Information fusion strategies and performance bounds in packet-drop networks

Alessandro Chiuso [a], Luca Schenato [b]

[a] *Dipartimento di Tecnica e Gestione dei Sistemi Industriali*
University of Padova, Vicenza, (Italy)

[b] *Department of Information Engineering*
University of Padova, Padova (Italy)

Abstract

In this paper we discuss suboptimal distributed estimation schemes for stable stochastic discrete time linear systems under the assumptions that (i) distributed sensors have computation capabilities, (ii) the communication between the sensors and the estimation center is subject to random packet loss, and (iii) there is no communication between sensors. We consider strategies which are based on raw measurement fusion (MF) as well as on fusing local estimates, such as local Kalman filters or other pre-processing rules. We show that the optimal mean square estimation error that can be achieved under packet loss, referred as infinite bandwidth filter (IBF), cannot be reached using a limited bandwidth channel; we also compare these strategies under specific noise regimes. We also propose novel mathematical tools to derive analytical upper and lower bounds for the expected estimation error covariance of the MF and the IBF strategies assuming identical sensors. The theoretical findings are complemented with simulation results.

Key words: Distributed estimation, sensor fusion, packet loss, minimum square estimators, multiple sensors, smart sensors, bounds.

1 Introduction

The rapid growth of large wireless sensor networks capable of sensing and computation promises the design of novel applications, but it is also posing several challenges due to the unavoidable lossy nature of the wireless channel. These challenges are particularly evident in control and estimation applications since packet loss and random delay degrade the overall system performance, thus motivating the development of novel tools and algorithms, as illustrated in the survey [9]. In this work we focus on the problem of estimating a stochastic discrete time linear system observed by a number of sensors which can preprocess sensor data and communicate this information to a central node via a wireless lossy channel.

[1] This paper was not presented at any IFAC meeting. This work has been partially supported by European Union project FP7-ICT-223866-FeedNetBack, by the Italian CaRiPaRo Foundation project "Wise-Wai" and by the Progetto di Ateneo CPDA090135/09 funded by the University of Padova. Corresponding author Alessandro Chiuso Ph. +390498277709

Email addresses: alessandro.chiuso@unipd.it (Alessandro Chiuso), schenato@dei.unipd.it (Luca Schenato).

There is a vast literature regarding distributed estimation and sensor fusion with perfect communication links (see for example [4], and references therein). In particular, there are two classes of problems that are relevant to this work. The first class addresses the problem of distributing computational burden from the central node, where the decision process takes place, to the distributed sensors, under the assumption of perfect communication, i.e. packets arrive with no delay or with a known constant delay. In this context, Willsky, Levy et al. [16] [10] showed that it is possible to reconstruct the centralized Kalman filter (CKF) estimate from local Kalman filter estimates generated by each sensor. In particular, the CKF can be obtained as the output of a linear filter which uses the local Kalman estimates as inputs. More recently Wolfe et al. [17] showed that the computational load of the central node can be reduced even further by running on each sensor a local filter which generates a partial estimate of the state so that the central node just needs to sum the contribution from each node together to recover the CKF estimate. The main difference between [16] [10] and [17] is that in the latter approach all local sensors need to know the whole system dynamics, while in the former approaches only the central node needs to know the dynamical model of the whole system.

The other class of works is related to estimation subject to packet loss and variable delay between the sensor and the estimation center. This problem is particularly relevant in moving target tracking applications based on radar and GPS measurements [4]. For example in [11] the problem of optimal estimation from randomly delayed measurements from multiple channels has been addressed; in [3] and [18] the authors showed how to perform optimal estimation with time-varying delay and out-of-order packets without requiring the storage of large memory buffers and the inversion of many matrices. More recently, in [15] the authors provided lower and upper bounds for the optimal mean square estimator subject to random measurement loss, and in [13] the upper bound was extended to multiple distributed sensors subject to simultaneous packet loss and random delay. Finally, the recent papers [14][1] analyze some tradeoffs between communication, computation and estimation performance in multi-hop tree networks.

However, there are only few scattered results concerned with distributed estimation subject to packet loss when sensors are provided with computation capabilities to preprocess data before transmitting it to the estimation center. A recent result in this direction is given by Gupta et al. [8] who showed that when there is only one sensor, the optimal strategy for the sensor in the presence of packet loss is to send the local Kalman estimate rather than the raw measurement. This is because the local estimate includes the information about all previous measurements, therefore as soon as the central node receives the local estimate it can reconstruct the optimal estimate even if some previous packets were lost. Along the same lines, Robinson et al. [12] showed under what conditions a linear combination of the past measurements can improve estimation performance. Unfortunately, these results do not generalize to multiple sensors each provided with its own lossy communication channel. Differently, a notable work which explicitly focuses on multiple sensors with lossy communication is given by Gupta et al. [7] who proposed a computationally and bandwidth efficient fusion strategy which can guarantee to achieve the same performance of the optimal strategy if each sensor knows the history of the packet loss sequence of all other sensors, i.e. under the assumption that sensors can communicate.

The contribution of this work can be summarized as follows:

- We show that the optimal mean square estimation error that can be achieved under packet loss, referred as infinite bandwidth filter (IBF) (Section 3), cannot be achieved using a limited bandwidth channel (see Theorem 1). As a consequence, we consider several suboptimal strategies with different computational and communication requirements by either fusing measurements (Section 4), or local estimates (Section 5). We also compare these strategies under specific noise regimes namely low process noise and low measurement noise (Section 6). It is proved that no strategy is superior to the others in all scenarios. This is investigated also via simulations confirming that the relative performance depends on the packet loss probability and noise scenarios. Partial results have appeared in [5], more specifically the statement of Theorem 1, the statements and proofs of Theorems 2 and 3, and the example of Section 9.1.
- We derive analytical expressions to compute upper and lower bounds of performance of these estimators assuming i.i.d. Bernoulli packet loss probabilities. Finding bounds on performance turns out to be particularly challenging due to the fact that the estimation error covariance of the different estimators at the central node depends nonlinearly on the specific packet loss sequence of all sensors, therefore computing expected error covariance a-priori given the packet loss statistics becomes a combinatorial problem that explodes with time. In particular, we derive upper and lower bounds in the scenario where all sensors are identical for two specific strategies: the measurement fusion (MF) strategy and the infinite bandwidth filter (IBF) strategy. The MF strategy is based on optimally fusing the raw measurements received by the central station from the sensors, while the IBF strategy is based on the assumption that each nodes sends to the base station not only the current measurement but also all previous measurements in a single packet. We also show through some simulations that some of these bounds are rather tight and can be used to estimate in advance the expected error of the different strategies. Preliminary results have appeared in [6], more specifically the Lemmas 6 and 8, Theorems 9 and 10 , and part of Section 9.2.

The structure of the paper is as follows: Section 2 contains the mathematical formulation of the problem; in Section 3 the optimal strategy under packet loss is presented. The measurement fusion strategy is presented in Section 4 while several strategies based on fusing local estimates are discussed in Section 5. Section 6 contains comparative results under different noise regimes, namely low process noise and low measurement noise. Bounds on the achievable performance are found in Section 7, while Section 8 discusses complexity issues. Simulation results illustrate the theoretical derivation in Section 9 and conclusions end the paper in Section 10.

2 Problem formulation

2.1 Modeling

We consider a stable discrete time linear stochastic systems observed by N sensors:

$$\begin{aligned} x_{t+1} &= Ax_t + w_t \\ y_t^i &= C_i x_t + v_t^i, \quad i = 1, \dots, N \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $y_i \in \mathbb{R}^{m_i}$, A has all eigenvalues inside the unit circle, w_t and v_t^i are uncorrelated, zero-mean, white Gaussian noises with covariances $\mathbb{E}[w_t w_t^\top] = Q$, and $\mathbb{E}[v_t^i (v_t^j)^\top] = R_{ij}\delta_{ij}$, i.e. we assume uncorrelated measurement noise unless differently stated. More compactly, if we define the compound measurement column noise vector $v_t = (v_t^1, \dots, v_t^N) \in$

$\mathbb{R}^m, m = \sum_i m_i$, we have $\mathbb{E}[v_t v_s^\top] = R\delta(t-s)$, where the (i,j)-th block of the matrix $R \in \mathbb{R}^{m\times m}$ is $[R]_{ij} = R_{ij} \in \mathbb{R}^{m_i \times m_j}$. The initial condition x_0 is again a zero-mean Gaussian random variable uncorrelated with the noises and covariance $\mathbb{E}[x_0 x_0^\top] = P_0$, and for convenience we define the matrix $C^\top = [C_1^\top \; C_2^\top \; \dots \; C_N^\top]$. We also assume that $R > 0$ unless differently stated. Note that A being stable guarantees the existence of stable estimators even in the presence of packet loss.

The sensors are not directly connected with each other and can send messages to a common central node through a lossy communication channel, i.e. there is a non zero probability that the message is not delivered correctly. We model the packet dropping events through a binary random variable $\gamma_t^i \in \{0,1\}$ such that:

$$\gamma_t^i = \begin{cases} 0 & \text{if packet sent at time } t \text{ by node } i \text{ is lost} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

Each sensor is provided with computational and memory resources to (possibly) preprocess information before sending it to the central node. More formally, at each time instant t each sensor i sends the preprocessed information $z_t^i \in \mathbb{R}^\ell$:

$$z_t^i = f_t^i(y_1^i, y_2^i, \dots, y_t^i) = f_t^i(y_{1:t}^i) \tag{3}$$

where ℓ is bounded and $f_t^i()$ are causal functions of the local measurements. Natural choices are $z_t^i = y_t^i$, i.e. the latest measurement, or the output of a (time varying) linear filter:

$$\begin{aligned} \xi_t^i &= F_t^i \xi_{t-1}^i + G_t^i y_t^i \\ z_t^i &= H_t^i \xi_t^i + D_t^i y_t^i \end{aligned}$$

as for example a local Kalman filter.

The objective is to design a state estimator at the central node given the information arrived up to time t. More formally, let us define the information set available at the central node as

$$\mathscr{I}_t = \bigcup_{i=1}^N \mathscr{I}_t^i, \quad \mathscr{I}_t^i = \{z_k^i \,|\, \gamma_k^i = 1, k = 1, \dots, t\} \tag{4}$$

Based on this set, we want to design an estimator as follows

$$\hat{x}_{t|t}^g = g_t(\mathscr{I}_t) \tag{5}$$

such that its error $P_{t|t}^g = \mathrm{var}(x_t - \hat{x}_{t|t}^g \,|\, \mathscr{I}_t) = \mathbb{E}[(x_t - \hat{x}_{t|t}^g)(x_t - \hat{x}_{t|t}^g)^\top \,|\, \mathscr{I}_t]$ is small. Depending on the choice of the sensor preprocessing functions f_t^i and the estimator functions g_t, we get different strategies. Note that the estimator error covariance $P_{t|t}^g$ is a function of $\mathscr{I}_t$, and therefore also of the specific packet loss sequence, i.e. it is a random variable. In the following of this section, we propose three different strategies, based on natural choices for the functions f_t^i and g_t. Other choices are obviously possible, as in [12], [7] and [5].

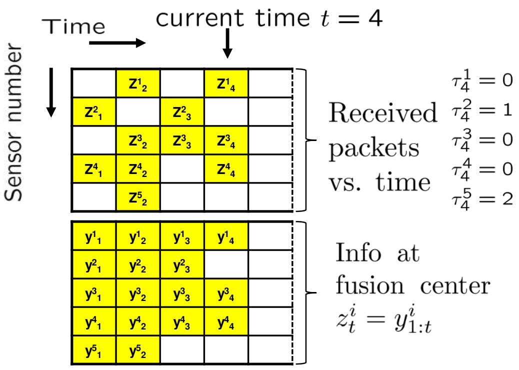


Fig. 1. Snapshot at time $t = 4$ of the information flow for the Infinite Bandwidth Filter (IBF). Top: packets arrived at the fusion center. The value of the i-th row, t-th column is the information received from the fusion center at time t from node i; an empty cell means that the corresponding packet has been lost. Bottom: corresponding information available at the fusion center.

3 Infinite Bandwidth Filter (IBF)

Here we consider the optimal filter in mean square sense that we can obtain if we assume infinite bandwidth in the communication channel when a packet is sent successfully, i.e. each sensor sends to the central node all measurements up to current time:

$$z_t^i = y_{1:t}^i \tag{6}$$

where $y_{1:t}^i = (y_t^i, y_{t-1}^i, \dots, y_1^i)$. This is illustrated in Figure 1. The estimator at the central node is given by

$$\hat{x}_{t|t}^{IBF} = \mathbb{E}[x_t \,|\, \mathscr{I}_t] = \mathbb{E}[x_t \,|\, y^1_{1:t-\tau_t^1}, \dots, y^N_{1:t-\tau_t^N}] \tag{7}$$

where τ_t^i is the time elapsed since the most recent received packet from the i-th sensor at time t, as shown in the top right corner of Fig. 1. This filter is optimal among all possible strategies i.e., more formally,

$$P_{t|t}^{IBF} \le P_{t|t}^g, \quad \forall f_t^i(), \forall \gamma_t^i, \forall g_t()$$

where $P_{t|t}^{IBF} = \mathrm{var}(x_t - \hat{x}_{t|t}^{IBF} \,|\, \mathscr{I}_t)$ is the error covariance of the infinite bandwidth filter. In other words, this filter sets a bound on the achievable performance of any other filter. Unfortunately there is no hope to find a strategy which achieves the same performance with a more parsimonious use of the channel. This finding is formally stated in the following theorem.

Theorem 1 *Let us consider the state estimate $\hat{x}_{t|t}^g$ and $\hat{x}_{t|t}^{IBF}$ defined as above. Then there do not exist (possibly nonlinear) functions $z_t^i = f_t^i(y_{1:t}^i) \in \mathbb{R}^\ell$ with bounded size $\ell < \infty$,* *and functions $g_t(\mathscr{I}_t)$, such that $P^g_{t|t} = P^{IBF}_{t|t}$ for any possible packet loss sequence, i.e.*

$$\nexists f^i_t(), g_t() \mid P^g_{t|t} = P^{IBF}_{t|t}, \ \forall \gamma^i_t$$

The previous theorem states that there is no hope to find a preprocessing with bounded message size which can achieve the error covariance $P^{IBF}_{t|t}$ of the infinite bandwidth filter (IBF) since it is not possible to know in advance what the packet loss event will be. This fact raises the problem of how to find bandwith-efficient strategies with good estimation performance. Here we propose two suboptimal estimation strategies which provide the optimal solution in the special case of perfect communication link, i.e. when there is no packet loss.

4 Measurement Fusion (MF)

The first estimation strategy, referred as measurement fusion (MF), consists in sending the raw measurements

$$z^i_t = y^i_t \tag{8}$$

from each sensor node, and to find the best mean square state estimator with the arrived measurements at the central node:

$$\hat{x}^{MF}_{t|t} = \mathbb{E}[x_t \mid \mathscr{I}_t, i = 1, \ldots, N] \tag{9}$$

where the information set in this case corresponds to $\mathscr{I}^i_t = \{y^i_k \mid \gamma^i_k = 1, k = 1, \ldots, t\}$.

It is possible to explicitly compute the MF filter as follows. Let us first define the following variables:

$$\bar{C}_t = \begin{bmatrix} \gamma^1_t C_1 \\ \vdots \\ \gamma^N_t C_N \end{bmatrix}, \quad \bar{y}_t = \begin{bmatrix} \gamma^1_t y^1_t \\ \vdots \\ \gamma^N_t y^N_t \end{bmatrix}$$

which can be obtained from the centralized matrix C and from the lumped column measurement vector $y_t = (y^1_t y^2_t \ldots y^N_t)^\top$ by replacing the rows and columns corresponding to the lost packets with zeros. It was shown in [13] that the state estimate for the measurement fusion strategy is given by:

$$\hat{x}^{MF}_{t|t} = (I - \bar{L}_t \bar{C}_t) A \hat{x}^{MF}_{t-1|t-1} + \bar{L}_t \bar{y}_t \tag{10}$$

$$P^{MF}_{t|t} = P_{t|t-1} - P_{t|t-1} \bar{C}^\top_t (\bar{C}_t P_{t|t-1} \bar{C}^\top_t + R)^\dagger \bar{C}_t P_{t|t-1} \tag{11}$$

$$= (P^{-1}_{t|t-1} + \bar{C}^\top_t R^{-1} \bar{C}_t)^{-1} = (P^{-1}_{t|t-1} + \textstyle\sum_{i=1}^N \gamma^i_t C^\top_i R^{-1}_{ii} C_i)^{-1} \tag{12}$$

$$\bar{L}_t = P_{t|t-1} \bar{C}^\top_t (\bar{C}_t P_{t|t-1} \bar{C}^\top_t + R)^\dagger \tag{13}$$

$$= P^{MF}_{t|t} \left[\gamma^1_t C^\top_i R^{-1}_{11} \ \ldots \ \gamma^N_t C^\top_i R^{-1}_{NN} \right] \tag{14}$$

$$P_{t+1|t} = A P^{MF}_{t|t} A^\top + Q \tag{15}$$

where the symbol † indicates the Moore-Penrose pseudoinverse, and Eqns. (12),(14) are valid with the additional assumption that $R > 0$ and $P_{t|t-1} > 0$. The previous equations correspond to a time-varying Kalman filter which depends on the packet loss sequence. Note that only measurements that have arrived are used for the computation of the estimate $\hat{x}^{MF}_{t|t}$, i.e. the dummy zero measurement in $\bar{y}_t$ are not used as if they were real measurements, but are discarded.

The measurement fusion strategy has the advantage to be computed recursively and exactly with the inversion of one matrix of (at most) the size of the state vector, as it can be inferred from Eqns. (12) and (14) which correspond to the implementation of the Kalman Filter via the Information Filter [16]. On the other hand, if a packet is lost, then the information conveyed by the measurement in that packet is lost forever, while sending filtered version of the output, as described in the next section, might partially recovered it.

This strategy has been shown to provide good performance in simulations under different noise regimes [5], however, intuitively, it should provide almost optimal performance in a scenario with high ratio between process noise and measurement noise. In fact, if the process noise is large as compared to the measurement noise only most recent measurements convey relevant information, therefore there is no much gain in filtering the past measurements at the sensors. Although this seems to be case in many simulations, there are choices for the system dynamics for which the MF strategy is not optimal even under noiseless measurements, as shown later in Section 6.

5 Fusion of Local Filter Estimates (EF)

The second estimation strategy, named estimate fusion (EF), is based on the fusion of local filtered version of the measurements. According to this strategy, the *i*-th node sends an “estimate” of the state computed via

$$z^i_t \ = \ \Gamma^i_t z^i_{t-1} + G^i_t y^i_t \tag{16}$$

and the central node performs the following fusion rule

$$\hat{x}^{EF}_{t|t} = \mathbb{E}[x_t \mid z^i_{t-\tau^i_t}, i = 1, \ldots, N] = \sum_{i=1}^N \Phi^i_t z^i_{t-\tau^i_t} \tag{17}$$

where $z^i_{t-\tau^i_t}$ is the most recent estimate received by the central node from the sensor node *i*, i.e. τ^i_t is the time elapsed since the most recent packet at time *t* from node *i*. The conditional expectation will be computed assuming a Gaussian measure [1].

[1] Alternatively one could think of $\mathbb{E}[\cdot \mid \cdot]$ as being the best *linear* estimator.

Besides computing the coefficients Φ_t^i, one has also to decide how each node processes its own measurements, i.e. how Γ_t^i and G_t^i are chosen.

Before discussing these choices, we first describe how the gains Φ_t^i can be computed. Let us define:

$$\Phi_t := \left[\Phi_t^1 \dots \Phi_t^N\right] \quad \text{and} \quad z_{t,\tau} := \begin{bmatrix} z^1_{t-\tau_t^1} \\ \vdots \\ z^N_{t-\tau_t^N} \end{bmatrix}.$$

Of course, the optimal fusion coefficients of Eqn. (17) can be computed as:

$$\Phi_t = \mathbb{E}\left[x_t\, z_{t,\tau}^\top\right] \mathbb{E}\left[z_{t,\tau}\, z_{t,\tau}^\top\right]^{-1} \tag{18}$$

A procedure based on a standard state-augmentation argument which allows to compute the covariance matrices $\mathbb{E}\left[x_t\, z_{t,\tau}^\top\right]$ and $\mathbb{E}\left[z_{t,\tau}\, z_{t,\tau}^\top\right]$ is illustrated in Appendix A. The conditional variance of $\tilde{x}_{t|t}^{EF} = x_t - \hat{x}_{t|t}^{EF}$ given the sequence $\{\gamma_s^i\}_{s=1,..,t}$ can be computed using the standard formula for the error covariance

$$P_{t|t}^{EF} = \text{var}\{\tilde{x}_{t|t}^{EF}\,|\,\gamma_s^i,\ s\le t\} = \text{var}\{x_t\} - \Phi_t \mathbb{E}\left[z_{t,\tau}\, z_{t,\tau}^\top\right]\Phi_t^\top \tag{19}$$

This equation will be useful in evaluating the performance of different choices of the local pre-processing strategies Γ_t^i and G_t^i. Of course it can also be used to monitor on-line the performance of the estimator $\hat{x}_{t|t}^{EF}$.

Note that the error covariance of EF, $P_{t|t}^{EF}$ is based only on the latest packet received from each sensor node, therefore is potentially larger than the error covariance that could be obtained by using all received packets, $P_{t|t} = \text{var}(x_t\,|\,\mathscr{I}_t)$, i.e.:

$$P_{t|t}^{IBF} \le P_{t|t} \le P_{t|t}^{EF} \ \ \forall \gamma_t^i.$$

However, the computational price to pay in this case is much larger. The optimal choice of the "local" filter matrices Γ_t^i and G_t^i in Eqn. (16) is far from being a trivial task even if topology and statistics of the model are completely known. Therefore, in order to gain some further intuition, we explore and compare some sensible choices of the matrices Γ_t^i and G_t^i.

5.1 Kalman Estimate Fusion (KEF)

A natural choice for the matrices Γ_t^i and G_t^i, is given by running a local Kalman filter on each sensor, i.e. by computing the best estimate given its own measurements, which is local in nature. More formally:

$$\begin{aligned} \hat{z}_t^{i,l} &= F_t^i \hat{z}_{t-1}^{i,l} + L_t^{i,l} y_t^i \\ F_t^i &= (I - L_t^{i,l} C_i) A \end{aligned}$$

where the gains [2] $L_t^{i,l}$ are the local Kalman filter gains computed as

$$\begin{aligned} P_{t+1}^i &= (A - K_t^{i,l} C_i) P_t^i (A - K_t^{i,l} C_i)^\top + \\ &\quad + K_t^{i,l} R_{ii} \left(K_t^{i,l}\right)^\top + Q \\ L_t^{i,l} &= P_t^i C_i^\top (C_i P_t^i C_i^\top + R_{ii})^{-1} \\ K_t^{i,l} &= A L_t^{i,l} \end{aligned}$$

We shall call the optimal estimate based on the received data $z^{i,l}_{t-\tau_t^i}$, the optimal Kalman estimate fusion (KEF):

$$\hat{x}_{t|t}^{KEF} = \mathbb{E}[x_t\,|\,z^{i,l}_{t-\tau_t^i}, i = 1,\dots,N] = \sum_{i=1}^N \Phi_t^{i,KEF} z^{i,l}_{t-\tau_t^i} \tag{20}$$

Unfortunately, as discussed in [16], even in the absence of packet losses and with uncorrelated measurement noise, the optimal estimate, i.e. the CKF, cannot in general be recovered as a static linear function of the most recent z_t^i only.

5.2 Partial Estimate Fusion (PEF)

This strategy is suggested by the observation that, in the absence of packet losses, one could compute the gains in a centralized manner and distribute the computations to each sensor. To be more precise, assume that all measurements are available to a common location, i.e. that there are no packet losses. We shall denote with $x_{t|t}^{CKF} := \mathbb{E}[x_t\,|\,y_{1:t}^i, i = 1,..,N]$ the centralized Kalman filter (CKF). Its evolution is governed by the equations:

$$\begin{aligned} \hat{x}_{t|t}^{CKF} &= F_t \hat{x}_{t-1|t-1}^{CKF} + L_t y_t \\ F_t &= (I - L_t C) A \end{aligned} \tag{21}$$

where the gain $L_t = [L_t^1\, L_t^2 \cdots L_t^N]$ is the centralized Kalman filter gain computed as

$$\begin{aligned} P_{t+1} &= (A - K_t C) P_t (A - K_t C)^\top + K_t R K_t^\top + Q \\ L_t &= P_t C^\top (C P_t C^\top + R)^{-1} \\ K_t &= A L_t \end{aligned}$$

[2] The superscript ^{i,l} reminds that $z_t^{i,l}$ is the local estimate of the state at the $i-$th sensor, where the gain $L^{i,l}$ is computed using the *local* Kalman filter equations.

Note now that, defining z_t^i to be the solution of

$$z_t^i = F_t z_{t-1}^i + L_t^i y_t^i, \qquad (22)$$

the CKF estimate $\hat{x}_{t|t}^{CKF}$ is given by $\hat{x}_{t|t}^{CKF} = \sum_{i=1}^N z_t^i$. For these reason we shall call the z_t^i's "partial estimates". This strategy was suggested in [17] for distributed estimation to the purpose of reducing the power consumption. Note that Eqn. (22) falls in the class Eqn. (16) with $\Gamma_t^i := F_t$ and $G_t^i := L_t^i$.

Similarly to the KEF strategy, the central node performs the optimal fusion of the most recent packet from each sensor $z_{t-\tau_t^i}^i$ as follows:

$$\hat{x}_{t|t}^{PEF} = \mathbb{E}[x_t \,|\, z_{t-\tau_t^i}^i, i = 1, \ldots, N] = \sum_{i=1}^N \Phi_t^{i,PEF} z_{t-\tau_t^i}^i \qquad (23)$$

where the superscript ^{PEF} stands for the optimal partial estimate fusion and the coefficients Φ_t^i are computed as described in the previous section.

Differently from KEF, in the absence of packet losses this strategy is guaranteed to recover the performance of the CKF even with correlated measurements noise [17].

5.3 Open-loop Partial Estimate Fusion (OPEF)

This strategy is similar to PEF since the sensor nodes perform the same filtering given by Eqn. (22), i.e. they send the partial state estimates according to the centralized Kalman filter gains. However, differently from PEF, the central node rather than computing the optimal gains Φ_t^i given by Eqn. (23), it compensates the packet loss by using the open loop partial state estimate based on the latest received packet from each node, i.e.:

$$x_{t|t}^{OPEF} = \sum_{i=1}^N \Phi_t^{i,OPEF} z_{t-\tau_t^i}^i = \sum_{i=1}^N A^{\tau_t^i} z_{t-\tau_t^i}^i \qquad (24)$$

where τ_t^i is the time elapsed since the most recent packet received from node i at time instant t. Although this looks like a naive solution, it dramatically reduces computational complexity at the central node, and in Section 6 it will be shown to achieve the optimal performance in the small process noise to measurement noise regime.

6 Analysis under special regimes

Even though it seems not possible to perform a rigorous comparative analysis of all the strategies presented in Sections 3, 4 and 5 in full generality, there are two special yet important regimes which deserve some attention. These are the two extreme scenarios in which either the process noise is zero or the measurement noise is zero.

The scenario with zero process noise, i.e. $Q = 0$, corresponds to the case in which a very accurate model is available for the state evolution. In these circumstances the state estimation problem essentially boils down to estimation of the initial condition. The first remarkable but not trivial fact is that the IBF can be computed by a static fusion of the local Kalman filters (KEF) as well as of the partial estimates (PEF). It is also rather intuitive that, in the absence of process noise, there is no loss in propagating estimators just using the system dynamics (i.e. in open loop): this gives also optimality of OPEF. Last it is clear that MF does not use information from lost measurements, and thus cannot be optimal. This is formalized in the next theorem:

Theorem 2 (Small process noise) *Let $Q = 0$ and $R = \mathrm{diag}\{R_1, .., R_N\} > 0$. Then*

$$P_{t|t}^{IBF} = P_{t|t}^{PEF} = P_{t|t}^{KEF} = P_{t|t}^{OPEF} < P_{t|t}^{MF}$$

Differently, in the scenario with zero measurement noise, i.e. $R = 0$, one may think that optimally fusing the latest received measurements would yield optimal performance. This is indeed true for scalar systems, as shown in [12], but it fails to be so for general multivariable systems; in fact, if one considers a system which is observable in n steps, the strategy that performs best depends upon the process noise and the specific loss sequence as discussed in the following theorem:

Theorem 3 (Small measurement noise) *Let us consider $R = 0$ and $Q > 0$. Then there exist scenarios in terms of packet loss sequences and systems dynamics parameters A, C for which for which $P_{t|t}^{MF} > P_{t|t}^{PEF}$ and scenarios for which $P_{t|t}^{MF} < P_{t|t}^{PEF}$.*

Remark 4 *Theorems 2 and 3 study, respectively, the cases $Q = 0, R > 0$ and $R = 0, Q > 0$. It can be proven that the minimum variance estimator of the state given the measurements is a continuous function w.r.t. changes in $\|Q\|$ (Theorem 2) and $\|R\|$ (Theorem 3). This can be verified using the following argument [3]: the covariance matrix Σ_{YY} of the measurements $Y := [y_1^\top, y_1^\top, ..., y_t^\top]^\top$, where $y_k := [(y_k^1)^\top, ..., (y_k^N)^\top]$, is positive definite (in fact bounded away from zero) under the assumptions of both Theorems 2 and 3. In addition both the covariance $\Sigma_{xY} := cov\{x_t, Y\}$ and the variance Σ_{YY} are continuous functions w.r.t. Q and R. Therefore the minimum variance estimator $\hat{x}_{t|t} = \Sigma_{xY}\Sigma_Y^{-1} Y$ is a continuous function w.r.t. to Q and R under both scenarios. This implies that our analysis will give insights also for either small process noise or small measurement noise scenarios, as confirmed by the simulations in Section 9.1.*

[3] We omit the details in the interest of space.

7 Performance bounds

In this Section we turn our attention to computing an analytical upper bound for the performance (state estimation error variance) of the MF, and lower bounds for both MF and IBF. The upper bound for MF is computed resorting to a suboptimal (and hence with larger variance) MF estimation strategy, while for the lower bounds we shall need to study in some detail the structure of the Riccati update which is, to the best of our knowledge, a novel contribution.

7.1 Upper Bound for Measurement Fusion (MF)

An upper bound on the state estimation error variance can be found by computing the error variance for a suboptimal measurement fusion procedure. As discussed in Section 4 the measurement fusion strategy is nothing but a time varying Kalman filter, for which the optimal gain L_t in Eqn. (10)-(15) can be computed on-line and depends on which packets have been received. Of course one could instead consider a suboptimal strategy in which the estimator gain $\bar{L}$ does not depend on the packet loss history. This suboptimal filter, introduced in [13], can be written as:

$$\check{x}^{MF}_{t|t} = (I - \bar{L}_t\bar{C}_t)A\check{x}^{MF}_{t-1|t-1} + \bar{L}_t\bar{y}_t \tag{25}$$

where $\bar{L}_t = [\gamma_t^1\bar{L}^1\, \gamma_t^2\bar{L}^2 \cdots \gamma_t^N\bar{L}^N]$, and $\bar{L}^i$ are constant gains. It has been shown in [13] that the steady state minimum expected error covariance for this filter provides an upper bound for the measurement fusion strategy, as summarized in the following theorem:

Theorem 5 ([13]) *Let us consider the systems of Eqn. (1) with possibly unstable dynamics A and correlated measurement noises, i.e. $R_{ij} \neq 0$, and the filter defined in Eqn. (25). Let*

$$S = \min_{\bar{L}^1,\ldots,\bar{L}^N} \lim_{t\to\infty} \mathbb{E}_\gamma[\mathrm{var}(x_t - \check{x}^{MF}_{t|t-1})]$$

Then S is given by the unique fixed point of the following operator:

$$\begin{aligned}\Psi_\lambda(S) &= ASA^\top + Q - \\ &- \lambda ASC^\top(\lambda CSC^\top + (1-\lambda)S_C + R)^{-1}CSA^\top \\ S_C &= \mathrm{diag}\{C_1SC_1^\top, \ldots, C_NSC_N^\top\}\end{aligned}$$

i.e. $S = \Psi_\lambda(S)$ and has the property that

$$\lim_{t\to\infty} \sup_t \mathbb{E}_\gamma[P^{MF}_{t|t-1}] \leq S$$

The previous theorem basically describes how to compute the best filter among the class of all (suboptimal) filters with constant gains $\bar{L}^i$. Since this filter is suboptimal, it provides also an upper bound for the error estimation error covariance of the MF strategy. Incidentally, being MF suboptimal as compared to the IBF, then it also provides a computable upper bound for the performance of the IBF strategy.

7.2 Lower bounds for the Riccati Equation: identical sensors and stable dynamics

In order to compute lower bounds for the estimation error covariance we first need to study in some detail the structure of the Riccati update for the estimation error covariance. We shall also consider only the case in which there are N identical sensors. More precisely, $C_i = C$ for all i, and $R_{ij} = R\delta_{ij}$ for all i and j.

Let us define:

$$\mathscr{G}(P,L,\ell) := (I - LC)P(I - LC)^\top + \frac{1}{\ell}LRL^\top \tag{26}$$

This is the (filtering) state estimation error using the gain L when the initial state estimate has variance P and measurements from ℓ sensors are utilized. The optimal Kalman gain can be obtained by minimizing $\mathscr{G}(P,L,\ell)$ with respect to L, obtaining

$$L^*(P,\ell) := \arg\min_L \mathscr{G}(P,L,\ell) = PC^\top\left(CPC^\top + \frac{1}{\ell}R\right)^{-1} \tag{27}$$

The corresponding optimal prediction error is given by

$$\begin{aligned}\Phi_f(P,\ell) &:= \mathscr{G}(P,L^*,\ell) \\ &= (I - L^*C)P(I - L^*C)^\top + \tfrac{1}{\ell}L^*R(L^*)^\top \\ &= P - PC^\top\left(CPC^\top + \tfrac{1}{\ell}R\right)^{-1}CP\end{aligned} \tag{28}$$

For future use we also define the prediction error variance update

$$\Phi(P,\ell) := A\Phi_f(P,\ell)A^\top + Q \tag{29}$$

Lemma 6 *The functions $\Phi_f(P,\ell)$ and $\Phi(P,\ell)$ are concave as a function of P and convex as a function of ℓ.*

In the following we shall also make extensive use of a lower bound of the Riccati operator $\Phi(P,\ell)$ as follows. Consider the convex set $\mathscr{P} := \{P = P^\top : P \geq P_m, P \leq P_M\}$. We would like to find a linear function of P, say $G(P,\ell)$ such that $G(P,\ell) \leq \Phi_f(P,\ell)\ \forall P \in \mathscr{P}$.

The matrices P_m and P_M define the set $\mathscr{P}$ over which the linear lower bounds holds. In the rest of the paper we shall always use $P_m = \Phi(P_m,N)$, i.e. the lowest achievable steady state prediction error variance when all N sensors are utilized, and $P_M = AP_MA^\top + Q$, i.e. the open loop steady state variance, which is the upper bound of the state prediction error when no information is available. It is a standard fact to show that, provided $P \in \mathscr{P}$, also $\Phi(P,\ell) \in \mathscr{P}$.

Of course a trivial (constant) lower bound is $G(P,\ell) = \Phi_m(\ell) := \Phi(P_m,\ell)$. This follows from the fact that $\Phi(P,\ell)$ is monotonically non-decreasing in P, i.e. $\Phi(P_2,\ell) \geq \Phi(P_1,\ell)$ holds whenever $P_2 \geq P_1$.

A tighter (linear in P) bound can be taken of the form:

$$\mathcal{L}(P,\ell) = \Phi(P_m,\ell) + \sum_{j=1}^{J} \Psi_j (P - P_m) \Psi_j^\top \tag{30}$$

where Ψ_j have to be chosen so that $\mathcal{L}(P,\ell) \leq \Phi(P,\ell)$, $\forall P \in \mathscr{P}$.

The following theorem discusses the choice of Ψ_j, $j = 1,..,J$ in Eqn. (30):

Theorem 7 *Let us define*

$$\begin{aligned} \bar{A}_m &:= A - K_m C \\ K_m &:= AL^*(P_m,\ell) = AP_mC^\top \left(AP_mC^\top + \tfrac{R}{\ell}\right)^{-1} \end{aligned} \tag{31}$$

The linear (in P) functions in Eqn. (30) are lower bounds for the Riccati update $\Phi(P,\ell)$ provided $J = 1$ and $\Psi_1 = \alpha \bar{A}_m$ for a suitable $\alpha \in [0,1]$. The tightest bound in this class is obtained for $\Psi_1^o := \alpha_0 \bar{A}_m$ where

$$\alpha_0 := \arg\max_{\alpha \in [0,1]} \alpha \quad \text{s.t. } \mathcal{L}(P,\ell,\alpha) \leq \Phi(P,\ell) \ \ \forall P \in \mathscr{P} \tag{32}$$

With a completely analogous argument it can be seen that the linear lower bound for the filtering update

$$\mathcal{L}_f^{LB}(P,\ell) \leq \Phi_f(P,\ell)$$

can be taken of the form

$$\mathcal{L}_f^{LB}(P,\ell) := \Phi_f(P_m,\ell) + \alpha_0^2 (I - L_mC)(P - P_m)(I - L_mC)^\top. \tag{33}$$

where $L_m := L^*(P_m,\ell) = P_mC^\top \left(AP_mC^\top + \frac{R}{\ell}\right)^{-1}$. It is useful to observe that

$$\mathcal{L}^{LB}(P,\ell) = A\mathcal{L}_f^{LB}(P,\ell)A^\top + Q.$$

The following lemma gives a very simple expression of this lower bound for scalar state space systems. The corresponding functions are graphically portrayed in Figure 2.

Lemma 8 *For system with scalar state spaces, i.e. $n = \dim\{x\} = 1$, the function $\mathcal{L}_f^{LB}(P,\ell)$ admits the very simple closed form expression*

$$\mathcal{L}_f^{LB}(P,\ell) = \Phi_f(P_m,\ell) + \beta(P - P_m) \tag{34}$$

where

$$\beta := \frac{\Phi_f(P_M,\ell) - \Phi_f(P_m,\ell)}{P_M - P_m}$$

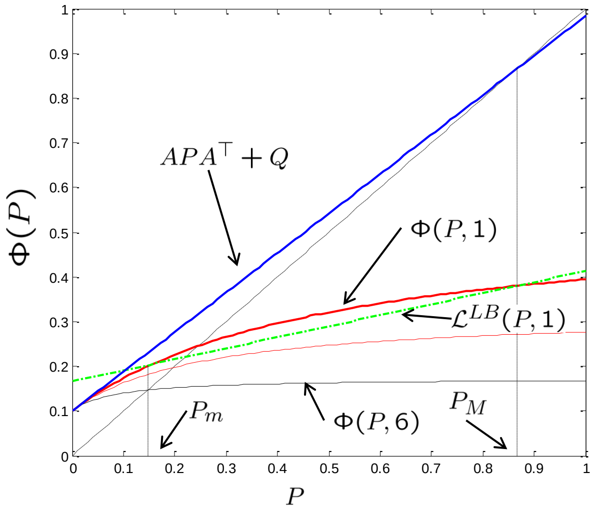


Fig. 2. Graphical representation of bounding functions for scalar system with $A = 0.94, C = 1, Q = 0.1, R = 0.5, N = 6$.

7.3 Lower bound for Measurement Fusion (MF)

The following theorem gives a lower bound on the expected (average) state estimation error for the measurement fusion approach

Theorem 9 *Let P_{MF} and P_{MF}^f respectively the asymptotic prediction and filtering state estimation errors*[4] *. Then*

$$\mathbb{E}[P_{MF}] \geq \bar{P}_{MF}^{LB} \quad \mathbb{E}[P_{MF}^f] \geq \bar{P}_{MF}^{f,LB}$$

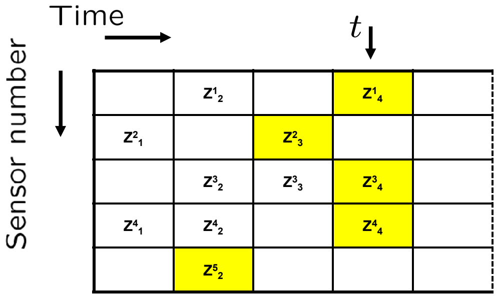
where $\bar{P}_{MF}^{LB}$ is the unique stationary solution of $\bar{P}_{MF}^{LB} = \mathcal{L}^{LB}(\bar{P}_{MF}^{LB}, \mathbb{E}\ell)$ and $\bar{P}_{MF}^{f,LB} = \mathcal{L}_f^{LB}(\bar{P}_{MF}^{LB}, \mathbb{E}\ell)$.

7.4 Lower Bound for Infinite Bandwidth Filter (IBF)

The estimator $\hat{x}_{t|t}^{IBF}$ is characterized by the variables $\tau_t^1, ..., \tau_t^N$; the value of τ_t^i is the number of steps elapsed since the last packet from node i has been received at time t. Under the assumption of identical sensors, the performance of the estimator $\hat{x}_{t|t}^{IBF}$ depends only upon the numbers $h_t^0, h_t^1, h_t^2, \ldots$ where h_t^i can be defined as follows: let us consider, for each node i, only the last packet which has been successfully received; according to the definition above this has happened at time $t - \tau_t^i$. The variable h_t^m represents the number of these packets which have been received at time $t - m$. In formulas:

$$h_t^m := \sum_{i=1}^{N} \delta(\tau_t^i - m)$$

[4] To be rigorous, the asymptotic variances P_{MF} and P_{MF}^f should be defined as the lim-inf of the sequences $P_{t|t-1}^{MF}$ and $P_{t|t}^{MF}$. With a little bit of abuse, we neglect this in the interest of clarity.



$h_t^3 = 0 \quad h_t^2 = 1 \quad h_t^1 = 1 \quad h_t^0 = 3$
$\ell_t^3 = 5 \quad \ell_t^2 = 5 \quad \ell_t^1 = 4 \quad \ell_t^0 = 3$

Fig. 3. Schematics for the definition of h_t^k and ℓ_t^k. The entries are the received packets; highlighted (yellow) are the last packets received from each sensor node: 3 packets received at time $t = 4$, 1 at time $t = 3$ and 1 at time $t = 2$.

where $\delta(\cdot)$ is the Kronecker delta. Let us now fix the time t; the IBF estimator $\hat{x}_{t|t}^{IBF}$ can be computed using the equations for the measurement fusion filter assuming that the equivalent number of packets ℓ_t^m arrived at time $t-m$ is defined, recursively, by the relation

$$\begin{cases} \ell_t^0 = h_t^0 \\ \ell_t^m = \ell_t^{m-1} + h_t^m \quad m = 1, 2, \ldots \end{cases}$$

The definition of h_t^m and ℓ_t^m is graphically illustrated in Figure 3. It is fairly easy to see that the joint probability density function of the variables variables ℓ_t^m can be written in terms of the conditional densities $p(\ell_t^{m+1}|\ell_t^m, \ell_t^{m-1}, \ldots, \ell_t^0) = p(\ell_t^{m+1}|\ell_t^m)$, which have the expression

$$p(\ell_t^{m+1} = \ell|\ell_t^m) = \binom{N - \ell_t^m}{\ell - \ell_t^m} \lambda^{N-\ell}(1-\lambda)^{\ell - \ell_t^m}$$

where λ is the packet loss probability, i.e. $\lambda = \mathbb{E}[\gamma_t^i = 0]$. Based on this we shall now construct a sequence of lower bounds as follows. Let us now fix an integer k and assume that

$$p(\ell_t^k = \ell|\ell_t^{k-1}) = \delta(\ell - N), \tag{35}$$

i.e. that $\ell_t^k = N$. This means that at time $t-k$ all previous measurements from all sensors are available and hence the state filtering error variance at time $t-k$ is the variance of the centralized Kalman estimator $P_{t+1-k|t-k}^{CKF}$. For simplicity of exposition we shall assume that[5] $t-k$ is "large" and hence $P_{t+1-k|t-k}^{CKF}$ converges to P_m, which is the (steady state) prediction error variance for the centralized setting, i.e. when all measurements are available. Of course this is does not happen in practice w.p.1, and hence this assumption will provide a lower bound. Let P_{IBF}^f and P_{IBF} be the (steady state) state filtering and prediction error variance using the IBF. Let us denote with $P_{IBF}^f(t, k, \ell_t^{k-1}, .., \ell_t^0)$ the state filtering error variance at time t and with $P_{IBF}(t+1, k, \ell_t^{k-1}, .., \ell_t^0)$ the state prediction error variance at time $t+1$ assuming the conditional distribution of ℓ_t^k in (35) (and hence $\ell_t^k = N$) and with subsequent cumulative number of arrived packets $\ell_t^{k-1}, .., \ell_t^0$.

It is clear that

$$\begin{aligned} \bar{P}_{IBF}^f(k) &:= \mathbb{E}\left[P_{IBF}^f(t, k, \ell_t^{k-1}, .., \ell_t^0)\right] \\ \bar{P}_{IBF}(k) &:= \mathbb{E}\left[P_{IBF}(t+1, k, \ell_t^{k-1}, .., \ell_t^0)\right] \\ &= A\bar{P}_{IBF}^f(k)A^\top + Q \end{aligned} \tag{36}$$

are increasing functions of k and provide a sequence of lower bounds for $\mathbb{E}[P_{IBF}^f]$ and $\mathbb{E}[P_{IBF}]$, i.e.

$$\mathbb{E}[P_{IBF}^f] = \bar{P}_{IBF}^f(\infty) \geq \cdot\cdot \bar{P}_{IBF}^f(k+1) \geq \bar{P}_{IBF}^f(k) \cdot\cdot \geq \bar{P}_{IBF}^f(1)$$
$$\mathbb{E}[P_{IBF}] = \bar{P}_{IBF}(\infty) \geq \cdot\cdot \bar{P}_{IBF}(k+1) \geq \bar{P}_{IBF}(k) \cdot\cdot \geq \bar{P}_{IBF}(1)$$

The following theorem provides computable lower bounds for the above quantities:

Theorem 10 *The matrices $\bar{P}_{IBF}^f(k)$ and $\bar{P}_{IBF}(k)$ defined in Eqn.* (36) *are lower bounded by $\bar{P}_{IBF}^f(k) \geq \bar{P}_{IBF}^{f,LB}(k)$ and $\bar{P}_{IBF}(k) \geq \bar{P}_{IBF}^{LB}(k)$ where:*

$$\begin{aligned} \bar{P}_{IBF}^{f,LB}(1) &= \Phi_f(P_m, \mathbb{E}\ell_t^0) = \Phi_f(P_m, N(1-\lambda)) \\ \bar{P}_{IBF}^{f,LB}(2) &= \mathbb{E}\left[\mathcal{L}_f^{LB}(\Phi(P_m, \mathbb{E}[\ell_t^1|\ell_t^0]), \ell_t^0)\right] \\ \bar{P}_{IBF}^{f,LB}(k) &= \mathbb{E}\left[\mathcal{L}_f^{LB} \circ \cdots \circ \mathcal{L}^{LB} \circ \Phi(P_m, \mathbb{E}[\ell_t^{k-1}|\ell_t^{k-2}])\right]. \end{aligned}$$

and

$$\begin{aligned} \bar{P}_{IBF}^{LB}(1) &= \Phi(P_m, \mathbb{E}\ell_t) = \Phi(P_m, N(1-\lambda)) \\ \bar{P}_{IBF}^{LB}(2) &= \mathbb{E}\left[\mathcal{L}^{LB}(\Phi(P_m, \mathbb{E}[\ell_t^1|\ell_t^0]), \ell_t^0)\right] \\ \bar{P}_{IBF}^{LB}(k) &= \mathbb{E}\left[\mathcal{L}^{LB} \circ \cdots \circ \mathcal{L}^{LB} \circ \Phi(P_m, \mathbb{E}[\ell_t^{k-1}|\ell_t^{k-2}])\right]. \end{aligned}$$

where $\mathbb{E}[\ell_t^k|\ell_t^{k-1}] = \ell_t^{k-1} + (1-\lambda)(N - \ell_t^{k-1})$ and P_m is the solution of $P_m = \Phi(P_m, N)$, i.e. the optimal (steady state) error variance when measurements from N sensors are received at all times.

In practice one can compute $\bar{P}_{IBF}^{f,LB}(k)$ for increasing values of k until convergence.

It is worth stressing that the lower bound for the IBF provides also a lower bound for MF. Therefore one can use, as a lower bound for MF, whichever is larger among $\bar{P}_{MF}^{f,LB}$ and $\bar{P}_{IBF}^{f,LB}$.

[5] The bounds computed this way will hence be valid for t "large".

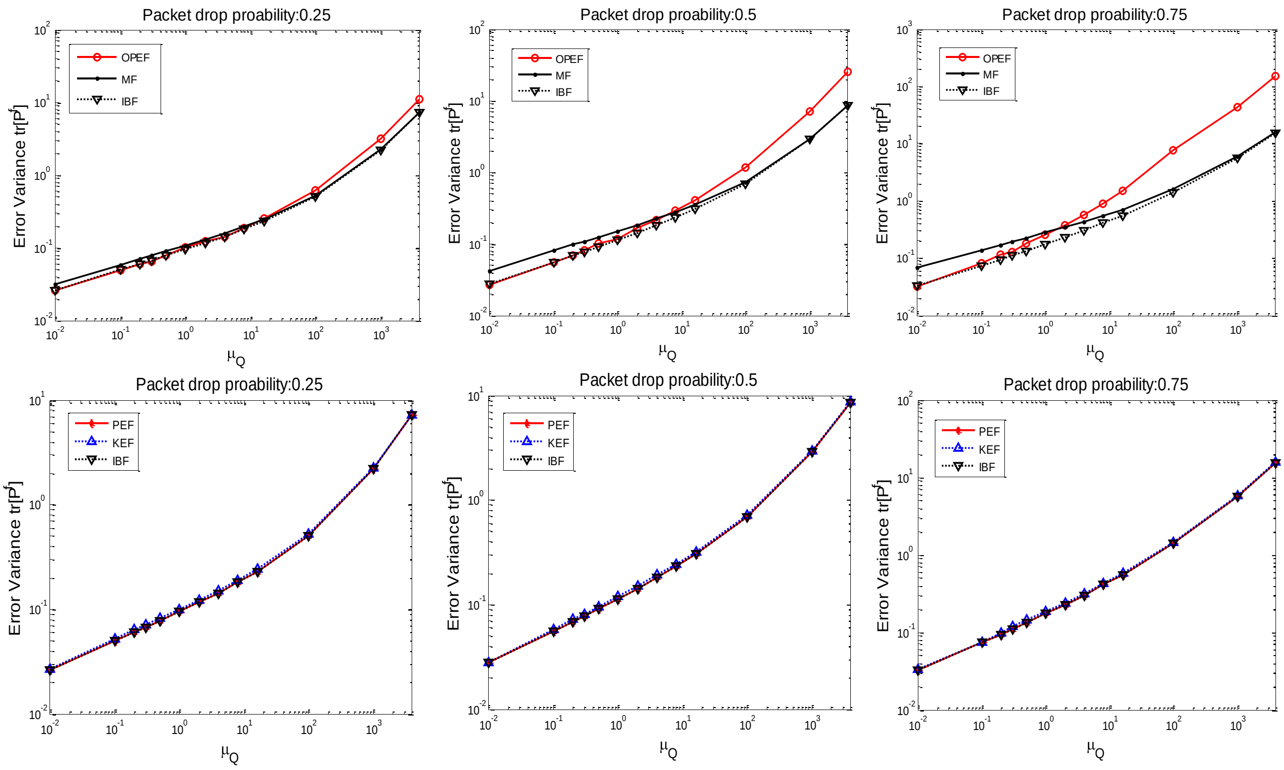


Fig. 4. Trace of error variance vs. μ_Q as defined in Eqn. (37) ($\mu_Q \propto \frac{\|Q\|_2}{\|R\|_2}$) computed via Montecarlo runs for three different packet loss probabilities ($\lambda = 0.25$, $\lambda = 0.5$ and $\lambda = 0.75$). Top: IBF, MF and OPEF. Bottom: KEF, PEF and IBF

8 Complexity Considerations

Besides performance considerations also computational complexity has to be taken into account when implementing estimation algorithms. In fact, complexity influences both computational time as well as energy consumption, which may be a critical issue when using battery powered devices. To derive some quantitative results, let n be the size of the state vector, m_i the size of each measurement vector y^i, N the number of sensors, and k the maximum time elapsed since all most recent packets are received by the central node from each sensor. The computational complexity at the sensor node is none for the MF since the raw measurement is sent, while it is $O(\max(n^2, nm_i))$ for the EF (KEF, PEF, OPEF) due to the computation of the state estimates. At the central node, the computational complexity is $O(n^3)$ for MF since[6] it is necessary to invert a matrix of at most size n, it is $O(N^3n^3k)$ for KEF and PEF, and $O(Nn)$ for OPEF since just a sum is required.

9 Simulation Results

We shall consider two simulation setups in order to illustrate the theoretical findings. In particular:

(1) The results in Section 6 concerning different regimes in terms of ratio between the model noise variance Q and the measurement noise variance R, are verified on a specific example in Section 9.1.
(2) The theoretical bounds for IBF and MF computed under the assumption that all sensors are identical and the dynamics is stable, are illustrated in Section 9.2.

9.1 *Comparison under different noise regimes*

We shall consider the following simulation example with 7 sensors generated by Eqn. (1) with parameters

$$\begin{aligned} A &= \begin{bmatrix} 0.99 & 1 \\ 0 & 0.99 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0.4 & 1 & 1 & 0.4 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top, \\ R &= \mathrm{diag}\{10, 20, 40, 0.5, 2, 1, 40\}, \\ Q &= \mu_Q \,\mathrm{diag}\{10^{-3}, 10^{-3}\} \end{aligned} \tag{37}$$

The parameter μ_Q will be varied to study the behavior under different regimes, i.e. different ratios between the model and the measurements noises.

Figure 4 reports the steady state error variance of the first component of the state as a function of μ_Q, where each point is computed by averaging the (filtering) variance $P_{t|t}$

[6] Using the information form of the Kalman filter as in [1]

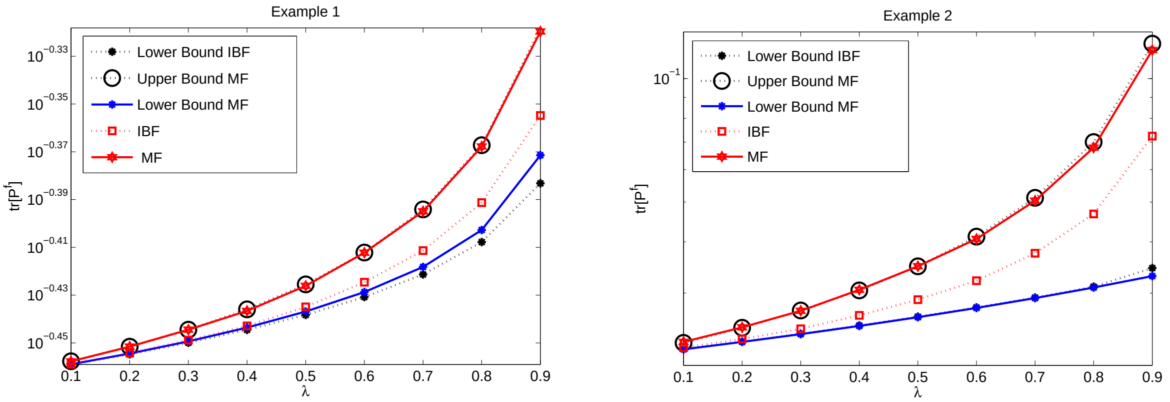


Fig. 5. Trace of Error Variance vs. packet loss probability λ: Montecarlo average over 1000 experiments (MF and IBF) vs. analytical bounds. Left: Example 1, Right: Example 2.

over a Montecarlo run of 5000 time steps packet loss sequence, i.e. $\lim_{t\to\infty}\mathbb{E}_\gamma[P_{t|t}] \approx \frac{1}{T}\sum_{t=1}^{T} P_{t|t}$ for $T = 5000$,[7] for packet drop probabilities $\lambda := \mathbb{P}[\gamma_t^i = 0] \in \{0.25, 0.5, 0.75\}$. For small values of μ_Q, i.e. under the small process noise regime, the OPEF behaves very similarly to PEF. This is reasonable since, for small process noise, it make sense to “trust” the model and hence to propagate estimates in open loop. Note also that MF is the worst strategy for small μ_Q. This is also in line with the results in Section 6 predicting that PEF is better than MF for $Q = 0$. On the opposite regime, i.e. under the small measurement noise, corresponding to large values of μ_Q, the MF is almost undistinguishable from the IBF. Although, this might seem in contrast with Theorem 3, it is important to remark that Figure 4 shows the average performance, while Theorem 3 focuses on a single realization. Therefore, these simulations suggest that from an empirical perspective, the MF behaves optimally under the small measurement noise regime. Differently, note how OPEF performs very poorly in this regime, since it equally weights old and recent estimates. Finally, it is interesting to observe that both KEF and PEF perform very well under all regimes, thus suggesting that how fusion is performed at the sensor nodes, i.e. the choice of the filter parameters Γ_t^i and G_t^i in Eqn. (16), is not too critical, since the central node can extract most the useful information by taking into account the exact correlation among all received local estimates, at the price of high computational cost.

	MF	KEF	PEF	OPEF
Performance for $\frac{\|\|Q\|\|}{\|\|R\|\|} \to 0$	Good	Optimal	Optimal	Optimal
Performance for $\frac{\|\|R\|\|}{\|\|Q\|\|} \to 0$	Almost optimal	Almost optimal	Almost optimal	Very poor
Complexity at sensor	None	Modest	Modest	Modest
Complexity at base station	Moderate	High	High	Very low

Table 1
Summary of the results for prosed strategies: Measurement Fusion (MF), Kalman Estimates Fusion (KEF), Partial Estimate Fusion (PEF), and Open-loop Partial Estimate Fusion (OPEF).

9.2 Bounds for identical sensors

We shall consider the two examples described by the following parameters:

Example 1

$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 0.9 \end{bmatrix}, \; C_i = [\,1 \;\; 0\,], \; Q = \mathrm{diag}\{10^{-2}, 10^{-1}\}, \; R_{ii} = 1$$

for $i = 1, \ldots, 25$.

Example 2:

$$A = \begin{bmatrix} 0.99 & 1 \\ 0 & 0.99 \end{bmatrix}, \; C_i = [\,1 \;\; 0\,], \; Q = \mathrm{diag}\{10^{-3}, 10^{-3}\}, \; R_{ii} = 1$$

for $i = 1, \ldots, 25$.

The packet loss probability λ is varied in the range $\lambda \in [0.1, 0.9]$.

[7] The conditional variance given the packet drop sequence $\{\gamma_t^i\}$ has been computed in closed form as discussed in Section 5 for all methods except OPEF. The unconditional variance is obtained simulating a sufficiently long sequence of packet drop sequence and averaging the conditional variance over that sequence. The same could also have been done for the OPEF; however this is rather involved from a computational point of view and hence the variance for OPEF has been computed purely by Monte Carlo simulations.

The results of a simulation are reported in Figure 5. The variances for MF and IBF are computed as in the previous simulations by averaging the (filtering) variance $P_{t|t}$ over a Montecarlo run of 1000 time steps. For the example considered, we stopped at $k = 3$ for the computation of the lower bound for IBF $\underline{P}_{IBF}^{LB}(k)$. Only marginal improvements could be noticed by increasing k further.

In the specific example, the true performance of the MF algorithm is indistinguishable from its analytical upper bound, while the two lower bounds become less tight as the packet loss probability increases.

10 Conclusions

In this paper we showed that it is not possible to design a bandwidth-limited distributed estimation fusion algorithm which achieves the same estimation error performance of the infinite bandwidth filter when random packet loss occurs. Consequently, we proposed some suboptimal strategies for which we derived some analytical upper and lower performance bounds under different regimes and we studied their computational complexity, as summarized in Table 1. This work and [7] suggest that distributed estimation and fusion with multiple sensors subject to random packet loss require the development of new design strategies as well as novel mathematical tools, and much research needs to be done. Moreover, further research directions include the extension to more complex communication topologies like trees or graphs [1], more realistic packet loss models which include loss correlation and transmission delay, and analytical performance bounds for unstable systems.

Appendix A

The covariance matrices $\mathbb{E}\left[x_t z_{t,\tau}^\top\right]$ and $\mathbb{E}\left[z_{t,\tau} z_{t,\tau}^\top\right]$ can be computed using a standard state-augmentation argument as follows: let us define the augmented state vector $s_t := \left(x_t, z_t^1, .., z_t^N\right)$. By combining Eqn. (1) and Eqn. (16), it is immediate to see that

$$s_t = \Psi_t s_{t-1} + B_t^w w_{t-1} + B_t^v v_t \tag{A.38}$$

where

$$\Psi_t := \begin{bmatrix} A & 0 & \dots & 0 \\ G_t^1 C_1 A \Gamma_t^1 & \Gamma_t^1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ G_t^N C_N A \Gamma_t^N & 0 & 0 & \Gamma_t^N \end{bmatrix}$$

$$B_t^w := \begin{bmatrix} I \\ G_t^1 C_1 \\ \vdots \\ G_t^M C_M \end{bmatrix} \quad B_t^v := \begin{bmatrix} 0 & \dots & 0 \\ G_t^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & G_t^M \end{bmatrix}$$

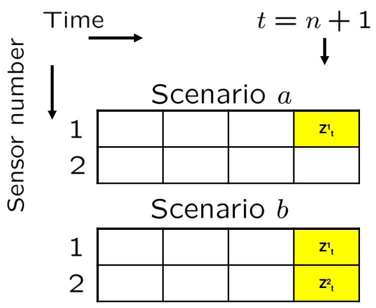


Fig. 6. Illustration of the two scenarios. Scenario a (top): all packets have been lost except the last one sent by sensor 1. Scenario b (bottom): all packets have been lost except the last ones sent by both sensor 1 and 2.

From this equation the covariance function $\Sigma_{h,k} := \mathbb{E}[s_h s_k^\top]$ can be easily computed, starting from the initial condition

$$\Sigma_{0,0} := \begin{bmatrix} \mathbb{E}[x_0 x_0^\top] & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Observe now that all the elements of $\mathbb{E}\left[x_t z_{t,\tau}^\top\right]$ and $\mathbb{E}\left[z_{t,\tau} z_{t,\tau}^\top\right]$ are indeed elements of $\Sigma_{h,k}$ for suitable values of h and k.

Appendix B

Proof of Theorem 1. We will prove the theorem by providing a family of counterexamples: We start by observing that the optimal fusion strategy at the fusion center is given by:

$$g_t^*(\mathscr{I}_t) := \mathbb{E}[x_t \mid \mathscr{I}_t]$$

independently of the choice of the functions $f_t^i()$, i.e. $var(x_t - g_t^*(\mathscr{I}_t)|\mathscr{I}_t) \leq \text{var}(x_t - g_t(\mathscr{I}_t)|\mathscr{I}_t), \forall (g_t, \mathscr{I}_t)$. Let us consider the following dynamical systems with two (identical) sensors:

$$\begin{aligned} x_{t+1} &= A x_t + w_t \\ y_t^1 &= C x_t + v_t^1 \\ y_t^2 &= C x_t + v_t^2 \end{aligned}$$

where $w_t \in \mathbb{R}^n, y^1 \in \mathbb{R}, y^2 \in \mathbb{R}$ and x_0, w_t, v_t^1, v_t^2 are uncorrelated zero-mean white random variables with covariances $\text{var}\{x_0\} = P_0 > 0$, $Var\{w_t\} = Q > 0$, $\sigma_{v^1} = \sigma_{v^2} = r$, respectively. We consider, for any $t = n+1$, two different packet arrival scenarios:

$$\begin{aligned} &a: \{\gamma_1^i = .. = \gamma_n^i = 0, i = 1,2, \gamma_{n+1}^1 = 1, \gamma_{n+1}^2 = 0\}, \\ &b: \{\gamma_1^i = .. = \gamma_n^i = 0, i = 1,2, \gamma_{n+1}^1 = \gamma_{n+1}^2 = 1\} \end{aligned}$$

i.e. at time $t = n+1$ in scenario (a) only the last packet from the first sensor arrived successfully at the central node, while in scenario (b) both packets corresponding to time $t = n+1$ were received but the packets corresponding to times $1,..,n$ were lost.

Let us first consider the expressions for the infinite bandwidth filter (IBF): Under scenario (a) we have

$$\hat{x}^{IBF,a}_{t|t} \;=\; \mathbb{E}[x_t\,|y^1_k, k \le t] = \textstyle\sum_{k=0}^t g^{a,1}_k y^1_k \quad g^{a,1}_k \in \mathbb{R}^n$$

where $g^{a,1}_k$ are computed through the standard Riccati recursions which we write compactly as

$$g^{a,1}_k = \Psi(A,C,P_0,Q,r,k), \tag{B.39}$$

which has the following meaning: $g^{a,1}_k$ is the coefficient vector at time k, obtained from the Riccati recursions for the model with parameters A,C, model noise variance Q, measurement noise variance r and initial variance P_0.

Under scenario (b) we have:

$$\hat{x}^{IBF,b}_{t|t} \;=\; \mathbb{E}[x_t\,|y^1_k, y^2_k; k \le t] = \textstyle\sum_{i=1}^2 \sum_{k=0}^t g^{b,i}_k y^i_k \quad g^{b,i}_k \in \mathbb{R}^n$$

Note also that, since the two sensors are assumed to be identical and with independent noises, the estimator in scenario (b) satisfies $g^{b,1}_k = g^{b,2}_k =: g^b_k$, $\forall k$ and can be written as

$$\hat{x}^{IBF,b}_{t|t} \;=\; \textstyle\sum_{k=0}^t 2 g^b_k \frac{y^1_k + y^2_k}{2}$$

This simply means that $2g^b_k = 2g^{b,1}_k = 2g^{b,2}_k$ is the impulse response of the estimator one would obtain with just one sensor whose measurement is the mean of the measurements of the two sensors (and with halved measurement noise variance).

Thus, with the notation introduced in (B.39), we have that

$$g^b_k = g^{b,1}_k = g^{b,2}_k = \frac{1}{2}\Psi(A,C,P_0,Q,r/2,k). \tag{B.40}$$

We start by showing that there do not exist *linear* functions of the measurement $z^i_t = f^i_t(y^i_{1:t}) = \sum_{k=1}^t \alpha^i_{t,k} y^i_k$ of size n (the state dimension), i.e. $z^i_t \in \mathbb{R}^n$, that can retrieve the optimal mean square estimate $\hat{x}^{IBF}_{t|t}$ for both the scenarios just illustrated. For this to be true the following would have to hold for some suitable matrices $T_{1,a}, T_{1,b}, T_{2,a} \in \mathbb{R}^{n\times n}$:

$$\begin{aligned} \mathbb{E}[x_t|z^1_t] \;&=\; T_{1,a} z^1_t = \hat{x}^{IBF,a}_{t|t} \\ \mathbb{E}[x_t|z^1_t, z^2_t] \;&=\; T_{1,b} z^1_t + T_{2,b} z^2_t = \hat{x}^{IBF,b}_{t|t} \end{aligned}$$

From the first equation we see that, w.l.o.g., we can take $z^1_t = \hat{x}^{IBF,a}_{t|t} = \sum_{k=0}^t g^{a,1}_k y^1_k$. With a symmetric argument, repeated for sensor 2, also $z^2_t = \sum_{k=0}^t g^{a,1}_k y^2_k$ holds true. In order for the second equation to be satisfied, there must exist matrices $T_{1,b}$ and $T_{2,b}$ such that

$$T_{1,b} \sum_{k=0}^t g^{a,1}_k y^1_k + T_{2,b} \sum_{k=0}^t g^{a,1}_k y^2_k = \sum_{k=0}^t 2 g^b_k \frac{y^1_k + y^2_k}{2}$$

which can only happen if $g^{a,1}_k = T g^b_k$ for some T. Note however that $g^{a,1}_k$ and g^b_k satisfy (B.39) and (B.40) respectively. Since the Riccati equation is not linear in the noise variance r, it is not possible to find T such that $g^{a,1}_k = \Psi(A,C,P_0,Q,r,k) = \frac{T}{2}\Psi(A,C,P_0,Q,r/2,k) = T g^b_k$, $\forall k$. For instance, it is straightforward to verify that, for

$$A = \begin{bmatrix} 0.9 & 1 \\ 0 & 0.9 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

and $Q = I$, $r = 1$, $t = n+1 = 3$, we have that

$$\min_{T \in R^{2\times 2}} \sqrt{\sum_{k=1}^3 \|g^{a,1}_k - T g^b_k\|_2^2} = 0.0698$$

confirming that it is not possible to find T such that $g^{a,1}_k = T g^b_k$ for $k = 1,2,3$.

This concludes the proof that there do not exist linear functions of dimension n that allow to retrieve the optimal estimate for all possible packet loss sequences.

These results continue to hold even if we consider more general *nonlinear* functions $z^i_t = f^i_t(y^i_{1:t})$. In fact, as shown in the specific example above, in order to retrieve the optimal estimate, starting from z^1_2 it has to be possible to reconstruct $\sum_{k=0}^{n+1} g^{a,1}_k y^1_k$ under scenario (a) and $\sum_{k=0}^{n+1} g^b_k y^1_k$ under scenario (b). Since, as shown above, $\nexists T$ s.t. $g^{a,1}_k = T g^b_k, \forall k \le n+1$, the central node can also reconstruct $y^1_1, y^1_2, .., y^1_{n+1}$ from z^1_{n+1}. This is equivalent to saying that the function $z^1_{n+1} = f^1_{n+1}(y^1_1, .., y^1_{n+1})$ maps $n+1$ real numbers into n real numbers, and that the central node can reconstruct the $n+1$ real numbers from the n real number z^1_{n+1}, which is clearly impossible [8].

The proof for arbitrary but finite packet size ℓ, i.e $z^i_t \in \mathbb{R}^\ell$ can be obtained similarly by properly constructing different

[8] Of course one could argue that in an infinite bandwidth setup there is essentially no limitation on the number ℓ in (3); however, when bandwidth limitations come into play, resolution requirements would of course impose an upper bound on ℓ. It would also be possible to consider "smart" coding schemes which, however, would have to depend also on the specific packet loss sequence.

packet loss scenarios for which the gains of the optimal linear combination of the measurements are linearly independent, which means that there do not exists linear functions $f_t^i()$ which always recover the optimal mean square estimate $x_{t|t}^{IBF}$. Also similarly to the proof above, this can be extended to general nonlinear functions $f_t^i()$. □

Proof of Theorem 2. We shall give the proof when the matrix A is invertible. If A is singular, then the proof can be easily adapted by first considering a basis transformation and subsequently by restricting to the subspace which corresponds to the non-zero eigenvalues of A.

Let us first consider the IBF given by

$$\begin{aligned}\hat{x}_{t|t}^{IBF} &:= \mathbb{E}[x_t|y^i_{1:t-\tau_t^i}, i=1,..,N] \\ &= A^t\mathbb{E}[x_0|y^i_{1:t-\tau_t^i}, i=1,..,N]\end{aligned}$$

If we denote by

$$\mathscr{O}_t^i := \begin{bmatrix} C_iA \\ C_iA^2 \\ \vdots \\ C_iA^t \end{bmatrix} \quad Y_t^i := \begin{bmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_t^i \end{bmatrix}$$

than a standard formula from linear minimum variance estimation [2] yields:

$$\begin{aligned}\hat{x}_{t|t}^{IBF} = A^t &\left(\textstyle\sum_{i=1}^N (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1}\mathscr{O}^i_{t-\tau_t^i} + P_0^{-1}\right)^{-1} \cdot \\ &\cdot \textstyle\sum_{i=1}^N (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1} Y^i_{t-\tau_t^i}\end{aligned} \tag{B.41}$$

Note also that the i-th local state estimator, i.e. the best estimator that the i-th node can construct based solely on its own measurements, is given by

$$\begin{aligned}z^{i,l}_{t-\tau_t^i} &:= \mathbb{E}[x_{t-\tau_t^i}|y^i_{1:t-\tau_t^i}] \\ &= A^{t-\tau_t^i}\left((\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1}\mathscr{O}^i_{t-\tau_t^i} + P_0^{-1}\right)^{-1} \cdot \\ &\quad \cdot (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1} Y^i_{t-\tau_t^i}\end{aligned}$$

Therefore, using the assumption that A is invertible,

$$\begin{aligned}\hat{x}_{t|t}^{IBF} = A^t &\left(\textstyle\sum_{i=1}^N (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1}\mathscr{O}^i_{t-\tau_t^i} + P_0^{-1}\right)^{-1} \cdot \\ &\cdot \textstyle\sum_{i=1}^N \left((\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1}\mathscr{O}^i_{t-\tau_t^i} + P_0^{-1}\right) A^{-t+\tau_t^i} z^{i,l}_{t-\tau_t^i}\end{aligned} \tag{B.42}$$

holds true. Since the right hand side is a linear function of $z^{i,l}_{t-\tau_t^i}$, also

$$\begin{aligned}\hat{x}_{t|t}^{KEF} &:= \mathbb{E}[x_t|z^{i,l}_{t-\tau_t^i}, i=1,..,N] \\ &= \mathbb{E}\left[\mathbb{E}[x_t|y^i_{1:t-\tau_t^i}, i=1,..,N]\,|\,z^{i,l}_{t-\tau_t^i}, i=1,..,N\right] \\ &= \mathbb{E}[\hat{x}_{t|t}^{IBF}|z^{i,l}_{t-\tau_t^i}, i=1,..,N] \\ &= \hat{x}_{t|t}^{IBF}\end{aligned}$$

holds, thus proving that $P_{t|t}^{KEF} = P_{t|t}^{IBF}$.

Let us now turn our attention to $\hat{x}_{t|t}^{PEF}$. By first computing $\hat{x}_{t|t}^{CKF} := \mathbb{E}[x_t|y^i_{1:t}, i=1,..,N]$ it is simple to observe that the partial estimate z_s^i, $s = t-\tau_t^i$ (see equations (22), (17)) is given by

$$\begin{aligned}z_s^i &= A^s\left(\textstyle\sum_{i=1}^N (\mathscr{O}_s^i)^\top R_i^{-1}\mathscr{O}_s^i + P_0^{-1}\right)^{-1}(\mathscr{O}_s^i)^\top R_i^{-1} Y_s^i \\ &= A^s\left(\textstyle\sum_{i=1}^N (\mathscr{O}_s^i)^\top R_i^{-1}\mathscr{O}_s^i + P_0^{-1}\right)^{-1} \cdot \\ &\quad \cdot \left((\mathscr{O}_s^i)^\top R_i^{-1}\mathscr{O}_s^i + P_0^{-1}\right) A^{-s} z_s^{i,l}\end{aligned}$$

The last equality proves that z_s^i are linear and invertible functions of $z_s^{i,l}$ and therefore

$$\begin{aligned}x_{t|t}^{PEF} &:= \mathbb{E}[x_t|z^i_{t-\tau_t^i}, i=1,..,N] \\ &= \mathbb{E}[x_t|z^{i,l}_{t-\tau_t^i}, i=1,..,N] \\ &= x_{t|t}^{KEF}\end{aligned}$$

thus implying also $P_{t|t}^{PEF} = P_{t|t}^{KEF}$.

If we now consider the open loop strategy $\hat{x}_{t|t}^{OPEF}$, recall that

$$\begin{aligned}\hat{x}_{t|t}^{OPEF} &= \textstyle\sum_{i=1}^N A^{\tau_t^i} z^i_{t-\tau_t^i} \\ &= A^t\left(\textstyle\sum_{i=1}^N (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1}\mathscr{O}^i_{t-\tau_t^i} + P_0^{-1}\right)^{-1} \cdot \\ &\quad \cdot \textstyle\sum_{i=1}^N (\mathscr{O}^i_{t-\tau_t^i})^\top R_i^{-1} Y^i_{t-\tau_t^i}\end{aligned}$$

Note now that the last term on the right hand side is indeed $\hat{x}_{t|t}^{IBF}$ given in Eqn. (B.41), thus proving that $\hat{x}_{t|t}^{OPEF} = \hat{x}_{t|t}^{IBF}$. This yields also the last equality $P_{t|t}^{OPEF} = P_{t|t}^{IBF}$.

Finally, note that $\hat{x}_{t|t}^{MF}$ computes the best estimate given only the measurements which have indeed reached the fusion center; hence its variance is strictly larger (for a generic choice of the dynamics governing the state evolution) than the variance of $\hat{x}_{t|t}^{IBF}$ (IBF), which is the lower bound on the achievable accuracy for any given packet drop sequence. □

Proof of Theorem 3. We start by showing that there exists a scenario for which $P^{MF}_{t|t} > P^{PEF}_{t|t}$. Let us consider the following systems:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad C_1 = [\,1 \;\; 0\,], \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 0$$

where $P_0 = I$, i.e. we consider a single sensor. Suppose that $\gamma^1_1 = 0, \gamma^1_2 = 1$, i.e. the first packet is lost, while the second is received successfully. It is easy to verify that $\hat{x}^{IBF}_{2|2} = \hat{x}^{PEF}_{2|2} = \alpha^1_1 y^1_1 + \alpha^1_2 y^1_2$, where $0 \neq \alpha^1_k \in \mathbb{R}^{2\times 1}, k = 1,2$. Since $\mathbb{E}[y^1_1 | y^1_2] \neq y^1_1$, it follows that $\hat{x}^{MF}_{2|2} \neq \hat{x}^{IBF}_{2|2}$, therefore $P^{MF}_{2|2} > P^{PEF}_{2|2}$. This result is not too surprising since it is already known that $\hat{x}^{IBF}_{t|t} = \hat{x}^{PEF}_{t|t}$ is always true when there is a single sensor [8].

We now prove that there exists a scenario for which $P^{MF}_{t|t} < P^{PEF}_{t|t}$. Consider the same dynamics of the previous example to which we add a second sensor with observation matrix $C_2 = [0 \; 1]$. It is easy to verify that the outputs of the local filter on each sensor according to the PEF strategy are $z^1_t = [y^1_t \;\; 0]^\top$ and $z^2_t = [0 \;\; y^2_t]^\top$. Let us consider the following packet loss sequence $\gamma^1_1 = \gamma^1_2 = \gamma^2_1 = 1, \gamma^2_2 = 0$, therefore $\hat{x}^{IBF}_{2|2} = \mathbb{E}[x_2 | y^1_1, y^1_2, y^2_1] = \hat{x}^{MF}_{2|2}$, while $\hat{x}^{PEF}_{2|2} = \mathbb{E}[x_2 \,|\, z^1_2, z^2_1] = \mathbb{E}[x_2 \,|\, y^1_2, y^2_1]$. It is also possible to verify that $\mathbb{E}[y^1_1 | y^1_2, y^2_1] \neq y^1_1$ since the covariance matrix $\Sigma = \mathbb{E}[\xi\xi^\top]$, where $\xi = [y^1_1 y^1_2 y^2_1]^\top$, is not singular. This implies that $\hat{x}^{PEF}_{2|2} \neq \hat{x}^{IBF}_{2|2}$, therefore $P^{MF}_{2|2} < P^{PEF}_{2|2}$. □

Proof of Lemma 6. Let us first consider $\Phi_f(P,\ell)$. Concavity in P follows rather easily from the fact that $\Phi_f(P,\ell) = \min_L \mathscr{G}(P,L,\ell)$. As far as convexity in ℓ, the following argument can be used: assume h is (positive) real variable and consider the derivatives $\frac{d\Phi_f(P,\ell)}{d\ell}$ and $\frac{d^2\Phi_f(P,\ell)}{d\ell^2}$. It is easy to verify that $\frac{d\Phi_f(P,\ell)}{d\ell} < 0$ and $\frac{d^2\Phi_f(P,\ell)}{d\ell^2} > 0$. The conclusion for $\Phi(P,\ell)$ follows from the fact that $\Phi(P,\ell)$ is an affine function of $\Phi_f(P,\ell)$. This completes the proof. □

Proof of Theorem 7. The structure of the Riccati update imposes some constraints on the matrix Ψ. In particular $\Phi(P,\ell)$ satisfies

$$\Phi(P,\ell) - \Phi(P_m,\ell) \leq \bar{A}_m (P - P_m) \bar{A}_m^\top \tag{B.43}$$

where $\bar{A}_m$ is defined in Eqn. (31). Note that in order to prove (B.43)

$$\Phi(P,\ell) \leq \bar{A}_m P \bar{A}_m^\top + K_m \frac{R}{\ell} K_m^\top + Q \tag{B.44}$$

has been used. It follows from (30) and (B.43) that

$$\mathcal{L}(P,\ell) \leq \Phi(P,\ell) \leq \Phi(P_m,\ell) + \bar{A}_m (P - P_m) \bar{A}_m^\top$$

from which Ψ_j have to be chosen so that [9]

$$\sum_{j=1}^{J} \Psi_j (P - P_m) \Psi_j^\top \leq \bar{A}_m (P - P_m) \bar{A}_m^\top \quad \forall P \in \mathscr{P} \tag{B.45}$$

In particular note that when $P - P_m$ is singular also $\Phi(P,\ell) - \Phi(P_m,\ell)$ (and hence $\mathcal{L}(P,\ell) - \Phi(P_m,\ell)$) is so.

Consider now rank one increments $\Delta^1_P = P - P_m$; it follows that for all Δ^1_P (positive semidefinite and of rank 1)

$$\sum_{j=1}^{J} \Psi_j \Delta^1_P \Psi_j^\top \leq \bar{A}_m \Delta^1_P \bar{A}_m^\top$$

must hold. This implies, in particular, that the range of $\Psi_j \Delta^1_P$ coincide, for all j, with the range of $\bar{A}_m \Delta^1_P$. Since Δ^1_P is an arbitrary rank 1 positive semidefinite matrix, this implies that

$$\Psi_j = \alpha_j \bar{A}_m \quad \forall j \in [1,J]$$

Therefore $\sum_{j=1}^{J} \Psi_j (P - P_m) \Psi_j^\top = \sum_{j=1}^{J} \alpha_j \bar{A}_m (P - P_m) \bar{A}_m^\top$. Thus, w.l.o.g., we can take $J = 1$ and

$$\Psi_1 = \alpha \bar{A}_m \tag{B.46}$$

Moreover, since $\alpha^2 \bar{A}_m (P - P_m) \bar{A}_m \leq \bar{A}_m (P - P_m) \bar{A}_m$ must hold, $\alpha \in [0,1]$ follows. At this point we would like to choose the tightest (linear) lower bound of the form

$$\mathcal{L}(P,\ell,\alpha) = \Phi(P_m,\ell) + \alpha^2 \bar{A}_m (P - P_m) \bar{A}_m^\top \tag{B.47}$$

which is equivalent to maximizing α under the constraint that $\mathcal{L}(P,\ell,\alpha)$ bounds from below $\Phi(P,\ell)$ in the set $\mathscr{P}$, i.e.

$$\alpha_0 := \arg\max_{\alpha \in [0,1]} \alpha \quad \text{s.t. } \mathcal{L}(P,\ell,\alpha) \leq \Phi(P,\ell) \;\; \forall P \in \mathscr{P}.$$

Hence the lower bound is $\mathcal{L}^{LB}(P,\ell) := \mathcal{L}(P,\ell,\alpha_0)$. □

Proof of Lemma 8. The proof is just based on the observation that Eqn. (34) is nothing but the line going through the points of coordinates $(P_m, \Phi_f(P_m,\ell))$ and $(P_M, \Phi_f(P_M,\ell))$. Of course concavity of $\Phi_f(P,\ell)$ guarantees that this line is below $\Phi_f(P,\ell)$ for all $P \in \mathscr{P}$. This is indeed the "optimal" approximation from below, i.e. the linear function in P with the largest slope which goes trough $(P_m, \Phi_f(P_m,\ell))$ and always remains below $\Phi_f(P)$ $\forall P \in \mathscr{P} = [P_m, P_M]$. □

Proof of Theorem 9. The (prediction) state estimation error using the measurement fusion approach satisfies the recursive equation $P_{t+1} = \Phi(P_t, \ell_t)$. From convexity of $\Phi(P,\ell)$ in

[9] Note that this is only a necessary condition for $\mathcal{L}(P,\ell) \leq \Phi(P,\ell)$ to hold.

ℓ, it follows that

$$\mathbb{E}[P_{t+1}|P_t] \geq \Phi(P_t, \mathbb{E}\ell_t)$$

where independence of ℓ_t and P_t has been used. Using the lower bound $\Phi(P,\ell) \geq \mathcal{L}^{LB}(P,\ell)$ it follows that

$$\mathbb{E}[P_{t+1}|P_t] \geq \mathcal{L}^{LB}(P_t, \mathbb{E}\ell_t).$$

Since $\mathcal{L}^{LB}$ is linear in P_t, also

$$\mathbb{E}[P_{t+1}] \geq \mathcal{L}^{LB}(\mathbb{E}P_t, \mathbb{E}\ell_t) \tag{B.48}$$

follows. Using the fact $\mathcal{L}^{LB}(P,\ell)$ is non-decreasing as a function of P, i.e. $\mathcal{L}^{LB}(P_2,\ell) \geq \mathcal{L}^{LB}(P_1,\ell)$ whenever $P_2 \geq P_1$ and using stationarity of ℓ_t (implying $\mathbb{E}\ell_t = \mathbb{E}\ell$), Eqn. (B.48) can be iterated yielding

$$\mathbb{E}[P_{MF}] \geq \bar{P}_{MF}^{LB}, \qquad \bar{P}_{MF}^{LB} = \mathcal{L}^{LB}(\bar{P}_{MF}^{LB}, \mathbb{E}\ell)$$

The bound for the filtering solution is easily obtained observing that $P_t^f = \Phi_f(P_t,\ell_t) \geq \mathcal{L}_f^{LB}(P_t,\ell)$ so that

$$\mathbb{E}P_t^f \geq \mathcal{L}_f^{LB}(\mathbb{E}P_t, \mathbb{E}\ell_t)$$

and therefore

$$\mathbb{E}P_{MF}^f \geq \bar{P}_{MF}^{f,LB} := \mathcal{L}_f^{LB}(\bar{P}_{MF}^{LB}, \mathbb{E}\ell).$$

□

Proof of Proposition 10. For $k=1$, $P_{IBF}(t+1,1,\ell_t^0) = \Phi(P_m,\ell_t^0)$. Then, using convexity of Φ in ℓ_t^0 it follows that

$$\bar{P}_{IBF}(1) = \mathbb{E}P_{IBF}(t+1,1,\ell_t^0) \geq \Phi(P_m, \mathbb{E}\ell_t^0).$$

When $k=2$,

$$\begin{aligned} P_{IBF}(t+1,2,\ell_t^1,\ell_t^0) &= \Phi(\Phi(P_m,\ell_{t-1}),\ell_t) \\ &\geq \mathcal{L}^{MF}(\Phi(P_m,\ell_t^1),\ell_t^0) \end{aligned}$$

holds. Using linearity of $\mathcal{L}^{LB}$ and convexity of Φ in ℓ_t^1 one obtains that (a.s.)

$$\mathbb{E}[P_{IBF}(t+1,2,\ell_t^1,\ell_t^0)|\ell_t^0] \geq \mathcal{L}^{MF}(\Phi(P_m,\mathbb{E}\ell_t^1|\ell_t^0),\ell_t^0)$$

from which,

$$\begin{aligned} \bar{P}_{IBF}(2) &:= \mathbb{E}[\mathbb{E}[P_{IBF}(t,2,\ell_t^1,\ell_t^0)|\ell_t^0]] \\ &\geq \mathbb{E}\left[\mathcal{L}^{MF}(\Phi(P_m,\mathbb{E}\ell_t^1|\ell_t^0),\ell_t^0)\right] \end{aligned}$$

The proofs for $k>2$ and for $P_{IBF}^f(k)$ follow the same lines and are therefore omitted. □

References

[1] A. Agnoli, A. Chiuso, P. D'Errico, A. Pegoraro, and L. Schenato. Sensor fusion and estimation strategies for data traffic reduction in rooted wireless sensor networks. In *Proc. of IEEE ISCCSP'08*, pages 677–682, La Valletta, Malta, 2008.

[2] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, 1979.

[3] Y. Bar-Shalom, H. Chen, and M. Mallick. One-step solution for the general out-of-sequence measurement problem in tracking. *IEEE Transactions on Aerospace and Electronics Systems*, 40(1):27–37, 2004.

[4] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[5] A. Chiuso and L. Schenato. Information fusion strategies from distributed filters in packet-drop networks. In *Proc. of CDC 08*, pages 1079–1084, 2008.

[6] A. Chiuso and L. Schenato. Performance bounds for information fusion strategies in packet-drop networks. In *Proc. of ECC 09*, pages 4326–4331, 2009.

[7] V. Gupta, N. C. Martins, and J. S. Baras. Stabilization over erasure channels using multiple sensors. *IEEE Transactions on Automatic Control*, 57(7):1463–1476, July 2009.

[8] V. Gupta, D. Spanos, B. Hassibi, and R. M. Murray. Optimal LQG control across a packet-dropping link. *Systems and Control Letters*, 56(6):439–446, 2007.

[9] J.P. Hespanha, P. Naghshtabrizi, and Y. Xu. A survey of recent results in networked control systems. *Proceedings of the IEEE, special issue*, 95(1):138–162, January 2007.

[10] B.C. Levy, D.A. Castañon, G.C. Verghese, and A.S. Willsky. A scattering framework for decentralized estimation problems. *Automatica*, 19(4):373–384, 1983.

[11] A.S. Matveev and A.V. Savkin. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Transactions on Automatic Control*, 48(4):670–676, April 2003.

[12] C. Robinson and P. R. Kumar. Sending the most recent observation is not optimal in networked control: Linear temporal coding and towards the design of a control specific transport protocol. In *Proc. of IEEE Conf. on Decision and Control*, pages 334–339, New Orleans, U.S.A., 2007.

[13] L. Schenato. Optimal sensor fusion for distributed sensors subject to random delay and packet loss. In *Proc. of IEEE Conf. on Decision and Control*, pages 1547–1552, New Orleans, U.S.A., 2007.

[14] L. Shi, K. H. Johansson, and R.M. Murray. Estimation over wireless sensor networks: Tradeoff between communication, computation and estimation qualities. In *Proc. of IFAC World Congress*, volume 17, Seoul, Korea, 2008.

[15] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan, and S.S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, pages 1453 – 1464, September 2004.

[16] A.S. Willsky, M.G. Bello, D.A. Castanon, B.C. Levy, and G.C. Verghese. Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *IEEE Transactions on Automatic Control*, 27(4):799–813, 1982.

[17] J.D. Wolfe and J.L. Speyer. A low-power filtering scheme for distributed sensor networks. In *Proceeding of IEEE Conference on Decision and Control (CDC'03)*, pages 6325–6326, 2003.

[18] K.S. Zhang, X. R. Li, and Y.M. Zhu. Optimal update with out-of-sequence observations for distributed filtering. *IEEE Transactions on Signal Processing*, 53(6):1992–2004, 2005.