A stability with optimality analysis of consensus-based distributed filters for discrete-time linear systems

S. Battilotti [a] F. Cacace [b] M. d'Angelo [a]

[a] *DIAG, Università "Sapienza" di Roma, Roma, Italy, (e-mail: battilotti@diag.uniroma1.it, mdangelo@diag.uniroma1.it).*

[b] *Università Campus Bio-Medico di Roma, Roma, Italy (e-mail: f.cacace@unicampus.it).*

Abstract

In this paper we investigate how stability and optimality of consensus-based distributed filters depend on the number of consensus steps in a discrete-time setting for both directed and undirected graphs. By introducing two new algorithms, a simpler one based on dynamic averaging of the estimates and a more complex version where local error covariance matrices are exchanged as well, we are able to derive a complete theoretical analysis. In particular we show that dynamic averaging alone suffices to approximate the optimal centralized estimate if the number of consensus steps is large enough and that the number of consensus steps needed for stability can be computed in a distributed way. These results shed light on the advantages as well as the fundamental limitations shared by all the existing proposals for this class of algorithms in the basic case of linear time-invariant systems, that are relevant for the analysis of more complex situations.

Key words: Discrete time filters; Kalman filters; Filtering theory; Consensus filters

1 Introduction

The availability of low-cost sensors and the diffusion of wireless networks have contributed in recent years to the development of applications based on wireless sensor networks for target tracking and estimation in a broad range of areas such as environmental monitoring (?), airborne target tracking (?), space situation awareness (?), spacecraft navigation (?) etc. Mainly due to this surge of interest in the application of wireless sensor networks to large-scale estimation and control problems, distributed estimation and filtering has become one of the most active topics in the area of filtering theory (?).

Distributed estimation is based on the usage of multiple sensor nodes to cooperatively perform large-scale sensing tasks that cannot be accomplished by individual devices. The use of redundant and cooperating sources of information in a completely distributed architecture may enhance flexibility of deployment, efficiency, robustness and accuracy of the estimates. A distributed algorithm dictates the way in which the information is exchanged and elaborated by the nodes of the network in order to reach a shared estimate of the target systems state under the constraint that each node can communicate only with its neighbors. Additional challenges include the limited processing power of the single nodes, that usually are low-cost devices, limited communication bandwidth, dynamic network topology, the presence of unreliable communications channel and/or communication delays, heterogeneity of the sensor with respect to the measurements that they have of the target system, and limited energy (?). It is often assumed that nodes can either be communication nodes, that only have communication and processing capabilities, or sensor nodes that in addition have sensing capabilities (?). In view of this, a distributed estimation algorithm is evaluated with respect to: a) *accuracy of estimates*, the optimal reference being, in the case of linear Gaussian systems, the centralized Kalman filter that makes use of the measurements of all the sensors; b) *consensus on estimates* among nodes, that is particularly relevant when the estimates are used for control purposes (??); c) *communication cost*, expressed as the amount of information exchanged among adjacent nodes; d) capability of guaranteeing *stability* under minimal requirements of network connectivity and system collective observability (?).

In this paper we consider linear time invariant systems in discrete-time. The case of linear time-varying or nonlinear systems is studied among many others in (????????). We also do not consider the presence of model uncertainties (?), intermittent observations (?), delays (??), asynchronous communication (?), unreliable communication links (**??**), sensor bias (**?**), etc.

In consensus-based filters the information exchanged among adjacent nodes can either be the local state estimates (*consensus on estimates*, CE) as in the Kalman Consensus Filter (KCF) (**?**), see also **?** and **?**, or the information (matrix-vector) pair (*consensus on information*, CI (**?**)) or the innovation information pair (*consensus on measurements*, CM). The CI filters (**???**) reduce to the covariance-intersection method (**?**) when only one consensus step is performed. They do not usually converge to optimal centralized estimates whatever the number of consensus steps and fusion strategies due to the fact that they underweight the innovation terms, but have good stability properties and generate unbiased and consistent local estimates even for a limited number of consensus iterations. CM filters include the CM Kalman filter (CMKF) originally proposed in (**?**) and its related variants (**???**) and gossip versions (**?**), as well as the Iterative Consensus Fiter (ICF, **?**). These filters recover the optimal centralized performance when the number of consensus steps tends to infinity but they do not preserve consistency of local estimates. To address this problem the Hybrid CM and CI (HCMCI) filter of **?** proposes a double consensus iteration on both the priors estimates and the new measurements, thus attaining error stability and consistency with any number of consensus steps. Finally, in a distributed optimization perspective **?** derive a CI filter that converges to the optimal centralized estimate.

Motivated by the lack of complete theoretical results in the literature, in this paper we investigate some basic properties of consensus filters, namely (i) convergence to the optimal estimate when the number of consensus steps increases; (ii) impact of covariance matrices exchange on performance and related stability issues; (iii) computation of the minimal number of consensus steps to achieve error stability. This analysis highlights fundamental limitations and trade-offs that remain valid for more general classes of systems.

The contributions and novelties of the paper are summarized below.

(1) We describe two CE filters. In the simpler one (Distributed Kalman Filter, DKF) only estimates are exchanged. In the other one (Steady-state Modified Distributed Kalman Filter, SMDKF) also covariance matrix are exchanged. By means of simulations we show that SMDKF yields the same performance as CM filters, thus establishing an equivalence between CE and CM consensus filters.
(2) We provide a complete theoretical analysis of these filters: stability and performance and their dependence on the number of consensus steps. Since we show that none of the algorithms proposed in the literature has acceptable performance with only one consensus step, we argue that in practice this result is more significant than establishing solely the stability of the error with any number of consensus steps.
(3) A prominent and new result of this analysis is that for a large number of consensus steps the simpler algorithm that only exchanges estimates tends to the optimal centralized estimate, in analogy with the continuous-time case (**?**).
(4) We prove that, as expected, algorithms that exchange covariance matrices reduce the minimum number of consensus steps needed for mean square stability of the estimation error.
(5) We prove that for static networks it is possible to compute in a distributed way the minimal number of consensus steps needed for stability.

The problem is formalized in Section 2. The DKF and its properties are the object of Section 3. We consider directed graphs in Section 4 and the deterministic case in Section 5. Section 6 introduces and studies the SMDKF. A comparison based on simulations among a few distributed consensus filters is presented in Section 7 for a marginally stable and an unstable system. This comparison validates the theoretical analysis and highlights common limitations of the existing approaches.

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 Schur stable if $\sigma(A)$ lies inside the unit circle, i.e. it is in $\mathbb{S}^1$ ($\mathbb{S}^1$ the open unit circle in the complex plane). $\|A\|$, $A \in \mathbb{R}^{n\times m}$, denotes the matrix operator norm. $\mathbb{E}\{\cdot\}$ 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. $\mathcal{S}(n) \in \mathbb{R}^{n\times n}$ is 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 $\mathcal{S}(n)$. I_n is the identity matrix in $\mathbb{R}^n$ and $U_n = 1_n 1_n^\top$, $1_n = \mathrm{col}_{i=1}^n(1)$, is the square matrix of size n having 1 in each entry.

2 Problem formulation and preliminaries

The undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ describes the information exchange between the agents. $\mathcal{V} = \{1, 2 \ldots, N\}$ 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, with $i \neq j$, are neighbors to each other if $(i, j) \in \mathcal{E}$. The set of neighbors of node i is denoted by $\mathcal{N}^{(i)} := \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$. 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 $(i, j) \in \mathcal{E}$ and 0 otherwise. The degree matrix $\mathcal{D}$ of $\mathcal{G}$ is a diagonal matrix whose i-th entry is the cardinality of $\mathcal{N}^{(i)}$, denoted $\#\mathcal{N}^{(i)}$. The Laplacian of $\mathcal{G}$ is the $N \times N$ 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 1_N. Consider the process

$$\mathbf{x}_{k+1} = A\mathbf{x}_k + \mathbf{f}_k, \tag{1}$$

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

where $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{y}_k^{(i)} \in \mathbb{R}^{q_i}$, $q_i \geqslant 0$, and $\mathbf{f}_k$ and $\mathbf{g}_k^{(i)}$, $i = 1, \ldots, N$, are zero-mean white noises, mutually independent with covariance respectively $Q \in \mathcal{P}(n)$, $R_i \in \mathcal{P}_+(q_i)$ $i = 1, \ldots, N$. The matrix $R = \mathrm{diag}_i(R_i)$ is nonsingular. $\mathbf{x}_0$ is a random variable with mean $\bar{\mathbf{x}}_0 := \mathbb{E}\{\mathbf{x}_0\}$ and covariance $\Sigma_{\mathbf{x}_0}$. $C = \mathrm{col}_i(C_i)$ is the aggregate matrix of the output maps. Throughout the paper we assume that the couple (C, A) is observable and the couple $(A, Q^{\frac{1}{2}})$ is controllable. We denote $\mathbf{y}_k = \mathrm{col}_i(\mathbf{y}_k^{(i)})$, each $\mathbf{y}_k^{(i)}$ represents the data available at node i, $i = 1, \ldots, N$, in the network. We consider the problem of designing an optimal distributed state estimator for the system (1) with the topology $\mathcal{G}$ of the network. The estimator consists of N local estimators, one for each node, that exchange local information with the neighbors.

2.1 Centralized Kalman filter (CKF)

The equations of the centralized Kalman filter (CKF) for (1) are

$$\begin{aligned}
\hat{\mathbf{x}}_{0|-1} &= \bar{\mathbf{x}}_0, \; \mathbf{P}_{0|-1} = \Sigma_{\mathbf{x}_0}, \\
\hat{\mathbf{x}}_{k+1|k} &= A\hat{\mathbf{x}}_{k|k}, \; \mathbf{P}_{k+1|k} = A\mathbf{P}_{k|k}A^\top + Q, \\
\mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k}(I + C^\top R^{-1} C\mathbf{P}_{k+1|k})^{-1} \\
\mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k+1}C^\top R^{-1}, \\
\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - C\hat{\mathbf{x}}_{k+1|k}).
\end{aligned} \tag{3}$$

The matrix $\mathbf{P}_{k+1|k+1}$ represents the covariance of the estimation error $\mathbb{E}\{(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1})^\top\}$ and $\mathbf{P}_{k+1|k}$ the covariance of the one-step prediction error. The CKF is optimal in the sense that it computes the conditional expectation $\mathbb{E}\{\mathbf{x}_{k+1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_{k+1}\}$, i.e. the projection of the state $\mathbf{x}_{k+1}$ onto the σ-algebra generated by the output sequence $\mathbf{y}_1, \ldots, \mathbf{y}_{k+1}$. We have that the covariance $\mathbf{P}_{k+1} \in \mathcal{P}_+(n)$ is bounded for all $k \geqslant 0$ and $\mathbf{P}_{k+1|k+1} \to P_\infty$ as $t \to +\infty$ with $P_\infty \in \mathcal{P}_+(n)$ the unique solution of

$$\begin{aligned}
P_\infty = A_C P_\infty A_C^\top + (I - K_\infty C)Q(I - K_\infty C)^\top \\
+ P_\infty C^\top R^{-1} C P_\infty
\end{aligned} \tag{4}$$

with

$$A_C := (I - K_\infty C)A, \quad K_\infty = P_\infty C^\top R^{-1}. \tag{5}$$

The covariance equation in (3) can be also written using the matrix inversion lemma and $\mathbf{P}_{k+1|k}$ as

$$\begin{aligned}
\mathbf{P}_{k+1|k+1} &= (I - \mathbf{K}_{k+1}C)\mathbf{P}_{k+1|k} \\
&= \mathbf{P}_{k+1|k} - (CR^{-1}C + \mathbf{P}_{k+1|k}^{-1})^{-1} \times \\
&\times CR^{-1}C\mathbf{P}_{k+1|k} = (CR^{-1}C + \mathbf{P}_{k+1|k}^{-1})^{-1}.
\end{aligned}$$

From (3), we also obtain the asymptotically optimal CKF

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + K_\infty(\mathbf{y}_{k+1} - C\hat{\mathbf{x}}_{k+1|k}), \tag{6}$$

where K_∞ is defined in (5).

3 Asymptotically optimal distributed Kalman filter (DKF)

In this section we prove that dynamic averaging alone allows to approximate arbitrarily well the optimal centralized estimate for a sufficiently large number of consensus step. The result is interesting because the derived algorithm is very simple as it does not require to exchange covariance matrices among nodes.

3.1 DKF algorithm

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

$$\hat{\mathbf{x}}_{k+1|k}^{(i)} = A\hat{\mathbf{x}}_{k|k}^{(i)} \tag{7}$$

$$\begin{cases}
\mathbf{z}_{k+1,0}^{(i)} = \hat{\mathbf{x}}_{k+1|k}^{(i)} + K_i(\mathbf{y}_{k+1}^{(i)} - C_i\hat{\mathbf{x}}_{k+1|k}^{(i)}), \\
\mathbf{z}_{k+1,h+1}^{(i)} = \mathbf{z}_{k+1,h}^{(i)} + \frac{1}{\delta}\sum_{j\in\mathcal{N}^{(i)}}(\mathbf{z}_{k+1,h}^{(j)} - \mathbf{z}_{k+1,h}^{(i)})
\end{cases} \tag{8}$$

$$\hat{\mathbf{x}}_{k+1|k+1}^{(i)} = \mathbf{z}_{k+1,\gamma}^{(i)}, \tag{9}$$

where $h = 0, \ldots, \gamma - 1$, and $K_i := NP_\infty {C_i}^\top {R_i}^{-1}$ and P_∞ is the solution of (4), $\delta > \lambda_N(\mathcal{L})$ and $\gamma \in \mathbb{N}$ is a parameter to be chosen as we shall point out later. The filter consists of a one-step prediction (7), a dynamic averaging step (8), in which the neighbor estimates are mixed up at each node dynamically through γ iterations and the estimate update (9). Define

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

and

$$A_D(\gamma) := J^\gamma \mathrm{diag}_i(A_i), \quad J := \left(I_N - \frac{\mathcal{L}}{\delta}\right) \otimes I_n. \quad (11)$$

The parameter γ is chosen accordingly with the next proposition.

Proposition 1 *There exists $\gamma_0 \in \mathbb{N}$ such that, for all integers $\gamma > \gamma_0$, $A_D(\gamma)$ is Schur stable.*

Proof. There clearly exists a orthonormal transformation $T = \begin{pmatrix} V \\ W \end{pmatrix}$, with $V = \frac{1}{\sqrt{N}} 1_N^\top$ and $W \in \mathbb{R}^{(N-1)\times N}$, such that

$$T\mathcal{L}T^\top = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda \end{pmatrix},$$

where $\Lambda = \mathrm{diag}(\lambda_2(\mathcal{L}), \ldots, \lambda_N(\mathcal{L}))$, and

$$\mathcal{L}V^\top = 0, \quad (12)$$

$$\mathcal{L}W^\top = W^\top \Lambda. \quad (13)$$

Moreover,

$$(V \otimes I_n)\mathrm{diag}_i(A_i)(V^\top \otimes I_n) = A_C, \quad (14)$$

and by (12) and (13) it easy to prove by induction the following relations:

$$\left(I_N - \frac{\mathcal{L}}{\delta}\right)^\gamma V^\top = V^\top, \quad (15)$$

$$\left(I_N - \frac{\mathcal{L}}{\delta}\right)^\gamma W^\top = W^\top \left(I_{N-1} - \frac{\Lambda}{\delta}\right)^\gamma. \quad (16)$$

Consider now the matrix

$$S := I_{N-1} - \frac{\Lambda}{\delta}.$$

By direct computations using (14) and (16), we obtain

$$(T \otimes I_n)A_D(\gamma)(T^\top \otimes I_n) = \begin{pmatrix} A_C & H_{12} \\ H_{21}(\gamma) & H_{22}(\gamma) \end{pmatrix}, \quad (17)$$

where

$$\begin{aligned} H_{12} &= (V \otimes I_n)\mathrm{diag}_i(A_i)(W^\top \otimes I_n)), \\ H_{21}(\gamma) &= (S^\gamma W \otimes I_n)\mathrm{diag}_i(A_i)(V^\top \otimes I_n), \\ H_{22}(\gamma) &= (S^\gamma W \otimes I_n)\mathrm{diag}_i(A_i)(W^\top \otimes I_n)). \end{aligned}$$

Notice that by (16) and since $\delta > \lambda_N(\mathcal{L})$, $S^\gamma \to 0$ as $\gamma \to +\infty$, so that

$$H_{21}(\gamma) \to 0 \text{ and } H_{22}(\gamma) \to 0 \quad \text{as } \gamma \to +\infty. \quad (18)$$

Thus, since A_C is Schur stable by construction, it follows that there exists $\gamma_0 \in \mathbb{N}$ such that for all integer $\gamma > \gamma_0$ $A_D(\gamma)$ is Schur stable. □

Remark 1 *The choice $\delta > \lambda_N(\mathcal{L})$ in (8) is such that S is Schur stable. However, other less conservative choices of δ are feasible with this property: for example, a good choice for δ is the maximum number of neighbors for any node, i.e. $\max_{i\in\mathcal{V}}\{\#\mathcal{N}^{(i)}\}$.*

3.2 Properties of the DKF

In order to study the asymptotic properties of the DKF we introduce at each node the local estimation error $\mathbf{e}_{k|k}^{(i)} := \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i)}$, the total estimation error $\mathbf{e}_{k|k} := \mathrm{col}_i(\mathbf{e}_{k|k}^{(i)})$, the total measurement noise error vector $\mathbf{g}_k := \mathrm{col}_i(\mathbf{g}_k^{(i)}) \in \mathbb{R}^{\sum_{i=1}^N q_i}$, $\mathbf{h}_k := J^\gamma \mathrm{col}_i(\mathbf{h}_k^{(i)}) \in \mathbb{R}^{nN}$ where

$$\mathbf{h}_k^{(i)} := -K_i \mathbf{g}_{k+1}^{(i)} + (I - K_i C_i)\mathbf{f}_k$$

with covariance $\Psi := \mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^\top\}$, and the estimation error covariance matrix $\mathbf{X}_{k|k} := \mathbb{E}\{\mathbf{e}_{k|k}\mathbf{e}_{k|k}^\top\}$. Clearly, $\mathbf{X}_{k|k}$ depends on γ, but we omit this dependence for notational simplicity. After some lengthy manipulations we obtain the following error equations

$$\mathbf{e}_{k+1|k+1} = A_D(\gamma)\mathbf{e}_{k|k} + \mathbf{h}_k \quad (19)$$

with

$$\begin{aligned} \Psi = J^\gamma \Big\{ &\mathrm{diag}_i(I - K_iC_i)(U_N \otimes Q)\mathrm{diag}_i((I - K_iC_i)^\top) \\ &+ \mathrm{diag}_i(K_i) R \, \mathrm{diag}_i(K_i^\top) \Big\} J^\gamma, \end{aligned} \quad (20)$$

where U_N is a matrix with all entries 1. Using uncorrelation between $\mathbf{e}_{k|k}$ and $\mathbf{h}_k$, the error covariance equation turns out to be

$$\mathbf{X}_{k+1|k+1} = A_D(\gamma)\mathbf{X}_{k|k}A_D^\top(\gamma) + \Psi.$$

Proposition 2 *For all $\gamma > \gamma_0$ the estimation error covariance matrix $\mathbf{X}_{k|k}$ is uniformly bounded in time k and*

$$\lim_{k\to+\infty} \mathbf{X}_{k|k} = \mathbf{X}_\infty(\gamma) \quad (21)$$

where $\mathbf{X}_\infty(\gamma)$ is the unique solution of

$$\mathbf{X}_\infty(\gamma) = A_D(\gamma)\mathbf{X}_\infty(\gamma)A_D^\top(\gamma) + \Psi. \quad (22)$$

Proof. The result follows by standard arguments from the fact that $A_D(\gamma)$ is Schur stable for $\gamma > \gamma_0$ and Ψ defined in (20) is constant. The covariance matrix $\mathbf{X}_{k|k}$ obeys for all $k \geqslant 0$ to

$$\mathbf{X}_{k+1|k+1} = A_D(\gamma)\mathbf{X}_{k|k}A_D^\top(\gamma) + \Psi,$$

and its asymptotic value $\mathbf{X}_\infty$ is the solution of (22). □

Our purpose is to show the key result that $\mathbf{X}_\infty(\gamma) \to U_N \otimes P_\infty$ when $\gamma \to \infty$ (recall that P_∞ is the asymptotic error covariance of the CKF). Let $\mathbf{X}_\infty^C := U_N \otimes P_\infty$ be the the asymptotic error covariance of N identical CKFs implemented at each node and using the whole output $\mathbf{y}_t$. Recalling that $(\mathcal{L}\otimes P_\infty)\mathbf{X}_\infty^C = (\mathcal{L}\otimes P_\infty)(U_N \otimes P_\infty) = 0$ and using (5), we obtain that $\mathbf{X}_\infty^C$ satisfies

$$\begin{aligned}\mathbf{X}_\infty^C =& \mathrm{diag}_i(A_C)\mathbf{X}_\infty^C\mathrm{diag}_i(A_C^\top)\\ &+ \mathrm{diag}_i(I - K_\infty C)(U_N \otimes Q)\mathrm{diag}_i((I - K_\infty C)^\top)\\ &+ \mathrm{diag}_i(K_\infty)(U_N \otimes R)\mathrm{diag}_i(K_\infty^\top). \qquad (23)\end{aligned}$$

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

$$\mathbf{E}(\gamma) = A_D(\gamma)\mathbf{E}(\gamma)A_D^\top(\gamma) + \Sigma(\gamma), \qquad (24)$$

where

$$\begin{aligned}\Sigma(\gamma) :=& A_D(\gamma)\mathbf{X}_\infty^C A_D^\top(\gamma) - \mathrm{diag}_i(A_C)\mathbf{X}_\infty^C\mathrm{diag}_i(A_C^\top)\\ &+J^\gamma\Big\{\mathrm{diag}_i(I - K_iC_i)(U_N \otimes Q)\mathrm{diag}_i((I - K_iC_i)^\top)\\ &+\mathrm{diag}_i(K_i)R\,\mathrm{diag}_i(K_i^\top)\Big\}J^\gamma\\ &-\mathrm{diag}_i(I - K_\infty C)(U_N \otimes Q)\mathrm{diag}_i((I - K_\infty C)^\top)\\ &-\mathrm{diag}_i(K_\infty)(U_N \otimes R)\mathrm{diag}_i(K_\infty^\top).\end{aligned}$$

Our main result can thus be stated as follows.

Proposition 3 *As $\gamma \to +\infty$, the covariance matrix of the estimation error of the DKF* (7)–(9) *tends to the covariance matrix of the estimation error of the CKF* (6) *when $k \to +\infty$. In other words, we have*

$$\lim_{\gamma\to+\infty}\lim_{k\to+\infty}\mathbf{X}_{k|k} = \mathbf{X}_\infty^C := U_N \otimes P_\infty.$$

Proof. On account of Proposition 2, it is sufficient to prove that

$$\lim_{\gamma\to+\infty}\mathbf{E}(\gamma) = 0. \qquad (25)$$

Let $T = \begin{pmatrix} V \\ W \end{pmatrix}$, with $V = \frac{1}{\sqrt{N}}1_N^\top$ and $W \in \mathbb{R}^{(N-1)\times N}$, be the same orthonormal transformation we considered in the proof of proposition 1. After some lengthy computations and using (15), (16) and (17) , we obtain that

$$(T \otimes I_n)\Sigma(\gamma)(T^\top \otimes I_n) = \begin{pmatrix} 0 & D_{12}(\gamma) \\ D_{21}(\gamma) & D_{22}(\gamma) \end{pmatrix} \qquad (26)$$

where T is as in the proof of Proposition 1 and the matrices $D_{12}(\gamma)$, $D_{21}(\gamma)$, and $D_{22}(\gamma)$ can be computed similarly as before and are such that

$$D_{12}(\gamma), D_{21}(\gamma), D_{22}(\gamma) \to 0 \text{ as } \gamma \to +\infty. \qquad (27)$$

From (17), (24) and (26)

$$\begin{aligned}\widetilde{\mathbf{E}}(\gamma) &:= (T \otimes I_n)\mathbf{E}(\gamma)(T^\top \otimes I_n)\\ &= \begin{pmatrix} A_C & H_{12} \\ H_{21}(\gamma) & H_{22}(\gamma) \end{pmatrix}\widetilde{\mathbf{E}}(\gamma)\begin{pmatrix} A_C^\top & H_{21}^\top(\gamma) \\ H_{12}^\top & H_{22}^\top(\gamma) \end{pmatrix}\\ &+ \begin{pmatrix} 0 & D_{12}(\gamma) \\ D_{21}(\gamma) & D_{22}(\gamma) \end{pmatrix}, \qquad (28)\end{aligned}$$

where $H_{12}, H_{21}(\gamma)$ and $H_{22}(\gamma)$ are as in the proof of Proposition 1. On account of (18) and (27) the unique solution $\widetilde{\mathbf{E}}(\gamma)$ of (28) tends as $\gamma \to +\infty$ to the unique solution $\widetilde{\mathbf{E}}_\infty$ of the equation

$$\widetilde{\mathbf{E}}_\infty = \begin{pmatrix} A_C & H_{12} \\ 0 & 0 \end{pmatrix}\widetilde{\mathbf{E}}_\infty\begin{pmatrix} A_C^\top & 0 \\ H_{12}^\top & 0 \end{pmatrix}$$

and since A_C is Schur stable it follows that $\widetilde{\mathbf{E}}_\infty = 0$, which implies (25). □

3.3 A distributed computation of P_∞ and $C^\top R^{-1}C$

In order to implement the DKF, each node i needs to compute (or to know) the value of the matrix P_∞, that depends on all the nodes of the graph. This may seem to prevent a truly distributed computation, thus the aim of this section is to show how the DKF can be implemented in a completely distributed manner.

In the first place it is worth remarking that computing P_∞ by solving (4) is trivial, since P_∞ is the solution of a matrix equation in $\mathbb{R}^{n\times n}$ that does not depend on the size of the graph. Thus, also nodes with limited computational power 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 $\sum_{j=1}^N C_i^\top R_i^{-1}C_i$, 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 (**?**). In Fig. 1 we report an

Algorithm Broadcast Push-Sum

1: In all nodes set $s_{0,i} = C_i^\top R_i^{-1} C_i$ and $w_{0,i} = 0$, except for $w_{0,1} = 1$.
2: At time 0 each nodes sends $(s_{0,i}, w_{0,i})$ to itself.
3: At time t each node executes:
   1. Let $\{s_r, w_r\}$ be the pairs sent to i in round $t-1$.
   2. Let $s_{t,i} = \sum_r s_r$, $w_{t,i} = \sum_r w_r$.
   3. Send to all neighbors and to i (yourself):

$$\left(\frac{1}{|\mathcal{N}^{(i)}|+1} s_{t,i},\ \frac{1}{|\mathcal{N}^{(i)}|+1} w_{t,i}\right)$$

   4. $s_{t,i}/w_{t_i}$ is the estimate of G at step t (if $w_{t,i} = 0$ the estimate is not specified or 0).

Fig. 1. A modified version of the *Push-Sum* algorithm of (**?**) for the distributed computation of G.

algorithm derived from the *Protocol Push-Sum* of (**?**) to compute G in a distributed way. The main difference is that (**?**) is a gossip algorithm with peer-to-peer communication, whereas the algorithm in Fig. 1 is a diffusion protocol with the node that broadcasts messages to all its neighbors. The speed of convergence of the local estimate to the true value of G can be analyzed in the light of the results of (**?**). This estimation phase can be executed before the filtering phase for static graphs, or it can be kept running during the execution of the filter in order to adjust the value of G in presence of a dynamical graph where nodes appear or disconnect. Finally, the value of N can be computed by the same distributed algorithm when it is not known at the nodes.

3.4 Lower bounds for γ

In this section we obtain a lower bound for γ, the number of iterations in the DKF (7)–(9) to ensure the stability of the overall error dynamics. A lower bound for γ can be obtained by guaranteeing the error system

$$\mathbf{e}_{k+1|k+1} = A_D(\gamma)\mathbf{e}_{k|k} \tag{29}$$

is asymptotically stable. To this aim, we will reason on the dual global error system

$$\mathbf{e}_{k+1|k+1} = A_D^\top(\gamma)\mathbf{e}_{k|k}.$$

If

$$(T\otimes I_n)\mathbf{e}_{k|k} =: \widetilde{\mathbf{e}}_{k|k} = \begin{pmatrix}\widetilde{\mathbf{e}}^{(p)}_{k|k}\\ \widetilde{\mathbf{e}}^{(t)}_{k|k}\end{pmatrix},$$

where T is as in Proposition 1, we obtain

$$\begin{pmatrix}\widetilde{\mathbf{e}}^{(p)}_{k+1|k+1}\\ \widetilde{\mathbf{e}}^{(t)}_{k+1|k+1}\end{pmatrix} = \begin{pmatrix}A_C^\top & H_{21}^\top(\gamma)\\ H_{12}^\top & H_{22}^\top(\gamma)\end{pmatrix}\begin{pmatrix}\widetilde{\mathbf{e}}^{(p)}_{k+1|k}\\ \widetilde{\mathbf{e}}^{(t)}_{k+1|k}\end{pmatrix}.$$

It is easy to see that

$$\begin{pmatrix}A_C^\top & H_{21}^\top(\gamma)\\ H_{12}^\top & H_{22}^\top(\gamma)\end{pmatrix} = \begin{pmatrix}I_n & 0\\ 0 & (S\otimes I_n)^\gamma\end{pmatrix}(T\otimes I_n)\bar{A}_D^\top\times$$
$$\times (T^\top\otimes I_n)\begin{pmatrix}I_n & 0\\ 0 & (S\otimes I_n)^\gamma\end{pmatrix}, \tag{30}$$

where

$$\bar{A}_D = (I_N\otimes A) - N\mathrm{diag}_i(P_\infty C_i^\top R_i^{-1} C_i A).$$

On account of (4), pick $\lambda\in(0,1)$ such that

$$A_C P_\infty A_C^\top \leqslant \lambda P_\infty. \tag{31}$$

By using the weighted norms

$$\|N\|_M := \sup_{z\in\mathbb{R}^m}\sqrt{\frac{z^\top N^\top M N z}{z^\top M z}}, \tag{32}$$

$N\in\mathbb{R}^{m\times m}$, $M\in\mathcal{P}_+(m)$, we can write

$$\begin{aligned}&\|(T\otimes I_n)\bar{A}_D^\top\times(T^\top\otimes I_n)\|_{I_N\otimes P_\infty}\\ &=\|(I_N\otimes P_\infty^{1/2})(T\otimes I_n)\bar{A}_D^\top(T^\top\otimes I_n(I_N\otimes P_\infty^{-1/2})\|\\ &=\|(T\otimes I_n)(I_N\otimes P_\infty^{1/2})\bar{A}_D^\top(I_N\otimes P_\infty^{-1/2})(T^\top\otimes I_n\|\\ &=\|\bar{A}_D^\top\|_{I_N\otimes P}.\end{aligned} \tag{33}$$

Moreover, notice that

$$\begin{aligned}\|S\otimes I_n\| &= 1-\frac{\lambda_2(\mathcal{L})}{\delta} =: \theta,\\ \|S\otimes I_n\|_{I_{N-1}\otimes P_\infty} &= \|S\otimes I_n\|,\\ \|(S\otimes I_n)^\gamma\|_{I_{N-1}\otimes P} &\leqslant \|S\otimes I_n\|^\gamma = \theta^\gamma.\end{aligned}$$

By using (30) and (33) we have

$$\begin{aligned}&\|\widetilde{\mathbf{e}}^{(t)}_{k+1|k+1}\|_{P_\infty}\leqslant\|A_C^\top\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{P_\infty}+\|H_{21}^\top(\gamma)\widetilde{\mathbf{e}}^{(t)}_{k|k}\|_{P_\infty}\\ &\leqslant\sqrt{\lambda}\|\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{P_\infty}+\Big\|\begin{pmatrix}I_n&0\\0&0\end{pmatrix}(T\otimes I_n)A_D^\top(\gamma)(T^\top\otimes I_n)\times\\ &\times\begin{pmatrix}0&0\\0&I_{(N-1)n}\end{pmatrix}\begin{pmatrix}0\\(S\otimes I_n)^\gamma\end{pmatrix}\widetilde{\mathbf{e}}^{(t)}_{k|k}\Big\|_{I_N\otimes P_\infty}\\ &\leqslant\sqrt{\lambda}\|\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{P_\infty}+\Big\|\begin{pmatrix}I_n&0\\0&0\end{pmatrix}(T\otimes I_n)A_D^\top(\gamma)(T^\top\otimes I_n)\times\\ &\times\begin{pmatrix}0&0\\0&I_{(N-1)n}\end{pmatrix}\Big\|_{I_N\otimes P_\infty}\Big\|\begin{pmatrix}0\\(S\otimes I_n)^\gamma\end{pmatrix}\widetilde{\mathbf{e}}^{(t)}_{k|k}\Big\|_{I_N\otimes P_\infty}\\ &\leqslant\sqrt{\lambda}\|\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_P+\theta^\gamma\|\bar{A}_D^\top\|_{I_N\otimes P_\infty}\|\widetilde{\mathbf{e}}^{(t)}_{k|k}\|_{I_{N-1}\otimes P_\infty}\end{aligned}$$

Similarly,

$$\|\widetilde{\mathbf{e}}^{(p)}_{k+1|k+1}\|_{I_{N-1}\otimes P_\infty} \leqslant \|H_{12}^\top \widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{I_{N-1}\otimes P_\infty} + \|H_{22}^\top(\gamma)\widetilde{\mathbf{e}}^{(t)}_{k|k}\|_{I_{N-1}\otimes P_\infty} \leqslant \theta^\gamma \|\bar{A}_D^\top\|_{I_N\otimes P_\infty}\|\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{P_\infty} + \theta^{2\gamma}\|\bar{A}_D^\top\|_{I_N\otimes P_\infty}\|\widetilde{\mathbf{e}}^{(t)}_{k|k}\|_{I_{N-1}\otimes P_\infty}.$$

Finally, we conclude

$$\left\| \begin{matrix} \|\widetilde{\mathbf{e}}^{(p)}_{k+1|k+1}\|_{P_\infty} \\ \|\widetilde{\mathbf{e}}^{(t)}_{k+1|k+1}\|_{I_{N-1}\otimes P_\infty} \end{matrix} \right\| \leqslant \rho(A_R(\gamma)) \left\| \begin{matrix} \|\widetilde{\mathbf{e}}^{(p)}_{k|k}\|_{P_\infty} \\ \|\widetilde{\mathbf{e}}^{(t)}_{k|k}\|_{I_{N-1}\otimes P_\infty} \end{matrix} \right\|$$

where ρ denotes the spectral radius and

$$A_R(\gamma) := \begin{pmatrix} \sqrt{\lambda} & \theta^\gamma \|\bar{A}_D^\top\|_{I_N\otimes P_\infty} \\ \theta^\gamma \|\bar{A}_D^\top\|_{I_N\otimes P_\infty} & \theta^{2\gamma}\|\bar{A}_D^\top\|_{I_N\otimes P_\infty} \end{pmatrix}. \quad (34)$$

For the stability of (29) it is sufficient that the spectral radius of $A_R(\gamma)$ is < 1. Thus, we conclude with the following result for a lower bound γ_0 for γ.

Proposition 4 *An integer $\gamma_0 > 0$ such that $A_D(\gamma)$ is Schur stable for all $\gamma \geqslant \gamma_0$ is given by any γ_0 such that the spectral radius of $A_R(\gamma)$ is < 1 for all $\gamma \geqslant \gamma_0$.*

We remark that the computation of γ_0 through the stability of $A_R(\gamma)$ can be performed by each node of the network in a complete distributed way. Indeed, this task requires the knowledge of an upper bound for $\|\bar{A}_D^\top\|_{I_N\otimes P_\infty}$: see (34). Since

$$\|\bar{A}_D^\top\|_{I_N\otimes P_\infty} \leqslant N[\|A^\top\|_{P_\infty} + \sqrt{\lambda_{max}(AP_\infty A^\top)\lambda_{max}(P_\infty)} \sum_{i=1}^N \|C_i^\top R_i^{-1} C_i\|]$$

(λ_{max} denotes the largest eigenvalue) and since a distributed computation of P_∞ and $\sum_{i=1}^N \|C_i^\top R_i^{-1} C_i\|$ is possible (see Section 3.3), then it is also possible the the distributed computation of the lower bound of γ_0.

4 Directed graphs

In this section we outline a generalization of DKF to weighted directed graphs. Consider graphs denoted by $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{A})$, where $\mathcal{A} \in \mathbb{R}^{N\times N}$ denotes the adjacency matrix. The (i,j)-th entry $\mathcal{A}_{i,j}$ is the weight associated with the edge (i,j). We have $\mathcal{A}_{i,j} \neq 0$ if and only if $(i,j) \in \mathcal{E}$. Otherwise $\mathcal{A}_{i,j} = 0$. The graph is said to be *directed* if it has the property that $(i,j) \in \mathcal{E}$ does not imply $(j,i) \in \mathcal{E}$ for all $i,j \in \mathcal{N}$. We will assume that the graph is *simple*, i.e. $\mathcal{A}_{i,i} = 0$ for all $i \in \mathcal{N}$. A directed path from node i_1 to node i_l is a sequence of edges (i_k, i_{k+1}), $k = 1, 2, \ldots, l-1$. A directed graph $\mathcal{G}$ is *strongly connected* if between any pair of distinct nodes i and j in $\mathcal{G}$, there exists a directed path from i to j, $i, j \in \mathcal{N}$.

For weighted directed graphs, the *Laplacian* $\mathcal{L} \in \mathbb{R}^{N\times N}$ is defined as $\mathcal{L} := \mathcal{M} - \mathcal{A}$ where the i-th diagonal entry of the diagonal matrix $\mathcal{M}$ is given by $m_i = \sum_{j=1}^N \mathcal{A}_{i,j}$. By construction $\mathcal{L}$ has a zero eigenvalue with an associated right eigenvector $\mathbf{1}_N$ (i.e. such that $\mathcal{L}\mathbf{1}_N = 0$) and if the graph is strongly connected all the other eigenvalues lie in the open right-half complex plane. A symmetrizable graph with Laplacian $\mathcal{L}$ is one for which there is a diagonal matrix M such that $M\mathcal{L} = \mathcal{L}_D$, where $\mathcal{L}_D$ is the Laplacian of an undirected graph. The Laplacian $\mathcal{L}$ can be decomposed in two components: $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1$ where $\mathcal{L}_0$ is the Laplacian of a symmetrizable graph and $\mathcal{L}_1$ is the Laplacian of a graph for which we have only one-way direct links between nodes. If the graph is symmetrizable the matrix M for which $M\mathcal{L} = \mathcal{L}_D$, the Laplacian of an undirected graph, is computed as $M = \mathrm{diag}\{m_1^{-1}, \ldots, m_N^{-1}\}$ where $m = (m_1, \ldots, m_N)$ is a left eigenvector of $\mathcal{L}$ associated to the eigenvalue $\lambda = 0$.

With this in mind, in order to extend our distributed filter DKF to directed symmetrizable graphs, it is sufficient to modify the equations of the local filter at the i-th sensor node as follows:

$$\hat{\mathbf{x}}^{(i)}_{k+1|k} = A\hat{\mathbf{x}}^{(i)}_{k|k} \quad (35)$$

$$\begin{cases} \mathbf{z}^{(i)}_{k+1,0} = \hat{\mathbf{x}}^{(i)}_{k+1|k} + K_i(\mathbf{y}^{(i)}_{k+1} - C_i\hat{\mathbf{x}}^{(i)}_{k+1|k}), \\ \mathbf{z}^{(i)}_{k+1,h+1} = \mathbf{z}^{(i)}_{k+1,h} - \frac{m_i}{\delta}\sum_{j=1}^N [\mathcal{L}]_{i,j}\mathbf{z}^{(j)}_{k+1,h} \end{cases} \quad (36)$$

$$\hat{\mathbf{x}}^{(i)}_{k+1|k+1} = \mathbf{z}^{(i)}_{k+1,\gamma} \quad (37)$$

and the stability and optimality analysis can be performed exactly as in the case of the DKF for undirected graphs. For not symmetrizable graphs, more complex structures for the local filter at the sensor nodes must be conceived (we defer this discussion elsewhere).

5 A distributed Luenberger observer

In the absence of noise, the convergence rate of the DKF can be arbitrarily increased by arbitrarily assigning the eigenvalues of the filter. In this case we have a distributed Luenberger observer which is of interest in its own. The equations of the observer at the i-th sensor node are:

$$\hat{\mathbf{x}}^{(i)}_{k+1|k} = A\hat{\mathbf{x}}^{(i)}_{k|k} \quad (38)$$

$$\mathbf{z}^{(i)}_{k+1,h+1} = \mathbf{z}^{(i)}_{k+1,h} + \frac{1}{\delta}\sum_{j\in\mathcal{N}^{(i)}} (\mathbf{z}^{(j)}_{k+1,h} - \mathbf{z}^{(i)}_{k+1,h}) \quad (39)$$

$$\mathbf{z}^{(i)}_{k+1,0} = \hat{\mathbf{x}}^{(i)}_{k+1|k} + K_i(\mathbf{y}^{(i)}_{k+1} - C_i\hat{\mathbf{x}}^{(i)}_{k+1|k}) \quad (40)$$

$$\hat{\mathbf{x}}^{(i)}_{k+1|k+1} = \mathbf{z}^{(i)}_{k+1,\gamma}, \quad (41)$$

where $h = 0, \ldots, \gamma - 1$, $K_i := N L_i C_i^\top R_i^{-1}$ and $\delta > \lambda_N(\mathcal{L})$ and $L_i \in \mathbb{R}^{n \times q_i}$ and $\gamma \in \mathbb{N}$ are parameters to be chosen.

Define A_i as in (10) with K_i replaced by L_i, and $A_D(\gamma)$ as in (11). The parameters $L_i \in \mathbb{R}^{n \times q_i}$ and $\gamma \in \mathbb{N}$ are chosen as pointed out in the following proposition: in particular, the eigenvalues of $A_D(\gamma)$ can be selected arbitrarily close to the origin.

Proposition 5 *There exist $L = \mathrm{col}_i(L_i)$, $L_i \in \mathbb{R}^{n \times q_i}$, $i = 1, \ldots, N$, and $\gamma_0 \in \mathbb{N}$ such that $A_D(\gamma)$ is Schur stable for all integer $\gamma > \gamma_0$. Moreover, it holds*

$$\lim_{\gamma \to \infty} \sigma(A_D(\gamma)) = \sigma(A_C) \cup \mathcal{Z},$$

with $\mathcal{Z} = \{0, \ldots, 0\}$ a set of $(n-1)N$ zero.

In other words, Proposition 5 states that n eigenvalues of $A_D(\gamma)$ tend, as $\gamma \to +\infty$, to the n eigenvalues of $A_C := A - LC$ and the remaining $(N-1)n$ eigenvalues tend to zero. The proof is omitted and follows the one of Proposition 1. In the absence of noise, the global error equation is

$$\mathbf{e}_{k+1|k+1} = A_D(\gamma)\mathbf{e}_{k|k}, \qquad (42)$$

and by Proposition 5 it follows that the convergence rate of the distributed Luenberger observer can be arbitrarily selected by a suitable choice of $L = \mathrm{col}_i(L_i)$, $L_i \in \mathbb{R}^{n \times q_i}$, $i = 1, \ldots, N$, and $\gamma \in \mathbb{N}$.

6 Distributed Kalman filters with local P_i

The filter DKF is guaranteed to be stable only for $\gamma \geqslant \gamma_0$. Also, the term N in the gains K_i has a downgrading effect on the performances of DKF for small values of γ. It is possible to partially compensate this downgrading effect while retaining convergence to the optimal estimate for large γ. This modified distributed Kalman filter consists of two consensus steps, one for the local estimate and another for the local error covariance:

$$\begin{aligned}
\hat{\mathbf{x}}^{(i)}_{k+1|k} &= A\hat{\mathbf{x}}^{(i)}_{k|k},\\
\mathbf{P}^{(i)}_{k+1|k} &= A\mathbf{P}^{(i)}_{k|k}A^\top + Q,\\
\mathbf{z}^{(i)}_{k+1,0} &= (\mathbf{P}^{(i)}_{k+1|k})^{-1}\hat{\mathbf{x}}^{(i)}_{k+1|k} + NC_i^\top R_i^{-1}\mathbf{y}^{(i)}_{k+1},\\
\mathbf{Z}^{(i)}_{k+1,0} &= (\mathbf{P}^{(i)}_{k+1|k})^{-1} + NC_i^\top R_i^{-1}C_i,\\
\mathbf{z}^{(i)}_{k+1,h+1} &= \mathbf{z}^{(i)}_{k+1,h} + \frac{1}{\delta}\sum_{j \in \mathcal{N}^{(i)}} (\mathbf{z}^{(j)}_{k+1,h} - \mathbf{z}^{(i)}_{k+1,h}),\\
\mathbf{Z}^{(i)}_{k+1,h+1} &= \mathbf{Z}^{(i)}_{k+1,h} + \frac{1}{\delta}\sum_{j \in \mathcal{N}^{(i)}} (\mathbf{Z}^{(i)}_{k+1,h} - \mathbf{Z}^{(i)}_{k+1,h}),\\
\mathbf{P}^{(i)}_{k+1|k+1} &= (\mathbf{Z}^{(i)}_{k+1,\gamma})^{-1},\\
\hat{\mathbf{x}}^{(i)}_{k+1|k+1} &= (\mathbf{Z}^{(i)}_{k+1,\gamma})^{-1}\mathbf{z}^{(i)}_{k+1,\gamma},
\end{aligned} \qquad (43)$$

where $h = 0, \ldots, \gamma - 1$, with $\mathbf{P}^{(i)}_{0|0} = \Psi_{\mathbf{x}_0}$. Notice that the product $(\mathbf{Z}^{(i)}_{k+1,\gamma})^{-1}\mathbf{z}^{(i)}_{k+1,\gamma}$, through which we get the updated estimate $\hat{\mathbf{x}}^{(i)}_{k+1|k+1}$, is aimed to compensate for the downgrading effect of the term N in the filter gain (previously denoted by K_i) on the performance: indeed, as it can be seen from (43), the parameter N directly influences both $\mathbf{Z}^{(i)}_{k+1,0}$ and $\mathbf{z}^{(i)}_{k+1,0}$ (and therefore $\mathbf{Z}^{(i)}_{k+1,\gamma}$ and $\mathbf{z}^{(i)}_{k+1,\gamma}$).

We consider the steady state version of the latter modified distributed Kalman filter, we call SMDKF:

$$\begin{aligned}
\hat{\mathbf{x}}^{(i)}_{k+1|k} &= A\hat{\mathbf{x}}^{(i)}_{k|k},\\
\mathbf{z}^{(i)}_{k+1,0} &= (\mathbf{P}^{(i)}_{p,\infty})^{-1}\hat{\mathbf{x}}^{(i)}_{k+1|k} + K_i\mathbf{y}^{(i)}_{k+1},\\
\mathbf{z}^{(i)}_{k+1,h+1} &= \mathbf{z}^{(i)}_{k+1,h} + \frac{1}{\delta}\sum_{j \in \mathcal{N}^{(i)}} (\mathbf{z}^{(j)}_{k+1,h} - \mathbf{z}^{(i)}_{k+1,h}),\\
\hat{\mathbf{x}}^{(i)}_{k+1|k+1} &= \mathbf{P}^{(i)}_{\infty}\mathbf{z}^{(i)}_{k+1,\gamma}.
\end{aligned} \qquad (44)$$

with $K_i = NC_i^\top R_i^{-1}$, $\mathbf{P}^{(i)}_{p,\infty} = A\mathbf{P}^{(i)}_{\infty}A^\top + Q$, where $\mathbf{P}^{(i)}_{\infty}$ is meant as the limit of $\mathbf{P}^{(i)}_{k|k}$ as $k \to +\infty$.

6.1 *Stability analysis: properties of the solutions $\mathbf{P}^{(i)}_{k|k}$*

By considering the dynamics of $\mathbf{P}^{(i)}_{k|k}$ of the filter (43), we obtain the global equation

$$\mathrm{col}_i(\mathbf{M}^{(i)}_{k|k}) := J^{-\gamma}\mathrm{col}_i(\mathbf{P}^{(i)}_{k|k}) = \mathrm{col}_i\left((\mathbf{P}^{(i)}_{k+1|k})^{-1} + NG_i\right),$$

where $G_i := C_i^\top R_i^{-1}C_i$. With $\mathbf{\Sigma}_k := \mathrm{col}_i(\mathbf{\Sigma}^{(i)}_k)$, where $\mathbf{\Sigma}^{(i)}_k := (\mathbf{P}^{(i)}_{k|k})^{-1}$, the above equation can be rewritten as

$$\mathbf{\Sigma}_{k+1} := J^\gamma \mathrm{col}_i((A(\mathbf{\Sigma}^{(i)}_k)^{-1}A^\top + Q)^{-1} + NG_i). \qquad (45)$$

The solutions $\mathbf{\Sigma}^{(i)}_k$, $i = 1, \ldots, N$, of (45) are continuous functions of the parameter γ but we will omit this dependence in what follows for simplicity. Moreover, the solutions $\mathbf{\Sigma}^{(i)}_k$, $i = 1, \ldots, N$, of (45) are symmetric and positive definite for all $k \geqslant 0$ (this fact easily follows by induction on k). Taking into account that $J^\gamma = (I_N - \frac{\mathcal{L}}{\delta})^\gamma \otimes I_n$ is symmetric and have all either positive (at least one) or zero entries by construction (the choice $\delta > \lambda_N(\mathcal{L})$).

An important result is the monotonicity of the solutions $\mathbf{\Sigma}^{(i)}_k$, $i = 1, \ldots, N$, of (45).

Proposition 6 *If there exists $k^* \geqslant 0$ such that*

$$\mathbf{\Sigma}^{(i)}_{k^*} \lesseqgtr \mathbf{\Sigma}^{(i)}_{k^*+1} \quad \text{for all } i = 1, \ldots, N,$$

then

$$\mathbf{\Sigma}_k^{(i)} \lesseqgtr \mathbf{\Sigma}_{k+1}^{(i)}, \forall k \geqslant k^*, \quad \text{for all } i = 1, \ldots, N.$$

Proof. We prove only the "$\leqslant$" part, since the "$\geqslant$" part follows exactly in the same way. Define

$$S(\mathbf{\Sigma}_k) := J^\gamma \mathrm{col}_i((S^{(i)}(\mathbf{\Sigma}_k^{(i)}))^{-1} + NG_i),$$
$$S^{(i)}(\mathbf{\Sigma}_k^{(i)}) := A\mathbf{\Sigma}_k^{(i)} A^\top + Q.$$

Notice that $\mathbf{\Sigma}_k$ satisfies for all $k \geqslant 0$

$$\mathbf{\Sigma}_{k+1} = S(\mathbf{\Sigma}_k). \tag{46}$$

Since

$$S(\mathbf{\Sigma}_k + \Delta) - S(\mathbf{\Sigma}_k) = \int_0^1 \frac{\partial S}{\partial \lambda}(\mathbf{\Sigma}_k + \lambda\Delta)\mathrm{d}\lambda,$$

where $\Delta := \mathrm{col}_i(\Delta^{(i)})$, $\Delta^{(i)}$ any positive semidefinite matrix, and

$$\begin{aligned}&\frac{\partial S}{\partial \lambda}(\mathbf{\Sigma}_k + \lambda\Delta)\\ &\quad = J^\gamma \mathrm{col}_i\Big((S^{(i)}(\mathbf{\Sigma}_k^{(i)}))^{-1} A(\mathbf{\Sigma}_k^{(i)} + \Delta^{(i)})^{-1}\Delta^{(i)} \times\\ &\qquad \times (\mathbf{\Sigma}_k^{(i)} + \Delta^{(i)})^{-1} A^\top (S^{(i)}(\mathbf{\Sigma}_k^{(i)}))^{-1}\Big),\end{aligned}$$

each matrix

$$\begin{aligned}&(S^{(i)}(\mathbf{\Sigma}_k^{(i)}))^{-1} A(\mathbf{\Sigma}_k^{(i)} + \Delta^{(i)})^{-1}\Delta^{(i)} \times\\ &\times (\mathbf{\Sigma}_k^{(i)} + \Delta^{(i)})^{-1} A^\top (S^{(i)}(\mathbf{\Sigma}_k^{(i)}))^{-1}\end{aligned}$$

is positive semidefinite. Moreover, the rows of J^γ have all either positive (at least one) or zero entries by construction and each matrix is positive semidefinite. It follows that $\frac{\partial S}{\partial \lambda}(\mathbf{\Sigma}_k + \lambda\Delta)$ is the a linear combination with nonnegative coefficients of positive semidefinite matrices, which proves that it is positive semidefinite and $S(\mathbf{\Sigma}_k + \Delta) \geqslant S(\mathbf{\Sigma}_k)$. By (46) this implies that $\mathbf{\Sigma}_{k^*+2} := S(\mathbf{\Sigma}_{k^*+1}) \geqslant \mathbf{\Sigma}_{k^*+1} := S(\mathbf{\Sigma}_{k^*})$ if $\mathbf{\Sigma}_{k^*+1} \geqslant \mathbf{\Sigma}_{k^*}$ for some $k^* \geqslant 0$. By repeating this argument, we prove the claim of the proposition. □

Remark 2 *Since $X \geqslant Y$ for any symmetric and invertible matrices X, Y implies that $X^{-1} \leqslant Y^{-1}$, Proposition 6 also implies the same monotonicity property on the matrices $\mathbf{P}_{k|k}^{(i)} := (\mathbf{\Sigma}_k^{(i)})^{-1}$: if there exists $k^* \geqslant 0$ such that*

$$\mathbf{P}_{k^*|k^*}^{(i)} \lesseqgtr \mathbf{P}_{k^*+1|k^*+1}^{(i)} \quad \text{for all } i = 1, \ldots, N,$$

then

$$\mathbf{P}_{k|k}^{(i)} \lesseqgtr \mathbf{P}_{k+1|k+1}^{(i)}, \forall k \geqslant k^*, \quad \text{for all } i = 1, \ldots, N.$$

A consequence is that, for instance, an initial condition $\mathbf{P}_{0|0}^{(i)} = 0$, $i = 1, \ldots, N$, triggers a nondecreasing sequence $\{\mathbf{P}_{k|k}^{(i)}\}$, $k \geqslant 0$.

The second important result of this section is the boundedness of the sequence $(\mathbf{P}_{k|k}^{(i)})^{-1}$.

Proposition 7 *Let $\mathbf{P}_{k|k}^{(i)}$, $i = 1, \ldots, N$, be a positive definite solution of (48). The sequence $\{\mathbf{\Sigma}_k^{(i)}\} = \{(\mathbf{P}_{k|k}^{(i)})^{-1}\}$, $i = 1, \ldots, N$, is bounded for all $k \geqslant 1$.*

Proof. Rewrite (45) with $\mathbf{S}_k := \mathrm{col}_i(\mathbf{S}_k^{(i)}) := \mathrm{col}_i(\frac{1}{N}\mathbf{\Sigma}_k^{(i)})$:

$$\mathbf{S}_{k+1} = J^\gamma \mathrm{col}_i\left(\frac{1}{N}\left(A\left(\frac{1}{N}\mathbf{S}_k^{(i)}\right)^{-1} A^\top + Q\right)^{-1} + G_i\right).$$

Since the entries of J^γ are all nonnegative and less or equal than 1 and G_i and $(A(\frac{1}{N}\mathbf{S}_k^{(i)})^{-1}A^\top + Q)^{-1}$, $i = 1, \ldots, N$, are all positive semidefinite matrices, we get

$$\mathbf{S}_{k+1} \leqslant \mathrm{col}_i\left(\frac{1}{N}\sum_{i=1}^N \left(A\left(\frac{1}{N}\mathbf{S}_k^{(i)}\right)^{-1} A^\top + Q\right)^{-1} + G\right).$$

By backward substitutions, since $N \geqslant 1$ and $\mathbf{S}_0 = \frac{1}{N}\Psi_{x_0}^{-1}$ and since $X \geqslant Y$ for any symmetric and invertible matrices X, Y implies that $X^{-1} \leqslant Y^{-1}$, we obtain

$$\begin{aligned}&\mathbf{S}_{k+1} \leqslant \overline{\mathbf{S}}_{k+1},\ k \geqslant 1,\\ &\overline{\mathbf{S}}_{k+1} = (A\overline{\mathbf{S}}_k A^\top + Q)^{-1} + G,\ k \geqslant 1,\end{aligned} \tag{47}$$

where the sequence $\{\overline{\mathbf{S}}_k\}$ satisfies the same equation as that of $(\mathbf{P}_{k|k})^{-1}$ ($\mathbf{P}_{k|k}$ being the error covariance of the CKF) which is clearly bounded for all $k \geqslant 0$. As a consequence, from the inequality of (47) we get the boundedness of $\{\mathbf{S}_k\} = \left\{\frac{1}{N}(\mathbf{P}_k^{(i)})^{-1}\right\}$ or what is the same of $\left\{(\mathbf{P}_k^{(i)})^{-1}\right\}$. □

Combining the monotonicity of the sequence $\left\{(\mathbf{P}_{k|k}^{(i)})^{-1}\right\}$, $i = 1, \ldots, N$, with its boundedness from above and below ($k = 0$), we can conclude that

$$\lim_{k \to +\infty} \left(\mathbf{P}_{k|k}^{(i)}\right)^{-1} = \left(\mathbf{P}_\infty^{(i)}\right)^{-1},\ i = 1, \ldots, N,$$

for some solution $\left\{(\mathbf{P}_\infty^{(i)})^{-1}\right\}$, $i = 1, \ldots, N$, (depending on the initial condition $\left\{(\mathbf{P}_{0|0}^{(i)})^{-1}\right\}$, $i = 1, \ldots, N$) of the equation

$$\mathrm{col}_i\left(\left(\mathbf{P}_\infty^{(i)}\right)^{-1}\right) = J^\gamma \mathrm{col}_i\left((A\mathbf{P}_\infty^{(i)}A^\top + Q)^{-1} + NG_i\right). \tag{48}$$

or, equivalently, that the solutions $\mathbf{\Sigma}_k^{(i)}$, $i = 1,\dots,N$, of (45) have a well-defined steady state (positive definite) value.

6.2 *Stability analysis: properties of the stationary solutions* $\mathbf{P}_\infty^{(i)}$

In this section we want to study the stabilizing properties of the asymptotic solutions $\mathbf{\Sigma}_k^{(i)}$, $i = 1,\dots,N$, of (45). To this aim, let us consider the steady state equation (45)

$$\mathbf{\Sigma}_\infty := J^\gamma \mathrm{col}_i((A(\mathbf{\Sigma}_\infty^{(i)})^{-1}A^\top + Q)^{-1} + NG_i) \tag{49}$$

or equivalently (48), since $\mathrm{col}_i(\mathbf{P}_\infty^{(i)})^{-1}) = \mathrm{col}_i(\mathbf{\Sigma}_\infty^{(i)})$. The solutions $\mathbf{P}_\infty^{(i)}$, $i = 1,\dots,N$, to (49) are continuous functions of the parameter γ but we will omit this dependence in what follows for simplicity. The positive definite solutions of (48) tend as $\gamma \to +\infty$ to the steady state error covariance P_∞ of the centralized CKF.

Proposition 8 *Let* $\mathbf{P}_\infty^{(i)}$, $i = 1,\dots,N$, *be a positive definite solution of (48). We have*

$$\lim_{\gamma\to+\infty} \mathbf{P}_\infty^{(i)} = \mathbf{P}_\infty, \ \forall i = 1,\dots,N.$$

Proof. From the error covariance of the centralized CKF and using $J^\gamma 1_N = 1_N$ for all $\gamma \geqslant 1$,

$$\mathrm{col}_i(\mathbf{X}_\infty) = J^\gamma \mathrm{col}_i((A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} + G), \tag{50}$$

with $\mathbf{X}_\infty := P_\infty^{-1}$. On the other hand, (48) gives

$$\mathrm{col}_i(\mathbf{X}_\infty^{(i)}) = J^\gamma \mathrm{col}_i((A(\mathbf{X}_\infty^{(i)})^{-1}A^\top + Q)^{-1} + NG_i), \tag{51}$$

with $\mathbf{X}_\infty^{(i)} := (\mathbf{P}_\infty^{(i)})^{-1}$. Subtracting (51) from (50) with $\Delta_\infty^{(i)} := \mathbf{X}_\infty - \mathbf{X}_\infty^{(i)}$, we get

$$\begin{aligned}\mathrm{col}_i(\Delta_\infty^{(i)}) &:= J^\gamma \mathrm{col}_i((A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} \\ &\quad - (A(\mathbf{X}_\infty - \Delta_\infty^{(i)})^{-1}A^\top + Q)^{-1} + G - NG_i).\end{aligned}$$

On the other hand, define as usual $T := \begin{pmatrix} V \\ W \end{pmatrix}$ then

$$(T\otimes I_n)J^\gamma \mathrm{col}_i(G - NG_i) = \begin{pmatrix} 0 \\ M(\gamma) \end{pmatrix},$$

where

$$M(\gamma) := -N((S^\gamma W^T)\otimes I_n)\mathrm{col}_i(G_i) \to 0 \text{ as } \gamma \to +\infty$$

and

$$\begin{aligned}&(T\otimes I_n)J^\gamma \mathrm{col}_i((A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} \\ &\quad - (A(\mathbf{X}_\infty - \Delta_\infty^{(i)})^{-1}A^\top + Q)^{-1}) = \begin{pmatrix} N_1(\gamma) \\ N_2(\gamma) \end{pmatrix},\end{aligned}$$

where

$$\begin{aligned}N_1(\gamma) &= \frac{1}{\sqrt{N}}\Big[N(A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} \\ &\quad - \sum_{j=1}^{N}(A(\mathbf{X}_\infty - \Delta_\infty^{(i)})^{-1}A^\top + Q)^{-1})\Big] \\ N_2(\gamma) &= -((S^\gamma W^T)\otimes I_n)\times \\ &\quad \times \mathrm{col}_i(A(\mathbf{X}_\infty - \Delta_\infty^{(i)})^{-1}A^\top + Q)^{-1}).\end{aligned}$$

By setting $\widetilde{\Delta}_\infty := \mathrm{col}_i(\widetilde{\Delta}_\infty^{(i}) := (T\otimes I_n)\Delta_\infty$, we get

$$\widetilde{\Delta}_\infty = \begin{pmatrix} N_1(\gamma) \\ M(\gamma) + N_2(\gamma) \end{pmatrix}.$$

Notice that, since $\Delta_\infty^{(i)}$ (and therefore $\widetilde{\Delta}_\infty^{(i)}$) is a bounded function of γ for each $i = 1,\dots,N$, also $N_2(\gamma) \to 0$ as $\gamma \to +\infty$. Therefore, as $\gamma \to +\infty$, the solution $\widetilde{\Delta}_\infty^{(i)}$, $i = 1,\dots,N$, tends to the unique solution $S_\infty^{(i)}$, $i = 1,\dots,N$, of

$$\begin{aligned}S_\infty^{(1)} &= \sqrt{N}\Big[(A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} \\ &\quad - (A(\mathbf{X}_\infty - \frac{1}{\sqrt{N}}S_\infty^{(1)})^{-1}A^\top + Q)^{-1})\Big] \\ S_\infty^{(j)} &= 0, \ j = 2,\dots,N.\end{aligned} \tag{52}$$

Indeed, the solution $S_\infty^{(i)}$, $i = 1,\dots,N$, of (52) is unique and equal to 0: indeed, the first equation of (52) can be rewritten as

$$\overline{S} = (A\overline{S}^{-1}A^\top + Q)^{-1} + G$$

with $\overline{S} := \mathbf{X}_\infty - \frac{1}{\sqrt{N}}S_\infty^{(1)}$ and since $-(A\mathbf{X}_\infty^{-1}A^\top + Q)^{-1} + \mathbf{X}_\infty = G$. By our standing assumptions on A, C and Q

the above equation has a unique positive definite solution [1]. This ends the proof of the proposition. □

The goal of the next proposition is to prove, under some mild assumptions, that the solutions $\mathbf{P}_\infty^{(i)}$, $i = 1, \ldots, N$, of (48) are unique, positive definite and stabilizing. Let

$$\mathrm{col}_i\left(\mathbf{M}_\infty^{(i)}\right) := J^{-\gamma}\mathrm{col}_i\left((\mathbf{P}_\infty^{(i)})^{-1}\right).$$

Proposition 9 *Let $\mathbf{P}_\infty^{(i)}$, $i = 1, \ldots, N$, be a positive definite solution of (48). If $\gamma \geqslant 1$ is such that*

$$J^\gamma \mathrm{diag}_i\left(\mathbf{M}_\infty^{(i)}\right) J^\gamma \leqslant \mathrm{diag}_i\left((\mathbf{P}_\infty^{(i)})^{-1}\right). \tag{53}$$

the matrix

$$A_D := \mathrm{diag}_i\left(\mathbf{P}_\infty^{(i)}\right) J^\gamma \mathrm{diag}_i(\mathbf{M}_\infty^{(i)} - NG_i)(I_N \otimes A) \tag{54}$$

is Shur stable.

Proof. Let $\mathbf{P}_\infty^{(i)}$, $i = 1, \ldots, N$, be positive semidefinite solutions of (48) and assume that $Q = FF^\top$ with full row rank F. From (48) after some manipulations we get for each $i = 1, \ldots, N$

$$\mathbf{M}_\infty^{(i)} - NG_i = (\mathbf{M}_\infty^{(i)} - NG_i)AP_\infty A^\top(\mathbf{M}_\infty^{(i)} - NG_i)^\top + (\mathbf{M}_\infty^{(i)} - NG_i)Q(\mathbf{M}_\infty^{(i)} - NG_i)^\top.$$

By collecting the above equations altogether and multiplying on the left by $\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma$ and on the right by $J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})$, we have

$$A_D\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})A_D^\top + H(I_N \otimes Q)H^\top = \mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{M}_\infty^{(i)} - NG_i)J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)}),$$

with

$$H := \mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{M}_\infty^{(i)} - NG_i),$$

and finally

$$A_D\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})A_D^\top + H(I_N \otimes Q)H^\top + N\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma\mathrm{diag}_i(G_i)J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)}) = \mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{M}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)}).$$

Assume that A_D is not Shur stable. There exist $\lambda \notin \mathbb{S}^1$ ($\mathbb{S}^1$ the open unit circle in the complex plane) and $x \in \mathbb{R}^{nN}\backslash\{0\}$ such that

$$A_D x = \lambda x.$$

Set $V(x) = x^*\Pi x$ with

$$\Pi := \mathrm{diag}_i(\mathbf{P}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{M}_\infty^{(i)})J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)}).$$

Clearly, Π is symmetric and positive definite, being a product of nonsingular matrices. We have

$$V(x) = |\lambda|^2 x^*\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})x + \|(I_N \otimes F^\top)H^\top x\|^2 + N\|\mathrm{diag}_i((R^{(i)})^{-1/2}C_i)J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})x\|^2),$$

so that, using (53),

$$\|(I_N \otimes F^\top)H^\top x\|^2 + N\|\mathrm{diag}_i((R^{(i)})^{-1/2}C_i)J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})x\|^2) \leqslant (1 - |\lambda|^2)\mathrm{diag}_i(\mathbf{P}_\infty^{(i)}) \leqslant 0.$$

It follows that

$$(I_N \otimes F^\top)H^\top x = 0,$$
$$\mathrm{diag}_i(C_i)J^\gamma\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})x = 0.$$

The above equations in particular imply that

$$\left(\lambda I_{nN} - (I_N \otimes A^\top)H^\top\right)x = 0.$$

By the Hautus criterion, the equation above contradicts the fact that the pair

$$\left((I_N \otimes F^\top)H^\top, (I_N \otimes A^\top)H^\top\right)$$

is observable (or equivalently $\left((I_N \otimes F^\top), (I_N \otimes A^\top)\right)$ which follows from F^T being full column rank and H being invertible). This ends the proof of the Proposition. □

We conclude this section with a comment on the condition (53): by multiplying on the left by $T \otimes I_n$ and on the right by $T^\top \otimes I_n$ after some manipulations and using the properties of J and $T := \left(V^\top W^\top\right)^\top$ (see (13)–(16)), we obtain that inequality (53) is equivalent to

$$(S^\gamma W \otimes I_n)\mathrm{diag}_i((A\mathbf{P}_\infty^{(i)}A^\top + Q)^{-1} + NG_i)\times \times (W^\top S^\gamma \otimes I_n) \leqslant (W \otimes I_n)\mathrm{diag}_i(\mathbf{P}_\infty^{(i)})^{-1})(W^\top \otimes I_n).$$

Clearly, since $S^\gamma \to 0$ as $\gamma \to +\infty$ and $(\mathbf{P}_\infty^{(i)})^{-1})$ are bounded functions of γ with $(\mathbf{P}_\infty^{(i)})^{-1} \to \mathbf{P}_\infty^{-1} > 0$ as $\gamma \to +\infty$, the last inequality (and therefore (53)) will be satisfied for some sufficiently large γ.

[1] By the matrix inversion lemma we can rewrite the underlying algebraic Riccati equation in the standard form (**?**).

6.3 Stability properties of the SMDKF filter

In order to study the asymptotic properties of the SMDKF filter (44), we introduce at each node the local estimation error $\mathbf{e}^{(i)}_{k|k} := \mathbf{x}_k - \hat{\mathbf{x}}^{(i)}_{k|k}$, the total estimation error $\mathbf{e}_{k|k} := \mathrm{col}_i(\mathbf{e}^{(i)}_{k|k})$, the total measurement noise error vector

$$\begin{aligned}\mathbf{h}_k := &-\mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{diag}_i(K_i)\mathbf{g}_{k+1}\\ &+ \big(I - \mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{col}_i(K_iC_i)\big)(1_N \otimes \mathbf{f}_k)\end{aligned}$$

and the estimation error covariance matrix $\mathbf{X}_{k|k} := \mathbb{E}\{\mathbf{e}_{k|k}\mathbf{e}^\top_{k|k}\}$. Similarly to the computations for (19), we obtain

$$\mathbf{e}_{k+1|k+1} = A_D(\gamma)\mathbf{e}_{k|k} + \mathbf{h}_k$$

with

$$A_D(\gamma) := \mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{diag}_i(\mathbf{M}^{(i)}_\infty - NG_i)(I_N \otimes A)$$

(the matrix $A_D(\gamma)$ was introduced in (54)). Using uncorrelation among $\mathbf{e}_{k|k}$, $\mathbf{g}_{k+1}$ and $\mathbf{f}_k$,

$$\begin{aligned}\Psi := \mathbb{E}\{\mathbf{h}_k\mathbf{h}_k^\top\} := &\,\mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{diag}_i(K_i)R\times\\ &\times \mathrm{diag}_i(K_i)J^\gamma \mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)\\ &+ \Big(I - \mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{col}_i(K_iC_i)\Big)(U_N \otimes Q)\times\\ &\times \Big(I - \mathrm{diag}_i(\mathbf{P}^{(i)}_\infty)J^\gamma \mathrm{col}_i(K_iC_i)\Big)^T\end{aligned}$$

where U_N is a matrix with all entries 1 and

$$\mathbf{X}_{k+1|k+1} = A_D(\gamma)\mathbf{X}_{k|k}A_D^\top(\gamma) + \Psi.$$

By introducing the covariance mismatch $\mathbf{E}_k(\gamma) := \mathbf{X}_{k|k}(\gamma) - \mathbf{X}^C_\infty$ and recalling (23) we obtain after some manipulations

$$\mathbf{E}_k(\gamma) = A_D(\gamma)\mathbf{E}_k(\gamma)A_D^\top(\gamma) + \Phi(\gamma),$$

where

$$\begin{aligned}\Phi(\gamma) := &\,\Psi - \mathrm{diag}_i(A_C)\mathbf{X}^C_\infty \mathrm{diag}_i(A_C^\top)\\ &- \mathrm{diag}_i(I - K_\infty C)(U_N \otimes Q)\mathrm{diag}_i((I - K_\infty C)^\top)\\ &- \mathrm{diag}_i(K_\infty)(U_N \otimes R)\mathrm{diag}_i(K_\infty^\top).\end{aligned}$$

By Proposition 7 the matrix $A_D(\gamma)$ is Schur stable. The main result of this section can be stated as follows and proved as the corresponding result of Proposition 3, taking into account Proposition 7.

Proposition 10 *The error covariance of each filter* (44) *tends as $\gamma \to +\infty$ and $k \to +\infty$ to the steady state error covariance of the centralized CKF, namely*

$$\lim_{\gamma\to+\infty}\lim_{k\to+\infty}\mathbf{X}_{k|k} = \mathbf{X}^C_\infty := U_N \otimes P_\infty.$$

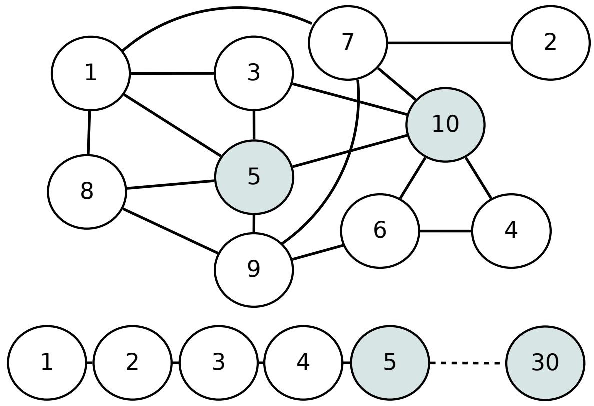

Fig. 2. Graphs used in the first (top) and in the second (bottom) simulation scenario. Sensor nodes are shaded.

7 Simulation results

The aim of this section is to validate the theoretical conclusions of the paper and to show that the existing proposals share basically the same drawbacks. We consider the problem of tracking a planar system with a discretization interval τ, by means of sensors each of which may estimate only one component of the position, that is

$$A = \begin{pmatrix} I_2 & \tau I_2\\ 0 & I_2\end{pmatrix}, \quad C_1 = \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix} \quad C_2 = \begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}$$

$$Q = \sigma^2 \begin{pmatrix} \frac{1}{3}\tau^3 I_2 & \frac{1}{2}\tau^2 I_2\\ \frac{1}{2}\tau^2 I_2 & I_2\end{pmatrix}, \quad R_i = \sigma_g^2.$$

We consider two scenarios:

(1) The first one (Fig. 2, top) consists of a connected graph with $N = 10$ nodes and 17 randomly chosen arcs. Nodes 5 and 10 have measurement matrices C_1 and C_2 respectively, the remaining nodes have communication capabilities only.
(2) In the second scenario (Fig. 2, bottom) there are $N = 30$ nodes connected in a chain. Node 5, 10, 15, 20 and 30 are sensors with measurement matrix C_1, whereas node 25 has measurement matrix C_2. The remaining 24 nodes have communication capabilities only.

With $\tau = 0.25$, $\sigma = 0.05$, $\sigma_g = 0.1$ we performed $N_s = 100$ simulations of $\bar{k} = 250$ points for several values of γ. We compare the performance of the following algorithms: the centralized KF, the CIKF of **?** and **?** in the

		marginally stable, $N = 10$			marginally stable, $N = 30$			unstable, $N = 30$		
	γ	1	2	20	1	10	10^3	1	10	10^3
CKF	tr(P)	$2.76 \cdot 10^{-2}$			$2.14 \cdot 10^{-2}$			$2.92 \cdot 10^{-2}$		
	mse	$2.80 \cdot 10^{-2}$			$2.17 \cdot 10^{-2}$			$2.92 \cdot 10^{-2}$		
DKF	mse	$7.6 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	∞	∞	$2.2 \cdot 10^{-2}$	∞	∞	$2.9 \cdot 10^{-2}$
	st.dev.	$1.8 \cdot 10^{-2}$	$8.5 \cdot 10^{-2}$	$4.1 \cdot 10^{-6}$	∞	∞	$3.0 \cdot 10^{-7}$	∞	∞	$7.5 \cdot 10^{-7}$
SMDKF	mse	$3.6 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	2.2	$7.0 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	246	$1.7 \cdot 10^{-1}$	$2.9 \cdot 10^{-2}$
	st.dev.	$5.3 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$1.4 \cdot 10^{-6}$	3.7	$4.6 \cdot 10^{-2}$	$2.0 \cdot 10^{-7}$	636	$1.8 \cdot 10^{-1}$	$6.0 \cdot 10^{-7}$
CIKF	mse	$4.5 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$3.8 \cdot 10^{-2}$	9.2	$1.7 \cdot 10^{-1}$	$4.7 \cdot 10^{-2}$	$6.5 \cdot 10^{4}$	$6.1 \cdot 10^{-1}$	$5.9 \cdot 10^{-2}$
	st.dev.	$1.0 \cdot 10^{-2}$	$3.6 \cdot 10^{-3}$	$2.0 \cdot 10^{-6}$	15	$1.6 \cdot 10^{-1}$	$1.6 \cdot 10^{-4}$	$1.9 \cdot 10^{5}$	$9.0 \cdot 10^{-1}$	$2.4 \cdot 10^{-4}$
ICF	mse	$3.6 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	2.2	$7.0 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	708	$1.7 \cdot 10^{-1}$	$2.9 \cdot 10^{-2}$
	st.dev.	$5.2 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$	$1.4 \cdot 10^{-6}$	3.6	$4.6 \cdot 10^{-2}$	$2.0 \cdot 10^{-7}$	$1.9 \cdot 10^{3}$	$1.8 \cdot 10^{-1}$	$6.0 \cdot 10^{-7}$
HCMCI-1	mse	$3.7 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$	$2.9 \cdot 10^{-2}$	7.7	$9.7 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$3.8 \cdot 10^{4}$	$3.0 \cdot 10^{-1}$	$3.7 \cdot 10^{-2}$
	st.dev.	$2.6 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$5.3 \cdot 10^{-7}$	13	$9.2 \cdot 10^{-2}$	$6.8 \cdot 10^{-5}$	$1.1 \cdot 10^{5}$	$4.3 \cdot 10^{-1}$	$9.8 \cdot 10^{-5}$
HCMCI-2	mse	$3.5 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	5.6	$6.5 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$1.4 \cdot 10^{4}$	$1.6 \cdot 10^{-1}$	$2.9 \cdot 10^{-2}$
	st.dev.	$6.6 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$1.8 \cdot 10^{-6}$	9.7	$5.1 \cdot 10^{-2}$	$2.4 \cdot 10^{-6}$	$4.1 \cdot 10^{4}$	$2.0 \cdot 10^{-1}$	$2.9 \cdot 10^{-6}$

Table 1
Mean square error and its standard deviation across nodes (consensus) as a function of γ, the number of consensus steps.

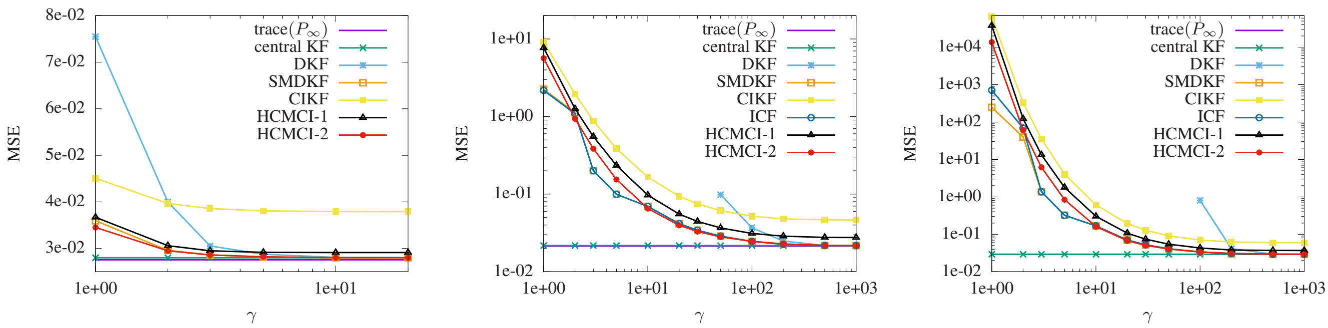

Fig. 3. Comparison of MSE at varying γ for a same marginally stable system, under a connected graph topology with $N = 10$ (left, linear scale on the vertical axis) and a loosely connected graph topology $N = 30$ (center, logarithmic scale on the vertical axis) and an unstable system with $N = 30$ (right, logarithmic scale on the vertical axis).

version of **?**, the ICF of **?**, the HCMCI-1 and HCMCI-2 of **?**, the proposed DKF of Section 3 and finally the proposed SMDKF of Section 6. The MSE is computed as the norm of the estimation error averaged over all the nodes and times $k \in [\frac{1}{5}\bar{k}, \ldots, \bar{k}]$ (to avoid transient effects). The results for the two scenarios are reported in Table 1 and Fig. 3. In the first scenario (Fig. 3, left) the MSE of all the filters is bounded even for $\gamma = 1$. This is in accordance with the fact that the lower bound γ_0 for the number of consensus steps that ensure stability of the mean square error, computed as in Section 3.4, is $\gamma_0 = 1$. When γ increases all the filters but the CIKF, and at a minor extent the HCMCI-1, converge to the MSE of the centralized KF, and a value $\gamma = 4$ is sufficient to obtain a good approximation. The MSE of the ICF is not reported since it is identical to the value for the SMDKF. In the second scenario (Fig. 3, center) the variance of the estimation error is much larger at low values of γ, due to looser graph connectivity. In this case, $\gamma_0 = 31$ and the stability of DKF for $\gamma > \gamma_0$ is confirmed by the plot of Fig. 3 (center). Again, ICF and SMDKF have nearly identical values.

In the last three columns of Table 1 and Fig. 3 (right) we plot the variance of the error for the unstable system obtained by replacing A with $A + \frac{1}{10}I_4$ on a time horizon of $\bar{k} = 150$ points. The loosely connected chain with $N = 30$ is used. All the filters but CIKF, and to minor extent the HCMICI-1, converge to the MSE of the centralized KF for large γ, but this time a value of at least $\gamma = 20$ is needed to attain acceptable performance. In this case, the results in Section 3.4 yield $\gamma_0 = 71$ and the stability of DKF for $\gamma > \gamma_0$ is confirmed by the plot. Since ICF and SMDKF have again the same performance we can conclude that SMDKF (a consensus on estimates filter) is equivalent ICF (a consensus on measurement filter). We notice that at $\gamma \leqslant 20$ the performance of all the filters is not acceptable. Even if the error variance of DKF, SMDKF and ICF tends to infinity as time grows, whereas that of CIKF, HCMCI-1/2 asymptotically reaches a finite value, in practice none of the filters works at small γ, thus in practice the difference may not be significant.

8 Conclusions

The results presented in this paper show that stability and performance of consensus-based distributed filters critically depends on the number of consensus steps as well as on the density of sensor nodes. It is therefore important to be able to compute the asymptotic variance of the estimation error as a function of the system structure, the network topology and the number of consensus steps. The results presented in this work may then be important for the analysis and design of distributed filters in more challenging situations such as system uncertainties, non-linear behavior, unreliable communications, etc.