Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 53-2 (2020) 3242–3247

Asymptotically Optimal Distributed Filtering of Continuous-Time Linear Systems

S. Battilotti * F. Cacace ** M. d'Angelo * A. Germani ***

* *DIAG, Università "Sapienza" di Roma, Roma, Italy, (e-mail: s.battilotti@diag.uniroma1.it, mdangelo@diag.uniroma1.it).*
** *Università Campus Bio-Medico di Roma, Italy; CNR-IASI Biomathematics Laboratory, Rome, Italy (e-mail: f.cacace@unicampus.it).*
*** *DISIM, Università degli Studi dell'Aquila, L'Aquila, Italy (e-mail: alfredo.germani@univaq.it).*

Abstract: In this paper we prove the following new and unexpected result: it is possible to design a continuous-time distributed filter for linear systems that asymptotically tends at each node to the optimal centralized filter. The result concerns distributed estimation over a connected undirected graph and it only requires to exchange the estimates among adjacent nodes. We exhibit an algorithm containing a consensus term with a parametrized gain and show that when the parameter becomes arbitrarily large the error covariance at each node becomes arbitrarily close to the error covariance of the optimal centralized Kalman filter.

Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

Keywords: Continuous time filters; Kalman filters; Filtering theory; Consensus filters

1. INTRODUCTION

Distributed filtering involves state estimation using a set of local filters that communicate with all other nodes through a communication network that constraints the information exchange to the neighbors in the network. In the last years distributed filtering techniques are becoming increasingly popular in the sensor network community due to their scalability for large networks and high fault tolerance and spurred a great amount of research to investigate their design and performance. They are particularly relevant in the context of cooperative multi-agent systems, a framework that is attracting a great deal of attention (see Lewis et al. (2013); Qu (2009) for comprehensive treatments). Surveys of distributed state estimation approaches and comparison with centralized and decentralized methods can be found in Chong (2017); Olfati-Saber et al. (2007); Taj and Cavallaro (2011). The most relevant issues for distributed estimation techniques include: (i) *accuracy*, i.e., reducing the estimation error at each node; (ii) *consensus*, i.e. convergence of estimates across nodes that is essential for cooperation of multi-agent systems; (iii) *communication*, i.e. reducing the amount of communication burden among nodes; (iv) *observability*, i.e. the capability of dealing with sensors with limited or null system measurements (Kamal et al., 2013). It is generally assumed the existence of trade-offs among these features, whose relative relevance may also vary in different application areas. For example, strong consensus may be obtained at the expenses of estimation accuracy at some nodes (Battistelli et al., 2014), and increased accuracy might require more intensive communication across nodes. It is moreover tacitly assumed that the accuracy of distributed schemes is always worse than accuracy of a centralized optimal algorithm that uses all the available information at the same time.

From a theoretical perspective the results presented in this work show that, at least for the relevant case of continuous-time systems and undirected network topology, these trade-offs and tacit assumption are apparent. We exhibit an algorithm that attains perfect consensus and optimal accuracy with least communication burden and requires only collective observability (i.e. the system must be observable from the union of the sensor nodes in the network). From a more practical perspective the proposed algorithm extends the consensus distributed DKF algorithm proposed Olfati-Saber et al. (2007) to the case of networks with nodes that have limited or null measurement capabilities and provide a more accurate estimate. Our filter, named Asymptotic Distributed Kalman Filter (ADKF) is introduced in Section 4 after formally describing the framework in Section 2 and the centralized optimal approach in Section 3. The proof of the asymptotic optimality of ADKF is derived in Section 5. Finally, in Section 6 and Section 7 we compare, from the theoretical and numerical perspective, the proposed algorithm with the centralized solution and other 3 recent approaches, namely the already mentioned DKF of Olfati-Saber et al. (2007), the Multi-Agent Consensus Filter of Ren et al. (2005); Ji et al. (2017), and the distributed information-weighted Kalman consensus filter (IKCF) of Ji et al. (2017).

Notation. $\mathbb{R}$ and $\mathbb{C}$ denote real and complex numbers. For a square matrix A, $\mathrm{tr}(A)$ is the trace and $\sigma(A)$ is the spectrum. A is said to be Hurwitz stable if $\sigma(A) \subset \mathbb{C}_-$,

2405-8963 Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2020.12.1124

the set of complex numbers with negative real part. $\mathbb{E}\{\cdot\}$ denotes expectation. $\otimes$ is the Kronecker product between vectors or matrices. The operators $\mathrm{row}_i()$, $\mathrm{col}_i()$, $\mathrm{diag}_i()$ denote respectively the horizontal, vertical and diagonal compositions of matrices and vectors indexed by i. Let $\mathcal{S}(n) \in \mathbb{R}^{n\times n}$ be the set of symmetric matrices of size n. $\mathcal{P}(n)$ (resp., $\mathcal{P}_+(n)$) $\subset \mathcal{S}(n)$ denotes the set of positive semi-definite (definite) matrices in $\mathcal{S}(n)$. We denote I_n the identity matrix of size n and by $U_n = \mathbf{1}_n\mathbf{1}_n^\top$, $\mathbf{1}_n = \mathrm{col}_{i=1}^n(1)$, the square matrix of size n having 1 in each entry.

2. SYSTEM AND PROBLEM FORMULATION

We use a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to describe the information exchange between the N nodes, where $\mathcal{V} = \{1, 2 \ldots, N\}$ is the set of vertices representing the N agents and $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$ is the set of edges of the graph. An edge of $\mathcal{G}$ is denoted by (i, j), representing that nodes i and j can exchange information between them. The graph is undirected, that is, the edges (i, j) and $(j, i) \in \mathcal{E}$ are considered to be the same. Two nodes i and j are neighbors to each other if $(i, j) \in \mathcal{E}$. The set of neighbors of node i is denoted by $\mathcal{N}^{(i)} := \{j \in \mathcal{V} : (j, i) \in \mathcal{E},\ j \neq i\}$. A path is a sequence of connected edges in a graph. A graph is connected if there is a path between every pair of vertices. The adjacency matrix $\mathcal{A}$ of a graph $\mathcal{G}$ is an $N \times N$ matrix, whose (i, j)-th entry is 1 if (i, j) is an edge of $\mathcal{G}$ and 0 otherwise. The degree matrix $\mathcal{D}$ of $\mathcal{G}$ is a diagonal matrix whose i-th diagonal element is equal to the cardinality of $\mathcal{N}_i$. The Laplacian of $\mathcal{G}$ is defined to be a $N \times N$ matrix $\mathcal{L}$ such that $\mathcal{L} = -\mathcal{A} + \mathcal{D}$. $\mathcal{L}$ is symmetric if and only if the graph is undirected. Moreover, $0 = \lambda_1(\mathcal{L}) < \lambda_2(\mathcal{L}) \leqslant \cdots \leqslant \lambda_N(\mathcal{L})$, where $\lambda_i(\mathcal{L})$ denotes an eigenvalue of $\mathcal{L}$, if and only if the graph is connected. An eigenvector associated to $\lambda_1(\mathcal{L})$ is $\mathbf{1}_N$.

Consider the process

$$\dot{\mathbf{x}}_t = A\mathbf{x}_t + \mathbf{f}_t, \tag{1}$$

$$\mathbf{y}_t^{(i)} = C_i\mathbf{x}_t + \mathbf{g}_t^{(i)}, \quad i = 1, \ldots, N, \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^n$, $\mathbf{y}_t^{(i)} \in \mathbb{R}^{q_i}$, $q_i \geqslant 0$, and $\mathbf{f}_t$ and $\mathbf{g}_t^{(i)}$, $i = 1, \ldots, N$, are zero-mean white Gaussian noises, mutually independent with covariance respectively $Q \in \mathcal{P}_+(n)$, $R_i \in \mathcal{P}_+(q_i)$ $i = 1, \ldots, N$. The matrices Q and $R = \mathrm{diag}_i(R_i)$ are nonsingular. Also $\mathbf{x}_0$ is random with mean $\bar{\mathbf{x}}_0 := \mathbb{E}\{\mathbf{x}_0\}$ and covariance $\Sigma_{\mathbf{x}_0} := \mathbb{E}\{(\mathbf{x}_0 - \bar{\mathbf{x}}_0)(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^\top\}$.

If $C = \mathrm{col}_i(C_i)$ we assume that (C, A) is observable. Also, we use the notation $\mathbf{y}_t = \mathrm{col}_i(\mathbf{y}_t^{(i)})$. Each $\mathbf{y}_t^{(i)}$ represents the data available at node i, $i = 1, \ldots, N$, in the network.

We will design an optimal distributed state estimator for the system (1) with the given topology of the network $\mathcal{G}$. The distributed estimator will consists of N local estimators, one for each node, which exchange their local estimate with the neighbors.

3. THE CENTRALIZED KALMAN-BUCY OPTIMAL FILTER (CKBF)

The equations of the centralized Kalman-Bucy optimal filter (CKBF) for (1) are

$$\begin{aligned}
\dot{\hat{\mathbf{x}}}_\mathbf{t} &= A\hat{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{y}_t - C\hat{\mathbf{x}}_t),\\
\hat{\mathbf{x}}_0 &= \bar{\mathbf{x}}_0,\\
\dot{\mathbf{P}}_t &= A\mathbf{P}_t + \mathbf{P}_tA^\top + Q - \mathbf{P}_tC^\top R^{-1}C\mathbf{P}_t,\\
\mathbf{P}_0 &= \Sigma_{\mathbf{x}_0},
\end{aligned} \tag{3}$$

with $\mathbf{K}_t = \mathbf{P}_tC^\top R^{-1}$. The matrix $\mathbf{P}_t$ represents the covariance of the estimation error $\mathbb{E}\{(\mathbf{x}_t - \hat{\mathbf{x}}_t)(\mathbf{x}_t - \hat{\mathbf{x}}_t)^\top\}$. We have $\mathbf{P}_t \in \mathcal{P}_+(n)$, bounded for all $t \geqslant 0$ and $\mathbf{P}_t \to P_\infty$ as $t \to +\infty$ with $P_\infty \in \mathcal{P}_+(n)$ the unique solution of

$$0 = AP_\infty + P_\infty A^\top + Q - P_\infty C^\top R^{-1}CP_\infty. \tag{4}$$

This Riccati equation can be also written as

$$0 = A_CP_\infty + P_\infty A_C^\top + Q + P_\infty C^\top R^{-1}CP_\infty \tag{5}$$

where

$$A_C := A - K_\infty C. \tag{6}$$

From (3) we also obtain the asymptotically optimal CBKF (ACKBF)

$$\dot{\hat{\mathbf{x}}}_{\mathbf{ss,t}} = A\hat{\mathbf{x}}_{ss,t} + K_\infty(\mathbf{y}_t - C\hat{\mathbf{x}}_{ss,t}), \tag{7}$$

with

$$K_\infty = P_\infty C^\top R^{-1}. \tag{8}$$

4. THE ASYMPTOTICALLY OPTIMAL DISTRIBUTED KALMAN FILTER (ADKF)

Our distributed Kalman filter (ADKF) consists of one filter for each sensor node of the network. The equations for the ADKF at the i-th sensor node are:

$$\begin{aligned}
\dot{\hat{\mathbf{x}}}_t^{(i)} &= A\hat{\mathbf{x}}_t^{(i)} + K_i(\mathbf{y}_t^{(i)} - C_i\hat{\mathbf{x}}_t^{(i)})\\
&\quad + \gamma P_\infty \sum_{j\in\mathcal{N}^{(i)}} (\hat{\mathbf{x}}_t^{(j)} - \hat{\mathbf{x}}_t^{(i)}),
\end{aligned} \tag{9}$$

with $K_i = NP_\infty {C_i}^\top {R_i}^{-1}$ and $\gamma > 0$ a parameter to be tuned as pointed out in the following proposition. Define

$$A_i := A - K_iC_i,\ i = 1, \ldots, N, \tag{10}$$

and

$$A_D(\gamma) := \mathrm{diag}_i(A_i) - \gamma(\mathcal{L}\otimes P_\infty). \tag{11}$$

Notice that $A_D(\gamma) \in \mathbb{R}^{nN\times nN}$.

Proposition 1. There exists $\gamma_0 > 0$ such that for all $\gamma > \gamma_0$ $A_D(\gamma)$ is Hurwitz stable.

Proof. We prove that there exists a positive definite symmetric matrix $X_\infty \in \mathcal{P}_+(nN)$ and $\gamma_0 > 0$ such that for all $\gamma > \gamma_0$

$$X_\infty A_D(\gamma) + A_D^\top(\gamma)X_\infty < 0. \tag{12}$$

If $X_\infty := I_N \otimes P_\infty^{-1}$ we have

$$X_\infty A_D(\gamma) + A_D^\top(\gamma)X_\infty = \mathrm{diag}_i(W_i) - 2\gamma(\mathcal{L}\otimes I_n) \tag{13}$$

where $W_i := P_\infty^{-1}A_i + A_i^\top P_\infty^{-1}$. Since

$$\begin{aligned}
v \in \mathbb{R}^{nN}\backslash\{0\}:\quad & v^\top(\mathcal{L}\otimes I_n)v = 0\\
&\Rightarrow v = \mathbf{1}_N \otimes \bar{v},\ \bar{v} \in \mathbb{R}^n\backslash\{0\},
\end{aligned} \tag{14}$$

it follows that

$$\begin{aligned}
&v \in \mathbb{R}^{nN}\backslash\{0\}:\quad v^\top(\mathcal{L}\otimes I_n)v = 0\\
&\Rightarrow v^\top \mathrm{diag}_i(W_i)v = N\bar{v}^\top(P_\infty^{-1}A_C + A_C^\top P_\infty^{-1})\bar{v} < 0.
\end{aligned}$$

By Finsler's lemma there exists $\gamma_0 > 0$ such that, for all $\gamma > \gamma_0$, $\mathrm{diag}_i(W_i) - 2\gamma(\mathcal{L}\otimes I_n) < 0$, which proves (12). □

5. PROPERTIES OF THE ADKF

In order to study the asymptotic properties of the ADKF we introduce the local (at the node) estimation error $\mathbf{e}_t^{(i)} := \mathbf{x}_t - \widehat{\mathbf{x}}_t^{(i)}$, the total estimation error $\mathbf{e}_t := \mathrm{col}_i(\mathbf{e}_t^{(i)})$, the total measurement noise vector $\mathbf{g}_t := \mathrm{col}_i(\mathbf{g}_t^{(i)}) \in \mathbb{R}^{\sum_{i=1}^n q_i}$, the noise vector $\mathbf{h}_t := \mathrm{col}_i(\mathbf{f}_t - K_i\mathbf{g}_t^{(i)})$ with covariance $\Psi_h := \mathbb{E}\{\mathbf{h}_t\mathbf{h}_t^\top\}$, and the estimation error covariance matrix $\mathbf{X}_t := \mathbb{E}\{\mathbf{e}_t\mathbf{e}_t^\top\}$. Clearly, $\mathbf{X}_t$ depends on γ, but we omit this dependence for notational simplicity. We have

$$\dot{\mathbf{e}}_t = A_D(\gamma)\mathbf{e}_t + \mathbf{h}_t \tag{15}$$

$$\Psi_h = U_N \otimes Q + \mathrm{diag}_i(K_i) R\, \mathrm{diag}_i(K_i^\top), \tag{16}$$

where U_N is a matrix with all entries 1 (see notation). Notice that the estimates $\widehat{\mathbf{x}}^{(i)}$ are (asymptotically) unbiased since $A_D(\gamma)$ is Hurwitz stable (for $\gamma > \gamma_0$).

Proposition 2. For all $\gamma > \gamma_0$ the estimation error covariance matrix $\mathbf{X}_t$ is uniformly bounded in time.

Proof. The result follows form standard arguments from the fact that $A_D(\gamma)$ is Hurwitz for $\gamma > \gamma_0$ and Ψ_h is constant. The covariance $\mathbf{X}_t$ obeys for all $t \geqslant 0$

$$\dot{\mathbf{X}}_t = A_D(\gamma)\mathbf{X}_t + \mathbf{X}_t A_D^\top(\gamma) + \Psi_h \tag{17}$$

and its asymptotic value is solution of

$$A_D(\gamma)\mathbf{X}_\infty + \mathbf{X}_\infty A_D^\top(\gamma) = -\Psi_h. \tag{18}$$

□

Our purpose is to show the key result that $\mathbf{X}_\infty \to U_N \otimes P_\infty$ when $\gamma \to \infty$ (recall that P_∞ is the asymptotic error covariance of the CKBF). The matrix $\mathbf{X}_\infty^C := U_N \otimes P_\infty$ is the asymptotic error covariance for N identical CBKFs implemented at each node and using the whole output $\mathbf{y}_t$. Recalling that $(\mathcal{L} \otimes P_\infty)\mathbf{X}_\infty^C = (\mathcal{L} \otimes P_\infty)(U_N \otimes P_\infty) = 0$, we obtain that $\mathbf{X}_\infty^C$ satisfies

$$\begin{aligned} 0 = &\left(\mathrm{diag}_{i=1}^N(A_C) - \gamma(\mathcal{L} \otimes P_\infty)\right)\mathbf{X}_\infty^C \\ &+ \mathbf{X}_\infty^C\left(\mathrm{diag}_{i=1}^N(A_C) - \gamma(\mathcal{L} \otimes P_\infty)\right)^\top + U_N \otimes Q \\ &+ \mathrm{diag}_{i=1}^N(K_\infty)(U_N \otimes R)\mathrm{diag}_{i=1}^N(K_\infty^\top). \end{aligned} \tag{19}$$

Let $G_i := C_i^\top R_i^{-1} C_i$, $G_d := \mathrm{diag}_i(G_i)$, and $G := C^\top R^{-1} C$. Notice that

$$G = \sum_{i=1}^N G_i. \tag{20}$$

By introducing the covariance mismatch $\mathbf{E}_t := \mathbf{X}_t - \mathbf{X}_\infty^C$ we obtain after some manipulations

$$\dot{\mathbf{E}}_t = A_D(\gamma)\mathbf{E}_t + \mathbf{E}_t A_D^\top(\gamma) + \Sigma, \tag{21}$$

where

$$\begin{aligned} \Sigma := &N^2(I_N \otimes P_\infty)G_d(I_N \otimes P_\infty) + U_N \otimes (P_\infty G P_\infty) \\ &- N(I_N \otimes P_\infty)G_d(U_N \otimes P_\infty) - N(U_N \otimes P_\infty)G_d(I_N \otimes P_\infty). \end{aligned}$$

Our main result can thus be stated as follows.

Proposition 3. $\lim_{\gamma\to+\infty}\lim_{t\to+\infty}\mathbf{X}_t = \mathbf{X}_\infty^C := U_N \otimes P_\infty$.

In order to prove the above proposition, we notice that it is always possible to construct a transformation T such that

$$T = \begin{pmatrix} \frac{\mathbf{1}_N^\top}{\sqrt{N}} \\ t_2^\top \\ \cdots \\ t_N^\top \end{pmatrix}^\top,\ T\mathcal{L}T^T = \mathrm{diag}\{0, \lambda_2, \ldots, \lambda_N\}$$

where $\lambda_2, \ldots, \lambda_n > 0$ are the positive eigenvalues of $\mathcal{L}$ (see Section 2) and

$$t_i \mathbf{1}_N = \sum_{j=1}^N t_{i,j} = 0, \quad t_i t_j^\top = \delta_{i,j},\ i,j = 2, \ldots, N,$$

i.e. T is orthornormal. Define

$$S := T \otimes I_n \tag{22}$$

and let $\widetilde{\mathbf{E}}_t := S\mathbf{E}_t S^\top$. We have after some manipulations and taking into account (20)

$$\dot{\widetilde{\mathbf{E}}}_t = \widetilde{A}_D(\gamma)\widetilde{\mathbf{E}}_t + \widetilde{\mathbf{E}}_t\widetilde{A}_D^\top(\gamma) + N^2\widetilde{\Sigma} \tag{23}$$

where

$$\widetilde{A}_D(\gamma) = \begin{pmatrix} A_C & \Pi_{1,2} & \cdots & \Pi_{1,N} \\ \Pi_{2,1} & \Pi_{2,2} - \gamma\lambda_2 P_\infty & \cdots & \Pi_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \Pi_{N,1} & \Pi_{N,2} & \cdots & \Pi_{N,N} - \gamma\lambda_N P_\infty \end{pmatrix} \tag{24}$$

$$\Pi_{1,j} = \Pi_{j,1} := -\sqrt{N}\sum_{l=1}^N t_{j,l} P_\infty G_l, \quad j = 2, \ldots, N, \tag{25}$$

$$\Pi_{i,j} = \Pi_{j,i} := \delta_{i,j}A - N\sum_{l=1}^N t_{i,l}t_{j,l}P_\infty G_l,\ i,j = 2, \ldots, N, \tag{26}$$

and

$$\widetilde{\Sigma} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \widetilde{\Sigma}_{2,2} & \cdots & \widetilde{\Sigma}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \widetilde{\Sigma}_{N,2} & \cdots & \widetilde{\Sigma}_{N,N} \end{pmatrix} \tag{27}$$

$$\widetilde{\Sigma}_{i,j} = \widetilde{\Sigma}_{j,i} := \sum_{l=1}^N t_{i,l}t_{j,l}P_\infty G_l P_\infty. \tag{28}$$

The matrix $\widetilde{A}_D(\gamma)$ is Hurwitz for all $\gamma > \gamma_0$ (since $A_D(\gamma)$ is Hurwitz for all $\gamma > \gamma_0$ by Proposition 1). Notice also that since $\sigma(\widetilde{A}_D(\gamma)) \cap \sigma(-\widetilde{A}_D^\top(\gamma)) = \varnothing$ for all $\gamma > \gamma_0$, for each $\gamma > \gamma_0$ there exists a unique symmetric solution $\widetilde{\mathbf{E}}_\infty(\gamma)$ to

$$0 = \widetilde{A}_D(\gamma)\widetilde{\mathbf{E}}_\infty(\gamma) + \widetilde{\mathbf{E}}_\infty(\gamma)\widetilde{A}_D^\top(\gamma) + N^2\widetilde{\Sigma} \tag{29}$$

The proof of Proposition 3 follows directly from the next two lemmas.

Lemma 1. For each $\gamma > \gamma_0$

$$\lim_{t\to+\infty}\widetilde{\mathbf{E}}_t = \widetilde{\mathbf{E}}_\infty(\gamma) \tag{30}$$

Proof. Let $\mathbf{\Delta}_t := \widetilde{\mathbf{E}}_t - \widetilde{\mathbf{E}}_\infty(\gamma)$. Since $\widetilde{A}(\gamma)$ is Hurwitz for all $\gamma \geqslant \gamma_0$,

$$\mathbf{\Delta}_t = e^{\widetilde{A}_D(\gamma)t}\mathbf{\Delta}_0 e^{\widetilde{A}_D^\top(\gamma)t} \to 0 \tag{31}$$

as $t \to +\infty$. □

Lemma 2. $\lim_{\gamma\to+\infty}\widetilde{\mathbf{E}}_\infty(\gamma) = 0$.

Proof. The solution $\widetilde{\mathbf{E}}_\infty(\gamma)$ of (29) is unique and it can be parametrized in γ as follows. Let

$$W_1 := \mathrm{row}_{i=2}^N \Pi_{1i} \quad W_2 := \mathrm{col}_{i=2}^N \Pi_{1i} \tag{32}$$

$$W_0 := \begin{pmatrix} \Pi_{2,2} & \cdots & \Pi_{2,N} \\ \vdots & \ddots & \vdots \\ \Pi_{N,2} & \cdots & \Pi_{N,N} \end{pmatrix}, \quad \Lambda := \begin{pmatrix} \widetilde{\Sigma}_{2,2} & \cdots & \widetilde{\Sigma}_{2,N} \\ \vdots & \ddots & \vdots \\ \widetilde{\Sigma}_{N,2} & \cdots & \widetilde{\Sigma}_{N,N} \end{pmatrix}. \tag{33}$$

With this definitions the equation (29) reads out as

$$\begin{pmatrix} 0 & 0 \\ 0 & \Lambda \end{pmatrix} = \begin{pmatrix} A_C & W_1 \\ W_2 & W_0 - \gamma D \otimes P_\infty \end{pmatrix} \widetilde{\mathbf{E}}_\infty(\gamma) + \widetilde{\mathbf{E}}_\infty(\gamma) \begin{pmatrix} A_C & W_1 \\ W_2 & W_0 - \gamma D \otimes P_\infty \end{pmatrix}^\top. \quad (34)$$

$$D = \mathrm{diag}_{i=2}^{N}(\lambda_i) \quad (35)$$

The solution $\widetilde{\mathbf{E}}_\infty(\gamma)$ is analytic in $\gamma > 0$ and can be written (using a Taylor expansion) as

$$\widetilde{\mathbf{E}}_\infty(\gamma) = \frac{1}{\gamma} \begin{pmatrix} Y_{1,1} + O\left(\frac{1}{\gamma^2}\right) & \frac{1}{\gamma} Y_{2,1} + O\left(\frac{1}{\gamma^2}\right) \\ \frac{1}{\gamma} Y_{2,1}^\top + O\left(\frac{1}{\gamma^2}\right) & Y_{3,1} + \frac{1}{\gamma} Y_{3,2} + O\left(\frac{1}{\gamma^2}\right) \end{pmatrix} \quad (36)$$

where $Y_{3,1}$ is the unique (since $\sigma(D \otimes P_\infty) \cap \sigma(-D \otimes P_\infty) = \varnothing$) solution of

$$Y_{3,1}(D \otimes P_\infty) + (D \otimes P_\infty) Y_{3,1} = N^2 \Lambda,$$

$Y_{3,2}$ is the unique (since $\sigma(D \otimes P_\infty) \cap \sigma(-D \otimes P_\infty) = \varnothing$) solution of

$$(D \otimes P_\infty) Y_{3,2} + Y_{3,2}(D \otimes P_\infty) = W_0 Y_{3,1} + Y_{3,1} W_0^\top,$$

$Y_{2,1}$ is defined as

$$Y_{2,1} := W_1 Y_{3,1} (D \otimes P_\infty)^{-1}$$

and $Y_{1,1}$ is the unique (since $\sigma(A_C) \cap \sigma(-A_C^\top) = \varnothing$) solution of

$$A_C Y_{1,1} + Y_{1,1} A_C^\top = -(W_1 Y_{2,1}^\top + Y_{2,1} W_1^\top).$$

From (36) it follows that $\lim_{\gamma \to +\infty} \widetilde{E}_\infty(\gamma) = 0$. □

Clearly, from the above lemmas we conclude that

$$\lim_{\gamma \to +\infty} \lim_{t \to +\infty} \widetilde{\mathbf{E}}_t = \lim_{\gamma \to +\infty} \lim_{t \to +\infty} \mathbf{E}_t = 0 \quad (37)$$

which proves Proposition 3.

An important consequence of proposition 3 is that the error covariance of each filter at the sensor node tends (as $\gamma \to +\infty$ and $t \to +\infty$) to the optimal steady state error covariance P_∞ of the centralized filter.

Corollary 1.

$$\lim_{\gamma \to +\infty} \lim_{t \to +\infty} [\mathbf{X}_t]_{j,j} = P_\infty \quad (38)$$

for each $j = 1, \ldots, N$.

Remark 1. It is worth remarking that in any discrete-time implementation of the ADKF the value of γ cannot be chosen arbitrarily large, due to numerical issues. Roughly speaking, a larger γ requires a smaller integration step that constraints the communication lag among nodes. Consequently, any implementable version of the ADKF will suffer a certain performance degradation with respect to the CKBF, in accordance with what happens in the discrete time framework.

Remark 2. The ADKF is extremely simple to implement and the information exchange among nodes is reduced to a minimum. The matrix P_∞ can be computed (off-line) by solving (4), a matrix equation with size n that does not depend on the graph structure. Clearly, with many sensor nodes the size of C and R can be large, but $C^\top R^{-1} C$ is a $n \times n$ matrix. The nodes can easily solve (4) provided that the value of $G = C^\top R^{-1} C$ is available. When measurement noises are independent G is expressed as in (20), that is, the sum of the matrices $C_i^\top R_i^{-1} C_i$ all over the graph. A distributed computation of G can thus be achieved by resorting to distributed algorithms to compute aggregate functions over graphs (Kempe et al. (2003)).

We want to add that also the calculation of the lower bound γ_0 for γ can be implemented in a distributed way. Indeed, it is possible to show that γ_0 can be lower bounded by some quantity which depend on G and P_∞ and, as noticed above, these quantities can be obtained through a distributed computation.

6. COMPARISON WITH RELATED APPROACHES

In this section we review some recent proposals of distributed filters for the continuous-time case, in order to highlight the similarities with the ADKF and the respective application scenarios.

We have reported the results about the stability of the errors of these algorithms. It is worth mentioning that none of them provides the exact variance of the estimation error at each node, an information that in the case of the ADKF is readily available by computing $\mathrm{tr}([\mathbf{X}_\infty]_{i,i})$.

6.1 DKF

The following continuous-time distributed filter was proposed in Olfati-Saber (2007). It modifies a previous proposal of a *consensus* filter of the same author (Olfati-Saber, 2005) with the aim of enhancing consensus. The DKF algorithm needs to exchange only the estimates among adjacent nodes and has equation

$$\dot{\hat{\mathbf{x}}}_t^{(i)} = A \hat{\mathbf{x}}_t^{(i)} + K_i \left(\mathbf{y}^{(i)} - C_i \hat{\mathbf{x}}_t^{(i)} \right) + \gamma P_i \sum_{j \in \mathcal{N}^{(i)}} \left(\hat{\mathbf{x}}_t^{(i)} - \hat{\mathbf{x}}_t^{(i)} \right) \quad (39)$$

$$\dot{P}_i = A P_i + P_i A^\top + Q - P_i C_i^\top R_i^{-1} C_i P_i \quad (40)$$

$$K_i = P_i C_i^\top R_i^{-1}, \quad (41)$$

with γ a positive parameter. The DKF algorithm Notice the similarity with (9). The only difference is that P_∞ is replaced by P_i. Clearly, P_i is bounded when (C_i, A) is an observable pair. A slight extension is to apply the algorithm by replacing $\mathbf{y}^{(i)}$ with $\mathrm{col}_{j \in \mathcal{N}^{(i)}}(\mathbf{y}^{(j)})$ and C_i with $\mathrm{col}_{j \in \mathcal{N}^{(i)}}(C_j)$. In any case, the DKF can be applied only when local observability conditions hold, and in particular $C_i \neq 0$, that is, all the sensor nodes (or their immediate neighbors) have measurements. In this condition it is possible to obtain a result about the boundedness of the estimation error variance.

6.2 MKCF

The Multi-agent Consensus Filter (MKCF) was originally proposed in Ren et al. (2005) for systems without dynamical equations. Its extension to our context was carried out in Ji et al. (2017). Here it is assumed that only the nodes for which $g_i = 1$ have access to the system output and in this case $C_i = I$. In other words, the measurement equation is

$$y^{(i)}(t) = g_i z_i = g_i (x + v_i). \quad (42)$$

The other nodes (*i.e.*, those with $g_i = 0$) must rely on the estimates of the neighbors as it is commonly assumed for

wireless sensor networks. The MKCF algorithm needs to exchange both the estimates and the matrices P_i among adjacent nodes and its equations are

$$\begin{aligned}\dot{\widehat{\mathbf{x}}}_t^{(i)} =& A\widehat{\mathbf{x}}_t^{(i)} + g_i P_i R_i^{-1}(z_i - \widehat{\mathbf{x}}_t^{(i)}) \\ &+ P_i \sum_{j\in\mathcal{N}^{(i)}} P_j^{-1}\left(\widehat{\mathbf{x}}_t^{(j)} - \widehat{\mathbf{x}}_t^{(i)}\right) \end{aligned} \tag{43}$$

$$\dot{P}_i = AP_i + P_iA^\top + Q - P_i\left(\sum_{j\in\mathcal{N}^{(i)}} P_j^{-1} + g_iR_i^{-1}\right)P_i. \tag{44}$$

There are no theoretical stability results for the estimation error of the MKCF. MKCF can applied to a wider sets of network with respect to DKF but it is limited by the assumption $C_i = I$.

6.3 IKCF

The information-wieghted Kalman consensus filter (IKCF) was proposed in Ji et al. (2017) as a modification of the MKCF, with which it shares the same assumptions except for the fact that nodes have a partial access to the state, *i.e.* $C_i \neq I$. The measurement equation is

$$y^{(i)}(t) = g_i z_i = g_i(C_i x + v_i). \tag{45}$$

In addition, IKCF explicitly models communication noises among nodes over directed graphs. The IKCF algorithm needs to exchange both the estimates and the matrices P_i among adjacent nodes and, for the case without communication noises, its equations are

$$\begin{aligned}\dot{\widehat{\mathbf{x}}}_t^{(i)} =& A\widehat{\mathbf{x}}_t^{(i)} + g_i P_i C_i^\top R_i^{-1}(z_i - C_i\widehat{\mathbf{x}}^{(i)}) \\ &+ P_i \sum_{j\in\mathcal{N}^{(i)}} P_j^{-1}\left(\widehat{\mathbf{x}}_t^{(j)} - \widehat{\mathbf{x}}_t^{(i)}\right) \end{aligned} \tag{46}$$

$$\begin{aligned}\dot{P}_i =& (A+\nu_i I)P_i + P_i(A+\nu_i I)^\top + Q - g_i^2 P_i C_i^\top R_i^{-1} C_i P_i \\ &- P_i\left(\sum_{j\in\mathcal{N}^{(i)}}\left(P_j^{-1} - P_i^{-1}\right)\right)P_i. \end{aligned} \tag{47}$$

The proof of the boundedness of the estimation error provided in Ji et al. (2017) contains some technical issues, related to the fact that the P_is do not actually correspond to the covariance of the estimation error at each node. Thus, the claim in Ji et al. (2017) that the IKCF provides an optimal estimate is not correct (Cacace, 2019).

7. SIMULATION RESULTS

Consider system (1) with

$$A = \begin{bmatrix} -0.1 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 \\ 1.5 & 0 & -0.2 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix} \tag{48}$$

We consider several scenarios that range from full state measurement for each node to sparse partial measurement (the latter case is typical of wireless sensor networks). We choose $N = 5$ with the five nodes connected in a chain, i.e. the edges are $(i, i+1)$, $i \in \{1, 2, 3, 4\}$.

Scenario 1: complete information. The system state is available to all nodes, *i.e.* $C_i = I_4$.

Scenario 2: local observability. The system output is available to all nodes and $C_i = [1,\ 1,\ 1,\ 1]$. Note that (C_i, A) is observable.

Scenario 3: collective observability. The system output is available to all nodes, with $C_1 = [1,\ 0,\ 1,\ 0]$, $C_2 = [0,\ 1,\ 0,\ 0]$, $C_3 = [1,\ 0,\ 0,\ 0]$, $C_4 = [1,\ 0,\ 1,\ 1]$, $C_5 = [0,\ 1,\ 1,\ 0]$. Note that (C_i, A) is never observable, but (C, A) is observable.

Scenario 4: sparse state availability. The system state is available only to nodes in $\{1, 5\}$, *i.e.* $C_1 = C_5 = I_4$, $C_2 = C_3 = C_4 = 0$.

Scenario 5: sparse measurement availability. The system output is available only to nodes in $\{1, 5\}$, with

$$C_1 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad C_5 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \tag{49}$$

$C_2 = C_3 = C_4 = 0$. Note that (C_1, A) is not observable but (C_2, A) is observable.

In order to compare the performance of the filters we have performed 100 simulations with $t \in [0, 50]$ and integration step $dt = 2 \cdot 10^{-3}$. The noise amplitude of the state noise has been chosen as $Q = q_f I_4$ with $q_f = 0.3$ and $R_i = r_f I_i$, where $r_f = 0.6$ and I_i has the same number of rows as C_i. For ADKF and DKF the parameter γ was set to $\gamma = 100$.

The results are summarized in Table 1. Notice that when it can be applied, the DKF attains almost optimal performance and a larger consensus than ADKF. This fact however may depend on the value of γ. To investigate this issue we have computed the variance of the estimation error $\mathrm{tr}([\mathbf{X}_\infty]_{i,i})$ at each node as a function of γ in Scenario 3 by using (18). The results are shown in Fig. 1 (left). The plot confirms the result of Proposition 3: when γ increases the variance of the estimation error of all the nodes converge to the optimal value $\mathrm{tr}(P_\infty)$. Fig. 1 (right) shows that the consensus, expressed as the standard deviation across nodes of the variance of the estimation error, is a linear function of γ.

Remark 3. DKF can be used only when local observability conditions hold, i.e., in scenarios 1 and 2. This is clear from the equation (40) for P that may become unstable when (C_i, A) is not as observable pair. However, when there are no measurements at nodes in which (C_i, A) is not observable, the value K_i is not needed. The filter can therefore be implemented by omitting the (potentially unstable) computation of P_i and replacing P_i in (39) with an arbitrary matrix, for example I. This is the approach that we have used in scenario 4. Clearly, this is not possible in scenario 3 where the gain K_i is needed.

Remark 4. MKCF needs $C_i = I$, thus it can be natively used in scenarios 1 and 4 because P_i and R_i must have the same size (see (44). However, when R_i is scalar the terms are still congruent, thus we used MKCF also for scenarios 2, 3.

8. CONCLUSIONS

As mentioned in Remark 1 the results described in this work for the case of continuous-time filters do not extend automatically to the discrete-time case. Thus it is of interest to derive a discrete-time implementation of the ADKF and to characterize the loss of accuracy and consensus of this discrete-time counterpart with respect to the optimal case as a function of the size of the discretization

		$\mathrm{tr}(P_\infty)$	CKBF	ADKF	DKF	MKCF	IKCF
Scenario 1	mse	0.319	0.344	0.394	0.408	7.374	1.544
	st.dev.			0.031	0.004	0.411	0.155
Scenario 2	mse	0.797	0.868	0.814	0.820	2.539	$> 10^3$
	st.dev.			0.015	0.003	0.078	$> 10^3$
Scenario 3	mse	0.553	0.595	0.608	N.A.	17.207	2.818
	st.dev.			0.029	N.A.	3.632	0.703
Scenario 4	mse	0.532	0.570	0.610	0.727	$> 10^2$	1.406
	st.dev.			0.064	$< 10^{-3}$	> 10	0.481
Scenario 5	mse	0.582	0.626	0.672	N.A.	N.A.	2.519
	st.dev.			0.050	N.A.	N.A.	0.519

Table 1. Comparison of **mse** and **consensus** among filters in five different scenarios. Consensus is measured by the mse standard deviation across nodes (smaller values are better).

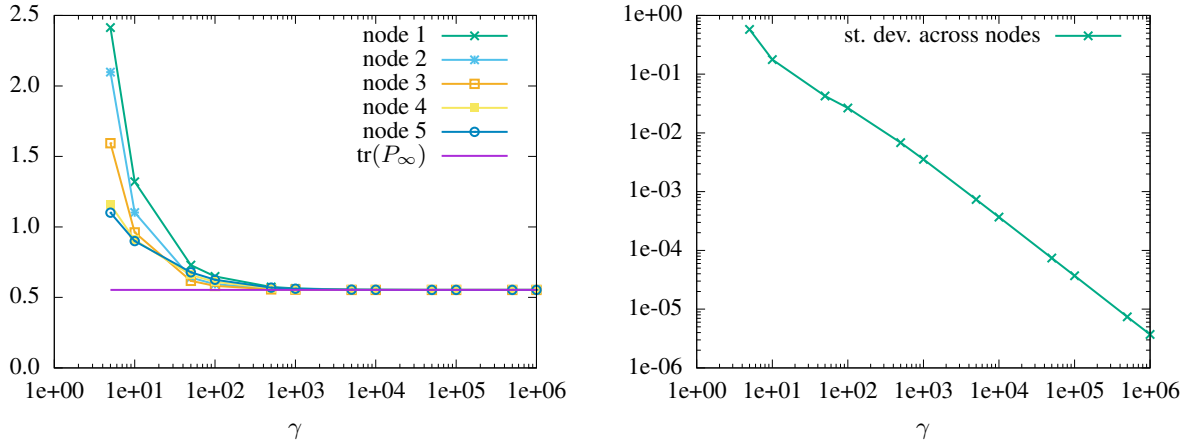

Fig. 1. $\mathrm{tr}([\mathbf{X}_\infty]_{i,i})$ for the nodes of Scenario 3, $i \in \{1,2,3,4,5\}$, compared with the optimal variance $\mathrm{tr}(P_\infty)$ as a function of γ (left). Standard deviation across nodes of the estimation error variances shows that consensus increases with γ (right).

interval. Additional extensions may include connectivity hypotheses for the case of directed network topology and the explicit modeling of communications delay and/or disturbances.

REFERENCES

Battistelli, G., Chisci, L., Mugnai, G., Farina, A., and Graziano, A. (2014). Consensus-based linear and nonlinear filtering. *IEEE Transactions on Automatic Control*, 60(5), 1410–1415.

Cacace, F. (2019). Comments on "distributed information-weighted Kalman consensus filter for sensor networks". *Automatica*, 109, 108552.

Chong, C.Y. (2017). Forty years of distributed estimation: A review of noteworthy developments. In *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 1–10. IEEE.

Ji, H., Lewis, F.L., Hou, Z., and Mikulski, D. (2017). Distributed information-weighted Kalman consensus filter for sensor networks. *Automatica*, 77, 18–30.

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.

Kempe, D., Dobra, A., and Gehrke, J. (2003). Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, 482–491. IEEE.

Lewis, F.L., Zhang, H., Hengster-Movric, K., and Das, A. (2013). *Cooperative control of multi-agent systems: optimal and adaptive design approaches*. Springer Science & Business Media.

Olfati-Saber, R. (2005). Distributed Kalman filter with embedded consensus filters. In *Proceedings of the 44th IEEE Conference on Decision and Control*, 8179–8184. IEEE.

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

Olfati-Saber, R., Fax, J.A., and Murray, R.M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1), 215–233.

Qu, Z. (2009). *Cooperative control of dynamical systems: applications to autonomous vehicles*. Springer Science & Business Media.

Ren, W., Beard, R.W., and Kingston, D.B. (2005). Multi-agent Kalman consensus with relative uncertainty. In *Proceedings of the 2005, American Control Conference, 2005.*, 1865–1870. IEEE.

Taj, M. and Cavallaro, A. (2011). Distributed and decentralized multicamera tracking. *IEEE Signal Processing Magazine*, 28(3), 46–58.