$$

is an  $nr \times nr$  positive definite symmetric block matrix, whose  $n \times n$  matrix entries  $P_{k/k}^{ij}$  are given in Theorems 7 and 12.

The error covariance matrices of the distributed weighted fusion filtering estimators are computed by

$$P_{k/k}^D = (e^T \Sigma_{k/k}^{-1} e)^{-1}, \quad k \geq 0, \quad (54)$$

and the following inequality holds:  $P_{k/k}^D \leq P_{k/k}^{ii}, i = 1, 2, \dots, r$ .

*Proof.* The proof is omitted because it follows directly from the optimal information criterion weighted by matrices in the linear minimum variance sense [30].  $\square$

*Remark 14.* The proposed distributed optimal LS linear fusion filter requires the computation of an  $nr \times nr$  inverse matrix, with  $n$  the dimension of the system state and  $r$  the number of sensors. Consequently, the proposed distributed fusion method has a computational complexity of  $O[(nr)^3]$ , equal to that of the distributed Kalman-type filter in [16] and less than that of the distributed fusion filters based on the state augmentation approach. Hence, our distributed fusion method is superior to the filter proposed in [16] (since it has the same computation burden but better accuracy) and also to the distributed fusion filters based on state augmentation (since it has less computational complexity).

## 5. Numerical Simulation Example

In this section, a numerical simulation example is presented to illustrate the effectiveness of the centralized and distributed filtering algorithms proposed in this paper. Consider a scalar first-order autoregressive model with missing measurements coming from two sensors with autocorrelated and cross-correlated noises. According to the proposed observation model, two different independent sequences of random variables with a certain probability distribution over the interval  $[0, 1]$  are used to model the missing phenomenon. Specifically, the following model is considered as follows:

$$x_k = 0.95x_{k-1} + w_{k-1}, \quad k \geq 1 \quad (55)$$

$$y_k = \theta_k^i x_k + v_k^i, \quad k \geq 1, \quad i = 1, 2,$$

where the initial state  $x_0$  is a zero-mean Gaussian variable with variance  $P_0 = 1$ . The noise processes  $\{w_k; k \geq 0\}$  and  $\{v_k^i; k \geq 1, i = 1, 2\}$ , are defined by

$$w_k = 0.6(\eta_{k+1} + \eta_{k+2}), \quad (56)$$

$$v_k^i = c_i(\eta_k + \eta_{k+1}), \quad i = 1, 2,$$

where the sequence of variables  $\{\eta_k; k \geq 1\}$  is a zero-mean Gaussian white process with variance 0.5. Clearly, according

to assumption (ii), the additive noises  $\{w_k\}$  and  $\{v_k^i\}$  are one-step autocorrelated and two-step cross-correlated with

$$\begin{aligned} Q_{k,k} &= 0.36, & Q_{k,k+1} &= 0.18, \\ S_{k,k}^i &= 0.3c_i, & S_{k-1,k}^i &= 0.6c_i, & S_{k-2,k}^i &= 0.3c_i, \\ R_{k,k}^{ii} &= c_i^2, & R_{k,k+1}^{ii} &= 0.5c_i^2, \\ R_{k,k}^{ij} &= c_i c_j, & R_{k,k+1}^{ij} &= 0.5c_i c_j. \end{aligned} \quad (57)$$

The phenomenon of missing measurements for each sensor is described as follows.

- (1) In the first sensor, a sequence of independent and identically distributed (i.i.d.) random variables,  $\{\theta_k^1; k \geq 1\}$ , is considered, with probability distribution given by

$$\begin{aligned} P[\theta_k^1 = 0] &= 0.1, & P[\theta_k^1 = 0.5] &= 0.5, \\ P[\theta_k^1 = 1] &= 0.4. \end{aligned} \quad (58)$$

If  $\theta_k^1 = 0$ , which occurs with probability 0.1, the state  $x_k$  is missing and the observation  $y_k^1$  contains only noise  $v_k^1$ ; if  $\theta_k^1 = 0.5$ , only partial information of the state  $x_k$  is missing in such observation, which happens with probability 0.5; and, finally, the state is present in the observation  $y_k^1$  with probability 0.4 when  $\theta_k^1 = 1$ . The mean and variance of these variables are easily calculated, being  $\bar{\theta}_k^1 = 0.65$  and  $V_k^{\theta^1} = 0.1025$ , for all  $k$ .

- (2) In the second sensor, a sequence of i.i.d. Bernoulli random variables,  $\{\theta_k^2; k \geq 1\}$ , is considered, with  $P[\theta_k^2 = 1] = p$ ; in this case, if  $\theta_k^2 = 1$  the state  $x_k$  is present in the measurement  $y_k^2$  with probability  $p$ , whereas if  $\theta_k^2 = 0$  such observation only contains additive noise,  $v_k^2$ , with probability  $1-p$ . So, no partial missing information is considered in this sensor.

Clearly, for all  $k$ ,  $\bar{\theta}_k^2 = p$  and  $V_k^{\theta^2} = p(1-p)$ .

To illustrate the feasibility and effectiveness of the proposed estimators we ran a program in MATLAB, in which fifty iterations of the proposed algorithms have been performed considering different values of  $c_i$  and  $p$ . Using simulated values of the state and the corresponding observations, both distributed and centralized filtering estimates of the state are calculated, as well as the corresponding error variances, which provide a measure of the estimation accuracy.

Firstly, for  $p = 0.8$ , the local, centralized, and distributed filtering error variances are displayed in Figure 1 considering the values  $c_1 = 1$  and  $c_2 = 0.5$ . According to Theorem 13, this figure corroborates that the optimal fusion distributed filter performs quite better than each local filter, but lightly worse than the centralized filter. Nevertheless, although the distributed fusion filter has a bit lower accuracy than the centralized one, both filters perform similarly and provide good estimations. Moreover, this slight difference is compensated because the distributed fusion structure is in general

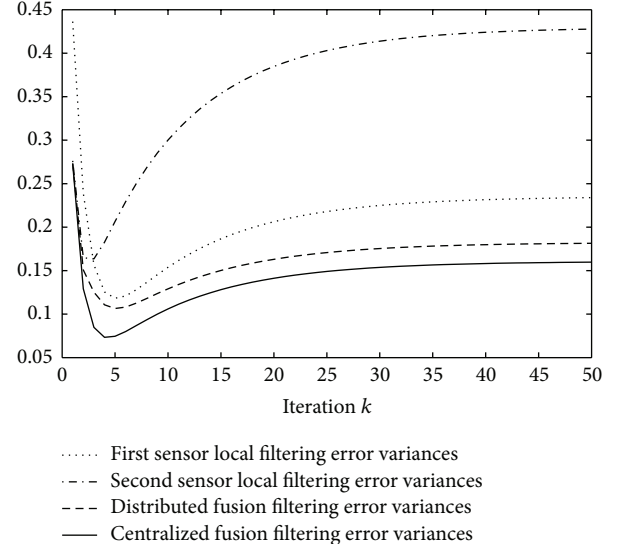


FIGURE 1: Local, centralized, and distributed fusion filtering error variances.

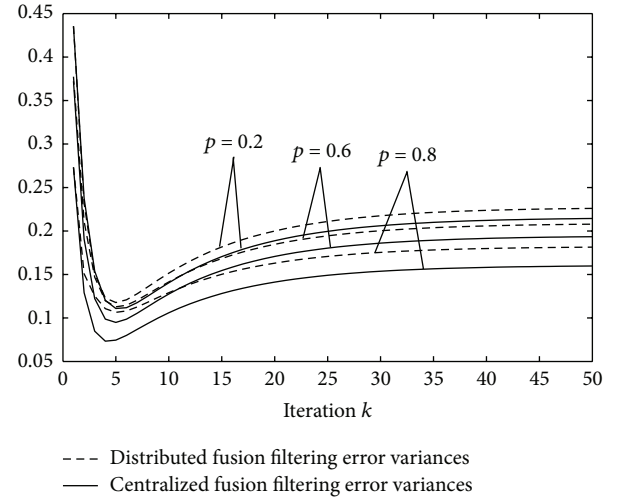


FIGURE 2: Centralized and distributed fusion filtering error variances for  $p = 0.2, 0.6, 0.8$ , when  $c_1 = 1, c_2 = 0.5$ .

more robust, reduces the computational cost, and improves the reliability due to its parallel structure. For these reasons, the distributed filter is generally preferred in practice.

Next, to analyze the performance of the proposed estimators versus the probability that the state  $x_k$  is present in the measurements of the second sensor, the centralized and distributed filtering error variances have been calculated for  $c_1 = 1, c_2 = 0.5$  and different values of the probability  $p = 0.2, 0.6$  and  $0.8$ . The results are displayed in Figure 2; analysis of this figure reveals that as  $p$  increases (or, equivalently, the probability  $1-p$  that the state is missing in the observations from the second sensor decreases), the filtering error variances become smaller and, hence, better estimations are obtained. Also, this figure shows that, for all the considered probability values, the error variances corresponding to the centralized filter are always less than

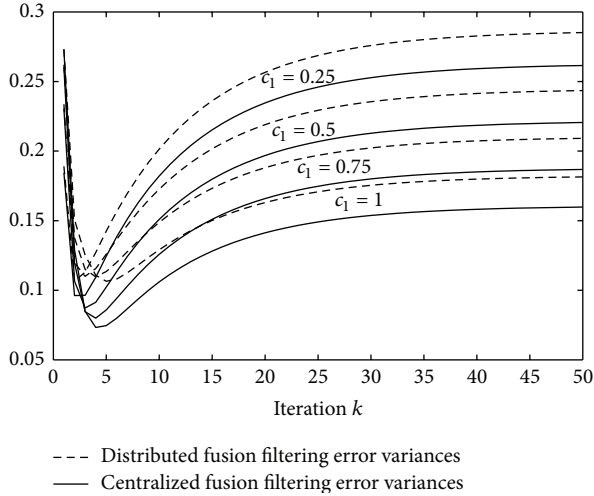


FIGURE 3: Centralized and distributed fusion filtering error variances for  $c_1 = 0.25, 0.5, 0.75, 1$ , when  $c_2 = 0.5$  and  $p = 0.8$ .

those of the distributed filter. Analogous results are obtained for other values of  $c_1$ ,  $c_2$  and the probability  $p$ .

On the other hand, to compare the performance of the estimators for different degrees of correlation between the state and the observation noises, the centralized and distributed filtering error variances have been calculated considering  $c_2 = 0.5$ ,  $p = 0.8$  and different values of  $c_1$ , specifically,  $c_1 = 0.25, 0.5, 0.75$ , and  $1$ . These values provide different correlations between the noise process  $\{w_k\}$  and the first sensor observation noise  $\{v_k^1\}$  and, consequently, different correlations,  $E_k^1$ , between the state and the first sensor observation noise. The error variances are displayed in Figure 3, from which it is inferred that the error variances are smaller (and, consequently, the performance of the estimators is better) as the value  $c_1$  is greater; these results were expected, since the correlation between the state and observations increases with  $c_1$ . Analogous results are obtained for different values of  $c_2$  and other values of the probability  $p$ .

Now, completing the results of the two previous figures, the performance of the filters is analyzed when  $c_2 = 0.5$ , the probability  $p$  is varied from  $0.1$  to  $0.9$ , and the values  $c_1 = 0.25, 0.5, 0.75, 1, 1.25$ , and  $1.5$  are considered. It must be noted that in all the cases examined, the error variances present insignificant variation from a certain iteration on and, consequently, only the values at a specific iteration (viz.,  $k = 50$ ) are shown. The results are presented in Figure 4 which, for the sake of clarity, only displays the distributed filtering error variances. Agreeing with the comments about Figures 2 and 3, this figure shows that, for a fixed value of  $c_1$ , the performance of the estimators improves as  $p$  becomes greater, and for a fixed value of  $p$ , also more accurate estimations are obtained as  $c_1$  increases. Hence, from this figure it is gathered that, as  $c_1$  or  $p$  decreases (which means that the correlation between the state and the first sensor observation noise decreases or the probability that the state is present in the second sensor measurements decreases, resp.), the filtering error variances become greater and, consequently, worse estimations are obtained.

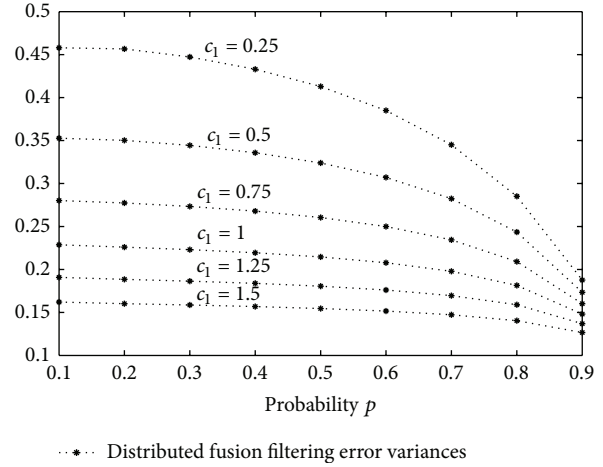


FIGURE 4: Distributed fusion filtering error variances versus  $p$ , with  $c_1 = 0.25, 0.5, 0.75, 1, 1.25, 1.5$ , when  $c_2 = 0.5$ .

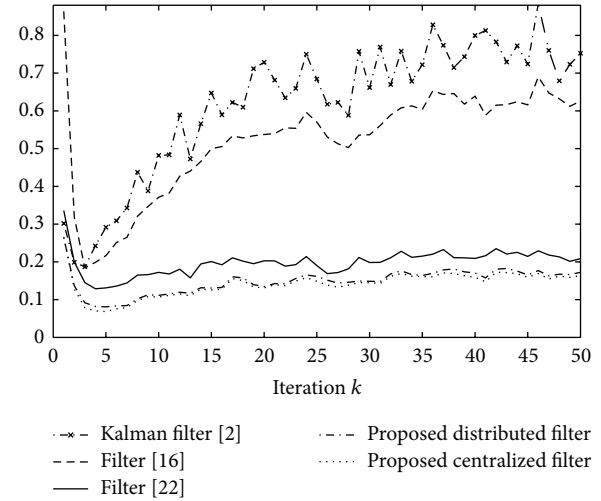


FIGURE 5: Comparison of MSE for different filters.

Finally, a comparative analysis is presented between the classical Kalman filter [2], the Kalman-type filter with correlated and cross-correlated noises given in [16], the filter proposed in [23] for systems with different failure rates in multisensor networks, and the centralized and distributed filters proposed in this paper. For the comparison, the same parameter values as in Figure 1 are considered ( $c_1 = 1$ ,  $c_2 = 0.5$ , and  $p = 0.8$ ).

On the basis of one thousand independent simulations of the mentioned algorithms, a comparison between the different filtering estimates is performed using the mean square error (MSE) criteria. For  $s = 1, \dots, 1000$ , let  $\{x_k^{(s)}, k = 1, \dots, 50\}$  denote the  $s$ th set of artificially simulated data (which is taken as the  $s$ th set of true values of the state), and  $\hat{x}_{k/k}^{(s)}$  the filtering estimate at the sampling time  $k$  in the  $s$ th simulation run. For each algorithm, the filtering MSE at time  $k$  is calculated by  $MSE_k = (1/1000) \sum_{s=1}^{1000} (x_k^{(s)} - \hat{x}_{k/k}^{(s)})^2$ .

The values  $MSE_k$ , for  $k = 1, \dots, 50$ , are displayed in Figure 5 which shows that, for all  $k$ , the proposed centralized

and distributed filters have approximately the same  $MSE_k$  values, which in turn are smaller than the  $MSE_k$  values of the filter in [23] and considerably less than those of the filters [2, 16]. Hence, we can conclude that, according to the MSE criterion, the proposed filtering estimates perform significantly better than other filters in the literature.

## 6. Conclusions

The LS linear estimation problem from missing measurements has been investigated for multisensor linear discrete-time systems with autocorrelated and cross-correlated noises. The main contributions are summarized as follows.

- (1) Using both centralized and distributed fusion methods to process the measurement data from the different sensors, recursive optimal LS linear filtering algorithms are derived by an innovation approach.
- (2) At each sensor, the possibility of missing measurements (i.e., observations containing only partial information about the state or even only noise) is modelled by a sequence of independent random variables taking discrete values over the interval  $[0, 1]$ .
- (3) The multisensor system model considered in the current paper covers those situations where the sensor and process noises are one-step autocorrelated and two-step cross-correlated. Also, one-step cross-correlations between different sensor noises is considered. This correlation assumption is valid in a wide spectrum of applications, for example, in target tracking systems with process and measurement noises dependent on the system state, or situations where a target is observed by multiple sensors and all of them operate in the same noisy environment. Nevertheless, the current study can be extended to more general systems involving finite-step autocorrelated and cross-correlated noises with no difficulty, except for a greater complexity in the mathematical expressions.
- (4) The applicability of the proposed centralized and distributed filtering algorithms is illustrated by a numerical simulation example, where a scalar state process generated by a first-order autoregressive model is estimated from missing measurements coming from two sensors with autocorrelated and cross-correlated noises. The results confirm that centralized and distributed fusion estimators have approximately the same error variances, with a slight inferiority of the distributed one which is compensated by a reduced computational burden and reduced communication demands for the sensor networks. Also, compared with some existing estimation methods, the proposed algorithms provide better estimations in the mean square error sense.

## Appendix

### Proof of Theorem 5

From (10), expression (15) for the state filter  $\hat{x}_{k/k}$  in terms of the one-stage predictor  $\hat{x}_{k/k-1}$  is immediately clear.

Expression (16) for the state predictor  $\hat{x}_{k/k-1}$  is obtained as follows:

$$\begin{aligned}
 \hat{x}_{k/k-1} &= \sum_{s=1}^{k-1} E \left[ x_k \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s \\
 &= \sum_{s=1}^{k-1} E \left[ (F_{k-1} x_{k-1} + w_{k-1}) \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s \\
 &= \sum_{s=1}^{k-1} F_{k-1} E \left[ x_{k-1} \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s \\
 &\quad + E \left[ w_{k-1} \mu_{k-1}^T \right] \Pi_{k-1,k-1}^{-1} \mu_{k-1} \\
 &= F_{k-1} \hat{x}_{k-1/k-1} + \mathcal{W}_{k-1,k-1} \Pi_{k-1,k-1}^{-1} \mu_{k-1}, \quad k \geq 2,
 \end{aligned} \tag{A.1}$$

and clearly,  $\hat{x}_{1/0} = E[x_1] = F_0 E[x_0] = F_0 \hat{x}_{0/0}$ .

Now we show expression (17) for the innovation,  $\mu_k = y_k - \hat{y}_{k/k-1}$ , for which it is enough to obtain an expression for  $\hat{y}_{k/k-1}$ . A similar reasoning to that used to prove (16) leads to

$$\begin{aligned}
 \hat{y}_{k/k-1} &= \sum_{s=1}^{k-1} E \left[ y_k \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s \\
 &= \sum_{s=1}^{k-1} E \left[ (\Theta_k H_k x_k + v_k) \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s \\
 &= \bar{\Theta}_k H_k \sum_{s=1}^{k-1} E \left[ x_k \mu_s^T \right] \Pi_{s,s}^{-1} \mu_s + E \left[ v_k \mu_{k-1}^T \right] \Pi_{k-1,k-1}^{-1} \mu_{k-1} \\
 &= \bar{\Theta}_k H_k \hat{x}_{k/k-1} + \mathcal{V}_{k,k-1} \Pi_{k-1,k-1}^{-1} \mu_{k-1}, \quad k \geq 2,
 \end{aligned} \tag{A.2}$$

with  $\hat{y}_{1/0} = E[y_1] = \bar{\Theta}_1 H_1 E[x_1] = \bar{\Theta}_1 H_1 \hat{x}_{1/0}$ . Hence,

$$\begin{aligned}
 \hat{y}_{k/k-1} &= \bar{\Theta}_k H_k \hat{x}_{k/k-1} + \mathcal{V}_{k,k-1} \Pi_{k-1,k-1}^{-1} \mu_{k-1}, \quad k \geq 2, \\
 \hat{y}_{1/0} &= \bar{\Theta}_1 H_1 \hat{x}_{1/0}
 \end{aligned} \tag{A.3}$$

and expression (17) for the innovation is clear.

Next, expression (18) for the matrix  $\mathcal{X}_{k,k} = E[x_k y_k^T] - E[x_k \hat{y}_{k/k-1}^T]$  is derived. From (6) and the independence assumption, it is clear that

$$E \left[ x_k y_k^T \right] = D_k H_k^T \bar{\Theta}_k + E_k, \quad k \geq 1. \tag{A.4}$$

From expression (A.3) for  $\hat{y}_{k/k-1}$  and the OPL,  $E[x_k \hat{y}_{k/k-1}^T]$  is calculated as follows:

$$\begin{aligned}
& E[x_k \hat{y}_{k/k-1}^T] \\
&= E[x_k \hat{x}_{k/k-1}^T] H_k^T \bar{\Theta}_k + E[x_k \mu_{k-1}^T] \Pi_{k-1, k-1}^{-1} \mathcal{V}_{k, k-1}^T \\
&= E[\hat{x}_{k/k-1} \hat{x}_{k/k-1}^T] H_k^T \bar{\Theta}_k + \mathcal{X}_{k, k-1} \Pi_{k-1, k-1}^{-1} \mathcal{V}_{k, k-1}^T \\
&= (D_k - P_{k/k-1}) H_k^T \bar{\Theta}_k + \mathcal{X}_{k, k-1} \Pi_{k-1, k-1}^{-1} \mathcal{V}_{k, k-1}^T, \\
& \hspace{15em} k \geq 2, \\
& E[x_1 \hat{y}_{1/0}^T] = (D_1 - P_{1/0}) H_1^T \bar{\Theta}_1.
\end{aligned} \tag{A.5}$$

By subtraction of the above expectations, expression (18) for  $\mathcal{X}_{k, k} = E[x_k y_k^T] - E[x_k \hat{y}_{k/k-1}^T]$  is obtained. From (1), expression (19) for  $\mathcal{X}_{k, k-1} = E[x_k \mu_{k-1}^T]$  is immediately clear.

Expression (20) for the prediction error covariance matrix,  $P_{k/k-1}$  is easily obtained by using (1) and (16); and, from (1) and (15), expression (21) for  $\mathcal{F}_k = E[\tilde{x}_{k/k} w_k^T] = E[x_k w_k^T] - E[\hat{x}_{k/k} w_k^T]$  is also obvious. Expression (22) for the filtering error covariance matrix,  $P_{k/k}$ , is immediately derived by using (15).

Finally, we prove expression (23) for the innovation covariance matrix  $\Pi_{k, k} = E[y_k y_k^T] - E[\hat{y}_{k/k-1} \hat{y}_{k/k-1}^T]$ . From (6) and using (9), we have that

$$\begin{aligned}
E[y_k y_k^T] &= E[\theta_k \theta_k^T] \circ (H_k D_k H_k^T) + R_{k, k} \\
&+ \bar{\Theta}_k H_k E_k + E_k^T H_k^T \bar{\Theta}_k.
\end{aligned} \tag{A.6}$$

Using now (A.3) for  $\hat{y}_{k/k-1}$  and property (9), and taking into account that, from the OPL,  $E[\hat{x}_{k/k-1} \mu_{k-1}^T] = E[x_k \mu_{k-1}^T] = \mathcal{X}_{k, k-1}$ , the following identity holds:

$$\begin{aligned}
E[\hat{y}_{k/k-1} \hat{y}_{k/k-1}^T] &= (\bar{\Theta}_k \bar{\Theta}_k^T) \circ (H_k (D_k - P_{k/k-1}) H_k^T) \\
&+ \mathcal{V}_{k, k-1} \Pi_{k-1, k-1}^{-1} \mathcal{V}_{k, k-1}^T \\
&+ \bar{\Theta}_k H_k \mathcal{X}_{k, k-1} \Pi_{k-1, k-1}^{-1} \mathcal{V}_{k, k-1}^T \\
&+ \mathcal{V}_{k, k-1} \Pi_{k-1, k-1}^{-1} \mathcal{X}_{k, k-1}^T H_k^T \bar{\Theta}_k.
\end{aligned} \tag{A.7}$$

From the above expectations and after some manipulations, expression (23) for the innovation covariance matrix  $\Pi_{k, k}$  is obtained.

## Acknowledgments

This research is supported by Ministerio de Ciencia e Innovación (Programa FPU and Grant no. MTM2011-24718) and Junta de Andalucía (Grant no. P07-FQM-02701).

## References

- [1] C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*, vol. 17 of *Springer Series in Information Sciences*, Springer, Berlin, Germany, 1999.
- [2] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering D*, vol. 82, pp. 35–45, 1960.
- [3] J. D. Jiménez-López, J. Linares-Pérez, S. Nakamori, R. Caballero-Águila, and A. Hermoso-Carazo, "Signal estimation based on covariance information from observations featuring correlated uncertainty and coming from multiple sensors," *Signal Processing*, vol. 88, no. 12, pp. 2998–3006, 2008.
- [4] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Least-squares linear estimators using measurements transmitted by different sensors with packet dropouts," *Digital Signal Processing*, vol. 22, no. 6, pp. 1118–1125, 2012.
- [5] D. Y. Kim and M. Jeon, "Robust distributed Kalman filter for wireless sensor networks with uncertain communication channels," *Mathematical Problems in Engineering*, vol. 2012, Article ID 238597, 12 pages, 2012.
- [6] J. Sun and C. Zhang, "Distributed optimal fusion estimators for multi-sensor systems with bounded random measurement delays and packet dropouts," *Journal of Computational Information Systems*, vol. 8, no. 10, pp. 4087–4094, 2012.
- [7] M. J. García-Ligero, A. Hermoso-Carazo, and J. Linares-Pérez, "Distributed and centralized fusion estimation from multiple sensors with Markovian delays," *Applied Mathematics and Computation*, vol. 219, no. 6, pp. 2932–2948, 2012.
- [8] W. A. Zhang, G. Feng, and L. Yu, "Multi-rate distributed fusion estimation for sensor networks with packet losses," *Automatica*, vol. 48, no. 9, pp. 2016–2028, 2012.
- [9] J. Ma and S. Sun, "Information fusion estimators for systems with multiple sensors of different packet dropout rates," *Information Fusion*, vol. 12, no. 3, pp. 213–222, 2011.
- [10] E. Song, Y. Zhu, J. Zhou, and Z. You, "Optimal Kalman filtering fusion with cross-correlated sensor noises," *Automatica*, vol. 43, no. 8, pp. 1450–1456, 2007.
- [11] X. Wang, S. L. Sun, K. H. Ding, and J. Y. Xue, "Weighted measurement fusion white noise deconvolution filter with correlated noise for multisensor stochastic systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 257619, 16 pages, 2012.
- [12] J. Feng and M. Zeng, "Optimal distributed Kalman filtering fusion for a linear dynamic system with cross-correlated noises," *International Journal of Systems Science*, vol. 43, no. 2, pp. 385–398, 2012.
- [13] E. Song, Y. Zhu, and Z. You, "The Kalman type recursive state estimator with a finite-step correlated process noises," in *Proceedings of the IEEE International Conference on Automation and Logistics (ICAL '08)*, pp. 196–200, Qingdao, China, September 2008.
- [14] A. Fu, Y. Zhu, and E. Song, "The optimal Kalman type state estimator with multi-step correlated process and measurement noises," in *Proceedings of the International Conference on Embedded Software and Systems (ICCESS '08)*, pp. 215–220, Sichuan, China, July 2008.
- [15] J. Feng, Z. Wang, and M. Zeng, "Optimal robust non-fragile Kalman-type recursive filtering with finite-step autocorrelated noises and multiple packet dropouts," *Aerospace Science and Technology*, vol. 15, no. 6, pp. 486–494, 2011.
- [16] J. Feng, Z. Wang, and M. Zeng, "Distributed weighted robust Kalman filter fusion for uncertain systems with autocorrelated and cross-correlated noises," *Information Fusion*, vol. 14, pp. 78–86, 2013.

- [17] S. L. Sun and J. Ma, "Optimal filtering and smoothing for discrete-time stochastic singular systems," *Signal Processing*, vol. 87, no. 1, pp. 189–201, 2007.
- [18] J. Ma and S. Sun, "Optimal linear estimators for systems with random sensor delays, multiple packet dropouts and uncertain observations," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5181–5192, 2011.
- [19] Z. Wang, F. Yang, D. W. C. Ho, and X. Liu, "Robust finite-horizon filtering for stochastic systems with missing measurements," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 437–440, 2005.
- [20] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Fixed-interval smoothing algorithm based on covariances with correlation in the uncertainty," *Digital Signal Processing*, vol. 15, no. 2, pp. 207–221, 2005.
- [21] Z. Wang, F. Yang, D. W. C. Ho, and X. Liu, "Robust  $H_\infty$  filtering for stochastic time-delay systems with missing measurements," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2579–2587, 2006.
- [22] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "New recursive estimators from correlated interrupted observations using covariance information," *International Journal of Systems Science*, vol. 36, no. 10, pp. 617–629, 2005.
- [23] F. O. Hounkpevi and E. E. Yaz, "Robust minimum variance linear state estimators for multiple sensors with different failure rates," *Automatica*, vol. 43, no. 7, pp. 1274–1280, 2007.
- [24] X. Qu and J. Zhou, "The optimal robust finite-horizon Kalman filtering for multiple sensors with different stochastic failure rates," *Applied Mathematics Letters*, vol. 26, no. 1, pp. 80–86, 2013.
- [25] R. Caballero-Águila, A. Hermoso-Carazo, and J. Linares-Pérez, "Linear and quadratic estimation using uncertain observations from multiple sensors with correlated uncertainty," *Signal Processing*, vol. 91, no. 2, pp. 330–337, 2011.
- [26] B. Shen, Z. D. Wang, and Y. S. Hung, "Distributed  $H_\infty$ -consensus filtering in sensor networks with multiple missing measurements: the finite-horizon case," *Automatica*, vol. 46, no. 10, pp. 1682–1688, 2010.
- [27] M. Liu, D. Qi, S. Zhang, M. Qiu, and S. Zheng, "Optimal  $H_\infty$  fusion filters for a class of discrete-time intelligent systems with time delays and missing measurement," *Neurocomputing*, vol. 74, no. 18, pp. 3741–3752, 2011.
- [28] G. Wei, Z. Wang, and H. Shu, "Robust filtering with stochastic nonlinearities and multiple missing measurements," *Automatica*, vol. 45, no. 3, pp. 836–841, 2009.
- [29] J. Hu, Z. Wang, H. Gao, and L. K. Stergioulas, "Extended Kalman filtering with stochastic nonlinearities and multiple missing measurements," *Automatica*, vol. 48, no. 9, pp. 2007–2015, 2012.
- [30] S. L. Sun and Z. L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

