Communication-Efficient Distributed Kalman Filtering Using ADMM

Muhammad Iqbal, Kundan Kumar, and Simo Särkkä, *Senior Member, IEEE*

***Abstract*—This article addresses the problem of optimal linear filtering in a network of local estimators, commonly referred to as distributed Kalman filtering (DKF). The DKF problem is formulated within a distributed optimization framework, where coupling constraints require the exchange of local state and covariance updates between neighboring nodes to achieve consensus. To address these constraints, the problem is transformed into an unconstrained optimization form using the augmented Lagrangian method. The distributed alternating direction method of multipliers is then applied to derive update steps that achieve the desired performance while exchanging only the primal variables. Notably, the proposed method enhances communication efficiency by eliminating the need for dual variable exchange. We show that the design parameters depend on the maximum eigenvalue of the network's Laplacian matrix, yielding a significantly tighter bound compared to existing results. A rigorous convergence analysis is provided, proving that the state estimates converge to the true state and that the covariance matrices across all local estimators converge to a globally optimal solution. Numerical results are presented to validate the efficacy of the proposed approach.**

***Index Terms*—Alternating direction method of multipliers (ADMM), distributed filtering, Kalman filtering.**

I. Introduction

The Kalman filter, introduced by Kalman [1], remains one of the most significant contributions to engineering, with widespread applications in science, medicine [2], economics [3], and various engineering domains [4]. Traditionally, a single estimator is employed to infer the state of a dynamical system. However, large-scale systems—such as bridges [5], smart grids [6], forest fire monitoring [7], phased-array systems [8], and complex cyber-physical systems [9]—demand a network of local estimators for accurate state estimation. In such contexts, distributed algorithms provide scalable, modular, and robust solutions. The complexity of state estimation is further heightened by the presence of process and measurement noise, particularly when heterogeneous sensor models are involved [10].

Distributed Kalman filtering (DKF) is a solution for estimating the state of a linear dynamical system, along with its associated uncertainty (covariance matrix), observed by a sensor network in the presence of process and measurement noise [11], [12], [13], [14], [15], [16], [17], [18], [19]. In DKF, the prediction step is performed using only local information, similar to a single estimator, while the update step incorporates both local information and information exchanged from neighboring nodes to achieve consensus [10], [11], [17]. Numerous studies have developed DKF algorithms for state estimation under uncertainty [11], [17], [20], [21], [22]. It is well-known that for linear Gaussian models, one can present optimal filtering problem as a maximum a posterior (MAP) problem, using Bayesian framework, to design Kalman filter [4]. Similarly, for linear Gaussian models, distributed state estimation using a sensors network, can be presented as distributed optimization problem using MAP [20], [23], [24] in a distributed Bayesian framework [25].

In the seminal work presenting the DKF problem as a distributed optimization problem [20], [23], the authors employed the dual-ascent method to estimate the state of a dynamical system and its covariance matrix in a fully distributed manner. An alternating direction method of multipliers (ADMM)-based approach is used to derive the update rule for DKF [26], [27]. In [26], the number of dual variables equals the number of edges, leading to increased computational complexity. In addition, the consensus process slows down due to the selection of a smaller step size in updating the information rate matrix, where the step size is inversely proportional to the degree of a node. Furthermore, Wang and Dekorsy [27] assumed a complete graph and requires the exchange of measurement matrices and measurement noise covariances, which is a strong assumption. In addition, neither Wang et al. [26] nor Wang and Dekorsy [27] provided a convergence analysis. In [20] and [23], both primal and dual variables are exchanged to achieve consensus. The design parameter in [23] is upper bounded by the inverse of the square of the maximum eigenvalue of the Laplacian matrix multiplied by the norm of the information rate matrix for each edge, leading to a small design parameter that slows the consensus process. Although Ryu and Back [23] improved this upper bound, it is still related to the square of the maximum eigenvalue of the Laplacian matrix.

The contributions of this article are as follows.

1) We introduce a variant of distributed ADMM to solve the DKF problem that does not require the exchange of dual variables in the update step, unlike the methods in [20] and [23], thus reducing the communication burden.
2) We derive upper bounds on the design parameters for updating the posterior state estimate and posterior covariance estimate, which decay linearly with $\lambda_{\max}(\mathcal{L})$, where $\lambda_{\max}(\mathcal{L})$ denotes the largest eigenvalue of the graph Laplacian. In contrast, the design parameters in [20] decay quadratically with $\lambda_{\max}(\mathcal{L})$. Consequently, in networks where $\lambda_{\max}(\mathcal{L})$ increases with size, our algorithm permits larger values of design parameters to expedite the consensus process.
3) In the proposed method, the update of information rate matrix does not require sub-iterations.
4) We show that the local estimators at each node provide unbiased estimates as time approaches infinity.

In addition, the distributed ADMM algorithm in this paper differs from [26], [27] because the augmented Lagrangian is designed in a unique way that produces an update step, which reduces the number of dual variables to the number of nodes. In contrast, Wang et al. [26] introduced dual variables equal to the number of edges, which increases computational complexity. Furthermore, the consensus process slows down due to the algorithm's structure, particularly in dense networks.

Received 28 June 2025; accepted 21 September 2025. Date of publication 29 September 2025; date of current version 2 March 2026. This work was supported in part by the Finnish Center for Artificial Intelligence (FCAI) and in part by the Research Council of Finland (RCF). Recommended by Associate Editor A. Tanwani. *(Corresponding author: Muhammad Iqbal.)*

The authors are with the Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland (e-mail: muhammad.iqbal@aalto.fi, iqbal.salarzai@gmail.com; kundan.kumar@aalto.fi; simo.sarkka@aalto.fi).

Digital Object Identifier 10.1109/TAC.2025.3615237

In addition, we provide convergence analysis, which is not given in [26] and [27].

II. Problem Formulation

Consider a network of $N \geq 2$ sensor nodes measuring the output of a discrete-time dynamical system

$$x_{t+1} = Fx_t + w_t$$
$$y_t = Hx_t + v_t \quad (1)$$

where $x_t \in \mathbb{R}^n$ is the state vector at time $t \in \{0, 1, \ldots\}$, and $y_t = [y_{1,t}^\top, \ldots, y_{N,t}^\top]^\top \in \mathbb{R}^m$ is the aggregated measurement vector of all sensors. Each sensor $i \in \{1, 2, \ldots, N\}$ provides measurements $y_{i,t} \in \mathbb{R}^{m_i}$, with F as the state-transition matrix and $H = [H_1^\top, \ldots, H_N^\top]^\top$ as the measurement matrix. The process noise w_t and measurement noise $v_t = [v_{1,t}^\top, \ldots, v_{N,t}^\top]^\top$ are zero-mean, and white Gaussian, satisfying the following properties:

$$\mathbb{E}\{w_t w_l^\top\} = Q\delta_{tl}, \quad \mathbb{E}\{v_t v_l^\top\} = \bar{R}\delta_{tl}$$
$$\mathbb{E}\{w_t v_{i,l}^\top\} = 0, \quad i = 1, \ldots, N$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator, Q and $\bar{R} = \text{diag}\{R_1, \ldots, R_N\}$ are positive definite matrices, and δ_{tl} is the Kronecker delta. The initial state $x_o \sim \mathcal{N}(\mathbb{E}\{x_o\}, P_o)$ is uncorrelated with w_t and v_t, where $\mathcal{N}$ represents the Gaussian distribution.

The sensor network is represented as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of nodes, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ defines the edges. The graph's adjacency matrix A satisfies $a_{ij} = 1$ if node i is receiving information from node j, and $a_{ij} = 0$ otherwise. For undirected graphs, $a_{ij} = a_{ji}$. The set of neighbors of node i is denoted as $\mathcal{N}_i = \{j \mid a_{ij} = 1, \ j \in \mathcal{V}\}$.

The objective is to design a distributed algorithm that enables each node to estimate the state x_t of a dynamical system using its local measurements and information from its neighbors. Specifically, we propose a distributed ADMM-based approach for the correction step in DKF. The algorithm ensures that the local estimators at each node satisfy

$$\lim_{t \to \infty} \mathbb{E}\{x_t - \xi_{i,t}\} = 0, \quad \lim_{t \to \infty} \|P^* - P_{i,t|t-1}\|_F = 0$$

where $\xi_{i,t}$ is the posterior estimate of the state at node i, $\|\cdot\|_F$ represents the Frobenius norm, $P_{i,t|t-1}$ is the local prior covariance matrix at node i, and P^* is the unique positive definite solution to the discrete-time algebraic Riccati equation

$$P^* = FP^*F^\top - FP^*H^\top(HP^*H^\top + \bar{R})^{-1}HP^*F^\top + Q$$

in the case of centralized Kalman filter [20]. The posterior covariance matrix is defined as $P_{i,t} = \text{Cov}(x_t \mid y_{i,t})$.

III. Distributed Kalman Filtering

Observing the state of a complex dynamical system using a single sensor is often impractical. The centralized filtering methods can be used to estimate the state of a complex dynamical system but increases the communication burden and fragility. Instead, multiple sensors can be deployed, without having an anchor node, to observe different parts of the state vector of the physical process. To address this, we relax the assumption that the entire system is observable from a single sensor. In this framework, multiple local estimators collaborate, sharing local information to estimate the entire state vector, while each observes only a partial state. To ensure the sensor network can collectively estimate the state of the system, we impose the following assumptions.

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

Assumption 2: The network $\mathcal{G}$ is static and connected.

Assumption 1 indicates that (F, H_i) is not necessarily observable for any individual sensor. These assumptions are standard and can be found in [20] and the references therein.

In a distributed estimation setting, the ith estimator has access to its own measurement $y_{i,t}$ at time t, the measurement matrix H_i, and the measurement covariance matrix R_i. The matrices F and Q are assumed to be known to all estimators.

In DKF, the prediction step is identical to that of centralized Kalman filtering. The local prediction step for each estimator is given by

$$\hat{x}_{i,t|t-1} = F\hat{x}_{i,t-1|t-1}$$
$$P_{i,t|t-1} = FP_{i,t-1|t-1}F^\top + Q \quad (2)$$

where $\hat{x}_{i,t|t-1}$ denotes the predicted mean at time t based on the posterior mean $\hat{x}_{i,t-1|t-1}$ at time $t-1$. Similarly, $P_{i,t|t-1}$ is the predicted error covariance matrix of agent i, computed using the posterior error covariance $P_{i,t-1|t-1}$ at time $t-1$.

In the correction step, each estimator updates its state estimate using local information and information exchanged with its neighbors. For a linear Gaussian system, this step corresponds to solving the maximum a posteriori (MAP) estimation problem in a distributed manner [20]. To develop the solution, we first formulate the MAP problem for the ith node by ignoring coupling constraints. This serves as a basis for understanding the local estimation problem, which is later extended to incorporate the coupling constraints for the complete distributed formulation. The MAP problem for the ith node, without coupling constraints, is formulated as

$$\hat{x}_{i,t|t} = \arg\max_{x_t} p(y_{i,t} \mid x_t) p(x_t \mid y_{i,1:t-1}) \quad (3)$$

where $p(y_{i,t} \mid x_t) = \mathcal{N}(y_{i,t} \mid H_i x_t, R_i)$ denotes the likelihood function with conditional mean $H_i x_t$ and covariance R_i, and $p(x_t \mid y_{i,1:t-1}) = \mathcal{N}(x_t \mid \hat{x}_{i,t|t-1}, P_{i,t|t-1})$ is the prior distribution of the state x_t of the ith estimator, given its local measurements up to time $t-1$.

Using the monotonicity of the logarithmic function, (3) is equivalent to

$$\hat{x}_{i,t|t} = \arg\max_{x_t} \ln\left(p(y_{i,t} \mid x_t) p(x_t \mid y_{i,1:t-1})\right). \quad (4)$$

The terms in (4) are expanded as follows:

$$\ln p(y_{i,t} \mid x_t) = -\frac{1}{2}(y_{i,t} - H_i x_t)^\top R_i^{-1}(y_{i,t} - H_i x_t)$$
$$- \frac{1}{2}\ln((2\pi)^{m_i} \det(R_i))$$
$$\ln p(x_t \mid y_{i,1:t-1}) = -\frac{1}{2}(\hat{x}_{i,t|t-1} - x_t)^\top P_{i,t|t-1}^{-1}(\hat{x}_{i,t|t-1} - x_t)$$
$$- \frac{1}{2}\ln((2\pi)^n \det(P_{i,t|t-1})).$$

Substituting $\ln p(y_{i,t} \mid x_t)$ and $\ln p(x_t \mid y_{i,1:t-1})$ in (4), the optimization problem in (4) takes the following form:

$$\hat{x}_{i,t|t} = \arg\min_{\xi_{i,t}} f_{i,t}(\xi_{i,t}) \quad (5)$$

where $\xi_{i,t} \in \mathbb{R}^n$ and

$$f_{i,t}(\xi_{i,t}) = \frac{1}{2}(\hat{x}_{i,t|t-1} - \xi_{i,t})^\top P_{i,t|t-1}^{-1}(\hat{x}_{i,t|t-1} - \xi_{i,t})$$
$$+ \frac{1}{2}(y_{i,t} - H_i\xi_{i,t})^\top R_i^{-1}(y_{i,t} - H_i\xi_{i,t}).$$

Sensor networks are typically heterogeneous, making (5) suboptimal for the entire network since the decision variable is local. To achieve network-wide optimality, we reformulate (5) as

$$\underset{\xi_t}{\text{minimize}} \quad \sum_{i=1}^{N} f_{i,t}(\xi_t) \tag{6}$$

where $\xi_t = [\xi_{1,t}^\top, \xi_{2,t}^\top, \ldots, \xi_{N,t}^\top]^\top$. However, since Estimator i does not have access to the entire ξ_t, we adopt a distributed optimization approach based on [28, Lemma 3.1] and [20]

$$\begin{aligned} \underset{\xi_{1,t},\ldots,\xi_{N,t}}{\text{minimize}} \quad & \sum_{i=1}^{N} f_{i,t}(\xi_{i,t}) \\ \text{subject to} \quad & \mathbb{L}\xi_t = 0_{Nn} \end{aligned} \tag{P1}$$

where $\mathbb{L} = \mathcal{L} \otimes I_n$, $\mathcal{L} = D - A$, with $\mathcal{L}$ denoting the graph Laplacian, and $D \in \mathbb{R}^{N\times N}$ is the degree matrix of $\mathcal{G}$. The constraint $\mathbb{L}\xi_t = 0_{Nn}$ ensures $\xi_{1,t} = \xi_{2,t} = \cdots = \xi_{N,t}$, as the kernel of $\mathcal{L}$ is spanned by 1_N. To express $f_{i,t}(\xi_{i,t})$ in compact quadratic form, define $\mathbf{z}_{i,t} = [y_{i,t}; \hat{x}_{i,t|t-1}]$, $\mathbf{H}_i = [H_i; I_n]$, and $\mathbf{S}_{i,t} = \text{diag}(R_i, NP_{i,t|t-1})$, where N is introduced to average out the summation of all $P_{i,t|t-1}$. Then

$$f_{i,t}(\xi_{i,t}) = \frac{1}{2}(\mathbf{z}_{i,t} - \mathbf{H}_i\xi_{i,t})^\top \mathbf{S}_{i,t}^{-1}(\mathbf{z}_{i,t} - \mathbf{H}_i\xi_{i,t}).$$

Finally, define $\mathbf{z}_t = [\mathbf{z}_{1,t}^\top, \ldots, \mathbf{z}_{N,t}^\top]^\top$, $\bar{\mathbf{H}} = \text{diag}(\mathbf{H}_1, \ldots, \mathbf{H}_N)$, and $\bar{\mathbf{S}}_t = \text{diag}(\mathbf{S}_{1,t}, \ldots, \mathbf{S}_{N,t})$. The matrix $\mathcal{H}_t = \mathbb{1}_N^\top \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1} \bar{\mathbf{H}} \mathbb{1}_N$, with $\mathbb{1}_N = 1_N \otimes I_n$, is symmetric positive definite.

Next, we present the distributed correction step for the covariance update using a distributed optimization framework. Let $\Omega_{t|t} = P_{t|t}^{-1}$ and $P_{t|t}^{-1}\hat{x}_{t|t}$ represent the information matrix and the information vector, respectively. We also define $\Omega_{t|t-1} = P_{t|t-1}^{-1}$. For a centralized estimator, the information matrix prediction and correction steps are given as

$$\begin{aligned} \Omega_{t|t-1} &= (F\Omega_{t-1|t-1}^{-1}F^\top + Q)^{-1} \\ \Omega_{t|t} &= \Omega_{t|t-1} + H^\top \bar{R}^{-1} H. \end{aligned} \tag{7}$$

The convergence of $\Omega_{t|t-1}$ to P^{*-1} is established in [29, Lemma 9.5.1 and Prob. 9.17]. In a distributed setting, the convergence of $\Omega_{t|t-1}$ to P^{*-1} remains valid if the global information rate matrix $H^\top \bar{R}^{-1} H$ is available to each estimator. To obtain $H^\top \bar{R}^{-1} H$ at each estimator, we solve the following consensus optimization problem

$$\begin{aligned} \underset{\theta_1,\ldots,\theta_N}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{N} \|N\omega_i^\delta - \theta_i\|^2 \\ \text{subject to} \quad & (\mathcal{L} \otimes I_{n_{\text{cov}}})\theta = 0_{Nn_{\text{cov}}} \end{aligned} \tag{P2}$$

where $\|\cdot\|$ denotes the Euclidean norm, $\omega_i^\delta = \text{vech}(H_i^\top R_i^{-1} H_i) \in \mathbb{R}^{n_{\text{cov}}}$, $\text{vech}(\cdot)$ is the half-vectorization of the symmetric matrix $H_i^\top R_i^{-1} H_i$, $n_{\text{cov}} = \frac{n(n+1)}{2}$, and $\theta_i \in \mathbb{R}^{n_{\text{cov}}}$ is the decision variable of ith estimator.

IV. DKF Using ADMM

In this section, we derive the distributed Kalman filter algorithm by solving an optimization problem using distributed ADMM. To achieve this, we solve (P1) by considering the following augmented Lagrangian:

$$L_{\text{est},t}(\xi_t, \lambda_t) = \sum_{i=1}^{N} f_{i,t}(\xi_{i,t}) + \lambda_t^\top \sqrt{\mathbb{L}}\xi_t + \frac{\mu}{2}\|\sqrt{\mathbb{L}}\xi_t\|^2. \tag{8}$$

Instead of $\mathbb{L}$, we use $\sqrt{\mathbb{L}}$ in (8) as the null space of both $\mathbb{L}$ and $\sqrt{\mathbb{L}}$ are the same; essentially the problem remains the same. The reason of this reformulation is to derive update laws that are both communication and computation efficient. In the framework of distributed optimization, the same computation and communication efficient update laws can be derived by linearizing the cost function [30]. An alternative way is to derive distributed ADMM-based update laws so that each node i maintains $\mathcal{O}(|\mathcal{V}|)$ variables [31, Algorithm 1], instead of $\mathcal{O}(|\mathcal{E}|)$ variables, which is the case in the standard distributed ADMM. This also improves the computation and communication complexity, however, in [31] dual variable needs to be exchanged, whereas in [30], dual variables are not exchanged. The regularization term in (8) helps to penalize the consensus error, and allow us to directly apply ADMM.

Taking the gradient of (8) with respect to λ_t and ξ_t, we obtain

$$\begin{aligned} \nabla_{\lambda_t} L_{\text{est},t} &= \sqrt{\mathbb{L}}\xi_t \\ \nabla_{\xi_t} L_{\text{est},t} &= \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1} \bar{\mathbf{H}}\xi_t - \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t + \sqrt{\mathbb{L}}\lambda_t + \mu\mathbb{L}\xi_t. \end{aligned} \tag{9}$$

Using (9), the update step for $\lambda_{t,l}$ and $\xi_{t,l}$ can be written as

$$\begin{aligned} \lambda_{t,l+1} &= \lambda_{t,l} + \alpha_\lambda \sqrt{\mathbb{L}}\xi_{t,l} \\ \xi_{t,l+1} &= K_t\left(\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t - \sqrt{\mathbb{L}}\lambda_{t,l+1} - \mu\mathbb{L}\xi_{t,l}\right). \end{aligned} \tag{10}$$

Due to the structure of $\sqrt{\mathbb{L}}$, the update law in (10) cannot be implemented in a fully distributed manner. To enable distributed implementation, we define an auxiliary variable

$$\tilde{\lambda}_{t,l} = \sqrt{\mathbb{L}}\lambda_{t,l}. \tag{11}$$

Premultiplying the dual variable update in (10) by $\sqrt{\mathbb{L}}$, and then replacing α_λ with $\alpha_\lambda K_t^{-1}$, and μ with μK_t^{-1}, where $K_t = (\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1} \bar{\mathbf{H}})^{-1}$, the DKF algorithm for state estimation using distributed ADMM becomes

$$\begin{aligned} \tilde{\lambda}_{t,l+1} &= \tilde{\lambda}_{t,l} + \alpha_\lambda K_t^{-1}\mathbb{L}\xi_{t,l} \\ \xi_{t,l+1} &= K_t\left(\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t - \tilde{\lambda}_{t,l+1} - \mu K_t^{-1}\mathbb{L}\xi_{t,l}\right). \end{aligned} \tag{12}$$

The integral feedback term $\tilde{\lambda}_{t,l+1}$ in (12) reduces the steady-state error in the consensus process. One can see in (12) that only primal variable needs to be exchanged, unlike [20], where both primal and dual variables need to be exchanged.

Similarly, for the covariance matrix, we solve the distributed optimization problem given in (P2). To this end, we propose the following augmented Lagrangian:

$$L_{\text{cov}}(\theta, \nu) = \frac{1}{2}(N\omega^\delta - \theta)^\top (N\omega^\delta - \theta) + \nu^\top \sqrt{\tilde{\mathbb{L}}}\theta + \frac{\alpha_\nu}{2}\|\sqrt{\tilde{\mathbb{L}}}\theta\|^2 \tag{13}$$

where $\omega^\delta = [\omega_1^\delta; \cdots; \omega_N^\delta]$, $\tilde{\mathbb{L}} = \mathcal{L} \otimes I_{n_{\text{cov}}}$ and $\alpha_\nu > 0$ is a positive constant. To derive the update laws for θ and ν, we take the gradient of the augmented Lagrangian in (13), yielding

$$\begin{aligned} \nabla_\theta L &= -(N\omega^\delta - \theta) + \sqrt{\tilde{\mathbb{L}}}\nu + \alpha_\nu \tilde{\mathbb{L}}\theta \\ \nabla_\nu L &= \sqrt{\tilde{\mathbb{L}}}\theta. \end{aligned} \tag{14}$$

To enable a distributed update law for minimizing (13), we introduce an auxiliary variable

$$\tilde{\nu}_{t,l} = \sqrt{\tilde{\mathbb{L}}}\nu_{t,l}. \tag{15}$$

The distributed update laws for the primal and dual variables are then given as

$$\begin{aligned} \tilde{\nu}_{t,l+1} &= \tilde{\nu}_{t,l} + \alpha_\nu \tilde{\mathbb{L}}\theta_{t,l} \\ \theta_{t,l+1} &= N\omega^\delta - \tilde{\nu}_{t,l+1} - \alpha_\nu \tilde{\mathbb{L}}\theta_{t,l}. \end{aligned} \tag{16}$$

Next, we parameterize the solution of (8) using the saddle-point equation.

Lemma 1: Let $P_{i,t|t-1}$ be a positive definite and symmetric matrix. Then, the solution of (P1), considering the augmented Lagrangian (8) is parameterized as $(\xi_t^*, \lambda_t^*) = ((1_N \otimes I_n)\xi_t^\dagger, (1_N \otimes I_n)\hat{\lambda}_t + \bar{\lambda}_t)$, where $\xi_t^\dagger = \mathcal{H}_t^{-1}\mathbb{1}_N^\top \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t$, $\hat{\lambda}_t \in \mathbb{R}^n$ is an arbitrary vector, and $\bar{\lambda}_t \in \mathbb{R}^{Nn}$.

Proof: Let (ξ_t^*, λ_t^*) be the solution of (8). Using the Karush–Kuhn–Tucker (KKT) condition [32, Thm. 12.1], (9) can be written as

$$\begin{bmatrix} \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}} + \mu\mathbb{L} & \sqrt{\mathbb{L}} \\ \sqrt{\mathbb{L}} & \mathbb{0} \end{bmatrix} \begin{bmatrix} \xi_t^* \\ \lambda_t^* \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t \\ 0_{Nn} \end{bmatrix}. \tag{17}$$

From the primal feasibility condition in (17), we have

$$\sqrt{\mathbb{L}}\xi_t^* = 0_{Nn}$$

implying ξ_t^* lies in the nullspace of $\sqrt{\mathbb{L}}$, which is spanned by $1_N \otimes I_n$. Thus, $\xi_t^* = (1_N \otimes I_n)\xi_t^\dagger$, where $\xi_t^\dagger \in \mathbb{R}^n$. For the dual variable, the dual feasibility condition (17) can be written as

$$\begin{aligned} (\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}} + \mu\mathbb{L})(1_N \otimes I_n)\xi_t^\dagger + \sqrt{\mathbb{L}}\lambda_t^* &= \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t \\ \sqrt{\mathbb{L}}\lambda_t^* &= \bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t - (\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}} + \mu\mathbb{L})(1_N \otimes I_n)\xi_t^\dagger \\ \mathbb{L}\lambda_t^* &= \sqrt{\mathbb{L}}(\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t - (\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}} + \mu\mathbb{L})(1_N \otimes I_n)\xi_t^\dagger). \end{aligned} \tag{18}$$

Let $b = \sqrt{\mathbb{L}}(\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\mathbf{z}_t - (\bar{\mathbf{H}}^\top \bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}} + \mu\mathbb{L})(1_N \otimes I_n)\xi_t^\dagger)$. We know that $\mathcal{L}U = U\Lambda$ where $\Lambda = \mathrm{diag}(0, \lambda_2, \ldots, \lambda_N)$, $U = [u_N \ \bar{U}]$ with $u_N = \frac{1}{\sqrt{N}}1_N$, $1_N^\top \bar{U} = 0_N^\top$, and $\bar{U}^\top \bar{U} = I_{N-1}$, we get $\lambda_t^* = U \otimes I_n[\hat{\lambda}_t; \tilde{\Lambda}^{-1}\bar{U}^\top b]$, where $\tilde{\Lambda} = \mathrm{diag}(\lambda_2, \ldots, \lambda_N)$. ■

Next, we parameterize the solution of (P2) using the KKT condition.

Lemma 2: The solution of (P2) is parameterized as $(\theta_t^*, \nu_t^*) = ((1_N \otimes I_{n_{\mathrm{cov}}})\theta_t^\dagger, (1_N \otimes I_{n_{\mathrm{cov}}})\tilde{\nu} + \bar{\nu})$, where $\theta_t^\dagger = \mathrm{vech}(\Theta^\dagger)$, $\Theta^\dagger = H^\top \bar{R}^{-1}H = \sum_{i=1}^N H_i^\top R_i^{-1}H_i$ $\tilde{\nu} = (\bar{U} \otimes I_{n_{\mathrm{cov}}})(\tilde{\Lambda}^{-1/2} \otimes I_{n_{\mathrm{cov}}})(\bar{U} \otimes I_{n_{\mathrm{cov}}})^\top(N\omega^\delta)$, and $\bar{\nu} \in \mathbb{R}^{n_{\mathrm{cov}}}$ is an arbitrary vector.

Proof: The proof of the Lemma 2 follows the same path as the proof of Lemma 1, thus omitted. ■

A pseudocode implementation of the proposed filtering method is provided in Algorithm 1 for T time steps. We notice that the update step for the state estimate in (12) is independent of the update step of the information rate matrix given in (16). Thus, in the following, we first show the boundedness and convergence of the covariance matrix for all the local estimators. Thenceforth, we show the convergence of the state estimate for all local estimators.

V. Stability Analysis

In this section, we analyze the stability of the proposed DKF algorithm. The update laws for the state estimate (12) and the posterior covariance (16) are modeled as discrete-time dynamical systems [33]. Using tools from system theory, we derive conditions on the design parameters that guarantee the asymptotic stability of the update laws given in (12) and (16). Finally, we show that all local estimators are unbiased, that is, $\lim_{t\to\infty} \mathbb{E}[x_t - \xi_{i,t}] = 0$.

Theorem 1: Let the communication network $\mathcal{G}$ of local estimators be undirected and connected, and let $\mathcal{L}$ be the Laplacian matrix of $\mathcal{G}$. If $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$, then the sequence $\{\Theta_{i,t} = \mathrm{vech}^{-1}(\theta_{i,t})\}_t$ generated by (16) converges to the global information rate matrix $H^\top \bar{R}^{-1}H$.

Proof: The update law for $\theta_{t,l}$ in (16) can be written as

$$\theta_{t,l+1} = (\mathbb{I} - 2\alpha_\nu\tilde{\mathbb{L}})\theta_{t,l} + \alpha_\nu\tilde{\mathbb{L}}\theta_{t,l-1}. \tag{19}$$

To analyze the stability, we apply the coordinate transformation

$$\theta_{t,l} = \mathbb{U}^\top\phi_{t,l}, \quad \tilde{\nu}_{t,l} = \mathbb{U}^\top\psi_{t,l} \tag{20}$$

Algorithm 1: DKF Using ADMM.

Initialization: Take any arbitrary $\hat{x}_{i,0}$, $P_{i,0} > 0$. Set $\theta_{i,0} = \mathrm{vech}(H_i^\top R_i^{-1}H_i)$, and $\tilde{\lambda}_{i,0} = 0_{Nn}$, $\tilde{\nu}_{i,0} = 0_{Nn_{\mathrm{cov}}}$, $t = 1$.
function $[\hat{x}_{i,t|t},\ P_{i,t|t}] = \mathrm{DKF}(\hat{x}_{i,0|0}, P_{i,0|0})$.
  for $t = 1, \ldots, T$ **do**
    Compute the prior mean at node i, $\hat{x}_{i,t|t-1} = F\hat{x}_{i,t-1|t-1}$.
    Evaluate the prior error covariance at node i, $P_{i,t|t-1} = FP_{i,t-1|t-1}F^\top + Q$.
    $\xi_{i,t,0} = \hat{x}_{i,t|t-1}$, $\tilde{\lambda}_{i,t,0} = 0_{Nn}$.
    for $l = 0, \ldots, L-1$ **do**
      $\tilde{\lambda}_{i,t,l+1} = \tilde{\lambda}_{i,t,l} + \alpha_\lambda K_{i,t}^{-1}\sum_{j\in\mathcal{N}_i} a_{ij}(\xi_{i,t,l} - \xi_{j,t,l})$, where $K_{i,t}^{-1} = H_i^\top R_i^{-1}H_i + \frac{1}{N}P_{i,t|t-1}^{-1}$.
      $\xi_{i,t,l+1} = K_{i,t}(H_i^\top R_i^{-1}y_{i,t} + \frac{1}{N}P_{i,t|t-1}^{-1}\hat{x}_{i,t|t-1}$ $-\tilde{\lambda}_{i,t,l+1} - \mu K_{i,t}^{-1}\sum_{j\in\mathcal{N}_i} a_{ij}(\xi_{i,t,l} - \xi_{j,t,l}))$.
    end for
    $\tilde{\nu}_{i,t} = \tilde{\nu}_{i,t-1} + \alpha_{\nu,i}\sum_{j\in\mathcal{N}_i} a_{ij}(\theta_{i,t} - \theta_{j,t})$.
    $\theta_{i,t} = N\omega_i^\delta - \tilde{\nu}_{i,t} - \alpha_\nu\sum_{j\in\mathcal{N}_i} a_{ij}(\theta_{i,t} - \theta_{j,t})$.
    $\hat{x}_{i,t|t} = \xi_{i,t,L}$, $P_{i,t|t} = (P_{i,t|t-1}^{-1} + \Theta_{i,t})^{-1}$, where $\Theta_{i,t} = \mathrm{vech}^{-1}(\theta_{i,t})$.
  end for
end Function

where $\mathbb{U} = U \otimes I_{n_{\mathrm{cov}}}$. Substituting (20) into (19), we obtain

$$\begin{aligned} \psi_{t,l+1} &= \psi_{t,l} + \alpha_\nu(\Lambda \otimes I_{n_{\mathrm{cov}}})\phi_{t,l} \\ \phi_{t,l+1} &= (I - 2\alpha_\nu(\Lambda \otimes I_{n_{\mathrm{cov}}}))\phi_{t,l} + \alpha_\nu(\Lambda \otimes I_{n_{\mathrm{cov}}})\phi_{t,l-1}. \end{aligned} \tag{21}$$

Let $\psi_{t,l} = [\tilde{\psi}_{t,l}^\top,\ \bar{\psi}_{t,l}^\top]^\top$ and $\phi_{t,l} = [\tilde{\phi}_{t,l}^\top,\ \bar{\phi}_{t,l}^\top]^\top$, where $\tilde{\psi}_{t,l}$ and $\tilde{\phi}_{t,l}$ correspond to the nullspace of $\tilde{\mathbb{L}}$, whereas $\bar{\psi}_{t,l}$ and $\bar{\phi}_{t,l}$ correspond to the range space of $\tilde{\mathbb{L}}$. The dynamics in (21) decompose as

$$\begin{aligned} \tilde{\psi}_{t,l+1} &= \tilde{\psi}_{t,l} \\ \bar{\psi}_{t,l+1} &= \bar{\psi}_{t,l} + \alpha_\nu(\tilde{\Lambda} \otimes I_{n_{\mathrm{cov}}})\bar{\phi}_{t,l} \\ \tilde{\phi}_{t,l+1} &= \tilde{\phi}_{t,l} \\ \bar{\phi}_{t,l+1} &= (I - 2\alpha_\nu(\tilde{\Lambda} \otimes I_{n_{\mathrm{cov}}}))\bar{\phi}_{t,l} + \alpha_\nu(\tilde{\Lambda} \otimes I_{n_{\mathrm{cov}}})\bar{\phi}_{t,l-1} \end{aligned} \tag{22}$$

where $\tilde{\Lambda} = \mathrm{diag}(\lambda_2, \ldots, \lambda_N)$. Defining $\hat{\phi}_{t,l} = [\bar{\phi}_{t,l}^\top, \bar{\phi}_{t,l-1}^\top]^\top$, we write the dynamics of $\bar{\phi}_{t,l}$ as

$$\hat{\phi}_{t,l+1} = M\hat{\phi}_{t,l} \tag{23}$$

where

$$M = \begin{bmatrix} I - 2\alpha_\nu(\tilde{\Lambda} \otimes I_{n_{\mathrm{cov}}}) & \alpha_\nu(\tilde{\Lambda} \otimes I_{n_{\mathrm{cov}}}) \\ I & \mathbb{0} \end{bmatrix}.$$

The eigenvalues of M determine the stability of the system given in (23). For each eigenvalue $\lambda_i(\tilde{\Lambda})$, we define

$$M_i = \begin{bmatrix} 1 - 2\alpha_\nu\lambda_i(\tilde{\Lambda}) & \alpha_\nu\lambda_i(\tilde{\Lambda}) \\ 1 & 0 \end{bmatrix}.$$

The eigenvalues of M_i can be written as

$$\lambda(M_i) = \frac{1}{2}\left(1 - 2\alpha_\nu\lambda_i \pm \sqrt{(1 - 2\alpha_\nu\lambda_i)^2 + 4\alpha_\nu\lambda_i}\right).$$

For the asymptotic stability of (23), $|\lambda(M_i)| < 1$. This condition holds if $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$. To ensure the stability of the entire system, we analyze the boundedness of $\bar{\psi}_{t,l}$. From (22), the dynamics of $\bar{\psi}_{t,l}$ are given by

$$\bar{\psi}_{t,l+1} = \sum_{k=0}^{l} \alpha_\nu (\tilde{\Lambda} \otimes I_{n_{\text{cov}}}) \bar{\phi}_{t,k}. \tag{24}$$

Since $\bar{\phi}_{t,l} \to 0$ exponentially as $l \to \infty$, there exist constants $C < \infty$ and $0 < \rho < 1$ such that

$$\|\bar{\phi}_{t,l}\| \leq C\rho^l. \tag{25}$$

Using (25), the norm of $\bar{\psi}_{t,l}$ is bounded as

$$\|\bar{\psi}_{t,l}\| \leq \alpha_\nu \|\tilde{\Lambda} \otimes I_{n_{\text{cov}}}\| \frac{C}{1-\rho}. \tag{26}$$

Thus, $\bar{\psi}_{t,l}$ remains bounded. Since $\tilde{\phi}_{t,l}$ corresponds to the nullspace of $\tilde{\mathbb{L}}$, it remains constant. Specifically, we have

$$\tilde{\phi}_{t,\infty} = (u_N^\top \otimes I_{n_{\text{cov}}}) N\omega^\delta = \sqrt{N} \sum_{i=1}^{N} \text{vech}(H_i^\top R_i^{-1} H_i).$$

Thus, under the condition $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$ and using the transformation of $\theta_{t,l}$ given in (20), $\{\Theta_{i,t}\}_t$ converges to $H^\top \bar{R}^{-1} H$ by Lemma 2, completing the proof. ∎

Notice that the distributed optimization problem in (P2) is static. Consequently, the update law in (16) can be expressed without subiterations as (as in Algorithm 1)

$$\begin{aligned} \tilde{\nu}_{t+1} &= \tilde{\nu}_t + \alpha_\nu \tilde{\mathbb{L}} \theta_t \\ \theta_{t+1} &= N\omega^\delta - \tilde{\nu}_{t+1} - \alpha_\nu \tilde{\mathbb{L}} \theta_t. \end{aligned} \tag{27}$$

The sequences generated by (27) converge under the same conditions stated in Theorem 1, which we summarize in the following corollary.

Corollary 1: Let the communication network $\mathcal{G}$ of local estimators be undirected and connected, and let $\mathcal{L}$ be the Laplacian matrix of $\mathcal{G}$. If $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$, then the sequence $\{\Theta_{i,t} = \text{vech}^{-1}(\theta_{i,t})\}_t$ generated by (27) converges to the global information rate matrix $H^\top \bar{R}^{-1} H$.

Proof: We begin by applying the same coordinate transformation as in Theorem 1

$$\theta_t = \mathbb{U}^\top \phi_t, \quad \tilde{\nu}_t = \mathbb{U}^\top \psi_t. \tag{28}$$

Substituting (28) into (27), the transformed dynamics are given by

$$\begin{aligned} \psi_{t+1} &= \psi_t + \alpha_\nu (\Lambda \otimes I_{n_{\text{cov}}}) \phi_t \\ \phi_{t+1} &= (I - 2\alpha_\nu (\Lambda \otimes I_{n_{\text{cov}}}))\phi_t + \alpha_\nu (\Lambda \otimes I_{n_{\text{cov}}}) \phi_{t-1}. \end{aligned} \tag{29}$$

The above dynamics are identical to those analyzed in Theorem 1, except that the update law in (27) does not include subiterations. Since the distributed optimization problem is static, the stability and convergence analysis in Theorem 1 directly applies. Therefore, for brevity, details are omitted here. ∎

Next, we show the boundedness of the posterior covariance matrix $P_{i,t}$ and the prior covariance matrix $P_{i,t|t-1}$.

Lemma 3: Consider the ADMM algorithm for DKF, as described in Algorithm 1. Assume that Assumptions 1 and 2 hold. If $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$, then there exist positive symmetric matrices $\bar{P} < \infty$ and $\underline{P} > 0$ such that $\underline{P} < P_{i,t} < \bar{P} <$ and $\underline{P} < P_{i,t|t-1} < \bar{P}$.

Proof: The proof follows the same path as given in [20, Lemma 8], thus omitted. ∎

Next, we show that the local covariance matrices $P_{i,t|t-1}$ for all $i \in \{1, 2, \ldots, N\}$ generated by Algorithm 1 converge to P^*, which is a unique solution of the following algebraic Riccati equation:

$$P = F\{P - PH^\top (H^\top PH + \bar{R})^{-1} HP\} F^\top + Q. \tag{30}$$

Theorem 2: Consider Algorithm 1 and let Assumptions 1 and 2 hold. Let P^* be the unique solution of (30). Let $0 < \alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$. Then the covariance matrices $P_{i,t|t-1}$ for all $i \in \{1, 2, \ldots, N\}$ generated by Algorithm 1 converge to P^*.

Proof: The proof follows the same path as in [20, Thm. 9], thus omitted. ∎

Next, we establish the stability of the sequence $\{\xi_{t,l}\}$ generated by Algorithm 1. To this end, we reformulate the dynamics of $\xi_{t,l}$ independently of $\tilde{\lambda}_{t,l}$. This reformulation allows us to use stability tools from dynamical systems theory, providing bounds on the design parameters α_λ and μ. Specifically, we define the error terms as

$$e_{t,l}^{\xi} = \xi_t^* - \xi_{t,l}, \quad e_{t,l}^{\tilde{\lambda}} = \tilde{\lambda}_t^* - \tilde{\lambda}_{t,l}. \tag{31}$$

The update step for $\xi_{t,l}$ in (12) can be expressed as a second-order discrete-time dynamical system

$$\xi_{t,l+1} = \xi_{t,l} + \mu \mathbb{L} \xi_{t,l-1} - (\alpha_\lambda + \mu) \mathbb{L} \xi_{t,l}. \tag{32}$$

Substituting the error definitions from (31) into (12) and (32), the error dynamics are given by

$$\begin{aligned} e_{t,l+1}^{\tilde{\lambda}} &= e_{t,l}^{\tilde{\lambda}} + \alpha_\lambda K_t^{-1} \mathbb{L} e_{t,l}^{\xi} \\ e_{t,l+1}^{\xi} &= e_{t,l}^{\xi} - (\alpha_\lambda + \mu) \mathbb{L} e_{t,l}^{\xi} + \mu \mathbb{L} e_{t,l-1}^{\xi}. \end{aligned} \tag{33}$$

The next result establishes conditions on the design parameters α_λ and μ in terms of $\lambda_{\max}(\mathcal{L})$. In contrast to [20], where the parameters are related to $\lambda_{\max}^2(\mathcal{L})$, the bounds provided here mitigate the slowdown of the consensus process, reducing the need for a larger L to achieve improved performance.

Theorem 3: Consider the discrete-time dynamical system given in (33) with $\alpha_\lambda > 0$ and $\mu > 0$ such that $\alpha_\lambda + 2\mu < \frac{2}{\lambda_{\max}(\mathcal{L})}$. Let Assumptions 1 and 2 hold. Then the state vector $e_{t,l}^{\xi}$ converges to zero asymptotically for sufficiently large L, and $e_{t,l}^{\tilde{\lambda}}$ remains bounded for any l. Consequently, $\xi_{t,l}$ in (12) asymptotically converges to ξ_t^*.

Proof: To analyze the stability of the equilibrium of (33), we introduce the following coordinate transformation:

$$e_{t,l}^{\xi_T} = \mathbb{U}^\top e_{t,l}^{\xi}. \tag{34}$$

Substituting (33) into (34), the discrete-time dynamical system can be written in the transformed coordinates as

$$\begin{aligned} \tilde{e}_{t,l+1}^{\xi_T} &= \tilde{e}_{t,l}^{\xi_T} \\ \bar{e}_{t,l+1}^{\xi_T} &= \bar{e}_{t,l}^{\xi_T} - (\alpha_\lambda + \mu) \tilde{\bar{\Lambda}} \bar{e}_{t,l}^{\xi_T} + \mu \tilde{\bar{\Lambda}} \bar{e}_{t,l-1}^{\xi_T} \end{aligned} \tag{35}$$

where $\tilde{\bar{\Lambda}} = \tilde{\Lambda} \otimes I_n$ with $\tilde{\Lambda} = \text{diag}(\lambda_2(\mathcal{L}), \ldots, \lambda_N(\mathcal{L}))$. Letting $\hat{e}_{t,l}^{\xi_T} = [\bar{e}_{t,l}^{\xi_T \top}, \bar{e}_{t,l-1}^{\xi_T \top}]^\top$, we rewrite (35) in compact form

$$\hat{e}_{t,l+1}^{\xi_T} = \bar{M} \hat{e}_{t,l}^{\xi_T}, \quad \text{where} \quad \bar{M} = \begin{bmatrix} \mathbb{I} - (\alpha_\lambda + \mu)\tilde{\bar{\Lambda}} & \mu \tilde{\bar{\Lambda}} \\ \mathbb{I} & \mathbb{O} \end{bmatrix}. \tag{36}$$

To show the stability of (36), we ensure that $\bar{M}$ is Schur stable. The eigenvalues of $\bar{M}$ are determined by the eigenvalues of the following 2×2 matrix for each eigenvalue $\lambda_i(\tilde{\Lambda})$ of $\tilde{\Lambda}$

$$\bar{M}^i = \begin{bmatrix} \bar{\alpha}_{\lambda\mu} & \mu\lambda_i \\ 1 & 0 \end{bmatrix}, \text{ where } \bar{\alpha}_{\lambda\mu} = 1 - (\alpha_\lambda + \mu)\lambda_i.$$

The eigenvalues of $\bar{M}^i$ are given by

$$\lambda(\bar{M}^i) = \frac{1}{2}\left(\bar{\alpha}_{\lambda\mu} \pm \sqrt{\bar{\alpha}_{\lambda\mu}^2 + 4\mu\lambda_i}\right).$$

For $\bar{M}^i$ to be Schur stable, we require $|\lambda(\bar{M}^i)| < 1$. This condition is satisfied if $\alpha_\lambda > 0$, $\mu > 0$, and $\alpha_\lambda + 2\mu < \frac{2}{\lambda_{\max}(\mathcal{L})}$ for all $i \in \{1, \ldots, N-1\}$. Under these conditions, $\bar{M}$ is Schur stable, and $\hat{e}_{t,l}^{\xi_T} \to 0$ as $l \to \infty$. Next, we show the convergence of $\tilde{e}_{t,l}^{\xi_T}$ given in (35). Using the transformation given (34), we have

$$e_{t,l}^{\xi} = (u_N \otimes I_n)\bar{e}_{t,l}^{\xi_T} + (\bar{U} \otimes I_n)\tilde{e}_{t,l}^{\xi_T}. \quad (37)$$

As $l \to \infty$, $e_{t,\infty}^{\xi} = (u_N \otimes I_n)\bar{e}_{t,0}^{\xi_T}$. From the transformation given in (34) and the definition of $e_{t,l}^{\xi}$ given in (33), we conclude using the results in Lemma 1 that $\xi_{t,l} \to \xi_t^*$ as $l \to \infty$. To establish the boundedness of $e_{t,l}^{\tilde{\lambda}}$, consider its update equation in (33)

$$e_{t,l+1}^{\tilde{\lambda}} = e_{t,l}^{\tilde{\lambda}} + \alpha_\lambda K_t^{-1}\mathbb{L}e_{t,l}^{\xi}.$$

Since $e_{t,l}^{\xi} \to 0$ as $l \to \infty$, $e_{t,l}^{\tilde{\lambda}}$ satisfies the following bound:

$$\|e_{t,l}^{\tilde{\lambda}}\| \le \|e_{t,0}^{\tilde{\lambda}}\| + \alpha_\lambda\|K_t^{-1}\|\frac{C}{1-\rho} \quad (38)$$

where $C < \infty$ and $0 < \rho < 1$ are constants determined by the exponential decay of $e_{t,l}^{\xi}$. Thus, $e_{t,l}^{\tilde{\lambda}}$ remains bounded. Under the conditions $\alpha_\lambda > 0$, $\mu > 0$, and $\alpha_\lambda + 2\mu < \frac{2}{\lambda_{\max}(\mathcal{L})}$, $e_{t,l}^{\xi} \to 0$ asymptotically, and $e_{t,l}^{\tilde{\lambda}}$ remains bounded. This completes the proof. ■

Next, we show the stability of the state estimates as t approaches infinity. We use the following definitions and lemma in the stability analysis. Define $\mu_t^{\xi} = \mathbb{E}\{e_{t,l}^{\xi}\} = \mathbb{E}\{\xi_t^* - \xi_{t,l}\}$, $\mu_t^{\tilde{\lambda}} = \mathbb{E}\{e_t^{\tilde{\lambda}}\} = \mathbb{E}\{\tilde{\lambda}_t^* - \tilde{\lambda}_{t,l}\}$, and $\mu_t^{\dagger} = \mathbb{E}\{x_t - \xi_t^{\dagger}\}$, where $\xi_t^{\dagger} = \mathcal{H}_t^{-1}\mathbb{1}_N^\top\bar{\mathbf{H}}^\top\bar{\mathbf{S}}_t^{-1}\mathbf{z}_t$ with $\mathcal{H}_t = \mathbb{1}_N^\top\bar{\mathbf{H}}^\top\bar{\mathbf{S}}_t^{-1}\bar{\mathbf{H}}\mathbb{1}_N$. Next we provide the following lemma.

Lemma 4: Consider the following function:

$$V_t^{\dagger} = \mu_t^{\dagger^\top}\mathcal{H}_t\mu_t^{\dagger}. \quad (39)$$

Let Assumption 1 hold and that there exists $\bar{f} > 0$ such that $\|F\|_F \le \bar{f}$. Then, there exist $c^{\dagger} > 0$, $\bar{c} > 0$, $c_o > 0$, and $c_1 > 0$ such that

$$V_{t+T}^{\dagger} - V_t^{\dagger} \le -c^{\dagger}\|\mu_t^{\dagger}\| + c_2\|\mu_t^{\xi}\| \quad (40)$$

where $c_2 = \left(\frac{c_o^2(T-1)\tilde{c}}{4c^{\dagger}} + c_1\right)(T-1)\tilde{c} + 1$ with $\tilde{c} = \max\{1, \bar{c}^{(T-1)}\}$.

Proof: The proof follows the similar path as given in Theorem 12 [20], therefore omitted. ■

Next, we provide the stability analysis.

Theorem 4: Let the Assumptions 1 and 2 hold. Then, the sequence generated by Algorithm 1 satisfies the following:

$$\lim_{t\to\infty}\mathbb{E}\{x_t - \xi_{i,t}\} = 0. \quad (41)$$

Proof: Defining $\eta_t^{\hat{e}} = \mathbb{E}\{\bar{e}_{t,l}^{\xi_T}\}$ and taking the expectation of (36), we write

$$\eta_{t+1}^{\hat{e}} = \bar{M}\eta_t^{\hat{e}}. \quad (42)$$

Next, we define $V_t^{\eta} = \eta_t^{\hat{e}^\top}\mathbb{P}\eta_t^{\hat{e}}$. Using the results in Theorem 3, we conclude that there exists a positive definite matrix Γ such that

$$V_{t+1}^{\eta} - V_t^{\eta} = -\eta_t^{\hat{e}^\top}\Gamma\eta_t^{\hat{e}}. \quad (43)$$

To see the behavior of $\eta_t^{\hat{e}}$ over T time steps, we take

$$V_{t+T}^{\eta} - V_t^{\eta} = -\eta_t^{\hat{e}^\top}\left\{\sum_{t=0}^{T-1}(\bar{M}^\top)^t\Gamma(\bar{M})^t\right\}\eta_t^{\hat{e}}. \quad (44)$$

By [34, Lemma 2.1 and Thm. 2.1], there exists a positive definite matrix $\bar{\Gamma}$ such that

$$V_{t+T}^{\eta} - V_t^{\eta} = -\eta_t^{\hat{e}^\top}\bar{\Gamma}\eta_t^{\hat{e}}. \quad (45)$$

Using the transformation (34), the following inequality can be derived from (45)

$$V_{t+T}^{\eta} - V_t^{\eta} \le -c_3\lambda_{\mathbb{R}_{>0}}(\bar{\Gamma})\|\mu_t^{\xi}\|^2 \quad (46)$$

where $\lambda_{\mathbb{R}_{>0}}(\bar{\Gamma}) > 0$ is the smallest positive eigenvalue of $\bar{\Gamma}$, and $c_3 > 0$. Next, consider the following Lyapunov function candidate:

$$V_t = V_t^{\dagger} + \gamma V_t^{\eta}. \quad (47)$$

To see the behavior of $\mu_t^{\dagger}$ and μ_t^{ξ} over T time steps, we take

$$\begin{aligned} V_{t+T} - V_t &= V_{t+T}^{\dagger} - V_t^{\dagger} + \gamma(V_{t+T}^{\eta} - V_t^{\eta}) \\ &\le -c^{\dagger}\|\mu_t^{\dagger}\|^2 + c_2\|\mu_t^{\xi}\|^2 + \gamma(-\lambda_{\mathbb{R}_{>0}}(\bar{\Gamma})c_3\|\mu_t^{\xi}\|^2 \\ &\le -c^{\dagger}\|\mu_t^{\dagger}\|^2 - (\gamma\lambda_{\mathbb{R}_{>0}}(\bar{\Gamma})c_3 - c_2)\|\mu_t^{\xi}\|^2. \end{aligned} \quad (48)$$

We choose $\gamma > 0$ and $\bar{\Gamma}$ such that $\gamma\lambda_{\mathbb{R}_{>0}}(\bar{\Gamma})c_3 - c_2 > 0$. Next, to show the asymptotic convergence of (41), we consider the following Lyapunov function candidate over a sliding window of T time step:

$$V_t^{w} = \sum_{j=0}^{T-1} V_{tT+j}. \quad (49)$$

The time-difference of V_t^{w} can be written as

$$\begin{aligned} V_{t+1}^{w} - V_t^{w} &= V_{tT+T} - V_{tT} + V_{tT+1+T} - V_{tT+1} + \cdots \\ &\quad + V_{tT+T-1+T} - V_{tT+T-1}. \end{aligned} \quad (50)$$

By letting $\tilde{t} = tT$ in (50), we see from (48) that $V_{t+1}^{w} - V_t^{w}$ is negative definite along the trajectories of the dynamics of $\mathbb{E}\{e_{t,l}^{\xi}\}$ and $\mu_t^{\dagger}$, and the result follows. ■

VI. Simulation Results

In this section, we validate the theoretical results of the proposed distributed ADMM-based DKF algorithm (ADMM-DKF) by simulating a network of 100 and 10 sensor nodes tracking the trajectory of a car moving with constant velocity, as described in [4, pp. 99–101]. We also compare the results with the distributed dual-ascent-based DKF (DA-DKF) [20]. The discrete-time dynamical model of a car is given below:

$$x_{t+1} = Fx_t + w_t, \quad w_t \sim \mathcal{N}(0, Q) \quad (51)$$

where $x_t = [x_1, x_2, x_3, x_4]^\top$ is the state vector

$$F = \begin{bmatrix} I_2 & \delta t\, I_2 \\ \mathbb{0} & I_2 \end{bmatrix}, \quad \text{and} \quad Q = \begin{bmatrix} \frac{q\,\delta t^3}{3} I_2 & \frac{q\,\delta t^2}{2} I_2 \\ \frac{q\,\delta t^2}{2} & q\,\delta t\, I_2 \end{bmatrix}.$$

Here, (x_1, x_2) represents the position of the car in the (x, y) plane, and (x_3, x_4) represent the corresponding velocities. For the numerical simulation, we consider the process noise intensity $q = 1$ and sampling time $\delta t = 0.1$ second. The measurement at the ith sensor node is either x_1 or x_2 randomly at each time step. Measurement noise covariance is selected as per [4, pp. 99–101]. The simulation spans a total duration of 10 s.

First, we consider a network of 100 sensor nodes. The design parameters for ADMM-DKF are $\alpha_\lambda = \alpha_\nu = \mu = \frac{2}{3\lambda_{\max}(\mathcal{L})+\epsilon}$, where $\epsilon = 0.001$, whereas the design parameters for DA-DKF are $\alpha_\lambda = \alpha_\nu = \frac{2}{(\lambda_{\max}(\mathcal{L})+\epsilon)^2}$. The estimator at the ith node is initialized with randomly selected $\hat{x}_{i,0|0}$ and $P_{i,0|0}$. For the state estimation of (51), the

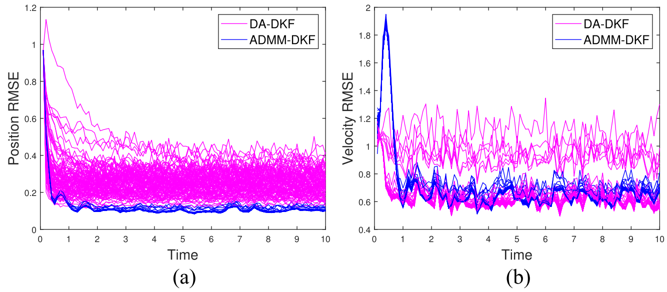

Fig. 1. Position and velocity RMSE of the DA-DKF and proposed ADMM-DKF for 100 nodes obtained from 50 MC runs.

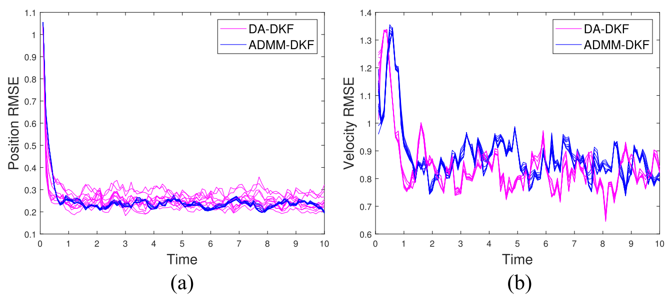

Fig. 2. Position and velocity RMSE of the DA-DKF and proposed ADMM-DKF for ten nodes obtained from 50 MC runs.

proposed ADMM-DKF algorithm is implemented with 20 subiterations per time step ($L = 20$), whereas the DA-DKF is implemented with $L = 2000$ subiterations per time step. The filtering performance is evaluated in terms of the root-mean-squared error (RMSE) of the position and velocity estimates. In Fig. 1, we plot the RMSE of the position and velocity estimates for all estimators for both DA-DKF and ADMM-DKF, averaged over 50 Monte Carlo (MC) runs. Fig. 1 demonstrates that the position and velocity RMSE values converge. The error of the proposed ADMM-DKF is small compared to that of DA-DKF. Recall that we use $L = 20$ for ADMM-DKF and $L = 2000$ for DA-DKF to ensure that a low number of iterations does not cause the difference. However, the error of DA-DKF can be reduced by decreasing the number of sensor nodes, as shown below.

Next, we consider a network of ten sensor nodes to estimate the state of (51). The number of subiterations for ADMM-DKF and DA-DKF is set to 40 and 1500, respectively. Fig. 2 shows the position and velocity RMSE for both DA-DKF and ADMM-DKF. From the figure, we observe that both distributed estimation algorithms converge and achieve similar performance across all nodes.

VII. Conclusion

We developed a consensus-based ADMM algorithm to derive the correction step for the distributed filtering. A new augmented Lagrangian formulation for the DKF problem was proposed, enabling a fully distributed implementation of the correction step for the posterior state and covariance estimates. The proposed consensus-based ADMM avoids exchanging dual variables, significantly reducing communication between nodes. The algorithm yields much tighter upper bounds, specifically $\alpha_\nu < \frac{2}{3\lambda_{\max}(\mathcal{L})}$ and $\alpha_\lambda + 2\mu < \frac{2}{\lambda_{\max}(\mathcal{L})}$. Larger design parameter values improve the convergence rate, allowing consensus to be achieved with fewer subiterations and greater accuracy. In addition, we observed that solving the distributed optimization problem for each node's covariance matrix is a static optimization task, eliminating the need for subiterations. We provided a stability analysis of the ADMM-based DKF. Furthermore, we showed all local estimators are unbiased. The performance of the proposed method was demonstrated on a car tracking problem using connected networks with 100 and ten sensor nodes, and compared against the DA-DKF algorithm presented in [20]. In the future, we aim to develop continuous-time DKF using ADMM [15] with time-varying network topologies.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[2] U. L. Mohite and H. G. Patel, "Optimization assisted kalman filter for cancer chemotherapy dosage estimation," *Artif. Intell. Med.*, vol. 119, 2021, Art. no. 102152.

[3] M. Athans, "The importance of Kalman filtering methods for economic systems," in *Annals of Economic and Social Measurement*, vol. 3. Cambridge, MA, USA: NBER, 1974, pp. 49–64.

[4] S. Särkkä and L. Svensson, *Bayesian Filtering Smoothing*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2023.

[5] M. Morgese, C. Wang, T. Taylor, M. Etemadi, and F. Ansari, "Distributed detection and quantification of cracks in operating large bridges," *J. Bridge Eng.*, vol. 29, no. 1, 2024, Art. no. 04023101.

[6] S. Kar, G. Hug, J. Mohammadi, and J. M. Moura, "Distributed state estimation and energy management in smart grids: A consensus innovations approach," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1022–1038, Dec. 2014.

[7] R. Olfati-Saber and P. Jalalkamali, "Coupled distributed estimation and control for mobile sensor networks," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2609–2614, Oct. 2012.

[8] M. Rashid and J. A. Nanzer, "High accuracy distributed Kalman filtering for synchronizing frequency and phase in distributed phased arrays," *IEEE Signal Process. Lett.*, vol. 30, pp. 688–692, 2023.

[9] L. An and G. H. Yang, "Distributed secure state estimation for cyber–physical systems under sensor attacks," *Automatica*, vol. 107, pp. 526–538, 2019.

[10] U. A. Khan and J. M. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4919–4935, Oct. 2008.

[11] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. 46th IEEE Conf. Decis. Control*, IEEE, 2007, pp. 5492–5498.

[12] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.

[13] D. Marelli, M. Zamani, M. Fu, and B. Ninness, "Distributed Kalman filter in a network of linear systems," *Syst. Control Lett.*, vol. 116, pp. 71–77, 2018.

[14] S. Das and J. M. Moura, "Consensus innovations distributed Kalman filter with optimized gains," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 467–481, Jan. 2017.

[15] A. Tanwani, "Suboptimal filtering over sensor networks with random communication," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5456–5463, Oct. 2022.

[16] S. P. Talebi and S. Werner, "Distributed Kalman filtering and control through embedded average consensus information fusion," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 4396–4403, Oct. 2019.

[17] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, Sep. 2010.

[18] S. Battilotti, A. Borri, F. Cacace, and M. d'Angelo, "Optimal discrete-time distributed Kalman filter with reduced communication," *IEEE Trans. Autom. Control*, vol. 70, no. 4, pp. 2754–2761, Apr. 2025.

[19] J. Yan, X. Yang, Y. Mo, and K. You, "A distributed implementation of steady-state Kalman filter," *IEEE Trans. Autom. Control*, vol. 68, no. 4, pp. 2490–2497, Apr. 2023.

[20] K. Ryu and J. Back, "Consensus optimization approach for distributed Kalman filtering: Performance recovery of centralized filtering," *Automatica*, vol. 149, 2023, Art. no. 110843.

[21] Q. Luo, S. Li, X. Yan, C. Wang, Z. Zhou, and G. Jia, "An improved two-phase robust distributed Kalman filter," *Signal Process.*, vol. 220, 2024, Art. no. 109438.

[22] X. Yan and L. Jin, "Distributed Kalman filter through trace proximity and covariance intersection," *IEEE Signal Process. Lett.*, vol. 31, pp. 1299–1303, 2024.
[23] K. Ryu and J. Back, "Distributed Kalman-filtering: Distributed optimization viewpoint," in *Proc. IEEE 58th Conf. Decis. Control*, IEEE, 2019, pp. 2640–2645.
[24] H. Li and Z. Lin, "Revisiting extra for smooth distributed optimization," *SIAM J. Optim.*, vol. 30, no. 3, pp. 1795–1821, 2020.
[25] P. Paritosh, N. Atanasov, and S. Martinez, "Distributed Bayesian estimation of continuous variables over time-varying directed networks," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 2545–2550, 2022.
[26] S. Wang, H. Paul, and A. Dekorsy, "Distributed optimal consensus-based Kalman filtering and its relation to MAP estimation," in *Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process.*, IEEE, 2018, pp. 3664–3668.
[27] S. Wang and A. Dekorsy, "Distributed consensus-based extended Kalman filtering: A Bayesian perspective," in *Proc. 27th Eur. Signal Process. Conf.*, 2019, pp. 1–5.
[28] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.
[29] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ, USA: Prentice Hall, 2000.
[30] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4051–4064, Aug. 2015.
[31] A. Makhdoumi and A. Ozdaglar, "Convergence rate of distributed ADMM over networks," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5082–5095, Oct. 2017.
[32] J. Nocedal and S. J. Wright, *Numerical Optimization*. Berlin, Germany: Springer, 1999.
[33] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 343–352.
[34] B. D. Anderson and J. B. Moore, *Optimal Filtering*. North Chelmsford, MA, USA: Courier Corporation, 2012.