

Distributed Moving Horizon Estimation for Linear Constrained Systems

Marcello Farina, Giancarlo Ferrari-Trecate and Riccardo Scattolini

Abstract—This paper presents a novel distributed estimation algorithm based on the concept of moving horizon estimation. Under weak observability conditions we prove convergence of the state estimates computed by any sensors to the correct state even when constraints on noise and state variables are taken into account in the estimation process. Simulation examples are provided in order to show the main features of the proposed method.

Index Terms—Distributed estimation, moving horizon estimation, consensus algorithms.

I. INTRODUCTION

A sensor network consists of a set of electronic devices, with sensing and computational capabilities, which coordinate their activity through a communication network. They can be employed in wide range of applications, such as monitoring, exploration, surveillance or to track targets over specific regions. The diffusion of sensor networks is partly due to the recent developments in wireless communications and to the availability of low cost devices. Many theoretical and technological challenges have still to be tackled in order to fully exploit their potentialities. Among the open problems, the use of sensor networks for distributed state estimation is of paramount importance. The problem can be described as follows. Assume that any sensors of the network measures some variables, computes a local estimate of the overall state of the system under monitoring, and transmits to its neighbors both the measured values and the computed state estimation. Then, the main challenge is to provide a methodology which guarantees that all the sensor asymptotically reach a common reliable estimate of the state variables, i.e. the local estimates reach a *consensus*. This goal must be achieved even if the measurements performed by any sensor are not sufficient to guarantee observability of the process state (i.e., *local observability*), provided that all the sensors, if put together, guarantee such property (i.e., *collective observability*). The transmission of measurements and of estimates among the sensors must lead to the twofold advantage of enhancing the property of observability of the sensors and of reducing the uncertainty of state estimates computed by each node.

Manuscript received March 18, 2009. This work has been supported by the European 7th framework STREP project “Hierarchical and distributed model predictive control (HD-MPC)”, contract number INFOS-ICT-223854.

M. Farina and R. Scattolini are with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio 34/5, 20133, Milan, Italy {farina,riccardo.scattolini}@elet.polimi.it

G. Ferrari-Trecate is with the Dipartimento di Informatica e Sistemistica, Università degli Studi di Pavia, Via Ferrata 1, 27100 Pavia, Italy {giancarlo.ferrari}@unipv.it

Consensus algorithms for distributed state estimation based on Kalman filters have recently been proposed in [1], [2], [3], [4], [5], [6], [7]. In particular, in [3], [4], [5], *consensus on measurements* is used to reduce their uncertainty and Kalman filters are applied by each agent. In [6], three algorithms for distributed filtering are proposed. The first algorithm is similar to the one described in [4], save for the fact that sensors exploit only partial measurements of the state vector. The second approach relies on communicating the state estimates among neighboring agents (*consensus on estimates*). The third algorithm, named *iterative Kalman consensus filter*, is based on the discrete-time version of a continuous-time Kalman filter plus a *consensus step* on the state estimates, which is proved to be stable. However, stability has not been proved for the discrete-time version of the algorithm and optimality of the estimates has not been addressed. Recently, convergence in mean of the local state estimates obtained with the algorithm presented in [4] has been proved in [7], provided that the observed process is stable.

In [2] consensus on the estimates is used together with Kalman filters. The weights of the sensors’ estimates in the consensus step and the Kalman gain are optimized in order to minimize the estimation error covariance. A two-step procedure is also used in [1], where the considered observed signal is a *random walk*. A two-step algorithm is proposed, where filtering and consensus are performed subsequently, and the estimation error is minimized with respect to both the observer gain and the consensus weights. This guarantees optimality of the solution.

More in general, the issue of distributed sensor fusion has been widely studied in the past years e.g., [8], [9]. The paper [8] provides an algorithm accounting for dynamically changing interconnections among sensors, unreliable communication links, and faults, where convergence of the estimates to the true values is proved, under suitable hypothesis of “dynamical” graph connectivity, while in [9] the authors propose a minimum variance estimator for distributed tracking of a noisy time-varying signal.

Other studies focused on the design of decentralized Kalman filters based on system decomposition. Different solutions can be classified according to the model used by each subsystem for state-estimation purposes and the topology of the communication network among subsystems. Early works, e.g. [10], [11] require all-to-all communication and assume each subsystem has full knowledge of the whole dynamics. Subsystems with overlapping states are also studied, e.g. in [12], where a fully decentralized scheme is presented.

In this paper we propose a distributed algorithm based on the concept of Moving Horizon Estimation (MHE), which has

been proposed for discrete-time linear [13], [14], nonlinear [15], [16], [17], [18] and hybrid systems [19]. This approach has many advantages; first of all, the observer displays optimality properties, since a suitable minimization problem must be solved on-line at each time instant. Furthermore, we prove that, under weak observability conditions, convergence of the state estimate is guaranteed in a deterministic framework. Finally, constraints on the noise and on the state are taken into account, as it is common in receding horizon approaches in control and estimation, [20].

The paper is structured as follows. In Section II we introduce the observed dynamical system, the structure of the sensor network, and we define a number of observability properties. In Section III we describe the distributed state estimation algorithm. In Section IV we investigate the convergence properties of the algorithm, and in Section V we discuss on how to select the design parameters in order to guarantee the applicability of the main results. In Section VI we present a simulation example, while Section VII reports some concluding remarks. For the sake of clarity, the proofs are reported in the Appendix.

Notation. I_n and $\mathbf{0}_{\nu \times \mu}$ denote the $n \times n$ identity matrix and the $\nu \times \mu$ matrix of zero elements, respectively. Given a set \mathcal{S} , $|\mathcal{S}|$ denotes its cardinality. The notation $\|z\|_S^2$ stands for $z^T S z$, where S is a symmetric positive-semidefinite matrix. The symbol \otimes denotes the Kronecker product, and $\mathbb{1}_M$ is the M -dimensional column vector whose entries are all equal to 1. The matrix $\text{diag}(\eta_1, \dots, \eta_s)$ is block-diagonal with blocks η_i . Finally, we use the short-hand $\mathbf{v} = (v_1, \dots, v_s)$ to denote a column vector with s (not necessarily scalar) components.

II. SYSTEM AND SENSOR NETWORK

We assume that the observed process obeys to the linear dynamics

$$x_{t+1} = A x_t + w_t, \quad (1)$$

where $x_t \in \mathbb{X} \subseteq \mathbb{R}^n$ is the state vector and the term $w_t \in \mathbb{W} \subseteq \mathbb{R}^n$ represents a white noise with covariance equal to Q . We assume that the sets \mathbb{X} and \mathbb{W} are convex and contain the origin. The initial condition $x_0 \in \mathbb{X}$ is a random variable with mean μ and covariance Π_0 . The pair (A, \sqrt{Q}) is stabilizable. Measurements on the state vector are performed by M sensors, according to the sensing model (in general different from sensor to sensor)

$$y_t^i = C^i x_t + v_t^i, \quad i = 1, \dots, M \quad (2)$$

where the term $v_t^i \in \mathbb{R}^{p_i}$ represents white noise with covariance equal to R_i .

The communication network among sensors is described by the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes in $\mathcal{V} = \{1, 2, \dots, M\}$ represent the sensors and the edge (j, i) in the set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ models that sensor j can transmit information to sensor i . We assume $(i, i) \in \mathcal{E}, \forall i \in \mathcal{V}$. We denote with \mathcal{V}_i^k the set of k -th order neighbors to node i , i.e., $\mathcal{V}_i^k = \{j \in \mathcal{V} : \text{there is a path of length at most } k \text{ from } j \text{ to } i \text{ in } \mathcal{G}\}$. We will also use the shorthand $\mathcal{V}_i = \mathcal{V}_i^1$.

We introduce now the definition of *isolated* subgraph. If the graph \mathcal{G} is not strongly connected (i.e., it is *reducible*),

one can partition \mathcal{G} into l nonempty irreducible subgraphs $\mathcal{G}_i = (\mathcal{N}_i, \mathcal{A}_i), i = 1, \dots, l$ (see e.g. [21]). If, for all $p \in \mathcal{N}_i, q \in \mathcal{V}_p$ implies that $q \in \mathcal{N}_i$ we say that \mathcal{G}_i is *isolated*. Remark that if \mathcal{G} is strongly connected, it is also isolated.

We associate to the graph \mathcal{G} the stochastic matrix $K \in \mathbb{R}^{M \times M}$, with entries

$$k_{ij} \geq 0 \text{ if } (j, i) \in \mathcal{E} \quad (3a)$$

$$k_{ij} = 0 \text{ otherwise} \quad (3b)$$

$$\sum_{j=1}^M k_{ij} = 1, \quad \forall i = 1, \dots, M \quad (3c)$$

Any matrix K with entries satisfying (3) is said to be compatible with \mathcal{G} . Given a graph \mathcal{G} , there are many degrees of freedom for the choice of K , which will be exploited to guarantee the convergence of the state estimator described in the following and/or to reduce the uncertainty of the estimates.

It is assumed that, at a generic time instant t , sensor i collects the measurements produced by itself and its neighboring sensors. Moreover, each sensor transmits and receives information once within a sampling interval. This means that measurements available to node i are y_t^j , with $j \in \mathcal{V}_i$.

Three types of quantities can be distinguished: *local*, *regional*, and *collective*. Specifically, a quantity is *local* (with respect to sensor i) when it is related to the node i solely. A quantity is *regional* (with respect to sensor i) if it is related to the nodes in \mathcal{V}_i . Finally, a quantity is *collective*, if it is related to the whole network. For the sake of clarity, we use different notations for local, regional and collective variables. Namely, given a variable z , z^i represents its local version, \bar{z}^i is its regional counterpart, and \mathbf{z} the collective one. For instance, we refer to y_t^i in (2) as local measurement. On the other hand, if $\mathcal{V}_i = \{j_1^i, \dots, j_{\bar{p}_i}^i\}$, the regional measurement of node i is given by

$$\bar{y}_t^i = \bar{C}^i x_t + \bar{v}_t^i \quad (4)$$

where $\bar{y}_t^i = (y_t^{j_1^i}, \dots, y_t^{j_{\bar{p}_i}^i})$, $\bar{C}^i = [(C^{j_1^i})^T \dots (C^{j_{\bar{p}_i}^i})^T]^T$, and $\bar{v}_t^i = (v_t^{j_1^i}, \dots, v_t^{j_{\bar{p}_i}^i})$. The dimension of vectors \bar{y}_t^i and \bar{v}_t^i , and the number of rows of matrix \bar{C}^i is $\bar{p}_i = \sum_{k=1}^{\bar{p}_i} p_{j_k^i}$. Furthermore, we denote by \bar{R}_i , the covariance matrix related to the regional noise \bar{v}_t^i on sensor i , i.e., $\bar{R}_i = \text{diag}(R_{j_1^i}, \dots, R_{j_{\bar{p}_i}^i})$.

According to the adopted terminology, three different observability notions can be introduced.

Definition 1: The system is *locally observable* by sensor i (sensor i is locally observable) if the pair (A, C^i) is observable. The system is *regionally observable* by sensor i (sensor i is regionally observable) if the pair (A, \bar{C}^i) is observable. The system is *collectively observable* if the pair (A, \mathbf{C}^*) is observable, where $\mathbf{C}^* = [(C^1)^T \dots (C^M)^T]^T$. \square

Notice that, for a given sensor i , local observability implies regional observability, and regional observability of any sensor implies collective observability, while all opposite implications are false. We partition the set \mathcal{V} into the subsets $\mathcal{V}_O = \{j \in \mathcal{V} : (A, \bar{C}^j) \text{ is an observable pair}\}$, $\mathcal{V}_{NO} = \{j \in \mathcal{V} : (A, \bar{C}^j) \text{ is an unobservable pair}\}$.

Given a single sensor model (1)-(2), the i -th sensor regional

observability matrix $\bar{\mathcal{O}}_n^i$ is

$$\bar{\mathcal{O}}_n^i = [(\bar{\mathcal{C}}^i)^T \quad (\bar{\mathcal{C}}^i A)^T \quad \dots \quad (\bar{\mathcal{C}}^i A^{n-1})^T]^T \quad (5)$$

Let \bar{P}_{NO}^i be the orthogonal projection matrix on $\ker(\bar{\mathcal{O}}_n^i)$, that is the regionally unobservable subspace. Similarly, let \bar{P}_O^i be the orthogonal projection on the regional observability subspace $\ker(\bar{\mathcal{O}}_n^i)^\perp$. Next, we recall how \bar{P}_O^i and \bar{P}_{NO}^i can be computed. Let $r_i = \text{rank}(\bar{\mathcal{O}}_n^i)$ and denote with $\xi_{r_i+1}, \dots, \xi_n$ an orthonormal basis of $\ker(\bar{\mathcal{O}}_n^i)$. Let also ξ_1, \dots, ξ_{r_i} be an orthonormal basis of $\ker(\bar{\mathcal{O}}_n^i)^\perp$ and define the orthonormal and non-singular matrix $\bar{T}^i = [\xi_1^i, \dots, \xi_n^i]$. Defining the matrices \bar{S}_O^i and \bar{S}_{NO}^i as

$$\bar{S}_O^i = \begin{bmatrix} I_{r_i} \\ 0_{(n-r_i) \times r_i} \end{bmatrix}, \quad \bar{S}_{NO}^i = \begin{bmatrix} 0_{r_i \times (n-r_i)} \\ I_{n-r_i} \end{bmatrix},$$

we have $\bar{P}_O^i = \bar{T}^i \bar{S}_O^i (\bar{S}_O^i)^T (\bar{T}^i)^{-1}$ and $\bar{P}_{NO}^i = \bar{T}^i \bar{S}_{NO}^i (\bar{S}_{NO}^i)^T (\bar{T}^i)^{-1}$. Furthermore, defining $\mathbf{T} = \text{diag}(\bar{T}_1, \dots, \bar{T}_M)$, $\mathbf{S}_O = \text{diag}(\bar{S}_O^1, \dots, \bar{S}_O^M)$, and $\mathbf{S}_{NO} = \text{diag}(\bar{S}_{NO}^1, \dots, \bar{S}_{NO}^M)$, the collective projection matrices are $\mathbf{P}_O = \mathbf{T} \mathbf{S}_O \mathbf{T}^T$ and $\mathbf{P}_{NO} = \mathbf{T} \mathbf{S}_{NO} \mathbf{T}^T$. Note that \bar{S}_{NO}^i is empty when the system is regionally observable by sensor i . In this case we assume that $\bar{P}_{NO}^i = \mathbf{0}_{n \times n}$.

III. THE DISTRIBUTED ESTIMATION ALGORITHM

Our aim is to design, for a generic sensor $i \in \mathcal{V}$, an algorithm for computing an estimate of the system state based on regional measurements \bar{y}_t^i and further pieces of information provided by sensors $j \in \mathcal{V}_i$. The proposed solution relies on MHE, in view of its capability to handle state and noise constraints. More specifically, we propose a Distributed MHE (DMHE) scheme where each sensor solves a MHE problem.

A. The local minimization problem

For a given estimation horizon $N \geq 1$, each node $i \in \mathcal{V}$ at time t determines the estimates \hat{x}^i and \hat{w}^i of x and w , respectively, by solving the constrained minimization problem (MHE- i)

$$\Theta_t^{*i} = \min_{\hat{x}_{t-N}^i, \{\hat{w}_k^i\}_{k=t-N}^{t-1}} J^i(t-N, t, \hat{x}_{t-N}^i, \hat{w}^i, \hat{v}^i, \Gamma_{t-N}^i) \quad (6)$$

under the constraints

$$\hat{x}_{k+1}^i = A \hat{x}_k^i + \hat{w}_k^i, \quad k = t-N, \dots, t \quad (7a)$$

$$\bar{y}_k^i = \bar{\mathcal{C}}^i \hat{x}_k^i + \hat{v}_k^i \quad (7b)$$

$$\hat{w}_k^i \in \mathbb{W} \quad (7c)$$

$$\hat{x}_k^i \in \mathbb{X} \quad (7d)$$

The local cost function J^i is given by

$$J^i(t-N, t, \hat{x}_{t-N}^i, \hat{w}^i, \hat{v}^i, \Gamma_{t-N}^i) = \frac{1}{2} \sum_{k=t-N}^t \|\hat{v}_k^i\|_{\bar{\mathbf{R}}^i}^2 + \frac{1}{2} \sum_{k=t-N}^{t-1} \|\hat{w}_k^i\|_{\mathbf{Q}^i}^2 + \Gamma_{t-N}^i(\hat{x}_{t-N}^i; \hat{x}_{t-N/t-1}^i) \quad (8)$$

We denote with $\hat{x}_{t-N/t}^i$ and with $\{\hat{w}_{k/t}^i\}_{k=t-N}^{t-1}$ the optimizers to (6) and with $\hat{x}_{k/t}^i$, $k = t-N, \dots, t$ the local state sequence stemming from $\hat{x}_{t-N/t}^i$ and $\{\hat{w}_{k/t}^i\}_{k=t-N}^{t-1}$. Furthermore

$$\hat{x}_{t-N/t-1}^i = \sum_{j=1}^M k_{ij} \hat{x}_{t-N/t-1}^j \quad (9)$$

denotes the weighted average state estimates produced by sensors $j \in \mathcal{V}_i$. In (8), the function $\Gamma_{t-N}^i(\hat{x}_{t-N}^i; \hat{x}_{t-N/t-1}^i)$ is the so called *initial penalty*, defined as follows

$$\Gamma_{t-N}^i(\hat{x}_{t-N}^i; \hat{x}_{t-N/t-1}^i) = \frac{1}{2} \|\hat{x}_{t-N}^i - \hat{x}_{t-N/t-1}^i\|_{(\Pi_{t-N/t-1}^i)^{-1}}^2 + \Theta_{t-1}^{*i} \quad (10)$$

where Θ_{t-1}^{*i} is the optimal cost defined in (6) and the positive-definite symmetric weighting matrix $\Pi_{t-N/t-1}^i$ appearing in (10) plays the role of a covariance matrix whose choice will be discussed in details in the next paragraphs. The term Θ_{t-1}^{*i} is a constant in (10) and could be neglected when solving (6). However, since it plays a major role in establishing the main convergence properties of DMHE, it is here maintained for clarity of presentation.

Note that, in view of the definition of k_{ij} in (3), $\Gamma^i(\cdot)$ depends only upon regional quantities and, since also the cost (8) and the constraints (7) depend only upon regional variables, the overall estimation scheme is decentralized. Finally, notice that $\Gamma^i(\cdot)$ embodies a *consensus-on-estimates* term, in the sense that it penalizes deviations of $\hat{x}_{t-N/t-1}^i$ from $\hat{x}_{t-N/t-1}^i$. Consensus, besides increasing accuracy of the local estimates, is fundamental to guarantee convergence of the state estimates to the state of the observed system even if regional observability does not hold. In other words, it allows sensor i to reconstruct components of the state that cannot be estimated by the i -th regional model.

B. The collective minimization problem

The local estimation problems (6)-(10) can be given a collective form more suitable for the following developments. To this end, let \mathbf{J} be the collective cost function given by

$$\mathbf{J}(\cdot) = \sum_{i=1}^M J^i(t-N, t, \hat{x}_{t-N}^i, \hat{w}^i, \hat{v}^i, \Gamma_{t-N}^i) \quad (11)$$

Define the collective vectors $\hat{\mathbf{x}}_t = (\hat{x}_t^1, \dots, \hat{x}_t^M)$, $\hat{\mathbf{v}}_t = (\hat{v}_t^1, \dots, \hat{v}_t^M)$, $\hat{\mathbf{w}}_t = (\hat{w}_t^1, \dots, \hat{w}_t^M)$, the quantities $\Theta_{t-1}^{*i} = \sum_{i=1}^M \Theta_{t-1}^{*i}$, $\mathbf{K} = K \otimes I_n$,

$$\mathbf{\Pi}_{t_1/t_2} = \text{diag}(\Pi_{t_1/t_2}^1, \dots, \Pi_{t_1/t_2}^M) \quad (12)$$

and the collective initial penalty

$$\Gamma_{t-N}(\hat{\mathbf{x}}_{t-N}; \hat{\mathbf{x}}_{t-N/t-1}) = \Gamma_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) + \Theta_{t-1}^* \quad (13)$$

where $\Gamma_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) = \frac{1}{2} \|\hat{\mathbf{x}}_{t-N} - \mathbf{K} \hat{\mathbf{x}}_{t-N/t-1}\|_{\mathbf{\Pi}_{t-N/t-1}^{-1}}^2$. Then, using the matrices $\bar{\mathbf{R}} = \text{diag}(\bar{\mathbf{R}}^1, \dots, \bar{\mathbf{R}}^M)$, $\mathbf{Q} = \text{diag}(\mathbf{Q}^1, \dots, \mathbf{Q}^M) \in \mathbb{R}^{nM \times nM}$, the collective cost function $\mathbf{J}(\cdot)$ can be rewritten as

$$\mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\mathbf{v}}, \Gamma_{t-N}) = \frac{1}{2} \sum_{k=t-N}^t \|\hat{\mathbf{v}}_k\|_{\bar{\mathbf{R}}}^2 + \frac{1}{2} \sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_k\|_{\mathbf{Q}}^2 + \Gamma_{t-N}(\hat{\mathbf{x}}_{t-N}; \hat{\mathbf{x}}_{t-N/t-1}) \quad (14)$$

Defining $\mathbf{A} = \text{diag}(A, \dots, A) \in \mathbb{R}^{nM \times nM}$ and $\bar{\mathbf{C}} = \text{diag}(\bar{\mathcal{C}}^1, \dots, \bar{\mathcal{C}}^M)$, also the constraints (7) can be written in the following collective form

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A} \hat{\mathbf{x}}_k + \hat{\mathbf{w}}_k, \quad k = t-N, \dots, t \quad (15a)$$

$$\bar{\mathbf{y}}_k = \bar{\mathbf{C}} \hat{\mathbf{x}}_k + \hat{\mathbf{v}}_k \quad (15b)$$

$$\hat{\mathbf{w}}_k \in \mathbb{W}^M \quad (15c)$$

$$\hat{\mathbf{x}}_k \in \mathbb{X}^M \quad (15d)$$

It is important to note that solving the problem

$$\Theta_t^* = \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \left\{ \mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\mathbf{v}}, \mathbf{\Gamma}_{t-N}) \text{ subject to (15)} \right\} \quad (16)$$

is equivalent to solve the *MHE-i* problems (6), in the sense that $\hat{\mathbf{x}}_{t-N/t}^i, \{\hat{\mathbf{w}}_{k/t}^i\}_{k=t-N}^{t-1}$ is a solution to (6) if and only if $\hat{\mathbf{x}}_{t-N/t}, \{\hat{\mathbf{w}}_{k/t}\}_{k=t-N}^{t-1}$ is a solution to (16), where $\hat{\mathbf{w}}_{k/t} = (\hat{\mathbf{w}}_{k/t}^1, \dots, \hat{\mathbf{w}}_{k/t}^M)$.

Let t_1 verify $t-N \leq t_1 \leq t$. We define the *transit cost* of a generic state $\mathbf{z} \in \mathbb{R}^{nM}$, computed at instant t as

$$\mathfrak{E}_{t_1/t}(\mathbf{z}) = \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \left\{ \mathbf{J}(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\mathbf{v}}, \mathbf{\Gamma}_{t-N}) \text{ subject to (15) and } \hat{\mathbf{x}}_{t_1} = \mathbf{z} \right\} \quad (17)$$

As discussed in [20], $\mathfrak{E}_{t_1/t}(\mathbf{z})$ provides a measure of the likelihood that $\hat{\mathbf{x}}_{t_1}$ is equal to \mathbf{z} given the data $\bar{\mathbf{y}}_k, k=t-N, \dots, t$ and the prior likelihood $\mathbf{\Gamma}_{t-N}(\cdot)$ on \mathbf{x}_{t-N} . Specifically, the lower $\mathfrak{E}_{t_1/t}(\mathbf{z})$, the more likely the equality $\hat{\mathbf{x}}_{t_1} = \mathbf{z}$. The prior $\mathbf{\Gamma}_{t-N}(\cdot)$ can be interpreted as an approximation of $\mathfrak{E}_{t-N/t-1}(\cdot)$. The key condition involving these two terms, that will be fundamental for proving convergence of DMHE (see Appendix B), is that, for all $z \in \mathbb{X}$

$$\mathbf{\Gamma}_{t-N}(\mathbb{1}_M \otimes z; \hat{\mathbf{x}}_{t-N/t-1}) \leq \mathfrak{E}_{t-N/t-1}(\mathbb{1}_M \otimes z) \quad (18)$$

Equation (18) is similar to the assumption (C2) in [15] for centralized MHE. However, in (18), the transit cost instead of the arrival cost appears. In fact $\mathfrak{E}_{t-N/t-1}$ is a smoothing term, since it takes into account data up to time t , in order to enforce consensus (in [13] this approach is called *smoothing update*). An explicit formula for a lower bound to $\mathfrak{E}_{t-N/t-1}(\mathbf{z})$ (which coincides with $\mathfrak{E}_{t-N/t-1}(\mathbf{z})$ for unconstrained estimation problems, see the proof of Lemma 1 in Appendix B) is given by a quadratic cost function, i.e.

$$\mathfrak{E}_{t-N/t-1}(\mathbf{z}) \geq \frac{1}{2} \|\mathbf{z} - \hat{\mathbf{x}}_{t-N/t-1}\|_{(\tilde{\mathbf{\Pi}}_{t-N/t-1})^{-1}}^2 + \Theta_{t-1}^* \quad (19)$$

for a suitable choice of $\tilde{\mathbf{\Pi}}_{t-N/t-1}$. The computation of $\tilde{\mathbf{\Pi}}_{t-N/t-1}$, and a procedure for updating the matrix $\mathbf{\Pi}_{t-N/t-1}$ in (13) satisfying (18) are given in the next section.

C. Update of the weighting matrices

As remarked in the previous section, the first step for updating matrices $\mathbf{\Pi}_{t-N/t-1}^i$, is to compute $\tilde{\mathbf{\Pi}}_{t-N/t-1}$ in (19), with the following diagonal structure

$$\tilde{\mathbf{\Pi}}_{t-N/t-1} = \text{diag} \left(\tilde{\mathbf{\Pi}}_{t-N/t-1}^1, \dots, \tilde{\mathbf{\Pi}}_{t-N/t-1}^M \right) \quad (20)$$

where the update of $\tilde{\mathbf{\Pi}}_{t-N/t-1}^i$ is carried out by the sensor i , based on regional pieces of information. For this reason, this step is denoted *regional weights update*. Specifically, the matrix $\tilde{\mathbf{\Pi}}_{t-N/t-1}^i, i \in \mathcal{V}$, is given by one iteration of the difference Riccati equation associated to a Kalman filter for the system

$$\begin{cases} x_{t-N} &= Ax_{t-N-1} + w_{t-N-1} \\ \bar{z}_{t-N} &= \bar{\mathcal{O}}_N^i x_{t-N} + \bar{V}_{t-N}^i \end{cases}$$

where matrix $\bar{\mathcal{O}}_N^i$ is defined in (5). If we define

$$\mathcal{C}_N^i = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \bar{C}^i & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{C}^i A^{N-2} & \bar{C}^i A^{N-3} & \dots & \bar{C}^i \end{bmatrix} \in \mathbb{R}^{\bar{p}_i N \times n(N-1)} \quad (21)$$

$$\bar{R}_N^i = \text{diag}(\bar{R}^i, \dots, \bar{R}^i) \in \mathbb{R}^{\bar{p}_i N \times \bar{p}_i N} \quad (22)$$

$$Q_{N-1} = \text{diag}(Q, \dots, Q) \in \mathbb{R}^{n(N-1) \times n(N-1)} \quad (23)$$

$$\text{Cov}[w_t] = Q \quad (24)$$

$$\text{Cov}[\bar{V}_t^i] = \bar{R}_N^i = \bar{R}_N^i + \mathcal{C}_N^i Q_{N-1} (\mathcal{C}_N^i)^T \quad (25)$$

and set the covariance of the estimate $\hat{\mathbf{x}}_{t-N-1}^i$ as

$$\mathbf{\Pi}_{t-N-1/t-2}^{*i} = ((\mathbf{\Pi}_{t-N-1/t-2}^i)^{-1} + (\bar{C}^i)^T (\bar{R}^i)^{-1} \bar{C}^i)^{-1} \quad (26)$$

the resulting Riccati recursion is given by

$$\begin{aligned} \tilde{\mathbf{\Pi}}_{t-N/t-1}^i &= \mathcal{R}^i \left(\mathbf{\Pi}_{t-N-1/t-2}^{*i}; Q, \bar{R}_N^{*i} \right) \\ &= A \mathbf{\Pi}_{t-N-1/t-2}^{*i} A^T + Q - A \mathbf{\Pi}_{t-N-1/t-2}^{*i} (\bar{\mathcal{O}}_N^i)^T \times \\ &\quad \times \left(\bar{\mathcal{O}}_N^i \mathbf{\Pi}_{t-N-1/t-2}^{*i} (\bar{\mathcal{O}}_N^i)^T + \bar{R}_N^{*i} \right)^{-1} \bar{\mathcal{O}}_N^i \mathbf{\Pi}_{t-N-1/t-2}^{*i} A^T \end{aligned} \quad (27)$$

Once the matrices $\tilde{\mathbf{\Pi}}_{t-N/t-1}^i$ have been computed, we perform a *consensus weights update*, in order to compute the matrices $\mathbf{\Pi}_{t-N/t-1}^i$ appearing in (10), which must satisfy the fundamental inequality (18). As stated in Lemma 1 in Appendix B, (18) is verified if $\mathbf{\Pi}_{t-N/t-1}$ fulfills the Linear Matrix Inequality (LMI)

$$\mathbf{\Pi}_{t-N/t-1} \geq \mathbf{K} \tilde{\mathbf{\Pi}}_{t-N/t-1} \mathbf{K}^T \quad (28)$$

The LMI (28) deserves a few comments. In order to make the initial penalty $\mathbf{\Gamma}_{t-N}(\cdot)$ a good approximation of the transit cost $\mathfrak{E}_{t-N/t-1}(\cdot)$, one would require the matrix $\mathbf{\Pi}_{t-N/t-1}$ to be “as close as possible” to $\mathbf{K} \tilde{\mathbf{\Pi}}_{t-N/t-1} \mathbf{K}^T$. Therefore, in our case, one would make the matrix $\mathbf{\Pi}_{t-N/t-1}$ “as small as possible”, subject to the constraint (28). A way for achieving this is to solve the LMI problem

$$\min \left(\text{trace}(\mathbf{\Pi}_{t-N/t-1}) \right), \text{ subject to (28),} \quad (29)$$

where $\mathbf{\Pi}_{t-N/t-1}$ has the block-diagonal structure (12). Notice that (29) could be solved by each sensor since, similarly to the formula for updating covariances in Kalman filtering, the computation of $\mathbf{\Pi}_{t-N/t-1}^i$ does not depend upon the collected measurements. However, problem (29) has a centralized flavor. This limitation is severe since, for instance, the LMI (28) has size $n \times M$ which implies that the computational burden for solving (29) scales with the number of sensors, hence hampering the application of DMHE to large networks. The next proposition provides a way to circumvent this problem.

Proposition 1: The matrices $\mathbf{\Pi}_{t-N/t-1}^i$ which satisfy, $\forall i \in \mathcal{V}$

$$\mathbf{\Pi}_{t-N/t-1}^i \geq 2 \sum_{j=1}^M k_{ij}^2 \tilde{\mathbf{\Pi}}_{t-N/t-1}^j \quad (30)$$

also satisfy the LMI (28).

Proof: See Appendix A. ■

Notice that, in the solution provided by Proposition 1, each node computes $\Pi_{t-N/t-1}^i$ solely on the basis of $\tilde{\Pi}_{t-N/t-1}^j$, provided by its neighbors, $j \in \mathcal{V}_i$. In view of this, the LMI (28) can be solved in a decentralized fashion by setting

$$\Pi_{t-N/t-1}^i = 2 \sum_{j=1}^M k_{ij}^2 \tilde{\Pi}_{t-N/t-1}^j \quad (31)$$

D. DMHE algorithm

In the following we sketch the steps that have to be carried out, in practice, in order to apply the proposed DMHE algorithm.

- Initialization: at $t = 0$ all nodes store the matrix Π_0 and the estimate $\hat{x}_{0/0} = \mu$ of x_0 , where μ is given. Recall that Π_0 is the covariance matrix related to the initial condition x_0
- if $1 \leq t \leq N$, the estimation horizon N is reduced to $\tilde{N} = t$ and node $i \in \mathcal{V}$ performs the following steps
 - compute $\Pi_{t-\tilde{N}/t-1}^i = \Pi_{0/t-1}^i$ from Π_0 according to (31), for all $i \in \mathcal{V}$,
 - solve the problem *MHE-i*, with initial penalty

$$\Gamma_{t-\tilde{N}}^i = \frac{1}{2} \|\hat{x}_{0/t-1}^i - \hat{x}_{0/t-1}^i\|_{(\Pi_{0/t-1}^i)^{-1}}^2$$

- if $t > N$, at each time instant, every node $i \in \mathcal{V}$,
 - computes $\Pi_{t-N/t-1}^i$ from $\Pi_{t-N-1/t-2}^i$ according to (26), (27) and (31),
 - solves the problem *MHE-i*, with initial penalty

$$\Gamma_{t-N}^i = \frac{1}{2} \|\hat{x}_{t-N}^i - \hat{x}_{t-N/t-1}^i\|_{(\Pi_{t-N/t-1}^i)^{-1}}^2$$

IV. CONVERGENCE PROPERTIES OF DMHE

The main purpose of this section is to extend the convergence results of [13] for centralized MHE to the proposed DMHE scheme.

Definition 2: Let Σ be system (1) with $w = 0$ and denote by $x_\Sigma(t, x_0)$ the state reached by Σ at time t starting from initial condition x_0 . Assume that the trajectory $x_\Sigma(t, x_0)$ is feasible, i.e., $x_\Sigma(t, x_0) \in \mathbb{X}$ for all t . DMHE is *convergent* if $\|\hat{x}_{t/t}^i - x_\Sigma(t, x_0)\| \xrightarrow{t \rightarrow \infty} 0$ for all $i \in \mathcal{V}$. \square

Note that, as in [13], convergence is defined assuming that the model generating the data is noiseless, but the possible presence of noise is taken into account in the state estimation algorithm. Now, defining the collective vector $\mathbf{x}_\Sigma(t, x_0) = \mathbf{1}_M \otimes x_\Sigma(t, x_0)$ and $\boldsymbol{\varepsilon}_{k/t} = \hat{\mathbf{x}}_{k/t} - \mathbf{x}_\Sigma(k, x_0)$, the following result can be stated.

Theorem 1: If: (i) matrices $\Pi_{t-N/t-1}^i$ are computed according to (26), (27) and (28), (ii) $\Pi_{t-N/t-1}^i$ are bounded for all t , and for all $i \in \mathcal{V}$, (iii) $N \geq n - 1$ and $N \geq 1$, then

- there exists an asymptotically vanishing sequence α_t (i.e., $\|\alpha_t\| \xrightarrow{t \rightarrow \infty} 0$) such that the dynamics of the state estimation error provided by the DMHE scheme is given by

$$\boldsymbol{\varepsilon}_{t-N/t} = \Phi \boldsymbol{\varepsilon}_{t-N-1/t-1} + \alpha_t \quad (32)$$

where $\Phi = \mathbf{P}_{NO} \mathbf{KAP}_{NO}$;

- if (iv) Φ is Schur, then DMHE is convergent.

Proof: See Appendix B. \blacksquare

In Section V we will provide a method to choose a matrix K compatible with \mathcal{G} such that conditions (ii) and (iv) of Theorem 1 are satisfied.

We highlight that Condition (iv) does not require the asymptotic stability of system (1). Moreover, Theorem 1 does not hinge on observability properties. In fact, convergence of the estimation error can be achieved even if a weaker detectability property is satisfied, i.e. if matrix Φ inherits only stable eigenvalues of A . However, it is of interest to determine conditions guaranteeing that the matrix Φ does not inherit any (non-zero) eigenvalue of A . The reason is twofold. First, this is tantamount to requiring that the unobservable dynamics of all regional systems are affected by the communication network. Second, the study is a preliminary step towards the goal of choosing K and, when possible, the network topology, in order to assign the eigenvalues of Φ at will. Let λ_A^i and v_A^i be the eigenvalues and the eigenvectors of A , respectively, with $i = 1, \dots, n$. Then, the eigenvalues of \mathbf{A} are λ_A^i ($i = 1, \dots, n$), each one with multiplicity M . Moreover, denoting by e_j , $j = 1, \dots, M$ the canonical basis vectors of \mathbb{R}^M , the eigenspace related to λ_A^i is $\text{span}(e_1 \otimes v_A^i, \dots, e_M \otimes v_A^i)$. In view of the previous discussion, we want to investigate the following property.

Property 1: If λ_i is a non-zero eigenvalue of A , for all $\mathbf{x} \in \text{span}(e_1 \otimes v_A^i, \dots, e_M \otimes v_A^i)$, λ_A^i and \mathbf{x} are not an eigenvalue/eigenvector pair for Φ . \square

Conditions guaranteeing that Property 1 holds are given in the following Theorem, which is illustrated in Fig. 1.

Theorem 2: Consider a partition of \mathcal{G} into the irreducible subgraphs $\mathcal{G}_i = (\mathcal{N}_i, \mathcal{A}_i)$, $i = 1, \dots, l$. If for all the isolated strongly connected subgraphs \mathcal{G}_i it holds

$$\bigcap_{j \in \mathcal{N}_i} \ker(\bar{\theta}_n^j) = 0 \quad (33)$$

then Property 1 is verified.

Proof: See Appendix C. \blacksquare

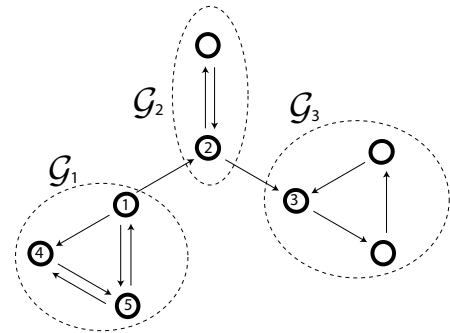


Fig. 1. The graph is decomposed into three connected subgraphs \mathcal{G}_1 , \mathcal{G}_2 and \mathcal{G}_3 . Notice that the node 2 of \mathcal{G}_2 is a neighbor to node 3 of \mathcal{G}_3 . Therefore, graph \mathcal{G}_3 is not isolated. Analogously, the subgraph \mathcal{G}_2 is not isolated, while the subgraph \mathcal{G}_1 is isolated. Condition (33) states that collective observability is required for the subgraph \mathcal{G}_1 , i.e., the pair $(A, \mathbf{C}_{\mathcal{G}_1}^*)$ is observable, where $\mathbf{C}_{\mathcal{G}_1}^* = [(C^1)^T (C^4)^T (C^5)^T]^T$.

In the case of strongly connected graphs we have the following result.

Corollary 1: If \mathcal{G} is strongly connected and the system is collectively observable, then Property 1 is verified.

Proof: See Appendix C. ■

As a trivial case, assume that all sensors are regionally observable and arranged in a strongly connected graph \mathcal{G} . This yields $\mathbf{P}_{NO} = \Phi = \mathbf{0}_{nM \times nM}$ and convergence of DMHE follows from Theorem 1. Moreover, Property 1 trivially holds.

V. SELECTION OF THE DESIGN PARAMETER K

The key assumption of Theorem 1 that the sequence $\{\prod_{t=N/t-1}^{\infty}\}$ is bounded is not a-priori guaranteed by formula (31). However, under weak assumptions, boundedness of $\{\prod_{t=N/t-1}^{\infty}\}$ can be enforced by properly choosing the entries k_{ij} , $\forall (i, j) \in \mathcal{E}$ of K . Interestingly, we will also prove that the proposed choice of K results in assigning all the eigenvalues of Φ equal to zero, that guarantees convergence of DMHE and Property 1.

Theorem 3: If \mathcal{V}_O is non-empty and, for all $i \in \mathcal{V}_{NO}$, there exists $k > 0$ such that $\mathcal{V}_i^k \cap \mathcal{V}_O \neq \emptyset$, then there exists K , compatible with \mathcal{G} , such that matrices $\prod_{t=N/t-1}^i$ ($i = 1, \dots, M$), resulting from (27) and (31) are bounded for all $i \in \mathcal{V}$.

Proof: See Appendix D. ■

The assumption of Theorem 3 that, for each node in \mathcal{V}_{NO} , there exists an incoming directed path stemming from a node in \mathcal{V}_O requires that at least one sensor is regionally observable. This condition, although not necessary to guarantee the existence of a suitable K , allows one to identify at least a “reference” node, which provides reliable estimates even without communication, see the proof of Theorem 3 in Appendix D. The proof of Theorem 3 is constructive and leads to the following algorithm for computing the matrix K .

- Algorithm 1:* 1) for each $i \in \mathcal{V}_O$, set $k_{ii} = 1$;
 2) for each $i \in \mathcal{V}_{NO}$, select $k_{ii} < \frac{1}{\sqrt{2\sigma^{(i)}(A)}}$, where $\sigma^{(i)}(A) = \max\{|\lambda_j^{(i)}(A)| : \lambda_j^{(i)}(A) \text{ is an unobservable eigenvalue for the pair } (A, \bar{C}^i)\}$;
 3) for each $i \in \mathcal{V}_{NO}$ select a node $j \in \mathcal{V}_O$ and a path from j to i , in such a way that each node in the path has at most one neighbor. We denote with \mathcal{E}^* the set of edges selected in this way;
 4) for all edges $(i, j) \in \mathcal{E}^*$, choose $k_{ij} = 1 - k_{ii}$, while for all edges $(i, j) \in \mathcal{E} \setminus \mathcal{E}^*$, set $k_{ij} = 0$. □

Algorithm 1 is illustrated in Fig. 2. Given the availability of methods for computing paths with a computational complexity that scales polynomially with $|\mathcal{V}|$ [22], the overall algorithm is polynomial. Moreover, if \mathcal{G} is complete graph, Algorithm 1 provides a method for designing a not a-priori fixed communication network. Furthermore, Algorithm 1 implicitly provides a rule for connecting a new regionally observable/unobservable sensor to the network without spoiling the boundedness of the sequence $\{\prod_{t=N/t-1}^i\}_{i=0}^{\infty}$.

Finally, by selecting K according to Algorithm 1, the following result holds.

Corollary 2: Under the assumption of Theorem 3, if K is selected according to Algorithm 1, then Φ has all the eigenvalues equal to zero.

Proof: See Appendix E. ■

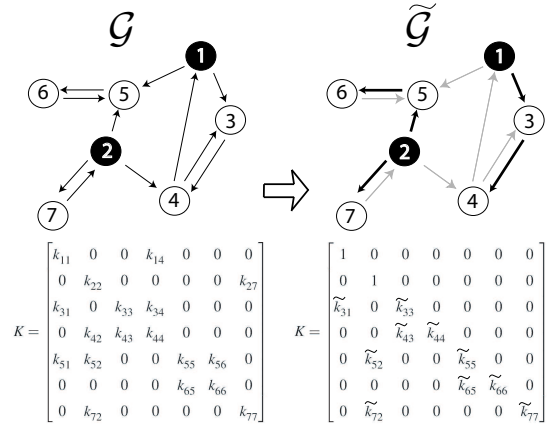


Fig. 2. The original graph \mathcal{G} (left panel) presents two types of sensors: nodes 1 and 2 are regionally observable (black circles), and nodes 3-7 are regionally unobservable (white circles). Therefore, $\mathcal{V}_O = \{1, 2\}$ and $\mathcal{V}_{NO} = \{3, 4, 5, 6, 7\}$. The graph $\tilde{\mathcal{G}}$ (right panel) is defined by selecting a subset of edges (black ones) of the original graph, according to step 2 of Algorithm 1. Below each graph we show the corresponding matrix K .

A final remark is due. Under the assumption of Theorem 3, the choice of a matrix K is not unique and details on the available degrees of freedom in the definition of a suitable K (see Remark 1 after the proof of Theorem 3) can be used to reduce the conservativeness imposed by Algorithm 1. In fact the generated matrix K is lower triangular, up to a permutation of the node indexes. However, the same arguments of the proof of Theorem 3 can be used to show that boundedness of $\prod_{t=N/t-1}^i$ is guaranteed also by any stochastic matrix \bar{K} compatible with \mathcal{G} with: (i) the same diagonal elements of the matrix K obtained with Algorithm 1; (ii) arbitrary (non a-priori zero) elements in the lower triangular part; (iii) sufficiently small (non a-priori zero) elements in the upper triangular part. Details on point (iii) are given in Remark 1 in Appendix D. This choice allows for a full exploitation of the communication links. In view of this, and the fact that connected components of the graph produced by Algorithm 1 can be linked through arcs, one expects to increase convergence rates of the estimates to a common value. Moreover, the presence of more links results in an increased reliability against communication faults.

VI. EXAMPLE

We consider the fourth-order system

$$x_{t+1} = \begin{bmatrix} 0.9962 & 0.1949 & 0 & 0 \\ -0.1949 & 0.3819 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1.21 & 1.98 \end{bmatrix} x_t + w_t \quad (34)$$

where $x_t = [x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}]^T$. Notice that the eigenvalues of the matrix A are 0.9264, 0.4517, $0.99 \pm 0.4795i$ and, since $|0.99 \pm 0.4795i| > 1$, the system is unstable.

Let $e_t \in \mathbb{R}^4$, be white noise with covariance $Q_e = \text{diag}(0.0012, 0.038, 0.0012, 0.038)$. In the following we consider two cases

- A. $w_t = e_t$, $Q = Q_e$ and $\mathbb{W} = \mathbb{R}^4$ (unconstrained input noise)
- B. $w_t = |e_t|$, $Q = Q_e$ and $\mathbb{W} = \mathbb{R}_{\geq 0}^4$ (constrained input noise)

In both cases, we set $\mu = [0\ 0\ 0\ 0]^T$, $\Pi_0 = 100I_n$ and $N = 2$ in the DMHE algorithm.

The state of (34) is measured by $M = 4$ sensors with sensing model

$$\begin{aligned} y_t^i &= [1\ 0\ 0\ 0]x_t + v_t^i & \text{if } i = 1, 2 \\ y_t^i &= [0\ 0\ 1\ 0]x_t + v_t^i & \text{if } i = 3, 4 \end{aligned}$$

where $\text{Var}(v_t^i) = R_i = 1$, $i = 1, \dots, 4$. Sensors are connected according to the graph in Fig. 3, where the matrix K is also given. It is apparent that the information available, at each

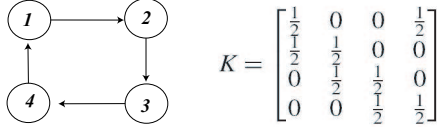
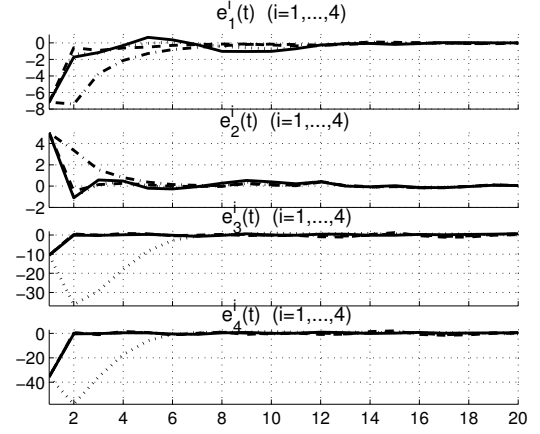


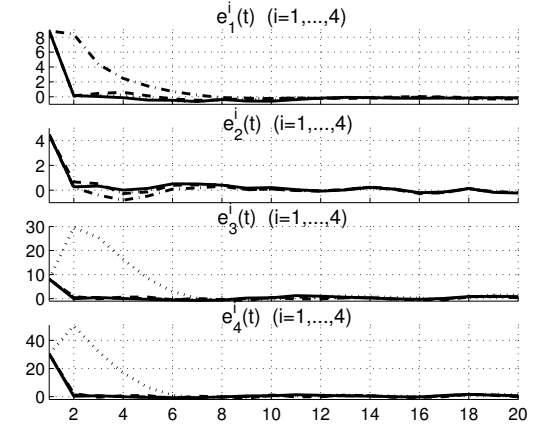
Fig. 3. Communication network and associated matrix K

instant, to node 1 consists of the measurements of $x_{1,t}$ and $x_{3,t}$ (transmitted by sensor 4). Analogously, the information available to node 3 consists of $x_{1,t}$ (transmitted by sensor 2) and $x_{3,t}$. It is easy to check that the system is regionally observable by sensors 1 and 3. On the other hand, at each time instant sensor 2 can only use two different measurements of $x_{1,t}$ (produced by sensors 1 and 2). Similarly, sensor 4 can only use two different measures of $x_{3,t}$ (produced by sensors 3 and 4). Therefore, the system is not regionally observable by sensors 2 and 4. In fact, $\bar{P}_{NO}^2 = \text{diag}(0, 0, 1, 1)$, $\bar{P}_{NO}^4 = \text{diag}(1, 1, 0, 0)$. The eigenvalues of the matrix Φ are 0, 0.4632, 0.2258 and $0.4950 \pm 0.2397i$. Since Φ is Schur, convergence of DMHE is guaranteed by Theorem 1. Moreover, since the graph is strongly connected and collective observability holds, Corollary 1 guarantees that also Property 1 holds.

In Fig. 4(a) the estimation errors produced by all sensors in the case A are shown. It is worth noticing that the estimates produced by sensors 2 [resp. 4], relative to states $x_{3,t}$, $x_{4,t}$ [resp. $x_{1,t}$, $x_{2,t}$] display big errors for $t < 6$. In fact, these states cannot be observed by these sensors using regional measurements. Nonetheless, the estimation errors of all sensors asymptotically tend to the same values, thanks to the consensus action embodied in the DMHE scheme. The estimation errors for case B are depicted in Fig. 4(b). Analogously to case A, convergence of DMHE can be noticed. Fig. 5 depicts the evolution of the eigenvalues of matrices $\Pi_{t/t+N-1}^i$ over time. Note that these matrices are the same in the cases A and B. Indeed, the update procedure described in Section III-C does not depend on the estimates and can be run off-line. Further simulation experiments have been performed (results not shown for space limitations), in order to assess the effect of the variation of the horizon length N on the estimation performances. As expected, the larger the horizon length, the more accurate the results. In fact, as N increases, a larger set of data is taken into account in the optimization problem. However, the need of increasing N for optimality reasons is conflicting with the need of reducing as much as possible the computational load.



(a) Case A



(b) Case B

Fig. 4. Components of the estimation error $[e_{1,t}^i, e_{2,t}^i, e_{3,t}^i, e_{4,t}^i]^T = x_t - \hat{x}_{t/t}^i$ of the different sensors. Solid line $i = 1$, dotted line $i = 2$, dashed line $i = 3$, dash-dotted line $i = 4$.

VII. CONCLUSIONS

We proposed a distributed moving horizon estimation algorithm for constrained discrete-time linear systems. Under suitable assumptions we proved convergence of the estimates to a common value. Many generalizations of the DMHE scheme can be considered. First, one can study how to further explore the degrees of freedom in the choice of the graph matrix K in order to improve the rate of convergence of the state estimates provided by any sensor. Second, each sensor i , beside knowing measurements of its neighbors in the time window $[t-N, t]$, can also know some past measurements of sensors in \mathcal{V}_i^k , $k > 1$. A suitable way of using this piece of information would reduce or totally eliminate unobservability problems. Third, in practical applications it might be possible to perform multiple transmissions within a sampling time and one can study how to use this feature in order to weaken the observability requirements for the convergence of the method.

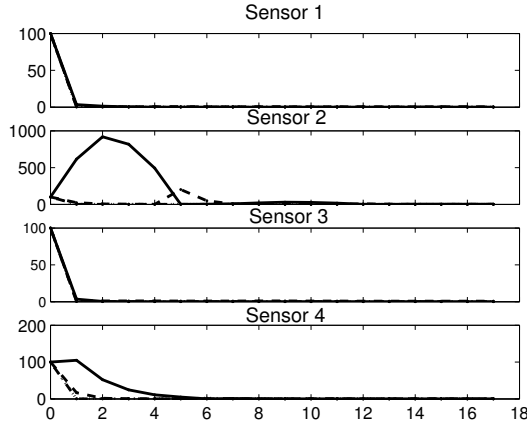


Fig. 5. Evolution of the four eigenvalues (for each sensor) of the matrices $\Pi_{t/t+N-1}^i$, $i = 1, \dots, 4$. Solid line $i = 1$, dotted line $i = 2$, dashed line $i = 3$, dash-dotted line $i = 4$.

These issues have partially been explored in [23].

APPENDIX

A. Proof of Proposition 1

Proof: For all vectors $\mathbf{x} = [x_1^T \dots x_M^T]^T \in \mathbb{R}^M$, from (28) it holds that

$$\mathbf{x}^T \Pi_{t-N/t-1} \mathbf{x} \geq \mathbf{x}^T \mathbf{K} \tilde{\Pi}_{t-N/t-1} \mathbf{K}^T \mathbf{x} \quad (35)$$

Notice that the j -th block of $\mathbf{K}^T \mathbf{x}$ corresponds to $\sum_{i=1}^M k_{ij} x_i$, so that, in view of (20), the right hand side of equation (35) can be written as

$$\|\mathbf{K}^T \mathbf{x}\|_{\tilde{\Pi}_{t-N/t-1}}^2 = \sum_{j=1}^M \left\| \sum_{i=1}^M k_{ij} x_i \right\|_{\tilde{\Pi}_{t-N/t-1}^j}^2$$

Using the triangle inequality we obtain

$$\begin{aligned} \sum_{j=1}^M \left\| \sum_{i=1}^M k_{ij} x_i \right\|_{\tilde{\Pi}_{t-N/t-1}^j}^2 &\leq 2 \sum_{j=1}^M \sum_{i=1}^M k_{ij}^2 \|x_i\|_{\tilde{\Pi}_{t-N/t-1}^j}^2 \\ &= \sum_{i=1}^M \|x_i\|_{2 \sum_{j=1}^M k_{ij}^2 \tilde{\Pi}_{t-N/t-1}^j}^2 \end{aligned}$$

which proves that matrices $\Pi_{t-N/t-1}^i$ verifying (30) also verify (28). \blacksquare

B. Proof of Theorem 1

The proof of Theorem 1 uses classical results for MHE, [24], [16], [13], [15] and additional results we provide next.

Lemma 1: If (28) is satisfied then, for $z \in \mathbb{X}$, (18) is fulfilled.

Proof of Lemma 1: Let $\mathbf{z} = \mathbf{1}_M \otimes z$. We define the ‘‘unconstrained’’ transit cost as

$$\begin{aligned} \mathfrak{E}_{t-N+1/t}^u(\mathbf{z}) &= \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \{J(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}, \hat{\mathbf{v}}, \Gamma_{t-N}) \\ &\text{subject to (15a), (15b) and } \hat{\mathbf{x}}_{t-N+1} = \mathbf{z}\} \end{aligned}$$

that, differently from $\mathfrak{E}_{t-N+1/t}$ in (17), does not account for input and state constraints. Notice that

$$\mathfrak{E}_{t-N+1/t}^u(\mathbf{z}) = \sum_{i=1}^M \mathfrak{E}_{t-N+1/t}^{i,u}(\mathbf{z}) \quad (36)$$

where the unconstrained transit cost associated to sensor i is

$$\begin{aligned} \mathfrak{E}_{t-N+1/t}^{i,u}(\mathbf{z}) &= \min_{\hat{\mathbf{x}}_{t-N}, \{\hat{\mathbf{w}}_k\}_{k=t-N}^{t-1}} \{J^i(t-N, t, \hat{\mathbf{x}}_{t-N}, \hat{\mathbf{w}}^i, \hat{\mathbf{v}}^i, \Gamma_{t-N}) \\ &\text{subject to (7a), (7b) and } \hat{\mathbf{x}}_{t-N+1}^i = \mathbf{z}\} \quad (37) \end{aligned}$$

We first compute explicitly $\mathfrak{E}_{t-N+1/t}^{i,u}(\mathbf{z})$. Recalling (7) we can write

$$\tilde{\mathbf{V}}_i^{[t-N+1,t]} = \tilde{\mathbf{Y}}_i^{[t-N+1,t]} - \tilde{\mathcal{O}}_N^i \hat{\mathbf{x}}_{t-N+1}^i - \mathcal{C}_N^i W_i^{[t-N+1,t-1]}$$

where matrices \mathcal{C}_N^i and $\tilde{\mathcal{O}}_{N+1}^i$ are defined in (21) and (5), respectively, $\tilde{\mathbf{V}}_i^{[t-N+1,t]} = [(\hat{\mathbf{v}}_{t-N+1}^i)^T, \dots, (\hat{\mathbf{v}}_t^i)^T]^T$, $\tilde{\mathbf{Y}}_i^{[t-N+1,t]} = [(\bar{\mathbf{y}}_{t-N+1}^i)^T, \dots, (\bar{\mathbf{y}}_t^i)^T]^T$, and $W_i^{[t-N+1,t-1]} = [(\hat{\mathbf{w}}_{t-N+1}^i)^T, \dots, (\hat{\mathbf{w}}_{t-1}^i)^T]^T$. We can rewrite the i -th sensor’s cost function as

$$\begin{aligned} 2(J^i - \Theta_{t-1}^{*i}) &= \|\tilde{\mathbf{Y}}_i^{[t-N+1,t]} - \tilde{\mathcal{O}}_N^i \hat{\mathbf{x}}_{t-N+1}^i - \mathcal{C}_N^i W_i^{[t-N+1,t-1]}\|_{(\tilde{\mathbf{R}}_N^i)^{-1}}^2 \\ &+ \|W_i^{[t-N+1,t-1]}\|_{(Q_{N-1})^{-1}}^2 + \|\hat{\mathbf{w}}_{t-N}^i\|_{Q^{-1}}^2 + \\ &+ \underbrace{\|\hat{\mathbf{x}}_{t-N}^i - \sum_{j=1}^M k_{ij} \hat{\mathbf{x}}_{t-N/t-1}^j\|_{(\Pi_{t-N/t-1}^i)^{-1}}^2 + \|\bar{\mathbf{y}}_{t-N}^i - \tilde{\mathbf{C}}^i \hat{\mathbf{x}}_{t-N}^i\|_{(\tilde{\mathbf{R}})^{-1}}^2}_{(*)} \quad (38) \end{aligned}$$

where matrices $\tilde{\mathbf{R}}_N^i$ and Q_{N-1}^i are defined in (22) and (23), respectively. Minimizing the partial cost $(*)$ in (38) with respect to $\hat{\mathbf{x}}_{t-N}^i$ gives

$$\hat{\mathbf{x}}_{t-N/t-1}^{*i} = \Pi_{t-N/t-1}^{*i} [(\Pi_{t-N/t-1}^i)^{-1} \sum_{j=1}^M k_{ij} \hat{\mathbf{x}}_{t-N/t-1}^j + (\tilde{\mathbf{C}}^i)^T (\tilde{\mathbf{R}}^i)^{-1} \bar{\mathbf{y}}_{t-N}^i]$$

with $\Pi_{t-N/t-1}^{*i}$ as in (26). Therefore one has that the term $(*)$ in (38) is equal to

$$\|\hat{\mathbf{x}}_{t-N}^i - \hat{\mathbf{x}}_{t-N/t-1}^{*i}\|_{(\Pi_{t-N/t-1}^{*i})^{-1}}^2 \quad (39)$$

up to a constant term.

We denote with $L^i(\cdot)$ the minimum of $J^i(\cdot)$ with respect to vector $W_i^{[t-N+1,t-1]}$, i.e.,

$$L^i = \min_{\{\hat{\mathbf{w}}_k\}_{k=t-N+1}^{t-1}} J^i(\cdot) \quad (40)$$

We compute $\frac{\partial J^i(\cdot)}{\partial W_i^{[t-N+1,t-1]}} = 0$. The vector $W_i^{[t-N+1,t-1],opt}$ which solves the minimization problem (40) is

$$\begin{aligned} W_i^{[t-N+1,t-1],opt} &= ((\mathcal{C}_N^i)^T (\tilde{\mathbf{R}}_N^i)^{-1} \mathcal{C}_N^i + (Q_{N-1}^i)^{-1})^{-1} \times \\ &\times (\mathcal{C}_N^i)^T (\tilde{\mathbf{R}}_N^i)^{-1} (\tilde{\mathbf{Y}}_i^{[t-N+1,t]} - \tilde{\mathcal{O}}_N^i \hat{\mathbf{x}}_{t-N+1}^i) \quad (41) \end{aligned}$$

Replacing (41) into (38) and using (39) one obtains

$$\begin{aligned} L^i &= \|\tilde{\mathbf{Y}}_i^{[t-N+1,t]} - \tilde{\mathcal{O}}_N^i \hat{\mathbf{x}}_{t-N+1}^i\|_{(\tilde{\mathbf{R}}_N^i + \mathcal{C}_N^i Q_{N-1}^i (\mathcal{C}_N^i)^T)^{-1}}^2 \\ &+ \|\hat{\mathbf{w}}_{t-N}^i\|_{Q^{-1}}^2 + \|\hat{\mathbf{x}}_{t-N}^i - \hat{\mathbf{x}}_{t-N/t-1}^{*i}\|_{(\Pi_{t-N/t-1}^{*i})^{-1}}^2 \end{aligned}$$

up to an additive constant term. The solution of the optimization problem (37) can be computed through a Kalman filter recursion with respect to the modified dynamical system

$$\begin{cases} \hat{\mathbf{x}}_{t-N+1}^i &= A \hat{\mathbf{x}}_{t-N}^i + w_{t-N} \\ \tilde{\mathbf{Y}}_i^{[t-N+1,t]} &= \tilde{\mathcal{O}}_N^i \hat{\mathbf{x}}_{t-N+1}^i + \tilde{\mathbf{V}}_i^{[t-N+1,t]} \end{cases} \quad (42)$$

where w_t has covariance equal to Q , the covariance of $\tilde{\mathbf{V}}_i^{[t-N+1,t]}$ is $\tilde{\mathbf{R}}_N^i + \mathcal{C}_N^i Q_{N-1}^i (\mathcal{C}_N^i)^T$, and $\Pi_{t-N/t-1}^{*i}$ in (26) is the

uncertainty of the initial condition guess. In this way we can write the unconstrained transit cost as follows (see [20])

$$\mathfrak{E}_{t-N+1/t}^{i,u}(z) = \frac{1}{2} \|z - \hat{x}_{t-N+1/t}^{i,u}\|_{(\tilde{\Pi}_{t-N+1/t}^i)^{-1}}^2 + \Theta_t^{*,i,u} \quad (43)$$

where $\hat{x}_{t-N+1/t}^{i,u}$ minimizes the unconstrained problem, and $\Theta_t^{*,i,u}$ is the optimal solution of the unconstrained minimization problem, and $\tilde{\Pi}_{t-N+1/t}^i$ is computed as in (27). Remark that the regionally unobservable subspaces of system (42) and system (1)-(2) coincide.

From (43) and (36) one has that

$$\mathfrak{E}_{t-N+1/t}^u(\mathbf{z}) = \frac{1}{2} \|\mathbf{z} - \hat{\mathbf{x}}_{t-N+1/t}^u\|_{(\tilde{\Pi}_{t-N+1/t}^u)^{-1}}^2 + \Theta_t^{*,u} \quad (44)$$

where $\Theta_t^{*,u} = \sum_{i=1}^M \Theta_t^{*,i,u}$ and $\tilde{\Pi}_{t-N+1/t}^u = \text{diag}(\tilde{\Pi}_{t-N+1/t}^1, \dots, \tilde{\Pi}_{t-N+1/t}^M)$. We also define $\Theta_{t-N+1/t}^u(\mathbf{x}; \hat{\mathbf{x}}_{t-N+1/t}^u) = \frac{1}{2} \|\mathbf{x} - \hat{\mathbf{x}}_{t-N+1/t}^u\|_{(\tilde{\Pi}_{t-N+1/t}^u)^{-1}}^2$ in such a way that $\mathfrak{E}_{t-N+1/t}^u(\mathbf{z}) = \Theta_{t-N+1/t}^u(\mathbf{z}; \hat{\mathbf{x}}_{t-N+1/t}^u) + \Theta_t^{*,u}$. Let us finally consider the case of constrained estimation. Following the rationale of the proof of Lemma 4 in [13] one has that, since \mathbf{z} lies in the feasibility region by assumption, one obtains (19). Notice that the initial penalty term $\Gamma_{t-N+1}(\cdot)$, computed as in (13) in \mathbf{z} , is

$$\begin{aligned} \Gamma_{t-N+1}(\mathbf{z}; \hat{\mathbf{x}}_{t-N+1/t}) &= \frac{1}{2} \|\mathbf{z} - \mathbf{K} \hat{\mathbf{x}}_{t-N+1/t}\|_{\Pi_{t-N+1/t}^{-1}}^2 + \Theta_t^* \\ &= \frac{1}{2} \|\mathbf{z} - \hat{\mathbf{x}}_{t-N+1/t}\|_{\mathbf{K}^T \Pi_{t-N+1/t}^{-1} \mathbf{K}}^2 + \Theta_t^* \end{aligned} \quad (45)$$

where the second equality holds because $\mathbf{K} \mathbf{z} = \mathbf{z}$. Using Schur complement, the LMI (28) is equivalent to

$$\begin{bmatrix} \Pi_{t-N/t-1} & \mathbf{K} \\ \mathbf{K}^T & \tilde{\Pi}_{t-N/t-1}^{-1} \end{bmatrix} \geq 0 \quad (46)$$

and, being matrices $\Pi_{t-N/t-1}$ and $\tilde{\Pi}_{t-N/t-1}$ positive definite, (46) is equivalent to

$$\mathbf{K}^T \Pi_{t-N/t-1}^{-1} \mathbf{K} \leq \tilde{\Pi}_{t-N/t-1}^{-1} \quad (47)$$

From (19) and (45), (47) implies (18). ■

Lemma 2: If (28) is satisfied, then

$$\Theta_t^* \leq \Gamma_0(\mathbf{x}_0; \mathbf{x}_{0/0}) \text{ for all } t \geq 0 \quad (48)$$

where $\mathbf{x}_0 = [x_0^T \dots x_0^T]^T \in \mathbb{R}^{nM}$ and $\mathbf{x}_{0/0} = \mathbf{1}_M \otimes \boldsymbol{\mu}$.

Proof of Lemma 2: First notice that, in view of Definition 2, the sequence $\mathbf{x}_\Sigma(t, x_0)$ verifies the constraints (15d). In view of Lemma 1, equation (18) holds for $\mathbf{z} = \mathbf{x}_\Sigma(t, x_0)$, for all t . By optimality, we have

$$\Theta_t^* \leq \mathfrak{E}_{t-N+1/t}(\mathbf{x}_\Sigma(t-N+1, x_0)), \forall t \geq 0$$

Furthermore

$$\mathfrak{E}_{t-N+1/t}(\mathbf{x}_\Sigma(t-N+1, x_0)) \leq \mathbf{J}(t-N, t, \mathbf{x}_\Sigma(t-N, x_0), 0, 0, \Gamma_{t-N})$$

Note that, from (14), one has $\mathbf{J}(t-N, t, \mathbf{x}_\Sigma(t-N, x_0), 0, 0, \Gamma_{t-N}) = \Gamma_{t-N}(\mathbf{x}_\Sigma(t-N, x_0); \hat{\mathbf{x}}_{t-N/t-1})$ and in view of (18), $\Theta_t^* \leq \mathfrak{E}_{t-N/t-1}(\mathbf{x}_\Sigma(t-N, x_0)) \leq \Gamma_{t-N-1}(\mathbf{x}_\Sigma(t-N-1, x_0); \hat{\mathbf{x}}_{t-N-1/t-2})$. We can further iterate this procedure in order to prove (48). ■

Lemma 3: Assume that (a) $N \geq n-1$, with $N \geq 1$, (b) $\exists \bar{\Pi}$ such that $\Pi_{t-N/t-1}^i < \bar{\Pi}$, for all t , for all $i \in \mathcal{V}$, and (c)

$$\max_{k=t-N, \dots, t} (\|\hat{\mathbf{v}}_{k/t}^i\|, \|\hat{\mathbf{w}}_{k/t}^i\|, \|\Gamma_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1})\|) \xrightarrow{t \rightarrow \infty} 0 \quad (49)$$

Then the dynamics of the state estimation error provided by the DMHE scheme is given by (32).

Proof of Lemma 3: In the noiseless case, for any sensor $i \in \mathcal{V}$ at any t , the output signal is $\bar{y}_k^i = \bar{C}^i x_\Sigma(t, x_0)$. Similarly to Lemma 4.3 in [24],

$$\begin{aligned} \sum_{k=t-N}^t \|\hat{\mathbf{v}}_{k/t}^i\| &= \sum_{k=t-N}^t \|\bar{y}_k^i - \bar{C}^i \hat{x}_{k/t}^i\| \geq \\ &\geq \sum_{k=t-N}^t (\|\bar{y}_k^i - \bar{C}^i A^{k-(t-N)} \hat{x}_{t-N/t}^i\| - \|\bar{C}^i A^{k-(t-N)} \hat{x}_{t-N/t}^i - \bar{C}^i \hat{x}_{k/t}^i\|) \end{aligned} \quad (50)$$

The first term at the right hand side of (50) is

$$\sum_{k=t-N}^t \|\bar{y}_k^i - \bar{C}^i A^{k-(t-N)} \hat{x}_{t-N/t}^i\| = \|\bar{\mathcal{O}}_{N+1}^i(x_\Sigma(t-N, x_0) - \hat{x}_{t-N/t}^i)\| \quad (51)$$

where $\bar{\mathcal{O}}_{N+1}^i$ is the ‘‘extended’’ regional observability matrix of $N+1$ rows defined by replacing n with $N+1$ in (5). From (7), one has

$$\hat{x}_{k/t}^i = A^{k-(t-N)} \hat{x}_{t-N/t}^i + \sum_{j=1}^{k-(t-N)} A^j \hat{w}_{k-j/t}^i$$

The second term at the right hand side of (50) can be bounded as

$$\begin{aligned} \sum_{k=t-N}^t \|\bar{C}^i A^{k-(t-N)} \hat{x}_{t-N/t}^i - \bar{C}^i \hat{x}_{k/t}^i\| &\leq \\ &\leq \sum_{k=t-N}^t \|\bar{C}^i \sum_{j=1}^{k-(t-N)} A^j \hat{w}_{k-j/t}^i\| \\ &\leq \|\bar{C}^i\| \sum_{k=t-N}^t \sum_{j=1}^{k-(t-N)} \|A\|^j \|\hat{w}_{k-j/t}^i\| \end{aligned} \quad (52)$$

By replacing equations (51) and (52) into (50), one obtains

$$\begin{aligned} \|\bar{\mathcal{O}}_{N+1}^i(x_{t-N/t}^i - x_\Sigma(t-N, x_0))\| &\leq \\ &\leq \sum_{k=t-N}^t \|\hat{\mathbf{v}}_{k/t}^i\| + \|\bar{C}^i\| \sum_{k=t-N}^t \sum_{j=1}^{k-(t-N)} \|A\|^j \|\hat{w}_{k-j/t}^i\| \end{aligned} \quad (53)$$

Note that the matrix $\bar{\mathcal{O}}_{N+1}^i$ at the left-hand side of (53) selects the observable part of $(\hat{x}_{t-N/t}^i - x_\Sigma(t-N, x_0))$. Therefore, from (49), equation (53) leads to

$$\|\mathbf{P}_0(\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0))\| \xrightarrow{t \rightarrow \infty} 0, \quad (54)$$

$$\Gamma_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) \xrightarrow{t \rightarrow \infty} 0 \quad (55)$$

In view of assumption (b), it follows that

$$\Gamma_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) \geq \|\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0)\|_{\text{diag}(\bar{\Pi}^{-1}, \dots, \bar{\Pi}^{-1})}$$

Hence, from (55)

$$\|\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0)\| \xrightarrow{t \rightarrow \infty} 0 \quad (56)$$

Note that, for $k = t-N, \dots, t-1$

$$\hat{\mathbf{x}}_{k+1/t} = \mathbf{A} \hat{\mathbf{x}}_{k/t} + \hat{\mathbf{w}}_{k/t} \quad (57)$$

and that, in view of (49), one has $\hat{\mathbf{w}}_{k/t} \rightarrow 0$ as $t \rightarrow \infty$. Therefore one also has

$$\hat{\mathbf{x}}_{k+1/t} - \mathbf{A} \hat{\mathbf{x}}_{k/t} \rightarrow 0 \text{ as } t \rightarrow \infty \quad (58)$$

From now on, we introduce, for simplicity of notation, terms α_t^j to indicate asymptotically vanishing variables, *i.e.*, $\|\alpha_t^j\| \xrightarrow{t \rightarrow \infty} 0$, for all $j \in \mathcal{V}$. Formulae (54), (56) and (57) are equivalent to

$$\mathbf{P}_O \hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O \mathbf{x}_\Sigma(t-N, x_0) + \alpha_t^1 \quad (59a)$$

$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{K} \hat{\mathbf{x}}_{t-N/t-1} + \alpha_t^2 \quad (59b)$$

$$\hat{\mathbf{x}}_{t-N+1/t} = \mathbf{A} \hat{\mathbf{x}}_{t-N/t} + \alpha_t^3 \quad (59c)$$

Recall that, by definition, $\mathbf{P}_O + \mathbf{P}_{NO} = \mathbf{I}$. Therefore,

$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O \hat{\mathbf{x}}_{t-N/t} + \mathbf{P}_{NO} \hat{\mathbf{x}}_{t-N/t} \quad (60)$$

In (60), we replace terms $\mathbf{P}_O \hat{\mathbf{x}}_{t-N/t}$ and $\mathbf{P}_{NO} \hat{\mathbf{x}}_{t-N/t}$ according to (59a) and (59b), premultiplied by \mathbf{P}_{NO} , respectively, we get

$$\hat{\mathbf{x}}_{t-N/t} = \mathbf{P}_O \mathbf{x}_\Sigma(t-N, x_0) + \mathbf{P}_{NO} \mathbf{K} \hat{\mathbf{x}}_{t-N/t-1} + \alpha_t^4 \quad (61)$$

Since $\mathbf{P}_O + \mathbf{P}_{NO} = \mathbf{I}$, we write $\mathbf{P}_O \mathbf{x}_\Sigma(t-N, x_0) = \mathbf{x}_\Sigma(t-N, x_0) - \mathbf{P}_{NO} \mathbf{x}_\Sigma(t-N, x_0)$, and obtain

$$\hat{\mathbf{x}}_{t-N/t} - \mathbf{x}_\Sigma(t-N, x_0) = \mathbf{P}_{NO} (\mathbf{K} \hat{\mathbf{x}}_{t-N/t-1} - \mathbf{x}_\Sigma(t-N, x_0)) + \alpha_t^4 \quad (62)$$

First recall that, since \mathbf{K} is stochastic and $\mathbf{x}_\Sigma(t-N, x_0) = \mathbf{1}_M \otimes x_\Sigma(t-N, x_0)$, $\mathbf{K} \mathbf{x}_\Sigma(t-N, x_0) = \mathbf{x}_\Sigma(t-N, x_0)$. Then notice that $\mathbf{x}_\Sigma(t-N, x_0) = \mathbf{A} \mathbf{x}_\Sigma(t-N-1, x_0)$. From (59c) one obtains

$$\begin{aligned} \boldsymbol{\varepsilon}_{t-N/t} &= \mathbf{P}_{NO} \mathbf{K} \mathbf{A} \boldsymbol{\varepsilon}_{t-N-1/t-1} + \alpha_t^5 \\ &= \mathbf{P}_{NO} \mathbf{K} \mathbf{A} (\mathbf{P}_{NO} + \mathbf{P}_O) \boldsymbol{\varepsilon}_{t-N/t} + \alpha_t^5 \end{aligned} \quad (63)$$

Equation (59a) implies that the term $\mathbf{P}_O \boldsymbol{\varepsilon}_{t-N/t}$ is asymptotically vanishing and equation (32) follows from (63). ■

Proof of Theorem 1: By direct calculation, for all $t \geq 0$ one has

$$\boldsymbol{\Theta}_t^* - \boldsymbol{\Theta}_{t-1}^* = \frac{1}{2} \sum_{k=t-N}^t \|\hat{\mathbf{v}}_{k/t}\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_{k/t}\|_{\mathbf{Q}^{-1}}^2 + \boldsymbol{\Gamma}_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1})$$

Furthermore, (48) follows from Lemma 2 and (28). Therefore it follows that $\frac{1}{2} \sum_{k=t-N}^t \|\hat{\mathbf{v}}_{k/t}\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \sum_{k=t-N}^{t-1} \|\hat{\mathbf{w}}_{k/t}\|_{\mathbf{Q}^{-1}}^2 + \boldsymbol{\Gamma}_{t-N}^o(\hat{\mathbf{x}}_{t-N/t}; \hat{\mathbf{x}}_{t-N/t-1}) \xrightarrow{t \rightarrow \infty} 0$ and hence (49) holds. This, in turn, implies (using Lemma 3) that the dynamics of state estimation error provided by the DMHE scheme is given by (32).

Furthermore, from (32), convergence of the error to zero is guaranteed if Φ is Schur. ■

C. Proof of Theorem 2 and Corollary 1

Proof of Theorem 2: If the graph \mathcal{G} is not strongly connected it can be partitioned into k irreducible subgraphs \mathcal{G}_1^* , \mathcal{G}_2^* , ..., \mathcal{G}_k^* of cardinality m_1 , ..., m_k , and $\sum_{i=1}^k m_i = M$. Without loss of generality, (*i.e.* by permuting sensor indexes) the matrix \mathbf{K} can be brought in a block lower triangular form (with k square diagonal blocks K_{11}, \dots, K_{kk} , of dimensions m_1 , ..., m_k , respectively).

Notice that the block K_{ii} is stochastic if and only if $K_{ij} = 0$ for $j < i$. In this case, the nodes of the subgraph \mathcal{G}_i^* have no neighbors belonging to other subgraphs and \mathcal{G}_i^* is *isolated*. Moreover, if a subgraph \mathcal{G}_i^* is isolated, the block K_{ii} is

stochastic and it has a single Frobenius eigenvalue equal to 1. On the other hand, if a graph \mathcal{G}_i^* is not isolated, K_{ii} is irreducible but not stochastic (specifically, the sum of the entries of at least a row is smaller than 1) and its Frobenius eigenvalue has absolute value smaller than 1¹. Notice that the eigenvalues of \mathbf{K} are the eigenvalues of K_{11}, \dots, K_{kk} . So, the number of eigenvalues of \mathbf{K} equal to 1 equals the number of isolated graphs in the network.

Note that $\mathbf{T}^{-1} \mathbf{A} \mathbf{T} = \mathbf{A}_{KO} = \text{diag}(A_{KO}^1, \dots, A_{KO}^M)$ where A_{KO}^i is the ‘‘regional’’ observability Kalman decomposition of \mathbf{A} associated to sensor i , that is

$$A_{KO}^i = \begin{bmatrix} A_O^i & 0 \\ A_{21}^i & A_{NO}^i \end{bmatrix} \quad (64)$$

Since $\mathbf{P}_{NO} = \mathbf{T} \mathbf{S}_{NO} \mathbf{S}_{NO}^T \mathbf{T}^{-1}$ one has

$$\Phi = \mathbf{P}_{NO} \mathbf{K} \tilde{\mathbf{A}}$$

where $\tilde{\mathbf{A}} = \mathbf{T} \tilde{\mathbf{A}}_{KO} \mathbf{T}^{-1}$, $\tilde{\mathbf{A}}_{KO} = \text{diag}(\tilde{A}_{KO}^1, \dots, \tilde{A}_{KO}^M)$, and

$$\tilde{A}_{KO}^i = \begin{bmatrix} 0 & 0 \\ 0 & A_{NO}^i \end{bmatrix}$$

Now we prove that $\mathbf{x} \in \text{span}(e_1 \otimes v_A^i, \dots, e_M \otimes v_A^i)$ is not an eigenvector of Φ associated to a non-zero eigenvalue λ_A^i . In general, given a vector $\alpha \in \mathbb{R}^M$, with $\alpha \neq 0$, one has that the eigenvector \mathbf{x} of \mathbf{A} can be written as $\mathbf{x} = \alpha \otimes v_A^i$. We obtain

$$\tilde{\mathbf{A}} \mathbf{x} = \text{diag}(\tilde{A}^1, \dots, \tilde{A}^M) \begin{bmatrix} \alpha_1 v_A^i \\ \vdots \\ \alpha_M v_A^i \end{bmatrix} = \begin{bmatrix} \alpha_1 \tilde{A}^1 v_A^i \\ \vdots \\ \alpha_M \tilde{A}^M v_A^i \end{bmatrix}$$

By construction, $\tilde{A}^j v_A^i = \lambda_A^i v_A^i$ if v_A^i belongs to the regionally unobservable subspace of sensor j . Otherwise $\tilde{A}^j v_A^i = 0$. We write, in general $\tilde{A}^j v_A^i = f_{ij} \lambda_A^i v_A^i$, where

$$f_{ij} = \begin{cases} 1 & \text{if } v_A^i \text{ belongs to the regionally unobservable} \\ & \text{subspace of sensor } j \\ 0 & \text{otherwise.} \end{cases}$$

Defining $f_i = [f_{i1}, \dots, f_{iM}]^T$, we can write

$$\tilde{\mathbf{A}} \mathbf{x} = \lambda_A^i (\text{diag}(f_i) \alpha) \otimes v_A^i$$

From $\mathbf{K} = \mathbf{K} \otimes \mathbf{I}_n$ we obtain

$$\Phi \mathbf{x} = \lambda_A^i \mathbf{P}_{NO} (\mathbf{K} \otimes \mathbf{I}_n) [(\text{diag}(f_i) \alpha) \otimes v_A^i]$$

Recall that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \mathbf{C}) \otimes (\mathbf{B} \mathbf{D})$ and hence

$$\Phi \mathbf{x} = \lambda_A^i \mathbf{P}_{NO} (\mathbf{K} \text{diag}(f_i) \alpha) \otimes v_A^i = \lambda_A^i (\text{diag}(f_i) \mathbf{K} \text{diag}(f_i) \alpha) \otimes v_A^i$$

Finally, we obtain

$$\Phi(\alpha \otimes v_A^i) = \lambda_A^i (\text{diag}(f_i) \mathbf{K} \text{diag}(f_i) \alpha) \otimes v_A^i$$

from which it is apparent that $\alpha \otimes v_A^i$ is an eigenvector of Φ , with eigenvalue $\lambda_A^i \neq 0$ if and only if $\text{diag}(f_i) \mathbf{K} \text{diag}(f_i) \alpha = \alpha$.

¹This follows from the third Gerschgorin theorem [25], dealing with irreducible matrices. Specifically, an eigenvalue of an irreducible matrix (in our case K_{ii}), which is on the boundary of a Gerschgorin circle, is located on the boundary of all the Gerschgorin circles. Since there is at least a row of K_{ii} such that the sum of its entries is smaller than 1, 1 cannot be an eigenvalue of K_{ii} , and hence all the eigenvalues of K_{ii} are strictly inside the unit circle (from the first Gerschgorin theorem).

Moreover, there exists α satisfying the previous equation if and only if $\text{diag}(f_i)K\text{diag}(f_i)$ has at least one eigenvector equal to 1. This occurs if and only if $f_{ij} = 1$ for all j belonging to an isolated subgraph. This means that all the sensors of an isolated subgraph have at least a common regionally unobservable eigenvector. Hence, $\mathbf{x} \in \text{span}(e_1 \otimes v_A^i, \dots, e_M \otimes v_A^i)$ can not be an eigenvector of Φ if (33) holds. This completes the proof. \blacksquare

Proof of Corollary 1: Recalling Definition 1, collective observability holds if and only if the observability matrix \mathbf{O}^* of the pair (A, \mathbf{C}^*) is such that

$$\ker(\mathbf{O}^*) = 0 \quad (65)$$

Notice that, up to a permutation of the rows of \mathbf{O}^* , we have $[(\bar{\mathcal{O}}_n^1)^T \ \dots \ (\bar{\mathcal{O}}_n^M)^T]^T$. Therefore (65) is equivalent to

$$\bigcap_{i \in \mathcal{V}} \ker(\bar{\mathcal{O}}_n^i) = 0$$

which is equivalent to (33) when the graph is strongly connected. This concludes the proof. \blacksquare

D. Proof of Theorem 3

To prove Theorem 3, a number of intermediate results are needed. First, we address the problem of the stability of Riccati equations with respect to perturbations. This problem has been scarcely explored in the literature, with the exception of [26] where stability is proved with respect to small perturbations. In the following, we explore the issue under the lead of Theorem 4.1 in [27], and provide global stability results.

Given a pair (A, C) , and matrices $Q \geq 0, R > 0$ of appropriate size, consider the following Riccati equation, affected by an exogenous perturbation term Δ_k

$$\Pi_{k+1}^o = (A - G_k^o C)(\Pi_k^o + \Delta_k)(A - G_k^o C)^T + Q + G_k^o R (G_k^o)^T \quad (66)$$

where Π_0^o is the initial condition and matrix G_k^o is the Kalman gain

$$G_k^o = A(\Pi_k^o + \Delta_k)C^T (C(\Pi_k^o + \Delta_k)C^T + R)^{-1} \quad (67)$$

Assuming that the pair (A, C) is detectable and that the pair (A, \sqrt{Q}) is stabilizable, there exists a unique solution $\bar{\Pi} \geq 0$ of the algebraic Riccati equation associated to (66) with $\Delta_k = 0$.

In the sequel, we will denote with \vec{X}_τ the sequence of matrices X_k , with $k = 0, \dots, \tau$. In [28] the following definition of \mathcal{L} -stability of system (66) is given.

Definition 3: System (66) is \mathcal{L} -stable from input Δ_k if, for a given norm \mathcal{L} , there exist $\gamma > 0$ and $\beta > 0$ such that

$$\|\vec{\Pi}_\tau - \bar{\Pi}\|_{\mathcal{L}} \leq \gamma \|\vec{\Delta}_\tau\|_{\mathcal{L}} + \beta, \quad \forall \Delta \in \mathcal{L}, \forall \tau \in [0, \infty).$$

From now on, we denote with $\|\vec{X}_\tau\|_{\mathcal{L}_\infty}$ the ∞ -norm of the sequence $\|X_t\|_2$, with $t = 0, \dots, \tau$.

Lemma 4: Given a detectable pair (A, C) , system (66) is \mathcal{L}_∞ -stable from a positive semi-definite input $\Delta_k \geq 0$.

Proof: We define a sequence Π_k (with $\Pi_0 = \Pi_0^o$) as follows

$$\Pi_{k+1} = (A - GC)(\Pi_k + \Delta_k)(A - GC)^T + Q + GRG^T, \quad (68)$$

where G is an arbitrary gain such that $F = A - GC$ is Hurwitz. Notice that G always exists, since (A, C) is detectable. From (68), we obtain, for $k \geq 1$,

$$\Pi_{k+1} - \Pi_k = F(\Pi_k - \Pi_{k-1})F^T + F(\Delta_k - \Delta_{k-1})F^T$$

and hence, for $i \geq 1$,

$$\Pi_{i+1} - \Pi_i = F^i(\Pi_1 - \Pi_0)(F^T)^i + \sum_{j=1}^i F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j$$

Then, for $k > 1$,

$$\begin{aligned} \Pi_k &= \Pi_0 + \sum_{i=0}^{k-1} (\Pi_{i+1} - \Pi_i) \\ &= \sum_{i=0}^{k-1} F^i(\Pi_1 - \Pi_0)(F^T)^i + \Pi_0 + \\ &\quad + \sum_{i=1}^{k-1} \sum_{j=1}^i F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j \end{aligned} \quad (69)$$

Notice that, assuming $\Delta_0 = 0$ in (69) one has

$$\sum_{i=1}^{k-1} \sum_{j=1}^i F^j(\Delta_{i+1-j} - \Delta_{i-j})(F^T)^j = \sum_{i=1}^{k-1} F^i \Delta_{k-i} (F^T)^i$$

and (69) gives

$$\Pi_k = \sum_{i=0}^{k-1} F^i(\Pi_1 - \Pi_0)(F^T)^i + \Pi_0 + \sum_{i=1}^{k-1} F^i \Delta_{k-i} (F^T)^i \quad (70)$$

Let us set $\|\Pi_1 - \Pi_0\|_2 = \alpha$ and $\|\vec{\Delta}_\infty\|_{\mathcal{L}_\infty} = \bar{\delta}$. Since F is Hurwitz, there exists $\mu > 0$ and $0 < \nu < 1$ such that $\|F^i\|_2 \leq \mu \nu^i$. Remark that, since $\Delta_k \geq 0$, from optimality of Π_k^o [27] one has $0 \leq \Pi_k^o \leq \Pi_k, \forall k \geq 0$, and hence $\|\Pi_k^o\|_2 \leq \|\Pi_k\|_2$. Furthermore, from (70)

$$\begin{aligned} \|\Pi_k^o\|_2 &\leq \|\Pi_k\|_2 \leq \sum_{i=0}^{k-1} \|F^i\|_2^2 \|\Pi_1 - \Pi_0\|_2 + \|\Pi_0\|_2 + \\ &\quad + \sum_{i=1}^{k-1} \|F^i\|_2^2 \|\Delta_{k-i}\|_2 \\ &\leq \alpha \mu^2 \sum_{i=0}^{k-1} \nu^{2i} + \|\Pi_0\|_2 + \bar{\delta} \mu^2 \sum_{i=1}^{k-1} \nu^{2i} \\ &\leq \alpha \mu^2 \frac{1 - \nu^{2k}}{1 - \nu^2} + \|\Pi_0\|_2 + \bar{\delta} \mu^2 \nu^2 \frac{1 - \nu^{2(k-1)}}{1 - \nu^2} \\ &\leq \alpha \mu^2 \frac{1}{1 - \nu^2} + \|\Pi_0\|_2 + \mu^2 \nu^2 \frac{1}{1 - \nu^2} \bar{\delta} \end{aligned} \quad (71)$$

The proof is concluded by applying Definition 3 with $\beta = \alpha \mu^2 \frac{1}{1 - \nu^2} + \|\Pi_0\|_2 + \|\bar{\Pi}\|_2$ and $\gamma = \mu^2 \nu^2 \frac{1}{1 - \nu^2}$. \blacksquare

The proof of Theorem 3 can now be completed by applying Lemma 4 and the small gain result for interconnected systems reported in [29].

Proof of Theorem 3: First we show, by applying (27) and (31), that it holds that:

$$\tilde{\Pi}_{t-N/t-1}^i \leq \mathcal{R}^i \left(2 \sum_{j=1}^M k_{ij}^2 \tilde{\Pi}_{t-N-1/t-2}^j; Q, \bar{R}_N^{*i} \right) \quad (72)$$

From (26), one has $\Pi_{t-N-1/t-2}^{*i} \leq \Pi_{t-N-1/t-2}^i$. Then, by applying (27) and by optimality [27],

$$\tilde{\Pi}_{t-N/t-1}^i = \mathcal{R}^i \left(\Pi_{t-N-1/t-2}^{*i}; Q, \bar{R}_N^{*i} \right) \leq \mathcal{R}^i \left(\Pi_{t-N-1/t-2}^i; Q, \bar{R}_N^{*i} \right)$$

and from (31) we obtain (72). Now, with reference to the i -th sensor characterized by the pair $(A, \bar{\mathcal{O}}_N^i)$, we define the following sequence of matrices Π_k^i

$$\Pi_{k+1}^i = \mathcal{R}^i \left(2 \sum_{j=1}^M k_{ij}^2 \Pi_k^j; Q, \bar{R}_N^{*i} \right) \quad (73)$$

with initial condition $\Pi_0^i = \bar{\Pi}_{0/N-1}^i$. From optimality we obtain that $\bar{\Pi}_{k-N/k-1}^i \leq \Pi_{k-N}^i$ for all $k \geq N$. Therefore, in order to prove boundedness of $\Pi_{k-N/k-1}^i$, it is sufficient to show that the sequence Π_k^i is bounded. This is the aim of the remainder of the proof.

If we define $\Delta_k^i = \frac{1}{k_{ii}^2} \sum_{j=1, j \neq i}^M \Pi_k^j$, (73) can be written as

$$\begin{aligned} \Pi_{k+1}^i &= (A - G_k^i \bar{\mathcal{O}}_N^i) 2k_{ii}^2 (\Pi_k^i + \Delta_k^i) (A - G_k^i \bar{\mathcal{O}}_N^i)^T + \\ &\quad + Q + G_k^i \bar{R}_N^{*i} (G_k^i)^T \\ &= (\sqrt{2}k_{ii}A - G_k^i \sqrt{2}k_{ii} \bar{\mathcal{O}}_N^i) (\Pi_k^i + \Delta_k^i) \times \\ &\quad \times (\sqrt{2}k_{ii}A - G_k^i \sqrt{2}k_{ii} \bar{\mathcal{O}}_N^i)^T + Q + G_k^i \bar{R}_N^{*i} (G_k^i)^T \end{aligned} \quad (74)$$

where G_k^i is the optimal Kalman gain computed as

$$G_k^i = A(\Pi_k^i + \Delta_k^i) (\bar{\mathcal{O}}_N^i)^T (\bar{\mathcal{O}}_N^i (\Pi_k^i + \Delta_k^i) (\bar{\mathcal{O}}_N^i)^T + \bar{R}_N^{*i})^{-1}$$

First we show that system (74) is \mathcal{L}_∞ -stable. To this aim, we use Lemma 4. In order to satisfy the assumption of Lemma 4, one must guarantee that the pairs $(\sqrt{2}k_{ii}A, \sqrt{2}k_{ii} \bar{\mathcal{O}}_N^i)$ are detectable, for all $i \in \mathcal{V}$, which turns out to be a condition on the pairs $(A, \bar{\mathcal{C}}^i)$, and on the weights k_{ii} . First notice that, by definition of $\bar{\mathcal{O}}_N^i$, the pair $(\sqrt{2}k_{ii}A, \sqrt{2}k_{ii} \bar{\mathcal{O}}_N^i)$ is detectable if and only if the pair $(\sqrt{2}k_{ii}A, \sqrt{2}k_{ii} \bar{\mathcal{C}}^i)$ is detectable. The assumption of Theorem 3 is sufficient to guarantee that, for any regionally unobservable nodes, there exists a path stemming from a regionally observable node i.e., for which the assumption of Lemma 4 is satisfied for any arbitrary value of k_{ii} . In particular, in step 1 of Algorithm 1 $k_{ii} = 1$ is chosen, for all $i \in \mathcal{V}_O$.

On the other hand, if $(A, \bar{\mathcal{C}}^i)$ is not observable, the assumption of Lemma 4 can be verified if the pair $(\sqrt{2}k_{ii}A, \bar{\mathcal{C}}^i)$ is detectable. This leads to the choice of k_{ii} in step 2 of Algorithm 1. Then, by Lemma 4, (74) is a finite gain \mathcal{L}_∞ -stable system from input $\Delta_k^i \geq 0$, and there exist $\gamma_i > 0, \beta_i > 0$ such that

$$\|\Pi_k^i - \bar{\Pi}^i\|_2 \leq \gamma_i \|\vec{\Delta}_k^i\|_{\mathcal{L}_\infty} + \beta_i, \quad \forall \Delta_k^i \in \mathcal{L}_\infty, \forall k \in [0, \infty) \quad (75)$$

From the definition of Δ_k^i we get

$$\|\Pi_k^i - \bar{\Pi}^i\|_2 \leq \frac{\gamma_i}{k_{ii}^2} \sum_{j=1, j \neq i}^M k_{ij}^2 \|\bar{\Pi}_k^j\|_{\mathcal{L}_\infty} + \beta_i \quad (76)$$

$\forall \Pi_k^j \in \mathcal{L}_\infty, \forall k \in [0, \infty), \forall j \in \mathcal{V}^i$. Given (76), we resort to Theorem 8² in [29] for guaranteeing that Π_k^i are bounded if the matrix

$$\Psi = \text{diag} \left(\frac{\gamma_1}{k_{11}^2}, \dots, \frac{\gamma_M}{k_{MM}^2} \right) (K \odot K - \text{diag}(k_{11}^2, \dots, k_{MM}^2)) \quad (77)$$

²Note that, Theorem 8 in [29] can be directly applied when the constant β_i in (75) is replaced by a \mathcal{K}_∞ function of the initial conditions. Furthermore, Theorem 8 deals with global stability rather than \mathcal{L}_∞ -stability. However, a careful examination of the proof reveals that Theorem 8 holds with reference to \mathcal{L}_∞ -stability when β is constant, provided that the gain map (represented by matrix Ψ in our context) is linear. We also highlight that, although Theorem 8 is for continuous-time systems, Proposition 15 in [29] guarantees that it holds for discrete-time systems as well.

is Schur. In (77) the symbol \odot represents the element-wise matrix product.

To conclude the proof, we show that, under the assumptions of Theorem 3, it is possible to find a matrix K , compatible with the graph topology, such that Ψ is Schur.

First, from the graph $(\mathcal{V}, \mathcal{E})$, we derive a subgraph $\mathcal{G}^* = (\mathcal{V}, \mathcal{E}^*)$, by selecting edges $(i, j) \in \mathcal{E}^* \subseteq \mathcal{E}$ according to Algorithm 1.

By construction, the graph \mathcal{G}^* is a forest [30], i.e. a graph composed by a number of mutually disjoint trees. Moreover, the root of each tree is a regionally observable node while all other nodes are regionally unobservable. It follows that each row of the matrix K produced by Algorithm 1 has only one off-diagonal element that is different from zero³.

Up to a permutation of the node indexes, K is lower triangular (see, e.g. Fig. 2). It follows that the matrix Ψ defined in (77) is triangular, with zero diagonal entries and hence, for any choice of $\gamma_i, I = 1, \dots, M$, Ψ is Schur. This concludes the proof. ■

Remark 1: The matrix K generated by Algorithm 1 is lower triangular, up to a permutation of the node indexes. The same arguments of the above proof can be used to show, by continuity, that boundedness of $\Pi_{t-N|t-1}^i$ is guaranteed by any stochastic matrix \bar{K} compatible with $(\mathcal{V}, \mathcal{E})$ with: (i) the same diagonal elements of the matrix K ; (ii) arbitrary elements in the lower triangular part; (iii) sufficiently small elements in the upper triangular part so as to guarantee the matrix Ψ defined in (77) is Schur.

E. Proof of Corollary 2

Proof: Recall that, from Algorithm 1, K is lower triangular up to a permutation of the sensor indexes. Hence, $\mathbf{K} = K \otimes I_n$ is a block lower triangular matrix. Recalling that \mathbf{P}_{NO} and \mathbf{A} are block diagonal matrices, $\Phi = \mathbf{P}_{NO} \mathbf{K} \mathbf{A} \mathbf{P}_{NO}$ is a block lower triangular matrix as well. Accordingly, the eigenvalues of Φ correspond to the eigenvalues of the M diagonal blocks of Φ , denoted as $\Phi_i, i \in \mathcal{V}$, and defined as

$$\Phi_i = k_{ii} \bar{T}^i (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i (\bar{T}^i)^{-1} A \bar{T}^i (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i (\bar{T}^i)^{-1}$$

Let A_{KO}^i be defined as in (64). One has

$$\Phi_i = k_{ii} \bar{T}^i (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i A_{KO}^i (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i (\bar{T}^i)^{-1} \quad (78)$$

and according to the definition of \bar{S}_{NO}^i , one also obtains

$$\bar{A}_{KO}^i = (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i A_{KO}^i (\bar{S}_{NO}^i)^T \bar{S}_{NO}^i = \begin{bmatrix} 0 & 0 \\ 0 & A_{NO}^i \end{bmatrix}$$

Therefore, from (78), $\Phi_i = \bar{T}^i \frac{1}{k_{ii}} \bar{A}_{KO}^i (\bar{T}^i)^{-1}$. It is thus clear that the non-zero eigenvalues of Φ are also eigenvalues of A_{NO}^i , for some $i \in \mathcal{V}_{NO}$. If $i \in \mathcal{V}_O$, $\Phi_i = \bar{A}_{KO}^i = \mathbf{0}_{n \times n}$. On the other hand, if $i \in \mathcal{V}_{NO}$, recall that, from step 2 of Algorithm 1, we have $k_{ii} < \frac{1}{\sqrt{2\sigma^{(i)}(A)}}$. Therefore, $|\lambda_j(\Phi_i)| \leq k_{ii} \sigma^{(i)}(A) < \frac{1}{\sqrt{2}} < 1$, for all $j = 1, \dots, n$, for all $i \in \mathcal{V}_{NO}$. The Schureness of Φ then follows from the Schureness of Φ_i .

Finally, notice that the assumptions of Theorem 3 imply that, in all the isolated strongly connected subgraphs of \mathcal{G} , there is

³We also highlight that the matrix K produced by Algorithm 1 is compatible with the graph $(\mathcal{V}, \mathcal{E})$.

at least one observable node, and hence (33) holds. Therefore, by Theorem 2, Property 1 is verified. This concludes the proof. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous Reviewers for their constructive comments and criticism.

REFERENCES

- [1] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, no. 4, pp. 622 – 633, 2008.
- [2] P. Alriksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," in *Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, 2006.
- [3] R. Olfati-Saber and J. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," *Proc. 44th IEEE Conference on Decision and Control - European Control Conference*, pp. 6698 – 6703, 2005.
- [4] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," *Proc. 44th IEEE Conference on Decision and Control - European Control Conference*, pp. 8179 – 8184, 2005.
- [5] D. Spanos, R. Olfati-Saber, and R. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," *Fourth International Symposium on Information Processing in Sensor Networks*, pp. 133 – 139, 2005.
- [6] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," *Proc. 46th IEEE Conference on Decision and Control*, pp. 5492 – 5498, 2007.
- [7] M. Kamgarpour and C. Tomlin, "Convergence properties of a decentralized Kalman filter," *Proc. 47th IEEE Conference on Decision and Control*, pp. 3205 – 3210, 2008.
- [8] G. Calafiore and F. Abrate, "Distributed linear estimation over sensor networks," *International Journal of Control*, vol. 82, no. 5, pp. 868 – 882, 2009.
- [9] A. Speranzon, C. Fischione, K. H. Johansson, and A. L. Sangiovanni-Vincentelli, "A distributed minimum variance estimator for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 609–621, 2008.
- [10] H. Hashemipour, S. Roy, and A. Laub, "Decentralized structures for parallel Kalman filtering," *IEEE Trans. on Automatic Control*, vol. 33, no. 1, pp. 88 – 94, Jan. 1988.
- [11] B. Rao and H. Durrant-Whyte, "Fully decentralised algorithm for multisensor Kalman filtering," in *IEE Proc. on Control Theory and Applications*, D, vol. 138, no. 5, Sept. 1991, pp. 413 – 420.
- [12] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. on Signal Processing*, vol. 56, no. 10, pp. 4919 – 4935, Oct. 2008.
- [13] C. V. Rao, J. Rawlings, and J. Lee, "Constrained linear state estimation - a moving horizon approach," *Automatica*, vol. 37, pp. 1619 – 1628, 2001.
- [14] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for discrete-time linear systems," *IEEE Trans. on Automatic Control*, vol. 48, no. 3, pp. 473 – 478, 2003.
- [15] C. V. Rao, J. Rawlings, and D. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Trans. on Automatic Control*, vol. 48, no. 2, pp. 246 – 258, 2003.
- [16] C. V. Rao and J. Rawlings, "Nonlinear moving horizon state estimation," in *F. Allgöwer and A. Zheng, editors, Nonlinear Model Predictive Control, Progress in Systems and Control Theory, Birkhauser*, pp. 45–70, 2000.
- [17] A. Alessandri, M. Baglietto, and G. Battistelli, "Moving-horizon state estimation for nonlinear discrete-time systems: new stability results and approximation schemes," *Automatica*, vol. 44, pp. 1753 – 1765, 2008.
- [18] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems," *IEEE Trans. on Automatic Control*, vol. 44, no. 11, pp. 2028 – 2042, 1999.
- [19] G. Ferrari-Trecate, D. Mignone, and M. Morari, "Moving horizon estimation for hybrid systems," *IEEE Trans. on Automatic Control*, vol. 47, no. 10, pp. 1663 – 1676, Oct. 2002.
- [20] G. Goodwin, M. Seron, and J. De Doná, *Constrained Control and Estimation*. New Jersey: Springer, 2005.
- [21] L. Farina and S. Rinaldi, *Positive Linear Systems*. J. Wiley and Sons, 2000.
- [22] D. Bertsekas, *Linear network optimization: algorithms and codes*. MIT Press, 1991.
- [23] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "A moving horizon scheme for distributed state estimation," in *Proc. 1st IFAC Workshop on Estimation and Control of Networked Systems, Venice, Italy*, 2009, pp. 126 – 131.
- [24] C. V. Rao, J. Rawlings, and J. Lee, "Stability of constrained linear moving horizon estimation," *Proc. American Control Conference*, pp. 3387 – 3391, 1999.
- [25] J. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, England: Clarendon Press, 1965.
- [26] M. Verhaegen and P. Van Dooren, "Numerical aspects of different Kalman filter implementations," *IEEE Trans. on Automatic Control*, vol. 31, no. 10, pp. 907 – 917, 1986.
- [27] P. Caines and D. Mayne, "On the discrete-time matrix Riccati equation for optimal control," *International Journal of Control*, vol. 12, no. 5, pp. 785–794, 1970.
- [28] H. K. Khalil, *Nonlinear systems - Third edition*. Prentice Hall, 2000.
- [29] S. Dashkovskiy, B. S. Rüffer, and F. Wirth, "An ISS small gain theorem for iss general networks," *Mathematics of Control, Signals, and Systems*, vol. 19, no. 2, pp. 93–122, 2007.
- [30] B. Bollobás, *Modern graph theory*, ser. Graduate texts in mathematics. New York: Springer, 1998, vol. 184.



Marcello Farina was born in Treviglio, Italy, in 1978. He received the "Laurea" degree in Electronic Engineering in 2003 and the Ph.D. degree in Information Engineering in 2007, both from the Politecnico di Milano. In 2005 he was visiting student at the Institute for Systems Theory and Automatic Control, Stuttgart, Germany. He is presently assistant professor at Dipartimento di Elettronica e Informazione, Politecnico di Milano. His research interests include distributed and decentralized state estimation, modeling and identification in systems biology, modeling and control of energy supply systems.



Giancarlo Ferrari-Trecate received the "Laurea" degree in Computer Engineering in 1995 and the Ph.D. degree in Electronic and Computer Engineering in 1999, both from the University of Pavia. In spring 1998 he was visiting researcher at the Neural Computing Research Group, University of Birmingham, UK. In fall 1998 he joined the Automatic Control Laboratory at ETH (Zurich, Switzerland) as a Postdoctoral fellow. In 1999 he won the grant "assegno di ricerca" at the University of Pavia and in 2000 he was appointed Oberassistent at ETH. In 2002, he joined INRIA (Rocquencourt, France) as a Research Fellow. From March to October 2005 he worked at the Politecnico di Milano supported by the a researcher mobility grant from the Italian Ministry of Education, University and Research. From November 2005 he is associate professor at DIS, University of Pavia. His research interests include distributed and decentralized control, modeling and analysis of biochemical networks, hybrid systems and Bayesian learning.



Riccardo Scattolini Riccardo Scattolini was born in Milano, Italy, in 1956. He received the "Laurea" degree in electrical engineering from the Politecnico di Milano, Milano, Italy, in 1979. He is presently Professor of automatic control with the Politecnico di Milano. From 1984 to 1985, he was a Visiting Researcher at the Department of Engineering Science, Oxford University, Oxford, U.K. He also spent one year working in industry on the simulation and control of chemical plants. His current research interests include modeling, identification, simulation and control of industrial plants, with emphasis on model predictive control of large-scale systems. Dr. Scattolini received the Heaviside Premium from the Institution of Electrical Engineers, U.K., in 1991.