Consensus-based linear and nonlinear filtering

Original Citation:

Availability:
This version is available at: 2158/911148 since: 2021-03-24T18:36:26Z

Published version:
DOI: 10.1109/TAC.2014.2357135

Terms of use:
Open Access
La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:
Consensus-based linear and nonlinear filtering

G. Battistelli, L. Chisci, Senior Member, IEEE, G. Mugnai, A. Farina Fellow, IEEE and A. Graziano,

Abstract—This note addresses Distributed State Estimation (DSE) over sensor networks. Two existing consensus approaches for DSE, i.e. consensus on information (CI) and consensus on measurements (CM), are combined to provide a novel class of hybrid consensus filters (named Hybrid CMCI) which enjoy the complementary benefits of CM and CI. Novel theoretical results, limitedly to linear systems, on the guaranteed stability of the Hybrid CMCI filters under collective observability and network connectivity are proved. Finally, the effectiveness of the proposed class of consensus filters is evaluated on a target tracking case-study with both linear and nonlinear sensors.

Index Terms—Distributed state estimation; sensor networks; consensus; nonlinear filtering.

I. INTRODUCTION

Advances in wireless sensor networks are making multi-agent (distributed) architectures ubiquitous in modern monitoring and control systems. This technological breakthrough, however, needs to be supported by theoretical research work finalized to redesign distributed estimation and control algorithms that preserve as much as possible the stability, performance and robustness requirements of their centralized counterparts. The present note specifically deals with Distributed State Estimation (DSE) over a sensor network with no fusion centre, by exploiting the consensus approach [1]. This topic has recently gathered a lot of attention in the context of both parameter estimation, consensus-based techniques with guaranteed performance can be found for example in [2], [3].

As for state estimation, several consensus-based approaches have been proposed in the literature both in linear and nonlinear settings. A first family of techniques, hereafter referred to as Consensus on Estimates (CE), is based on the idea of spreading the available information over the network by performing, at each time instant, a consensus averaging of the local state estimates/predictions [4]-[11]. The choice of averaging out only the state estimates serves the purpose of keeping the information exchange between neighboring nodes as limited as possible. However, since also the covariance matrices contain valuable information which can be used to improve performance, other, more elaborated approaches have been proposed. For instance, in [12]-[14], it is proposed to perform a consensus among the local measurements and innovation covariances so as to approximate, in a distributed way, the correction step of the centralized Kalman filter. This approach is referred to in this note as Consensus on Measurements (CM). A different point of view has been adopted in [15], giving rise to the so-called Consensus on Measurements (CM) approach. From an algorithmic viewpoint, in the linear case CI is nothing but consensus applied to the inverse covariance (information) matrix and to the information vector. From an information-theoretic viewpoint, CI can be interpreted as consensus on probability density functions in the Kullback-Leibler average sense.

G. Battistelli, L. Chisci and G. Mugnai are with Dipartimento di Ingegneria dell’Informazione, Università di Firenze, 50139, Firenze, Italy. e-mail: {giorgio.battistelli,luigi.chisci,giovanni.mugnai}@unifi.it

A. Farina and A. Graziano are with Selex ES, Rome, Italy. e-mail: {alfonso.farina,antonio.graziano}@selex-es.com

This work was supported in part by a grant of Selex ES.

The approach presented in this note moves from the observation that CM and CI exhibit complementary features. Specifically, CI guarantees stability for any number of consensus steps (even a single one [15]), but its mean-square estimation error performance can be hampered by the fact that the fusion rule adopts a conservative point of view by assuming the correlation between the estimates coming from different nodes to be completely unknown. On the other hand, CM avoids any conservative assumption on the correlation by fusing only the novel information, but it does not guarantee stability unless the number of consensus steps is sufficiently high. Clearly, this can be problematic whenever, for reduced communication cost and improved energy efficiency, only few consensus steps can be performed in each sampling interval. In this context, a novel family of DSE algorithms, named Hybrid CMCI (HCMCI), is proposed wherein CM and CI are combined so as to retain the positive features of both approaches. Different choices for the combination weights are discussed, leading to different properties. Following [16], where preliminary results on the subject can be found, the nonlinear case is treated by resorting to the Extended Kalman Filter (EKF) paradigm. The main contribution of this note is a stability analysis of the proposed family of HCMCI filters in the case of linear systems. In order to substantiate the analysis, a performance evaluation of the considered consensus-based filters on a target tracking simulation case-study with both linear and nonlinear sensors is presented. As a final remark, it is pointed out that, for a particular choice of the combination weights, the proposed algorithm is equivalent to the Information Consensus Filter (ICF), derived independently from our work and originally presented in [17], whose stability was observed experimentally in [18].

II. PROBLEM FORMULATION

This note addresses 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 among distant sensor nodes, in order to guarantee network connectivity while keeping the transmission power (and hence the communication range) of each sensor node at a moderate intensity, hence prolonging network lifetime. In the sequel, the network will be denoted by the triplet \((S,C,A)\): where \(\mathcal{S}\) is the set of sensor nodes, \(\mathcal{C}\) the set of communication nodes, \(\mathcal{N} = \mathcal{S} \cup \mathcal{C}, \mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}\) is the set of arcs (connections) such that \((i,j) \in \mathcal{A}\) if node \(j\) can receive data from node \(i\) (clearly \((i,i) \notin \mathcal{A}\) 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. \(\mathcal{N}_i \doteq \{j : (j,i) \in \mathcal{A}\}\).

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

\[
\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \mathbf{w}_t
\]

and a set of sensors \(S\) with measurement equations

\[
\mathbf{y}^i_t = \mathbf{h}_t^i(\mathbf{x}_t) + \mathbf{v}_t^i, \quad i \in \mathcal{S}.
\]

It is assumed that \(\mathbf{w}_t, \mathbf{v}^1_t, \mathbf{v}^2_t, \ldots\) are mutually uncorrelated zero-mean white noises with covariances \(Q_t = E[\mathbf{w}_t \mathbf{w}_t^T] > 0\) and

\[
\mathbf{R}_t = E[\mathbf{v}^i_t \mathbf{v}^i_t] > 0.
\]
TABLE I: Information C(E)KF Algorithm, to be implemented at each sampling interval $t = 1, 2, \ldots$ starting from initial conditions $\hat{x}_1(0), \Omega_1(0), q_1(0) = \Omega_1(0) \hat{x}_1(0)$.

<table>
<thead>
<tr>
<th>Correction (measurement-update):</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_t = \frac{\partial \hat{x}_1(t)}{\partial \hat{x}<em>r}(\hat{x}</em>{t-1}, t)$, \quad $i \in S$</td>
</tr>
<tr>
<td>$\Omega_{t</td>
</tr>
<tr>
<td>$y_{t</td>
</tr>
<tr>
<td>$q_{t</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction (time-update):</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{x}_{t+1</td>
</tr>
</tbody>
</table>

R = $E \left[ y_t^2 \right] > 0$. Then the objective is to have, at each time $t \in \{1, 2, \ldots\}$ and in each node $i \in N$, an estimate $\hat{x}_{t|t}$ of the state $x_t$ constructed only on the basis of the local measurements (when available) and of data received from all adjacent nodes $j \in N \setminus \{i\}$.

Centralized (Extended) Kalman Filter

A testbed for the performance of any DSE algorithm is the centralized (Extended) Kalman Filter, denoted as C(E)KF, which is assumed to simultaneously process all measurements $\{y_t, i \in S\}$. Hereafter, for convenience, the information filter form will be adopted. The information filter propagates, instead of the estimate $\hat{x}_{t|t-1}$ and covariance $P_{t|t-1}$, the information (inverse covariance) matrices

$$\Omega_{t|t} = P_{t|t-1}^{-1}, \quad \Omega_{t|t} = \Omega_{t|t-1}^{-1}$$

and the information vectors

$$q_{t|t-1} = P_{t|t-1}^{-1} \hat{x}_{t|t-1}, \quad q_{t|t} = P_{t|t}^{-1} \hat{x}_{t|t}.$$

Introducing the noise information matrices $W_t \triangleq Q_t^{-1}$ and $V_t = \left(R_t^T\right)^{-1}$ and supposing the functions $f_t(\cdot)$ and $h_i(\cdot)$ to be continuously differentiable for any $t$ and any $i \in S$, the recursive information filter of Table 1 can be derived.

Notice that the algorithm of Table I generalizes the Information Kalman Filter algorithm, corresponding to $f_t(x) = A_t x$ and $h_i(x) = C_i x$, to nonlinear systems (1) and/or sensors (2) via the Extended Kalman Filter paradigm of linearizing the state and measurement equations at the current estimate.

III. DISTRIBUTED STATE ESTIMATION ALGORITHMS

A. Consensus on Information and Consensus on Measurements

The covariance intersection fusion rule [19] suggests a possible consensus approach to DSE, namely consensus on the information (matrix-vector) pair. Let us assume that, at time $t$, each node $i \in N$ be provided with a local information pair $\{\Omega_{t|t-1}, q_{t|t-1}\}$. Then the Consensus on Information (CI) approach to DSE is summarized by the algorithm of Table II to be carried out at each sampling interval $t \geq 1$ in each node $i \in N$. Notice that, in this consensus iteration, each node $i$ computes a regional Kalman, that is a combination of the values in $N_i$ with suitable consensus weights $\omega^i_{i\ell}, j \in N_i$. In this note, a convex combination is adopted by supposing $\omega^i_{i\ell} \geq 0$ and $\sum_{j \in N_i} \omega^i_{i\ell} = 1, \forall i \in N$.

It is worth to point out that the CI algorithm of Table II reduces to the well known covariance intersection [19] for a single ($L = 1$) consensus step. It represents, therefore, a generalization of covariance intersection to multiple ($L > 1$) consensus steps which can be introduced in order to improve performance. Notice that $L$ linearly increases both computation and communication burdens; hence it should be selected as a suitable tradeoff between cost and performance.

As an alternative approach, in [12], [13] it has been proposed to exploit consensus in order to compute in a distributed way the quantities $\partial \Omega_i / \partial x_i$ and $\partial q_i / \partial x_i$ for distributed linear Kalman filtering. As shown in [16], this approach, that will be referred to as consensus on measurements (CM), can be extended to nonlinear systems by following the EKF paradigm. To this end, the information filter of Table I can be exploited, with the only difference that the virtual measurements $\bar{y}_i(t)$ has to be redefined in terms of the local state predictions $\hat{x}_{t|t-1}$ instead of the centralized one $\hat{x}_{t|t-1}$ (which is not available in a distributed setting). The interest reader is referred to [16] for a detailed derivation.

Notice that consensus provides, at convergence, the averages $\partial \Omega_i / \partial x_i$ and $\partial q_i / \partial x_i$, $N_i$ denoting cardinality of $N_i$, while the information filter update actually requires $\partial \Omega_i$ and $\partial q_i$. This drawback can be partially remedied by multiplying the consensus outcome by some suitable scalar weight $\omega^i$. Possible choices for the weights $\omega^i$ will be discussed in the next section. Summing up, the consensus-based DSE algorithm of Table III is obtained.

B. Hybrid Consensus

Consensus on measurements and consensus on information have complementary positive and negative features. To see this, let us denote by $\Pi$ the consensus matrix, whose elements are the consensus weights $\pi^i_{j\ell}, i, j \in N$. Further, let $\pi^i_{j\ell}$ be the $(i, j)$-th element of $\Pi^T$, i.e. the $\ell$-th power of the consensus matrix $\Pi$. Then, the correction step for the considered algorithms can be rewritten as in Table IV. Notice that, for the terms $\delta q_i$ and $\delta \Omega_i$, the summations are extended only to sensor nodes since, for all communication nodes, one has $\delta q_i = 0$ and $\delta \Omega_i = 0$.

TABLE II: Consensus on Information (CI) Algorithm

<table>
<thead>
<tr>
<th>Correction:</th>
</tr>
</thead>
<tbody>
<tr>
<td>if $i \in S$ then</td>
</tr>
<tr>
<td>$C^i_{t} = \frac{\partial \hat{x}_{i</td>
</tr>
<tr>
<td>$\Omega_i(0) = \Omega_i(t</td>
</tr>
<tr>
<td>sample the measurement $y_i^t$</td>
</tr>
<tr>
<td>$y_i^t = y_i - h_i(\hat{x}_{t</td>
</tr>
<tr>
<td>$q_i(0) = q_i(t</td>
</tr>
<tr>
<td>end if</td>
</tr>
<tr>
<td>if $i \in C$ then</td>
</tr>
<tr>
<td>$\Omega_i(0) = \Omega_i(t</td>
</tr>
<tr>
<td>end if</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consensus:</th>
</tr>
</thead>
<tbody>
<tr>
<td>for $\ell = 0, 1, \ldots, L - 1$ do</td>
</tr>
<tr>
<td>Fuse the quantities $q_i^\ell(t)$ and $\Omega_i^\ell(t)$ according to</td>
</tr>
<tr>
<td>$q_i^{\ell+1}(t) = \sum_{j \in N_i} \pi^i_{i\ell} q_j^\ell(t)$</td>
</tr>
<tr>
<td>$\Omega_i^{\ell+1}(t) = \sum_{j \in N_i} \pi^i_{i\ell} \Omega_j^\ell(t)$</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{x}_{t+1</td>
</tr>
<tr>
<td>$\Omega_{t+1</td>
</tr>
<tr>
<td>$q_{t+1</td>
</tr>
</tbody>
</table>
Consensus on Measurements (CM) Algorithm

Compute the local correction terms:

If \( \ell \in S \) then

\[
\begin{align*}
\Omega_i^\ell &= (C_i^T)^T V_i^\ell V_i^\ell C_i \\
\delta q_i^\ell &= C_i^T V_i^\ell y_i^\ell \\
\delta \Omega_j^\ell &= (C_j^T)^T V_j^\ell V_j^\ell C_j
\end{align*}
\]

end if

Consensus:

\[
\delta q_i^ℓ(0) = \delta q_i^ℓ, \quad \text{and} \quad \delta \Omega_i^ℓ(0) = \delta \Omega_i^ℓ
\]

for \( \ell = 0, 1, \ldots, L - 1 \) do

\[
\begin{align*}
\delta q_i^ℓ(\ell + 1) &= \sum_{j \in N_i} \pi_{ij}^\ell \delta q_j^ℓ(\ell) \\
\delta \Omega_i^ℓ(\ell + 1) &= \sum_{j \in S} \pi_{ij}^\ell \delta \Omega_j^ℓ(\ell)
\end{align*}
\]

end for

Correction:

\[
q_i^ℓ|_t = q_i^{ℓ|_{t-1}} + \omega_i^ℓ \delta q_i^ℓ(L) \\
\Omega_i^ℓ|_t = \Omega_i^{ℓ|_{t-1}} + \omega_i^ℓ \delta \Omega_i^ℓ(L) \\
\hat{x}_i^ℓ|_t = \left( \Omega_i^{ℓ|_{t-1}} \right)^{-1} q_i^ℓ|_t
\]

Prediction: as in Table II

As it can be seen from Table IV, in CI a convex combination of the local priors is computed. As shown in [15], this is important in order to ensure boundedness of the estimation error for any number (even a single one) of consensus steps. On the contrary, in CM no consensus on the local priors is performed since only the novel information, i.e. \( \delta q_i^ℓ \) and \( \delta \Omega_i^ℓ \), is combined. Hence, the stability of this family of distributed filters depends on a crucial way on the number \( L \) of consensus steps per sampling interval. To see this, let \( N_i(L) \) be the set of nodes from which node \( i \) can be reached in \( L \) hops. Then, since \( \pi_{ij}^ℓ \) can be different from 0 only if \( j \in N_i(L) \), it is immediate to see that CM can provide a bounded estimation error on a given node \( i \) only if observability (or at least detectability) from the set of nodes \( N_i(L) \) holds. In general, this becomes true only for a sufficiently large number of consensus steps (see, for instance, the example of Section V).

On the other hand, in CI the novel information is unavoidably underestimated, whereas in CM the multiplication by the scalar weights \( \omega_i^ℓ \) can, at least partially, remedy this drawback, thus leading to enhanced performance as the number of consensus steps increases. In order to combine the benefits and neutralize the drawbacks of both CI and CM, the two approaches can be combined so as to give rise to a hybrid consensus approach, hereafter called HCMCI (Hybrid CMCI) Algorithm.

Consensus on Measurements - Consensus on Information). HCMCI actually carries out in parallel, at each consensus step, the two types of consensus CI and CM described above. The resulting algorithm is detailed in Table V.

Notice that, actually, Table V provides a family of distributed filters corresponding to different choices of the scalar weights \( \omega_i^ℓ \). With this respect, a reasonable choice consists in setting \( \omega_i^ℓ = |N_i| \). In fact, when the consensus weights are chosen so that \( \pi_{ij}^ℓ \to 1/|N_i| \) as \( L \to \infty \), such a choice has the appealing feature of giving rise to a distributed algorithm converging to the centralized (E)KF as \( L \) tends to infinity. Notice that, when such a choice is adopted, the resulting algorithm is equivalent to the ICF proposed in [18] in a linear setting. Notice also that, in this case, it is possible to perform jointly the two parallel consensus algorithms of Table V so as to save bandwidth [18] (this is true whenever the weights \( \omega_i^ℓ \) are node-independent).

While asymptotically optimal, the choice \( \omega_i^ℓ = |N_i| \) may have some drawbacks when only a finite, possibly small, number of consensus steps is performed. In fact, a multiplication by \( |N_i| \) could actually lead in some nodes to an overestimation of \( \delta \Omega_i^ℓ \), a situation that one might want to avoid in order to preserve the consistency of each local filter. An alternative solution is to exploit consensus so as to compute, in a distributed way, a normalization factor able to improve the filter performance while preserving consistency of each local filter. For example, an estimate of the fraction \(|S|/|N_i|\) of sensor nodes in the network can be computed via the consensus algorithm

\[
b_i^ℓ(\ell + 1) = \sum_{j \in N_i} \pi_{ij}^ℓ b_i^ℓ(\ell), \quad \text{for} \quad \ell = 0, 1, \ldots, L - 1
\]

with the initialization \( b_i^ℓ(0) = 1 \) if \( i \in S \), and \( b_i^ℓ(0) = 0 \) otherwise. Then, it can be seen that the choice

\[
\omega_i^ℓ = \begin{cases} 
1/b_i^ℓ(L) & \text{if } b_i^ℓ(L) \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]

has the desirable property of preserving the consistency of each local filter. This can be seen by looking at the last row of Table IV and noting that the novel information is never overweighted. In Section V and noting that the novel information is never overweighted. In Table V and noting that the novel information is never overweighted. In

\[\delta q_i^ℓ = 0, \quad \text{and} \quad \delta \Omega_i^ℓ = 0\]

end if

Consensus:

\[
\delta q_i^ℓ(0) = \delta q_i^ℓ, \quad \text{and} \quad \delta \Omega_i^ℓ(0) = \delta \Omega_i^ℓ
\]

for \( \ell = 0, 1, \ldots, L - 1 \) do

\[
\begin{align*}
\delta q_i^ℓ(\ell + 1) &= \sum_{j \in N_i} \pi_{ij}^ℓ \delta q_j^ℓ(\ell) \\
\delta \Omega_i^ℓ(\ell + 1) &= \sum_{j \in S} \pi_{ij}^ℓ \delta \Omega_j^ℓ(\ell)
\end{align*}
\]

end for

Correction:

\[
q_i^ℓ|_t = q_i^{ℓ|_{t-1}} + \omega_i^ℓ \delta q_i^ℓ(L) \\
\Omega_i^ℓ|_t = \Omega_i^{ℓ|_{t-1}} + \omega_i^ℓ \delta \Omega_i^ℓ(L) \\
\hat{x}_i^ℓ|_t = \left( \Omega_i^{ℓ|_{t-1}} \right)^{-1} q_i^ℓ|_t
\]

Prediction: as in Table II

\[\Omega^{ℓ|_{t-1}}\]

TABLE IV: Expression of the information vector and matrix in the correction step after consensus for the considered algorithms

| \hline
| \hline
| CI & \sum_{j \in N_i} \pi_{ij}^ℓ q_j^{ℓ|_{t-1}} + \sum_{j \in S} \pi_{ij}^ℓ \delta q_j^ℓ | \sum_{j \in N_i} \pi_{ij}^ℓ \delta \Omega_j^ℓ | \hline
| CM & q_i^{ℓ|_{t-1}} + \omega_i^ℓ \sum_{j \in S} \pi_{ij}^ℓ \delta q_j^ℓ | \Omega_i^{ℓ|_{t-1}} + \omega_i^ℓ \sum_{j \in S} \pi_{ij}^ℓ \delta \Omega_j^ℓ | \hline
| HCMCI & \sum_{j \in N_i} \pi_{ij}^ℓ q_j^{ℓ|_{t-1}} + \omega_i^ℓ \sum_{j \in S} \pi_{ij}^ℓ \delta q_j^ℓ | \sum_{j \in N_i} \pi_{ij}^ℓ \delta \Omega_j^ℓ | \hline
| \hline

TABLE V: Hybrid CMCI (HCMCI) Algorithm

| \hline
| \hline
| Compute the local correction terms:
| \hline
| If \( \ell \in S \) then
| \hline
| \hline
| \hline
| end if
| \hline
| Consensus:
| \hline
| \hline
| end if
| \hline
| Correction:
| \hline
| \hline
| \hline
| Prediction: as in Table II
| \hline

Recall that a filter is said to be consistent when its estimated error covariance is an upper bound (in the positive definite sense) of the true error covariance [19].
\[ \omega_i^j \pi_{t+1}^i = \pi_t^i / b_i(L) \] which is guaranteed not to exceed 1 since, by construction, \( b_i(L) = \sum_{j \in \mathcal{S}} \pi_{t}^{i,j} \).

**IV. STABILITY ANALYSIS**

In this section, the stability properties of the proposed DSE algorithm are analyzed in a linear time-invariant setting. To this end, it is supposed that the system dynamics and measurement equations take the form

\[
\begin{align*}
x_{t+1} &= Ax_t + w_t \\
y_t^i &= C^i x_t + v_t & \text{if } i \in \mathcal{S}
\end{align*}
\]  

Further, the process and measurement noise covariances are supposed to be time-invariant as well, i.e., \( Q_i = Q > 0 \) and \( R_i = R^i > 0 \) for any \( t \) and \( i \). The following preliminary assumptions are needed.

**A1.** The system matrix \( A \) is invertible.

**A2.** The system is collectively observable, i.e. the pair \((A, C)\) is observable where \( C := \text{col} \{ C^i; i \in \mathcal{S} \} \).

**A3.** The consensus matrix \( \Pi \) is row stochastic and primitive.\(^2\)

**A4.** There exist two positive scalars \( \omega \) and \( \bar{\omega} \) such that \( 0 < \bar{\omega} \leq \omega \) for any \( i \in \mathcal{N} \) and \( t \geq 0 \).

Notice that assumption A1 is automatically satisfied in sampled-data systems wherein the matrix \( A \) is obtained by discretization of a continuous-time system matrix. As for assumption A2, the collective observability requirement can be relaxed to collective detectability by resorting to an observability decomposition in each network node as discussed in [15]. Assumption A3 is strictly related to the network connectivity. In fact, it can be satisfied if and only if the network is strongly connected (i.e., any node is reachable from any other node through a directed path). For instance, in this case, the Metropolis weights [2], [3] satisfy A3. While taking the consensus matrix \( \Pi \) row stochastic is sufficient for stability, a doubly stochastic \( \Pi \) would also ensure that all the elements of \( \Pi^L \) tends to \( 1/|\mathcal{N}| \) as \( L \to +\infty \). Finally, notice that both the choices suggested in the previous section for the weights \( \omega_t \) satisfy assumption A4.

We now state the main stability result of this note which will be proved step by step in the remaining of this section.

**Theorem 1:** Let assumptions A1-A4 hold. Further, let the HCMCI algorithm be adopted and let \( L \geq 1 \). Finally, suppose that, for every \( i \in \mathcal{N} \), the a priori information matrix \( \Omega^i_{t=0} \) is positive definite. Then, the estimation error \( e^i_t = x_t - \hat{x}^i_{t+1} \) is asymptotically bounded in mean square, i.e., \( \lim_{t \to \infty} \mathbb{E} \left\{ \|e^i_t\|^2 \right\} < +\infty \), for any \( i \in \mathcal{N} \).

In order to derive the above stability result, a first important step is the study of the properties of the information matrices \( \Omega^i_{t+1} \).

**Lemma 1:** Let the same assumptions as in Theorem 1 hold. Then, there exist positive definite matrices \( \Omega^i_{t} \), \( \Pi^i_{t} \), and \( \Omega^i_{t+1} \) such that \( 0 < \Omega^i_{t} \leq \Omega^i_{t+1} \leq \Pi^i_{t} \) for any \( i \in \mathcal{N} \) and \( t \geq 1 \).

**Proof:** Let us first focus on the matrices \( \Omega^i_{t+1} \). By recalling the identity in the last row of Table IV and the fact that \( \Delta \Omega^i_t = (C^i)^T V^j C^i \), it can be seen that assumption A4 implies that

\[
\begin{align*}
\Omega^i_{t+1} &\leq \sum_{j \in \mathcal{N}} \pi^i_{t,j} \Omega^i_{t-1} + \sum_{j \in \mathcal{S}} \pi^i_{t,j} \bar{\omega} (C^i)^T V^j C^i, \\
\Omega^i_{t+1} &\geq \sum_{j \in \mathcal{N}} \pi^i_{t,j} \Omega^i_{t-1} + \sum_{j \in \mathcal{S}} \pi^i_{t,j} \omega (C^i)^T V^j C^i.
\end{align*}
\]

Suppose now that a CI algorithm is carried out with noise covariance matrices \( R^i / \bar{\omega} \) and let \( \Omega^i_{t+1} \) and \( \Omega^i_{t} \) be the resulting information matrices. Similarly, let \( \Omega^i_{t+1} \) and \( \Omega^i_{t} \) denote the information matrices resulting from a CI algorithm with the matrices \( R^i / \omega \). Recalling the first row of Table IV, it can be readily seen that

\[
\begin{align*}
\Omega^i_{t+1} &= \sum_{j \in \mathcal{N}} \pi^i_{t,j} \Omega^i_{t+1} + \sum_{j \in \mathcal{S}} \pi^i_{t,j} \bar{\omega} (C^i)^T V^j C^i, \\
\Omega^i_{t+1} &= \sum_{j \in \mathcal{N}} \pi^i_{t,j} \Omega^i_{t+1} + \sum_{j \in \mathcal{S}} \pi^i_{t,j} \omega (C^i)^T V^j C^i.
\end{align*}
\]

Hence, provided that the same initializations are adopted for all the algorithms, i.e. \( \Omega^i_{0} = \Omega^i_{1} \), one can see by means of simple induction arguments that \( \Omega^i_{t} \leq \Omega^i_{t+1} \leq \Pi^i_{t} \) for any \( i \in \mathcal{N} \) and \( t \geq 1 \) (recall also that the prediction step of the Kalman filter recursion is monotone nondecreasing in the sense that \( \Omega^i_{t} \leq \Omega^i_{t+1} \) implies \( \Omega^i_{t+1} \leq \Omega^i_{t+1} \leq \Pi^i_{t} \)). Then, the first part of the proof is concluded by noting that, under assumptions A1-A3, there exist two positive definite matrices \( \Omega^i_{t} \) and \( \Pi^i_{t} \) such that \( \Omega^i_{t} \geq \Omega^i_{t+1} \) and \( \Omega^i_{t+1} \leq \Pi^i_{t} \) (by virtue of Theorem 2 of [15] and the fact that \( \Omega^i_{1} > 0 \)). As for the bounds on \( \Omega^i_{t} \), one can simply take \( \Pi^i_{t} = \mathbf{W} \) and \( \Omega^i_{t} = (\mathbf{A} \Omega^i_{t-1} \mathbf{A}^T + \mathbf{Q})^{-1} \).

**Lemma 1** ensures that the matrices \( \Omega^i_{t} \) are non-singular and, hence, that the estimation errors \( e^i_t \) are well defined. As it can be seen from the proof, this property crucially depends on network connectivity (assumption A3) and collective observability (assumption A2). The following result can now be stated.

**Proposition 1:** Let the same assumptions as in Theorem 1 hold. Then the estimation errors \( e^i_t \) obey the recursion

\[
e^i_{t+1} = \sum_{j \in \mathcal{N}} \Phi^i_{j} e^i_t + \sum_{j \in \mathcal{N}} \Gamma_{j} v^i_t + w_t
\]

for any \( i \in \mathcal{N} \), where \( \Phi^i_{j} = \pi^i_{j} A (\Omega^i_{t-1})^{-1} \Omega^i_{t-1} \) and \( \Gamma_{j} = \pi^i_{j} \omega (C^i)^T V^j C^i \).

**Proof:** Notice first that \( e^i_{t+1} = A(\mathbf{x}_t - \hat{x}^i_{t+1}) + w_t \). Further, the estimate \( \hat{x}^i_{t+1} \) can be expressed as \( \hat{x}^i_{t+1} = \sum_{j \in \mathcal{N}} \pi^i_{j} \Omega^i_{t-1} \mathbf{x}_t + \omega (C^i)^T V^j C^i \mathbf{y}^i_t \) with \( \mathbf{y}^i_t = C^i \mathbf{x}_t + v_t \) (recall again the last row of Table IV). Then, equation (11) can be derived with straightforward calculations by recalling the expression for \( \Omega^i_{t+1} \) in the last row of Table IV and noting that the identity \( \mathbf{x}_t = \sum_{j \in \mathcal{N}} \pi^i_{j} \Omega^i_{t-1} \mathbf{x}_t + \omega (C^i)^T V^j C^i \mathbf{y}^i_t \) holds.

Consider now the noise-free collective dynamics of the estimation errors

\[
e^i_{t+1} = \Phi^i_{t} e^i_t
\]

where \( e^i_t = \text{col} \{ e^i_t, i \in \mathcal{N} \} \) and \( \Phi^i_{t} \) is a block matrix whose block elements are given by the matrices \( \Phi^i_{j} \) defined in Proposition 1.

Then, the following result holds.

**Lemma 2:** Let the same assumptions as in Theorem 1 hold. Then the time-varying system (12) is uniformly exponentially stable.

**Proof:** Let \( p \) be the Perron-Frobenius left eigenvector of the matrix \( \Pi^L \). Notice that, by virtue of Assumption A3, such an eigenvector has strictly positive components \( p^i \), \( i \in \mathcal{N} \), and satisfies the equation \( \Pi^L p^i = p^i \), i.e., \( \sum_{j \in \mathcal{N}} p^j \pi^i_{j} = p^i \). Consider now the candidate Lyapunov function

\[
\gamma^i(e^i_t) = \sum_{i \in \mathcal{N}} p^i (e^i_t)^T \Omega^i_{t-1} e^i_t
\]
for system (12). In view of Lemma 1, it is immediate to see that there exist suitable strictly positive constants $\alpha_1$ and $\alpha_2$ such that

$$\alpha_1 \|e_t\|^2 \leq \gamma_t(e_t) \leq \alpha_2 \|e_t\|^2.$$  

Moreover, exploiting the fact that $\Omega^i_{\{t+1\}} \leq 2\beta A^{-1} \Omega^i_{\{t\}} A^{-1}$ for some positive real $\beta < 1$ (see point iii) in Lemma 1 of [15]), it turns out that

$$\left( e_{t+1}^i \right)^T \Omega^i_{\{t+1\}[e_{t+1}^i] = \left( \sum_{j \in N} \pi^i_{\{j\}} e_j^i \right)^T \Omega^i_{\{t+1\}} \left( \sum_{j \in N} \pi^i_{\{j\}} e_j^i \right) \leq \beta \sum_{j \in N} \pi^i_{\{j\}} \left( e_j^i \right)^T \Omega^i_{\{t\}} \left( e_j^i \right) \leq \beta \sum_{j \in N} \pi^i_{\{j\}} \left( e_j^i \right)^T \Omega^i_{\{t\} - 1} e_j^i$$

where the latter inequality follows from the fact that $\Omega^i_{\{t\}} \geq \sum_{j \in N} \pi^i_{\{j\}} \Omega^i_{\{t\} - 1}$ and from Lemma 2 of [15]. As a consequence, it is possible to write

$$\gamma_{t+1}(e_{t+1}) = \sum_{i \in N} p^i \left( e_{t+1}^i \right)^T \Omega^i_{\{t+1\}} \left( e_{t+1}^i \right) \leq \beta \sum_{i \in N} p^i \pi^i_{\{j\}} \left( e_j^i \right)^T \Omega^i_{\{t\}} \left( e_j^i \right) = \beta \sum_{j \in N} p^j \left( e_j^i \right)^T \Omega^i_{\{t\} - 1} e_j^i = \beta \gamma_t(e_t)$$

from which uniform exponential stability of (12) follows at once. Square error (PRMSE), averaged over time and over all the network nodes, has been computed as performance index. Fig 2 shows the comparison between the following state estimators: C(E)KF, CM with weights $\omega^i_1$ as in (4), CI, HCMCI with weights $\omega^i_1$ as in (4) (denoted hereafter as HCMCI-1), HCMCI with weights $\omega^i_1 = |N|$ (denoted hereafter as HCMCI-2). Notice that the minimum number of consensus steps for which the CM algorithm exhibits a stable behavior for the two networks is $L = 4$ and $L = 5$, respectively. Accordingly the PRMSEs of the CM for lower values of $L$ are omitted in the plots since out of scale. The simulation results confirm the theoretical analysis by showing that HCMCI, for both the considered weight choices, ensures stability already for $L = 1$ and outperforms CI thanks to the presence of the scalars $\omega^i_1$ weighting the novel information. Further, by comparing the PRMSEs of HCMCI-1 and HCMCI-2, it can be seen that, at least in the considered set-up, the choice $\omega^i_1 = |N|$, which is asymptotically optimal as $L \to +\infty$, is not necessarily the best one when only a limited number of consensus steps per iteration can be performed. Of course, different choices of the weights $\omega^i_1$ can be convenient depending on the particular setup.

V. SIMULATION EXAMPLES

The aim of this section is to investigate how different types of consensus-based distributed state estimators compare with each other and with the centralized state estimator. To this end, a single-target tracking case study will be considered, where the target motion is modelled by a linear (nearly constant velocity) model (see [15]) where $x_t = [x_t, \dot{x}_t, y_t, \dot{y}_t]^T$ is the kinematic target state at sampling time $t$ made up of the Cartesian coordinates of position $(x_t, y_t)$ and of velocity $(\dot{x}_t, \dot{y}_t)$; the sampling interval has been fixed to 1 s and the variance of the random fluctuations of target speed to 0.25 m/s$^2$. Two different simulation scenarios corresponding to two different sensor networks will be considered.

In the first scenario, the network is composed of 100 communication nodes and 5 linear position sensor nodes characterized by the measurement equation (6) with

$$C^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \forall i \in S.$$  

Conversely, in the second scenario two types of nonlinear position sensors measuring angle or, respectively, distance have been considered. 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:

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

TABLE VI: Percentage performance degradation

<table>
<thead>
<tr>
<th>$P_d$</th>
<th>$P_L$</th>
<th>HCMCI-1</th>
<th>HCMCI-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>0.9</td>
<td>2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>0.9</td>
<td>3</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.9</td>
<td>4</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

VI. CONCLUDING REMARKS

While the stability analysis carried out in this note pertains only to the LTI case, it turns out that it could be used as a starting point for studying the stability of the estimation error dynamics also in a nonlinear setting. In fact, under suitable continuity assumptions it is possible to write the estimation error dynamics in a suitable way so that the linearized part is separated from the nonlinear (higher-order) terms. Then, an analysis similar to the one of Section IV could be carried out to prove stability of the linearized part of the estimation error dynamics, which in turn is the key for achieving a local stability result for the overall dynamics. Results in this direction can be found in [20] to which the interested reader is referred.
Fig. 1: Linear (top) and nonlinear (bottom) sensor networks used in the simulations.

REFERENCES