Consensus optimization approach for distributed Kalman filtering: performance recovery of centralized filtering with proofs

Kunhee Ryu a, Juhoon Back a

*aSchool of Robotics, Kwangwoon University, Seoul 01890, Korea

Abstract

This paper investigates the distributed Kalman filtering (DKF) from distributed optimization viewpoint. Motivated by the fact that Kalman filtering is a maximum a posteriori estimation (MAP) problem, which is a quadratic optimization problem, we reformulate DKF problem as a consensus optimization problem, resulting in that it can be solved by many existing distributed optimization algorithms. A new DKF algorithm employing the dual ascent method is proposed, and its stability is proved under mild assumptions. The performance of the proposed algorithm is evaluated through numerical experiments.

Key words: Distributed Kalman filtering, distributed optimization, dual ascent method

1 Introduction

In order to monitor large scale systems or environments such as traffic networks, plants, sea, etc., distributed filtering using multiple estimators is preferred because it has advantages in terms of scalability, robustness to component loss, and computational cost. Although promising, developing fully distributed solutions with guaranteed stability and optimality is still challenging due to practical issues including heterogeneity of sensors, restriction on communication, uncertainty of network topology, etc. Against this backdrop, literature on distributed Kalman filtering (DKF) is expanding rapidly [2, 9, 12, 17, 25, 28, 32, 40, 42]; see also the survey [29] and references therein.

Recently, consensus based DKF algorithms have received particular attention since the seminal works [32–34] have been published. In [32], the author addressed the connection between the consensus problems and DKF problems. Two DKF algorithms, called Kalman-Consensus filters have been presented in [33]. In the first algorithm, each filter calculates the average of the measurements across all filters in a distributed way and then updates local estimate using it, while in the other, local estimates are obtained by applying standard Kalman filtering computation and then the filters draw consensus on these estimates. See the works [24, 33, 34] for the analysis on stability and optimality.

Notably, in [7], three average consensus based algorithms (called consensus on information, consensus on measurement, and their hybrid type) have been presented for collectively observable sensor networks. These algorithms perform subiterations to compensate for insufficient information from the unobservable subspace of individual sensor and to accelerate consensus. The consensus on measurement algorithm, for example, finds the weighted averages on measurements and information rate matrices (see Section 2 for the definition) through the consensus step (subiteration), and by using them, the estimates and covariances are corrected. Motivated by these works, various DKF algorithms have been developed; algorithms with consistency [5, 21] and asymptotic optimality [3, 4], and those for collectively detectable sensor networks [16, 27, 39]. There are also alternative approaches which are not directly connected to average consensus; e.g., diffusive DKF [13] and the dynamic consensus on pseudo-observation DKF [15]. See also the works on distributed Kalman-Bucy filtering [25, 36], partition-based DKF [19], DKF with state constraints [20], DKF for uncertain systems [43], etc.

To the best of the authors’ knowledge, most of the researches on DKF try to find a good way to fuse the outcomes of local Kalman filters, and this is done by modifying the filter structure or adding extra consensus pro-
In this paper, we reconsider the DKF problem from the distributed optimization perspective, motivated by the fact that Kalman filtering is basically an optimization problem [8, 14, 31, 38]. Under the assumption that the measurement noises of sensors are mutually uncorrelated, we observed that the cost function of centralized Kalman filtering (CKF) can be decomposed into parts so that each part depends on only one sensor. From this, it is shown that the DKF problem can be reformulated as a consensus optimization problem [11], which is the first contribution.

One important implication of the reformulation is that novel DKF algorithms employing distributed optimization methods can be developed. In view of this, a new DKF algorithm is proposed in this paper by employing the dual ascent method, which is the second contribution. It is noted that the proposed algorithm is an improved version of [37] in the sense that less information is required to choose estimator gains. In addition, it is proved that the proposed algorithm is unbiased, and that the variance of each estimator converges to the steady-state covariance of CKF [3, 4]. This ensures that the proposed algorithm asymptotically recovers the performance of CKF. The stability analysis is done under standard assumptions on the target system commonly made in the Kalman filtering [22, 23]. Precisely, we assume neither the local observability of the sensor network [13, 15, 34] nor the invertibility of the system matrix [6, 7, 18, 20, 21]. This contributes to the theoretical completeness of this study.

This paper is organized as follows. In Section 2, we connect DKF problem to a distributed optimization problem. A new DKF algorithm based on the dual ascent method is proposed in Section 3, and stability analysis is presented in Section 4. Numerical experiments are given in Section 5. Section 6 concludes the paper.

Notation: For matrices $A_1, \ldots, A_n$, $\text{diag}\{A_1, \ldots, A_n\}$ denotes the block diagonal matrix composed of $A_1, \ldots, A_n$. For vectors $a_1, \ldots, a_n$, $[a_1; \cdots; a_n]^T := [a_1, \cdots, a_n]^T$, and $[A_1; \cdots; A_n]$ with matrices $A_i$ defined similarly. $I_n \in \mathbb{R}^{n \times n}$ denotes the vector whose components are all 1, and $I_n \in \mathbb{R}^{n \times n}$ and $0_n \in \mathbb{R}^{n \times n}$ are the identity matrix and the zero matrix, respectively. For a symmetric matrix $A$, $A > 0$ ($A \succeq 0$, resp.) denotes that $A$ is a positive definite (semidefinite, resp.) matrix. We write $x \sim \mathcal{N}(\mu, \sigma^2)$ when $x$ is normally distributed with mean $\mu$ and variance $\sigma^2$. $\mathbb{E}\{x\}$ and $\text{Cov}(x)$ denote the expectation and covariance of $x$, respectively. For a symmetric matrix $M \in \mathbb{R}^{n \times n}$, $\text{vech}(M) \in \mathbb{R}^{(n+1)/2}$ denotes the half vectorization of $M$, a column vector obtained by using only the upper triangular part of $M$, and $\text{vech}^{-1}(\cdot)$ denotes the inverse of $\text{vech}(\cdot)$. For a function $f(x, y)$, $\nabla_x f(x, y)$ denotes the gradient vector with respect to $x$. $\delta_{kl}$ represents the Kronecker delta.

2 Distributed Kalman filtering and its Connection to Consensus Optimization

Consider a linear system with $N$ sensors given by

$$
\begin{align}
  x_{k+1} &= F x_k + w_k \\
  y_k &= H x_k + v_k
\end{align}
$$

where $x_k \in \mathbb{R}^n$ is the state vector, $y_k := [y_{1,k}; \cdots; y_{N,k}] \in \mathbb{R}^{nN}$ is the measurement vector, and $y_{i,k} \in \mathbb{R}^m_i$ is the measurement associated with sensor $i$ where $m_i$'s satisfy $\sum_{i=1}^N m_i = m$. $F$ is the system matrix and $H$ is the output matrix consisting of $H_i \in \mathbb{R}^{m_i \times n}$ that is the output matrix associated with sensor $i$. The process noise is denoted by $w_k$ and $v_{i,k}$ is the measurement noise on sensor $i$, which are zero-mean Gaussian. $w_k$ and $v_{i,k}$ are mutually uncorrelated jointly Gaussian and white, i.e., $\mathbb{E}\{w_k|y_{i,k}\} = Q\delta_{ik}$, $\mathbb{E}\{v_{i,k}|y_{j,k}\} = R_i\delta_{ij}\delta_{kl}$, and $\mathbb{E}\{w_k|y_{i,k}\} = 0$ for any $i, j = 1, \ldots, N$ and positive integers $k, l$. It is assumed that $Q > 0$ and $R := \text{diag}\{R_1, \ldots, R_N\} > 0$. In addition, let the initial state vector $x_0$ be Gaussian, with mean $\mathbb{E}\{x_0\}$ and covariance $P_0 > 0$, i.e., $x_0 \sim \mathcal{N}(\mathbb{E}\{x_0\}, P_0)$. It is supposed that $x_0$ is uncorrelated with $w_k$ and $v_k$.

Assumption 1 The pair $(F, H)$ is observable.

Remark 2 Assumption 1 means that it may not be possible to estimate the state of the system using the measurement from a single sensor, i.e., $(F, H)$ may not be observable, while the whole sensor network consisting of $N$ sensors satisfies the usual sense of observability.

If all the measurements from $N$ sensors are collected and processed altogether, the problem can be seen as the one with an imaginary sensor that measures $y_k$ with knowledge on $H$, thus called centralized Kalman filtering. The filtering consists of two steps, prediction and correction, and it is well known that the update rules can be derived from MAP (maximum a posteriori) approach in the Bayesian framework.

Let $Y_k := [y_0; \cdots; y_k]$. In the view of MAP [31], the optimal estimate of Kalman filtering is defined as $\hat{x}_k = \text{argmax}_{x_k} p(x_k|Y_k)$. Suppose that $\hat{x}_{k-1} = \mathbb{E}\{x_{k-1}|Y_{k-1}\}$ and $P_{k-1} = \text{Cov}(x_{k-1}|Y_{k-1})$ be the estimate of the state and its covariance at preceding time are given. In the prediction step, the predictive estimate $\hat{x}_{k|k-1}$ and covariance matrix $P_{k|k-1}$ are computed as $\hat{x}_{k|k-1} = F\hat{x}_{k-1}$ and $P_{k|k-1} = FP_{k-1}F^T + Q$.
By defining $z_{c,k} = [y_k; \hat{x}_{k|k-1}]$, $H_c = [H; I_n]$, $S_{c,k} = \text{diag}\{R, P_{k|k-1}\}$, the optimal estimate can be equivalently obtained as $\hat{x}_k = \arg\min_{\xi_k} f_{c,k}(\xi_k)$ where $\xi_k \in \mathbb{R}^n$ is the free variable and the cost function is given by

$$f_{c,k}(\xi_k) = \frac{1}{2}(z_{c,k} - H C \xi_k)^	op S_{c,k}^{-1}(z_{c,k} - H C \xi_k). \quad (2)$$

Since $f_{c,k}(\xi_k)$ is a convex function, provided that $P_{k|k-1} > 0$, $\hat{x}_k$ can be obtained from $\nabla_x f_{c,k}(\hat{x}_k) = 0$, from which we have the correction step as

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - H \hat{x}_{k|k-1})$$

$$P_k = P_{k|k-1} - P_{k|k-1} H^\top (H P_{k|k-1} H^\top + R)^{-1} H P_{k|k-1},$$

where $K_k = (H^\top H + P_{k|k-1})^{-1} H^\top H - 1$ that is the Kalman gain. For further details, see, e.g., [14, 22, 23, 31].

Now we consider the DKF problem. Each estimator in the network tries to find the optimal estimate by processing the local measurement and exchanging information with its neighbors. The communication network among estimators is modeled by a graph $G$, and $N$ and $N_i$ denote the node set and the neighbor set of estimator $i$, respectively. The Laplacian matrix associated with $G$ is denoted by $L \in \mathbb{R}^{N \times N}$ and $a_{ij}$ is a weight of the edge between nodes $i$ and $j$. It is known that $L$ has a simple zero eigenvalue corresponding to the unit eigenvector $u = \frac{1}{\sqrt{N}} 1_N$, and there exists an orthogonal matrix $U := [u \ W]$ such that $LU = UL$ where $W \in \mathbb{R}^{N \times (N-1)}$ is a matrix consisting of unit eigenvectors corresponding to the nonzero eigenvalues of $L$, denoted by $\sigma_2, \ldots, \sigma_N$, and $\Lambda = \text{diag}\{0, \Lambda\}$ with $\Lambda = \text{diag}\{\sigma_2, \ldots, \sigma_N\}$. To proceed, we define the following to simplify the notation.

$$1_N = 1_N \otimes I_n, \quad 1_N = I_N \otimes 1_n, \quad U = U \otimes I_n, \quad W = W \otimes I_n$$

$$\Lambda = \Lambda \otimes I_n, \quad \tilde{\Lambda} = \tilde{\Lambda} \otimes I_n.$$

For the network, we make the following assumption.

**Assumption 3** The network $G$ is undirected and connected, and the maximum eigenvalue of $L$, denoted by $\sigma_N$, is bounded by $\sigma$ which is known.

Under the setting (1), estimator $i$ acquires only the local measurement $y_{i,k}$, and the parameters $H_i$ and $R_i$ are kept private to estimator $i$. It is noted that the pair $(F, H_i)$ is not necessarily observable, and we assume that $F$ and $Q$ are open to all estimators, and $N$ is known.

Similar to CKF, DKF is performed in two steps, local prediction and distributed correction. In the local prediction step, each estimator predicts

$$\hat{x}_{i,k|k-1} = F \hat{x}_{i,k-1}, \quad P_{i,k|k-1} = FP_{i,k-1} F^\top + Q.$$

where $\hat{x}_{i,k|k-1}$ and $P_{i,k|k-1}$ are local estimates of $\hat{x}_{k|k-1}$ and $P_{k|k-1}$, respectively, that estimator $i$ holds.

The distributed correction step solves the MAP estimation problem in a distributed manner. First, we define $\xi_k = [y_k; \hat{x}_{i,k|k-1}]$, $H_c = [H_i; I_n]$, and $S_{i,k} = \text{diag}\{R_i, NP_{i,k|k-1}\}$ which depend on only local variables and parameters. Owing to the structure of $S_{i,k}$, the cost function $f_{c,k}(\xi_k)$ in (2) can be decomposed into

$$f_{c,k}(\xi_k) = \sum_{i=1}^N f_i(\xi_k)$$

where $f_i(\xi_k) = \frac{1}{2}(z_{i,k} - H_i \xi_k) S_{i,k}^{-1}(z_{i,k} - H_i \xi_k)$. It is noted that the cost becomes identical to that of CKF when the estimators reach a consensus on $\hat{x}_{i,k|k-1}$ and $P_{i,k|k-1}$ in the correction step at time $k-1$.

Allowing that each estimator holds its own optimization variable $\xi_{i,k} \in \mathbb{R}^n$ for $\xi_{i,k}$, DKF problem becomes a consensus optimization problem given by

$$\begin{aligned}
\min_{\xi_{1,k}, \ldots, \xi_{N,k}} & \quad \sum_{i=1}^N f_i(\xi_i) \\
\text{subject to} & \quad \xi_{1,k} = \cdots = \xi_{N,k}.
\end{aligned} \quad (P.1)$$

If there exists a distributed algorithm that finds a minimizer, we say that the algorithm solves DKF problem.

Since the kernel of Laplacian $L$ is span{1N}, the constraints of (P.1) can be written as $L \xi_{i,k} = 0$ where $\xi_{i,k} = [\xi_{i,1,k}; \cdots; \xi_{i,N,k}]$. We define the Lagrangian for (P.1) as

$$L_{\text{est},k}(\xi_k, \lambda_k) = \sum_{i=1}^N f_i(\xi_i) + \lambda_k^\top L \xi_{i,k}$$

where $\lambda_k \in \mathbb{R}^{Nn}$ is the Lagrange multiplier (dual variable) associated with the consensus constraint. Define

$$z_k = [z_{1,k}; \cdots; z_{N,k}], \quad \bar{H} = \text{diag}\{H_1, \ldots, H_N\}$$

$$S_k = \text{diag}\{S_{1,k}, \ldots, S_{N,k}\}, \quad H_k = H_N \bar{H} S^{-1}_k \bar{H}^\top \in \mathbb{R}^{n \times n}$$

Note that $H_k$ is a symmetric positive definite matrix.

We rewrite the Lagrangian (3) as $L_{\text{est},k}(\xi_k, \lambda_k) = \frac{1}{2}(z_k - H_k \xi_k) S_k^{-1}(z_k - H_k \xi_k) + \lambda_k^\top L \xi_k$ and compute its gradients with respect to $\xi_k$ and $\lambda_k$ as

$$\nabla_\xi L_{\text{est},k}(\xi_k, \lambda_k) = -H_k^\top S_k^{-1}(z_k - H_k \xi_k) + \lambda_k$$

$$\nabla_\lambda L_{\text{est},k}(\xi_k, \lambda_k) = L \xi_k.$$

From the optimality condition for $(\xi_k^*, \lambda_k^*)$, we have the saddle point equation (KKT conditions) given by

$$\begin{bmatrix}
-H_k^\top S_k^{-1} H_k - L
0
\end{bmatrix}
\begin{bmatrix}
\xi_k^*
\lambda_k^*
\end{bmatrix} =
\begin{bmatrix}
0
0
\end{bmatrix}$$

where $\xi_k^* := [\xi_{1,k}^*; \cdots; \xi_{N,k}^*]$ and $\lambda_k^* := [\lambda_{1,k}^*; \cdots; \lambda_{N,k}^*]$. 

Lemma 4 Suppose that $P_{i,k|k-1}$ is symmetric positive definite for all $i \in \mathcal{N}$. Then, the solutions to DKF problem are given by $(\xi_i^k, \lambda_i^k) = (\hat{1}_N, \hat{1}_N, \hat{1}_N \lambda_k + \tilde{\lambda}_k)$ where
\[
\xi_i^k = H_k^{-1}L_k^T \hat{S}_k^{-1} z_k,
\]
where $\tilde{\lambda}_k \in \mathbb{R}^n$ is a vector that is uniquely determined in terms of $W$, $\hat{\lambda}$, and $\xi_i^k$, and $\lambda_k \in \mathbb{R}^n$ is an arbitrary vector.

**PROOF.** We refer the reader to the paper [37]. □

The covariance correction step can also be formulated as an optimization problem. It is formulated under the framework of information filtering [38]. Let $\Omega_k := P_k^{-1}$ and $P_k^{-1} \hat{x}_k$ be the information matrix and information vector, respectively. We also define $\Omega_{k|k-1} = P_{k|k-1}^{-1}$.

At each time $k$, the information matrix is updated in two steps; prediction and correction, namely
\[
\Omega_{k|k-1} = (F \Omega_{k-1}^{-1} F^T + Q)^{-1},
\]
(9a)
\[
\Omega_k = \Omega_{k|k-1} + H^T \hat{R}^{-1} H.
\]
(9b)

It is noted that if the information matrix is updated by (9a) and (9b), then $\Omega_{k|k-1}$ converges to $P^{*-1}$ where $P^*$ is a unique positive definite solution to the discrete-time algebraic Riccati equation [1] given by
\[
P^* = FP^* F^T - FP^* H^T (H P^* H^T + \hat{R})^{-1} H P^* F^T + Q,
\]
(10)
and the limit of $\Omega_{k|k-1}$ is the same no matter what the initial condition $\Omega_0$ is chosen as long as $\Omega_0 > 0$.

Thanks to the convergence property, it is expected that $\Omega_{i,k|k-1}$ will converge to $P^{*-1}$ provided that the estimators can compute the global information rate matrix $H^T \hat{R}^{-1} H$ in a distributed way; the estimators need not choose the same initial condition for $\Omega_k$. In fact, this idea is widely used, see, e.g., [24, 32–34]. Accordingly, we formulate a consensus optimization problem as

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \sum_{i=1}^{N} \| N \omega_i^\delta - \theta_i \|^2 \\
\text{subject to} & \quad \theta_1 = \cdots = \theta_N
\end{align*}
\]
(P.2)

where $\omega_i^\delta := \text{vech} \left( H_i^T R_i^{-1} H_i \right) \in \mathbb{R}^{n_{\text{cov}}}$, $\theta_i \in \mathbb{R}^{n_{\text{cov}}}$ is estimator $i$’s decision variable, and $n_{\text{cov}} = n(n + 1)/2$.

To solve (P.2), we construct the Lagrangian as
\[
\mathcal{L}(\theta, \nu) = \frac{1}{2} (N\omega^\delta - \theta)^T (N\omega^\delta - \theta) + \nu^T \hat{L} \theta,
\]
where $\nu \in \mathbb{R}^{n_{\text{cov}}}$ is the Lagrange multiplier, $\omega^\delta = [\omega_1^\delta; \cdots; \omega_N^\delta]$, $\theta = [\theta_1; \cdots; \theta_N]$, and $\hat{L} = L \otimes I_{n_{\text{cov}}}$, and this leads to the following result, where we use the notation $\hat{1}_N$, $\hat{1}_N$, $\hat{1}_N$, $\hat{1}_N$, $\hat{1}_N$, $\hat{1}_N$, $\hat{1}_N$, and $\hat{A}$ defined similarly to $\hat{L}$.

Lemma 5 The solutions to (P.2) are parameterized as $(\theta^*, \nu^*) = (\hat{1}_N \theta^!, \hat{1}_N \hat{\nu} + \bar{v})$ where $\theta^! := \text{vech}(\theta^1)$, $\Theta^! := \sum_{i=1}^{N} (H_i^T R_i^{-1} H_i) = H^T \hat{R}^{-1} H$, $\hat{\nu} = \hat{\nu} \hat{A}^{-1} \hat{\nu}^T (N\omega^\delta - \hat{1}_N \theta^!)$, and $\bar{v} \in \mathbb{R}^{n_{\text{cov}}}$ is arbitrary.

**PROOF.** Similar to Lemma 4 and thus omitted. □

3 Dual Ascent Distributed Kalman Filtering

In this section, we propose a new DKF algorithm employing a distributed optimization method. The proposed DKF algorithm consists of two steps, prediction and correction, as CKF does. In the prediction step, each estimator predicts the estimate and covariance locally. The correction step, which is our main concern, is the process of finding the minimizers of (P.1) and (P.2) in a distributed fashion. For this, various distributed optimization algorithms can be employed allowing each estimator to reach the minimizers. We call this step distributed correction. In this paper, we employ the dual ascent method [11] for the distributed correction step, which is a well-known convex optimization method. The update of the dual variable is performed by using the gradient ascent and the primal variable is updated by finding the minimizer of the local Lagrangian with the updated dual variable. In each correction step, additional iterations can be conducted so that the primal variable of each estimator converges to the minimizer with small error. Such iteration is called subiteration and is indexed by the subscript $l$ hereafter.

For (P.1), the dual ascent based update rule is given by
\[
\lambda_{k,l+1} = \lambda_{k,l} + \hat{A}_k K_{k,l}^\text{dual} \xi_{k,l},
\]
(11a)
\[
\xi_{k,l+1} = K_{k,l}^\text{cons} (H_k^T \hat{S}_k^{-1} z_k - L \lambda_{k,l+1}),
\]
(11b)
\[
K_{k,l}^\text{cons} = (H_k^T \hat{S}_k^{-1} H_k)^{-1},
\]
\[K_{k,l}^\text{dual} = \text{diag}(K_{1,k,l}^\text{dual}, \cdots, K_{N,k,l}^\text{dual})
\]
\[
\hat{A}_k = A_k \otimes I_n, A_k = \text{diag}(\alpha_{\lambda,1}, \cdots, \alpha_{\lambda,N}), \alpha_{\lambda,i} > 0, i = 1, \ldots, N \text{ is the update gain and } \epsilon_i \text{ is an arbitrary positive scalar. When } l \text{ reaches } l^*, \text{ the number of subiterations, the distributed correction step stops and we update } \hat{x}_{i,k} = \xi_{i,k,l}.
\]

One advantage of using the dual ascent for consensus optimization is that the resulting algorithm has a distributed form owing to the structure of $\nabla \xi \mathcal{L}_{\text{est}, k}(\xi_k, \lambda_k)$ and $\nabla \lambda \mathcal{L}_{\text{est}, k}(\xi_k, \lambda_k)$: see (5) and (6). In addition, $(\xi_{k,l}, \lambda_{k,l})$ satisfies the dual feasibility equation for any $k$ and $l$, i.e., $-H_k^T \hat{S}_k^{-1} (z_k - H_k \xi_{k,l}) + L \lambda_{k,l} = 0, \forall k, l \in \mathcal{N}$.

It is emphasized that the update gain for the dual variable update (11a) is chosen as $\hat{A}_k K_{k,l}^\text{dual}$ rather than a scalar $\alpha_\lambda$. Using a scalar update gain is typical in dual ascent approach and it is also the case with the preliminary version of this paper [37]. In [37], the update gain
Algorithm 1 DA-DKF

1: Initialization: Take arbitrary $\hat{x}_{i,0}, P_{i,0} > 0$, and $\epsilon_i > 0$. Set $\theta_{i,0,i^*} = \text{vech} (H_i^T R_i^{-1} H_i)$, $v_{i,0,i^*} = 0$, $k = 1$.
2: repeat
3: // Local prediction
4: $\hat{x}_{i,k|k-1} = F \hat{x}_{i,k-1}, P_{i,k|k-1} = FP_{i,k-1}F^T + Q$
5: // Distributed correction
6: $\xi_{i,k,0} = \hat{x}_{i,k|k-1}, \lambda_{i,k,0} = 0$
7: $\theta_{i,k,0} = \theta_{i,k-1,i^*}, v_{i,k,0} = v_{i,k-1,i^*}$
8: while $l = 0, \ldots, l^* - 1$, do
9: // Distributed estimate update (11)
10: $\lambda_{i,k,l+1} = \lambda_{i,k,l} + \alpha_{i,k,l} K_{i,k}^{\text{dual}} \sum_{j \in \mathcal{N}_i} a_{ij} (\xi_{i,k,l} - \xi_{j,k,l})$
11: $\xi_{i,k,l+1} = \hat{x}_{i,k|k-1} + K_{i,k}^{\text{dual}} (y_{i,k} - H_i \hat{x}_{i,k|k-1} - \xi_{j,k,l})$
12: // Distributed information update (12)
13: $v_{i,k,l+1} = v_{i,k,l} + \alpha_{i,k,l} \sum_{j \in \mathcal{N}_i} a_{ij} (\theta_{i,k,l} - \theta_{j,k,l})$
14: $\theta_{i,k,l+1} = N \omega_i \theta_{i,k,l} + \sum_{j \in \mathcal{N}_i} a_{ij} (v_{i,k,l+1} - v_{j,k,l+1})$
15: end
16: $\hat{x}_{i,k} = \xi_{i,k,l^*}$, $P_{i,k} = (P_{i,k|k-1}^{-1} + \Theta_{i,k,l^*})^{-1}$
17: $k = k + 1$
18: until $k = \infty$

is chosen using the maximum norm of local covariance, which is not easy to obtain in advance. As will be seen in the stability proof (Lemma 10), the matrix update gain $A_{i,k} K_{i,k}^{\text{dual}}$ relaxes the dependence so that the new gain with $\alpha_{i,i^*} \leq 2/\sigma^2$ ($\sigma$ is the upper bound of the maximum eigenvalue of $L$) ensures the stability of the proposed approach. The rationale behind this choice is that with this new gain matrix, the dynamics of $\xi_{i,k}$ has $n$ simple eigenvalues at 1 while the other eigenvalues are stable for sufficiently large $k$, independently of local covariance.

Similarly to the problem (P.1), the distributed correction algorithm for the problem (P.2) can be obtained as

\begin{align}
\upsilon_{i,k,l+1} &= \upsilon_{i,k,l} + \bar{A}_i \hat{\Theta}_{i,k,l} \\
\theta_{i,k,l+1} &= N \omega_i \theta_{i,k,l} - \bar{\upsilon}_{i,k,l+1}
\end{align}

(12a) (12b)

where $\bar{A}_i$ is defined similarly to $A_i$ and $\alpha_{i,i^*} > 0$ is the update gain. At the end of the distributed correction, we have $P_{i,k} = (P_{i,k|k-1}^{-1} + \Theta_{i,k,l^*})^{-1}$. Updating the local covariance matrix with the exchanged information rate matrix is widely used in existing DKF algorithms [7, 30, 33, 35]. See also [43] for the case with system uncertainty.

Using (11) and (12), we propose DA-DKF (dual-ascent based distributed Kalman filtering) described in Algorithm 1. The design parameters are $\alpha_{i,i^*}$, $\alpha_{i,i^*}$, $\epsilon_i$, and $l^*$. Among these parameters, $\alpha_{i,i^*}$ and $\alpha_{i,i^*}$ should be selected so that the primal variable of each problem converges to the minimizer, as $l$ increases. Meanwhile, a small number of subiterations is preferred in practice in order to reduce communication and computation load.

In the next section, we provide a sufficient condition that ensures the stability of DA-DKF, i.e.,

\[
\lim_{k \to \infty} \mathbb{E} \{x_k - \xi_{i,k,l^*} \} = 0, \quad \lim_{k \to \infty} \| P^* - P_{i,k|k-1} \| = 0. \tag{13}
\]

where $P^*$ is the unique positive definite solution to (10). In fact, (13) is the key stability results in the case of CKF [22, 23], which implies that CKF is unbiased and converges to the steady-state CKF asymptotically. Therefore, if (13) holds true, we state that DA-DKF asymptotically recovers the performance of CKF.

4 Stability Analysis

We start the stability analysis by noting that in Algorithm 1 the update rules for the estimate and the covariance are of cascade type and the latter is autonomous. Based on this fact, we first state results on the local covariances, which include a sufficient condition on the update gain for (P.2) and the boundedness of the local covariances (Lemmas 6 and 8). The convergence of local covariances is proved in Theorem 9. Once the boundedness of covariance is guaranteed, we establish a sufficient condition for the update gain $\alpha_{i,i^*}$ that guarantees asymptotic convergence of the residual $e_{k,i}^* := \xi_k^\ast - \xi_{k,i}$ as $l$ increases. Then, we derive the dynamics of the expectation of estimate error, denoted by $\mathbb{E} \{ e_k \}$, where $e_k := [e_{k,i}^\ast, e_{k,i}^*, e_{k,i}^\dagger]$, $e_{k,i}^\ast := x_k - \hat{x}_{k,i}$, and $e_{k,i}^* := \hat{x}_{k,i}^\ast - \hat{x}_{k,i}$. Subsequently, convergence of $\mathbb{E} \{ e_k \}$ is proved in Theorem 12.

Lemma 6 Suppose that Assumption 3 holds true and let $l^*$ be any positive integer. Then, the sequence $\{ \xi_{i,k,l^*} := \text{vech}^{-1}(\theta_{i,k,l^*}) \}$ generated by Algorithm 1 converges to the global information rate matrix $H_i^T R_i^{-1} H_i$, as $k$ goes to infinity, if the update gain $\alpha_{i,i^*}$ is chosen such that

\[
0 < \alpha_{i,i^*} < 2/\sigma^2. \tag{14}
\]

Moreover, if $\alpha_{i,i^*}$ satisfies (14), then for any $0 < \kappa < 1$ there exists $k_0 > 0$ such that

\[
0 \leq (1 - \kappa) \Theta_i \leq \Theta_{i,k,l^*} \leq (1 + \kappa) \Theta_i, \quad \forall k \geq k_0. \tag{15}
\]

PROOF. See Appendix A.

Remark 7 Since $P_{i,k}$ is an estimate of covariance, it is meaningful when $P_{i,k} \geq 0$. In fact, in the case of CKF, the positive definiteness of $P_k, k \geq 1$, is guaranteed whenever $Q$ is positive definite. Unfortunately, the positive definiteness of $P_{i,k}$ generated by DA-DKF may not be preserved or may not even be well-defined if $\Theta_{i,k,l^*}$
is not positive semidefinite since $P_{i,k}$ is computed by $P_{i,k} = (P_{i,k_{k-1}}^{-1} + \Theta_{i,k,l})^{-1}$ (line 16 of Algorithm 1).

In fact, the positive definiteness can be guaranteed by making $\Theta_{i,k,l}$ positive semidefinite. This is because the iteration of $\Theta_{i,k,l}$ (lines 13 and 14 of the algorithm) is done autonomously and independently of $P_{i,k}$. Based on this observation and the result given in Lemma 6, we propose two simple ways to keep $P_{i,k}$ to be positive definite for all $k \geq 1$; i) using sufficiently large $i^*$ and ii) exception handling when $\Theta_{i,k,l}$ is not positive semidefinite.

For the former case, use a sufficiently large $i^*$ such that $\Theta_{i,k,l} \geq 0$ for any $k \geq 1, l^* \geq l^*$. Note that the existence of $l^*$ is clear from Lemma 6. One example of exception handling would be to use a projected matrix of $\Theta_{i,k,l}$ onto a set of positive semidefinite matrices, denoted by $\text{Proj}_{\geq 0}(\Theta_{i,k,l})$, to compute $P_{i,k}$, namely, modify line 16 of Algorithm 1 as $P_{i,k} = (P_{i,k_{k-1}}^{-1} + \text{Proj}_{\geq 0}(\Theta_{i,k,l}))^{-1}$. Obtaining $\text{Proj}_{\geq 0}(\Theta_{i,k,l})$ can be accomplished by formulating a semidefinite programming and solving it [11]. A simple and effective way to find that matrix, using an image of $P_{i,k_{k-1}}$, is introduced in [41].

Lemma 8 Consider Algorithm 1 (or the modified one discussed in Remark 7) and suppose that Assumptions 1 and 3 hold true. Then, if the update gain $\alpha_{v,i}$ is chosen such that (14) is satisfied, there exist symmetric positive definite matrices $\overline{P} > 0$ and $\overline{T} < \infty$ such that $\overline{P} \leq P_{i,k} \leq \overline{T}$ and $P_{i,k_{k-1}} \leq \overline{T}$ for all $k \geq 1$ and $i \in N$.

**Proof.** See Appendix B. \qed

In CKF, it is known that the covariance converges to $P^* > 0$ that is a unique positive definite solution of the discrete-time algebraic Riccati equation

$$P = F \left\{ P - PH^T (H^T PH + \overline{R})^{-1} HP \right\} F^T + Q \quad (16)$$

for any initial covariance $P_0 > 0$ [22]. In the following theorem, we state the asymptotic performance recovery of DKF to CKF in terms of covariance.

Theorem 9 Suppose that Assumptions 1 and 3 hold true, and the update gain $\alpha_{v,i}$ is chosen such that $0 < \alpha_{v,i} < 2/\sigma^2$. Then, the local covariance matrix $P_{i,k_{k-1}}$ generated by Algorithm 1 converges to the unique positive definite solution of (16), i.e., that of CKF.

**Proof.** From the covariance prediction rule given by $P_{i,k_{k+1}} = FP_{i,k}F^T + Q$, we have

$$P_{i,k_{k+1}} = F(P_{i,k_{k-1}}^{-1} + \Theta_{i,k,l})^{-1} F^T + Q = F(P_{i,k_{k-1}}^{-1} + \Theta^T)^{-1} F^T + Q + M_{i,k,l} \quad (17)$$

where $M_{i,k,l} = F(P_{i,k_{k-1}}^{-1} + \Theta_{i,k,l})^{-1} F^T - F(P_{i,k_{k-1}}^{-1} + \Theta^T)^{-1} F^T$. By Lemma 6, it holds that $\Theta_{i,k,l} \geq 0$ for all $k \geq k, i \in N$. Then, we can rewrite $M_{i,k,l}$ as $M_{i,k,l} = F(P_{i,k_{k-1}}^{-1} + \Theta_{i,k,l})^{-1} e_{i,k,l} - F(P_{i,k_{k-1}}^{-1} + \Theta^T)^{-1} F^T$

where $e_{i,k,l} = : \Theta^T - \Theta_{i,k,l}$, and this leads to $\|M_{i,k,l}\| \leq \bar{p} \|e_{i,k,l}\|$ where $\bar{p}$ is a positive scalar such that $\|\bar{T}\| \leq \bar{p}$. Since $\lim_{k \to \infty} e_{i,k,l} = 0$ by Lemma 6, we have

$$\lim_{k \to \infty} \|M_{i,k,l}\| = 0. \quad (18)$$

Let $\hat{P}_{i,k} = P^* - P_{i,k_{k-1}}$ where $P^*$ is the unique solution of (16). By the matrix inversion lemma, one has

$$P^* = F(P^{* - 1} + \Theta^T)^{-1} F^T + Q. \quad (19)$$

Then, we have from (17) and (19) that

$$\hat{P}_{i,k_{k+1}} = P^* - P_{i,k_{k+1}} = F \left( (P^{*- 1} + \Theta^T)^{-1} - (P_{i,k_{k-1}}^{-1} + \Theta^T)^{-1} \right) F^T - M_{i,k,l} \quad (20)$$

Noting that $\Theta^T = H^T \overline{R} \overline{H}$, one can derive

$$F(P^{*- 1} + \Theta^T)^{-1} P^{*- 1} = F(I_n - K^* H) \quad (21)$$

where $K^* = P^* H^T (\overline{R} + \overline{H} P^* H^T)^{-1}$ and $K_{i,k} = P_{i,k_{k-1}} H^T (\overline{R} + \overline{H} P_{i,k_{k-1}} H^T)^{-1}$ are the Kalman gains in the steady-state and at time $k$, respectively. Define $\Phi^* = F(I_n - K^* H)$ and $\Phi_{i,k} = F(I_n - K_{i,k} H)$. Then, from (20) and (21), we have

$$\hat{P}_{i,k_{k+1}} = \Phi^* \hat{P}_{i,k} \Phi_{i,k}^T - M_{i,k,l} \quad (22)$$

We now investigate $\Phi^*$ and $\Phi_{i,k}$. Since $(F, \sqrt{Q})$ is controllable and $(F, H)$ is observable, by Lemma D.2 in [23], $\Phi^*$ is Schur stable and there exists $a^* > 0$ such that

$$\|\Phi^*\| < a^* < 1. \quad (23)$$

Regarding $\Phi_{i,k}$, we rewrite (17) as

$$P_{i,k_{k+1}} = \Phi_{i,k} P_{i,k} \Phi_{i,k}^T + F P_{i,k_{k-1}} H^T K_{i,k} F^T - F K_{i,k} H P_{i,k_{k-1}} H^T K_{i,k} F^T + Q + M_{i,k,l} \quad (24)$$

Recalling that $K_{i,k} = P_{i,k_{k-1}} H^T (\overline{R} + \overline{H} P_{i,k_{k-1}} H^T)^{-1}$, we have $K_{i,k} H P_{i,k_{k-1}} H^T K_{i,k} = P_{i,k_{k-1}} H^T K_{i,k}^2 -$
Lemma 10 Suppose that Assumption 3 holds true. Consider Algorithm 1 with the update gains \( \alpha_{v,i} \) and \( \alpha_{\lambda,i} \) satisfying

\[
0 < \alpha_{v,i} < 2/\sigma^2, \quad 0 < \alpha_{\lambda,i} < 2/\sigma^2.
\]

Then, there exists \( k^* \geq 1 \) such that for any \( k \geq k^* \), \( \xi_{k,l} \) converges to \( \xi^*_l \) as \( l \) goes to infinity and \( \lambda_{k,l} \) is bounded for all \( l \). Moreover, there exist \( 0 < \mu < 1 \) and \( k^* \geq 1 \) such that \( \|\xi_k\| \leq \mu, \forall k \geq k^* \) where \( \tilde{\xi}_k = I_{N-1} - \sum_{j=1}^{N} K_{k,1}^\top \lambda_j \). Define \( \mathcal{P}_{k+1|k} = A_{k,1}^\top K_{k,1}^\top \lambda_j \). Recalling that \( H = \text{diag}(H_1, \ldots, H_N) \) and \( P_{k+1|k} = \text{diag}(P_{1,k+1|k}, \ldots, P_{N,k+1|k}) \), we rewrite the dynamics of \( e_{k+1}^\xi \) in (28) as

\[
\mathcal{P}_{k+1|k} H_{k+1} e_{k+1}^\xi = \mathcal{F}_{k+1}^\xi + e_{k+1}^\xi \mathcal{G}_{k+1}^\xi e_{k+1}^\xi.
\]

Lemma 11 Consider \( \xi_{k,l} \) generated by Algorithm 1. Then, the dynamics of \( e_k := [e_k^1; e_k^2] \) is given by

\[
\begin{bmatrix}
 e_{k+1}^1 \\
 e_{k+1}^2
\end{bmatrix} =
\begin{bmatrix}
 E_{k+1}^1 & E_{k+1}^2 \\
 0 & \bar{E}_{k+1}^2
\end{bmatrix}
\begin{bmatrix}
 e_k^1 \\
 e_k^2
\end{bmatrix}
\]

(28)

where \( E_k^1 = \mathcal{U}^\top (I_{N} - \frac{1}{N} H_{k+1}^\top \mathcal{F}_{k+1}^\xi) \mathcal{U} \), \( E_k^2 = \mathcal{U}^\top (I_{N} - \frac{1}{N} H_{k+1}^\top \mathcal{F}_{k+1}^\xi) \mathcal{U} \), and \( \bar{E}_{k+1}^2 = \mathcal{U}^\top (I_{N} - \frac{1}{N} H_{k+1}^\top \mathcal{F}_{k+1}^\xi) \mathcal{U} \). Moreover, if \( \alpha_{\lambda,i} \) is chosen such that \( 0 < \alpha_{\lambda,i} < 2/\sigma^2 \), then there exists \( k^* \geq 1 \) such that \( \bar{E}_{k+1}^2 (\tilde{\xi}_k) \) is Schur stable for all \( k \geq k^* \).

PROOF. See Appendix D.

Finally, we state the stability result on estimates.

Theorem 12 Suppose that Assumptions 1 and 3 hold true and consider Algorithm 1. If \( \alpha_{\lambda,i} \) and \( \alpha_{v,i} \) are chosen such that \( 0 < \alpha_{\lambda,i} < 2/\sigma^2 \) and \( 0 < \alpha_{v,i} < 2/\sigma^2 \), then there exists \( l^* \) such that for any \( l^* \geq l \), it holds that

\[
\lim_{k \to \infty} E\{ x_k - \xi_{k,l} \} = 0.
\]
We then add and subtract $e_k^\dagger F^T P_{k+1[k]} F e_k^\dagger$ to complete the square, namely,
$$V_{k+1}^\dagger = -e_k^\dagger H^T R^{-1} H e_k + u_{k+1}^\dagger P_{k+1[k]}^\dagger u_{k+1} + e_k^\dagger F^T P_{k+1[k]}^\dagger F e_k + 2e_k^\dagger P_{k+1[k]}^\dagger \xi e_k^\dagger. \tag{32}$$

By the matrix inversion lemma and the continuity argument, the third term $e_k^\dagger F^T P_{k+1[k]}^\dagger F e_k$ is bounded as $e_k^\dagger F^T P_{k+1[k]}^\dagger F e_k \leq e_k^\dagger \sum_{i=1}^N P_{i,k}^\dagger e_k^\dagger$. Moreover, from the fact that $P_{i,k}^\dagger = P_{i,k}^{-1} + \Theta_{i,k,l}$ and $\sum_{i=1}^N \Theta_{i,k,l} = N \sum_{j=1}^N H_i^T R^{-1} H_i$ for any $k \geq 1$ and $l^* \geq 1$ (see the proof of Lemma 6), we have
$$e_k^\dagger F^T P_{k+1[k]}^\dagger F e_k \leq V_k^\dagger. \tag{33}$$

Let $\bar{f} > 0$ be such that $\|F\| \leq \bar{f}$. Then, from (28), we know that there exist constants $c_0$ and $c_1$, which depend on $\|F\|$, $\|P\|$, and $\bar{f}$, such that
$$2e_k^\dagger P_{k+1[k]}^\dagger \xi e_k^\dagger \leq c_0 \|e_k^\dagger\|^2 \|e_k^\dagger\| + c_1 \|\xi\|^2. \tag{34}$$

Applying the relations (33) and (34) to (32) yields $V_{k+1}^\dagger - V_k^\dagger \leq -e_k^\dagger H^T R^{-1} H e_k + u_{k+1}^\dagger P_{k+1[k]}^\dagger u_{k+1} + c_0 \|e_k^\dagger\|^2 \|e_k^\dagger\| + c_1 \|\xi\|^2$, and by Young’s inequality, it follows that
$$V_{k+1}^\dagger - V_k^\dagger \leq -e_k^\dagger H^T R^{-1} H e_k + u_{k+1}^\dagger P_{k+1[k]}^\dagger u_{k+1} + \left(\frac{c_0}{2\varepsilon} + c_1\right) \|\xi\|^2 \tag{35}$$

where $\varepsilon > 0$ is a constant to be determined later.

Let us define $J_k = \sum_{j=0}^{n-1} (e_{k+1+j}^\dagger + u_{k+1+j}^\dagger P_{k+1[j]}^\dagger u_{k+1+j}^\dagger)$ and sum up (35) from $k$ to $k+n$ to have
$$V_{k+n}^\dagger - V_k^\dagger \leq -J_k + \frac{c_0}{2\varepsilon} \sum_{j=0}^{n-1} \|e_{k+j}\|^2 + \left(\frac{c_0}{2} + c_1\right) \sum_{j=0}^{n-1} \|\xi_{k+j}\|^2. \tag{36}$$

We would like to bound the function $J_k$. Let $\mathcal{E}_{k+1} = [e_{k+1}^\dagger; \ldots; e_{k+n}^\dagger]$ and $\mathcal{W}_{k+1} = [u_{k+1}^\dagger; \ldots; u_{k+n}^\dagger]$. Then, from (31), one obtains
$$\mathcal{E}_{k+1}^\dagger = \begin{bmatrix} F^\dagger & I \\ F^2 & F \\ \vdots & \vdots \\ F^n & \cdots \cdots \\ \end{bmatrix} \begin{bmatrix} e_k^\dagger \\ e_{k+1}^\dagger \\ \vdots \\ e_{k+n}^\dagger \\ \end{bmatrix} + \mathcal{W}_{k+1} \tag{37}$$

By letting $\mathcal{H} = I_n \otimes H$, $\mathcal{R} = I_n \otimes \bar{R}$, and $\mathcal{P}_{k+1[k]} = \text{diag}\{P_{k+1[k]}, \ldots, P_{k+n[k]}\}$, one can rewrite $J_k$ as
$$J_k = (\mathcal{F} e_k^\dagger + \mathcal{W}_{k+1}^\dagger F^\dagger \mathcal{H}^\dagger \mathcal{R}^\dagger \mathcal{H} (\mathcal{F} e_k^\dagger + \mathcal{W}_{k+1}^\dagger) + \mathcal{W}_{k+1}^\dagger \mathcal{P}_{k+1[k]}^\dagger \mathcal{W}_{k+1}).$$
Since $J_k$ is convex and quadratic with respect to $\mathcal{W}_{k+1}$, one can obtain $\mathcal{W}_{k+1}^*$ minimizing $J_k$ from $\nabla J_k = 0$, i.e., $\mathcal{W}_{k+1} = -(\mathcal{F}^\dagger \mathcal{H}^\dagger \mathcal{R}^\dagger \mathcal{H} (\mathcal{F} e_k^\dagger + \mathcal{W}_{k+1}) + \mathcal{W}_{k+1}^\dagger \mathcal{P}_{k+1[k]}^\dagger \mathcal{W}_{k+1})^{-1} \mathcal{F} e_k^\dagger \mathcal{H}^\dagger \mathcal{R}^\dagger \mathcal{H} e_k^\dagger$, where $\mathcal{O}$ is the observability matrix of the pair $(F, H)$. Since $\mathcal{O}$ has full column rank and $\mathcal{R} > 0$, there exists $c^1 > 0$ such that
$$2c^1 \|e_k^\dagger\|^2 \leq J_k^* \leq J_k. \tag{37}$$

From (28), there exists $c > 0$ such that $\|e_{k+1}^\dagger\|^2 \leq c\|e_k^\dagger\|^2 + c\|\xi\|^2$. Then, with $\tilde{c} = \max\{1, c^{n-1}\}$, the second term in (36) can be bounded as
$$\sum_{j=1}^{n-1} \|e_{k+j}\|^2 \leq (n-1)c\|e_k^\dagger\|^2 + \sum_{j=1}^{n-1} \|\xi_{k+j}\|^2. \tag{38}$$

In addition, from (28) and the structure of $\hat{E}_{k+1}(\bar{X}_{k+1})^*$, we have
$$\|e_{k+1}^\dagger\|^2 = e_{k+1}^\dagger F^\dagger (\bar{X}_{k+1})^* e_{k+1}^\dagger \hat{E}_{k+1}(\bar{X}_{k+1})^* F e_{k+1}^\dagger = e_{k+1}^\dagger F^\dagger \text{diag}(0, \ldots, 0) \hat{X}_{k+1})^* \hat{E}_{k+1}(\bar{X}_{k+1})^* F e_{k+1}^\dagger.$$

Since there exists $\mu > 0$ such that $\|\bar{X}_{k+1}\| \leq \mu < 1$ for all $k \geq k^*$ (see Lemma 10), we have
$$\|e_{k+1}^\dagger\|^2 \leq 2f^2\mu^{2\mu} \|e_k^\dagger\|^2, \tag{39}$$

which results in
$$\sum_{j=0}^{n-1} \|e_{k+j}\|^2 \leq \sum_{j=0}^{n-1} \left(2f^2\mu^{2\mu}\right)^j \|e_{k}^\dagger\|^2. \tag{40}$$

Take $\varepsilon = (n-1)c\tilde{c}/(2c^1)$ and substitute (37), (38) and (40) into (36). Then, one has
$$V_{k+n}^\dagger - V_k^\dagger \leq -c^1 \|e_k^\dagger\|^2 + c_2 \sum_{j=0}^{n-1} \left(2f^2\mu^{2\mu}\right)^j \|e_{k}^\dagger\|^2 \tag{41}$$

where $c_2 = (n-1)c\tilde{c}/(2c^1) + c_0/2 + c_1$.

Meanwhile, let $V_{k+1}^\xi = e_{k+1}^\dagger \xi$. Then, from (39), one has $V_{k+1}^\xi \leq 2f^2\mu^{2\mu} \|e_k^\dagger\|^2$, and this yields $V_{k+n}^\xi \leq 2f^2\mu^{2\mu} \|e_{k+n-1}^\dagger\|^2 \leq (2f^2\mu^{2\mu})^{n} \|e_k^\dagger\|^2$. Hence,
$$V_{k+n}^\xi - V_k^\xi \leq -\left(1 - (2f^2\mu^{2\mu})^n\right) \|e_k^\dagger\|^2. \tag{42}$$
We now consider a Lyapunov function given by $V_k = V_k^\dagger + \gamma V_k^\circ$, where $\gamma$ is a positive scalar to be determined later. Then, from (41) and (42), we have $V_{k+n} - V_k \leq -c\|e_k\|^2 + (c_2 - \gamma(1 - 2f^2\mu^{2\tau}))\sum_{j=0}^{n-1} (2f^2\mu^{2\tau}) \|\xi_{j}\|^2$ where the identity $(1 - 2f^2\mu^{2\tau})\sum_{j=0}^{n-1} (2f^2\mu^{2\tau}) \gamma = 1 - (2f^2\mu^{2\tau})^n$ is used. Choose $\gamma$ such that $2f^2\mu^{2\tau} < 1$ and take any $c^2 > 0$. Let $c_3 = \sum_{j=0}^{n-1} (2f^2\mu^{2\tau})^j$. Then, taking $\gamma = (c_2 + c^2/c_3)/(1 - 2f^2\mu^{2\tau})$, we obtain $V_{k+n} - V_k \leq -c\|e_k\|^2 - c^2\|\xi_{n}\|^2$. From this, the asymptotic stability is obtained with a Lyapunov function $V_k = \sum_{j=0}^{n-1} V_{nk+j}$. This completes the proof. \hfill \Box

5 Numerical Example

Consider a collectively observable sensor network consisting of one hundred estimators. The system matrix of target system is given by $F = \text{diag}\{F_1, F_2\}$ where

$$F_1 = \begin{bmatrix} 0.4 & 0.9 \\ -0.9 & 0.4 \end{bmatrix}$$

and $F_2 = \begin{bmatrix} 0.5 & 0.8 \\ -0.8 & 0.5 \end{bmatrix}$. The output matrices associated with the sensors, $H_i$, are chosen randomly where each element has $-1, 0, 1$ or 0. It is assumed that $Q = 0.05I_2$ and $R = 0.05I_{100}$. The network topology is depicted in Fig. 1 and all the edge weights are 1. The maximum eigenvalue of the Laplacian matrix associated to the network is 14.26, $\alpha_{Li}$ and $\alpha_{ui}$ are chosen as 0.009, and $\epsilon_i = 1, \forall i \in \mathcal{N}$. The initial estimate $\hat{x}_{i,0}$ is randomly selected, and we choose the initial covariance of each estimator as $I_4$.

Fig. 2 presents the estimation errors of the filters with $t^* = 1$. As shown in the figure, the estimation error for each filter is bounded in the steady-state with small error. The error between the local covariance and the steady-state covariance of CKF converges to zero. Fig. 3 shows the effect of $t^*$ on the estimation performance. This result has been obtained through 100 repeated experiments, and in all cases ($t^* = 1$ to $t^* = 7$) the mean squared error (MSE) $\sum_{i=1}^{\mathcal{N}} \|\hat{X}_k - \tilde{x}_{i,k}\|^2$ remains close to zero in the steady-state. The convergence performance improves as $t^*$ increases. In all cases, MSE of covariance converges to zero.

6 Conclusion

In this paper, we have formulated the DKF problem as a consensus optimization problem. It is expected that this new perspective enables us to develop novel DKF algorithms by employing various efficient distributed optimization techniques. As an instance, we have proposed DA-DKF, adopting the dual-ascent method. The unbiased property and the convergence of local covariances to the steady-state covariance of CKF are shown, and this implies that DA-DKF recovers the performance of CKF asymptotically.

Formulating a problem as an optimization problem enables us to deal with constraints very efficiently. This implies that DKF with constraints can be effectively handled under the proposed formulation, which is one of our future research topics. Another important research direction is to consider more general network topology such as directed and time varying graph, etc. In addition, relaxing the assumptions such as the known number of estimators or extending to collective detectability will improve the applicability of the proposed approach.

Appendices

A Proof of Lemma 6

For convenience, we proceed using $\Theta_{i,k,l} := \text{vech}^{-1}(\theta_{i,k,l})$ and $\Upsilon_{i,k,l} := \text{vech}^{-1}(\upsilon_{i,k,l})$ rather than $\theta_{i,k,l}$ and $\upsilon_{i,k,l}$. Let $\Theta_{k,l} = [\Theta_{1,k,l}; \cdots; \Theta_{N,k,l}] \in R^{Nn \times n}$ and $\Upsilon_{k,l} = [\Upsilon_{1,k,l}; \cdots; \Upsilon_{N,k,l}] \in R^{Nn \times n}$. For the case $k \geq 1$
and $l \geq 1$, it follows from (12) that
\[
\Theta_{k, l+1} = N \Omega^\delta - L \Upsilon_{k, l+1} = (I_N - L \bar{A}_e \bar{L}) \Theta_{k, l} \quad \text{(A.1)}
\]
where $\Omega^\delta := [\Omega_{1,1}^\delta \cdots \Omega_{N,1}^\delta]$ with $\Omega_i^\delta := \vech^{-1}(\omega_i^\delta) = H_i^T R_i^{-1} H_i$.

Recall that $\Theta^\dagger = H^T \bar{R}^{-1} H$ from Lemma 5, and let $e_{i, k}^{\Theta^\dagger} := \Theta^\dagger - \Theta_{i, k}$ and $\bar{e}_{i, k}^{\Theta^\dagger} := [e_{i, k, 1}^\Theta \cdots e_{i, k, l}^\Theta]$. Then, from (A.1), we have $\bar{e}_{i, k}^{\Theta^\dagger, l+1} = (I_N - L \bar{A}_e \bar{L}) e_{i, k}^{\Theta^\dagger, l}$.

Applying the coordinate transformation given by $e_{k, l}^{\Theta^\dagger} = \Upsilon^T e_{k, l}^{\Theta^\dagger}$, we have $e_{k, l+1}^{\Theta^\dagger} = (I_N - L \bar{A}_e \bar{L}) e_{k, l}^{\Theta^\dagger}$ equivalent to $e_{k, l+1}^{\Theta^\dagger} = \diag\{I_N, I_{N-1} - L \bar{A}_e \bar{L}, \cdots, I_{N-l} - L \bar{A}_e \bar{L}\} e_{k, l}^{\Theta^\dagger}$. If we choose $\alpha_{u, i}$ satisfying (14), it holds that $\bar{A}_e \bar{L}^T A_e \bar{A}_u \bar{L} < 2I_{N-1}$. Since $\bar{A}_e \bar{L}^T A_e \bar{L}$ is symmetric positive definite, we have $I_{N-l} - \bar{A}_e \bar{L}^T A_e \bar{L}$ is Schur stable.

Meanwhile, the elements of the first $n$ rows of $e_{k, l}^{\Theta^\dagger}$ remain the same for any $l \geq 1$, namely, $1_N e_{k, 1}^{\Theta^\dagger} = 1_N e_{k, l}^{\Theta^\dagger} = \Upsilon^T e_{k, 1}^{\Theta^\dagger} = \Upsilon^T (I_N \Omega^\delta - L \Upsilon_{k, 1})$. Since $\Theta^\dagger = \sum_{i=1}^N H_i^T R_i^{-1} H_i$ (Lemma 5) and $\Upsilon^T \Omega^\delta = \sum_{i=1}^N H_i^T R_i^{-1} H_i$, it follows that the first $n$ rows of $e_{k, l}^{\Theta^\dagger}$ are zero, i.e., $\Upsilon^T e_{k, l}^{\Theta^\dagger} = 0$. Thus, we have $e_{k, l+1}^{\Theta^\dagger} = (I_N - L \bar{A}_e \bar{L}) e_{k, l}^{\Theta^\dagger} = (I_N - L \bar{A}_e \bar{L}) e_{k, 0}^{\Theta^\dagger} = [0_{n \times i}; (I_N - L \bar{A}_e \bar{L}) e_{k, 0}^{\Theta^\dagger}]$.

We prove the second part, the existence of $k$. Suppose $k \geq 1$. From (A.1), one has $\Theta_{k, l+1} = (I_N \bar{A}_e \bar{L}) e_{k, l}^{\Theta^\dagger}$, which proves (15).

\[
\Theta_{1, 1} = N \Omega^\delta - L \bar{A}_e \bar{L} \Theta_{1, 0}.
\]

Since $\Omega^\delta = \Theta_{1, 0}$ by Algorithm 1, we have $\Theta_{1, 0} = (N \bar{A}_e \bar{L} \bar{A}_e \bar{L}) \Theta_{1, 0}$. Recalling that $-I < (I - L \bar{A}_e \bar{L}) < I$, we have $-N \bar{A}_e \bar{L} \bar{A}_e \bar{L} < N \bar{A}_e \bar{L} \bar{A}_e \bar{L}$. In addition, from the fact that $\Theta_{1, 0} < 0$, $\forall i \in N$, we obtain $\Theta_{1, 1} < N \Omega^\delta$. Since $I - L \bar{A}_e \bar{L}$ is Schur stable, there exists $0 < \rho < 1$ such that
\[
-N \rho^{l+1} \Theta^\dagger \leq e_{k, l+1}^{\Theta^\dagger} \leq N \rho^{l+1} \Theta^\dagger.
\]

Hence, for any $0 < \kappa < 1$, there exists $\bar{k}$ such that $-\kappa \Theta^\dagger \leq e_{k, l+1}^{\Theta^\dagger} \leq \kappa \Theta^\dagger$, $\forall k \geq \bar{k}$, and by recalling that $e_{k, l+1}^{\Theta^\dagger} = \Theta^\dagger - \Theta_{k, l+1}$, we have (15). This completes the proof.

**B Proof of Lemma 8**

Let $\kappa$ be given such that $0 < \kappa < 1$. By Lemma 6, there exists $\bar{k}$ such that the relation (15) holds true.

1) Existence of $P$: The proof is done by exploiting the monotonicity of the algebraic Riccati equation [10]. We take a symmetric positive definite matrix $P_k^\dagger$ such that $P_{i, k} \leq P_k^\dagger$ for all $i \in N$ and all $k \leq \bar{k}$. Then,
\[
P_{i, k+1|k} \leq P_{k+1|k}^\dagger, \quad \forall i \in N, \forall k \leq \bar{k} \quad \text{(B.1)}
\]
where $P_{k+1|k} = FP_k R_k^T + Q$ and $P_{k+1|k}^\dagger = FP_k^\dagger R_k^T + Q$.

Meanwhile, let $P$ be the unique positive definite solution to the algebraic Riccati equation given by
\[
\mathcal{T} = F \left( P^{-1} \right) \left( 1 - (\kappa) \Theta^\dagger \right)^{-1} \left( F + \bar{q} I \right) = \mathcal{T}.
\]

where $\bar{q}$ is a positive scalar such that $P_{k+1|k}^\dagger \leq \bar{q} I$. The existence of $P$ is guaranteed by the controllability of $(F, \sqrt{\bar{q}} I)$ and the observability of $(F, \sqrt{1 - \kappa} H)$ [10]. From the manner in which $\bar{q}$ is chosen, it follows from (B.2) that $P_{k+1|k}^\dagger \leq \bar{q} I \leq \mathcal{T}$. Then, using the relations (15), (B.1), (B.2), $P_{k+1|k} \leq \mathcal{T}$, and $Q \leq \bar{q} I$, we derive
\[
P_{i, k+2|k+1} = F \left( P_{i, k+1|k}^\dagger + \Theta_{i, k+1, i'} \right)^{-1} F + Q
\]
\[
\leq F \left( P_{k+1|k}^\dagger + \Theta_{i, k+1, i'} \right)^{-1} F + Q
\]
\[
\leq F \left( \mathcal{T}^{-1} \right) \left( 1 - (\kappa) \Theta^\dagger \right)^{-1} \left( F + \bar{q} I \right) = \mathcal{T}.
\]

2) Existence of $P$: From the prediction rule of covariance given by $P_{i, k+1|k} = FP_{i, k|k} R_k^T + Q$, it follows that
\[
Q \leq P_{i, k+1|k}.
\]

Moreover, since the prior covariance $P_{i, k}$ is obtained as $P_{i, k} = \left( P_{i, k+1|k} + \Theta_{i, k+1, i'} \right)^{-1}$, we have from (15) and (B.3) that $P_{i, k} \geq \left( \left( Q^{-1} + (1 + \kappa) \Theta \right)^{-1} \right) \left( Q^{-1} + (1 + \kappa) \Theta \right)^{-1} \leq P_{i, k+1|k} \leq \mathcal{T}$.

**C Proof of Lemma 10**

We first derive the dynamics of $e_{k, l}^\lambda := \lambda_k^* - \lambda_{k, l}$. From (11), we have
\[
e_{k, l+1}^\lambda = \lambda_k^* - \lambda_{k, l} - \bar{A}_k K_k^\text{dual}^T L e_{k, l} \]
\[
e_{k, l}^\lambda - \bar{A}_k K_k^\text{dual}^T L e_{k, l} \])
\]
α
calling

\[ L_{NP} \]

I

k

\[ \| \text{nonnegative real numbers.} \text{ This fact and the property } \| U \]

\[ U \]

of 

\[ \| P \]

k

\[ \| \text{symmetric and positive semidefinite, it fol-} \]

\[ \| \lambda, i \]

\[ \lambda, i \]

\[ \| \text{are symmetric and positive semidefinite, it fol-} \]

\[ \| \xi \]

\[ \| \text{at zero, it follows that there exists a nonsingu-} \]

\[ \| \text{lim}_{l \to \infty} \| e_{k,l}^\lambda \| = 0. \]

\[ \| e_{k,l}^\xi \| = -K_{k}^{cons} L_{e_{k,l}^\lambda}, \]

\[ \| e_{k,l}^\xi \| \text{ and by applying (C.3) we obtain that } \lim_{l \to \infty} \| e_{k,l}^\xi \| = 0. \]

\[ \| e_{k,l}^\xi \| = \| (1 - K_{k}^{cons} \hat{\Lambda}_{k} L_{k}^{dual} \hat{\Lambda}_{k}^{dual}) \| e_{k,l}^\xi \| = \| \hat{\xi} e_{k,l}^\xi \| \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ reads as } \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ and re-} \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ as } \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ first term in parentheses in (D.1) can be rearranged as follows.} \]

\[ (K_{k}^{cons})^{-1} \| N E \{ x_{k+1} \} \]

\[ = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N F E \{ x_{k} \}. \]

\[ \| = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N E \{ x_{k} \} \]

\[ \| e_{k,l}^\xi \| \text{ reads as } \]

\[ \| e_{k,l}^\xi \| \text{ first term in parentheses in (D.1) can be rearranged as follows.} \]

\[ (K_{k}^{cons})^{-1} \| N E \{ x_{k+1} \} \]

\[ = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N F E \{ x_{k} \}. \]

\[ \| = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N E \{ x_{k} \} \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ first term in parentheses in (D.1) can be rearranged as follows.} \]

\[ (K_{k}^{cons})^{-1} \| N E \{ x_{k+1} \} \]

\[ = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N F E \{ x_{k} \}. \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ first term in parentheses in (D.1) can be rearranged as follows.} \]

\[ (K_{k}^{cons})^{-1} \| N E \{ x_{k+1} \} \]

\[ = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N F E \{ x_{k} \}. \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \]

\[ \| \hat{\xi} e_{k,l}^\xi \| \text{ first term in parentheses in (D.1) can be rearranged as follows.} \]

\[ (K_{k}^{cons})^{-1} \| N E \{ x_{k+1} \} \]

\[ = H^{T} R^{-1} H \| N E \{ x_{k+1} \} + \frac{1}{N} P_{k+1}^{-1} \| N F E \{ x_{k} \}. \]
Subtracting (D.3) from (D.2) and applying the identities 
\[ \dot{x}_{k+1|k} = F_{k|k} \xi_{k|k} \] and 
\[ \mathbb{I}_N F \{ r_k \} = F \{ \xi_{k|k} \} + \mathbb{I}_N F_{k|k} = \mathbb{I}_N F_{k|k} + F \xi_{k|k} + \mathbb{I}_N F_{k|k} \mathbb{I}_N \mathbb{E} \{ x_{k+1|k} \} - \mathbb{E} \{ \mathbb{H}^T \mathbb{S}_{k+1|k} z_{k+1} \} = \frac{1}{N} P_{k+1|k}^{(1)} (1_N F_{k|k} + F \xi_{k|k}) . \] Then, (D.1) becomes 
\[ e_{k+1|k}^1 = E_{k+1|k}^1 e_{k|k}^1 + \mathbb{H}_{k+1|k}^{(1)} \frac{1}{N} P_{k+1|k}^{(1)} F e_{k|k}^1 . \] (D.4)

We now derive the dynamics of \( e_{k+1|k}^1 \). It holds from (C.5) that \( e_{k+1|k}^1 = (\bar{\xi}_{k+1|k,0})^T e_{k+1|k}^1 \). Recalling that \( \xi_k^1 = \mathbb{I}_N H_{k+1|k}^{(1)} \mathbb{H}^T \mathbb{S}_{k} \mathbb{z}_k \), one can obtain 
\[ e_{k+1|k}^1 = \mathbb{E} \{ \xi_{k+1|k,0} - \xi_{k|k,0} \} = \mathbb{I}_N H_{k+1|k}^{(1)} \mathbb{H}^T \mathbb{S}_{k} \mathbb{z}_k - \mathbb{E} \{ \xi_{k|k,0} \} . \] (D.5)

From the fact that \( \dot{x}_{k+1|k} = \bar{\xi}_{k+1|k,0} + F \xi_{k|k,0} \) and the relation (D.3), we have 
\[ \mathbb{E} \{ \mathbb{H}^T \mathbb{S}_{k+1|k} z_{k+1} \} = \mathbb{H}^T \mathbb{R} \mathbb{H} \mathbb{F} \{ x_k \} + \frac{1}{N} P_{k+1|k}^{(1)} F e_{k|k} . \] Subtracting (D.3) from (D.2) and applying the identities
\[ \dot{\xi}_{k+1|k} = (K_{k+1}^{(1)})^{-1} \xi_{k+1|k} \] and substituting this identity to (D.5) gives 
\[ e_{k+1|k}^1 = \mathbb{E} \{ \xi_{k|k,0} \} - \mathbb{I}_N H_{k+1|k}^{(1)} \mathbb{H}^T \mathbb{S}_{k} \mathbb{z}_k - \mathbb{E} \{ \xi_{k|k,0} \} . \]

Thus, if \( a_{\alpha,i} \) is chosen such that \( 0 < a_{\alpha,i} < 2/\alpha^2 \), then the matrix \( E_{k+1|k} (\bar{\xi}_{k+1|k})^T e_{k+1|k} \) is Schur stable for \( k \geq k^* \) by Lemma 10, which completes the proof.
[8] B. M. Bell and F. W. Cathey. The iterated Kalman filter update as a Gauss-Newton method. IEEE Trans. on Automatic Control, 38(2):294–297, 1993.

[9] T. M. Berg and H. Durrant-Whyte. General decentralized Kalman filters. In Proceedings of American Control Conference, volume 2, pages 2273–2274, 1994.

[10] R. Bitmead, M. Gevers, I. Petersen, and J. Kaye. Monotonicity and stabilizability-properties of solutions of the Riccati difference equation: Propositions, lemmas, theorems, fallacious conjectures and counterexamples. Systems & Control Letters, 5(5):309–315, 1985.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine learning, 3(1):1–122, 2011.

[12] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering based on consensus strategies. IEEE Journal on Selected Areas in communications, 26(4):622–633, 2008.

[13] F. S. Cattivelli and A. H. Sayed. Diffusion strategies for distributed Kalman filtering and smoothing. IEEE Trans. on Automatic Control, 55(9):2069–2084, 2010.

[14] Z. Chen et al. Bayesian filtering: From Kalman filters to particle filters, and beyond. Statistics, 182(1):1–69, 2003.

[15] S. Das and J. M. Moura. Distributed Kalman filtering with dynamic observations consensus. IEEE Trans. on Signal Processing, 63(17):4458–4473, 2015.

[16] Á. R. del Nozal, P. Millán, L. Orihuela, A. Seuret, and L. Zaccarian. Distributed estimation based on multi-hop subspace decomposition. Automatica, 99:213–230, 2019.

[17] R. Deshmukh, C. Kwon, and I. Hwang. Optimal discrete-time Kalman consensus filter. In Proceedings of American Control Conference, pages 5801–5806, 2017.

[18] P. Duan, Z. Duan, G. Chen, and L. Shi. Distributed Kalman filter with state equality constraints: A localized least-squares approach. Automatica, 117:109607, 2020.

[19] M. Farina and R. Carli. Partition-based distributed Kalman filter with plug and play features. IEEE Trans. on Control of Network Systems, 5(1):560–570, 2016.

[20] X. He, C. Hu, Y. Hong, L. Shi, and H.-T. Fang. Distributed Kalman filters with state equality constraints: Time-based and event-triggered communications. IEEE Trans. on Automatic Control, 65(1):28–43, 2019.

[21] X. He, W. Xue, and H. Fang. Consistent distributed state estimation with global observability over sensor network. Automatica, 92:162–172, 2018.

[22] A. H. Jazwinski. Stochastic processes and filtering theory. Courier Corporation, 2007.

[23] E. W. Kamen and J. K. Su. Introduction to optimal estimation. Springer Science & Business Media, 1999.

[24] M. Kamgarpour and C. Tomlin. Convergence properties of a decentralized Kalman filter. In Proceedings of Conference on Decision and Control, pages 3205–3210, 2008.

[25] J. Kim, H. Shim, and J. Wu. On distributed optimal Kalman-Bucy filtering by averaging dynamics of heterogeneous agents. In Proceedings of Conference on Decision and Control, pages 6309–6314, 2016.

[26] W. Li, Y. Jia, and J. Du. Distributed Kalman consensus filter with intermittent observations. Journal of the Franklin Institute, 352(9):3764–3781, 2015.

[27] W. Li, Z. Wang, D. W. Ho, and G. Wei. On boundedness of error covariances for Kalman consensus filtering problems. IEEE Trans. on Automatic Control, 65(6):2654–2661, 2020.

[28] W. Liu, P. Shi, and S. Wang. Distributed Kalman filtering through trace proximity. IEEE Trans. on Automatic Control, 2022.

[29] M. S. Mahmoud and H. M. Khalid. Distributed Kalman filtering: a bibliographic review. IET Control Theory & Applications, 7(4):483–501, 2013.

[30] D. Marelli, T. Sui, and M. Fu. Distributed Kalman estimation with decoupled local filters. Automatica, 130:109724, 2021.

[31] J. M. Mendel. Lessons in estimation theory for signal processing, communications, and control. Pearson Education, 1995.

[32] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In Proceedings of Conference on Decision and Control, pages 8179–8184, 2005.

[33] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In Proceedings of Conference on Decision and Control, pages 5492–5498, 2007.

[34] R. Olfati-Saber. Kalman-consensus filter: Optimality, stability, and performance. In Proceedings of Conference on Decision and Control, pages 7036–7042, 2009.

[35] J. Qian, P. Duan, Z. Duan, G. Chen, and L. Shi. Consensus-based distributed filtering with fusion step analysis. Automatica, 142:110408, 2022.

[36] W. Ren and U. M. Al-Saggaf. Distributed Kalman–Bucy filter with embedded dynamic averaging algorithm. IEEE Systems Journal, 12(2):1722–1730, 2017.

[37] K. Ryu and J. Back. Distributed Kalman-filtering: Distributed optimization viewpoint. In Proceedings of Conference on Decision and Control, pages 2640–2645, 2019.

[38] S. Thrun, W. Burgard, and D. Fox. Probabilistic robotics. MIT press, 2005.

[39] G. Wei, W. Li, D. Ding, and Y. Liu. Stability analysis of covariance intersection-based Kalman consensus filtering for time-varying systems. IEEE Trans. on Systems, Man, and Cybernetics: Systems, 50(11):4611–4622, 2020.

[40] J. Yan, X. Yang, Y. Mo, and K. You. A distributed implementation of steady-state Kalman filter. IEEE Trans. on Automatic Control, early access, 2022.

[41] S. Yi and M. Zorzi. Robust Kalman filtering under model uncertainty: the case of degenerate densities. IEEE Trans. on Automatic Control, 2021.

[42] D. Yu, Y. Xia, L. Li, Z. Xing, and C. Zhu. Distributed covariance intersection fusion estimation with delayed measurements and unknown inputs. IEEE Trans. on Systems, Man, and Cybernetics: Systems, 51(8):5165–5173, 2019.

[43] M. Zorzi. Distributed Kalman filtering under model uncertainty. IEEE Trans. on Control of Network Systems, 7(2):990–1001, 2019.