Applied Mathematics and Computation 226 (2014) 548–563

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Information fusion algorithms for state estimation in multi-sensor systems with correlated missing measurements

R. Caballero-Águila [a,*], I. García-Garrido [b], J. Linares-Pérez [b]

[a] Departamento de Estadística e I.O., Universidad de Jaén, Paraje Las Lagunillas s/n, 23071 Jaén, Spain
[b] Departamento de Estadística e I.O., Universidad de Granada, Campus Fuentenueva s/n, 18071 Granada, Spain

ARTICLE INFO

Keywords:
Least-squares estimation
Centralized fusion estimation
Distributed fusion estimation
Missing observations
Multi-sensor systems

ABSTRACT

In this paper, centralized and distributed fusion estimation problems in linear discrete-time stochastic systems with missing observations coming from multiple sensors are addressed. At each sensor, the Bernoulli random variables describing the phenomenon of missing observations are assumed to be correlated at instants that differ m units of time. By using an innovation approach, recursive linear filtering and fixed-point smoothing algorithms for the centralized fusion problem are derived in the least-squares sense. The distributed fusion estimation problem is addressed based on the distributed fusion criterion weighted by matrices in the linear minimum variance sense. For each sensor subsystem, local least-squares linear filtering and fixed-point smoothing estimators are given and the estimation error cross-covariance matrices between any two sensors are derived to obtain the distributed fusion estimators.

The performance of the proposed estimators is illustrated by numerical simulation examples where scalar and two-dimensional signals are estimated from missing observations coming from two sensors, and the estimation accuracy is analyzed for different missing probabilities and different values of m.

1. Introduction

During the past decades, the estimation problem in multi-sensor systems has motivated a significant amount of research due to its increasing application in many engineering fields (for example, in the fields of computer and communication) 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 (see for example [1] and references therein).

Although the use of sensor networks offers several advantages such as easier installation, simpler maintenance and reduced cost, since the measured data are sent to a processing center via a communication network, the unreliable network characteristics usually leads to other problems such as missing measurements (i.e. measured outputs containing noise only, also called uncertain observations), random communication packet losses and/or delays. These problems may occur in practical applications for many different reasons, such as random failures in the transmission mechanism, accidental loss of some measurements or data inaccessibility at certain times, etc.

The estimation problems in systems with only one or several of the aforementioned uncertainties has attracted considerable research attention (see e.g. [2–10] and references therein). To be more specific, the estimation problem in discrete-time nonlinear systems with uncertain observations has been studied in [2,3]; the estimation problem from measurements

* Corresponding author.
E-mail addresses: raguila@ujaen.es (R. Caballero-Águila), irenegarciag@ugr.es (I. García-Garrido), jlinares@ugr.es (J. Linares-Pérez).

0096-3003/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.10.068

subject to random delay which does not exceed one sampling period is addressed in [4,5]; modifications of conventional linear estimation algorithms for systems with packet dropouts have been proposed in [6,7]; the optimal linear estimation problem for systems with random delays and packet dropouts has been considered in [8], and for systems including the three sources of uncertainty in [9,10].

All the above papers consider a single sensor or multiple sensors with the same uncertainty characteristics. However, this is not a realistic assumption in several application fields, for instance, in networked communication systems involving heterogeneous measurement devices, and hence multiple-sensor systems whose statistical properties are not necessarily the same for all the sensors require the derivation of new estimation algorithms, as the conventional ones cannot be applied directly. A basic matter in systems with multiple sensors is how to fuse the measurement data from the different sensors to address the estimation problem. Mainly two methods are used to process the measured sensor data in estimation problems with sensor networks: 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; specifically, the observations from multiple sensors are stacked as one sensor measurement (with greater dimension) and, hence, it does not require a particular fusion rule. In [11,12] centralized linear minimum variance estimators are derived considering multiple sensors with different failure rates, and different delay rates are considered in [13,14]. The optimal centralized problem, also in linear minimum variance sense, is investigated in [15,16] for systems with multiple sensors of different packet dropout rates.

Nevertheless, as it is known, the centralized approach has several drawbacks due to augmentation, such as poor survivability, reliability, heavy communication and expensive computational cost, and various distributed fusion algorithms have been proposed to reduce these drawbacks. In the distributed fusion method, each sensor estimates the state based on its own measurement data, and then it sends such estimate to the fusion center for fusion according to a certain information fusion criterion. For example, under the assumption of normal distribution, a distributed fusion estimator is proposed in [17] based in maximum likelihood criterion, and the distributed fusion criterion weighted by matrices in the linear minimum variance sense is established in [18], which is equivalent to the maximum likelihood fusion criterion under normality assumption.

Recently, more attention has been paid to the distributed fusion estimation in networked systems with unreliable network transmission (see e.g. [19–23] and references therein). Distributed fusion estimators for multi-sensor systems with random delays were presented in [19,20], and for systems with packet dropouts in [15,21]. Simultaneous packet delays and dropouts are considered in [22,23].

Compared with the number of papers about multi-sensor systems with random communication packet delays and/or dropouts, to the best of the authors knowledge, the literature regarding distributed fusion estimation in multi-sensor systems with missing measurements is relatively scarcer, and most existing papers use independent Bernoulli variables to model the missing measurements [24–26].

In [12] centralized linear minimum variance estimators are obtained removing the assumption of independence of the Bernoulli variables describing the phenomenon of missing measurements. Specifically, different sequences of Bernoulli random variables correlated at consecutive sampling times are considered to model the uncertainty at each sensor. This form of correlation covers practical situations where the state cannot be missing in two successive observations and hence, transmission models with stand-by sensors, which are immediately substituted when a failure occurs, are appropriately managed with this model. However, the failed sensor may not be replaced immediately but after m instants of time; in such situations, correlation among the random variables modeling the missing measurements at times k and $k+m$ must be considered and new algorithms must be deduced.

In response to the above considerations, this paper deals with the centralized and distributed fusion estimation problems in multi-sensor systems with missing measurements when, at each sensor, the random variables modeling the missing measurements are correlated at instants that differ m units of time. The main contributions can be summarized as follows: (i) centralized fusion filtering and fixed-point smoothing algorithms are proposed in multi-sensor systems with correlated missing measurements and the correlation form considered covers certain models in which the state cannot be missing in $m+1$ consecutive observations, thus generalizing the results in [12]; (ii) the distributed fusion filtering and fixed-point smoothing problems are addressed in multi-sensor systems with missing measurements.

The paper is organized as follows: in Section 2 the problem formulation is described; more specifically, we introduce the linear state transition model perturbed by white noise, and the measurement model affected by additive white noise and multiplicative noise describing the phenomenon of missing measurements. Also, the pertinent assumptions to address the least-squares linear estimation problem are established. In Section 3, by using an innovation analysis approach and the orthogonal projection Lemma, recursive algorithms for the centralized fusion filter and fixed-point smoothers are presented (the derivation has been deferred to Appendices A and B). Next, in Section 4, the local least-squares linear estimators and the error cross-covariance matrices between any two local estimates are derived, then distributed fusion estimators are obtained based on the optimal fusion criterion weighted by matrices in the linear minimum variance sense. The performance of the proposed estimators is illustrated in Section 5 by two numerical simulation examples where local, distributed and centralized fusion estimators are compared. The paper ends with some concluding remarks in Section 6.

Notation: The notation used is standard. A^T represents the transpose of A, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, $\mathbb{R}^{m\times n}$ is the set of all real matrices of dimension $m\times n$, and I and 0 represent the identity matrix and zero matrix of appropriate dimension, respectively. The shorthand $Diag(M_1,\ldots,M_r)$ denotes a block diagonal matrix whose diagonal blocks are the matrices $M_1,\ldots,M_r$. If the dimension of a matrix is not explicitly stated, it is assumed to be compatible for algebraic

operations. For time-varying matrices $F_k,\ k \geqslant 0$, the product $F_{k-1}, \ldots, F_i$ is denoted by $\mathbb{F}_{k,i}$. The Hadamard product of matrices C and D is denoted by $\circ$ ($[C \circ D]_{ij} = C_{ij}D_{ij}$).

Also, for arbitrary random vectors α and β, the following notation is used throughout the paper: $\mathrm{Cov}[\alpha, \beta] = E\left[(\alpha - E[\alpha])(\beta - E[\beta])^T\right]$ and $\mathrm{Cov}[\alpha] = \mathrm{Cov}[\alpha, \alpha]$, where $E[\cdot]$ stands for the mathematical expectation operator. $\widehat{\alpha}$ denotes the estimator of α and $\widetilde{\alpha} = \alpha - \widehat{\alpha}$ the estimation error.

2. Problem formulation

The problem at hand is to determine the least-squares (LS) linear filtering and fixed-point smoothing estimators of the state in linear discrete-time stochastic systems with missing measurements coming from multiple sensors. In this section, we present the system model and the hypotheses about the state and noise processes involved.

Consider a class of discrete-time linear stochastic systems with missing measurements coming from r sensors; the phenomenon of missing measurements (that is, observations containing only noise) occurs randomly and, for each sensor, a different sequence of Bernoulli random variables is used to model this phenomenon. Specifically, the following system is considered

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \quad k \geqslant 1, \tag{1}$$

$$y_k^i = \theta_k^i H_k^i x_k + v_k^i, \quad k \geqslant 1, \quad i = 1, 2, \ldots, r, \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k^i \in \mathbb{R},\ i = 1, 2, \ldots, r$, is the measurement collected by sensor i at sampling time k, $\{w_k;\ k \geqslant 0\}$ and $\{v_k^i;\ k \geqslant 1\},\ i = 1, 2, \ldots, r$, are noise sequences, and $\{\theta_k^i;\ k \geqslant 1\},\ i = 1, 2, \ldots, r$, are Bernoulli random variables whose values – one or zero – indicate whether the state is present or missing in the corresponding measure. F_k and $H_k^i,\ i = 1, 2, \ldots, r$, are known time-varying matrices with compatible dimensions, superscript i denotes the ith sensor, and r is the number of sensors.

As is known, to address the LS linear estimation problem the state and the observations are required to have finite second-order moments; the following assumptions specify the first- and second-order moments required in the study of this problem, as well as the statistical properties assumed about the initial state and noise processes.

Assumption 1. The initial state x_0 is a random vector with $E[x_0] = \overline{x}_0$ and $Cov[x_0] = P_0$.

Assumption 2. The additive noises $\{w_k; k \geqslant 0\}$ and $\{v_k^i;\ k \geqslant 1\}$, $i = 1, 2, \ldots, r$, are zero-mean white sequences with covariances $Cov[w_k] = Q_k$ and $Cov[v_k^i] = R_k^i$, respectively.

Assumption 3. The multiplicative noises $\{\theta_k^i;\ k \geqslant 1\},\ i = 1, 2, \ldots, r$, are sequences of Bernoulli random variables with $P[\theta_k^i = 1] = \overline{\theta}_k^i$. For $i = 1, 2, \ldots, r$, the variables θ_k^i and θ_s^i are independent for $|k - s| \neq 0, m$, and $Cov[\theta_k^i, \theta_s^i] = K_{k,s}^{\theta^i}$ are known for $|k - s| = 0, m$.

Assumption 4. The initial state x_0 and the noise processes $\{w_k;\ k \geqslant 0\}$, $\{v_k^i;\ k \geqslant 1\}$ and $\{\theta_k^i;\ k \geqslant 1\}$, for $i = 1, 2, \ldots, r$, are mutually independent.

Remark 1. Note that, when $\theta_k^i = 1$, which occurs with known probability $\overline{\theta}_k^i$, the state x_k is present in the measure y_k^i coming from the ith sensor at time k, whereas if $\theta_k^i = 0$ the state is missing in the measured data at time k, which means that such observation only contains additive noise v_k^i with probability $1 - \overline{\theta}_k^i$. To model the phenomenon of missing measurements at each sensor, different sequences of Bernoulli random variables correlated at instants that differ m units of time are considered. This special form of correlation allows us to consider certain class of systems in which the state cannot be missing in $m + 1$ consecutive observations; specifically, sensor networks where sensor failures may happen and a failed sensor is substituted not immediately, but m sampling times after having failed. For instance, consider that, as in Section 5, $\theta_k^i = 1 - \gamma_{k+m}^i(1 - \gamma_k^i)$, with $\{\gamma_k^i;\ k \geqslant 1\}$ sequences of independent Bernoulli random variables. Hence, if $\theta_k^i = 0$, then $\gamma_{k+m}^i = 1$ and $\gamma_k^i = 0$, and consequently $\theta_{k+m}^i = 1$; this fact guarantees that, if the state is missing at time k, the output measurement at time $k + m$ necessarily contains the state. Therefore, there cannot be more than m consecutive measured data consisting of noise only.

Remark 2. From Assumption 3, θ_k^i and θ_s^i are independent for $|k - s| \neq 0, m$, and hence $K_{k,s}^{\theta^i} = 0$ for $|k - s| \neq 0, m$. Also, it is immediate that $K_{k,k}^{\theta^i} = \overline{\theta}_k^i(1 - \overline{\theta}_k^i)$ and $K_{k,k-m}^{\theta^i} = E[\theta_k^i\theta_{k-m}^i] - \overline{\theta}_k^i\overline{\theta}_{k-m}^i$.

Remark 3. From Assumption 4, the Bernoulli sequences as well as the observation noises are independent from sensor to sensor. This condition is not necessary to deduce the centralized estimators and clearly it is not involved in the derivation of the local estimators. Such condition is just used to obtain the cross-covariance matrices of the local estimation errors, which are necessary to determine the matrix weights of the distributed fusion estimators.

Our aim is to solve the LS estimation problem of the state x_k based on the received measurements $\{y_1^i, y_2^i, \ldots, y_k^i, \ldots, y_{k+N}^i\}$, $N \geqslant 0$, $i = 1, 2, \ldots, r$, by using centralized and distributed fusion methods to process the measured sensor data. More specifically, our aim can be stated as follows:

(i) *Centralized fusion estimation problem.* Consider that all measurement data coming from r sensors are transmitted to a fusion center for being processed, and our aim is to obtain the LS linear filter, $\widehat{x}_{k/k}$, and fixed-point smoother, $\widehat{x}_{k/k+N}$, $N \geqslant 1$, by recursive algorithms.

(ii) *Distributed fusion estimation problem.* Firstly, recursive algorithms to obtain local LS linear filters, $\widehat{x}_{k/k}^i$, and fixed-point smoothers, $\widehat{x}_{k/k+N}^i$, $N \geqslant 1$, for $i = 1, 2, \ldots, r$, are derived. Secondly, distributed matrix-weighted fusion estimators $\widehat{x}_{k/k+N}^0$, $N \geqslant 0$, are established by applying the optimal information fusion criterion weighted by matrices in the linear minimum variance sense [18].

Remark 4. In both cases, recursive algorithms for the LS linear estimators will be established using an innovation approach and the orthogonal projection Lemma (OPL). Since the observations are generally nonorthogonal vectors, through the *Gram-Schmidt orthogonalization procedure*, the set of observations is transformed into an equivalent set of orthogonal vectors, *innovations*, defined as the differences between each observation and the one-stage observation predictor. The fact that the innovation process is uniquely determined by the observations allows us to state that the LS linear estimator of the state based on the observations is equal to the LS linear estimator of the state based on the innovations. The advantage of this approach comes from the fact that the innovations constitute a white process, and the expression of the estimators as linear combination of the innovations provide the starting point to derive the recursive filtering and fixed-point smoothing algorithms.

3. Centralized fusion estimation

In this section, our aim is to obtain the optimal (under the LS criterion) linear estimator by the centralized fusion method, in which all the measurement data coming from r sensors are transmitted to a central site for being processed.

For this purpose, denoting $y_k = (y_k^1, \ldots, y_k^r)^T$, $v_k = (v_k^1, \ldots, v_k^r)^T$, $H_k = (H_k^{1T}, \ldots, H_k^{rT})^T$ and $\Theta_k = Diag(\theta_k^1, \ldots, \theta_k^r)$, Eq. (2) is equivalent to the following stacked measurement equation

$$y_k = \Theta_k H_k x_k + v_k, \quad k \geqslant 1. \tag{3}$$

Remark 5. The following properties of the noises in (3) are easily inferred from the model assumptions stated in the previous Section 2:

- The additive noise $\{v_k;\ k \geqslant 1\}$ is a zero-mean white process with covariance matrix $R_k = Diag(R_k^1, \ldots, R_k^r)$, $\forall k \geqslant 1$.
- The random matrices $\{\Theta_k;\ k \geqslant 1\}$ satisfy

$$E[\Theta_k] = \overline{\Theta}_k = Diag(\overline{\theta}_k^1, \ldots, \overline{\theta}_k^r),$$

$$E[(\Theta_k - \overline{\Theta}_k)^2] = \overline{\Theta}_k(I - \overline{\Theta}_k),$$

$$E[(\Theta_k - \overline{\Theta}_k)(\Theta_{k-m} - \overline{\Theta}_{k-m})] = Diag(K_{k,k-m}^{\theta^1}, \ldots, K_{k,k-m}^{\theta^r}).$$

- The initial state x_0 and the noise processes $\{w_k;\ k \geqslant 0\}$, $\{v_k;\ k \geqslant 1\}$ and $\{\Theta_k;\ k \geqslant 1\}$ are mutually independent.

Remark 6. By denoting $\theta_k = (\theta_k^1, \ldots, \theta_k^r)^T$, it is clear that $Cov[\theta_k] = K_{k,k}^{\theta} = \overline{\Theta}_k(I - \overline{\Theta}_k)$ and $Cov[\theta_k, \theta_{k-m}] = K_{k,k-m}^{\theta} = Diag(K_{k,k-m}^{\theta^1}, \ldots, K_{k,k-m}^{\theta^r})$. Moreover, for any random matrix G independent of $\{\Theta_k;\ k \geqslant 1\}$, using the Hadamard product, it is easily deduced [11] that $E[\Theta_k G \Theta_s] = E[\theta_k \theta_s^T] \circ E[G]$. Particularly, the next property (which will be needed later) is immediately clear

$$E[(\Theta_k - \overline{\Theta}_k)G(\Theta_{k-m} - \overline{\Theta}_{k-m})] = K_{k,k-m}^{\theta} \circ E[G]. \tag{4}$$

In the following theorems, using an innovation approach and the OPL, recursive algorithms for the linear filter, $\widehat{x}_{k/k}$, (Theorem 1) and the fixed-point smoother, $\widehat{x}_{k/k+N}$, for fixed k and $N \geqslant 1$, (Theorem 2) are derived.

Theorem 1. *For the system model (1) and measurement model (3), under Assumptions 1–4, the LS linear filter $\widehat{x}_{k/k}$ is obtained as*

$$\widehat{x}_{k/k} = \widehat{x}_{k/k-1} + S_{k,k}\Pi_k^{-1}\nu_k, \quad k \geqslant 1; \quad \widehat{x}_{0/0} = \overline{x}_0, \tag{5}$$

where the state predictor, $\widehat{x}_{k/k-1}$, is given by

$$\widehat{x}_{k/k-1} = F_{k-1}\widehat{x}_{k-1/k-1}, \quad k \geqslant 1. \tag{6}$$

The innovation, ν_k, satisfies

$$\begin{aligned} &\nu_k = y_k - \overline{\Theta}_k H_k \widehat{x}_{k/k-1}, \quad k \leqslant m, \\ &\nu_k = y_k - \overline{\Theta}_k H_k \widehat{x}_{k/k-1} + \Psi_{k,k-m}\left(\nu_{k-m} - \sum_{i=1}^{m-1} T_{k-i,k-m}^T \Pi_{k-i}^{-1}\nu_{k-i}\right), \quad k > m, \end{aligned} \tag{7}$$

where $\Psi_{k,k-m} = K_{k,k-m}^{\theta} \circ \left(H_k \mathbb{F}_{k,k-m} D_{k-m} H_{k-m}^T\right)\Pi_{k-m}^{-1}$, with $D_k = E[x_k x_k^T]$ recursively calculated from

$$D_k = F_{k-1} D_{k-1} F_{k-1}^T + Q_{k-1}, \quad k \geqslant 1; \quad D_0 = P_0 + \overline{x}_0 \overline{x}_0^T. \tag{8}$$

The matrices $T_{k,k-i}$ are determined by

$$\begin{aligned} &T_{k,k-i} = \overline{\Theta}_k H_k \mathbb{F}_{k,k-i} S_{k-i,k-i}, \quad 2 \leqslant k \leqslant m, \quad 1 \leqslant i \leqslant k-1, \\ &T_{k,k-i} = \overline{\Theta}_k H_k \mathbb{F}_{k,k-i} S_{k-i,k-i} - \Psi_{k,k-m} T_{k-i,k-m}^T, \quad k > m, \quad 1 \leqslant i \leqslant m-1. \end{aligned} \tag{9}$$

The matrix $S_{k,k}$ is calculated by

$$\begin{aligned} &S_{k,k} = P_{k/k-1} H_k^T \overline{\Theta}_k, \quad k \leqslant m, \\ &S_{k,k} = P_{k/k-1} H_k^T \overline{\Theta}_k - \left(\mathbb{F}_{k,k-m} S_{k-m,k-m} - \sum_{i=1}^{m-1} \mathbb{F}_{k,k-i} S_{k-i,k-i} \Pi_{k-i}^{-1} T_{k-i,k-m}\right)\Psi_{k,k-m}^T, \quad k > m, \end{aligned} \tag{10}$$

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

$$P_{k/k-1} = F_{k-1} P_{k-1/k-1} F_{k-1}^T + Q_{k-1}, \quad k \geqslant 1,$$

with $P_{k/k}$, the filtering error covariance matrix, given by

$$P_{k/k} = P_{k/k-1} - S_{k,k}\Pi_k^{-1} S_{k,k}^T, \quad k \geqslant 1; \quad P_{0/0} = P_0.$$

The innovation covariance matrix, Π_k, satisfies

$$\begin{aligned} &\Pi_k = K_{k,k}^{\theta} \circ \left(H_k D_k H_k^T\right) + R_k + \overline{\Theta}_k H_k S_{k,k}, \quad k \leqslant m, \\ &\Pi_k = K_{k,k}^{\theta} \circ \left(H_k D_k H_k^T\right) + R_k + \overline{\Theta}_k H_k S_{k,k} + S_{k,k}^T H_k^T \overline{\Theta}_k - \overline{\Theta}_k H_k P_{k/k-1} H_k^T \overline{\Theta}_k \\ &\qquad - \Psi_{k,k-m}\left(\Pi_{k-m} + \sum_{i=1}^{m-1} T_{k-i,k-m}^T \Pi_{k-i}^{-1} T_{k-i,k-m}\right)\Psi_{k,k-m}^T, \quad k > m. \end{aligned} \tag{11}$$

Proof. See Appendix A. □

Theorem 2. *For the system model (1) and measurement model (3), under Assumptions 1–4, the fixed-point smoothers, $\widehat{x}_{k/k+N}$, $N \geqslant 1$ are recursively obtained by*

$$\widehat{x}_{k/k+N} = \widehat{x}_{k/k+N-1} + S_{k,k+N}\Pi_{k+N}^{-1}\nu_{k+N}, \quad N \geqslant 1, \tag{12}$$

whose initial condition is the filter, $\widehat{x}_{k/k}$, given in Theorem 1.

The matrices $S_{k,k+N}$ are determined by

$$\begin{aligned} &S_{k,k+N} = \left(D_k \mathbb{F}_{k+N,k}^T - M_{k,k+N-1} F_{k+N-1}^T\right) H_{k+N}^T \overline{\Theta}_{k+N}, \quad k \leqslant m-N, \\ &S_{k,k+N} = \left(D_k \mathbb{F}_{k+N,k}^T - M_{k,k+N-1} F_{k+N-1}^T\right) H_{k+N}^T \overline{\Theta}_{k+N} - \left(S_{k,k+N-m} - \sum_{i=1}^{m-1} S_{k,k+N-i} \Pi_{k+N-i}^{-1} T_{k+N-i,k+N-m}\right)\Psi_{k+N,k+N-m}^T, \quad k > m-N. \end{aligned} \tag{13}$$

where the matrices $M_{k,k+N}$ are recursively obtained from

$$\begin{aligned} &M_{k,k+N} = M_{k,k+N-1} F_{k+N-1}^T + S_{k,k+N}\Pi_{k+N}^{-1} S_{k+N,k+N}^T, \\ &M_{k,k} = D_k - P_{k/k}. \end{aligned} \tag{14}$$

The fixed-point smoothing error covariance matrix, $P_{k/k+N}$, satisfies

$$P_{k/k+N} = P_{k/k+N-1} - S_{k,k+N}\Pi_{k+N}^{-1}S_{k,k+N}^T, \quad N \geqslant 1, \tag{15}$$

with initial condition $P_{k/k}$, the filtering error covariance matrix.

The innovations v_{k+N}, their covariance matrices Π_{k+N}, the matrices $T_{k+N,k+N-i}$, $\Psi_{k+N,k+N-m}$, D_k and $P_{k/k}$ are given in Theorem 1.

Proof. See Appendix B. □

Remark 7. As indicated in Remark 3, the assumption that the Bernoulli sequences and the observation noises are independent from sensor to sensor is not required to obtain the centralized estimators. If this assumption is suppressed, one should take into account that, in Theorem 1, the covariance matrices $K_{k,k}^{\theta}$, $K_{k,k-m}^{\theta}$ and R_k would not be necessarily diagonal, and clearly $K_{k,k}^{\theta} \neq \overline{\Theta}_k(I - \overline{\Theta}_k)$.

4. Distributed fusion estimation

Our aim in this section is to find optimal distributed fusion estimators, in the linear minimum variance sense, based on the information provided by local LS linear estimators.

This estimation problem is tackled in two-stage fusion structure. In the first fusion stage, each sensor provides its local estimator based on its own measurement data along with their estimation error covariance matrices. In the second fusion stage, the cross-covariance matrix of the estimation errors between any two sensors from the first fusion stage are determined, and then, these covariances along with the estimates and error covariance matrices of all local subsystems are fused to determine the optimal matrix weights and the optimal fusion estimators in the linear minimum variance sense.

4.1. Local LS linear estimators

This section is concerned with the problem of obtaining, for each sensor subsystem of system (1) and (2), the local LS linear filter, $\widehat{x}_{k/k}^i$, and fixed-point smoothers, $\widehat{x}_{k/k+N}^i$, $N \geqslant 1$, along with their corresponding error covariance matrices from recursive algorithms. By using an innovation approach, these algorithms are established in the following theorems.

Theorem 3. *For the ith sensor subsystem of system (1) and (2) under Assumptions 1–4, the local LS linear filter, $\widehat{x}_{k/k}^i$, is calculated by*

$$\widehat{x}_{k/k}^i = \widehat{x}_{k/k-1}^i + S_{k,k}^i\left(\Pi_{k,k}^{ii}\right)^{-1}v_k^i, \quad k \geqslant 1, \quad \widehat{x}_{0/0}^i = \overline{x}_0, \tag{16}$$

where the local state predictor, $\widehat{x}_{k/k-1}^i$, satisfies

$$\widehat{x}_{k/k-1}^i = F_{k-1}\widehat{x}_{k-1/k-1}^i, \quad k \geqslant 1. \tag{17}$$

The innovation, v_k^i, is given by

$$\begin{aligned} v_k^i &= y_k^i - \overline{\theta}_k^i H_k^i \widehat{x}_{k/k-1}^i, \quad k \leqslant m, \\ v_k^i &= y_k^i - \overline{\theta}_k^i H_k^i \widehat{x}_{k/k-1}^i - \Psi_{k,k-m}^i\left(v_{k-m}^i - \sum_{l=1}^{m-1} T_{k-l,k-m}^{iT}\left(\Pi_{k-l,k-l}^{ii}\right)^{-1}v_{k-l}^i\right), \quad k > m, \end{aligned} \tag{18}$$

where $\Psi_{k,k-m}^i = K_{k,k-m}^{\theta^i}H_k^i\mathbb{F}_{k,k-m}D_{k-m}H_{k-m}^{iT}\left(\Pi_{k-m,k-m}^{ii}\right)^{-1}$, with D_{k-m} given in Theorem 1.

The matrices $T_{k,k-l}^i$ are determined by

$$\begin{aligned} T_{k,k-l}^i &= \overline{\theta}_k^i H_k^i \mathbb{F}_{k,k-l} S_{k-l,k-l}^i, \quad 2 \leqslant k \leqslant m, \quad 1 \leqslant l \leqslant k-1, \\ T_{k,k-l}^i &= \overline{\theta}_k^i H_k^i \mathbb{F}_{k,k-l} S_{k-l,k-l}^i - \Psi_{k,k-m}^i T_{k-l,k-m}^{iT}, \quad k > m, \quad 1 \leqslant l \leqslant m-1. \end{aligned}$$

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

$$\begin{aligned} \Pi_{k,k}^{ii} &= \overline{\theta}_k^i(1-\overline{\theta}_k^i)H_k^i D_k H_k^{iT} + R_k^i + \overline{\theta}_k^i H_k^i S_{k,k}^i, \quad k \leqslant m, \\ \Pi_{k,k}^{ii} &= \overline{\theta}_k^i(1-\overline{\theta}_k^i)H_k^i D_k H_k^{iT} + R_k^i + \overline{\theta}_k^i H_k^i S_{k,k}^i + \overline{\theta}_k^i S_{k,k}^{iT} H_k^{iT} \\ &\quad - (\overline{\theta}_k^i)^2 H_k^i P_{k/k-1}^{ii} H_k^{iT} - \Psi_{k,k-m}^i\left(\Pi_{k-m,k-m}^{ii} + \sum_{l=1}^{m-1} T_{k-l,k-m}^{iT}\left(\Pi_{k-l,k-l}^{ii}\right)^{-1}T_{k-l,k-m}^i\right)\Psi_{k,k-m}^{iT}, \quad k > m. \end{aligned}$$

The matrix $S^i_{k,k}$ is derived by the following expression

$$S^i_{k,k} = \overline{\theta}^i_k P^{ii}_{k/k-1} H^{iT}_k, \quad k \leqslant m,$$

$$S^i_{k,k} = \overline{\theta}^i_k P^{ii}_{k/k-1} H^{iT}_k - \left(\mathbb{F}_{k,k-m} S^i_{k-m,k-m} - \sum_{l=1}^{m-1} \mathbb{F}_{k,k-l} S^i_{k-l,k-l} \left(\Pi^{ii}_{k-l,k-l} \right)^{-1} T^i_{k-l,k-m} \right) \Psi^{iT}_{k,k-m}, \quad k > m,$$

where $P^{ii}_{k/k-1}$, the prediction error covariance matrix, is obtained by

$$P^{ii}_{k/k-1} = F_{k-1} P^{ii}_{k-1/k-1} F^T_{k-1} + Q_{k-1}, \quad k \geqslant 1,$$

with $P^{ii}_{k/k}$, the filtering error covariance matrix, satisfying

$$P^{ii}_{k/k} = P^{ii}_{k/k-1} - S^i_{k,k} \left(\Pi^{ii}_{k,k} \right)^{-1} S^{iT}_{k,k}, \quad k \geqslant 1, \quad P^{ii}_{0/0} = P_0.$$

Proof. This proof is analogous to that of Theorem 1 and hence it is omitted. □

Theorem 4. *For the ith sensor subsystem of system (1) and (2) under Assumptions 1–4, the local LS linear fixed-point smoothers $\widehat{x}^i_{k/k+N}$, $N \geqslant 1$, are recursively calculated by*

$$\widehat{x}^i_{k/k+N} = \widehat{x}^i_{k/k+N-1} + S^i_{k,k+N} \left(\Pi^{ii}_{k+N,k+N} \right)^{-1} v^i_{k+N}, \quad N \geqslant 1, \tag{19}$$

whose initial condition is the local filter, $\widehat{x}^i_{k/k}$, given in Theorem 3.

The matrices $S^i_{k,k+N}$ satisfy the following expressions

$$S^i_{k,k+N} = \overline{\theta}^i_{k+N} \left(D_k \mathbb{F}^T_{k+N,k} - M^i_{k,k+N-1} F^T_{k+N-1} \right) H^{iT}_{k+N}, \quad k \leqslant m-N,$$

$$S^i_{k,k+N} = \overline{\theta}^i_{k+N} \left(D_k \mathbb{F}^T_{k+N,k} - M^i_{k,k+N-1} F^T_{k+N-1} \right) H^{iT}_{k+N} - \left(S^i_{k,k+N-m} - \sum_{l=1}^{m-1} S^i_{k,k+N-l} \left(\Pi^{ii}_{k+N-l,k+N-l} \right)^{-1} T^i_{k+N-l,k+N-m} \right) \Psi^{iT}_{k+N,k+N-m}, \quad k > m-N,$$

where the matrices $M^i_{k,k+N}$ are recursively obtained by

$$M^i_{k,k+N} = M^i_{k,k+N-1} F^T_{k+N-1} + S^i_{k,k+N} \left(\Pi^{ii}_{k+N,k+N} \right)^{-1} S^{iT}_{k+N,k+N},$$

$$M^i_{k,k} = D_k - P^{ii}_{k/k}.$$

The fixed-point smoothing error covariance matrices, $P^{ii}_{k/k+N}$, are given by

$$P^{ii}_{k/k+N} = P^{ii}_{k/k+N-1} - S^i_{k,k+N} \left(\Pi^{ii}_{k+N,k+N} \right)^{-1} S^{iT}_{k,k+N}, \quad N \geqslant 1,$$

with initial condition the filtering error covariance matrix, $P^{ii}_{k/k}$.

The innovations v^i_{k+N}, their covariance matrices $\Pi^{ii}_{k+N,k+N}$, the matrices $T^i_{k+N,k+N-l}$, $\Psi^i_{k+N,k+N-m}$, D_k and $P^{ii}_{k/k}$ are given in Theorem 3.

Proof. This proof is analogous to that of Theorem 2 and hence it is omitted. □

4.2. Distributed fusion estimators

Once the local LS linear filtering and fixed-point smoothing estimators given in Theorems 3 and 4 are available, we can easily obtain the distributed optimal weighted fusion estimators and their error covariance matrices, by applying the optimal information fusion criterion weighted by matrices in the linear minimum variance sense [18].

Theorem 5. *For the system model (1) and measurement model (2), under Assumptions 1–4, the distributed optimal fusion filter, $\widehat{x}^0_{k/k}$, and fixed-point smoother, $\widehat{x}^0_{k/k+N}$, $N \geqslant 1$, are given by*

$$\widehat{x}^0_{k/k+N} = A^1_{k,k+N} \widehat{x}^1_{k/k+N} + \cdots + A^r_{k,k+N} \widehat{x}^r_{k/k+N}, \quad N \geqslant 0,$$

where $\widehat{x}^i_{k/k+N}$, $N \geqslant 0$ $(i = 1, 2 \ldots, r)$ are calculated by the recursive algorithms established in Theorems 3 and 4.

The optimal matrix weights $A^i_{k,k+N}$ $(i = 1, 2, \ldots, r)$ are computed by

$$A_{k,k+N} = \Sigma^{-1}_{k/k+N} e \left(e^T \Sigma^{-1}_{k/k+N} e \right)^{-1},$$

where the matrices $A_{k,k+N} = \left[A^1_{k,k+N}, \ldots, A^r_{k,k+N}\right]^T$ *and* $e = [I, \ldots, I]^T$ *are both* $nr \times n$ *matrices, and*

$$\Sigma_{k/k+N} = E\left[\left(\tilde{x}^1_{k/k+N}, \ldots, \tilde{x}^r_{k/k+N}\right)\left(\tilde{x}^1_{k/k+N}, \ldots, \tilde{x}^r_{k/k+N}\right)^T\right] = (P^{ij}_{k/k+N})$$

is a symmetric positive definite matrix of dimension $nr \times nr$.

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

$$P^0_{k/k+N} = \left(e^T \Sigma^{-1}_{k/k+N} e\right)^{-1}, \quad N \geqslant 0$$

and the following inequality holds: $P^0_{k/k+N} \leqslant P^{ii}_{k/k+N}$, $i = 1, 2, \ldots, r$.

Proof. The proof follows directly from the optimal information criterion weighted by matrices in the linear minimum variance sense [18] and therefore it is omitted. □

To apply the above Theorem 5, besides the local estimators, $\widehat{x}^i_{k/k+N}$, $N \geqslant 0$ $(i = 1, 2 \ldots, r)$, and their error covariance matrices, $P^{ii}_{k/k+N}$, given in Theorems 3 and 4, we need to calculate the cross-covariance matrices, $P^{ij}_{k/k+N}$, between any two subsystems. Next, computation procedures for the cross-covariance matrices, $P^{ij}_{k/k+N}$, $N \geqslant 0$, $i \neq j$, $i, j = 1, 2 \ldots, r$, will be presented, before which some useful lemmas will be given. The assumptions and notation in these lemmas are those of Theorems 3 and 4.

Lemma 1. *For* $i \neq j$, $i, j = 1, 2 \ldots, r$, $L^{ij}_{k,k} = E\left[\widehat{x}^i_{k/k-1} v^{jT}_k\right]$, *is calculated by*

$$\begin{aligned} L^{ij}_{k,k} &= \overline{\theta}^j_k \left(P^{ij}_{k/k-1} - P^{ij}_{k/k-1}\right) H^{jT}_k, \quad k \leqslant m, \\ L^{ij}_{k,k} &= \overline{\theta}^j_k \left(P^{ij}_{k/k-1} - P^{ij}_{k/k-1}\right) H^{jT}_k - \left(L^{ij}_{k,k-m} - \sum_{l=1}^{m-1} L^{ij}_{k,k-l} \left(\Pi^{jj}_{k-l,k-l}\right)^{-1} T^j_{k-l,k-m}\right) \Psi^{jT}_{k,k-m}, \quad k > m, \end{aligned} \tag{20}$$

where $L^{ij}_{k,s} = E\left[\widehat{x}^i_{k/k-1} v^{jT}_s\right]$, $s < k$, *is recursively obtained by*

$$L^{ij}_{k,s} = F_{k-1} L^{ij}_{k-1,s} + F_{k-1} S^i_{k-1,k-1} \left(\Pi^{ii}_{k-1,k-1}\right)^{-1} \Pi^{ij}_{k-1,s}, \quad s < k. \tag{21}$$

Proof. Taking into account expression (18) for the innovation v^j_k, to obtain (20) for $L^{ij}_{k,k}$ it is enough to prove that

$$E\left[\widehat{x}^i_{k/k-1} y^{jT}_k\right] - \overline{\theta}^j_k E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{jT}_{k/k-1}\right] H^{jT}_k = \overline{\theta}^j_k \left(P^{ij}_{k/k-1} - P^{ij}_{k/k-1}\right) H^{jT}_k. \tag{22}$$

Using (2) for y^j_k and the OPL, we have

$$E\left[\widehat{x}^i_{k/k-1} y^{jT}_k\right] = \overline{\theta}^j_k E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{iT}_{k/k-1}\right] H^{jT}_k.$$

Now, since

$$E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{jT}_{k/k-1}\right] = P^{ij}_{k/k-1} - D_k + E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{iT}_{k/k-1}\right] + E\left[\widehat{x}^j_{k/k-1} \widehat{x}^{jT}_{k/k-1}\right],$$

$$E\left[\widehat{x}^j_{k/k-1} \widehat{x}^{jT}_{k/k-1}\right] = D_k - P^{jj}_{k/k-1},$$

we have that $E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{iT}_{k/k-1}\right] - E\left[\widehat{x}^i_{k/k-1} \widehat{x}^{jT}_{k/k-1}\right] = P^{jj}_{k/k-1} - P^{ij}_{k/k-1}$ and (22) is easily derived.

Finally, from (17) for $\widehat{x}^i_{k/k-1}$ and (16) for $\widehat{x}^i_{k-1/k-1}$, expression (21) is immediately obtained. □

Lemma 2. *For* $i \neq j$, $i, j = 1, 2 \ldots, r$, *the innovation cross-covariance matrix* $\Pi^{ij}_{k,s} = E\left[v^i_k v^{jT}_s\right]$ *satisfies*

$$\begin{aligned} \Pi^{ij}_{k,s} &= \overline{\theta}^i_k H^i_k \left(\mathbb{F}_{k,s} S^j_{s,s} - L^{ij}_{k,s}\right), \quad k \leqslant m, \quad 1 \leqslant s \leqslant k, \\ \Pi^{ij}_{k,s} &= \overline{\theta}^i_k H^i_k \left(\mathbb{F}_{k,s} S^j_{s,s} - L^{ij}_{k,s}\right) - \Psi^i_{k,k-m} \left(\Pi^{ij}_{k-m,s} - \sum_{l=1}^{m-1} T^{iT}_{k-l,k-m} \left(\Pi^{ii}_{k-l,k-l}\right)^{-1} \Pi^{ij}_{k-l,s}\right), \quad k > m, \quad k - m \leqslant s \leqslant k. \end{aligned} \tag{23}$$

Proof. By using (18) for the innovation v^i_k, and taking into account that, from (2),

$$E\left[y^i_k v^{jT}_s\right] = \overline{\theta}^i_k H^i_k E\left[x_k v^{jT}_s\right] = \overline{\theta}^i_k H^i_k \mathbb{F}_{k,s} S^j_{s,s},$$

expression (23) is obtained. □

Lemma 3. *For $i \neq j,\ i,j = 1,2\ldots,r,\ J^{ij}_{k/k+N-1,k+N} = E\left[\widehat{x}^i_{k/k+N-1} v^{jT}_{k+N}\right]$ satisfies*

$$J^{ij}_{k/k+N-1,k+N} = \overline{\theta}^j_{k+N}\left(E^{ii}_{k,k+N-1} - E^{ij}_{k,k+N-1}\right)F^T_{k+N-1}H^{jT}_{k+N},\quad k \leqslant m-N,$$

$$J^{ij}_{k/k+N-1,k+N} = \overline{\theta}^j_{k+N}\left(E^{ii}_{k,k+N-1} - E^{ij}_{k,k+N-1}\right)F^T_{k+N-1}H^{jT}_{k+N} - \left(J^{ij}_{k/k+N-1,k+N-m} - \sum_{l=1}^{m-1} J^{ij}_{k/k+N-1,k+N-l}\left(\Pi^{jj}_{k+N-l,k+N-l}\right)^{-1} T^{j}_{k+N-l,k+N-m}\right)\Psi^{jT}_{k+N,k+N-m},\quad k > m-N, \tag{24}$$

where, for $i,j = 1,2\ldots,r,\ E^{ij}_{k,k+N} = E\left[\widehat{x}^i_{k/k+N}\widehat{x}^{jT}_{k+N/k+N}\right]$ is recursively computed by

$$E^{ij}_{k,k+N} = E^{ij}_{k,k+N-1}F^T_{k+N-1} + J^{ij}_{k/k+N-1,k+N}\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k+N,k+N} + S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1}\left(L^{ji}_{k+N,k+N}\right)^T + S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1}\Pi^{ij}_{k+N,k+N}\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k+N,k+N},\quad N \geqslant 1, \tag{25}$$

with initial condition $E^{ij}_{k,k} = D_k + P^{ij}_{k/k} - P^{ii}_{k/k} - P^{ij}_{k/k}$.

For $l = 1,2\ldots,m,\ J^{ij}_{k/k+N-1,k+N-l} = E\left[\widehat{x}^i_{k/k+N-1} v^{jT}_{k+N-l}\right]$ satisfies

$$J^{ij}_{k/k+N-1,k+N-l} = J^{ij}_{k/k+N-2,k+N-l} + S^i_{k,k+N-1}\left(\Pi^{ii}_{k+N-1,k+N-1}\right)^{-1}\Pi^{ij}_{k+N-1,k+N-l}. \tag{26}$$

Proof. From (18) for the innovation v^j_{k+N}, in order to obtain (24) we just need to prove that

$$E\left[\widehat{x}^i_{k+N/k+N-1} y^{jT}_{k+N}\right] - \overline{\theta}^j_{k+N} E\left[\widehat{x}^i_{k/k+N-1}\widehat{x}^{jT}_{k+N/k+N-1}\right]H^{jT}_{k+N} = \overline{\theta}^j_{k+N}\left(E^{ii}_{k,k+N-1} - E^{ij}_{k,k+N-1}\right)F^T_{k+N-1}H^{jT}_{k+N}. \tag{27}$$

Using (2) for y^j_{k+N} and the OPL, we have

$$E\left[\widehat{x}^i_{k/k+N-1} y^{jT}_{k+N}\right] = \overline{\theta}^j_{k+N} E\left[\widehat{x}^i_{k/k+N-1}\widehat{x}^{jT}_{k+N/k+N-1}\right]H^{jT}_{k+N}$$

and since, from (17), $\widehat{x}^j_{k+N/k+N-1} = F_{k+N-1}\widehat{x}^j_{k+N-1/k+N-1}$, expression (27) is easily obtained.

On the other hand, by using (19) for $\widehat{x}^i_{k/k+N}$ and (16) for $\widehat{x}^j_{k+N/k+N}$, recursive expression (25) for $E^{ij}_{k,k+N}$ is immediately derived; its initial condition $E^{ij}_{k,k}$ is also easily obtained.

Finally, by using again (19) for $\widehat{x}^i_{k/k+N-1}$, recursive expression (26) is also immediately clear and the proof is completed. □

Remark 8. For $i = j$, since the innovation is a white process, it is clear that $L^{ii}_{k,k} = E\left[\widehat{x}^i_{k/k-1} v^{iT}_k\right] = 0$ and $J^{ii}_{k/k+N-1,k+N} = E\left[\widehat{x}^i_{k/k+N-1} v^{iT}_{k+N}\right] = 0$.

A set of recursive formulas to calculate the filtering and fixed-point smoothing error cross-covariance matrices $P^{ij}_{k/k+N},\ i \neq j,\ i,j = 1,2\ldots,r,\ N \geqslant 0$, is now derived in the following theorem based on Lemmas 1–4.

Theorem 6. *The cross-covariance matrices, $P^{ij}_{k/k+N},\ N \geqslant 1$, of the fixed-point smoothing errors between the ith and the jth sensor subsystems are recursively computed by*

$$P^{ij}_{k/k+N} = P^{ij}_{k/k+N-1} + S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1}\Pi^{ij}_{k+N,k+N}\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k,k+N} - \left(S^j_{k,k+N} - J^{ij}_{k/k+N-1,k+N}\right)\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k,k+N} - S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1}\left(S^i_{k,k+N} - J^{ji}_{k/k+N-1,k+N}\right)^T,\quad N \geqslant 1. \tag{28}$$

The initial condition, $P^{ij}_{k/k}$, the cross-covariance matrix of the filtering error between the ith and the jth sensor subsystems, satisfies

$$P^{ij}_{k/k} = P^{ij}_{k/k-1} + S^i_{k,k}\left(\Pi^{ii}_{k,k}\right)^{-1}\Pi^{ij}_{k,k}\left(\Pi^{jj}_{k,k}\right)^{-1} S^{jT}_{k,k} - \left(S^j_{k,k} - L^{ij}_{k,k}\right)\left(\Pi^{jj}_{k,k}\right)^{-1} S^{jT}_{k,k} - S^i_{k,k}\left(\Pi^{ii}_{k,k}\right)^{-1}\left(S^i_{k,k} - L^{ji}_{k,k}\right)^T,\quad k \geqslant 1,$$

$$P^{ij}_{k/k-1} = F_{k-1}P^{ij}_{k-1/k-1}F^T_{k-1} + Q_{k-1},\quad k \geqslant 1;\quad P^{ij}_{0/0} = P_0. \tag{29}$$

Proof. By using (19) for $\widehat{x}^i_{k/k+N}$ and $\widehat{x}^j_{k/k+N}$, we have

$$P^{ij}_{k/k+N} = P^{ij}_{k/k+N-1} - E\left[\left(x_k - \widehat{x}^i_{k/k+N-1}\right)v^{jT}_{k+N}\right]\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k,k+N} - S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1} E\left[v^i_{k+N}\left(x_k - \widehat{x}^j_{k+N/k+N-1}\right)^T\right] + S^i_{k,k+N}\left(\Pi^{ii}_{k+N,k+N}\right)^{-1}\Pi^{ij}_{k+N,k+N}\left(\Pi^{jj}_{k+N,k+N}\right)^{-1} S^{jT}_{k,k+N}.$$

Taking into account that $E\left[x_k v_{k+N}^{jT}\right] = S_{k,k+N}^{j}$, $E\left[\widehat{x}_{k/k+N-1}^{i} v_{k+N}^{jT}\right] = J_{k/k+N-1,k+N}^{ij}$, $E\left[v_{k+N}^{i} x_k^T\right] = S_{k,k+N}^{iT}$ and $E\left[v_{k+N}^{i} \widehat{x}_{k/k+N-1}^{jT}\right] = J_{k/k+N-1,k+N}^{jiT}$, recursive expression (28), for the cross-covariance matrices of the local fixed-point smoothing errors, is obtained.

Finally, by using (16), and following an analogous reasoning, it is easy to get (29) for the cross-covariance matrices of the local filtering errors. □

5. Numerical simulation examples

In this section, two numerical examples are presented to show the effectiveness of the proposed estimation algorithms. To test and compare the performance of the proposed estimators, we ran a program in MATLAB, simulating at each iteration the state and the measured values and providing the centralized and distributed fusion filter and fixed-point smoothers, as well as the corresponding error covariance matrices.

5.1. Example 1

In this example, for the simulation, we consider that the system state is given by a scalar process, $\{x_k;\ k \geqslant 0\}$, generated by the following first-order autoregressive model,

$$x_k = 0.95x_{k-1} + w_{k-1}, \quad k \geqslant 1,$$

where the initial state is a zero-mean Gaussian variable with $Var[x_0] = 1$ and $\{w_k;\ k \geqslant 0\}$ is a zero-mean white Gaussian noise with $Var[w_k] = 0.1$, for all k.

Consider missing measurements coming from two sensors and perturbed by independent sequences of Bernoulli random variables $\{\theta_k^i;\ k \geqslant 1\},\ i = 1, 2$, and by independent additive white noises, $\{v_k^i;\ k \geqslant 1\},\ i = 1, 2$, with zero-mean and variances $Var[v_k^1] = 1$ and $Var[v_k^2] = 1.5$, for all k.

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

According to our theoretical model, it is assumed that, for each sensor, the uncertainty at any sampling time $k \geqslant 1$ depends only on the uncertainty at the previous time $k - m$. The variables θ_k^i modeling this type of uncertainty correlation in the output measurements are defined based on two independent sequences of independent Bernoulli random variables, $\{\gamma_k^i;\ k \geqslant 1\},\ i = 1, 2$, with constant probabilities $P[\gamma_k^i = 1] = \gamma_i$. Specifically, the variables θ_k^i are defined as follows

$$\theta_k^i = 1 - \gamma_{k+m}^i(1 - \gamma_k^i), \quad i = 1, 2.$$

Thus, if $\theta_k^i = 0$, then $\gamma_{k+m}^i = 1$ and $\gamma_k^i = 0$, and hence, $\theta_{k+m}^i = 1$; this fact guarantees that, if the state is missing at time k, the output measurement at time $k + m$ necessarily contains the state. Therefore, there cannot be more than m consecutive measured data consisting of noise only.

Since the variables γ_k^i and γ_s^i are independent, θ_k^i and θ_s^i are also independent for $|k - s| \neq 0, m$. The mean of these variables is $\overline{\theta}^i = 1 - \gamma_i(1 - \gamma_i)$ and its covariance function is given by

$$K_{k,s}^{\theta} = E[(\theta_k^i - \overline{\theta}^i)(\theta_s^i - \overline{\theta}^i)] = \begin{cases} 0, & |k - s| \neq 0, m, \\ -(1 - \overline{\theta}^i)^2, & |k - s| = m, \\ \overline{\theta}^i(1 - \overline{\theta}^i), & |k - s| = 0. \end{cases}$$

To compare the effectiveness of the proposed estimators, fifty iterations of the proposed algorithms have been performed and the results obtained for different values of the uncertainty probability and several values of m have been analyzed.

Let us observe that the means, $\overline{\theta}^i$, for $i = 1, 2$, of the variables θ_k^i, are the same if $1 - \gamma_i$ is used instead of γ_i; for this reason, only the case $\gamma_i \leqslant 0.5$ will be considered here.

Assuming that the Bernoulli variables θ_k^i, for $i = 1, 2$, of the measurement outputs are correlated at sampling times that differ three units of time ($m = 3$), the error variances of local, centralized and distributed fusion filters will be compared considering fixed values of the probabilities γ_1 and γ_2; specifically, $\gamma_1 = 0.1,\ \gamma_2 = 0.2$. In Fig. 1, as mentioned in Theorem 5, we can see that the error variances of each local filter are higher than that of the distributed fusion filter. Although the distributed fusion filter has lower accuracy than the centralized one, this difference is slight. Besides, this is compensated by the fact that the distributed fusion structure is in general more robust, reduces the computational cost and improves the reliability due to its parallel structure.

Fig. 2 displays the filtering and fixed-point smoothing error variances ($N = 2, 5$) for the centralized and distributed fusion methods. It can be seen that the error variances corresponding to the fixed-point smoothers are less than those of the filters and, consequently, the fixed-point smoothing estimates are more accurate. It is also verified that centralized and distributed fusion filter and smoothers have a similar accuracy. If we compare the smoothing error variances at each fixed-point k for $N = 2$ and $N = 5$, we observe that these estimators become more accurate as the number of available observations increases.

Finally, in order to show more precisely the dependence of the error variances on the values γ_1 and γ_2, Fig. 3 displays the filtering error variances, at a fixed iteration (namely, $k = 50$) for $m = 3$, when both γ_1 and γ_2 are varied from 0.1 to 0.5, which

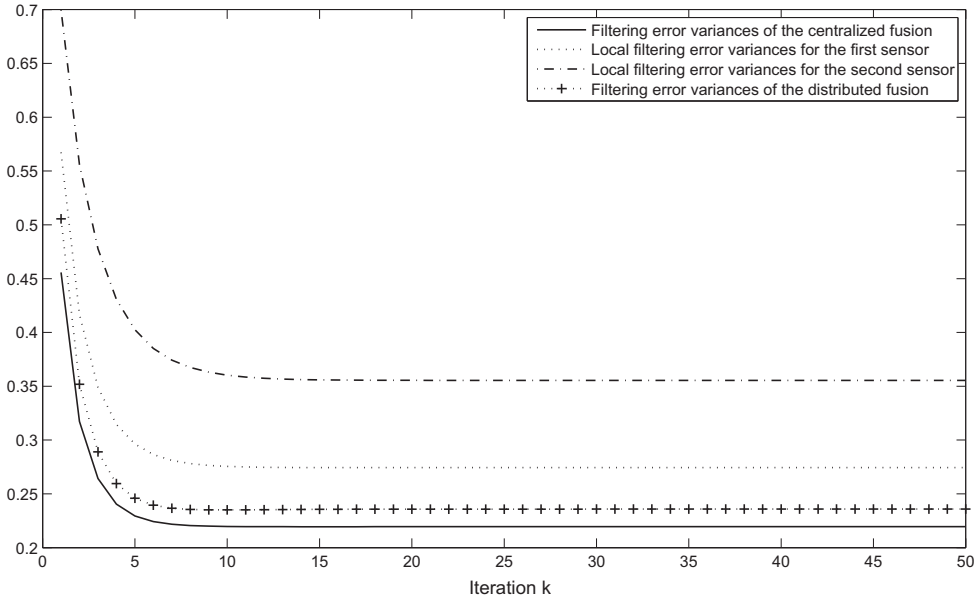

Fig. 1. Filtering error variances for the centralized and distributed fusion methods for $\gamma_1 = 0.1,\ \gamma_2 = 0.2$, when $m = 3$.

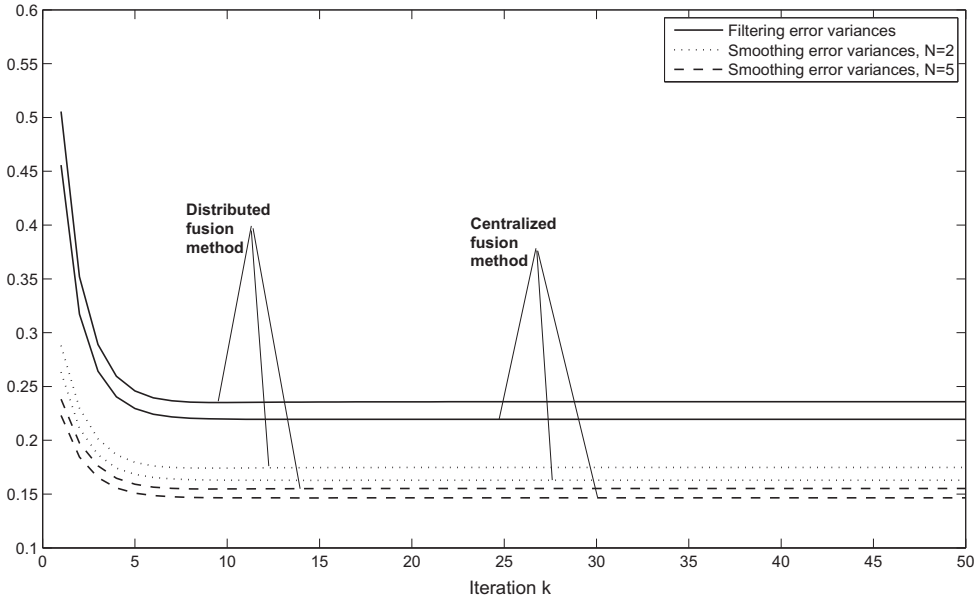

Fig. 2. Filtering and smoothing error variances for the centralized and distributed fusion methods for $\gamma_1 = 0.1,\ \gamma_2 = 0.2$, when $m = 3$.

provide different values of the probabilities $\overline{\theta}^1$ and $\overline{\theta}^2$. More specifically, we have considered the values $\gamma_i = 0.1, 0.2, 0.3, 0.4, 0.5$, which lead to the probabilities $\overline{\theta}^i = 0.91, 0.84, 0.78, 0.76, 0.75$, respectively.

In this figure, both graphs (corresponding to the centralized and distributed fusion filters, respectively) show that the performance of the filter diminishes as $\overline{\theta}^i$ becomes lower, due to the fact that the probability of observations containing the state decreases. Also, this figure confirms that both methods, centralized and distributed, have approximately the same accuracy, corroborating the previous results.

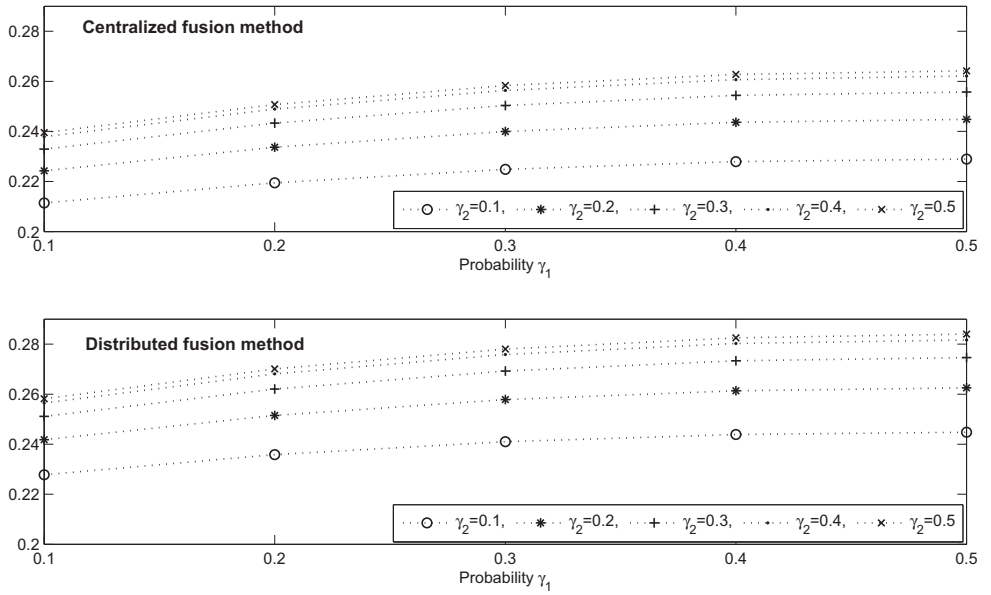

Fig. 3. Filtering error variances for the centralized and distributed fusion methods at $k = 50$ versus γ_1, with γ_2 varying from 0.1 to 0.5 when $m = 3$.

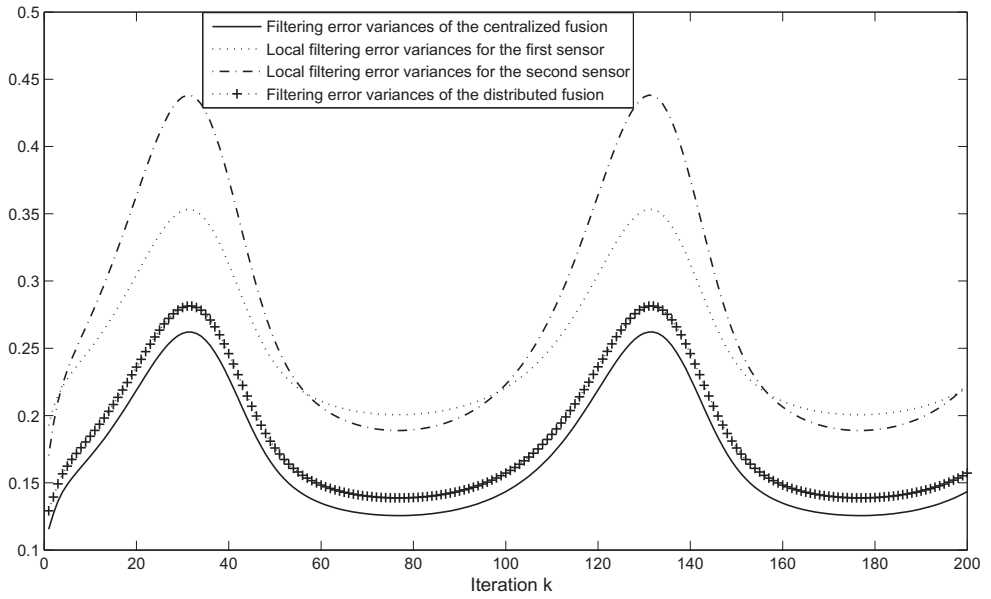

Fig. 4. Filtering error variances for the centralized and distributed fusion methods for the first state component for $\gamma_1 = 0.1,\ \gamma_2 = 0.2$, when $m = 3$.

5.2. Example 2

In this example, the following discrete-time system with missing measurements has been considered

$$x_k = \left(1 + 0.2\sin\left(\frac{(k-1)\pi}{50}\right)\right)\begin{pmatrix} 0.8 & 0 \\ 0.9 & 0.2 \end{pmatrix}x_{k-1} + w_{k-1}, \quad k \geqslant 1$$

$$y_k^i = \theta_k^i(1 \quad 1)x_k + v_k^i, \quad k \geqslant 1, \quad i = 1, 2$$

where the initial state, x_0, is a zero-mean Gaussian vector with covariance matrix given by $Cov[x_0] = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$, the processes $\{w_k;\ k \geqslant 0\}$ and $\{v_k^i;\ k \geqslant 1\},\ i = 1, 2$ are zero-mean white Gaussian noises with

Table 1
Filtering error variances for the centralized and distributed fusion methods for $\gamma_1 = 0.2,\ \gamma_2 = 0.4$ at $k = 30$ when $m = 2, 3, 4, 5$.

Component	Filtering error variances	$m = 2$	$m = 3$	$m = 4$	$m = 5$
First	Centralized	0.3310	0.3483	0.3628	0.3744
	Distributed	0.3561	0.3717	0.3827	0.3906
Second	Centralized	0.4014	0.4451	0.4835	0.5151
	Distributed	0.4340	0.4722	0.5007	0.5218

$Cov[w_k] = \begin{pmatrix} 0.36 & 0.3 \\ 0.3 & 0.25 \end{pmatrix}$, $Var[v_k^1] = 0.5$ and $Var[v_k^2] = 0.9,\ \forall k$, and the multiplicative noises $\{\theta_k^i;\ k \geqslant 1\},\ i = 1, 2$ are defined as in Example 1.

Firstly, our aim is to check that the accuracy of the optimal distributed fusion filter is higher than that of any local filter, but lower than that of the centralized fusion filter. For this, two hundred iterations of the proposed algorithms have been carried out and the results corresponding to the first state component for $m = 3$ and probabilities $\gamma_1 = 0.1$ and $\gamma_2 = 0.2$ are shown graphically in Fig. 4. As in Fig. 1, the error variances of each local filter are higher than that of the distributed fusion filter and the centralized and distributed filters have a similar accuracy. Analogous results for the second state component are obtained.

Also, analogous comments and conclusions to those made from Figs. 2 and 3 in Example 1 are deduced for the first and second components of the state in this example. For this reason, the corresponding figures have not been included.

Finally, for $\gamma_1 = 0.2,\ \gamma_2 = 0.4$ the performance of the estimators is compared for different values of m at a fixed iteration; specifically, for $m = 2, 3, 4, 5$ at $k = 30$, the filtering error variances of both state components are shown in Table 1. From this table it is gathered that the estimators are more accurate as the values of m are lower. In other words, a greater distance between the instants at which the variables are correlated (which means that the state can be missing in more consecutive observations) yields worse estimators. As expected, this table also shows that the estimators obtained by the centralized and distributed fusion methods have a very rough precision. It must be noticed that an analogous comparison has been performed in Example 1 and the results obtained are completely similar, so they have been omitted.

6. Conclusions

For multi-sensor linear discrete-time systems with missing measurements, the LS linear estimation problem has been addressed. The main contributions of the current paper can be summarized as follows:

1. Using both centralized and distributed fusion methods to process the measurement data from the different sensors, recursive filtering and fixed-point smoothing algorithms are derived by an innovation approach.
2. At each sensor, the possibility of missing measurements or uncertain observations (that is, observations containing no information about the state but only noise) is modeled by binary variables taking the values one or zero (Bernoulli variables), depending on whether the state is present or missing in the corresponding observation. Such variables are assumed to be correlated at instants that differ m units of time.
3. The basic model in which the Bernoulli variables describing the uncertainty in the observations at each sensor are independent is a particular case of the proposed model, just making $K_{k,s}^{\theta^i} = 0$ for $|k - s| = m$. Also, the model with correlation in consecutive sampling times is covered by the current study when $m = 1$. However, theses two assumptions can be unrealistic in many practical situations, and the estimation algorithms must be modified to incorporate the effect of different types of correlation. Specifically, the form of correlation considered in this paper is appropriate, in particular, to model situations where the state cannot be missing in $m + 1$ consecutive observations, as occurs, for instance, in sensor networks where sensor failures may happen and a failed sensor is substituted not immediately, but after m sampling times.
4. The multi-sensor system model considered in the current paper covers those situations where the additive observation noises and the Bernoulli variables involved are independent from sensor to sensor. This independence assumption simplifies the mathematical expression considerably and it is valid in a wide spectrum of applications, for example in wireless sensor networks which are characterized by sensor independence, limited storage capacity, lack of physical infrastructure and limited energy. Nevertheless, if such assumption is omitted, a similar technique to that used in this paper would allow us to extend the current study to this more general case with no difficulty, except for a greater complexity in the mathematical expressions.
5. Two numerical simulation examples illustrate the applicability of the current results to estimate a scalar state process generated by an AR model and a two-dimensional state, respectively, from uncertain observations coming from two sensors featuring correlation in the uncertainty. The results confirm that centralized and distributed fusion estimators have approximately the same accuracy. For different uncertainty probabilities and different values of m, both examples confirm the greater effectiveness of the fixed-point smoothing estimators in contrast to the filtering ones and conclude that more accurate estimations are obtained as the values of m are lower.

Acknowledgements

This research is supported by Ministerio de Ciencia e Innovación (Programa FPU and Grant No. MTM2011-24718).

Appendix A. Proof of Theorem 1

From the OPL, the LS linear estimators, $\widehat{x}_{k/L}$, expressed as combination linear of the innovations, are given by

$$\widehat{x}_{k/L} = \sum_{i=1}^{L} S_{k,i}\Pi_i^{-1}\nu_i, \tag{A.1}$$

where $\nu_i = y_i - \widehat{y}_{i/i-1}$ are the innovation vectors, with $\widehat{y}_{i/i-1}$ the one-stage observation predictor, $\Pi_i = E[\nu_i\nu_i^T]$, and $S_{k,i} = E[x_k\nu_i^T]$.

Using (A.1) with $L = k,\ k-1$, expression (5) for the filter is immediately derived. From (1) and OPL, expression (6) for the state predictor is easily obtained.

Now we show expression (7) for the innovation, $\nu_k = y_k - \widehat{y}_{k/k-1}$, for which it is enough to obtain an expression for $\widehat{y}_{k/k-1}$. From the OPL, it follows that $\widehat{y}_{k/k-1}$ is given by

$$\widehat{y}_{k/k-1} = \sum_{i=1}^{k-1} T_{k,i}\Pi_i^{-1}\nu_i, \quad k \geqslant 2, \quad T_{k,i} = E[y_k\nu_i^T].$$

Hence, we start by calculating $T_{k,i}$, for $i \leqslant k-1$. From the observation Eq. (3) and the model assumptions, it is clear that $T_{k,i} = E[\Theta_k H_k x_k \nu_i^T]$, for $i \leqslant k-1$, and $T_{k,i} = \overline{\Theta}_k H_k S_{k,i}$, for $k \leqslant m$ or $k > m$ and $i < k-m$. So, after some manipulations, we obtain:

(a) For $k \leqslant m$, using (A.1) for $L = k-1$, we have $\widehat{y}_{k/k-1} = \overline{\Theta}_k H_k \widehat{x}_{k/k-1}$.
(b) For $k > m$, the following equality is easily deduced

$$\widehat{y}_{k/k-1} = \overline{\Theta}_k H_k \sum_{i=1}^{k-1} S_{k,i}\Pi_i^{-1}\nu_i + \sum_{i=1}^{m}(T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i})\Pi_{k-i}^{-1}\nu_{k-i}, \tag{A.2}$$

where

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k \nu_{k-i}^T\right], \quad 1 \leqslant i \leqslant m \tag{A.3}$$

or equivalently,

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k y_{k-i}^T\right] - E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k \widehat{y}_{k-i/k-(i+1)}^T\right].$$

Using again (3) for y_{k-i}, property (4), and since from (1), $E[x_k x_{k-i}^T] = \mathbb{F}_{k,k-i}D_{k-i}$, it is concluded that

$$E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k y_{k-i}^T\right] = K_{k,k-i}^{\theta} \circ \left(H_k \mathbb{F}_{k,k-i} D_{k-i} H_{k-i}^T\right),$$

where $D_k = E[x_k x_k^T]$ can be clearly obtained by the recursive formula (8).

Summarizing, we have that

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = K_{k,k-i}^{\theta} \circ \left(H_k \mathbb{F}_{k,k-i} D_{k-i} H_{k-i}^T\right) - E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k \widehat{y}_{k-i/k-(i+1)}^T\right], \quad 1 \leqslant i \leqslant m. \tag{A.4}$$

On the one hand, for $i = m$, since Θ_k is independent of the innovations ν_i, for $i < k-m$, we have that $E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k \widehat{y}_{k-m/k-(m+1)}^T\right] = 0$, and from (A.4)

$$T_{k,k-m} - \overline{\Theta}_k H_k S_{k,k-m} = K_{k,k-m}^{\theta} \circ \left(H_k \mathbb{F}_{k,k-m} D_{k-m} H_{k-m}^T\right). \tag{A.5}$$

On the other hand, for $i < m,\ K_{k,k-i}^{\theta} = 0$ and, hence, from (A.4)

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = -E\left[(\Theta_k - \overline{\Theta}_k)H_k x_k \widehat{y}_{k-i/k-(i+1)}^T\right].$$

Now, using again that Θ_k is independent of ν_i, for $i \neq k-m$, it is deduced that

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = -E[\left(\Theta_k - \overline{\Theta}_k\right)H_k x_k \nu_{k-m}^T]\Pi_{k-m}^{-1} T_{k-i,k-m}^T$$

or, equivalently, from (A.3) for $i = m$, (A.5) and denoting

$$\Psi_{k,k-m} = K_{k,k-m}^{\theta} \circ \left(H_k \mathbb{F}_{k,k-m} D_{k-m} H_{k-m}^T\right)\Pi_{k-m}^{-1},$$

we have that

$$T_{k,k-i} - \overline{\Theta}_k H_k S_{k,k-i} = -\Psi_{k,k-m} T^T_{k-i,k-m}, \quad i < m. \tag{A.6}$$

Next, substituting this expression into (A.2) and using (A.1) for $\widehat{x}_{k/k-1}$, expression (7) is deduced. Moreover, using (3) and (A.6) and taking into account that, from (1), $S_{k,k-i} = \mathbb{F}_{k,k-i} S_{k-i,k-i}$, the matrices (9) are obtained.

Now, expression (10) for the matrix $S_{k,k} = E[x_k y_k^T] - E[x_k \widehat{y}^T_{k/k-1}]$ is derived. From (3) and the independence assumption, it is clear that $E[x_k y_k^T] = D_k H_k^T \overline{\Theta}_k, \ \forall k \geqslant 1$. To calculate $E[x_k \widehat{y}^T_{k/k-1}]$, the correlation assumption of the random variables θ_k must be taken into account and hence two cases must be considered:

- For $k \leqslant m$, from (7) we obtain $E[x_k \widehat{y}^T_{k/k-1}] = E[x_k \widehat{x}^T_{k/k-1}] H_k^T \overline{\Theta}_k$. From the OPL, $E[x_k \widehat{x}^T_{k/k-1}] = D_k - P_{k/k-1}$ where $P_{k/k-1}$ is the prediction error covariance matrix, and hence

$$E[x_k \widehat{y}^T_{k/k-1}] = (D_k - P_{k/k-1}) H_k^T \overline{\Theta}_k, \quad k \leqslant m.$$

- For $k > m$, from (7) it follows that

$$E[x_k \widehat{y}^T_{k/k-1}] = E[x_k \widehat{x}^T_{k/k-1}] H_k^T \overline{\Theta}_k + E[x_k v^T_{k-m}] \Psi^T_{k,k-m} - E\left[x_k \left(\sum_{i=1}^{m-1} T^T_{k-i,k-m} \Pi^{-1}_{k-i} \nu_{k-i} \right)^T \right] \Psi^T_{k,k-m},$$

hence, using again the OPL and taking into account that $E[x_k v^T_{k-i}] = S_{k,k-i}$, for $1 \leqslant i \leqslant m$, it is deduced that

$$E[x_k \widehat{y}^T_{k/k-1}] = (D_k - P_{k/k-1}) H_k^T \overline{\Theta}_k + S_{k,k-m} \Psi^T_{k,k-m} - \sum_{i=1}^{m-1} S_{k,k-i} \Pi^{-1}_{k-i} T_{k-i,k-m} \Psi^T_{k,k-m}, \quad k > m.$$

From the above expectations, expression (10) for $S_{k,k}$ is clear.

From (1), the expression for the prediction error covariance matrix, $P_{k/k-1}$ is immediately clear and, from (5), the expression for the filtering error covariance matrix, $P_{k/k}$, is also obvious.

Finally, we prove expression (11) for the innovation covariance matrix $\Pi_k = E[y_k y_k^T] - E[\widehat{y}_{k/k-1} \widehat{y}^T_{k/k-1}]$. From (3) and using (4), we have that

$$E[y_k y_k^T] = E[\theta_k \theta_k^T] \circ \left(H_k D_k H_k^T \right) + R_k, \quad k \geqslant 1.$$

Due to the correlation hypothesis of the Bernoulli variables θ_k, we need to distinguish two cases to calculate $E[\widehat{y}_{k/k-1} \widehat{y}^T_{k/k-1}]$. For $k \leqslant m$, from (7), (4) and the OPL, we have

$$E[\widehat{y}_{k/k-1} \widehat{y}^T_{k/k-1}] = (\overline{\theta}_k \overline{\theta}_k^T) \circ \left(H_k (D_k - P_{k/k-1}) H_k^T \right).$$

For $k > m$, using an analogous reasoning, applying the OPL and after some manipulations, we deduce that

$$\begin{aligned} E[\widehat{y}_{k/k-1} \widehat{y}^T_{k/k-1}] = {} & (\overline{\theta}_k \overline{\theta}_k^T) \circ \left(H_k (D_k - P_{k/k-1}) H_k^T \right) + \Psi_{k,k-m} \Pi_{k-m} \Psi^T_{k,k-m} + \Psi_{k,k-m} \sum_{i=1}^{m-1} T^T_{k-i,k-m} \Pi^{-1}_{k-i} T_{k-i,k-m} \Psi^T_{k,k-m} \\ & - \overline{\Theta}_k H_k \left(S_{k,k} - P_{k/k-1} H_k^T \overline{\Theta}_k \right) - \left(S^T_{k,k} - \overline{\Theta}_k H_k P_{k/k-1} \right) H_k^T \overline{\Theta}_k. \end{aligned}$$

So, from the above expectations, expression (11) for the innovation covariance matrix Π_k is obtained. □

Appendix B. Proof of Theorem 2

From the general expression (A.1), for each fixed $k \geqslant 1$, the recursive relation (12) is immediately clear.

Next, to prove (13) for $S_{k,k+N} = E[x_k y^T_{k+N}] - E[x_k \widehat{y}^T_{k+N/k+N-1}]$, it is necessary to calculate both expectations.

On the one hand, from Eq. (3), taking into account that $E[x_k x^T_{k+N}] = D_k \mathbb{F}^T_{k+N,k}$ and using that Θ_{k+N} and ν_{k+N} are independent of x_k, we obtain

$$E[x_k y^T_{k+N}] = D_k \mathbb{F}^T_{k+N,k} H^T_{k+N} \overline{\Theta}_{k+N}, \quad N \geqslant 1.$$

On the other hand, based on expression (7) for ν_{k+N}, which is different depending on wether $k+N \leqslant m$ or $k+N > m$, two options must be considered:

- From (7) for $k \leqslant m-N$, using (6) for $\widehat{x}_{k+N/k+N-1}$, we have that

$$E\left[x_k \widehat{y}^T_{k+N/k+N-1} \right] = M_{k,k+N-1} F^T_{k+N-1} H^T_{k+N} \overline{\Theta}_{k+N},$$

where $M_{k,k+N-1} = E\left[x_k \widehat{x}^T_{k+N-1/k+N-1} \right]$.

– A similar reasoning to the above one, but starting from (7) for $k > m - N$, yields

$$E[x_k \widehat{y}^T_{k+N/k+N-1}] = M_{k,k+N-1} F^T_{k+N-1} H^T_{k+N} \overline{\Theta}_{k+N} + \left(S_{k,k+N-m} - \sum_{i=1}^{m-1} S_{k,k+N-i} \Pi^{-1}_{k+N-i} T_{k+N-i,k+N-m} \right) \Psi^T_{k+N,k+N-m}.$$

Then, the replacement of the above expectations in $S_{k,k+N}$ leads to expression (13).

The recursive relation (14) for $M_{k,k+N} = E\left[x_k \widehat{x}^T_{k+N/k+N}\right]$ is immediately clear from (5) for $\widehat{x}_{k+N/k+N}$ and its initial condition $M_{k,k} = E[x_k \widehat{x}_{k/k}]$ is calculated taking into account that, from the orthogonality, $E[x_k \widehat{x}^T_{k/k}] = E[\widehat{x}_{k/k} \widehat{x}^T_{k/k}] = D_k - P_{k/k}$.

Finally, since $P_{k/k+N} = E\left[x_k x^T_k\right] - E\left[\widehat{x}_{k/k+N} \widehat{x}^T_{k/k+N}\right]$, using (12) and taking into account that $\widehat{x}_{k/k+N-1}$ is uncorrelated with v_{k+N}, expression (15) is deduced. □

References

[1] J. Hespanha, P. Naghshtabrizi, Yonggang Xu, A survey of recent results in networked control systems, Proc. IEEE 95 (1) (2007) 138–162.
[2] L. Ma, Z. Wang, J. Hu, Y. Bo, Z. Guo, Robust variance-constrained filtering for a class of nonlinear stochastic systems with missing measurements, Signal Process. 90 (2010) 2060–2071.
[3] A. Hermoso-Carazo, J. Linares-Pérez, Nonlinear estimation applying an unscented transformation in systems with correlated uncertain observations, Appl. Math. Comput. 217 (2011) 7998–8009.
[4] S. Zhou, G. Feng, H_∞ filtering for discrete-time systems with randomly varying sensor delays, Automatica 44 (2008) 1918–1922.
[5] S. Nakamori, A. Hermoso-Carazo, J. Linares-Pérez, Design of RLS Wiener estimators from randomly delayed observations in linear discrete-time stochastic systems, Appl. Math. Comput. 217 (2010) 3801–3815.
[6] S.L. Sun, Optimal estimators for systems with finite consecutive packet dropouts, IEEE Signal Process. Lett. 16 (7) (2009) 557–560.
[7] R. Caballero-Águila, A. Hermoso-Carazo, J. Linares-Pérez, Recursive least-squares quadratic smoothing from measurements with packet dropouts, Signal Process. 92 (2012) 931–938.
[8] S. Sun, W. Xiao, Optimal linear estimators for systems with multiple random measurement delays and packet dropouts, Int. J. Syst. Sci. (2011) 1–13.
[9] M. Moayedi, Y.K. Foo, Y.C. Soh, Adaptive Kalman filtering in networked systems with random sensor delays, multiple packet dropouts and missing measurements, IEEE Trans. Signal Process. 58 (3) (2010) 1577–1588.
[10] J. Ma, S.L. Sun, Optimal linear estimators for systems with random sensor delays, multiple packet dropouts and uncertain observations, IEEE Trans. Signal Process. 59 (11) (2011) 5181–5192.
[11] F.O. Hounkpevi, E.E. Yaz, Robust minimum variance linear state estimators for multiple sensors with different failure rates, Automatica 43 (2007) 1274–1280.
[12] R. Caballero-Águila, A. Hermoso-Carazo, J. Linares-Pérez, Linear and quadratic estimation using uncertain observations from multiple sensors with correlated uncertainty, Signal Process. 91 (2011) 330–337.
[13] F.O. Hounkpevi, E.E. Yaz, Minimum variance generalized state estimators for multiple sensors with different delay rates, Signal Process. 87 (2007) 602–613.
[14] J. Linares-Pérez, A. Hermoso-Carazo, R. Caballero-Águila, J.D. Jiménez-López, Least-squares linear filtering using observations coming from multiple sensors with one- or two-step random delay, Signal Process. 89 (2009) 2045–2052.
[15] J. Ma, S.L. Sun, Information fusion estimators for systems with multiple sensors of different packet dropout rates, Inf. Fusion 12 (2011) 213–222.
[16] R. Caballero-Águila, A. Hermoso-Carazo, J. Linares-Pérez, Least-squares linear estimators using measurements transmitted by different sensors with packet dropouts, Digital Signal Process. 22 (2012) 1118–1125.
[17] K.H. Kim, Development of track to track fusion algorithms, in: Proc. Am. Control Conf. Maryland, 1994, pp. 1037–1041.
[18] S.L. Sun, Z.L. Deng, Multi-sensor optimal information fusion Kalman filter, Automatica 40 (2004) 1017–1023.
[19] J. Feng, M. Zeng, Descriptor recursive estimation for multiple sensors with different delay rates, Int. J. Control 84 (3) (2011) 584–596.
[20] M.J. García-Ligero, A. Hermoso-Carazo, J. Linares-Pérez, Distributed and centralized fusion estimation from multiple sensors with Markovian delays, Appl. Math. Comput. 219 (2012) 2932–2948.
[21] W.-A. Zhang, G. Feng, L. Yu, Multi-rate distributed fusion estimation for sensor networks with packet losses, Automatica 48 (2012) 2016–2028.
[22] S. Sun, W. Xiao, Distributed weighted fusion estimators with random delays and packet dropping, Circuits Syst. Signal Process. 26 (4) (2007) 591–605.
[23] J. Sun, C. Zhang, Distributed optimal fusion estimators for multi-sensor systems with bounded random measurement delays and packet dropouts, J. Comput. Inf. Syst. 8 (10) (2012) 4087–4094.
[24] B. Shen, Z.D. Wang, Y.S. Hung, Distributed H_∞-consensus filtering in sensor networks with multiple missing measurements: the finite-horizon case, Automatica 46 (2010) 1682–1688.
[25] B. Chen, L. Yu, W.-A. Zhang, A. Liu, Distributed robust fusion Kalman filtering for uncertain stochastic systems with random delays and missing measurements, in: Proc. 2011 8th Asian Control Conf., 2011, pp. 1288–1293.
[26] M. Liu, D. Qi, S. Zhang, M. Qiu, S. Zheng, Optimal H_∞ fusion filters for a class of discrete-time intelligent systems with time delays and missing measurement, Neurocomputing 74 (2011) 3741–3752.