Stability of consensus extended Kalman filter for distributed state estimation

Stability of Consensus Extended Kalman Filter for Distributed State Estimation

Giorgio Battistelli [a] Luigi Chisci [a]

[a] *Università di Firenze, Dipartimento di Ingegneria dell'Informazione (DINFO), Via di Santa Marta 3, 50139 Firenze, Italy.*

Abstract

The paper addresses consensus-based networked estimation of the state of a nonlinear dynamical system. The focus is on a family of distributed state estimation algorithms which relies on the extended Kalman filter linearization paradigm. Consensus is exploited in order to fuse the information, both prior and novel, available in each network node. It is shown that the considered family of distributed Extended Kalman Filters enjoys local stability properties, under minimal requirements of network connectivity and system collective observability. A simulation case-study concerning target tracking with a network of nonlinear (angle and range) position sensors is worked out in order to show the effectiveness of the considered nonlinear consensus filter.

Key words: Networked systems; distributed state estimation; sensor fusion; nonlinear filters; stability analysis; consensus; Kalman filters.

1 Introduction

Consensus is a widely exploited tool for distributing computations over networks in a scalable way. An especially important application of consensus, which has recently received great attention, is networked state estimation, i.e., distributed estimation of the state of a dynamical system given measurements provided by a wireless sensor network. The literature on the subject is quite vast and includes approaches based on consensus Kalman filtering (Olfati-Saber, 2007; Kamgarpour and Tomlin, 2008; Kar and Moura, 2011; Cattivelli and Sayed, 2010; Li and Jia, 2012; Zhou et al., 2013; Wang et al., 2014), Luenberger-like consensus estimation (Stankovic et al., 2009; Matei and Baras, 2012; Millan et al., 2013, 2015), consensus H_∞ estimation (Ugrinovskii, 2011, 2013), distributed particle filtering (Mohammadi and Asif, 2013; Hlinka et al., 2013), and distributed moving-horizon estimation (Farina et al., 2010, 2012). The interested reader is referred to the above-cited papers as well so to the references therein for an overview of the different existing approaches. In the context of networked state estimation, the main challenge is to design distributed estimation algorithms that preserve as much as possible the stability, performance and robustness requirements of their centralized counterparts.

With this respect, significant advances have been made, in the last years, in the linear setting by developing distributed state estimation (DSE) algorithms able to guarantee stability under minimal requirements of network connectivity and system collective observability, i.e. observability from the whole network but not necessarily from individual sensors. Such algorithms include the *consensus on information* (CI) filter of Battistelli et al. (2011); Battistelli and Chisci (2014) and the *information weighted consensus filter* (ICF) of Kamal et al. (2012, 2013). The CI, in which consensus is carried out on the posterior information of the network nodes, can be interpreted in terms of consensus to the average of the local posteriors according to the pseudo-metric induced by the Kullback-Leibler average (Battistelli and Chisci, 2014). The ICF algorithm performs a consensus with a suitable weighting of the prior state and measurement information so as to ensure convergence to the centralized estimate as the number of consensus steps goes to infinity. Recently in (Battistelli et al., 2015), it was shown that both the CI and the ICF belong to a broader family of DSE algorithms, and a generalization

⋆ This paper is an expanded and revised version of "Stability of consensus extended Kalman filtering for distributed state estimation", presented at the 2014 IFAC World Congress held in Cape Town, South Africa.

Email addresses: `giorgio.battistelli@unifi.it` (Giorgio Battistelli), `luigi.chisci@unifi.it` (Luigi Chisci).

to a nonlinear setting was proposed by exploiting the Extended Kalman Filter (EKF) linearization argument. Hereafter, the family of DSE algorithms resulting from such a generalization will be referred to as Distributed EKFs (DEKFs).

The present paper provides a contribution by proving that the family of DSE algorithms of Battistelli et al. (2015) enjoy nice stability properties also in the more general nonlinear setting, provided that, similarly to the linear case, suitable connectivity and collective observability assumptions hold. In the lines of the classical results on stability of centralized EKF (La Scala et al., 1995; Reif and Unbehauen, 1999; Reif et al., 1999), the stability analysis is based on the idea of writing the estimation error dynamics in a suitable way so that the linearized part is separated from the nonlinear (higher-order) terms. Then, the stability of the linear part of the estimation error dynamics can be analyzed via Lyapunov-like methods, and local stability results can be derived for the overall estimation error dynamics. As a further contribution, an explicit connection is established between the boundedness of the filter covariance matrix and the invertibility of the collective observability mapping.

Thanks to this result, the considered family of DEKFs emerges as an effective tool for the solution of many practically relevant distributed nonlinear filtering problems like, e.g., distributed tracking of a moving object given measurements of angle, range and/or Doppler wireless communicating sensors spread over the area of interest; such sensors, in fact, are highly nonlinear and unable to individually guarantee observability.

The rest of the paper is organised as follows. Section 2 introduces the problem setting. Section 3 describes the considered family of DEKF algorithms for networked state estimation and section 4 analyses its stability properties. Section 5 demonstrates, via simulation experiments, the effectiveness of such a consensus filter in a nonlinear target tracking case-study. Section 6 ends the paper with concluding remarks. All mathematical proofs are reported in the appendix.

2 Problem setting

This paper addresses Distributed State Estimation (DSE) over a sensor network consisting of two types of nodes: *communication* nodes have only *processing* and *communication* capabilities, i.e. they can process local data as well as exchange data with neighboring nodes, while *sensor* nodes have also *sensing* capabilities, i.e. they can sense data from the environment. Notice that communication nodes are introduced to act as "relays" of information whenever sensor nodes are too far away to communicate. For insights on the importance of considering the effect of communication nodes when studying the properties of a distributed state estimation algorithm we refer the reader to (Kamal et al., 2013; Wang et al., 2014)(where this type of nodes are referred to as "naive nodes"). In the sequel, the network will be denoted by the triplet $(\mathcal{S}, \mathcal{C}, \mathcal{A})$ where: $\mathcal{S}$ is the set of sensor nodes, $\mathcal{C}$ the set of communication nodes, $\mathcal{N} = \mathcal{S} \bigcup \mathcal{C}$, $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of arcs (connections) such that $(i, j) \in \mathcal{A}$ if node j can receive data from node i (clearly $(i, i) \in \mathcal{A}$ for all $i \in \mathcal{N}$). Further, for each node $i \in \mathcal{N}$, $\mathcal{N}^i$ will denote the set of its in-neighbors (including i itself), i.e. $\mathcal{N}^i \triangleq \{j : (j, i) \in \mathcal{A}\}$.

The DSE problem over the sensor network $(\mathcal{S}, \mathcal{C}, \mathcal{A})$ can be formulated as follows. Consider a dynamical system

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{w}_t \tag{1}$$

and a set of sensors $\mathcal{S}$ with measurement equations

$$\mathbf{y}_t^i = \mathbf{h}^i(\mathbf{x}_t) + \mathbf{v}_t^i\,, \qquad i \in \mathcal{S}\,. \tag{2}$$

Notice that the above measurement equation is defined only for sensor nodes, since no measurement is supposed to be collected by the communication nodes. Then the objective is to have, at each time $t \in \{1, 2, \dots\}$ and in each node $i \in \mathcal{N}$, an estimate $\hat{\mathbf{x}}_{t|t}$ of the state $\mathbf{x}_t$ constructed only on the basis of the local measurements (when available) and of data received from all adjacent nodes $j \in \mathcal{N}^i \backslash \{i\}$.

2.1 Centralized Extended Kalman Filter

Before describing the family of DEKF algorithms under consideration, it is convenient to briefly recall the equations of the centralized Extended Kalman Filter, which is assumed to simultaneously process all measurements $\{y_k^i, i \in \mathcal{S}\}$. Hereafter, for convenience, the *information filter* form will be adopted. The information filter propagates, instead of the estimate $\hat{\mathbf{x}}_{t|t-1}$ and covariance $\mathbf{P}_{t|t-1}$, the *information* (inverse covariance) matrices

$$\mathbf{\Omega}_{t|t-1} \triangleq \mathbf{P}_{t|t-1}^{-1}, \quad \mathbf{\Omega}_{t|t} \triangleq \mathbf{P}_{t|t}^{-1}$$

and the vectors

$$\mathbf{q}_{t|t-1} \triangleq \mathbf{P}_{t|t-1}^{-1} \hat{\mathbf{x}}_{t|t-1}, \quad \mathbf{q}_{t|t} \triangleq \mathbf{P}_{t|t}^{-1} \hat{\mathbf{x}}_{t|t}$$

that will be referred to as *information vectors*. Then, the recursive information filter of Table 1 can be derived (Battistelli et al., 2015), where $\mathbf{W}$ and $\mathbf{V}^i$, $i \in \mathcal{N}$, are given positive definite matrices. A typical choice for such matrices is to take $\mathbf{W}$ as an estimate of the inverse covariance of the process disturbance $\mathbf{w}_t$, and each $\mathbf{V}^i$ as an estimate of the inverse covariance of the measurement noise $\mathbf{v}_t^i$ affecting the i-th sensor. Notice, however,

Table 1
Information CEKF Algorithm, to be implemented at each sampling interval $t = 1, 2, \ldots$ starting from initial conditions $\hat{\mathbf{x}}_{1|0}$, $\mathbf{\Omega}_{1|0}$, $\mathbf{q}_{1|0} = \mathbf{\Omega}_{1|0}\ \hat{\mathbf{x}}_{1|0}$.

Correction (measurement-update):

$$\mathbf{C}_t^i = \frac{\partial \mathbf{h}^i}{\partial \mathbf{x}}\left(\hat{\mathbf{x}}_{t|t-1}\right), \qquad i \in \mathcal{S}$$
$$\mathbf{\Omega}_{t|t} = \mathbf{\Omega}_{t|t-1} + \sum_{i \in \mathcal{S}} \left(\mathbf{C}_t^i\right)^\top \mathbf{V}^i\, \mathbf{C}_t^i$$
$$\overline{\mathbf{y}}_t^i = \mathbf{y}_t^i - \mathbf{h}^i\left(\hat{\mathbf{x}}_{t|t-1}\right) + \mathbf{C}_t^i \hat{\mathbf{x}}_{t|t-1}, \quad i \in \mathcal{S}$$
$$\mathbf{q}_{t|t} = \mathbf{q}_{t|t-1} + \sum_{i \in \mathcal{S}} \left(\mathbf{C}_t^i\right)^\top \mathbf{V}^i\, \overline{\mathbf{y}}_t^i$$

Prediction (time-update):

$$\hat{\mathbf{x}}_{t|t} = \mathbf{\Omega}_{t|t}^{-1}\mathbf{q}_{t|t}, \quad \text{and} \quad \mathbf{A}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\left(\hat{\mathbf{x}}_{t|t}\right)$$
$$\mathbf{\Omega}_{t+1|t} = \mathbf{W} - \mathbf{W}\mathbf{A}_t\left(\mathbf{\Omega}_{t|t} + \mathbf{A}_t^\top \mathbf{W}\mathbf{A}_t\right)^{-1}\mathbf{A}_t^\top \mathbf{W}$$
$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{f}\left(\hat{\mathbf{x}}_{t|t}\right), \quad \text{and} \quad \mathbf{q}_{t+1|t} = \mathbf{\Omega}_{t+1|t}\ \hat{\mathbf{x}}_{t+1|t}$$

that a specific choice of such matrices is immaterial for the subsequent developments.

The algorithm of Table 1 generalizes the *Information Kalman Filter* algorithm, corresponding to $\mathbf{f}(\mathbf{x}) = \mathbf{A}_t\mathbf{x}$ and $\mathbf{h}^i(\mathbf{x}) = \mathbf{C}_t^i\mathbf{x}$, to nonlinear systems (1) and/or sensors (2) via the Extended Kalman Filter paradigm of linearizing the state and measurement equations around the current estimate. With this respect, the following assumption is needed.

A1. The functions $\mathbf{f}$ and $\mathbf{h}^i$, $i \in \mathcal{S}$, are twice continuously differentiable on $\mathbb{R}^n$, where $n = \dim(\mathbf{x})$.

Notice that, in order to streamline the presentation, here and in the following it is supposed that the functions $\mathbf{f}$ and $\mathbf{h}^i$, $i \in \mathcal{S}$, are defined over the whole $\mathbb{R}^n$. However, all the results presented hereafter could be suitably modified to account for the case when the system trajectories are confined to a given set $\mathcal{X} \subset \mathbb{R}^n$.

3 Distributed Extended Kalman Filter

The focus of this paper is on a family of DSE algorithms proposed in (Battistelli et al., 2015) wherein each network node runs a local EKF and information is spread through the network by means of consensus. As well known, consensus is a widely exploited tool for distributed computation (e.g. minimization, maximization, averaging, etc.) over a network. The basic idea of consensus is to perform a collective computation over the whole network by iterating, in each node of the network, regional computations of the same type involving only the subset of neighboring nodes. In the specific context of this paper, consensus is applied to fuse the information on the state $\mathbf{x}_t$ available in each node with the one received from the neighbors.

To this end, let us assume that at time t each node $i \in \mathcal{N}$ be provided with a local information pair $\left(\mathbf{\Omega}_{t|t-1}^i, \mathbf{q}_{t|t-1}^i\right)$. Then, after a novel measurement $\mathbf{y}_t^i$ has been collected, the information available in node i consists of two information pairs:

(i) the information pair $\left(\mathbf{\Omega}_{t|t-1}^i, \mathbf{q}_{t|t-1}^i\right)$ which represents the *prior information*;
(ii) the information pair $\left(\delta\mathbf{\Omega}_t^i, \delta\mathbf{q}_t^i\right)$ which represents the *novel information*.

Notice that, for a generic sensor node $i \in \mathcal{S}$, one has $\delta\mathbf{\Omega}_t^i = \left(\mathbf{C}_t^i\right)^\top \mathbf{V}_t^i\mathbf{C}_t^i$ and $\delta\mathbf{q}_t^i = \left(\mathbf{C}_t^i\right)^\top \mathbf{V}_t^i\overline{\mathbf{y}}_t^i$, where the linearized output matrices $\mathbf{C}_t^i$, and hence the virtual measurements $\overline{\mathbf{y}}_t^i$, have to be redefined in terms of the local state predictions $\hat{\mathbf{x}}_{t|t-1}^i$ instead of the centralized one $\hat{\mathbf{x}}_{t|t-1}$, which is not available in a distributed setting. For each communication node $i \in \mathcal{C}$, we simply let $\delta\mathbf{q}_t^i = 0$ and $\delta\mathbf{\Omega}_t^i = 0$.

In the considered family of DEKFs, before combining the prior information with the novel one, two parallel consensus algorithms are carried out by iterating a certain number, say L, of regional averages on the two information pairs (prior and novel). Specifically, the consensus on prior information takes the form

$$\begin{aligned}
\mathbf{q}_{t|t-1}^i(\ell+1) &= \sum_{j \in \mathcal{N}^i} \pi^{i,j}\ \mathbf{q}_{t|t-1}^j(\ell) \\
\mathbf{\Omega}_{t|t-1}^i(\ell+1) &= \sum_{j \in \mathcal{N}^i} \pi^{i,j}\ \mathbf{\Omega}_{t|t-1}^j(\ell)
\end{aligned} \tag{3}$$

for $\ell = 0, 1, \ldots, L-1$ with the initialization $\mathbf{q}_{t|t-1}^i(0) = \mathbf{q}_{t|t-1}^i$ and $\mathbf{\Omega}_{t|t-1}^i(0) = \mathbf{\Omega}_{t|t-1}^i$. As for the novel information, L consensus steps of the type

$$\begin{aligned}
\delta\mathbf{q}_t^i(\ell+1) &= \sum_{j \in \mathcal{N}^i} \pi^{i,j}\ \delta\mathbf{q}_t^j(\ell) \\
\delta\mathbf{\Omega}_t^i(\ell+1) &= \sum_{j \in \mathcal{N}^i} \pi^{i,j}\ \delta\mathbf{\Omega}_t^j(\ell)
\end{aligned} \tag{4}$$

are performed, where $\ell = 0, 1, \ldots, L-1$. For each sensor node $i \in \mathcal{S}$, the initial vector $\delta\mathbf{q}_t^i(0)$ and the initial matrix $\delta\mathbf{\Omega}_t^i(0)$ are set equal to $\delta\mathbf{q}_t^i$ and $\delta\mathbf{\Omega}_t^i$, respectively.

Notice that in both algorithms, in each consensus iteration, each node i computes a regional average, that is a combination of the values in $\mathcal{N}^i$ with suitable consensus weights $\pi^{i,j}$, $j \in \mathcal{N}^i$. In this paper, a convex combination is adopted by supposing $\pi^{i,j} \geq 0$ and $\sum_{j \in \mathcal{N}^i} \pi^{i,j} = 1, \forall i \in \mathcal{N}$.

After the two consensus iterations have been carried out, the two fused information pairs $\left(\mathbf{\Omega}_{t|t-1}^i(L), \mathbf{q}_{t|t-1}^i(L)\right)$

Table 2
DEKF Algorithm

Compute the local correction terms
if $i \in \mathcal{S}$ **then**
sample the measurement $\mathbf{y}_t^i$
$\mathbf{C}_t^i = \dfrac{\partial \mathbf{h}^i}{\partial \mathbf{x}}\left(\hat{\mathbf{x}}_{t|t-1}^i\right)$
$\overline{\mathbf{y}}_t^i = \mathbf{y}_t^i - \mathbf{h}^i\left(\hat{\mathbf{x}}_{t|t-1}^i\right) + \mathbf{C}_t^i \hat{\mathbf{x}}_{t|t-1}^i$
$\delta \mathbf{q}_t^i = \left(\mathbf{C}_t^i\right)^\top \mathbf{V}_t^i \overline{\mathbf{y}}_t^i$
$\delta \mathbf{\Omega}_t^i = \left(\mathbf{C}_t^i\right)^\top \mathbf{V}^i \mathbf{C}_t^i$
else
$\delta \mathbf{q}_t^i = \mathbf{0}$, and $\delta \mathbf{\Omega}_t^i = \mathbf{0}$
end if
Consensus:
$\delta \mathbf{q}_t^i(0) = \delta \mathbf{q}_t^i, \quad \delta \mathbf{\Omega}_t^i(0) = \delta \mathbf{\Omega}_t^i,$
$\mathbf{q}_{t|t-1}^i(0) = \mathbf{q}_{t|t-1}^i, \quad \mathbf{\Omega}_{t|t-1}^i(0) = \mathbf{\Omega}_{t|t-1}^i,$
for $\ell = 0, 1, \ldots, L-1$ **do**
Fuse the quantities $\delta \mathbf{q}_t^j(\ell)$ and $\delta \mathbf{\Omega}_t^j(\ell)$ as in (4)
and the quantities $\mathbf{q}_{t|t-1}^j(\ell)$ and $\mathbf{\Omega}_{t|t-1}^j(\ell)$ as in (3)
end for
Correction:
$\mathbf{q}_{t|t}^i = \mathbf{q}_{t|t-1}^i(L) + \gamma_t^i \, \delta \mathbf{q}_t^i(L)$
$\mathbf{\Omega}_{t|t}^i = \mathbf{\Omega}_{t|t-1}^i(L) + \gamma_t^i \, \delta \mathbf{\Omega}_t^i(L)$
$\hat{\mathbf{x}}_{t|t}^i = \left(\mathbf{\Omega}_{t|t}^i\right)^{-1} \mathbf{q}_{t|t}^i$
Prediction:
$\hat{\mathbf{x}}_{t+1|t}^i = \mathbf{f}\left(\hat{\mathbf{x}}_{t|t}^i\right)$, and $\mathbf{A}_t^i = \dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}\left(\hat{\mathbf{x}}_{t|t}^i\right)$
$\mathbf{\Omega}_{t+1|t}^i = \mathbf{W} - \mathbf{W}\mathbf{A}_t^i \left(\mathbf{\Omega}_{t|t}^i + (\mathbf{A}_t^i)^\top \mathbf{W} \mathbf{A}_t^i\right)^{-1} (\mathbf{A}_t^i)^\top \mathbf{W}$
$\mathbf{q}_{t+1|t}^i = \mathbf{\Omega}_{t+1|t}^i \hat{\mathbf{x}}_{t+1|t}^i$

and $\left(\delta \mathbf{\Omega}_t^i(L), \delta \mathbf{q}_t^i(L)\right)$ are suitably combined in the correction step. Summing up, the DEKF algorithm of Table 2 is obtained. It is worth noting that, in the DEKF algorithm, each network node performs three main operations: EKF prediction, computation of the local correction term, and consensus. The first two operations make use only of the local information and hence have the same complexity of a local EKF. Further, the complexity of the consensus iteration depends on the number of neighbors and not on the total number $|\mathcal{N}|$ of network nodes. These properties, together with the fact (shown in the following) that stability is guaranteed for any number L of consensus steps irrespective of the network size, ensure the scalability of the approach. A few remarks on the considered algorithm are in order.

Remark 1 *First of all notice that, in the correction step of each network node i, the fused novel information resulting from the consensus iteration is multiplied by some suitable scalar weight γ_t^i. Hence, Table 2 actually provides a family of distributed filters corresponding to different choices of the scalar weights γ_t^i. For instance, when $\gamma_t^i = 1$, the consensus on information filter of Battistelli and Chisci (2014); Battistelli et al. (2011) is retrieved. If instead γ_t^i is taken equal to $|\mathcal{N}|$, the number of nodes in the network, then the resulting CEKF algorithm coincides with the EKF-based generalization of the information weighted consensus proposed by Kamal et al. (2012, 2013) in a linear setting.*

Remark 2 *When a large number of consensus steps per sampling interval is performed, consensus on novel information can be sufficient to ensure stability and performance. This is precisely the idea exploited in the approaches of Olfati-Saber (2007); Kamgarpour and Tomlin (2008); Li and Jia (2012) wherein consensus is performed only on the novel information so as to approximate, in a distributed way, the correction step of the centralized filter. However, this kind of approaches has an intrinsic limitation in the fact that stability can be guaranteed only when a sufficiently large number L of consensus steps is performed, so that the local information provided by the innovation pairs can spread throughout the whole network (since in this case information collected at time t is spread only at time t). See also (Battistelli and Chisci, 2014; Battistelli et al., 2015) for a discussion on this issue. On the other hand, by performing a consensus also on the prior information, the information collected at time t is spread also in the subsequent sampling intervals $t+1, t+2, \ldots$ through the predictions. Hence, as proved in the subsequent part of the paper, we can have stability for any number L of consensus steps.*

Remark 3 *As discussed by Battistelli and Chisci (2014), in the linear Gaussian case the consensus on the prior information has a meaningful interpretation, from the information-theoretic point of view, as a consensus among the local Gaussian prior probability density functions (PDFs) in the pseudo-metric defined by the Kullback-Leibler divergence. The EKF-paradigm allows to readily extend such concepts to a nonlinear setting by approximating each local prior PDF with a Gaussian (see also Battistelli et al., 2014).*

Remark 4 *The scalar weights γ_t^i are introduced in order to counteract the possible underweighting of novel information. To elaborate more on this issue, notice that in the correction step of the centralized EKF the novel information is represented by the pair $(\partial \mathbf{\Omega}_t, \partial \mathbf{q}_t)$ where $\partial \mathbf{\Omega}_t = \sum_{i \in \mathcal{S}} \left(\mathbf{C}_t^i\right)^\top \mathbf{V}_t^i \mathbf{C}_t^i$ and $\partial \mathbf{q}_t = \sum_{i \in \mathcal{S}} \left(\mathbf{C}_t^i\right)^\top \mathbf{V}_t^i \overline{\mathbf{y}}_t^i$. On the other hand, even assuming a balanced choice of the consensus weights, consensus on the novel information would provide at convergence the averages $\partial \mathbf{\Omega}_t / |\mathcal{N}|$ and $\partial \mathbf{q}_t / |\mathcal{N}|$. This implies the necessity for rescaling the consensus outcome by means of multiplication with the scalar weight γ_t^i. For example, the choice $\gamma_t^i = |\mathcal{N}|$, which has been proposed by Kamal et al. (2012, 2013), has the appealing feature of giving rise to a distributed algorithm converging to the centralized one as L tends to infinity. However, when only a moderate number of consensus steps per sampling interval can be afforded, other choices can be more convenient. For a discussion on this issue as well as an alternative choice for the weights γ_t^i, the interested reader is referred to Battistelli et al. (2014, 2015).*

Remark 5 *Whenever the weights γ_t^i are node-independent, i.e. $\gamma_t^i = \gamma_t$, $\forall i \in \mathcal{N}$, it is possible to perform jointly* *the two parallel consensus algorithms of Table 2 so as to save bandwidth by considering the combined information pair* $\left(\boldsymbol{\Omega}^i_{t|t-1} + \gamma_t \delta \boldsymbol{\Omega}^i_t, \mathbf{q}^i_{t|t-1} + \gamma_t \delta \mathbf{q}^i_t\right)$.

For the purposes of this paper, the particular choice of the scalars γ^i_t is immaterial as it is just sufficient to make the following assumption.

A2. There exist two positive scalars $\underline{\gamma}$ and $\bar{\gamma}$ such that $0 < \underline{\gamma} \le \gamma^i_t \le \bar{\gamma}$, for any $i \in \mathcal{N}$ and $t \ge 0$.

4 Stability analysis

In this section, the stability properties of the DEKF algorithm of Section 3 are analyzed. To this end, notice first that under assumption A1 the function $\mathbf{f}$ can be expanded as

$$\mathbf{f}(\mathbf{x}_t) - \mathbf{f}(\hat{\mathbf{x}}^i_{t|t}) = \mathbf{A}^i_t(\mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t}) + \boldsymbol{\varphi}(\mathbf{x}_t, \hat{\mathbf{x}}^i_{t|t}) \tag{5}$$

with $\mathbf{A}^i_t$ as in the DEKF algorithm and $\boldsymbol{\varphi}(\cdot)$ a suitable continuous function going to zero as $\hat{\mathbf{x}}_{t|t}$ tends to $\mathbf{x}_t$. Similarly, each function $\mathbf{h}^i$, $i \in \mathcal{S}$, can be expanded as

$$\mathbf{h}^i(\mathbf{x}_t) - \mathbf{h}^i(\hat{\mathbf{x}}^i_{t|t-1}) = \mathbf{C}^i_t(\mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t-1}) + \boldsymbol{\chi}^i(\mathbf{x}_t, \hat{\mathbf{x}}^i_{t|t-1}) \tag{6}$$

with $\mathbf{C}^i_t$ as in the DEKF algorithm and $\boldsymbol{\chi}^i(\cdot)$ a suitable continuous function going to zero as $\hat{\mathbf{x}}^i_{t|t-1}$ tends to $\mathbf{x}_t$. Here, the functions $\boldsymbol{\varphi}$ and $\boldsymbol{\chi}^i$ in (5) and (6) represent the remainders of the Taylor expansion of $\mathbf{f}$ and, respectively, $\mathbf{h}^i$ (see also Reif et al., 1999). By exploiting such expansions, it is possible to write the estimation error dynamics so that the linearized part is separated from the nonlinear terms. To this end, let us denote by $\boldsymbol{\Pi}$ the consensus matrix, whose elements are the consensus weights $\pi^{i,j}$ for any $i,j \in \mathcal{N}$. Further, let $\pi^{i,j}_\ell$ be the (i,j)-th element of $\boldsymbol{\Pi}^\ell$, i.e. the ℓ-th power of the consensus matrix $\boldsymbol{\Pi}$. Then the following result holds.

Proposition 1 *Let assumptions A1-A2 hold and let the DEKF algorithm be initialized at time $t=1$ with positive definite information matrices $\boldsymbol{\Omega}^i_{1|0}$. Then, for any i and any t, the matrices $\boldsymbol{\Omega}^i_{t|t}$ are invertible and the estimation errors $\mathbf{e}^i_t = \mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t-1}$ obey the recursion*

$$\mathbf{e}^i_{t+1} = \sum_{j\in\mathcal{N}} \boldsymbol{\Phi}^{i,j}_t \mathbf{e}^j_t + \mathbf{r}^i_t + \mathbf{s}^i_t \tag{7}$$

where

$$\begin{aligned}
\boldsymbol{\Phi}^{i,j}_t &= \pi^{i,j}_L\, \mathbf{A}^i_t \left(\boldsymbol{\Omega}^i_{t|t}\right)^{-1} \boldsymbol{\Omega}^j_{t|t-1}\,, \\
\mathbf{r}^i_t &= \boldsymbol{\varphi}(\mathbf{x}_t, \hat{\mathbf{x}}^i_{t|t}) \\
&\quad + \sum_{j\in\mathcal{S}} \pi^{i,j}_L \gamma^i_t \mathbf{A}^i_t \left(\boldsymbol{\Omega}^i_{t|t}\right)^{-1} (\mathbf{C}^j_t)^\top \mathbf{V}^j \boldsymbol{\chi}^j(\mathbf{x}_t, \hat{\mathbf{x}}^j_{t|t-1})\,, \\
\mathbf{s}^i_t &= \mathbf{w}_t - \sum_{j\in\mathcal{S}} \pi^{i,j}_L \gamma^i_t \mathbf{A}^i_t \left(\boldsymbol{\Omega}^i_{t|t}\right)^{-1} (\mathbf{C}^j_t)^\top \mathbf{V}^j \mathbf{v}^j_t\,.
\end{aligned}$$

□

In order to study the stability of the estimation error dynamics (7), the following assumption on the consensus weights is needed.

A3. The consensus matrix $\boldsymbol{\Pi}$ is row stochastic and primitive [1].

Notice that assumption A3 can always be satisfied provided that the network is connected. For instance, in this case, the Metropolis weights (Xiao et al., 2005; Calafiore and Abrate, 2009) satisfy A3. While taking the consensus matrix $\boldsymbol{\Pi}$ row stochastic is sufficient for stability, a doubly stochastic $\boldsymbol{\Pi}$ would also ensure that all the elements of $\boldsymbol{\Pi}^L$ tends to $1/|\mathcal{N}|$ as $L \to +\infty$.

Let now $\mathbf{p}$ denote the Perron-Frobenius left eigenvector of the matrix $\boldsymbol{\Pi}^L$ and let p^i denote its i-th component. Further, consider the candidate Lyapunov function

$$\mathscr{V}_t(\mathbf{e}_t) = \sum_{i\in\mathcal{N}} p^i \left(\mathbf{e}^i_t\right)^\top \boldsymbol{\Omega}^i_{t|t-1} \mathbf{e}^i_t \tag{8}$$

for the overall estimation error dynamics, where $\mathbf{e}_t = \mathrm{col}\left(\mathbf{e}^i_t,\, i \in \mathcal{N}\right)$. Notice that, by virtue of assumption A3, the eigenvector $\mathbf{p}$ has strictly positive components $p^i, i \in \mathcal{N}$, and satisfies the equation $\mathbf{p}^\top \boldsymbol{\Pi}^L = \mathbf{p}^\top$, i.e., $\sum_{j\in\mathcal{N}} p^j \pi^{j,i}_L = p^i$. The following result concerning the linearized part of the error dynamics can now be stated.

Lemma 1 *Let assumptions A1-A3 be satisfied. Further, suppose that the following conditions hold:*

i) there exist nonnegative reals $\bar{a}$ and $\bar{c}$ such that

$$\|\mathbf{A}^i_t\| \le \bar{a}, \quad \|\mathbf{C}^i_t\| \le \bar{c}$$

for any i and any t;

[1] Recall that a non-negative square matrix $\boldsymbol{\Pi}$ is row stochastic if all its rows sum up to 1. Further, it is primitive if there exists an integer m such that all the elements of $\boldsymbol{\Pi}^m$ are strictly positive.

ii) there exist positive reals $\underline{\omega}$, $\bar{\omega}$ such that

$$0 < \underline{\omega}\mathbf{I} \le \mathbf{\Omega}^i_{t|t} \le \bar{\omega}\mathbf{I}$$

for any i and any t;

iii) the matrix $\mathbf{A}^i_t$ is invertible for any i and any t.

Then, there exists a nonnegative scalar $\tilde{\beta} < 1$ such that, for any t, the candidate Lyapunov function defined in (8) satisfies

$$\mathscr{V}_{t+1}(\mathbf{\Phi}_t \mathbf{e}_t) \le \tilde{\beta}\, \mathscr{V}_t(\mathbf{e}_t)$$

where $\mathbf{\Phi}_t$ is the block matrix whose block elements are given by the matrices $\mathbf{\Phi}^{i,j}_t$ defined in Proposition 1. $\square$

Notice that conditions (i)-(iii) are taken preliminarily as hypotheses for deriving Lemma 1, which has to be regarded as an intermediate step for obtaining the main stability result of the paper which is Theorem 2. As shown in Section 4.1, such conditions automatically hold provided that system (1)-(2) enjoys some basic regularity and observability assumptions. For instance, condition i) holds when the functions $\mathbf{f}$ and $\mathbf{h}^i$, $i \in \mathcal{S}$, are globally Lipschitz or, in view of assumption A1, when the estimated trajectories $\hat{\mathbf{x}}^i_{k|k}$, $i \in \mathcal{N}$, are bounded. Further, condition ii) is closely related to the collective observability of the state $\mathbf{x}_t$ from the measurements $\mathbf{y}^i_t$, $i \in \mathcal{S}$, collected by all the available sensors, while condition iii) is related to the time reversibility of the system dynamics. For the sake of clarity, we defer the detailed discussion on these conditions to Section 4.1 and proceed now to derive another intermediate result concerning the overall estimation error dynamics.

To this end, recall that the functions $\boldsymbol{\varphi}$ and $\boldsymbol{\chi}^i$ in (5) and (6) represent the remainders of the Taylor expansion of $\mathbf{f}$ and, respectively, $\mathbf{h}^i$ and hence, under suitable assumptions, go to zero with order of convergence greater than 1. With this respect, in the lines of Reif et al. (1999), the following assumption is made.

A4. There exist positive reals $\epsilon_\varphi, \kappa_\varphi, \epsilon_{\chi^i}, \kappa_{\chi^i}$, $i \in \mathcal{S}$, such that the nonlinear functions $\boldsymbol{\varphi}$ and $\boldsymbol{\chi}^i$ in (5) and (6), respectively, are bounded as

$$\|\boldsymbol{\varphi}(\mathbf{x}, \hat{\mathbf{x}})\| \le \kappa_\varphi \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \tag{9}$$

$$\|\boldsymbol{\chi}^i(\mathbf{x}, \hat{\mathbf{x}})\| \le \kappa_{\chi^i} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \tag{10}$$

for any pair $\mathbf{x}, \hat{\mathbf{x}} \in \mathbb{R}^n$ such that $\|\mathbf{x} - \hat{\mathbf{x}}\| \le \epsilon_\varphi$ and $\|\mathbf{x} - \hat{\mathbf{x}}\| \le \epsilon_{\chi^i}$, respectively.

By exploiting Lemma 1 and assumption A4 the following local stability result can be derived.

Theorem 1 *Let assumptions A1-A4 be satisfied. Further, suppose that conditions i)-iii) of Lemma 1 hold. Then, the estimation error $\mathbf{e}^i_t$ turns out to be bounded in all the network nodes, i.e., there exists a positive real ϵ such that*

$$\limsup_{t\to\infty} \|\mathbf{e}^i_t\| \le \epsilon \tag{11}$$

for any i, provided that the initial estimation errors satisfy

$$\|\mathbf{e}^i_1\| \le \epsilon_0 \tag{12}$$

for some suitable constant $\epsilon_0 > 0$ and the disturbances satisfy

$$\|\mathbf{w}_t\| \le \epsilon_w, \quad \|\mathbf{v}^i_t\| \le \epsilon_{v^i},\ i \in \mathcal{S} \tag{13}$$

for suitable constants $\epsilon_w > 0$ and $\epsilon_{v^i} > 0$, $i \in \mathcal{S}$. $\square$

It is worth noting that, when the disturbance amplitudes $\epsilon_w > 0$ and $\epsilon_{v^i} > 0$ decrease, the asymptotic bound ϵ decreases as well and, in particular, the following corollary to Theorem 1 holds.

Corollary 1 *Let the system dynamics (1) and the measurement equations (2) be noise-free, i.e.,*

$$\mathbf{w}_t = 0, \quad \mathbf{v}^i_t = 0$$

for any i and any t. Then, under the same assumptions of Theorem 1, the estimation error goes to zero in all the network nodes, i.e.,

$$\lim_{t\to\infty} \|\mathbf{e}^i_t\| = 0$$

for any i, provided that the initial estimation errors satisfy

$$\|\mathbf{e}^i_1\| \le \epsilon_0$$

for some suitable constant $\epsilon_0 > 0$. $\square$

4.1 Connection with collective observability

This section is devoted to discussing how conditions i)-iii) of Lemma 1 can be related to specific properties of system (1)-(2). To this end, let $\mathbf{h} = \mathrm{col}(\mathbf{h}^i,\ i \in \mathcal{S})$ be the collective output function, and let $\mathbf{F}^{[M]}(\mathbf{x})$ be the collective observability mapping defined over a time window of length M, i.e.

$$\mathbf{F}^{[M]}(\mathbf{x}) = \begin{bmatrix} \mathbf{h}(\mathbf{x}) \\ \mathbf{h} \circ \mathbf{f}(\mathbf{x}) \\ \vdots \\ \mathbf{h} \circ \underbrace{\mathbf{f} \circ \cdots \circ \mathbf{f}}_{M \text{ times}}(\mathbf{x}) \end{bmatrix}$$

where $\circ$ denotes composition. In words, given a time window $\{t-M, \ldots, t\}$, $\mathbf{F}^{[M]}(\mathbf{x})$ coincides with the mapping from the state $\mathbf{x}$ at time $t-M$ to the vector made up of the noise-free collective outputs at times $t-M, \ldots, t$.

Supposing that the system trajectory lies within some compact set $\mathcal{X}$, the following assumptions are now needed.

A5. For any $\mathbf{x} \in \mathcal{X}$, $\partial \mathbf{f}(\mathbf{x})/\partial \mathbf{x}$ is non-singular.
A6. There exist a positive integer M such that, for any $\mathbf{x} \in \mathcal{X}$, $\operatorname{rank}\left\{\partial \mathbf{F}^{[M]}(\mathbf{x})/\partial \mathbf{x}\right\} = n$ where $n = \dim(\mathbf{x})$.

Notice that assumption A5 amounts to requiring that the state transition function $\mathbf{f}(\mathbf{x})$ is a diffeomorphism on $\mathcal{X}$ and, hence, reversible. Further, as well known, assumption A6 ensures that *collective observability*, in the sense of the invertibility of the mapping $\mathbf{F}^{[M]}(\mathbf{x})$, holds.

The following intermediate result can now be stated.

Lemma 2 *Let the system trajectory belong to a compact set $\mathcal{X}$, i.e. $\{\mathbf{x}_t\} \subset \mathcal{X}$, and suppose that assumptions A1-A6 are satisfied and that the DEKF algorithm is initialized at time $t = 1$ with positive definite information matrices $\mathbf{\Omega}^i_{1|0}$. Then, conditions i)-iii) of Lemma 1 hold provided that, for any i and t, the estimation errors satisfy*

$$\|\mathbf{e}^i_t\| \leq \hat{\epsilon} \tag{14}$$

for some suitable constant $\hat{\epsilon}$ and the disturbances satisfy

$$\|\mathbf{w}_t\| \leq \hat{\epsilon}_w, \quad \|\mathbf{v}^i_t\| \leq \hat{\epsilon}_{v^i},\ i \in \mathcal{S} \tag{15}$$

for suitable constants $\hat{\epsilon}_w > 0$ and $\hat{\epsilon}_{v^i} > 0$, $i \in \mathcal{S}$. □

Notice that in the above lemma the boundedness of the estimation error (14) is taken as an assumption just as an intermediate step. In fact, by exploiting Lemma 2 and Theorem 1, it turns out that the following stability result can be proven which summarizes all the foregoing derivations.

Theorem 2 *Let the system trajectory belong to $\mathcal{X}$, i.e., $\{\mathbf{x}_t\} \subset \mathcal{X}$, and suppose that assumptions A1-A6 are satisfied and that the DEKF algorithm is initialized at time $t = 1$ with positive definite information matrices $\mathbf{\Omega}^i_{1|0}$. Then, the estimation error $\mathbf{e}^i_t$ turns out to be bounded in all the network nodes, i.e., there exists a positive real $\tilde{\epsilon}$ such that*

$$\|\mathbf{e}^i_t\| \leq \tilde{\epsilon} \tag{16}$$

for any i, provided that the initial estimation errors satisfy

$$\|\mathbf{e}^i_1\| \leq \tilde{\epsilon}_0 \tag{17}$$

for some suitable constant $\tilde{\epsilon}_0 > 0$ and the disturbances satisfy

$$\|\mathbf{w}_t\| \leq \tilde{\epsilon}_w, \quad \|\mathbf{v}^i_t\| \leq \tilde{\epsilon}_{v^i},\ i \in \mathcal{S} \tag{18}$$

for suitable constants $\tilde{\epsilon}_w > 0$ and $\tilde{\epsilon}_{v^i} > 0$, $i \in \mathcal{S}$. □

5 Simulation results

The aim of this section is to corroborate the theoretical analysis by showing the effectiveness of the DEKF algorithm in a target tracking case study. To this end, the target motion is modeled by a linear (nearly constant velocity) model

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t$$

with

$$\mathbf{A} = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \frac{T_s^3}{3} & \frac{T_s^2}{2} & 0 & 0 \\ \frac{T_s^2}{2} & T_s & 0 & 0 \\ 0 & 0 & \frac{T_s^3}{3} & \frac{T_s^2}{2} \\ 0 & 0 & \frac{T_s^2}{2} & T_s \end{bmatrix} q$$

where: $\mathbf{x}_t = [x_t, \dot{x}_t, y_t, \dot{y}_t]^\top$ is the kinematic target state at sampling time t made up of the Cartesian coordinates of position (x_t, y_t) and of velocity $(\dot{x}_t, \dot{y}_t)$; T_s is the sampling interval; q is the variance of the random fluctuations of target speed and $\mathbf{Q}$ the covariance matrix of the disturbance $\mathbf{w}_t$.

The target position is measured by two types of nonlinear sensors measuring angle or, respectively, distance. These two sensors, from now on indicated by the acronyms DOA (Direction Of Arrival) and TOA (Time Of Arrival), are characterized by the following measurement functions:

$$\mathbf{h}^i(\mathbf{x}) = \begin{cases} \operatorname{atan2}\left(x - x^i, y - y^i\right), & \text{if } i \text{ is a DOA sensor} \\ \sqrt{\left(x - x^i\right)^2 + \left(y - y^i\right)^2}, & \text{if } i \text{ is a TOA sensor} \end{cases}$$

where atan2 is the 4-quadrant inverse tangent function and (x^i, y^i) denotes the position of the i-th sensor. Overall, the network consists of 100 communication nodes, 5 TOA sensor nodes, and 5 DOA sensor nodes. Graphical representations of the sensor network is provided in Fig. 1. Notice that each sensor is unable by itself to ensure observability of the whole state vector, but collective observability holds.

The measurement noise is assumed to have $\sigma_\theta = 2°$ standard deviation for DOA sensors, and $\sigma_r = 10\ m$ standard deviation for TOA sensors. Other parameters of the simulations are fixed to sampling interval $T_s = 1$ and $q = 0.5\ m^2/s^3$. The DEKF algorithm described in Section 3 is compared with the CEKF of Section 2.1. The matrices $\mathbf{W}$ and $\mathbf{V^i}$ for both algorithms are taken as the inverse of the disturbance and, respectively, measurement noise covariances. The consensus weights used in the simulations of the DEKF algorithm are set equal to the Metropolis weights (Xiao et al., 2005; Calafiore and

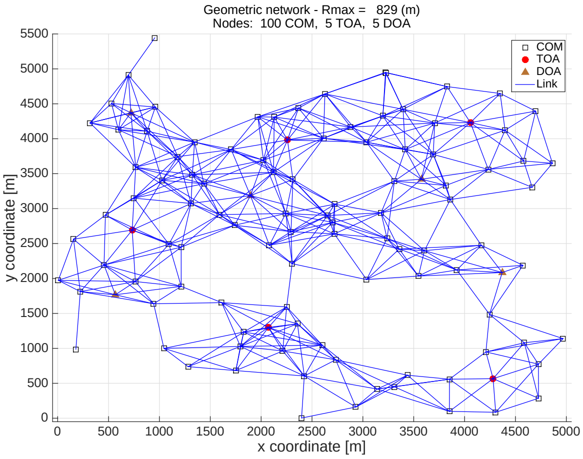

Fig. 1. Nonlinear sensor network used in the simulations.

Abrate, 2009). Further, for the sake of comparison, we consider different numbers L of consensus steps, ranging from 1 to 5, as well as three different choices for the weights γ_t^i: 1) $\gamma_t^i = 1$; 2) γ_t^i generated by means of the binary consensus algorithm described in (Battistelli et al., 2015); and 3) $\gamma_t^i = |\mathcal{N}|$. For brevity, hereafter, we denote the distributed filters resulting from such three different choices of γ_t^i as DEKF1, DEKF2, and DEKF3, respectively.

Overall, 200 independent Monte Carlo trials have been performed and the position root mean square error (PRMSE) has been computed as performance index. The resulting PRMSE time behaviors are reported in Fig. 2 for the considered filters. For the CEKFs, we report the results corresponding to $L = 1$ (in the top plot) and $L = 5$ (in the bottom one). Further, Table 3 shows the PRMSE averaged over time, Monte Carlo trials, and nodes for different values of L. For each filter and each L two values are reported: the top one refers to the PRMSE computed over all the simulation horizon, whereas the bottom one refers to the PRMSE computed after the transient period (i.e., in the last 50 time instants of the simulation horizon).

It can be seen that the DEKF provides satisfactory performance in all the considered settings, even for a low L, and that the lack of local observability does not compromise the stability of the estimation error. On the other hand, the choice of the parameters L and γ_t^i can affect the filter behavior. In fact, in general, performance improves as the number L of consensus steps increases. Further, in accordance with the discussion of Remark 4, by taking γ_t^i greater than 1 it is possible to substantially improve the filter performance both in the transient and in the asymptotic error (see also Battistelli et al., 2015, for additional insights on this issue).

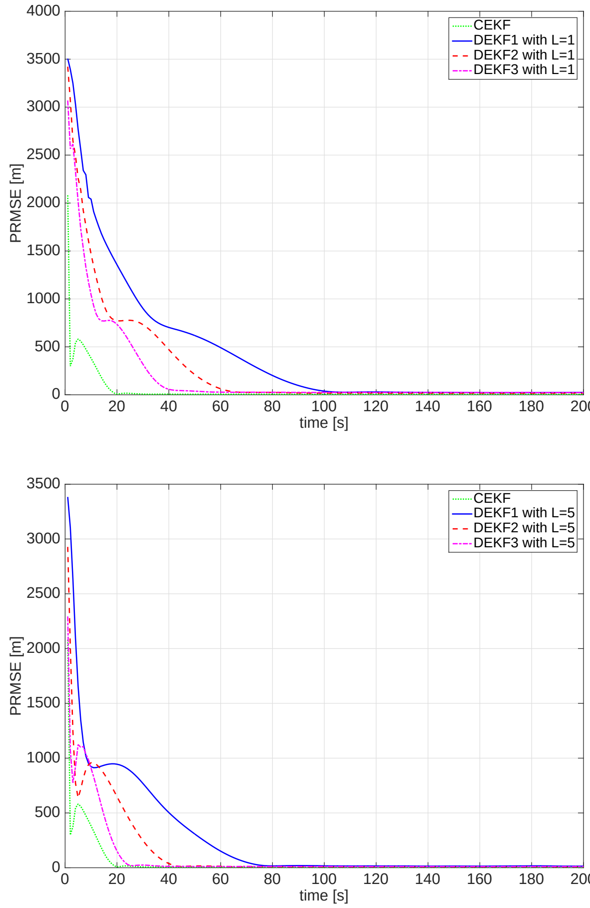

Fig. 2. Behavior of the PRMSE, averaged over all the nodes, for the considered filters. For the DEKFs, the number L of consensus steps is 1 in the top plot and 5 in the bottom one.

Table 3
Performance comparison

PRMSE $[m]$	DEKF1	DEKF2	DEKF3
$L = 1$	426.51	266.11	190.63
	22.60	16.74	23.11
$L = 2$	335.45	183.43	131.27
	18.61	12.44	15.61
$L = 3$	297.23	162.74	111.41
	17.09	10.89	12.79
$L = 4$	273.42	146.26	97.94
	17.02	10.54	11.40
$L = 5$	258.21	135.80	91.09
	15.33	9.20	10.26

6 Conclusions

A family of consensus EKFs for networked estimation has been analyzed, including the consensus on information filter of Battistelli and Chisci (2014) as well as the information weighted consensus filter of Kamal et al.

(2013). In the considered family of algorithms, consensus is applied both to the predicted and the novel information in the EKF information form. A collective Lyapunov function has been constructed for the dynamics of the estimation errors over the network. Then, it has been shown that, under network connectivity and collective observability, the proposed consensus EKF guarantee local stability in all network nodes. An open problem that deserves further investigation is whether similar, or even stronger, stability properties can be achieved by means of different distributed nonlinear state estimation techniques, for example based on the Unscented Kalman Filter (Julier and Uhlmann, 2004).

A Proofs

Proof of Proposition 1: Since the predicted and filtered covariances can be written as

$$\mathbf{\Omega}^i_{t+1|t} = \left(\mathbf{A}^i_t(\mathbf{\Omega}^i_{t|t})^{-1}(\mathbf{A}^i_t)^\top + \mathbf{W}^{-1}\right)^{-1} \tag{A.1}$$

and, respectively,

$$\mathbf{\Omega}^i_{t|t} = \sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1} + \gamma^i_t\sum_{j\in\mathcal{S}}\pi_L^{i,j}\left(\mathbf{C}^j_t\right)^\top \mathbf{V}^j\mathbf{C}^j_t \tag{A.2}$$

it is immediate to see that $\mathbf{\Omega}^i_{1|0} > 0, i\in\mathcal{N}$, implies that $\mathbf{\Omega}^i_{t|t} > 0, i \in \mathcal{N}$, for any finite t. Notice now that

$$\begin{aligned}\mathbf{e}^i_{t+1} &= \mathbf{f}(\mathbf{x}_t) - \mathbf{f}(\hat{\mathbf{x}}_{t|t}) + \mathbf{w}_t = \mathbf{A}^i_t(\mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t}) \\ &\quad + \boldsymbol{\varphi}(\mathbf{x}_t, \hat{\mathbf{x}}^i_{t|t}) + \mathbf{w}_t\,. \end{aligned}\tag{A.3}$$

Further, the estimate $\hat{\mathbf{x}}^i_{t|t}$ can be expressed as

$$\begin{aligned}\hat{\mathbf{x}}^i_{t|t} &= \left(\mathbf{\Omega}^i_{t|t}\right)^{-1} \\ &\quad\times\left[\sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}\hat{\mathbf{x}}^j_{t|t-1} + \gamma^i_t\sum_{j\in\mathcal{S}}\pi_L^{i,j}\left(\mathbf{C}^j_t\right)^\top \mathbf{V}^j\,\bar{\mathbf{y}}^j_t\right]\end{aligned}$$

with $\bar{\mathbf{y}}^j_t$ as in Table 2. Notice also that, by virtue of (A.2), the following identity holds

$$\begin{aligned}\mathbf{x}_t &= \left(\mathbf{\Omega}^i_{t|t}\right)^{-1} \\ &\quad\times\left[\sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}\mathbf{x}_t + \gamma^i_t\sum_{j\in\mathcal{S}}\pi_L^{i,j}\left(\mathbf{C}^j_t\right)^\top \mathbf{V}^j\mathbf{C}^j_t\mathbf{x}_t\right].\end{aligned}$$

Since

$$\begin{aligned}\mathbf{C}^j_t\mathbf{x}_t - \bar{\mathbf{y}}^j_t &= \mathbf{C}^j_t(\mathbf{x}_t - \hat{\mathbf{x}}^j_{t|t-1}) - \mathbf{y}^j_t + \mathbf{h}^j(\hat{\mathbf{x}}^j_{t|t-1}) \\ &= \mathbf{C}^j_t(\mathbf{x}_t - \hat{\mathbf{x}}^j_{t|t-1}) + \mathbf{h}^j(\hat{\mathbf{x}}^j_{t|t-1}) - \mathbf{h}^j(\mathbf{x}_t) - \mathbf{v}^j_t \\ &= \boldsymbol{\chi}^j(\mathbf{x}_t, \hat{\mathbf{x}}^j_{t|t-1}) - \mathbf{v}^j_t\,,\end{aligned}$$

with straightforward calculation one gets

$$\begin{aligned}\mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t} &= \left(\mathbf{\Omega}^i_{t|t}\right)^{-1}\left(\sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}^i_{t|t-1})\right. \\ &\quad\left. + \gamma^i_t\sum_{j\in\mathcal{S}}\pi_L^{i,j}(\mathbf{C}^j_t)^\top\mathbf{V}^j(\boldsymbol{\chi}^j(\mathbf{x}_t, \hat{\mathbf{x}}^i_{t|t-1}) - \mathbf{v}^j_t)\right).\end{aligned}\tag{A.4}$$

Then, equation (7) can be derived by combining (A.4) with (A.3). □

Proof of Lemma 1: Exploiting the fact that

$$\mathbf{\Omega}^i_{t+1|t} \le \tilde{\beta}(\mathbf{A}^i_t)^{-\top}\mathbf{\Omega}^i_{t|t}(\mathbf{A}^i_t)^{-1}$$

for some positive real $\tilde{\beta} < 1$ (see point iii) in Lemma 1 of Battistelli and Chisci (2014)), it turns out that

$$\begin{aligned}&\left(\sum_{j\in\mathcal{N}}\mathbf{\Phi}^{i,j}_t\mathbf{e}^j_t\right)^\top\mathbf{\Omega}^i_{t+1|t}\sum_{j\in\mathcal{N}}\mathbf{\Phi}^{i,j}_t\mathbf{e}^j_t \\ &\le \tilde{\beta}\left(\sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}\mathbf{e}^j_t\right)^\top\left(\mathbf{\Omega}^i_{t|t}\right)^{-1}\sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}\mathbf{e}^j_t \\ &\le \tilde{\beta}\sum_{j\in\mathcal{N}}\pi_L^{i,j}\left(\mathbf{e}^j_t\right)^\top\mathbf{\Omega}^j_{t|t-1}\mathbf{e}^j_t\end{aligned}$$

where the latter inequality follows from the fact that $\mathbf{\Omega}^i_{t|t} \ge \sum_{j\in\mathcal{N}}\pi_L^{i,j}\mathbf{\Omega}^j_{t|t-1}$ and from Lemma 2 of Battistelli and Chisci (2014). As a consequence, it is possible to write

$$\begin{aligned}\mathscr{V}_{t+1}(\mathbf{\Phi}_t\mathbf{e}_t) &= \sum_{i\in\mathcal{N}}p^i\left(\sum_{j\in\mathcal{N}}\mathbf{\Phi}^{i,j}_t\mathbf{e}^j_t\right)^\top\mathbf{\Omega}^i_{t+1|t}\sum_{j\in\mathcal{N}}\mathbf{\Phi}^{i,j}_t\mathbf{e}^j_t \\ &\le \tilde{\beta}\sum_{i,j\in\mathcal{N}}p^i\pi_L^{i,j}\left(\mathbf{e}^j_t\right)^\top\mathbf{\Omega}^j_{t|t-1}\mathbf{e}^j_t \\ &= \tilde{\beta}\sum_{j\in\mathcal{N}}p^j\left(\mathbf{e}^j_t\right)^\top\mathbf{\Omega}^j_{t|t-1}\mathbf{e}^j_t = \tilde{\beta}\,\mathscr{V}_t(\mathbf{e}_t)\end{aligned}$$

which concludes the proof. □

Proof of Theorem 1: Since the predicted covariance can be written as in (A.1), it is easy to verify that condition ii) implies also that there exist suitable positive constants $\underline{\omega}^+$, $\bar{\omega}^+$ such that $0 < \underline{\omega}^+ I \le \mathbf{\Omega}^i_{t+1|t} \le \bar{\omega}^+ I$ for any i and any t. This, in turn, implies that there exist suitable positive constants $\underline{\alpha}$, $\bar{\alpha}$ such that

$$\underline{\alpha}\,\|\mathbf{e}\|^2 \le \mathscr{V}_t(\mathbf{e}) \le \bar{\alpha}\,\|\mathbf{e}\|^2 \tag{A.5}$$

for any t and any $\mathbf{e}$ (recall that all p^i are positive). Observe now that, by triangular inequality, one has

$$
\begin{aligned}
&[\mathscr{V}_{t+1}(\mathbf{e}_{t+1})]^{1/2} \\
&\le [\mathscr{V}_{t+1}(\mathbf{\Phi}_t\mathbf{e}_t)]^{1/2} + [\mathscr{V}_{t+1}(\mathbf{r}_t)]^{1/2} + [\mathscr{V}_{t+1}(\mathbf{s}_t)]^{1/2} \\
&\le \left[\tilde{\beta}\mathscr{V}_t(\mathbf{e}_t)\right]^{1/2} + [\mathscr{V}_{t+1}(\mathbf{r}_t)]^{1/2} + [\mathscr{V}_{t+1}(\mathbf{s}_t)]^{1/2} \quad \text{(A.6)}
\end{aligned}
$$

where $\mathbf{r}_t = \mathrm{col}(\mathbf{r}_t^i,\ i \in \mathcal{N})$, $\mathbf{s}_t = \mathrm{col}(\mathbf{s}_t^i,\ i \in \mathcal{N})$, and the latter inequality follows from Lemma 1.

Suppose now that the disturbances satisfy (13) and let $\epsilon_d = \max\{\epsilon_w;\ \epsilon_{v^i},\ i \in \mathcal{S}\}$. Then, thanks to the boundedness of the time-varying quantities $\gamma_t^i, \mathbf{A}_t^i, (\mathbf{\Omega}_{t|t}^i)^{-1}$, and $\mathbf{C}_t^j$, one has that

$$
[\mathscr{V}_{t+1}(\mathbf{s}_t)]^{1/2} \le \sigma_1\, \epsilon_d
$$

for some positive scalar σ_1.

The only term that remains to be bounded in (A.6) is the one dependent on $\mathbf{r}_t$. To this end, let us suppose for the sake of argument that, at time t, all the estimates $\hat{\mathbf{x}}_{t|t}^i$, $i \in \mathcal{N}$, and $\hat{\mathbf{x}}_{t|t-1}^i$, $i \in \mathcal{S}$, satisfy the conditions

$$
\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i\| \le \kappa_\varphi\,, \quad \text{(A.7)}
$$

$$
\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^i\| \le \kappa_{\chi^i}\,. \quad \text{(A.8)}
$$

Then, by exploiting assumption A4, it is an easy matter to derive the upper bound

$$
\begin{aligned}
&[\mathscr{V}_{t+1}(\mathbf{r}_t)]^{1/2} \le \bar{\alpha}^{1/2}\|\mathbf{r}_t\| \le \bar{\alpha}^{1/2}\sum_{i\in\mathcal{N}}\|\mathbf{r}_t^i\| \\
&\le \bar{\alpha}^{1/2}\sum_{i\in\mathcal{N}}\left(\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i\|^2 + \kappa_1\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^i\|^2\right) \quad \text{(A.9)}
\end{aligned}
$$

for some positive scalar κ_1. Since $\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i$ can be written as in (A.4), then condition (A.8) implies that there exist positive scalars $\tilde{\kappa}_1, \tilde{\kappa}_2$, and $\bar{\sigma}$ such that

$$
\begin{aligned}
&\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i\| \\
&\le \sum_{j\in\mathcal{N}}\left(\tilde{\kappa}_1\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^j\| + \tilde{\kappa}_2\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}^j\|^2\right) + \bar{\sigma}\epsilon_d \\
&\le \tilde{\kappa}_1\|\mathbf{e}_t\| + \tilde{\kappa}_2\|\mathbf{e}_t\|^2 + \bar{\sigma}\epsilon_d \\
&\le \bar{\kappa}_1\,[\mathscr{V}_t(\mathbf{e})]^{1/2} + \bar{\kappa}_2\mathscr{V}_t(\mathbf{e}) + \bar{\sigma}\epsilon_d \quad \text{(A.10)}
\end{aligned}
$$

with $\bar{\kappa}_1 = \tilde{\kappa}_1/\underline{\alpha}^{1/2}$ and $\bar{\kappa}_2 = \tilde{\kappa}_2/\underline{\alpha}$. In view of (A.9) and (A.10), it is immediate to conclude that there exist positive scalars $\kappa_1, \kappa_2, \sigma_2$ such that

$$
[\mathscr{V}_{t+1}(\mathbf{r}_t)]^{1/2} \le \kappa_1\mathscr{V}_t(\mathbf{e}) + \kappa_2\,[\mathscr{V}_t(\mathbf{e})]^2 + \sigma_2\epsilon_d^2\,.
$$

Summing up, it has been proved that, under conditions (A.7) and (A.8), the following inequality holds

$$
\begin{aligned}
[\mathscr{V}_{t+1}(\mathbf{e}_{t+1})]^{1/2} \le \left[\tilde{\beta}\mathscr{V}_t(\mathbf{e}_t)\right]^{1/2} &+ \kappa_1\mathscr{V}_t(\mathbf{e}) + \kappa_2\,[\mathscr{V}_t(\mathbf{e})]^2 \\
&+ \sigma_1\epsilon_d + \sigma_2\epsilon_d^2\,. \quad \text{(A.11)}
\end{aligned}
$$

Suppose now that we can find a pair $(\bar{\epsilon}, \epsilon_d)$ such that

$$
\mathscr{V}_t(\mathbf{e}_t) \le \bar{\epsilon}^2 \quad \text{(A.12)}
$$

implies that conditions (A.7) and (A.8) are satisfied and also that

$$
\mathscr{V}_{t+1}(\mathbf{e}_{t+1}) \le \bar{\epsilon}^2\,. \quad \text{(A.13)}
$$

Then, if such a pair exists, it is sufficient to suppose that the initial estimates $\hat{\mathbf{x}}_{1|0}^i, i \in \mathcal{N}$, are such that condition (A.12) holds at time $t = 1$ in order to conclude, by induction, that it must hold also for any t. This, in turn, would imply that, by taking $\epsilon_0 = \bar{\epsilon}/(\bar{\alpha}\,|\mathcal{N}|)^{1/2}$ in (12), condition (11) would be satisfied with $\epsilon = \bar{\epsilon}/\underline{\alpha}^{1/2}$.

Hence, in order to conclude the proof, it is sufficient to show that a pair $(\bar{\epsilon}, \epsilon_d)$ enjoying such desired properties always exists. To see this, notice first that, when (A.12) holds, one has

$$
\begin{aligned}
\|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^i\| &\le \bar{\kappa}_1\bar{\epsilon} + \bar{\kappa}_2\bar{\epsilon}^2 + \bar{\sigma}\epsilon_d\,, \\
[\mathscr{V}_{t+1}(\mathbf{e}_{t+1})]^{1/2} &\le \tilde{\beta}^{1/2}\bar{\epsilon} + \kappa_1\bar{\epsilon}^2 + \kappa_2\bar{\epsilon}^4 + \sigma_1\epsilon_d + \sigma_2\epsilon_d^2
\end{aligned}
$$

by virtue of (A.10) and (A.11). Hence, in order for (A.7), (A.8), and (A.13) to jointly hold, it is sufficient that

$$
\begin{aligned}
&\bar{\kappa}_1\bar{\epsilon} + \bar{\kappa}_2\bar{\epsilon}^2 + \bar{\sigma}\epsilon_d \le \kappa_\varphi \\
&\bar{\epsilon}/\underline{\alpha}^{1/2} \le \kappa_{\chi^i}, \quad i \in \mathcal{S} \\
&\tilde{\beta}^{1/2}\bar{\epsilon} + \kappa_1\bar{\epsilon}^2 + \kappa_2\bar{\epsilon}^4 + \sigma_1\epsilon_d + \sigma_2\epsilon_d^2 \le \bar{\epsilon}\,.
\end{aligned}
$$

The proof is concluded by noting that, since $\tilde{\beta} < 1$, such a system of inequalities always admit strictly positive solutions $(\bar{\epsilon}, \epsilon_d)$ of sufficiently small norm. □

Proof of Corollary 1: Since the candidate Lyapunov function $V_t(\cdot)$ satisfies (A.5), the statement can be proved by showing that $\mathscr{V}_t(\mathbf{e}_t)$ goes to 0 as $t \to \infty$ provided that the initial estimation errors are sufficiently small. To see this, consider a positive scalar β such that $\tilde{\beta} < \beta < 1$. Then, the convergence to 0 of $\mathscr{V}_t(\mathbf{e}_t)$ can be ensured provided that there exist a positive scalar $\bar{\epsilon}$ such that, whenever $\mathscr{V}_t(\mathbf{e}_t) \le \bar{\epsilon}^2$, one has that conditions (A.7) and (A.8) are satisfied and also that

$$
\mathscr{V}_{t+1}(\mathbf{e}_{t+1}) \le \beta\mathscr{V}_t(\mathbf{e}_t)\,. \quad \text{(A.14)}
$$

Proceeding as in the proof of Theorem 1 (details are omitted for the sake of brevity), it can be shown that, in order for (A.7), (A.8), and (A.14) to jointly hold, it is sufficient that

$$\bar{\kappa}_1\bar{\epsilon}+\bar{\kappa}_2\bar{\epsilon}^2 \leq \kappa_\varphi \tag{A.15}$$
$$\bar{\epsilon}/\underline{\alpha}^{1/2} \leq \kappa_{\chi^i}, \quad i\in\mathcal{S} \tag{A.16}$$
$$(\tilde{\beta}^{1/2}+\kappa_1\bar{\epsilon}+\kappa_2\bar{\epsilon}^3)\,[\mathscr{V}_t(\mathbf{e}_t)]^{1/2} \leq [\beta\mathscr{V}_t(\mathbf{e}_t)]^{1/2} \tag{A.17}$$

which are always satisfied for suitably small values of $\bar{\epsilon}$, since $\tilde{\beta}<\beta$ by construction. □

Proof of Lemma 2: Condition i) is a straightforward consequence of assumption A1 and of the fact that, under condition (14), all the estimated state trajectories $\{\hat{\mathbf{x}}^i_{t|t}\}$ and $\{\hat{\mathbf{x}}^i_{t+1|t}=\mathbf{f}(\hat{\mathbf{x}}^i_{t|t})\}$, $i\in\mathcal{N}$, are guaranteed to lie on a compact set when $\{\mathbf{x}_t\}\subset\mathcal{X}$. As for condition iii), since $\partial\mathbf{f}/\partial\mathbf{x}$ is invertible for any $\mathbf{x}\in\mathcal{X}$ by virtue of assumption A5, the continuity of the matrix inverse ensures that there exists an open neighborhood $\mathcal{X}^+$ of $\mathcal{X}$ (i.e., an open set $\mathcal{X}^+$ with the property that $\mathcal{X}\subset\mathcal{X}^+$) such that $\partial\mathbf{f}/\partial\mathbf{x}$ is invertible on $\mathcal{X}^+$. This, in turn, ensures that $\mathbf{A}^i_t=\partial\mathbf{f}(\hat{\mathbf{x}}^i_{t|t})/\partial\mathbf{x}$ is invertible provided that $\hat{\mathbf{x}}^i_{t|t}\in\mathcal{X}^+$, which can always be ensured by choosing $\hat{\epsilon}$ suitably small in (14).

Hence, only condition ii) remains to be proven. The existence of an upper bound on $\mathbf{\Omega}^i_{t|t}$ readily follows from (A.2), from the fact that $\mathbf{\Omega}^i_{t+1|t}\leq\mathbf{W}$, for any i and $t>1$, and from condition i) (which holds as discussed above). As for the existence of a positive definite lower bound on $\mathbf{\Omega}^i_{t|t}$, recall that, as pointed out at the beginning of the proof of Proposition 1, $\mathbf{\Omega}^i_{1|0}>0, i\in\mathcal{N}$, implies that $\mathbf{\Omega}^i_{t|t}>0, i\in\mathcal{N}$, for any finite t. Hence, if we restrict our attention to a finite time interval $\{1,\dots,T\}$, it follows that condition ii) holds by taking $\underline{\omega}=\underline{\omega}_0\triangleq\min_{i\in\mathcal{N},t\leq T}\underline{\lambda}(\mathbf{\Omega}^i_{t|t})$, where $\underline{\lambda}(\cdot)$ denotes the minimum eigenvalue of a matrix. Consider now a time instant $t>T$ and observe that, due to the linearization, each $\mathbf{\Omega}^i_{t|t}$ can be written as a function of the estimates $\hat{\mathbf{x}}^j_{\tau|\tau}$ and $\hat{\mathbf{x}}^j_{\tau|\tau-1}$ for $\tau\in\{t-T,\dots,t\}$. Then, we can exploit the fact that such estimates are supposed to belong to a neighboorhood of the true system state and use a suitable continuity argument.

To this end, given the state $\mathbf{x}_{t-T}$ at time $t-T$, let us denote by $\mathbf{x}_{t-T,k}$ the k-step-ahead prediction

$$\mathbf{x}_{t-T,k}=\underbrace{\mathbf{f}\circ\cdots\circ\mathbf{f}}_{k\text{ times}}(\mathbf{x}_{t-T})\,.$$

We will show in the last part of the proof that when T is large enough and the estimates in the interval $\{t-T,\dots,t-1\}$ coincide with the predicted trajectory, i.e., $\hat{\mathbf{x}}^j_{t-T+k|t-T+k}=\hat{\mathbf{x}}^j_{t-T+k|t-T+k-1}=\mathbf{x}_{t-T,k}$ for any $k=0,\dots,T$, then it is possible to find a positive scalar $\underline{\omega}_1$ such that $\mathbf{\Omega}^i_{t|t}\geq\underline{\omega}_1\mathbf{I}$. In view of such a result, the lower bound in condition ii) can be easily derived. In fact, by taking $\hat{\epsilon}$, $\hat{\epsilon}_w$ and $\hat{\epsilon}_{v^i}$, $i\in\mathcal{S}$, suitably small in (14) and (15), respectively, it is possible to make the estimates in the interval $\{t-T,\dots,t-1\}$ arbitrarily close to the predicted trajectory. This, by continuity of $\mathbf{\Omega}^i_{t|t}$, implies that it is possible to choose $\hat{\epsilon}$ and $\hat{\epsilon}_w,\hat{\epsilon}_{v^i}$, $i\in\mathcal{S}$, so as to ensure that $\mathbf{\Omega}^i_{t|t}\geq\underline{\omega}_2\mathbf{I}$ with $0<\underline{\omega}_2<\underline{\omega}_1$. Hence, the lower bound in condition ii) can be established with $\underline{\omega}=\min\{\underline{\omega}_2,\underline{\omega}_0\}$.

In order to conclude the proof, let us now suppose that $\hat{\mathbf{x}}^j_{t-T+k|t-T+k}=\hat{\mathbf{x}}^j_{t-T+k|t-T+k-1}=\mathbf{x}_{t-T,k}$ for any $k=0,\dots,T$ and consider the resulting covariance matrix $\mathbf{\Omega}^i_{t|t}$. Notice that, in this case, the linearized matrices can be obtained as

$$\mathbf{A}^j_{t-M+k}=\partial\mathbf{f}(\mathbf{x}_{t-T,k})/\partial\mathbf{x}\,,\quad j\in\mathcal{N} \tag{A.18}$$
$$\mathbf{C}^j_{t-M+k}=\partial\mathbf{h}^j(\mathbf{x}_{t-T,k})/\partial\mathbf{x}\,,\quad j\in\mathcal{S}. \tag{A.19}$$

Hence, we can drop the superscript j from $\mathbf{A}^j_{t-M+k}$ and simply write $\mathbf{A}_{t-M+k}$. By recalling (A.2) and invoking fact ii) in Lemma 1 of Battistelli and Chisci (2014), one gets

$$\begin{aligned}\mathbf{\Omega}^i_{t|t}\geq{}&\check{\beta}\mathbf{A}^{-\top}_{t-1}\left(\sum_{j\in\mathcal{N}}\pi^{i,j}_L\mathbf{\Omega}^j_{t-1|t-1}\right)\mathbf{A}^{-1}_{t-1}\\&+\gamma^i_t\sum_{j\in\mathcal{S}}\pi^{i,j}_L\left(\mathbf{C}^j_t\right)^\top\mathbf{V}^j\mathbf{C}^j_t\end{aligned}$$

for some positive real $\check{\beta}$. By recursively applying such inequality T times, it is possible to write

$$\begin{aligned}\mathbf{\Omega}^i_{t|t}\geq{}&\check{\beta}^T\mathbf{A}^{-\top}_{t-T,t-1}\left(\sum_{j\in\mathcal{N}}\pi^{i,j}_{(T+1)L}\mathbf{\Omega}^j_{t-T|t-T-1}\right)\mathbf{A}^{-1}_{t-T,t-1}\\&+\sum_{\tau=t-T}^{t}\check{\beta}^{t-\tau}\mathbf{A}^{-\top}_{\tau,t-1}\left(\gamma^i_\tau\sum_{j\in\mathcal{S}}\pi^{i,j}_{(t-\tau+1)L}\left(\mathbf{C}^j_\tau\right)^\top\mathbf{V}^j\mathbf{C}^j_\tau\right)\\&\times\mathbf{A}^{-1}_{\tau,t-1}\end{aligned}$$

where, for the sake of compactness, we have defined

$$\mathbf{A}_{\tau,t-1}=\begin{cases}\mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_\tau & \text{if }\tau\leq t-1\\ \mathbf{I} & \text{otherwise}\end{cases}.$$

Since the consensus matrix $\mathbf{\Pi}$ is primitive, the elements $\pi^{i,j}_{(t-\tau+1)L}$ are all positive provided that $t-\tau+1$ is greater than a certain integer, say N. Then, if T is chosen so that $T>N+M$, with M as in assumption A6, it is an easy matter to see that there exists a positive real ρ such that

$$\Omega^i_{t|t} \geq \rho \sum_{\tau=t-T}^{t-T+M} \mathbf{A}^{-\top}_{\tau,t-1} \left(\sum_{j\in\mathcal{S}} \left(\mathbf{C}^j_\tau\right)^\top \mathbf{V}^j \mathbf{C}^j_\tau \right) \mathbf{A}^{-1}_{\tau,t-1} \,. \tag{A.20}$$

By defining $\mathbf{C}_k = \mathrm{col}(\mathbf{C}^j_k,\, j \in \mathcal{S})$ and recalling that we are supposing that (A.18)-(A.19) hold, it is immediate to see that

$$\frac{\partial \mathbf{F}^{[M]}(\mathbf{x}_{t-T})}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{C}_{t-T} \\ \mathbf{C}_{t-T+1}\mathbf{A}_{t-T} \\ \vdots \\ \mathbf{C}_{t-T+M}\mathbf{A}_{t-T,t-T+M-1} \end{bmatrix} .$$

Thus, (A.20) can be rewritten as

$$\Omega^i_{t|t} \geq \rho\, \mathbf{A}^{-\top}_{t-T,t-T+M-1} \left(\frac{\partial \mathbf{F}^{[M]}(\mathbf{x}_{t-T})}{\partial \mathbf{x}} \right)^\top \times \mathbf{V}^{[M]} \frac{\partial \mathbf{F}^{[M]}(\mathbf{x}_{t-T})}{\partial \mathbf{x}} \mathbf{A}^{-1}_{t-T,t-T+M-1} \tag{A.21}$$

where $\mathbf{V}^{[M]}$ is a block-diagonal matrix made up of $M+1$ copies of $\mathbf{V} = \mathrm{diag}(\mathbf{V}^i,\, i \in \mathcal{S})$. Then, the proof can be concluded by noting that under the stated assumptions the right-hand side of (A.21) is always positive definite for any $\mathbf{x}_{t-T} \in \mathcal{X}$ and, hence, by continuity there exists a positive scalar $\underline{\omega}_1$ such that $\Omega^i_{t|t} \geq \underline{\omega}_1 \mathbf{I}$. $\square$

Proof of Theorem 2: Let the constants $\hat{\epsilon}, \hat{\epsilon}_w > 0$, and $\hat{\epsilon}_{v^i} > 0$, $i \in \mathcal{S}$, be as in the statement of Lemma 2. Further, let us consider two constants $\check{\epsilon}$ and $\check{\epsilon}_d$ with the following properties:

(a) conditions (A.15)-(A.17) are satisfied for any $\bar{\epsilon} \leq \check{\epsilon}$ and for any $\epsilon_d \leq \check{\epsilon}_d$;
(b) $\check{\epsilon} \leq \underline{\alpha}^{1/2}\, \hat{\epsilon}$.

Notice that, as discussed at the end of the proof of Theorem 1, such constants can always be found. Suppose that the disturbances are bounded as in (18) with $\tilde{\epsilon}_w = \min\{\hat{\epsilon}_w, \check{\epsilon}_d\}$, and $\tilde{\epsilon}_{v^i} = \min\{\hat{\epsilon}_{v^i}, \check{\epsilon}_d\}, i \in \mathcal{S}$. Then, the proof can be given by induction. To see this, suppose that the inequalities

$$\mathscr{V}_\tau(\mathbf{e}_\tau) \leq \check{\epsilon}^2 \tag{A.22}$$

$$\|\mathbf{e}^i_t\| \leq \hat{\epsilon}, \quad \forall i \tag{A.23}$$

hold up to time τ, i.e., for $\tau = 1, 2, \ldots, t$. Then, by virtue of Lemma 2, we have that conditions i)-iii) hold up to time t as well. Then, by proceeding as in the proof of Theorem 1 and exploiting property (a) above, we can show that condition (A.22) holds also for $\tau = t+1$. Hence, by virtue of property (b), we also have that condition (A.23) holds for $\tau = t+1$ (recall inequality (A.5)). The induction argument can be concluded by noting that we can ensure the fulfillment of (A.22) and (A.23) for $\tau = 1$ by taking $\tilde{\epsilon}_0 = \min\{\hat{\epsilon}, \check{\epsilon}/(\bar{\alpha}|\mathcal{N}|)^{1/2}\}$ (recall again inequality (A.5)). Hence, Theorem 2 holds with $\tilde{\epsilon} = \hat{\epsilon}$. $\square$

References

Battistelli, G. and Chisci, L. (2014). Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, 50(3), 707–718.

Battistelli, G., Chisci, L., and Fantacci, C. (2014). Parallel consensus on likelihoods and priors for networked nonlinear filtering. *Signal Processing Letters*, 21(7), 787–791.

Battistelli, G., Chisci, L., Morrocchi, S., and Papi, F. (2011). An information-theoretic approach to distributed state estimation. In *Proc. 18th IFAC World Congress,* Milan, Italy, 12477–12482.

Battistelli, G., Chisci, L., Mugnai, G., Farina, A., and Graziano, A. (2015). Consensus-based linear and nonlinear filtering. *IEEE Transactions on Automatic Control*, 60(5), 1410–1415.

Calafiore, G.C. and Abrate, F. (2009). Distributed linear estimation over sensor networks. *International Journal of Control*, 82(5), 868–882.

Cattivelli, F. and Sayed, A. (2010). Diffusion strategies for distributed kalman filtering and smoothing. *IEEE Transactions on Automatic Control*, 55(9), 2069–2084.

Farina, M., Ferrari-Trecate, G., and Scattolini, R. (2010). Distributed moving horizon estimation for linear constrained systems. *IEEE Transactions on Automatic Control*, 55(11), 2462–2475.

Farina, M., Ferrari-Trecate, G., and Scattolini, R. (2012). Distributed moving horizon estimation for nonlinear constrained systems. *International Journal of Robust and Nonlinear Control*, 22(2), 123–143.

Hlinka, O., Hlawatsch, F., and Djurc, P. (2013). Distributed particle filtering in agent networks : A survey, classification, and comparison]. *IEEE Signal Processing Magazine*, 30(1), 61–81.

Julier, S.J. and Uhlmann, J.K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3), 401–422.

Kamal, A.T., Farrell, J.A., and Roy-Chowdhury, A.K. (2012). Information weighted consensus. In *Proc. 51st IEEE Conference on Decision and Control,* (*Maui, USA)*, 2732–2737.

Kamal, A.T., Farrell, J.A., and Roy-Chowdhury, A.K. (2013). Information weighted consensus filters and their application in distributed camera networks. *IEEE Transactions on Automatic Control*, 58(12), 3112–3125.

Kamgarpour, M. and Tomlin, C. (2008). Convergence properties of a decentralized Kalman filter. In *Proc. 47th IEEE Conference on Decision and Control,* Cancun, Mexico, 3205–3210.

Kar, S. and Moura, J. (2011). Gossip and distributed Kalman filtering: Weak consensus under weak detectability. *IEEE Transactions on Signal Processing*, 59(4), 1766–1784.

La Scala, B., Bitmead, R., and James, M. (1995). Conditions for stability of the extended Kalman filter and their application to the frequency tracking problem. *Mathematics of Control, Signals, and Systems*, 8(1), 1–26.

Li, W. and Jia, Y. (2012). Distributed consensus filtering for discrete-time nonlinear systems with non-Gaussian noise. *Signal Processing*, 92(10), 2464–2470.

Matei, I. and Baras, J. (2012). Consensus-based linear distributed filtering. *Automatica*, 48(8), 1776–1782.

Millan, P., Orihuela, L., Jurado, I., Vivas, C., and Rubio, F.R. (2015). Distributed estimation in networked systems under periodic and event-based communication policies. *International Journal of Systems Science*, 46(1), 139–151.

Millan, P., Orihuela, L., Vivas, C., Rubio, F.R., Dimarogonas, D.V., and Johansson, K. (2013). Sensor-network-based robust distributed control and estimation. *Control Engineering Practice*, 21(9), 1238–1249.

Mohammadi, A. and Asif, A. (2013). Distributed particle filter implementation with intermittent/irregular consensus convergence. *IEEE Transactions on Signal Processing*, 61(10), 2572–2587.

Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *Proc. 46th IEEE Conference on Decision and Control,* New Orleans, USA, 5492–5498.

Reif, K., Gunther, S., Yaz, E., and Unbehauen, R. (1999). Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transactions on Automatic Control*, 44(4), 714–728.

Reif, K. and Unbehauen, R. (1999). The extended Kalman filter as an exponential observer for nonlinear systems. *IEEE Transactions on Signal Processing*, 47(8), 2324–2328.

Stankovic, S.S., Stankovic, M.S., and Stipanovic, D.M. (2009). Consensus based overlapping decentralized estimation with missing observations and communication faults. *Automatica*, 45(6), 1397–1406.

Ugrinovskii, V. (2011). Distributed robust filtering with H_∞ consensus. *Automatica*, 47(1), 1–13.

Ugrinovskii, V. (2013). Conditions for detectability in distributed consensus-based observer networks. *IEEE Transactions on Automatic Control*, 58(10), 2659–2664.

Wang, S., Ren, W., and Li, Z. (2014). Information-driven fully distributed kalman filter for sensor networks in presence of naive nodes. *arXiv*, 1410.0411.

Xiao, L., Boyd, S., and Lall, S. (2005). A scheme for robust distributed sensor fusion based on average consensus. In *Proc. 4th Int. Symposium on Information Processing in Sensor Networks,* Los Angeles, USA, 63–70.

Zhou, Z., Fang, H., and Hong, Y. (2013). Distributed estimation for moving target based on state-consensus strategy. *IEEE Transactions on Automatic Control*, 58(8), 2096–2101.