Data-Driven Communication for State Estimation with Sensor Networks

Giorgio Battistelli [a] Alessio Benavoli [b] Luigi Chisci [a]

[a] *Dipartimento di Sistemi e Informatica, DSI-Università di Firenze*
Via Santa Marta 3, 50139 Firenze, Italy
(e-mail: {battistelli, chisci}@dsi.unifi.it).

[b] *Istituto "Dalle Molle" di Studi sull'Intelligenza Artificiale, 6928 Manno-Lugano, Switzerland*
(e-mail: alessio@idsia.ch)

Abstract

This paper deals with the problem of estimating the state of a discrete-time linear stochastic dynamical system on the basis of data collected from multiple sensors subject to a limitation on the communication rate from the sensors. More specifically, the attention is devoted to a centralized sensor network consisting of: (1) multiple remote nodes which collect measurements of the given system, compute state estimates at the full measurement rate and transmit data (either raw measurements or estimates) at a reduced communication rate; (2) a fusion node that, based on received data, provides an estimate of the system state at the full rate. Local data-driven transmission strategies are considered and issues related to the stability and performance of such strategies are investigated. Simulation results confirm the effectiveness of the proposed strategies.

Key words: State estimation; sensor network; transmission strategies.

1 Introduction

The present paper addresses estimation of the state of a discrete-time linear stochastic dynamical system

$$x_{k+1} = Ax_k + w_k \tag{1}$$

⋆ A preliminary version of this paper has been presented at the 1st IFAC Workshop on Estimation and Control of Networked Systems.

given measurements collected from multiple linear sensors

$$y_k^i = C^i x_k + v_k^i\,, \quad i = 1, \dots, s \tag{2}$$

under a limitation on the communication rate from each remote sensor unit to the state estimation unit. Specifically, a centralized network will be considered consisting of: s remote *sensing nodes* $1, \dots, s$ which collect noisy measurements $y_k^1, \dots, y_k^s$ of the given system, can process them to find filtered estimates $\hat{x}_{k|k}^1, \dots, \hat{x}_{k|k}^s$ and transmit either measurements or estimates to the fusion node; a *fusion node* F which receives data from the s sensors and, based on such data, should provide, in the best possible way, an estimate $\hat{x}_{k|k}$ of the system's state.

In the foregoing, we formalize the concept of *communication strategy (CS)* with fixed rate α^i for sensor i. To this end, let us introduce for each sensor i binary variables c_k^i such that $c_k^i = 1$ if sensor i transmits at time k or $c_k^i = 0$ otherwise. Then, a decision mechanism with rate $\alpha^i \in (0,1)$ can be formally defined as any, deterministic or stochastic, mechanism of generating c_k^i such that

$$\lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \mathbb{E}\left\{c_k^i\right\} \;=\; \alpha^i\,. \tag{3}$$

This indicates that, for each sensor, the averaged number of data transmission per time unit is constrained to take on a value α^i. Such a constraint can be used to model all those practical situations in which the sensing units and the monitoring unit are remotely dislocated with respect to each other and the communication rate between them is severely limited by energy, band and/or security concerns. This occurs, for instance, in wireless sensor networks wherein every transmission typically reduces the lifetime of the sensor devices, since wireless communication represents the major source of energy consumption (Feeney and Nilsson, 2001). Moreover, a reduction in the sensor data transmission rate can be crucial in networked control systems in order to reduce the network traffic and hopefully avoid congestion (Yook *et al.*, 2002).

The main assumption [1] underlying definition (3) is that infinite-precision data are transmitted over the communication channel while the bandwidth limitation is accomplished by imposing a suitable value of the transmission rate (Yook *et al.*, 2002; Xu and Hespanha, 2005; Hespanha *et al.*, 2007; Gupta *et al.*, 2007). The rationale behind such a definition is that "in most digital networks, data is transmitted in atomic units called packets and sending a single bit or several hundred bits consumes the same amount of network resources" (Hespanha *et al.*, 2007). Thus, while data precision (i.e., quantization) is certainly an important issue (see e.g. Li and Wong, 1996; Wong and Brockett, 1997), in some situations the number of transmitted data appears to be an even more critical issue. In this context, the focus will be on the choice of a transmission strategy for deciding which data transmit from the remote sensors.

[1] This assumption, though incompatible with the finite bandwidth, holds in practice provided that quantization errors are negligible with respect to measurement errors.

The idea of controlling data transmission so as to achieve a trade-off between communication costs and estimation performance, usually referred to as *controlled communication* (Hespanha *et al.*, 2007), has recently received great interest in the literature. For example, (Yook *et al.*, 2002) deals with a MIMO networked control system assuming measurement transmission and broadcast communication among the sensors (this implies that each sensor can compute locally a fused estimate). Further Suh *et al.* (2007) proposed a *send-on-delta* strategy wherein each sensor transmits its local measurement only if its value changes more than a specified threshold with respect to the last transmitted one. A similar strategy is proposed and analyzed by Xu and Hespanha (2005); Hespanha *et al.* (2007) for estimate transmission in the case of a single remote sensor and assuming unreliable communication (i.e., each transmitted data can be dropped by the network with a certain probability). It is worth noting that such transmission strategies are somehow related to the concept of *Lebesgue sampling* (Åström and Bernhardsson, 2002). Finally, Battistelli *et al.* (2008, 2009) considered probabilistic transmission strategies wherein the times between consecutive transmissions are random variables governed by a finite-state Markov chain. Besides controlled communication, interesting alternatives for energy-efficient state estimation are sensor scheduling (**?**) as well as the approach proposed by **?** based on the combined use of power and coding control.

In this work, the attention will be devoted to local *data-driven transmission strategies* wherein, at time k, each sensor i decides whether to transmit or not on the basis of the collected measurements and, possibly, of its past transmission pattern. Issues related to the stability and performance of such strategies are investigated. The main contributions of this paper concern: i) the development of a unified framework for analyzing the theoretical properties of data-driven transmission strategies; ii) the derivation of linear fusion rules for combining the information available at the fusion node; iii) the proposal of novel transmission strategies that are based on the idea of minimizing the volume of the non-transmission region and ensures mean-square stability of the estimation error at the fusion node.

Notations: $\mathbb{Z}_+$ is the set of positive integers; given a square matrix M, $\mathrm{tr}(M)$ denotes its trace; given a positive definite matrix M, $\mathcal{E}_M \triangleq \{\zeta : \zeta' M \zeta \leq 1\}$ denotes the ellipsoid centered at the origin associated with M; $\|\zeta\|_M \triangleq (\zeta' M \zeta)^{1/2}$ is the weighted norm of the vector ζ; $\mathbb{E}\{\cdot\}$ and $\mathbb{P}\{\cdot\}$ denote the expectation and, respectively, probability operators; given s square matrices $M^1, \ldots, M^s$, $\mathrm{diag}_i(M^i)$ denotes the block-diagonal matrix whose diagonal blocks are the matrices $M^1, \ldots, M^s$; further, given s matrices $M^1, \ldots, M^s$ with the same number of columns, $\mathrm{col}_i(M^i)$ is the matrix obtained by stacking the matrices $M^1, \ldots, M^s$ one on top of the other; given a generic sequence $\{\zeta_k;\, k = 0, 1, \ldots\}$ and two time instants $k_1 \leq k_2$, we define $\zeta_{k_1:k_2} \triangleq \{\zeta_{k_1}, \zeta_{k_1+1}, \ldots, \zeta_{k_2}\}$.

2 Data-driven transmission strategies for parameter estimation

For the sake of clarity, throughout this section, lower-case bold-face letters will be used to denote random variables, whereas non-bold lower-case letters will denote deterministic quantities as well as random variable realizations.

Before addressing state estimation, it is convenient to preliminarily analyze the properties of data-driven transmission strategies in the case of constant parameter estimation. To this end, consider a constant parameter $\boldsymbol{\theta} \in \mathbb{R}^n$ observed through a noisy measurement channel

$$\mathbf{y} = h(\boldsymbol{\theta}) + \mathbf{v} \tag{4}$$

where $\mathbf{v} \in \mathbb{R}^m$ is a measurement noise independent from $\boldsymbol{\theta}$ with PDF (probability density function) $p_{\mathbf{v}}(\cdot)$. Further, let $p_{\boldsymbol{\theta}}(\cdot)$ be a PDF representing the prior information about $\boldsymbol{\theta}$.

Let $\mathbf{c}$ denote a binary variable taking value 1 if the measurement $\mathbf{y}$ is transmitted and 0 otherwise. A data-driven transmission strategy takes the generic form

$$\mathbf{c} = \begin{cases} 0\,, & \text{if } \mathbf{y} - \tilde{y} \in \mathcal{Y} \\ 1\,, & \text{otherwise} \end{cases} \tag{5}$$

where $\tilde{y} \in \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ is a measurable set. Without loss of generality, $\mathcal{Y}$ is supposed to have its centroid at the origin, i.e., $\int_{\mathcal{Y}} \zeta \, d\zeta = 0$. Notice that the test (5) actually commands the transmission of observations $\mathbf{y}$ that are far away from a certain value $\tilde{y}$, in the sense that the difference $\mathbf{y} - \tilde{y}$ falls outside the region $\mathcal{Y}$.

It is supposed that the vector $\tilde{y}$ as well as the region $\mathcal{Y}$ are chosen so as to ensure a desired transmission rate α, i.e.,

$$\mathbb{P}(\mathbf{c} = 0) = \int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{y}}(\zeta) \, d\zeta = 1 - \alpha\,. \tag{6}$$

In the Bayesian approach, the solution of the estimation problem is given by the conditional PDF $p_{\boldsymbol{\theta}|\mathbf{c}}(\cdot|\cdot)$. When a measurement $\mathbf{y} = y$ is transmitted, i.e. $c = 1$, one has $p_{\boldsymbol{\theta}|\mathbf{c}}(\vartheta|1) \propto p_{\mathbf{v}}(y - h(\vartheta))\, p_{\boldsymbol{\theta}}(\vartheta)$. Conversely, when no measurement is received, i.e. $c = 0$, the posterior PDF can be obtained as

$$p_{\boldsymbol{\theta}|\mathbf{c}}(\vartheta|0) \propto \int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{v}}(\zeta - h(\vartheta)) \, d\zeta \; p_{\boldsymbol{\theta}}(\vartheta)\,. \tag{7}$$

The following lemma provides an interpretation of the posterior PDF (7) that will be very useful in studying the properties of the considered transmission strategy.

Lemma 1 *Consider the measurement channel (4) and the data-driven transmission strategy (5). Then, regardless of the choice of $\tilde{y}$, the posterior PDF of $\boldsymbol{\theta}$ conditioned to the fact that no data have been received (i.e., $\mathbf{c} = 0$) coincides with the posterior PDF of $\boldsymbol{\theta}$ conditioned to the fact that a measurement $\mathbf{z} = \tilde{y}$ has been originated from a virtual measurement channel*

$$\mathbf{z} = h(\boldsymbol{\theta}) + \mathbf{v} - \mathbf{u} \tag{8}$$

where $\mathbf{u} \in \mathbb{R}^m$ is a random variable independent from $\boldsymbol{\theta}, \mathbf{v}$ and uniformly distributed in $\mathcal{Y}$.

Proof: Recalling that the PDF of the sum of two independent random variables is the convolution of the two PDFs, one can write

$$p_{\mathbf{v}-\mathbf{u}}(v) = \int_{\mathbb{R}^m} p_{\mathbf{u}}(u)\, p_{\mathbf{v}}(v+u)\, du \propto \int_{\mathcal{Y}} p_{\mathbf{v}}(v+u)\, du\,.$$

Then, it is immediate to see that

$$p_{\mathbf{z}|\boldsymbol{\theta}}(\tilde{y}|\vartheta) = p_{\mathbf{v}-\mathbf{u}}(\tilde{y} - h(\vartheta)) \propto \int_{\mathcal{Y}} p_{\mathbf{v}}(\tilde{y} - h(\vartheta) + u)\, du$$

and consequently

$$p_{\boldsymbol{\theta}|\mathbf{z}}(\vartheta|\tilde{y}) \propto \int_{\mathcal{Y}} p_{\mathbf{v}}(\tilde{y} - h(\vartheta) + u)\, du\, p_{\boldsymbol{\theta}}(\vartheta)\,.$$

By the change of variables $\zeta = u + \tilde{y}$, the equivalence $p_{\boldsymbol{\theta}|\mathbf{z}}(\vartheta|\tilde{y}) = p_{\boldsymbol{\theta}|\mathbf{c}}(\vartheta|0)$ follows at once. □

It is important to stress that Lemma 1 does not mean that $\mathbf{y} - \tilde{y}$ is uniformly distributed. In fact, this cannot be guaranteed in general[2]. Lemma 1 just points out that, since the two situations give rise to the very same posterior PDF, the case of no transmission can be treated, at least from a Bayesian filtering perspective, as if a measurement $\mathbf{z} = \tilde{y}$ were received from the virtual measurement channel (8).

2.1 Optimal transmission strategy

It is natural to ask whether there is an optimal choice for the vector $\tilde{y}$ and the region $\mathcal{Y}$ that determine the data-driven transmission strategy (5). With this respect, one can see that a possible optimality criterion amounts to minimizing the volume of the non-transmission region $\{\tilde{y}\} + \mathcal{Y}$. Such a choice stems from the observation that, when no data is received, such a region represents the uncertainty on the unknown measurement $\mathbf{y}$ (see also equation (7)). It is worth noting that such a choice corresponds also to minimizing the entropy of the uniform noise $\mathbf{u}$ affecting the virtual measurement channel (8) which is equal to $\log |\mathcal{Y}|$, where $|\mathcal{Y}|$ denotes the volume of $\mathcal{Y}$. This is coherent with the fact that such an entropy provides a measure of the uncertainty associated to the virtual additive noise that is introduced in case of no transmission. Then, the following problem can be stated

$$\text{minimize} \quad |\mathcal{Y}| \qquad \text{subject to} \quad \textstyle\int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{y}}(\zeta)\, d\zeta = 1 - \alpha. \tag{9}$$

In general, the solution of problem (9) depends on the form of the PDF $p_{\mathbf{y}}(\cdot)$. In this connection, by imposing unimodality and radial symmetry of $p_{\mathbf{y}}(\cdot)$, the following result can be readily established.

[2] This state of affairs is analogous to what happens in vector quantization. With this respect, if this is of concern, the distribution of $\mathbf{y} - \tilde{y}$ can be modified, in the lines of classical results on dithered vector quantization (Zamir and Feder, 1995), by means of a random subtractive dither with a suitable distribution.

Proposition 1 *Suppose that the PDF $p_{\mathbf{y}}(\cdot)$ takes the form $p_{\mathbf{y}}(y) = \varphi\left(\|y-\bar{y}\|^2_{\Upsilon}\right)$ where Υ is a positive definite matrix and $\varphi(\cdot)$ is a monotonically non-increasing function. Then, the optimal solution of problem (9) is given by*

$$\tilde{y} = \bar{y}, \qquad \mathcal{Y} \; = \; \delta\, \mathcal{E}_{\Upsilon} \stackrel{\triangle}{=} \{y: \; y = \delta\zeta, \, \zeta \in \mathcal{E}_{\Upsilon}\} \, ,$$

for some scalar $\delta > 0$ depending on the tranmission rate α.

Proof: It can be given by showing that for any pair $(\tilde{y}, \mathcal{Y})$ with $|\mathcal{Y}| < \delta|\mathcal{E}_{\Upsilon}|$, the resulting transmission rate $1 - \int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{y}}(\zeta)\, d\zeta$ must necessarily exceed the desired rate α so that the constraint in problem (9) is violated.

To see this, consider the three sets $\mathcal{I}_1 = (\tilde{y}+\mathcal{Y}) \cap (\bar{y} + \delta|\mathcal{E}_{\Upsilon}|)$, $\mathcal{I}_2 = (\tilde{y}+\mathcal{Y}) \setminus \mathcal{I}_1$, and $\mathcal{I}_3 = (\bar{y} + \delta|\mathcal{E}_{\Upsilon}|) \setminus \mathcal{I}_1$. Then it is possible to write

$$1 - \int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{y}}(\zeta)\, d\zeta = 1 - \int_{\mathcal{I}_1} p_{\mathbf{y}}(\zeta)\, d\zeta - \int_{\mathcal{I}_2} p_{\mathbf{y}}(\zeta)\, d\zeta \geq 1 - \int_{\mathcal{I}_1} p_{\mathbf{y}}(\zeta)\, d\zeta - \varphi(\delta)|\mathcal{I}_2| \tag{10}$$

where the latter inequality follows from the fact that for any $\zeta \in \mathcal{I}_2$ it must be $\|\zeta - \bar{y}\|^2_{\Upsilon} \geq \delta$ and consequently $p_{\mathbf{y}}(\zeta) \leq \varphi(\delta)$. Similarly, it can be seen that

$$1 - \int_{\bar{y}+\delta\mathcal{E}_{\Upsilon}} p_{\mathbf{y}}(\zeta)\, d\zeta = 1 - \int_{\mathcal{I}_1} p_{\mathbf{y}}(\zeta)\, d\zeta - \int_{\mathcal{I}_3} p_{\mathbf{y}}(\zeta)\, d\zeta \leq 1 - \int_{\mathcal{I}_1} p_{\mathbf{y}}(\zeta)\, d\zeta - \varphi(\delta)|\mathcal{I}_3| \, . \tag{11}$$

Since tha pair $(\bar{y}, \delta\mathcal{E}_{\Upsilon})$ fulfills the transmission rate constraint, by combining (10) with (11) it turns out that $1 - \int_{\tilde{y}+\mathcal{Y}} p_{\mathbf{y}}(\zeta)\, d\zeta \geq \alpha + \varphi(\delta)(|\mathcal{I}_3| - |\mathcal{I}_2|)$. Thus, the proof can be concluded by noting that $|\mathcal{Y}| < \delta|\mathcal{E}_{\Upsilon}|$ implies $|\mathcal{I}_2| < |\mathcal{I}_3|$ which, in turn, leads to a violation of the transmission rate constraint for the pair $(\tilde{y}, \mathcal{Y})$. $\square$

Notice that, with such an optimal choice, the transmission condition $\mathbf{y} - \tilde{y} \notin \mathcal{Y}$ corresponds to $\|\mathbf{y} - \bar{y}\|^2_{\Upsilon} \geq \delta$ and the scalar δ plays the role of a transmission threshold.

It is immediate to see that the linear-Gaussian case falls within the framework of Proposition 1. In fact, supposing that $p_{\boldsymbol{\theta}}(\vartheta) = \mathcal{N}(\vartheta; \bar{\vartheta}, \Sigma)\,, \quad p_{\mathbf{v}}(v) = \mathcal{N}(v; 0, R)$ and that the measurement channel (4) takes the form $\mathbf{y} = H\boldsymbol{\theta} + \mathbf{v}$ one has $p_{\mathbf{y}}(y) = \mathcal{N}(y; H\bar{\vartheta}, S)$ where $S \stackrel{\triangle}{=} R + H\Sigma H'$ and $\mathcal{N}(\cdot; \mu, P)$ denotes the Gaussian PDF with mean μ and covariance P. Then, Proposition 1 yields the optimal transmission test $\|\mathbf{y} - H\bar{\vartheta}\|^2_{S^{-1}} \geq \delta$.

Further, in this case, the threshold δ can be readily determined for any given desired transmission probability α. In fact, since $\mathbf{y} - H\bar{\vartheta}$ is a zero-mean normal random variable with covariance S, the quadratic form $\|\mathbf{y} - H\bar{\vartheta}\|^2_{S^{-1}}$ turns out to be a χ^2 random variable with m degrees of freedom. As a consequence,

$$\mathbb{P}(\mathbf{c} = 0) = \mathbb{P}\left\{\|\mathbf{y} - H\bar{\vartheta}\|^2_{S^{-1}} < \delta\right\} = \gamma_m(\delta)$$

where $\gamma_m(\cdot)$ is the cumulative distribution function (CDF) of a χ^2 random variable with m degrees of freedom. Then, the transmission rate constraint is satisfied by imposing $\delta = \gamma_m^{-1}(1-\alpha)$.

3 State estimation: the case of measurement transmission

Let us now turn back our attention to the discrete-time linear dynamical system (1)-(2). In the spirit of the results of Section 2, it is supposed that each sensor i transmits its local measurement y_k^i on the basis of a data-driven strategy of the type

$$c_k^i = \begin{cases} 0\,, & \text{if } \|y_k^i - \tilde{y}_k^i\|^2_{W_k^i} \leq \delta^i \\ \\ 1\,, & \text{otherwise} \end{cases} \tag{12}$$

where the vectors $\tilde{y}_k^i, k \in \mathbb{Z}_+$, the positive definite weight matrices $W_k^i, k \in \mathbb{Z}_+$, and the positive reals δ^i have to be chosen so as to ensure that the transmission rate constraint (3) is satisfied. Hereafter, at a generic time k, $n_k^i \geq 0$ will denote the number of time instants elapsed from the last transmission of sensor i, i.e., n_k^i is such that: $c^i_{k-n_k^i} = 1$ and $c^i_{k-1} = \cdots = c^i_{k-n_k^i+1} = 0.$.

It is worth noting that the model (12) encompasses some data-driven transmission strategies considered in previous works. For example, the send-on-delta strategy proposed in (Suh *et al.*, 2007) for scalar sensors corresponds to (12) with $\tilde{y}_k^i = y^i_{k-n_k^i}$ and $W_k^i = 1$. The main contribution of this section with respect to available results concern: i) the development of a unifying framework (see the forthcoming Theorem 1) for analyzing the theoretical properties of data-driven transmission strategies that can be modelled as in (12); ii) the proposal of an iterative algorithm for generating the vectors $\tilde{y}_k^i$ and the weight matrices $W_k^i, k \in \mathbb{Z}_+$, that can significantly improve the performance of the estimate at the fusion node.

In order to ensure the well-posedness of the state estimation problem, the following assumptions are needed.

A1. The process noise w_k and the measurement noises v_k^i, $i = 1, \ldots, s$, are zero-mean stochastic processes with $\mathbb{E}\{w_k w_t'\} = Q\,\delta_{kt}$, $\mathbb{E}\{v_k^i w_t'\} = 0$, and $\mathbb{E}\{v_k^i (v_t^j)'\} = R^i\,\delta_{ij}\,\delta_{kt}$ where δ_{ij} denotes the Kronecker delta.

A2. $Q > 0\,,\ R^i > 0$ for $i = 1, 2, \ldots, s$, and (A, C) is detectable where $C \triangleq \mathrm{col}_i\left(C^i\right)$.

The first problem that needs to be addressed is how the estimate $\hat{x}_{k|k}$ at the fusion node F is to be computed when the data-driven transmission strategy (12) is adopted. With this respect, one can exploit once again Lemma 1 and treat the case of no transmission from sensor i at time k as if a virtual measurement $z_k^i = \tilde{y}_k^i$ were generated by the measurement channel

$$z_k^i = C^i x_k + v_k^i - u_k^i \tag{13}$$

where u_k^i is uniformly distributed in the ellipsoid $\delta^i \mathcal{E}^i_{W_k}$ and uncorrelated with v_k^i. Note that the $v_k^i - u_k^i$ has zero-mean and covariance matrix [3]

$$\mathbb{E}\{(v_k^i - u_k^i)(v_k^i - u_k^i)'\} = \mathbb{E}\{v_k^i (v_k^i)'\} + \mathbb{E}\{u_k^i (u_k^i)'\} = R^i + \frac{\delta^i}{m_i + 2}(W_k^i)^{-1},$$

[3] Recall that a uniform random variable taking value in an ellipsoid $\mathcal{E}_M \subset \mathbb{R}^m$ has covariance matrix equal to $[(m+2)M]^{-1}$.

where $m_i \triangleq \dim(y_k^i)$. Unfortunately, a closed-form solution for the resulting linear non-Gaussian state estimation problem cannot be found and a suitable suboptimal filter has to be sought. For example, by defining

$$z_k \triangleq \mathrm{col}_i \left(c_k^i y_k^i + (1-c_k^i)\tilde{y}_k^i \right) , \quad R_k \triangleq \mathrm{diag}_i \left(R^i + (1-c_k^i)\frac{\delta^i}{m_i+2}(W_k^i)^{-1} \right) ,$$

the estimate $\hat{x}_{k|k}$ at the fusion node F can be computed by means of the following standard Kalman filter recursion

$$\begin{cases} \hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} \\ P_{k|k-1} = AP_{k-1|k-1}A' + Q \\ S_k = R_k + CP_{k|k-1}C' \\ K_k = P_{k|k-1}C'S_k^{-1} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - C\hat{x}_{k|k-1}) \\ P_{k|k} = P_{k|k-1} - P_{k|k-1}C'S_k^{-1}CP_{k|k-1} \end{cases} \tag{14}$$

which represents the *best linear unbiased estimator* (BLUE) for the linear non-Gaussian virtual measurement model (13). As a further motivation for the above choice, the following stability result can now be stated.

Theorem 1 *Suppose that Assumptions A1 and A2 hold and that the data-driven transmission strategy (12) is adopted. If for each $i = 1, 2, \ldots, s$ the weight matrices W_k^i satisfy the condition*

$$W_k^i \geq \mu^i I \tag{15}$$

for some $\mu^i > 0$, then the estimation error $x_k - \hat{x}_{k|k}$ is uniformly bounded in mean square error for any possible choice of the sequences $\{\tilde{y}_k^i, k \in \mathbb{Z}_+\}$, i.e.,

$$\limsup_{k\to\infty} \mathbb{E}\{\|x_k - \hat{x}_{k|k}\|^2\} < +\infty .$$

Proof: Let $\bar{x}_{k|k}$ be the estimate obtained at time k by replacing z_τ with $\bar{z}_\tau \triangleq \mathrm{col}_i \left(y_\tau^i \right)$, for any $\tau \leq k$, in the Kalman filter recursion (14). Then, one can consider the decomposition $\hat{x}_{k|k} = \bar{x}_{k|k} + \nu_k$ with

$$\nu_k = \sum_{\tau=1}^{k} \left[\prod_{h=\tau+1}^{k} (I - K_h C)A \right] K_\tau (z_\tau - \bar{z}_\tau) \tag{16}$$

and obtain the upper bound

$$\mathbb{E}\{\|x_k - \hat{x}_{k|k}\|^2\} \leq 2\mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\} + 2\mathbb{E}\{\|\nu_k\|^2\} \tag{17}$$

where the latter inequality follows from the fact that

$$\|x_k - \bar{x}_{k|k} - \nu_k\|^2 \leq \left(\|x_k - \bar{x}_{k|k}\| + \|\nu_k\|\right)^2 + \left(\|x_k - \bar{x}_{k|k}\| - \|\nu_k\|\right)^2 = 2\|x_k - \bar{x}_{k|k}\|^2 + 2\|\nu_k\|^2 .$$

As for the first term in the right-hand side of (17), since $\bar{x}_{k|k}$ is computed on the basis of the true measurement vector $\bar{z}_\tau$, one has $\mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\} \leq \mathrm{tr}\left(P_{k|k}\right)$. Moreover, under condition (15), one has that $R_k \leq \bar{R}$ with

$$\bar{R} \triangleq \mathrm{diag}_i \left(R^i + \frac{\delta^i}{(m_i+2)\mu^i} \; I \right) .$$

Then, it is immediate to see that $\limsup_{k\to\infty} \mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\} \leq \limsup_{k\to\infty} \mathrm{tr}\left(P_{k|k}\right) \leq \mathrm{tr}\left(\bar{P}\right)$, where $\bar{P}$ is the unique positive definite solution of the Discrete Algebraic Riccati Equation (DARE)

$$\bar{P} = A\bar{P}A' + Q - A\bar{P}C'(C\bar{P}C' + \bar{R})^{-1}C\bar{P}A' .$$

Consider now the second term in the right-hand side of (17). Taking into account (16), one can write

$$\|\nu_k\| \leq \sum_{\tau=1}^{k} \left\| \left[\prod_{h=\tau+1}^{k} (I - K_h C)A \right] K_\tau \right\| \|z_\tau - \bar{z}_\tau\| .$$

Noting that condition (15) is equivalent to $\|y_k^i - \tilde{y}_k^i\|^2_{W_k^i} \geq \mu^i \|y_k^i - \tilde{y}_k^i\|^2$, then one can see that $c_k^i = 0$ implies $\|y_k^i - \tilde{y}_k^i\|^2 \leq \delta^i/\mu^i$. As a consequence, it turns out that

$$\|z_k - \bar{z}_k\|^2 \leq \sum_{i=1}^{s} \delta^i/\mu^i . \tag{18}$$

Further, since, for any possible transmission pattern, one has that $R_k \leq \bar{R}$, one can invoke Theorem 5.3 in (Anderson and Moore, 1981) and conclude that the Kalman filter recursion (14) yields an exponentially stable estimation error transition matrix $\prod_{h=\tau+1}^{k}(I - K_h C)A$. Thus, it follows that there exist a, b, with $0 \leq a < 1$ and $b \geq 0$, such that

$$\left\| \left[\prod_{h=\tau+1}^{k} (I - K_h C)A \right] K_\tau \right\| \leq b\, a^{k-\tau} . \tag{19}$$

Combining (18) with (19), one has

$$\|\nu_k\|^2 \leq \left(\sum_{\tau=0}^{k} b\, a^{k-\tau} \right)^2 \sum_{i=1}^{s} \delta^i/\mu^i \leq \left(\frac{b}{1-a} \right)^2 \sum_{i=1}^{s} \delta^i/\mu^i$$

and consequently

$$\mathbb{E}\{\|\nu_k\|^2\} \leq \left(\frac{b}{1-a} \right)^2 \sum_{i=1}^{s} \delta^i/\mu^i ,$$

which concludes the proof. □

While Theorem 1 ensures the estimation error at the fusion node to be bounded for any possible choice of the vectors $\tilde{y}_k^i$ and of the weight matrices W_k^i provided that condition (15) is satisfied, the performance of transmission strategy (12) may be significantly affected by the specific mechanism used for generating such quantities. In this connection, a sensible choice should fulfill the following three properties:

(a) in order to minimize the volume of the non-transmission region, the vector $\tilde{y}^i_k$ and the weight matrix W^i_k have to be tailored to the prior distribution of the measurement y^i_k so that $\tilde{y}^i_k$ coincides with the predicted value of y^i_k and W^i_k is proportional to the inverse innovation covariance (see Section 2.1);

(b) the mechanism for generating the vectors $\tilde{y}^i_k$ and the weight matrices W^i_k must use only information available to both the sensor i and the fusion node F;

(c) the weight matrices W^i_k must satisfy condition (15) in order to ensure boundedness of the estimation error at the fusion node.

In order to derive a predictor for the measurement y^i_k, it is convenient to put the system (1) in standard form by separating the part of the system that is unobservable from y^i_k. By invoking classical results on linear system theory (see **?**, Lemma 4.2), this can always be done by means of a suitable nonsingular transformation matrix T^i so that the transformed system matrices take the form

$$(T^i)^{-1}AT^i = \begin{bmatrix} A^i_1 & 0 \\ A^i_{21} & A^i_2 \end{bmatrix}, \quad C^iT^i = \begin{bmatrix} C^i_1 & 0 \end{bmatrix}, \quad (T^i)^{-1} = \begin{bmatrix} D^i_1 \\ D^i_2 \end{bmatrix}$$

with (A^i_1, C^i_1) completely observable by construction. Since the output y^i_k is not influenced at all by the unobservable part of the system, as far as prediction of y^i_k is concerned, one can restrict its attention to the observable subsystem with state $x^i_k \triangleq D^i_1 x_k$ and equations

$$x^i_{k+1} = A^i_1 x^i_k + D^i_1 w_k \tag{20}$$

$$y^i_k = C^i_1 x^i_k + v^i_k \,. \tag{21}$$

Then, by redefining $z^i_k \triangleq c^i_k y^i_k + (1 - c^i_k)\tilde{y}^i_k$ and $R^i_k \triangleq R + (1 - c^i_k)\frac{\delta^i}{m_i+2}(W^i_k)^{-1}$, the prediction $\hat{y}^i_{k|k-1}$ of the measurement y^i_k given the information available to both the sensor i and the fusion node F up to time $k-1$ can be computed by means of the following recursion:

$$\begin{cases} \hat{x}^i_{k|k-1} = A^i_1 \hat{x}^i_{k-1|k-1}, \\ \hat{y}^i_{k|k-1} = C^i_1 \hat{x}^i_{k|k-1}, \\ P^i_{k|k-1} = A^i_1 P^i_{k-1|k-1}(A^i_1)' + D^i_1 Q (D^i_1)', \\ S^i_k = R^i_k + C^i_1 P^i_{k|k-1}(C^i_1)' \\ K^i_k = P^i_{k|k-1}(C^i_1)'(S^i_k)^{-1} \\ \hat{x}^i_{k|k} = \hat{x}^i_{k|k-1} + K^i_k(z^i_k - C^i_1 \hat{x}^i_{k|k-1}), \\ P^i_{k|k} = P^i_{k|k-1} - P^i_{k|k-1}(C^i_1)' \left(S^i_k\right)^{-1} C^i_1 P^i_{k|k-1}. \end{cases} \tag{22}$$

Taking into account the desired properties (a)-(c), as well as the results of Section 2.1, a reasonable choice is

$$\tilde{y}_k^i = \hat{y}_{k|k-1}^i\,, \quad (W_k^i)^{-1} = \frac{1}{\mathrm{tr}(S_k^i)}\, S_k^i \tag{23}$$

The normalization factor $\mathrm{tr}(S_k^i)$ is a way for ensuring that condition (15) is satisfied in that $\frac{1}{\mathrm{tr}(S_k^i)}\, S_k^i \leq I$. Of course other choices for the normalization factor can be devised (e.g., the spectral radius of S_k^i). The main reason for preferring the trace is that it can be computed with little computational effort from the remote sensors.

4 State estimation: the case of estimate transmission

Let us now consider the case of estimate transmission, supposing that each sensor i transmits, instead of its local measurement y_k^i, a local estimate $\hat{x}_{k|k}^i$ on the basis of a data-driven strategy of the type

$$c_k^i = \begin{cases} 0\,, & \text{if } \|\hat{x}_{k|k}^i - \tilde{x}_k^i\|^2_{W_k^i} \leq \delta^i \\ 1\,, & \text{otherwise} \end{cases} \tag{24}$$

where the vectors $\tilde{x}_k^i, k \in \mathbb{Z}_+$, the positive definite weight matrices $W_k^i, k \in \mathbb{Z}_+$, and the positive reals δ^i have to be chosen so as to ensure that the transmission rate constraint (3) is satisfied.

The estimate $\hat{x}_{k|k}^i$ is supposed to be computed at full rate from the measurements $y_{0:k}^i$ by means of a Kalman filter recursion applied to the observable subsystem (20)-(21), i.e., (22) with the virtual measurement z_k^i replaced by the true measurement y_k^i and R_k^i replaced by R^i. Notice that, since each pair (A_1^i, C_1^i) is completely observable by construction, all the local Kalman filters are guaranteed to provide a mean-square-stable estimate of the locally observable state variables x_k^i.

As well known, different fusion algorithms can be devised to combine data collected from multiple sensors and related information in order to derive a fused estimate $\hat{x}_{k|k}$ of the true state x_k (see Li *et al.*, 2003, and the references therein). In what follows, a fusion algorithm based on a BLUE criterion is proposed. To this end, for the sake of simplicity, Assumption A2 is replaced by the following:

A2'. $Q > 0\,,\ R^i > 0$ for $i = 1, 2, \ldots, s$, and (A, C) is observable.

The case of (A, C) detectable but not completely observable could be dealt with by adopting more involved fusion algorithms (see, for instance, the BLUE fusion with prior in Li *et al.*, 2003).

Since each local estimate $\hat{x}_{k|k}^i$ summarizes the information collected by sensor i up to time k, it is reasonable to take into account only such quantities when computing the fused estimate $\hat{x}_{k|k}$. In this connection, one can interpret each local estimate $\hat{x}_{k|k}^i$ as a measurement z_k^i of the true state x_k collected through the (virtual) measurement channel

$$z_k^i = D_1^i\, x_k + (\hat{x}_{k|k}^i - x_k^i)\,. \tag{25}$$

Here, the estimation error $\hat{x}^i_{k|k} - x^i_k$ plays the role of a (virtual) measurement noise.

Of course, equation (25) can be directly applied only for the indices i for which an estimate has been received (i.e., for which $c^i_k = 1$). However, in view of the results of Section 2, one can treat the case of no transmission by replacing (25) with

$$z^i_k = D^i_1\, x_k + (\hat{x}^i_{k|k} - x^i_k) - u^i_k \tag{26}$$

where $z^i_k = \tilde{x}^i_k$ and u^i_k is uniformly distributed in the ellipsoid $\delta^i \mathcal{E}_{W_k}$ as well as uncorrelated with the estimation error $\hat{x}^i_{k|k} - x^i_k$.

Summing up, by defining $z_k \triangleq \mathrm{col}_i\left(c^i_k \hat{x}^i_{k|k} + (1-c^i_k)\tilde{x}^i_k\right)$, $\eta_k \triangleq \mathrm{col}_i\left(\hat{x}^i_{k|k} - x^i_k\right) - \mathrm{col}_i\left((1-c^i_k)u^i_k\right)$ and $H \triangleq \mathrm{col}_i\left(D^i_1\right)$, the information available at node F can be treated as originating from the measurement channel $z_k = Hx_k + \eta_k$. Then the fused estimate $\hat{x}_{k|k}$ can be obtained from z_k according to a BLUE criterion as

$$\hat{x}_{k|k} = \left(H'\Sigma_k^{-1}H\right)^{-1} H'\Sigma_k^{-1} z_k \tag{27}$$

where Σ_k is the covariance of the virtual measurement noise η_k. As for the computation of Σ_k, one can write

$$\Sigma_k = \begin{bmatrix} P^{1,1}_k & \cdots & P^{1,s}_k \\ \vdots & \ddots & \vdots \\ P^{s,1}_k & \cdots & P^{s,s}_k \end{bmatrix} + \mathrm{diag}_i\left((1-c^i_k)\frac{\delta^i}{m_i+2}(W^i_k)^{-1}\right)$$

where $P^{i,j}_k \triangleq \mathbb{E}\left\{(\hat{x}^i_{k|k} - x^i_k)(\hat{x}^j_{k|k} - x^j_k)'\right\}$. Clearly, one has $P^{i,i}_k = P^i_{k|k}$ for any $i = 1, \ldots, s$. For what concerns the cross-covariances $P^{i,j}_k$, noting that the estimation error $\hat{x}^i_{k|k} - x^i_k$ can be expressed in terms of $\hat{x}^i_{k-1|k-1} - x^i_{k-1}$ as

$$\hat{x}^i_{k|k} - x^i_k = (I - K^i_k C^i_1)[A^i_1(\hat{x}^i_{k-1|k-1} - x^i_{k-1}) - D^i_1 w_{k-1}] + K^i_k v^i_k\,,$$

each $P^{i,j}_k$ can be obtained from $P^{i,j}_{k-1}$ by means of the recursion

$$P^{i,j}_k = (I - K^i_k C^i_1)(A^i_1 P^{i,j}_{k-1}(A^j_1)' + D^i_1 Q (D^j_1)')(I - K^j_k C^j_1)'$$

being w_{k-1}, v^i_k, and v^j_k mutually uncorrelated by virtue of assumption A1.

It is worth noting that, in the worst case, the proposed fusion rule requires that $O(s^2)$ recursions be performed. Clearly, this is somehow unavoidable if one wants to take into account all the cross-covariances among the sensor estimates. However, in many situations, the actual complexity is rather smaller (e.g., if all the sensors are identical only one covariance and one cross-covariance has to be computed). Further, if computational complexity at the fusion node is a concern, different fusion rules can be adopted that disregard the information on the cross-covariances (see, e.g., the covariance intersection fusion rule which would require only $O(s)$ recursions (Julier and Uhlmann, 1997)).

Similarly to Section 3, the following theorem can be stated.

Theorem 2 *Suppose that Assumptions A1 and A2' hold and that each sensor transmits its local estimate $\hat{x}^i_{k|k}$ according to the data-driven strategy (24). If for each $i = 1, 2, \ldots, s$ the weight matrices W^i_k satisfy the condition*

$$W^i_k \geq \mu^i I \tag{28}$$

for some positive real μ^i, then the estimation error $x_k - \hat{x}_{k|k}$ is uniformly bounded in mean square error for any possible choice of the sequences $\{\tilde{x}^i_k, k \in \mathbb{Z}_+\}$, i.e.,

$$\limsup_{k\to\infty} \mathbb{E}\{\|x_k - \hat{x}_{k|k}\|^2\} < +\infty\,. \tag{29}$$

Proof: Let $\bar{x}_{k|k}$ be the estimate obtained at time k by replacing z_k with $\bar{z}_k \triangleq \mathrm{col}_i\left(\hat{x}^i_{k|k}\right)$ in equation (27). Then, one has

$$\hat{x}_{k|k} = \bar{x}_{k|k} + \left(H'\Sigma_k^{-1}H\right)^{-1} H'\Sigma_k^{-1}\left(z_k - \bar{z}_k\right)$$

and consequently

$$\mathbb{E}\{\|x_k - \hat{x}_{k|k}\|^2\} \leq 2\,\mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\} + 2\left\|\left(H'\Sigma_k^{-1}H\right)^{-1} H'\Sigma_k^{-1}\right\|^2 \mathbb{E}\{\|z_k - \bar{z}_k\|^2\}\,. \tag{30}$$

Consider the first term in the right-hand side of (30). Since $\bar{x}_{k|k}$ is computed on the basis of the true local estimate vector $\bar{z}_k$, it can be seen that

$$\mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\} \leq \mathrm{tr}\left(H'\Sigma_{k|k}^{-1}H\right)^{-1}\,. \tag{31}$$

For the sake of compactness, let

$$\overline{\Sigma}_{k|k} \triangleq \begin{bmatrix} P_k^{1,1} & \cdots & P_k^{1,s} \\ \vdots & \ddots & \vdots \\ P_k^{s,1} & \cdots & P_k^{s,s} \end{bmatrix}.$$

By defining $\Phi_k \triangleq \mathrm{diag}_i\left(A_1^i - K_k^i C_1^i A_1^i\right)$, the asymptotic behavior of $\overline{\Sigma}_{k|k}$ can be analyzed by noting that it obeys the Lyapunov recursion $\overline{\Sigma}_{k|k} = \Phi_k \overline{\Sigma}_{k-1|k-1}\Phi_k' + \Omega_k$, where Ω_k is a suitable matrix that accounts for the terms dependent on Q and R^i. Since such a recursion does not depend on the transmission pattern $\{c_k, k \in \mathbb{Z}_+\}$, it is an easy matter to check that, under assumption A2': (i) each Kalman gain K^i_k converges exponentially to a steady-state gain K^i; (ii) the matrix Φ_k converges exponentially to the strictly Schur matrix $\Phi \triangleq \mathrm{diag}_i\left(A_1^i - K^i C_1^i A_1^i\right)$; (iii) the matrix Ω_k converges exponentially to a positive definite matrix Ω.

Hence, as k tends to infinity, the covariance $\overline{\Sigma}_{k|k}$ exponentially converges to the unique positive definite solution $\overline{\Sigma}$

of the Algebraic Lyapunov Equation (ALE) $\overline{\Sigma} = \Phi\overline{\Sigma}\Phi' + \Omega$. This, in turn, implies that, under condition (28),

$$\limsup_{k\to\infty} \operatorname{tr}\left(\Sigma_{k|k}\right) \leq \operatorname{tr}\left(\overline{\Sigma}\right) + \sum_{i=1}^{s} \frac{\delta^i}{(m_i+2)\mu^i}$$

Then, uniform boundedness of $\mathbb{E}\{\|x_k - \bar{x}_{k|k}\|^2\}$ follows from (31) and from the fact that, under assumption A2', H is full-rank by construction. In order to conclude the proof, it is sufficient to note that the second term in the right-hand side of (30) remains bounded as well. This simply follows from the fact that, under condition (28), one always has $\|z_k - \bar{z}_k\|^2 \leq \sum_{i=1}^{s} \delta^i/\mu^i$ (which can be shown following the lines of the proof of Theorem 1). $\square$

As to the choice of the vectors $\tilde{x}_k^i$ and of the weight matrices W_k^i, similar considerations to those of Section 3 can be made (see points (a)-(c)). With this respect, notice that $\hat{x}^i_{k|k-n_k^i} = (A_1^i)^{n_k^i}\,\hat{x}_{k-n_k^i|k-n_k^i}$ provides the best prediction of the observable sub-state x_k^i on the basis of the information available to both the sensor i and the fusion node F up to time $k-1$. Thus, a natural choice corresponds to setting

$$\tilde{x}_k^i = \hat{x}^i_{k|k-n_k^i}\,, \quad (W_k^i)^{-1} = \frac{1}{\operatorname{tr}(\overline{P}^i_{k|k-1})}\,\overline{P}^i_{k|k-1}, \tag{32}$$

where

$$\begin{aligned} \overline{P}^i_{k|k-1} &= A_1^i\left[P^i_{k-1|k-1} + (1-c^i_{k-1})\tfrac{\delta^i}{m_i+2}(W^i_{k-1})^{-1}\right](A_1^i)' + D_1^i Q (D_1^i)' \\ &= P^i_{k|k-1} + (1-c^i_{k-1})\tfrac{\delta^i}{m_i+2}A_1^i(W^i_{k-1})^{-1}(A_1^i)'\,. \end{aligned}$$

Notice that the covariance matrix $\overline{P}^i_{k|k-1}$ takes into account the fact that, in case of no transmission at time $k-1$, the fusion node knows that the local estimate $\hat{x}^i_{k-1|k-1}$ is in a neighborhood of $\hat{x}^i_{k-1|k-n^i_{k-1}-1}$.

As a final remark, it is pointed out that data-driven transmission strategies similar to (24), with the choices $\tilde{x}_k^i = \hat{x}^i_{k|k-n_k^i}$ and $W_k^i = I$, have been already proposed by Xu and Hespanha (2005) and Hespanha *et al.* (2007) in the case of a *single* remote sensor supposing unreliable communication (i.e., each message can be dropped by the network with a certain probability). However, as evident from the foregoing developments, the extension to the multi-sensor case is not straightforward and requires many additional efforts, mainly due to the issues concerning the local observability decompositions and the correlations among the estimates. In this connection, the main contribution of this section lies in the development of a novel optimal fusion rule for the multi-sensor case ensuring boundedness of the estimation error at the fusion node.

5 A way to account for packet drops

The fusion algorithms for the data-driven strategies discussed in Sections 3 and 4 interpret a missed reception of data from a particular sensor at a given time instant as an intentional missed transmission, due to the failure of the transmission test, and thus exploit this accordingly considering suitable virtual measurement and measurement noise. Of course, this makes the performance of data-driven strategies somehow sensitive to packet drops (for instance due to congestion). This section discusses a possible way to tackle this problem. For the sake of brevity, only the case of estimate trasmission is considered; the case of measurement transmission could be handled in a similar way.

Let r_k^i be a binary variable which is equal to 1 if at time k the fusion node has received a packet from sensor i and zero otherwise. Hereafter, with a little abuse of notation, n_k^i will denote the number of time instants elapsed since the last data received from sensor i. In a communication channel without packet losses, it clearly results that $r_k^i = c_k^i$. Conversely, in presence of packet losses, it can happen that $r_k^i = 0$ even if $c_k^i = 1$. A probabilistic way to account for packet drops is by modelling the transmission channel as a discrete-time Markov chain with two states, i.e. loss (L) and no-loss (N), and transition probabilities $\mathbb{P}(N|L) = p_r$, $\mathbb{P}(L|L) = 1-p_r$, $\mathbb{P}(L|N) = p_c$, $\mathbb{P}(N|N) = 1-p_c$.[4] Notice that $r_k^i = c_k^i$ when the state of the communication channel is N, while $r_k^i = 0$ when the state of the communication channel is L. Hence, if $r_k^i = 1$ the fusion node knows that the channel is in the state N while if $r_k^i = 0$, because of the Markov chain assumption, it can compute the probability of the channel to be in the state L or N at time k. More specifically, let $\pi_{k|k}^i$ be the probability that the i-th communication channel is in the state L at time k conditioned to the sequence $r_{0:k}^i$, then when $r_k^i = 0$ application of the Bayes recursion yields

$$\pi_{k|k}^i = \frac{\mathbb{P}(r_k^i = 0|L)\pi_{k|k-1}^i}{\mathbb{P}(r_k^i = 0|L)\pi_{k|k-1}^i + \mathbb{P}(r_k^i = 0|N)(1-\pi_{k|k-1}^i)}, \tag{33}$$

where $\mathbb{P}(r_k^i = 0|L)$ and $\mathbb{P}(r_k^i = 0|N)$ denote the probabilities that no packet has been received from sensor i given that the i-th communication channel is in the state L and, respectively, N. Clearly, $\mathbb{P}(r_k^i = 0|L) = 1$ and $\mathbb{P}(r_k^i = 0|N) = \mathbb{P}(c_k^i = 0)$; the latter probability will be denoted by β_k^i. From (33) and the Markov chain assumption, it thus follows that

$$\pi_{k|k}^i = \frac{\pi_{k|k-1}^i}{\beta_k^i + (1-\beta_k^i)\pi_{k|k-1}^i}, \qquad \pi_{k+1|k}^i = (1-p_r)\pi_{k|k}^i + p_c(1-\pi_{k|k}^i). \tag{34}$$

Conversely, in the case $r_k^i = 1$, it results that $\pi_{k|k}^i = 0$. By updating and propagating $\pi_{k|k}^i$, $\mathbb{P}(c_k^i = 0|r_k^i = 0)$ can be computed as

$$\mathbb{P}(c_k^i = 0|r_k^i = 0) = \frac{\mathbb{P}(r_k^i = 0|c_k^i = 0)\mathbb{P}(c_k^i = 0)}{\mathbb{P}(r_k^i = 0|c_k^i = 0)\mathbb{P}(c_k^i = 0) + \mathbb{P}(r_k^i = 0|c_k^i = 1)\mathbb{P}(c_k^i = 1)} = \frac{\beta_k^i}{\beta_k^i + (1-\beta_k^i)\pi_{k|k}^i} \triangleq \lambda_k^i, \tag{35}$$

and $\mathbb{P}(c_k^i = 1|r_k^i = 0) = 1 - \lambda_k^i$.

Remark 1 *In principle, the probability β_k^i could be computed as the probability that the transmision test (24) is not satisfied for the considered choice of $\tilde{x}_k^i$ and W_k^i, conditioned to all the information shared by the fusion node and sensor i at time k. Unfortunately, the on-line calculation of β_k^i turns out to be cumbersome as it would require the determination of a possibly non-Gaussian distribution and its integration over an ellipsoid. In most situations, this has to be ruled out due to computational limitations of remote sensors. Hence, suitable approximations have*

[4] In the literature, this is usually referred to as Gilbert–Elliott channel model (Huang and Dey, 2007). Here, for simplicity, it is assumed that the transition probabilities are the same for each sensor.

to be adopted. A first simple approach is to set β_k^i equal to its asymptotic average $1-\alpha^i$. Clearly, the accuracy of this approximation depends on the specific case under study and is difficult to quantify. However, if necessary, more accurate approaches can be conceived. For instance, one can observe that, for the suggested choice (32) and when the local filter is in steady state, the probability β_k^i depends only on n_k^i, i.e., $\beta_k^i = \beta^i(n_k^i)$. Thus, since each local filter converges exponentially to the steady state, it seems reasonable to compute off line an approximation $\hat{\beta}^i(\cdot)$ of the function $\beta^i(\cdot)$ and then set $\beta_k^i = \hat{\beta}^i(n_k^i)$.

The probabilities λ_k^i can be used to account for packet drops in the fusion algorithms discussed in this paper by resorting to well-known results on estimation in the presence of multiple models (Bar-Shalom *et al.*, 2001). In fact, in the case $r_k^i = 0$ (i.e., no packet has been received from node i at time k), the two possible situations $c_k^i = 0$ and $c_k^i = 1$ can be interpreted as two possible models leading to two different estimate-covariance pairs $\left(m_k^{i,[0]}, V_k^{i,[0]}\right)$ and $\left(m_k^{i,[1]}, V_k^{i,[1]}\right)$, respectively. More specifically, as discussed in Section 4, the case of no transmission would correspond to

$$m_k^{i,[0]} = \tilde{x}_k^i\,, \qquad V_k^{i,[0]} = P_{k|k}^i + \frac{\delta^i}{m_i+2}\left(W_k^i\right)^{-1}\,.$$

Further, in case of packet drop, i.e., $c_k^i = 1$, an open-loop prediction should be used

$$m_k^{i,[1]} = \hat{x}_{k|k-n_k^i}^i\,, \qquad V_k^{i,[1]} = P_{k|k-n_k^i}^i\,.$$

Then, given the modal probabilities computed as in (35), the two model-conditioned estimate-covariance pairs can be combined as follows (see Bar-Shalom *et al.*, 2001, p. 443)

$$\begin{aligned} m_k^i &= \lambda_k^i\, m_k^{i,[0]} + (1-\lambda_k^i)\, m_k^{i,[1]}, \\ V_k^i &= \lambda_k^i\, V_k^{i,[0]} + (1-\lambda_k^i)\, V_k^{i,[1]} + \lambda_k^i\,(m_k^{i,[0]} - m_k^i)(m_{k|k}^{i,[0]} - m_k^i)^T + (1-\lambda_k^i)\,(m_k^{i,[1]} - m_k^i)(m_k^{i,[1]} - m_k^i)^T. \end{aligned} \tag{36}$$

Clearly, this implies that the virtual measurement vector z_k is redefined as $z_k \triangleq \mathrm{col}_i\left(r_k^i \hat{x}_{k|k}^i + (1-r_k^i) m_k^i\right)$ and that the virtual noise covariance is redefined according to (36).

As a final remark, notice that for the suggested choice (32) equation (36) becomes

$$m_k^i = \hat{x}_{k|k-n_k^i}^i\,, \qquad V_k^i = \lambda_k^i\, V_k^{i,[0]} + (1-\lambda_k^i)\, V_k^{i,[1]}. \tag{37}$$

6 Simulation results

The goal of this section is to evaluate the performance of the proposed transmission strategies. To this end, the centralized multi-sensor network is used to estimate the state of an object whose motion is described by the kinematic nearly-constant velocity model:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_k \ + \ \mathbf{w}_k, \qquad \mathbf{y}_k^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k \ + \ \mathbf{v}_k^i,$$

where: $\Delta = 0.1$ is the sampling time; the unknown state vector is given by the position and velocity components along the coordinate axes, i.e., $\mathbf{x} = [p_x, v_x, p_y, v_y]'$; $\mathbf{y}_k^i$ is the measurement of the object's position in Cartesian coordinates provided by the i-th sensor. The covariance matrices of measurement noise have been assumed equal to $\mathbf{R}^{i,j} = \mathrm{diag}(r, r)\delta_{ij}$ for each pair of sensors i, j, where $r > 0$ and δ_{ij} is the Kronecker delta. Conversely, the covariance matrix of process noise has been assumed equal to $\mathbf{Q} = \mathbf{G}q$ with

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0 \end{bmatrix}, \quad \mathbf{G}_0 = \begin{bmatrix} \Delta^3/3 & \Delta^2/2 \\ \Delta^2/2 & \Delta \end{bmatrix}$$

and $q > 0$. Transmission rates that are reciprocal of integers, i.e., α with $1/\alpha \in \mathbb{Z}_+$, have been considered. The following transmission strategies have been compared:

- **Periodic Measurement Transmission (PMT)**: sensors transmit periodically their measurements once every $1/\alpha$ time instants, the phase shift among the sensors being random.
- **Data-Driven Measurement Transmission (DDMT)**: the strategy of equations (12) and (23) for measurement transmission; the threshold δ has been tuned so as to obtain the desired transmission rate α.
- **Periodic Estimate Transmission (PET)**: sensors transmit periodically their estimates once every $1/\alpha$ time instants, the phase shift among the sensors being random.
- **Data-Driven Estimate Transmission (DDET)**: the strategy of equations (24) and (32) for estimate transmission; the threshold δ has been tuned so as to obtain the desired transmission rate α.

In order to investigate the robustness of data-driven strategies to packet loss, simulations have been performed considering an unreliable channel with transition probabilities $p_c = 0.1$ and $p_r = 0.9$ (i.e., there is a probability of 10% that a transmitted packet is not received from the fusion node). Thus, all the above transmission strategies have been modified as described in Section 5 to take into account the effect of packet drops. For the sake of simplicity, the probabilities β_k^i in (35) have been approximated by $1 - \alpha^i$. For periodic strategies, there is no practical change in the fusion algorithm to manage packet drops.

The comparison among the four strategies has been carried out via Monte Carlo simulations with independent runs obtained for random generated trajectories by varying the measurement and process noises realizations. The *time averaged square error* (TASE) at the fusion node F, $TASE = \frac{1}{T}\sum_{k=1}^{T}(x_k - \hat{x}_{k|k})'(x_k - \hat{x}_{k|k})$, has been computed for each simulation run and for each strategy. Then, the mean and the maximum of the TASE over the Monte Carlo runs have been considered as performance indices. The simulation time T and the number of Monte Carlo runs have been chosen equal to 600 and 1000, respectively. Finally, in all simulations the transmission rate α^i has been fixed to 0.1 for all sensors and communication strategies.

Fig. 1 shows the behavior of the mean TASE for the considered strategies as a function of the number of sensors for $q = 5 \cdot 10^{-4}$ and $r = 0.3$. By comparing the performance results of the four strategies, several conclusions can be drawn. First of all, transmitting the estimates is better than transmitting the measurements. In fact, PET and DDET have a lower TASE than PMT and DDMT. This is coherent with what has been already observed in (Xu and Hespanha, 2005; Gupta *et al.*, 2007). Second, data-driven strategies are better than periodic strategies. Notice, in fact, that DDMT provides a notable improvement with respect to PMT in terms of both mean and maximum TASE. In particular, the improvement of the mean TASE for DDMT w.r.t. PMT is almost 47%, when the number of sensors is lower than 5, and decreases to 38% in the 10-sensor case. Also DDET provides an improvement with respect to PET. The improvement ranges between 28% (1 sensor) and 9% (10 sensors) for the mean TASE and is always more than 13% for the maximum TASE. Finally, the performance of periodic w.r.t. data-driven strategies is worse in the worst-case (maximum TASE) than in the average case (mean TASE). This depends on the fact that, whenever the phase shift among the sensors is not well distributed (i.e., all sensors are almost in-phase), the estimation performance of the periodic strategies undergoes a notable degradation.

Further insights can be obtained by observing the trend of the trace of the covariance matrix $P^i_{k|k}$ for the two strategies PET and DDET (which is shown in Fig. 2 for a given sensor i in the case of no packet drops). Recall that $P^i_{k|k}$ is the covariance used in the fusion algorithm to combine the estimate from the i-th sensor with the estimates from the other sensors. For PET, the trend of the trace is periodic and varies linearly from a minimum value, obtained in correspondence of the instant in which an estimate is transmitted from sensor i, to a maximum value obtained at the instant before transmission. Conversely, for the data driven strategy, it can be noticed that: (i) the trend of the trace for DDET is aperiodic, since the transmissions are determined by the transmission test (24); (ii) trace remains constant between two consecutive transmissions. In fact, from (24) we can determine a region which includes the estimate of the i-th sensor (even if this estimate is not transmitted) and, thus, determine a bound for the variance of the estimate used in the fusion algorithm. This mainly determines the performance advantages of data driven over periodic strategies. It must also be pointed out that for data-driven strategies the transmission rate is achieved only on average, while for a short time window of a single run the transmissions are allowed to be more or less frequent (see fig. 2) based on necessity, as established by the transmission test (24).

From the above results, it can be concluded that data-driven strategies are particularly convenient in sensor networks characterized by a low communication rate and a small number of sensors. In this case, in fact, because of the low number of transmissions between the sensors and the fusion node, the estimation performance becomes very sensitive

(a) (b)

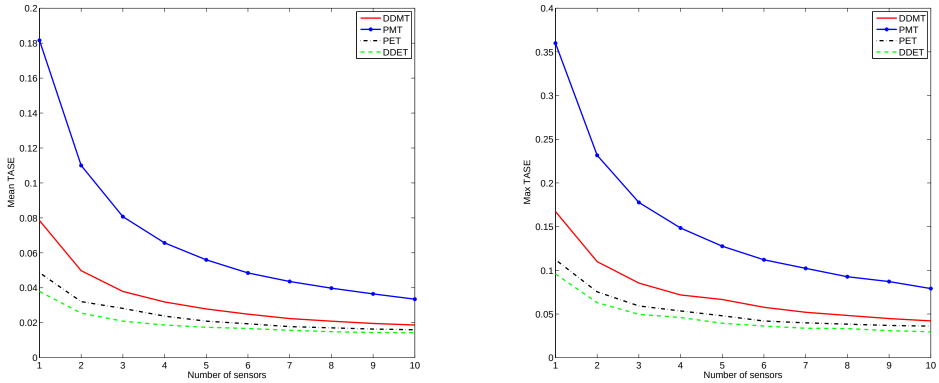

Fig. 1. Mean (a) and maximum (b) TASE for the considered strategies as a function of the number of sensors in the case $q = 5 \cdot 10^{-4}$ and $r = 0.3$.

(a) (b)

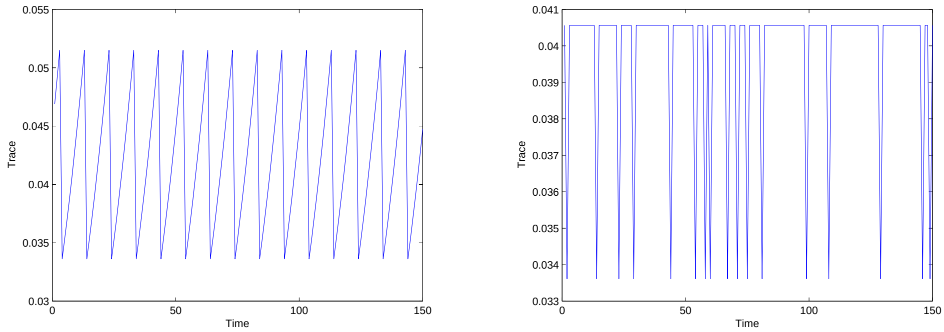

Fig. 2. Trace of the covariance matrix used in the fusion algorithm for PET (a) and DDET (b) in a single run and for a given sensor.

to the transmission strategy (i.e., deciding when transmitting an estimate or a measurement) and, thus, data-driven strategies can significantly improve performance.

7 Conclusions

The paper has addressed the state estimation problem in a centralized sensor network assuming that: (1) estimates are required at a distant location from the sensors connected via a communication link; (2) a limitation on the communication rate of each sensor is imposed; (3) each sensor node has enough processing capability to compute local state estimates. Data-driven strategies for deciding which data transmit have been investigated. Condition ensuring the mean square stability of the estimation error dynamics at the fusion node have been derived and novel data-driven transmission strategies have been proposed. Simulation results, concerning tracking of a moving object, have been presented to show the performance improvement with respect to periodic transmission strategies. The extensions of the proposed results in order to deal with nonlinear sensors, exploiting nonlinear observability decompositions, as well as distributed sensor networks (Carli *et al.*, 2008; Spanos and Murray, 2005) are currently under development.

References

Anderson, B. D. O. and J. B. Moore (1981). 'Detectability and stabilizability of time-varying discrete-time linear systems'. *SIAM Journal on Control and Optimization* **19**(1), 20–32.

Åström, K. and B. Bernhardsson (2002). Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In 'Proc. of the 41th Conf. on Decision and Control'. pp. 2011–2016.

Bar-Shalom, Y., X.R. Li and T. Kirubarajan (2001). *Estimation with Applications to Tracking and Navigation*. Wiley.

Battistelli, G., A. Benavoli and L. Chisci (2008). State estimation with a remote sensor under limited communication rate. In 'Proc. of the 3rd International Symposium on Communications, Control and Signal Processing'. pp. 654–659.

Battistelli, G., A. Benavoli and L. Chisci (2009). State estimation in a centralized sensor network under limited communication rate. In 'Proc. of the 2009 European Control Conference'.

Carli, R., A. Chiuso, L. Schenato and S. Zampieri (2008). 'Distributed Kalman filtering based on consensus strategies'. *IEEE Journal on Selected Areas in Communications* **26**, 622–633.

Feeney, L. M. and M. Nilsson (2001). Investigating the energy consumption of a wireless network interface in an ad hoc networking environment. In 'Proc. IEEE Infocom'. pp. 1548–1557.

Gupta, V., B. Hassibi and M. Murray (2007). 'Optimal LQG control across packet-dropping links'. *Systems & Control Letters* **56**, 439–446.

Hespanha, J. P., P. Naghshtabrizi and Y. Xu (2007). 'A survey of recent results in networked control systems'. *Proc. of IEEE, Special Issue on Technology of Networked Control Systems* **95**, 138–162.

Huang, M. and S. Dey (2007). 'Stability of kalman filtering with markovian packet losses'. *Automatica* **43**(4), 598–607.

Julier, S. and J.K. Uhlmann (1997). A non-divergent estimation algorithm in the presence of unknown correlations. In 'Proceedings of the 1997 American Control Conference'. Vol. 4. pp. 2369–2373.

Li, X. and W. S. Wong (1996). 'State estimation with communication constraints'. *Systems & Control Letters* **28**, 49–54.

Li, X. R., Y. M. Zhu, J. Wang and C. Z. Han (2003). 'Optimal linear estimation fusion-part I: unified fusion rules'. *IEEE Transactions on Information Theory* **49**, 2192–2208.

Spanos, D. P. and R. M. Murray (2005). Distributed sensor fusion using dynamic consensus. In 'Proc. of the 16th IFAC World Congress'.

Suh, Y. S., V. H. Nguyen and Y. S. Ro (2007). 'Modified Kalman filter for networked monitoring systems employing a send-on-delta method'. *Automatica* **43**, 332–338.

Wong, W. S. and R. W. Brockett (1997). 'Systems with finite communication bandwidth constraints - Part I: state estimation problem'. *IEEE Transactions on Automatic Control* **42**, 1294–1299.

Xu, Y. and J. P. Hespanha (2005). Estimation under uncontrolled and controlled communications in networked control systems. In 'Proc. of the 44th IEEE Conference on Decision and Control'. pp. 842–847.

Yook, J. K., D. M. Tilbury and N. R. Soparkar (2002). 'Trading computation for bandwidth: Reducing communication in distributed control systems using state estimators'. *IEEE Transactions on Control Systems Technology* **10**, 503–518.

Zamir, R. and M. Feder (1995). 'Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization'. *IEEE Transactions on Information Theory* **41**(1), 141–154.