Distributed Linear Filtering and Prediction of Time-varying Random Fields

Submitted in partial fulfillment for the requirements for
the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Subhro Das

B.Tech., Electronics & Electrical Comm. Engineering, Indian Institute of Technology Kharagpur
M.S., Electrical and Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

June 23, 2016

Abstract

We study distributed estimation of dynamic random fields observed by a sparsely connected network of agents/sensors. The sensors are inexpensive, low power, and they communicate locally and perform computation tasks. In the era of large-scale systems and big data, distributed estimators, yielding robust and reliable field estimates, are capable of significantly reducing the large computation and communication load required by centralized estimators, by running local parallel inference algorithms. The distributed estimators have applications in estimation, for example, of temperature, rainfall or wind-speed over a large geographical area; dynamic states of a power grid; location of a group of cooperating vehicles; or beliefs in social networks.

The thesis develops distributed estimators where each sensor reconstructs the estimate of the entire field. Since the local estimators have direct access to only local innovations, local observations or a local state, the agents need a consensus-type step to construct locally an estimate of their global versions. This is akin to what we refer to as distributed dynamic averaging. Dynamic averaged quantities, which we call pseudo-quantities, are then used by the distributed local estimators to yield at each sensor an estimate of the whole field. Using terminology from the literature, we refer to the distributed estimators presented in this thesis as *Consensus+Innovations*-type Kalman filters.

We propose three distinct types of distributed estimators according to the quantity that is dynamically averaged: (1) Pseudo-Innovations Kalman Filter (PIKF), (2) Distributed Information Kalman Filter (DIKF), and (3) *Consensus+Innovations* Kalman Filter (CIKF). The thesis proves that under minimal assumptions the distributed estimators, PIKF, DIKF and CIKF converge to unbiased and bounded mean-squared error (MSE) distributed estimates of the field. These distributed algorithms exhibit a Network Tracking Capacity (NTC) behavior – the MSE is bounded if the degree of instability of the field dynamics is below a threshold. We derive the threshold for each of the filters.

The thesis establishes trade-offs between these three distributed estimators. The NTC of the PIKF depends on the network connectivity only, while the NTC of the DIKF and of the CIKF depend also on the observation models. On the other hand, when all the three

estimators converge, numerical simulations show that the DIKF improves 2dB over the PIKF. Since the DIKF uses scalar gains, it is simpler to implement than the CIKF. Of the three estimators, the CIKF provides the best MSE performance using optimized gain matrices, yielding an improvement of 3dB over the DIKF.

Keywords: Kalman filter, distributed state estimation, multi-agent networks, sensor networks, distributed algorithms, consensus, innovation, asymptotic convergence, mean-squared error, dynamic averaging, Riccati equation, Lyapunov iterations, distributed signal processing, random dynamical systems.

Acknowledgments

I extend my heartfelt gratitude to my advisor Prof. José M. F. Moura for providing me the opportunity to pursue doctoral research under his supervision and guiding me through out my graduate studies. It has been a great learning experience. I am grateful to him for training me to become an independent researcher. I feel privileged to have him as my mentor.

I thank my committee members Prof. Rohit Negi, Prof. Bruno Sinopoli, Prof. Soummya Kar and Dr. Dmitry Malioutov for their insightful comments and suggestions. The last year of my research work has been shaped by those comments. I would also like to thank Prof. Franz Franchetti, Prof. Marija Ilic, Prof. Pulkit Grover and Prof. Gabriela Hug for various technical helps and advices.

I thank Carnegie Mellon University and Electrical and Computer Engineering Department for the opportunity to be a part of this institution and learn from the eminent faculties and fellow students. Thanks to the ECE staff members who have helped me in various ways over the last five years – Carol Patterson, Claire Bauerle, Elaine Lawrence, Samantha Goldstein and Nathan Snizaski.

Many thanks to my "Porter Basement" friends and colleagues - Kyle Anderson, Aurora Schmidt, Stephen Kruzick, Liangyan Gui, June Zhang, Joya Deri, Joel Harley, Jonathan Mei, Yuan Chen, Augusto Santos, Evgeny Toropov, Dusan Jakovetic, Nicholas O'Donoughue, Satwik Kottur, Mark Cheung, Shanghang Zhang, Anit Sahu, Brian Swenson, Steven Aday, Vishwanath S R V, Nipun Popli, Yaoqing Yang, Joao Saude, Javad Mohammadi, Jian Wang, Nicola Forti, Xiaoqi Yin, Siheng Chen, Stefanos Baros, Minhee Jun, Vinay Prabhu, Nikos Arechiga, Kyri Baker, Andrew Hsu, Jonathan Donadee, JY Joo, Sergio Pequito, Milos Cvetkovic and so many more, who made my work life full of fun.

A special thanks to my family – mom, dad and my sisters, who taught me to have dreams and helped me through every steps to make them come true. I am grateful to all of you for your love and sacrifices. I would like to thank my brother-in-law and my lovely niece who never fail to make me smile. Thanks to Meenakshi for always being there for me through thick and thin. You have helped make this an amazing journey. Thanks to all my friends Shubhaditya (housemate for five years), Arka, Pratiti, Anit, Amit, Sudipto, Namit, Shudhanshu, Manaal, Ankita, Utsav, Argha, Vaibhav and everyone else for making me feel at home in Pittsburgh.

I thank National Science Foundation (NSF), Air Force Office of Scientific Research (AFOSR) and Microsoft for the grants that made this thesis possible. This work was supported by NSF grants CCF1513936, CCF1011903 and CCF1018509, and AFOSR grant FA95501010291. Some of the simulations were run in Microsoft Azure cloud supported by Microsoft Azure for Research Grant.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

For decades, the Kalman-Bucy filter [1, 2] has played a key role in estimation, detection, or prediction of time-varying noisy signals. The Kalman filter is found in a wide variety of applications ranging from problems in navigation to environmental studies, computer vision to bioengineering, signal processing to econometrics. More recently, algorithms inspired by the Kalman filter have been applied to estimate random fields monitored by networks of sensors. In these problems, we distinguish two distinct layers: (a) the physical layer of the time-varying random field; and (b) the cyber layer of sensors observing the field.

A centralized approach to field estimation poses several challenges. It requires that all sensors communicate their measurements to a centralized fusion center. This is fragile to central node failure and severely taxes computationally the fusion center. Moreover, it also requires excessive communication bandwidth to and from the fusion center. Hence, the centralized approach is inelastic to estimation of large-scale time-varying random fields, like when estimating temperature, rainfall, or wind-speed over large geographical areas [3–6], estimating dynamic states of a power grid, [7–9], tracking a group of cooperating vehicles, [10, 11], or analyzing the diffusion of beliefs in social networks, [12]. The field is observed by a sparse network of sensors (agents) that are inexpensive, low power, communicate locally, and can compute. The goal of this thesis is to design distributed linear estimators of time-varying fields that use only local communication among the sensors such that the distributed estimates are asymptotically unbiased with bounded mean-squared errors (MSE).

1.2 Previous Work

We review related previous research on distributed estimation of time-varying random fields. We organize the prior work into two main classes, the second class being further sub-divided in two sub-classes: i) *Two time scale: Fast communication–slow dynamics and sensing*; and ii) *Single time scale*: ii-1) *Gossip Kalman filters*; and ii-2) *Consensus+innovations* estimators.

Two time scale: Fast communication–slow dynamics and sensing. Here, the linear distributed estimator reproduces a Kalman filter locally at each sensor with a consensus step on the observations. This is a two time scale algorithm because the sensing and the local Kalman filter updates occur at the same slow time scale of the process dynamics, while the consensus iterations happen at the fast time scale of communication among sensors. References [13–21] represent variations on this structure. Reference [15] modifies the local Kalman filters to account for the distributed nature of the field so that the local Kalman filters are all different since they operate on different dynamics (*local* dynamics). References [19, 20] assume local observability, i.e., the dynamical system is observable at each sensor; this is a much more stringent condition than global observability and is not practical in most distributed large-scale applications. The two time scale distributed estimator in [21] considers time-varying dynamics and observation matrices and expresses the measurements in information filter form. As a final comment, we may also include in this class [22, 23] that assume that the communication network is complete. Since every agent communicates with every other, every sensor acts as a centralized fusion center and the consensus step is trivial. In [23] the nodes broadcast only one bit based on the sign of the innovations.

Single time scale. In this class, sensing and communication occur once at every time step; the local estimators operate at the same time scale of the process dynamics. Operating at a single time scale reduces the onerous communication rounds required by the consensus step in the two time step estimators. The tradeoffs between different versions of these single time step algorithms are between the convergence rate and the asymptotic MSE of the estimators. Reference [24] proposed a *gossip Kalman filter*, a distributed estimator where each agent replicates a (local) Kalman filter processing its local observations. At every sensing time step, there is a communication round between a single pair of connected sensors, chosen according to the gossip protocol, [25], when these two sensors exchange their current state estimate and their current error covariance matrix provided by the

local Kalman filter Riccati equation. The reference shows through large deviations theory that, under appropriate conditions, asymptotically, each local estimate is unbiased and has bounded covariance, regardless of the degree of instability of the field dynamics. Reference [26] improved the convergence rate of [24] by allowing several pairs, rather than a single pair, of connected sensors to exchange their states. Although gossip Kalman filters can track any unstable field, they may pay a price on the communications (every round exchanges not only the state estimate but also the error covariance, an object on the order of M^2 that can be very large for spatially distributed fields), on the convergence rate, and on higher asymptotic MSE with respect to the *consensus+innovations* distributed estimator that we present here.

The *consensus+innovations* estimators, the second class of single time scale distributed estimators, extend to time varying dynamics the consensus+innovations distributed estimators introduced in [27, 28] for parameter estimation; see also [29]. The consensus+innovations distributed linear parameter estimator updates at every sensor the local state estimate by a consensus term on the current estimate and by the current local innovations–the difference between the current sensor observation and its local prediction. A number of approaches [30–33] have been developed to extend this structure to time varying fields. They add a term to the consensus+innovations distributed parameter estimator to account for the field time dynamics. In [30, 31], the innovation update at each sensor collects its own observations and those of its neighbors. These references show that the distributed estimate at every sensor is asymptotically unbiased and, if the degree of instability of the field dynamics is appropriately upper bounded (this bound is defined as the Network Tracking Capacity (NTC) [30, 31]), the distributed field estimator has bounded MSE. The NTC for the estimators in [30, 31] depends on the connectivity of the sensor network and on the assumed observation model. Further, [32, 33] developed structural system properties that the sensor network and the observation models satisfy such that the consensus+innovations type distributed estimator in [30] has bounded MSE performance. Reference [34] developed a single time-scale distributed estimator considering that there are no input or measurement noises. This reference provides a sufficient condition for the existence of an augmented observation model, which depends on the eigenvalues of the network Laplacian and the field dynamics matrix.

1.3 Contributions

In this section, we highlight the major challenges in distributed filtering and prediction of time-varying random fields and discuss the contributions of the thesis.

- **Time-scales of operation:** In distributed estimation, the three components of the problem framework are field dynamics, sensing, and communication. We denote the time constants of these components by τ_{dyn}, τ_{sens}, and τ_{comm}, respectively. In parameter estimation the field is almost static, i.e., $\tau_{\text{dyn}} \gg \tau_{\text{sens}}$ and $\tau_{\text{dyn}} \gg \tau_{\text{comm}}$. In dynamic field estimation, the time constants of field dynamics and sensing are comparable, $\tau_{\text{dyn}} \approx \tau_{\text{sens}}$. For this scenario, prior work on distributed Kalman filtering considered a two time-scale approach where the communication between agents is orders of magnitude faster than the dynamics and sensing, i.e., $\tau_{\text{comm}} \gg \tau_{\text{dyn}} \approx \tau_{\text{sens}}$. However, in most practical applications this is not the case; instead all the three time constants are of the same order of magnitude. To address this problem, we consider single time-scale distributed estimation where $\tau_{\text{comm}} \approx \tau_{\text{dyn}} \approx \tau_{\text{sens}}$.

- **Pseudo-innovations, pseudo-observations, and pseudo-state:** In centralized estimation, a fusion center receives the observations from all the agents and then combines them to compute the field estimates. In distributed inference, the agents do not have access to all the observations. Hence, the local distributed estimators at each agent lack the global innovations that are used by the centralized Kalman filter. Further, the dimensions of the local observations and of the local innovations are different at different agents. In contradistinction with the centralized solutions, we identify that the key components in the distributed estimators that we develop are a global average step of linearly transformed versions of one of the three quantities: innovations, observations, or state. We refer to these linearly transformed and normalized versions of the local innovations, local observations, and state as pseudo-innovations, pseudo-observations, and pseudo-state, respectively. The dimensions of the pseudo-innovations, pseudo-observations, and pseudo-state are the same at every agent.

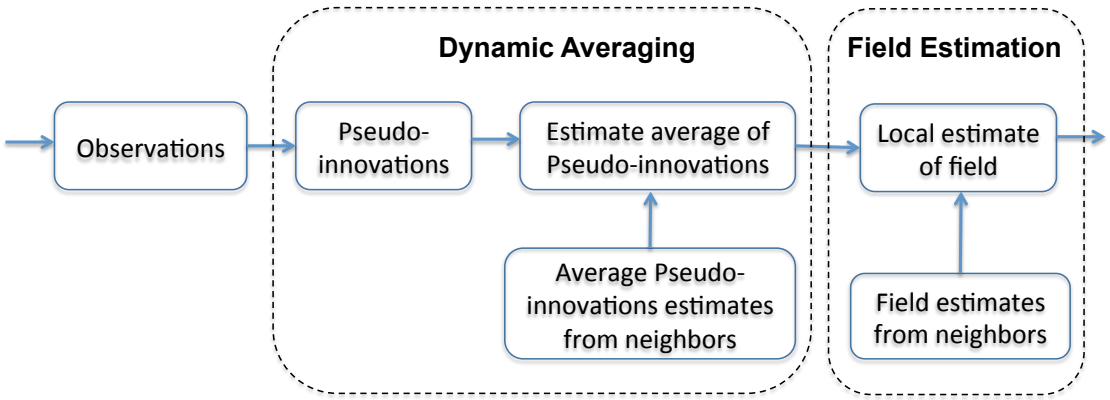
- **Dynamic averaging:** In a centralized solution, at every time iteration all the agents transmit their observations to a fusion center that can compute the average of the local pseudo-observations, pseudo-innovations, or pseudo-state. But in the distributed setup, each agent has

its own pseudo-observation/pseudo-innovations/pseudo-state and, at each time iteration, it can communicate only once with its neighbors. The distributed estimators compute the distributed estimates of the average of all the local quantities, namely, of the pseudo-observations, pseudo-innovations, or pseudo-state through a dynamic consensus step. This is similar to distributed dynamic averaging. Although distributed averaging is well-studied, [35–37], very limited literature is available when the inputs are time-varying [38–41]. In this thesis, we propose *consensus+innovations* approaches for dynamic averaging. Under the set-ups of distributed field estimation that we assume, we prove that the dynamic averaging steps yield unbiased distributed average estimates with bounded mean-squared error (MSE). The distributed field estimators use these average estimates in the filtering step to compute the field estimates.

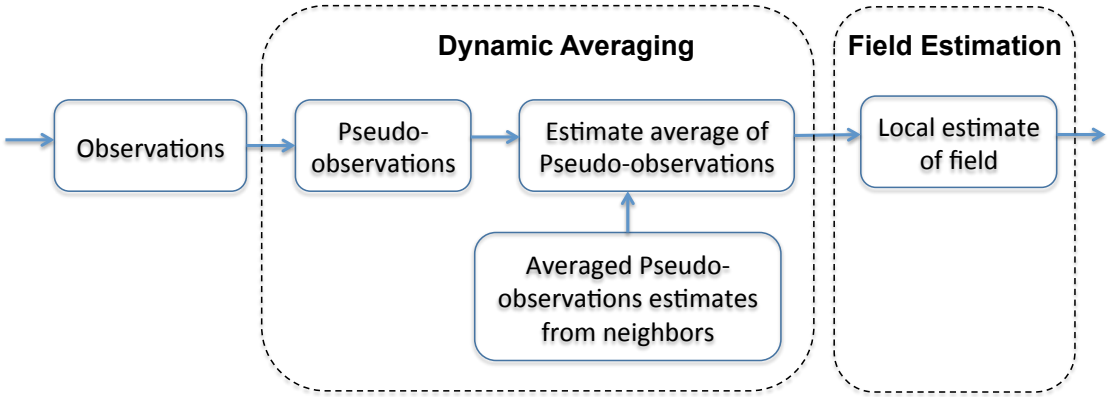
- **Distributed Kalman filter:** Apriori it is not known which of the three dynamic averaging estimators will yield better performance or what are the trade-offs in the computation complexity or what are the communications constraints of each algorithm. Hence, in this thesis, we develop three distinct distributed estimators:

  1. Pseudo-Innovations Kalman Filter (PIKF in Chapter 3) using average pseudo-innovations estimates;
  2. Distributed Information Kalman Filter (DIKF in Chapter 4) using average pseudo-observation estimates; and
  3. *Consensus+Innovations* Kalman Filter (CIKF in Chapter 5) using pseudo-state estimates.

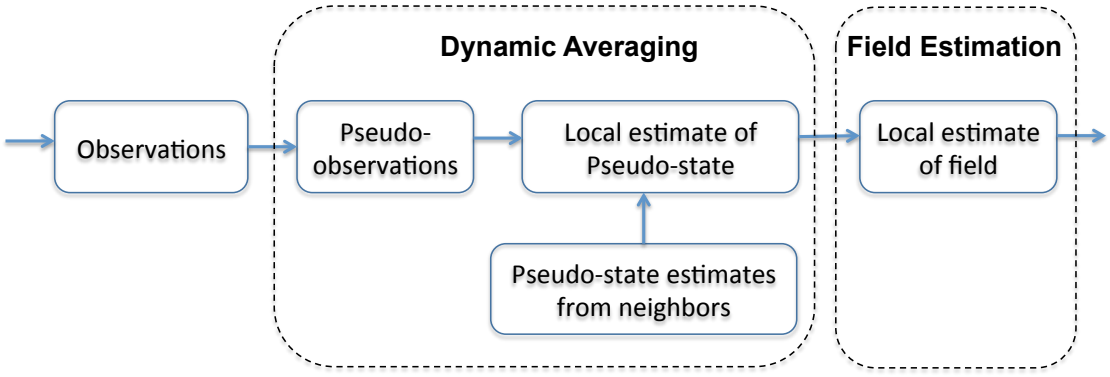
  The structure of these estimators involves two steps – a dynamic averaging step and a field estimation step. The details vary, see Fig 1.1. The structure of the dynamic averaging is *consensus+innovations* type for all the three estimators. So there are two gain parameters for dynamic averaging: a consensus gain and an innovations (pseudo-innovations/pseudo-observations/pseudo-state) gain. The structure of the field estimation step in the PIKF is also *consensus+innovations* type, i.e., it includes a consensus term and an innovations term, whereas in the DIKF and the CIKF it only includes an innovations term. In the field estimation step of the PIKF, there are again two gain parameters: a consensus gain and a



Pseudo-Innovations Kalman Filter (PIKF)



Distributed Information Kalman Filter (DIKF)



Consensus+Innovations Kalman Filter (CIKF)

Figure 1.1: Comparison between the three distributed estimators presented in the thesis: (top) PIKF, (middle) DIKF, and (bottom) CIKF.

filtering gain. With the DIKF and the CIKF, the field estimation step involves only one gain. We prove that these estimators achieve unbiased estimates with bounded MSE under the minimal assumptions of global detectability and network connectedness. We analyze all the error processes in vector form and derive the distributed version of the associated algebraic Riccati equation. The distributed algebraic Riccati equation captures the asymptotic behavior of the prediction error covariances, whose traces provide the MSE of the distributed estimators. We validate our theoretical results through numerical evaluations using Monte-Carlo simulations. We evaluate experimentally the sensitivity of these three distributed estimators to model parameters, noise statistics, gain variations, and network models. The steps of the distributed estimators, PIKF, DIKF and CIKF, are illustrated using a block diagram in Fig 1.1. The structure of these three estimators are similar but they exhibit trade-offs in terms of convergence guarantees, gain designs and MSE performance.

- **Tracking capacity:** In distributed parameter estimation and detection, global observability and mean connectedness of the network is sufficient for optimal asymptotic performance of *consensus+innovations* type algorithms [27, 28]. In two time-scale distributed estimation of dynamical fields, the agents are allowed to communicate among themselves multiple times between each observation cycle till they come to a consensus of the global average of all the local observations. Hence the two time-scale distributed Kalman filters also converge under the assumptions of global observability and mean connectedness of the network. But in this thesis, we consider single time-scale updates only, i.e., the agents are constrained to communicate with their neighbors only once between each observation of the dynamical system. In distributed estimation of dynamical fields, global observability and connectedness on average of the underlying network are sufficient for bounded MSE only for stable fields. Distributed tracking of unstable systems requires stricter conditions on the network topology and/or local observation models, in addition to global observability. References [30, 31] studied distributed state estimation of dynamical systems with Kalman filter type algorithms and introduced the concept of Network Tracking Capacity (NTC). NTC restricts the degree of instability of the dynamical systems that can be tracked by a distributed network of agents. In contrast with [30, 31], we show that the network tracking capacity of the PIKF in Chapter 3 is dependent only on the spectral gap of the network topology and is independent of the

model assumed for the local observations. For the DIKF and CIKF algorithms in Chapters 4 and 5, the NTC depends on both the connectivity of the agent communication network and on the local observation models. Hence the PIKF can still provide bounded MSE distributed estimates depending only if the network connectivity satisfies the NTC, having the flexibility of increasing the NTC just by rewiring the agent communication network.

- **Gain designs:** The distributed estimators have gains associated with each step – dynamic averaging and field estimation. The design of these gains directly impacts the MSE performance of the distributed estimation algorithms. It is imperative to choose the gain designs such that the MSE of the distributed solutions are bounded. In PIKF, we design a scalar constant-weight for the dynamic averaging consensus, a constant time-invariant gain matrix for the pseudo-innovations and a constant matrix gain for the field estimation step. These designs guaranty bounded MSE performance for the PIKF. In the dynamic averaging step of the DIKF, we consider scalar weights for the consensus and for the innovations. Then, given the distributed averaged pseudo-observation estimates, we design a filter gain matrix for the DIKF using the Gauss-Markov theorem, so that the MSE is minimized. We also provide a design with a scalar filter gain for the DIKF that shows comparable MSE performance and reduces the computation complexity significantly. In the CIKF, the consensus and innovations gain matrices for the dynamic averaging step and the gain matrix for the field estimation step are designed using the Gauss-Markov theorem so that the estimation MSE is minimized. The gain design is novel as it maintains the block-sparsity pattern similar to the graph Laplacian; these non-zero blocks are computed using the Gauss-Markov theorem. Designing the Gauss-Markov gains is particularly challenging for distributed estimators, as the gain matrices can not be full-matrices. We overcome this challenge and provide design details in Chapter 5. The DIKF provides the advantage of simpler implementation with scalar gains with a corresponding loss in performance with respect to the CIKF.

- **MSE Performance:** Developing distributed estimators for time-varying random fields ("distributed Kalman filter") has gained considerable attention over the last few years. The goal is to achieve MSE performance as close as possible to the optimal centralized Kalman filter. The design of the distributed optimal time-varying field estimator is, in general, NP-hard.

For distributed parameter estimation, references [27, 42] have shown asymptotic optimality, in the sense that the distributed parameter estimator is asymptotically unbiased, consistent, and efficient converging at the same rate as the centralized optimal estimator. However, estimation of time-varying random fields adds another degree of complication since, while information diffuses through the network, the field itself evolves. This lag causes a gap in MSE performance between the distributed and centralized field estimators. Numerical simulations show that when all these estimators meet the NTC condition, the CIKF is the best among the three in terms of MSE performance. The DIKF is 2dB better than the PIKF. The CIKF improves the performance by 3dB over the DIKF, reducing by half the gap to the centralized (optimal) Kalman filter, while showing a faster convergence rate than the DIKF.

1.4 Thesis Outline

In this section, we summarize the organization of the rest of the thesis. We detail the contents of each chapter.

- **Chapter 2: Distributed Field Estimation**

  In this chapter, we discuss the framework of distributed field estimation over multi-agent networks. The problem setup is classified into two layers: physical layer and cyber layer. The physical layer is the time-varying random field. The cyber layer consists of the network of sensors/agents. The sensors make local observations of the underlying field and they exchange their estimates with their neighbors by local communications. We state the modeling assumptions: Gaussian and uncorrelated input noise, observation noise, and initial conditions; and, the assumptions of global detectability and network connectedness that are required for bounded MSE distributed field estimation. We briefly discuss the centralized Kalman filter and the Information filter. This chapter briefly compares and contrasts the prior work on distributed estimation, both single time-scale and two time-scale, with the distributed estimators developed in this thesis.

- **Chapter 3: Pseudo-Innovations Kalman Filter**

  This chapter introduces pseudo-innovations that are linear transformations of innovations. The dimension of the local innovations are different for different agents, whereas the pseudo-

innovations of all the agents have the same dimension (equal to the dimension of the field). The Pseudo-Innovations Kalman Filter (PIKF) proposed in this chapter combines the pseudo-innovations from neighboring agents to obtain the distributed estimates of the global average of all the pseudo-innovations at each agent. Since the pseudo-innovations are themselves dynamic (time-varying), this step of the distributed estimator is akin to a dynamic averaging problem. We show, both theoretically and numerically, that the PIKF is unbiased with bounded MSE as long as the spectral norm of the field dynamics $\left(\|A\|_2\right)$ is less than the Network Tracking Capacity (NTC). The NTC depends on the diffusion rate of the agent communication network. Numerically, we show that it is possible to increase the NTC of the network by altering the communication links among the nodes.

- **Chapter 4: Distributed Information Kalman Filter**

  This chapter develops the Distributed Information Kalman Filter (DIKF) for the distributed estimation of the dynamic random field. This chapter introduces pseudo-observations, a normalized version of the observations, whose dimensions are the same for all the agents. Recall that the dimension of the local observations are different. The DIKF introduces a new *consensus+innovations* type Dynamic Consensus on Pseudo-Observations (DCPO) algorithm. The DCPO step computes the distributed estimate of the global average of the pseudo-observations, and this estimate turns out to yield more accurate field estimates than the global average of the pseudo-innovations. In the PIKF presented in Chapter 3, the NTC depends only on the network connectivity, but for the DIKF it depends on both the network connectivity and the local observation models. The MSE error performance of the DIKF is better than other contemporary single time-scale distributed estimators [24,30,31] in literature. We design the scalar consensus and innovations gain for the DCPO and the innovation gain matrices for the DIKF. In this chapter, we develop a distributed version of the algebraic Riccati equation that provides the trajectory of the asymptotic error covariance matrices. Through numerical evaluations, we validate our theoretical convergence and design claims, and demonstrate the impact of model parameters, noise, and network parameters on the performance of the DIKF.

- **Chapter 5: *Consensus+Innovations* Kalman Filter**

  In this chapter, we propose a *Consensus+Innovations* Kalman Filter (CIKF) that obtains unbiased minimized MSE distributed estimates of the pseudo-states and real-time employs them to obtain the unbiased distributed filtering and prediction estimates of the time-varying random state at each agent. The pseudo-state is a linear transformation of the dynamic state/field and it has the same dimension as the field. The filter update iterations are of the *Consensus+Innovations* type. Using the Gauss-Markov principle, we design the optimal consensus and innovation gain matrices that yield approximately 3dB improvement over the DIKF in Chapter 4. The method to design the gain matrices for distributed estimation and consensus are unique and novel. We simulate the CIKF with a large field (50 state variables) monitored by a network with 50 agents, and study its convergence characteristics.

- **Chapter 6: Conclusions**

  This chapter summarizes the theoretical and simulation results presented in the thesis. We revisit the contributions of the thesis towards the development and design of single time-scale distributed filtering and prediction algorithms for time-varying random fields. We state how this work can be used in applications and draw insights on how the thesis can lay the foundations for future research directions.

Each chapter contains related literature review, and compares the results, presented in the chapter, with existing work.

CHAPTER 2

Distributed Field Estimation

2.1 Introduction

In this chapter we introduce the problem statement for distributed estimation. We define the underlying framework for which we develop a distributed estimator. In Section 2.2, the formulation of the system, observation and communications models are presented. Section 2.2 states the modeling assumptions and the objectives. We briefly discuss the centralized estimators in Section 2.4. In Section 2.5 the existing distributed estimators are summarized. Finally, we conclude in Section 2.6.

2.2 System, Observation, and Communication

The distributed estimation framework consists of three components: dynamical system, local observation, and neighborhood communications. These three parts include two layers: the physical layer and the cyber layer. For the sake of simplicity, we motivate the model with the example of a time-varying temperature field over a large geographical area monitored by a sensor network.

2.2.1 Physical layer: dynamical system

Consider a time-varying temperature field distributed over a large geographical area, as shown in Fig 2.1. A first-order approximation and discretization of the temperature field provide spatio-temporal discretized temperature variables $x_i^j, j = 1, \cdots, M$, of M sites at discrete time indexes $i = 1, \cdots, .$ We stack the M field variables in a temperature state vector $\boldsymbol{x}_i = \left[x_i^1, \cdots, x_i^M\right]^T \in \mathbb{R}^M$. The evolution of the time-varying temperature field, $\boldsymbol{x}_i$, can be represented by a discrete time

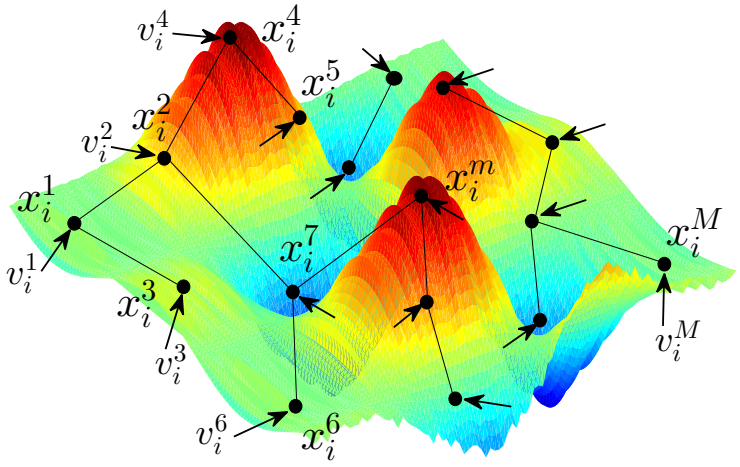


Figure 2.1: Dynamical system: random field

linear dynamical system:

$$\boldsymbol{x}_{i+1} = A\boldsymbol{x}_i + \boldsymbol{v}_i, \tag{2.1}$$

where the first-order dynamics matrix $A \in \mathbb{R}^{M\times M}$ contains the coupling effects between the M temperature variables, and the residual $\boldsymbol{v}_i = \left[v_i^1, \cdots, v_i^M\right]^T \in \mathbb{R}^M$ is the system noise driving the dynamical temperature field. At each of the M sites, the input noise $v_i^j, j = 1, \cdots, M$ accounts for the deviations in the temperature after the overall field dynamics. The field dynamics A incorporates the sparsity pattern and connectivity of the physical layer consisting of the dynamical system (2.1).

2.2.2 Cyber layer: local observations

The physical layer consisting of the field dynamics (2.1) is observed by a cyber layer consisting of a network of N agents (sensors). In Fig 2.2, we see that there are N sensors, where each sensor observes the temperatures of only a few sites. Denote the number of sites observed by sensor n by $M_n, M_n \ll M$, and its measurements at time i by $z_i^n \in \mathbb{R}^{M_n}$. The observations of the agents in the cyber layer can be represented by a linear model:

$$\boldsymbol{z}_i^n = H_n\boldsymbol{x}_i + \boldsymbol{r}_i^n, \qquad n = 1, \ldots, N, \tag{2.2}$$

where, the observation matrix $H_n \in \mathbb{R}^{M_n\times M}$ contains the observation pattern and strength information, and the observation noise $r_i^n \in \mathbb{R}^{M_n}$ reflects the inaccuracy in measurements due to sensor precision, high frequency fluctuations in temperature, and other unavoidable constraints.

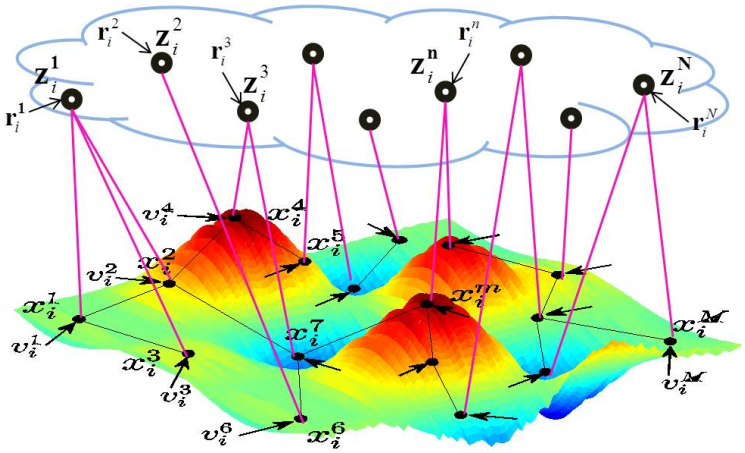


Figure 2.2: Agent network: local measurements

To illustrate, consider that sensor 1 is observing the temperature of 3 sites $\{x_i^1, x_i^2, x_i^3\}$, i.e., $M_1 = 3$. An example snapshot of the observation model of agent n is

$$\underbrace{\begin{bmatrix} z_i^{11} \\ z_i^{12} \\ z_i^{13} \end{bmatrix}}_{\boldsymbol{z}_i^1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 5 & 0 & \dots & 0 \\ 0 & 0 & 6 & \dots & 0 \end{bmatrix}}_{H_1} \underbrace{\begin{bmatrix} x_i^1 \\ \vdots \\ x_i^M \end{bmatrix}}_{\boldsymbol{x}_i} + \underbrace{\begin{bmatrix} r_i^{11} \\ r_i^{12} \\ r_i^{13} \end{bmatrix}}_{\boldsymbol{r}_i^1} = \begin{bmatrix} x_i^1 \\ 5x_i^2 \\ 6x_i^3 \end{bmatrix} + \begin{bmatrix} r_i^{11} \\ r_i^{12} \\ r_i^{13} \end{bmatrix}.$$

Now, for the ease of analysis, we aggregate the noisy local temperature measurements, $\boldsymbol{z}_i^1, \cdots \boldsymbol{z}_i^N$, of all the N agents in a global observation vector, $\boldsymbol{z}_i \in \mathbb{R}^{\sum_{n=1}^N M_n}$:

$$\underbrace{\begin{bmatrix} \boldsymbol{z}_i^1 \\ \vdots \\ \boldsymbol{z}_i^N \end{bmatrix}}_{\boldsymbol{z}_i} = \underbrace{\begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}}_{H} \boldsymbol{x}_i + \underbrace{\begin{bmatrix} \boldsymbol{r}_i^1 \\ \vdots \\ \boldsymbol{r}_i^N \end{bmatrix}}_{\boldsymbol{r}_i}, \tag{2.3}$$

where, the global observation matrix is $H \in \mathbb{R}^{\sum_{n=1}^N M_n \times M}$ and the stacked measurement noise is $r_i \in \mathbb{R}^{\sum_{n=1}^N M_n}$. Note that, in general, the temperature measurement model is non-linear. For non-linear cases, refer to distributed particle filter in [43] and the references cited therein. Here we perform a first-order approximation to obtain a linear observation sequence.

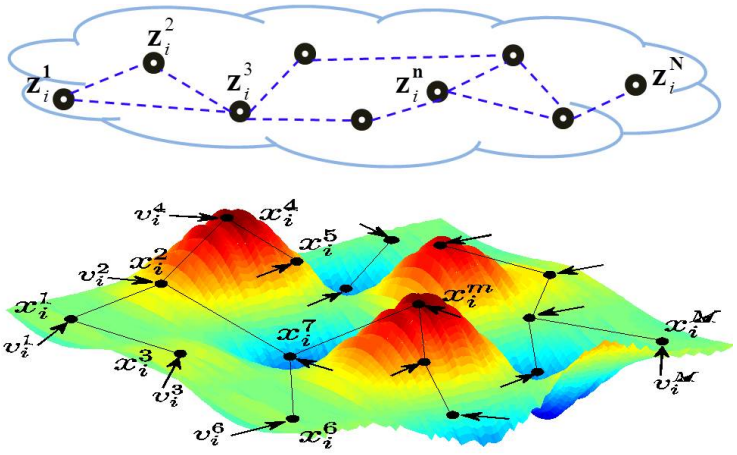


Figure 2.3: Agent network: neighborhood communications

2.2.3 Cyber layer: neighborhood communication

In the cyber layer, the agents exchange their temperature readings or current estimates with their neighbors. In many applications, to reduce communications costs, neighbors communicate only with their geographically nearest agents as shown in Fig 2.3.

Formally, the agent communication network is defined by a simple (no self-loops nor multiple edges), undirected, connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of sensors (nodes or agents) and $\mathcal{E}$ is the set of local communication channels (edges or links) among the agents. The open Ω_n and closed $\overline{\Omega}_n$ neighborhoods of agent n are:

$$\Omega_n \quad = \{l|(n,l) \in \mathcal{E}\}. \tag{2.4}$$

$$\overline{\Omega}_n \quad = n \cup \{l|(n,l) \in \mathcal{E}\}. \tag{2.5}$$

In Fig 2.3 the open and closed neighborhoods of agent 1 are $\Omega_1 = \{2,3\}$ and $\overline{\Omega}_1 = \{1,2,3\}$, respectively. The Laplacian matrix of $\mathcal{G}$ is denoted by L. The eigenvalues of the positive semi-definite matrix L are ordered as $0 = \lambda_1(L) \leq \lambda_2(L) \leq ... \leq \lambda_N(L)$. For details on graphs refer to [44]. The communication network is sparse and time-invariant.

In this thesis, we focus on developing a distributed estimator over a time-invariant network with time-invariant observation models, H_n, and time-invariant dynamics A. Future work can extend these results to make the solution robust to communication link failure, observation and/or node failures. In this section, we have laid the framework of the problem statement for the rest of this thesis. The complete problem setup is demonstrated in Fig. 2.4. The framework consists of two

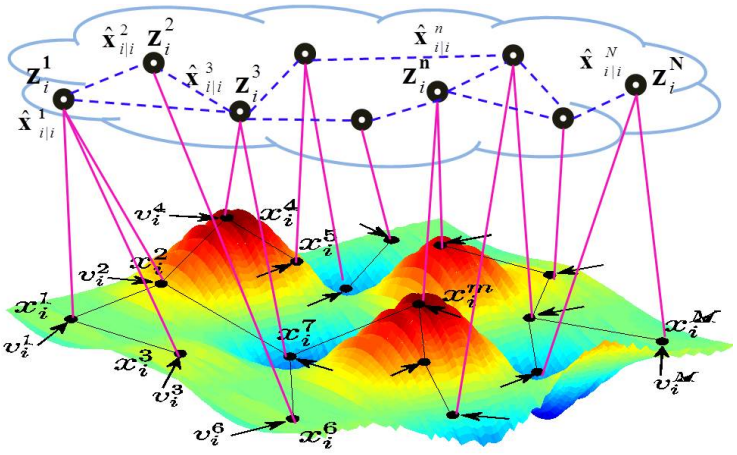


Figure 2.4: The evolution of a random field as a discrete-time linear dynamical system. The field is observed by a network of sensors (agents). The agents collaborate locally via agent communication network.

layers: (a) physical layer, consisting of the field dynamics, and (b) cyber layer, consisting of the sensor local observations and agent communication network.

2.3 Problem Statement

The goal of this work is to obtain distributed estimates of the time-varying random field $\boldsymbol{x}_i$ at each agent. The agents have access to their own observations $\boldsymbol{z}_i^n$ and exchange their estimates with theirs neighbors. In this section we state the objective of the distributed estimation, and the assumptions on the model.

2.3.1 Modeling assumptions

In this thesis, we make three assumptions on the model framework for the purpose of field estimation. The assumptions are formally stated as:

Assumption 1 (Gaussian processes)**.** *The system noise, $\boldsymbol{v}_i$, the observation noise, $\boldsymbol{r}_i$, and the initial condition of the system, $\boldsymbol{x}_0$, are Gaussian sequences, with*

$$\boldsymbol{v}_i \sim \mathcal{N}(\bar{0}, V), \quad \boldsymbol{r}_i^n \sim \mathcal{N}(\bar{0}, R_n), \quad \boldsymbol{x}_0 \sim \mathcal{N}(\bar{\boldsymbol{x}}_0, \Sigma_0),$$

where, $V \in \mathbb{R}^{M\times M}$, $R_n \in \mathbb{R}^{M_n\times M_n}$ and $\Sigma_0 \in \mathbb{R}^{M\times M}$ are the corresponding covariance matrices. The noise covariance matrix R of the global noise vector $\boldsymbol{v}_i$ in (2.3) *is block-diagonal, i.e., $R =$ blockdiag$\{R_1, \ldots, R_N\}$, and positive-definite, i.e., $R > 0$.*

Assumption 2 (Uncorrelated sequences)**.** *The system noise, the observation noise, and the initial condition:* $\{\{\boldsymbol{v}_i\}_i, \{\boldsymbol{r}_i\}_i, \boldsymbol{x}_0\}_{i\geq 0}$ *are uncorrelated random vector sequences.*

Assumption 3 (Prior information)**.** *Each agent in the cyber layer knows the system dynamics model,* A *and* V*, the initial condition statistics,* $\bar{x}_0$ *and* Σ_0*, the parameters of the observation model,* H *and* R*, and the communication network model,* $\mathcal{G}$*.*

Assumption 1 characterize the properties of the noises in the system and observations, and the initial conditions. For Gaussian noise with non-zero means, the problem setup can be reduced to a zero-mean case by simple linear transformation and all the results hold true. However, for distributed estimation under non-Gaussian noises refer to distributed particle filtering approaches in [45, 46] and the references cited therein. In this thesis, by Assumption 2 we consider that the noises and the initial conditions are uncorrelated. In the scenario of correlated noises, the derivations can be modified for distributed estimation as explained in [47]. In large-scale system applications, the dynamics, observation, and network Laplacian matrices, A, H_n, and L, are sparse, with $M_n \ll M$, and the agents communicate with only a few of their neighbors, $|\Omega_n| \ll N, \forall n$.

In the dynamics (2.1) and observations (2.2), we assume that there is no deterministic input.

Proposition 1. *It is sufficient to design a filter without any input and in the case of non-zero inputs the filter estimates are corrected with the known non-zero mean of the field.*

Proof. It is sufficient to design an estimator considering no deterministic inputs and then we can extend it to non-zero input case by adapting the following approach. Consider a discrete-time system and observation model:

$$\boldsymbol{x}_{i+1} = A\boldsymbol{x}_i + B\boldsymbol{u}_i + \boldsymbol{v}_i$$
$$\boldsymbol{z}_i = C\boldsymbol{x}_i + D\boldsymbol{u}_i + \boldsymbol{r}_i$$

where, i is the time index; $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are the state and observation of the system; the matrices A, B, C, D and the input sequence $\boldsymbol{u}_i$ are known; and, the mean and covariance of the random initial condition x_0 and random noise sequences v_i, r_i are known. The objective is to obtain the estimates, $\widehat{\boldsymbol{x}}_{i+1}$, of $\boldsymbol{x}_{i+1}$ given the observations $\boldsymbol{z}_i$. Now consider another discrete-time system and observation model where the inputs are zero, i.e., $B\boldsymbol{u}_i = 0$ and $D\boldsymbol{u}_i = 0$,

$$\overline{\boldsymbol{x}}_{i+1} = A\overline{\boldsymbol{x}}_i + \boldsymbol{v}_i$$

$$\overline{\boldsymbol{z}}_i = C\overline{\boldsymbol{x}}_i + \boldsymbol{r}_i$$

Then, we can show that,

$$\boldsymbol{x}_{i+1} = \overline{\boldsymbol{x}}_{i+1} + \sum_{j=0}^{i} A^j B \boldsymbol{u}_{i-j} \tag{2.6}$$

$$\overline{\boldsymbol{z}}_i = \boldsymbol{z}_i - \left(C \sum_{j=0}^{i-1} A^j B \boldsymbol{u}_{i-j-1} + D\boldsymbol{u}_i \right). \tag{2.7}$$

Let us assume that we have an algorithm (say, Algorithm XYZ) to compute the estimates, $\widehat{\overline{\boldsymbol{x}}}_{i+1}$, of $\overline{\boldsymbol{x}}_{i+1}$ given the observations $\overline{\boldsymbol{z}}_i$ when there are no inputs in the system. When there are inputs in the system, i.e., $B\boldsymbol{u}_i \neq 0$ and/or $D\boldsymbol{u}_i \neq 0$, then we can use the same algorithm to compute the estimates, $\widehat{\boldsymbol{x}}_{i+1}$, of $\boldsymbol{x}_{i+1}$ given the observations $\boldsymbol{z}_i$, by performing the following steps (Note that A, B, C, D and $\boldsymbol{u}_i$ are known):

Step 1: Given $\boldsymbol{z}_i$, compute $\overline{\boldsymbol{z}}_i$ using equation (2.7) stated above.

Step 2: Using Algorithm XYZ and the observations $\overline{\boldsymbol{z}}_i$, compute the estimates, $\widehat{\overline{\boldsymbol{x}}}_{i+1}$.

Step 3: Compute $\widehat{\boldsymbol{x}}_{i+1}$ from $\widehat{\overline{\boldsymbol{x}}}_{i+1}$ using the following equation,

$$\widehat{\boldsymbol{x}}_{i+1} = \widehat{\overline{\boldsymbol{x}}}_{i+1} + \sum_{j=0}^{i} A^j B \boldsymbol{u}_{i-j}.$$

□

Thus the results are readily extended if there is a known deterministic input. In Subsection 2.3.2, we formally state the objectives of distributed estimation and the underlying assumptions required to achieve those objectives.

2.3.2 Unbiased Estimates with bounded MSE

At time i, denote the n^{th} agent's distributed filter and prediction estimates of the state $\boldsymbol{x}_i$ by $\widehat{\boldsymbol{x}}^n_{i|i}$ and $\widehat{\boldsymbol{x}}^n_{i+1|i}$ respectively. Denote the filtering and prediction error processes $\boldsymbol{\epsilon}^n_{i|i}$ and $\boldsymbol{\epsilon}^n_{i+1|i}$ of the the state $\boldsymbol{x}_i$ at agent n by

$$\boldsymbol{\epsilon}^n_{i|i} = \boldsymbol{x}_i - \widehat{\boldsymbol{x}}^n_{i|i}, \tag{2.8}$$

$$\boldsymbol{\epsilon}^n_{i+1|i} = \boldsymbol{x}_{i+1} - \widehat{\boldsymbol{x}}^n_{i+1|i}. \tag{2.9}$$

The objective is to design distributed estimators such that the distributed estimates of the state at each agent are unbiased with bounded mean squared error (MSE), i.e.,

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}^{n}\right] = \mathbb{E}\left[\boldsymbol{\epsilon}_{i+1|i}^{n}\right] = \mathbf{0}, \qquad n = 1, \cdots, N, \tag{2.10}$$

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}^{n}\boldsymbol{\epsilon}_{i|i}^{n^T}\right] = \Sigma_{i|i}^{n} < \infty, \quad \mathbb{E}\left[\boldsymbol{\epsilon}_{i+1|i}^{n}\boldsymbol{\epsilon}_{i+1|i}^{n^T}\right] = \Sigma_{i+1|i}^{n} < \infty, \qquad n = 1, \cdots, N. \tag{2.11}$$

The unbiasedness of the estimates are reflected in (2.10) and (2.11) ensures that the distributed estimates converge with bounded MSE. To achieve unbiased distributed estimates with bounded MSE we make the following assumptions:

Assumption 4 (Global detectability)**.** *The dynamic state equation* (2.1) *and the observations model* (2.3) *are globally detectable, i.e., the pair* (A, H) *is detectable.*

Assumption 5 (Connectedness)**.** *The agent communication network is connected, i.e., the algebraic connectivity* $\lambda_2(L)$ *of the Laplacian matrix* L *of the graph* $\mathcal{G}$ *is strictly positive.*

By Assumption 4, the state-observation model (2.1)-(2.3) is globally detectable but not necessarily locally detectable, i.e., $(A, H_n), \forall n,$ are not necessarily detectable. Note that these two are minimal assumptions. Assumption 4 is mandatory even for a centralized system, and Assumption 5 is required for consensus algorithms to converge. In Section 2.4, we review centralized estimator followed by a overview of existing distributed estimators of time-varying random fields in Section 2.5.

2.4 Centralized Estimators

Although not practical in the context of the problem we study, we use the centralized information filter to benchmark our results on distributed estimator. In a centralized scheme, all the agents in the cyber layer communicate their measurements to a central fusion center, as depicted in Fig 2.5. The fusion center performs all needed computation tasks.

2.4.1 Centralized Kalman filter

At time i, denote the centralized filter and prediction estimates of $\boldsymbol{x}_i$ and the centralized filter and prediction error covariance matrices as $\widehat{\boldsymbol{x}}_{i|i}^{c}$, $\widehat{\boldsymbol{x}}_{i+1|i}^{c}$, $\Sigma_{i|i}^{c}$, and $\Sigma_{i+1|i}^{c}$, respectively. The filtering and prediction equations of the centralized Kalman filter are:

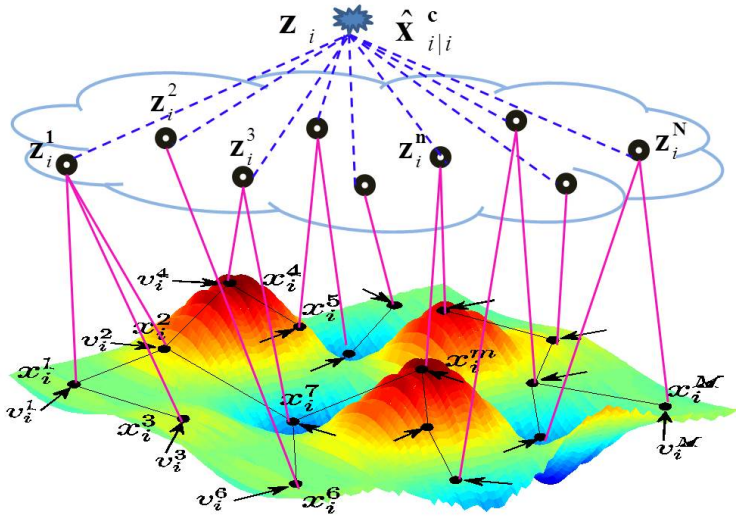


Figure 2.5: Centralized estimator

Step 1: *Filtering*

$$K_i^c = \Sigma_{i|i-1}^c H^T \left(H\Sigma_{i|i-1}^c H^T + R \right)^{-1} \tag{2.12}$$

$$\widehat{\boldsymbol{x}}_{i|i}^c = \widehat{x}_{i|i-1}^c + K_i^c \left(\boldsymbol{z}_i - H\widehat{\boldsymbol{x}}_{i|i-1}^c \right) \tag{2.13}$$

$$\Sigma_{i|i}^c = \Sigma_{i|i-1}^c - K_i^c \left(H\Sigma_{i|i-1}^c H^T + R \right) K_i^{c^T} \tag{2.14}$$

Step 2: *Prediction*

$$\widehat{\boldsymbol{x}}_{i+1|i}^c = A\widehat{\boldsymbol{x}}_{i|i}^c \tag{2.15}$$

$$\Sigma_{i+1|i}^c = A\Sigma_{i|i}^c A^T + V \tag{2.16}$$

where: K_i^c is the centralized Kalman gain; the initial conditions are $\widehat{x}_{0|-1}^c = \overline{x}_0$ and $\Sigma_{0|-1}^c = \Sigma_0$. The centralized Kalman filter (2.12)-(2.16) provides unbiased estimates of x_i with miminum MSE under the Assumption 4 of global detectability. The MSE is the trace of the error covariance matrix.

2.4.2 Centralized Information filter

Recall the information form [47] of the centralized Kalman filter (CKF). The filtering and prediction equations of the information form of the centralized Information filter are:

Step 1: *Global averaging*

$$\overline{z}_i = \frac{1}{N} \sum_{n=1}^{N} H_n^T R_n^{-1} z_i^n \tag{2.17}$$

Step 2: *Filtering*

$$K_i^c = \left(\frac{1}{N}(\Sigma_{i|i-1}^c)^{-1} + G\right)^{-1} \tag{2.18}$$

$$\widehat{x}_{i|i}^c = \widehat{x}_{i|i-1}^c + K_i^c\left(\overline{z}_i - G\widehat{x}_{i|i-1}^c\right) \tag{2.19}$$

$$\Sigma_{i|i}^c = \left(I_M - K_i^c G\right)\Sigma_{i|i-1}^c\left(I_M - GK_i^c\right) + \frac{1}{N}K_i^c G K_i^c \tag{2.20}$$

Step 3: *Prediction*

$$\widehat{x}_{i+1|i}^c = A\widehat{x}_{i|i}^c \tag{2.21}$$

$$\Sigma_{i+1|i}^c = A\Sigma_{i|i}^c A^T + V \tag{2.22}$$

where: K_i^c is the centralized gain; the initial conditions are $\widehat{x}_{0|-1}^c = \overline{x}_0$ and $\Sigma_{0|-1}^c = \Sigma_0$; and I_M denotes an identity matrix of dimension $M \times M$. The centralized Kalman filter (2.17)-(2.22) provides unbiased estimates of x_i with miminum MSE. Note that, in the information form of the Kalman filter, the global average of a linearly transformed version of the observations are used instead of the observations. The performance of the information filter is equivalent to that of the Kalman filter that receives the entire measurement vector z_i. The performance equivalence of the two centralized filters can be shown using Matrix Inversion Lemma [48]. However, the information filter has computational advantages and similarity to distributed approaches.

2.5 Distributed Estimators

We review related prior research on distributed estimation of time-varying random fields. We classify prior work into two categories based on the time-scales of operation: (a) two time-scale and (b) single time-scale.

2.5.1 Two time-scale

In two-time scale distributed estimators, see Fig 2.6, agents exchange their information multiple number of times between each dynamics/observations time-scale. Here, the linear distributed estimator reproduces a Kalman filter locally at each sensor with a consensus step on the observations. This is a two time scale algorithm because the sensing and the local Kalman filter updates occur at the same slow time scale of the process dynamics, while the consensus iterations happen at the fast time scale of communications among sensors. References [13–21] represent variations on this

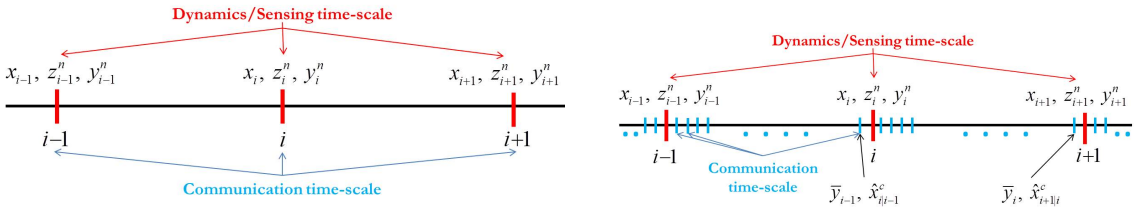


Figure 2.6: Time-scales of operation of dynamics, sensing and communications: single and two time-scales.

structure. Reference [15] modifies the local Kalman filters to account for the distributed nature of the field so that the local Kalman filters are all different since they operate on different dynamics (*local* dynamics). References [19, 20] assume local observability, i.e., the dynamical system is observable at each sensor, which is not practical in most distributed large-scale applications. The two time scale distributed estimator in [21] considers time-varying dynamics and observation matrices and expresses the measurements in information filter form. As a final comment, we may also include in this class [22, 23] that assume that the communication network is complete. Since every agent communicates with every other, every sensor acts as a centralized fusion center and the consensus step is trivial. In [23] the nodes broadcast only one bit based on the sign of the innovations.

2.5.2 Single time-scale

In this class, sensing and communication occur once at every time step; the local estimators operate at the same time scale of the process dynamics. Operating at a single time scale reduces the onerous communication rounds required by the consensus step in the two time step estimators. The trade-offs between different versions of these single time step algorithms are between the convergence rate and the asymptotic MSE of the estimators. Reference [24] proposed a *gossip Kalman filter*, a distributed estimator where each agent replicates a (local) Kalman filter processing its local observations. At every sensing time step, there is a communication round between a single pair of connected sensors, chosen according to the gossip protocol, [25], when these two sensors exchange their current state estimate and their current error covariance matrix provided by the local Kalman filter Riccati equation. The reference shows through large deviations theory that, under appropriate conditions, asymptotically, each local estimate is unbiased and has bounded covariance, regardless of the degree of instability of the field dynamics. Reference [26] improved the convergence rate of [24] by allowing several pairs, rather than a single pair, of connected sensors

to exchange their states. Although the gossip Kalman filters can track any unstable field, it may pay a price on the communications (every round exchanges not only the state estimate but also the error covariance, an object on the order of M^2 that can be very large for spatially distributed fields), on the convergence rate, and on higher asymptotic MSE with respect to the *consensus+innovations* distributed estimator that we present here.

The *consensus+innovations* estimators, the second class of single time scale distributed estimators, extend to time varying dynamics the consensus+innovations distributed estimators introduced in [27, 28] for parameter estimation; see also [29]. The consensus+innovations distributed linear parameter estimator updates at every sensor the local state estimate by a consensus term on the current estimate and by the current local innovations–the difference between the current sensor observation and its local prediction. A number of approaches [30–33] have been developed to extend this structure to time varying fields. They add a term to the consensus+innovations distributed parameter estimator to account for the field time dynamics. In [30, 31], the innovation update at each sensor collects its own observations and those of its neighbors. This reference shows that the distributed estimate at every sensor is asymptotically unbiased and, if the degree of instability of the field dynamics is appropriately upper bounded (the bound defined as the Network Tracking Capacity (NTC)), the distributed field estimator has bounded MSE. This NTC depends on the connectivity of the sensor network and on the observation model. Further, [32, 33] developed structural system properties that the sensor network and the observation models satisfy such that the consensus+innovations type distributed estimator in [30] has bounded MSE performance.

2.6 Conclusions

In this thesis we are interested in single time-scale solution. Developing distributed estimators for time-varying random fields ("distributed Kalman filter") has gained considerable attention over the last few years. The goal has been to achieve MSE performance as close as possible to the optimal centralized Kalman filter. The distributed optimal time-varying field estimation is, in general, NP-hard. The distributed parameter estimation [27, 42] have shown asymptotic optimality, in the sense that the distributed parameter estimator is asymptotically unbiased, consistent, and efficient converging at the same rate as the centralized optimal estimator. However, estimation of time-varying random fields adds another degree of complication since, while information diffuses

through the network, the field itself evolves. So, this lag causes a gap in performance between distributed and centralized field estimators. In this thesis, we develop distributed estimators that converge to a bounded MSE solution requiring minimal assumptions, namely global detectability and connected network.

CHAPTER 3

Pseudo-Innovations Kalman Filter

3.1 Introduction

In this chapter, we present Pseudo-Innovations Kalman Filter (PIKF), a single time-scale distributed estimator for time-varying random fields. We extend a *consensus+innovations* type algorithm for parameter estimation, as in [27] and [28], to linear dynamic fields. We define a modified version of the innovations, pseudo-innovations, that are different from the way defined in [49] and [50]. The PIKF consists of simultaneous computation of two companion distributed estimates: the field estimate and the pseudo-innovations estimate at each sensor. The pseudo-innovations estimate is a local distributed estimate of global average of all the pseudo-innovations. This part is similar to the dynamic averaging problem addressed in [41].

We show that the Pseudo-Innovations Kalman Filter (PIKF) is asymptotically unbiased and its mean-squared error (MSE) is bounded, as long as a particular eigen-ratio of the network topology is above a lower bound. In other words, the Network Tracking Capacity (NTC), as defined in [30] and [31], of the estimator depends only on the connectivity of the sensor network and is independent of the local measurement models (provided the dynamic field is globally observable). However in [13], [14] and [15], the agents are allowed to communicate among themselves multiple times between each observation cycle till they come to a consensus of global average of all the local observations. Under such circumstances the NTC of the distributed Kalman filter is infinity. But here we consider single time-scale updates only, i.e., the agents are constrained to communicate with their neighbors only once between each observation of the dynamical system.

The rest of the chapter is organized as follows. In Section 3.2 we state the preliminaries and

present the PIKF. Section 3.3 analyze the error processes of the estimator proposed in Section 3.2. We study the Network Tracking Capacity (NTC) of the estimator and provide designs of estimator gain matrices so that it converges with bounded MSE in Section 3.4. The simulation results presented in Section 3.5 establish our claims on the convergence of the PIKF and visually demonstrates the impact of the NTC on the performance of the estimator. Finally we conclude in Section 3.6.

3.2 Distributed Kalman Filter

The information form of the centralized Kalman filter, [47], has access to all the local innovations and computes their scaled average. Since the local innovations of the agents can have different dimensions, we normalize the local innovations by pre-multiplying them with the inverse of the observation noise covariance matrices and then pre-multiply the normalized innovations with the transpose of the observation matrix so that all of them have same the dimension M. We term these normalized form of the innovations of same dimension as pseudo-innovations, different from the un-normalized pseudo-innovations defined in [49] and [50]. The distributed Kalman filter has two companion estimation processes running in parallel - dynamical state estimation and dynamic global pseudo-innovations average estimation as considered in [51].

3.2.1 Notation

In this chapter, we represent the dynamical system $\theta(i)$ which evolves with discrete time i as:

$$\theta(i+1) = A\theta(i) + v(i) \tag{3.1}$$

where: $A \in \mathbb{R}^{M\times M}$ is the system matrix and $v(i) \in \mathbb{R}^M$ is the system noise. $v(i)$ is white Gaussian noise, i.e., $v(i) \sim \mathcal{N}(0, V)$. The initial condition of the state vector is $\theta(0) \sim \mathcal{N}(\bar{\theta}_0, \Sigma_0)$.

The local observations of the n^{th} agent, denoted by $z_n(i) \in \mathbb{R}^{M_n}$, are:

$$z_n(i) = H_n\theta(i) + r_n(i) \tag{3.2}$$

where: $H_n \in \mathbb{R}^{M\times M_n}$ is the observation matrix and $r_n(i)$ is the observation noise. The observation noises are also white Gaussian noises, i.e., $r_n(i) \sim \mathcal{N}(0, R_n)$. Note that each agent observes only a fraction $M_n \ll M$ of the dynamic state vector $\theta(i)$. We assume that the system noise, $v(i)$, observation noises, $r_n(i)$, and the initial condition, $\theta(0)$, of the state vector are statistically independent. The system and observation model satisfies Assumptions 1-3.

3.2.2 Distributed Algorithm

Denote $x_n(i)$ as the field estimate and $\widehat{\nu}_n(i)$ as the pseudo-innovations estimate at sensor n at time i given the measurements up to time $i-1$. Each sensor implements the following steps:

1. *Pseudo-innovations measurement:*

$$\nu_n(i) = H_n^T R_n^{-1}\Big(z_n(i) - H_n \sum_{l\in\Omega_n} w_{nl} x_l(i)\Big) \tag{3.3}$$

2. *Updates:*

$$\widehat{\nu}_n(i+1) = \sum_{l\in\Omega_n} w_{nl}\widehat{\nu}_l(i) + \Big(\nu_n(i) - C_n \sum_{l\in\Omega_n} w_{nl}\widehat{\nu}_l(i)\Big) \tag{3.4}$$

$$x_n(i+1) = A\Big(\sum_{l\in\Omega_n} w_{nl} x_l(i) + K_n(i)\widehat{\nu}_n(i+1)\Big) \tag{3.5}$$

where: $W = \{w_{nl}\}$ is the consensus weight matrix with the same sparsity as the graph Laplacian L, $K_n(i)$ is the local estimator gain matrix, and C_n is the local pseudo-innovations gain matrix. The initial conditions are $x_n(0) = \bar{\theta}_0$ and $\widehat{\nu}_n(0) = \nu_n(0), \forall n$. In vector notation, we combine equations (3.3) - (3.5) as:

$$\nu(i) = D_H^T R^{-1}\Big(z(i) - D_H(W\otimes I_M)x(i)\Big) \tag{3.6}$$

$$\widehat{\nu}(i+1) = (W\otimes I_M)\widehat{\nu}(i) + \Big(\nu(i) - C(W\otimes I_M)\widehat{\nu}(i)\Big) \tag{3.7}$$

$$x(i+1) = (I_N\otimes A)\Big((W\otimes I_M)x(i) + K(i)\widehat{\nu}(i+1)\Big) \tag{3.8}$$

where,

$$x(i) = \begin{bmatrix} x_1(i) \\ \vdots \\ x_N(i) \end{bmatrix},\ z(i) = \begin{bmatrix} z_1(i) \\ \vdots \\ z_N(i) \end{bmatrix},\ \widehat{\nu}(i) = \begin{bmatrix} \widehat{\nu}_1(i) \\ \vdots \\ \widehat{\nu}_N(i) \end{bmatrix},$$

the blockdiagonal matrices are

$$\begin{aligned} D_H &= \text{blockdiag}\{H_1, ..., H_N\} \\ R^{-1} &= \text{blockdiag}\{R_1^{-1}, ..., R_N^{-1}\} \end{aligned}$$

Algorithm 1 Pseudo-Innovations Kalman Filter (PIKF)

Input: Model parameters A, V, H, R, G, L, $\overline{\theta}_0$, Σ_0.
Initialize: $x_n(0) = \bar{\theta}_0$ and $\widehat{\nu}_n(0) = \nu_n(0)$.
Pre-compute: Gain matrices W, C_n and $K_n(i)$ using (3.30)-(3.34), (3.43).
while $i \geq 0$ **do**
Communications:
Broadcast $x_n(i)$ and $\widehat{\nu}_n(i)$ to all neighbors $l \in \Omega_n$.
Receive $\{x_l(i), \widehat{\nu}_l(i)\}_{l\in\Omega_n}$ from neighbors.
Observation:
Make measurement $z_n(i)$ of the state $\theta(i)$.
Transform $z_n(i)$ in pseudo-innovation $\nu_n(i)$ using (3.3).
Prediction updates:
Compute the estimate $\widehat{\nu}_n(i+1)$ of $\nu_{\text{avg}}(i)$ using (3.4).
Compute the estimate $x_n(i+1)$ of the state $\theta(i+1)$ using (3.5).
end while

$$\begin{aligned} K(i) &= \text{blockdiag}\{K_1(i), ..., K_N(i)\} \\ C &= \text{blockdiag}\{C_1, ..., C_N\} \end{aligned}$$

and $\otimes$ denotes Kronecker matrix product. This Pseudo-Innovations Kalman Filter (PIKF) is a modified version of the algorithm proposed in [49] and [50]. The two key properties that an estimate should satisfy are unbiasedness and bounded error. In this chapter we will show that the proposed PIKF is asymptotically unbiased and the mean-squared error of estimation is bounded. First we formulate and analyze the different errors of estimation, then design the different parameters in the PIKF such that it satisfy the desired two key properties.

3.3 Error Analysis

The proposed Pseudo-Innovations Kalman Filter (PIKF) has two companion update equations, (3.7) and (3.8). These kinds of algorithms inhibit the conventional way of error analysis, that is why we do not have a standalone equation for the error process. So we define the averaged or consensus estimate sequence and the centralized estimate sequence. Then we will show that the proposed PIKF (3.7) converges to the averaged or consensus estimator and the averaged estimator converges to the centralized estimator. This multistep methodology is introduced in [27]. The averaged linear

estimator is:

$$x_{\text{avg}}(i) = \frac{1}{N}\sum_{n=1}^{N} x_n(i) = \frac{1}{N}(1_N \otimes I_M)^T x(i). \tag{3.9}$$

Similarly, global dynamic average of the local pseudo-innovations of N agents $\nu_{\text{avg}}(i)$ and its estimate $\widehat{\nu}_{\text{avg}}(i)$ sequences are:

$$\nu_{\text{avg}}(i) = \frac{1}{N}\sum_{n=1}^{N} \nu_n(i) = \frac{1}{N}(1_N \otimes I_M)^T \nu(i), \tag{3.10}$$

$$\widehat{\nu}_{\text{avg}}(i) = \frac{1}{N}\sum_{n=1}^{N} \widehat{\nu}_n(i) = \frac{1}{N}(1_N \otimes I_M)^T \widehat{\nu}(i). \tag{3.11}$$

Define the error process $e^x_{\text{avg}}(i)$ as the error between the distributed field estimates $x(i)$ and the averaged estimates $x_{\text{avg}}(i)$

$$e^x_{\text{avg}}(i) = x(i) - 1_N \otimes x_{\text{avg}}(i) \tag{3.12}$$

and the pseudo-innovations error process $e^{\widehat{\nu}}_{\text{avg}}(i)$ as the error between the distributed pseudo-innovations estimates $\widehat{\nu}(i)$ and the averaged estimates $\widehat{\nu}_{\text{avg}}(i)$,

$$e^{\widehat{\nu}}_{\text{avg}}(i) = \widehat{\nu}(i) - 1_N \otimes \widehat{\nu}_{\text{avg}}(i). \tag{3.13}$$

Define

$$P_{_{NM}} = \frac{1}{N}(1_N \otimes I_M)(1_N \otimes I_M)^T \tag{3.14}$$

and note that

$$P_{_{NM}} x(i) = 1_N \otimes x_{\text{avg}}(i), \tag{3.15}$$

$$P_{_{NM}} \widehat{\nu}(i) = 1_N \otimes \widehat{\nu}_{\text{avg}}(i), \tag{3.16}$$

$$P_{_{NM}}(W \otimes I_M) = P_{_{NM}}. \tag{3.17}$$

The error process $e^x_{\text{avg}}(i)$ evolves as:

$$\begin{aligned}
&e^x_{\text{avg}}(i+1) = x(i+1) - 1_N \otimes x_{\text{avg}}(i+1) \\
&= (I_N \otimes A)\Big((W \otimes I_M)x(i) + K(i)\widehat{\nu}(i+1) - (W \otimes I_M)P_{_{NM}}x(i) - K(i)P_{_{NM}}\widehat{\nu}(i+1)\Big) \\
&= (I_N \otimes A)\Big((W \otimes I_M - P_{_{NM}})(x(i) - 1_N \otimes x_{\text{avg}}(i)) + K(i)(\widehat{\nu}(i+1) - 1_N \otimes \widehat{\nu}_{\text{avg}}(i+1))\Big)
\end{aligned}$$

$$= (I_N \otimes A)\left((W \otimes I_M - P_{_{NM}}) e^{x}_{\text{avg}}(i) + K(i) e^{\widehat{\nu}}_{\text{avg}}(i+1) \right) \tag{3.18}$$

and the error process $e^{\widehat{\nu}}_{\text{avg}}(i)$ evolves as:

$$\begin{aligned} e^{\widehat{\nu}}_{\text{avg}}(i+1) &= \widehat{\nu}(i+1) - 1_N \otimes \widehat{\nu}_{\text{avg}}(i+1) \\ &= (W \otimes I_M)\widehat{\nu}(i) + \Big(\nu(i) - C(W \otimes I_M)\widehat{\nu}(i)\Big) - P_{_{MN}}(W \otimes I_{_M})\widehat{\nu}(i) + \Big(P_{_{MN}}\nu(i) \\ &\qquad - C P_{_{MN}}(W \otimes I_{_M})\widehat{\nu}(i)\Big) \\ &= (I_{MN} - C)(W \otimes I_M - P_{_{NM}})(\widehat{\nu}(i) - 1_N \otimes \widehat{\nu}_{\text{avg}}(i)) + \Big(\nu(i) - P_{_{NM}}\nu(i)\Big) \\ &= (I_{_{MN}} - C)(W \otimes I_{_M} - P_{_{MN}}) e^{\widehat{\nu}}_{\text{avg}}(i) + \Big(I_{_{MN}} - P_{_{MN}}\Big)\nu(i) \end{aligned} \tag{3.19}$$

Note that the estimate $x_n(i)$ at the n^{th} sensor is the expected value of $\theta(i)$ given its observations and neighbors estimates upto time index $i-1$. Therefore the pseudo-innovations $\nu(i)$ is a zero-mean process. So far we have studied the error processes between the distributed estimates and the averaged estimates. Now we study the error processes between the averaged estimates and centralized estimates. The objective is to design the estimator in such a way that the averaged field estimates $x_{\text{avg}}(i)$ converge to the centralized estimates $x_c(i)$,

$$x_c(i+1) = A\Bigg(x_c(i) + K_c(i)\frac{1}{N}\sum_{n=1}^{N} H_n^T R_1^{-1}\Big(z_n(i) - H_n x_c(i)\Big)\Bigg) \tag{3.20}$$

and the averaged pseudo-innovations estimate $\widehat{\nu}_{\text{avg}}(i)$ converge to the global average, $\nu_{\text{avg}}(i)$, of all the pseudo-innovations. Define $e^x_c(i)$ as the error process between the averaged field estimates $x_{\text{avg}}(i)$ and the centralized estimates $x_c(i)$.

$$e^x_c(i) = 1_N \otimes x_{\text{avg}}(i) - 1_N \otimes x_c(i) \tag{3.21}$$

and $e^{\widehat{\nu}}_c(i)$ as the error process between the averaged pseudo-innovations estimates $\widehat{\nu}_{\text{avg}}(i)$ and the global average of all the pseudo-innovations $\nu_{\text{avg}}(i)$,

$$e^{\widehat{\nu}}_c(i+1) = 1_N \otimes \widehat{\nu}_{\text{avg}}(i+1) - 1_N \otimes \nu_{\text{avg}}(i). \tag{3.22}$$

The error process $e^x_c(i)$ evolves as:

$$e^x_c(i+1) = 1_N \otimes x_{\text{avg}}(i+1) - 1_N \otimes x_c(i+1)$$

$$
\begin{aligned}
&= (I_N \otimes A)\Big(P_{NM}x(i) + K(i)P_{NM}\widehat{\nu}(i+1) - 1_N \otimes x_c(i) \\
&\qquad -(I_N \otimes K_c(i))P_{NM}D_H^T R^{-1}(z(i) - D_H x_c(i))\Big) \\
&= (I_N \otimes A)\Bigg(e_c^x(i) + K(i)(I - C)P_{NM}(W \otimes I)\widehat{\nu}(i) + K(i)P_{NM}\overline{D}_H \\
&\qquad \times \Big(1_N \otimes (x_c(i) - x_{\text{avg}}(i)) + (1_N \otimes x_{\text{avg}}(i) - (W \otimes I)x(i))\Big)\Bigg) \\
&= (I_N \otimes A)\Bigg(\Big(I_{MN} - K(i)(I_N \otimes G)\Big)e_c^x(i) \\
&\qquad + K(i)P_{NM}\Big(\widehat{\nu}(i+1) - \nu(i)\Big) - K(i)P_{NM}\overline{D}_H(W \otimes I)e_{\text{avg}}^x(i)\Bigg) \\
&= (I_N \otimes A)\Bigg(\Big(I_{MN} - K(i)(I_N \otimes G)\Big)e_c^x(i) + K(i)e_c^{\widehat{\nu}}(i+1) \\
&\qquad - K(i)P_{NM}\overline{D}_H(W \otimes I)e_{\text{avg}}^x(i)\Bigg). \qquad (3.23)
\end{aligned}
$$

Note that here we choose $K_n(i) = K_c(i),\ \forall n$, and $\overline{D}_H = D_H^T R^{-1} D_H = \text{blkdiag}\{H_1^T R_1^{-1} H_1, \ldots, H_N^T R_N^{-1} H_N\}$. We discuss the design of the gain matrices later in Section 3.4. Now we study the evolution of the error process $e_c^{\widehat{\nu}}(i)$,

$$
\begin{aligned}
e_c^{\widehat{\nu}}(i+1) &= 1_N \otimes \widehat{\nu}_{\text{avg}}(i+1) - 1_N \otimes \nu_{\text{avg}}(i) \\
&= \Big(I - C\Big)(W \otimes I)P_{NM}\widehat{\nu}(i) + P_{NM}\nu(i) - P_{NM}\nu(i) \\
&= \Big(I - C\Big)P_{NM}\Big(\widehat{\nu}(i) - \nu(i-1)\Big) + \Big(I - C\Big)P_{NM}\nu(i-1) \\
&= \Big(I - C\Big)e_c^{\widehat{\nu}}(i) + \Big(I - C\Big)P_{NM}\nu(i-1). \qquad (3.24)
\end{aligned}
$$

The error process that is still left to be derived is the one between the centralized field estimates $x_c(i)$ and the field dynamics $\theta(i)$. Denote this error process by $e_\theta^c(i)$,

$$
\begin{aligned}
e_\theta^c(i+1) &= 1_N \otimes x_c(i+1) - 1_N \otimes \theta(i+1) \\
&= (I_N \otimes A)\Big(e_\theta^c(i) + (I_N \otimes K_c(i))P_{NM}D_H^T R^{-1}(z(i) - D_H x_c(i))\Big) - 1_N \otimes v(i) \\
&= (I_N \otimes A)\Big(e_\theta^c(i) + (I_N \otimes K_c(i))P_{NM}\overline{D}_H e_\theta^c(i)\Big)
\end{aligned}
$$

$$
\begin{aligned}
& & & +(I_{_N} \otimes A)(I_N \otimes K_c(i))P_{_{NM}} D_H^T R^{-1} r(i) - 1_N \otimes v(i) \\
&= (I_N \otimes A)\left(I_{MN} - I_N \otimes \Big(K_c(i) G \Big) \right) e_\theta^c(i) & & \\
& & & + \underbrace{(I_{_N} \otimes A)(I_{_N} \otimes K_c(i))P_{_{NM}} D_H^T R^{-1} r(i) - 1_{_N} \otimes v(i)}_{\delta(i)} \qquad (3.25)
\end{aligned}
$$

where, G is

$$
G = \frac{1}{N} \sum_{n=1}^{N} H_n^T R_n^{-1} H_n. \qquad (3.26)
$$

The noise process $\delta(i)$, which is a linear combination of zero-mean field noise $v(i)$ and measurement noises $r(i)$, is also a zero-mean noise. Now we assemble all the error processes together to analyze the overall stability of the PIKF. Combining equations (3.18), (3.19), (3.23), (3.24) and (3.26),

$$
\underbrace{\begin{bmatrix} e_\theta^c(i+1) \\ e_c^x(i+1) \\ e_{\text{avg}}^x(i+1) \\ e_{\text{avg}}^{\widehat{\nu}}(i+1) \\ e_c^{\widehat{\nu}}(i+1) \end{bmatrix}}_{\tilde{e}(i+1)} = \tilde{A}(i) \underbrace{\begin{bmatrix} e_\theta^c(i) \\ e_c^x(i) \\ e_{\text{avg}}^x(i) \\ e_{\text{avg}}^{\widehat{\nu}}(i) \\ e_c^{\widehat{\nu}}(i) \end{bmatrix}}_{\tilde{e}(i)} + \eta(i) \qquad (3.27)
$$

where, $\tilde{A}(i)$ and $\eta(i)$ are

$$
\tilde{A}(i) = \begin{bmatrix}
(I \otimes A)\Big(I - I \otimes (K_c(i)G)\Big) & 0 & 0 & 0 & 0 \\
0 & (I \otimes A)\Big(I - K(i)(I \otimes G)\Big) & -(I \otimes A)K(i)P\overline{D}_H & 0 & (I \otimes A)K(i)(I - C) \\
0 & 0 & (I \otimes A)(W \otimes I - P) & (I \otimes A)K(i)(I - C)(W \otimes I - P) & 0 \\
0 & 0 & 0 & (I - C)(W \otimes I - P) & 0 \\
0 & 0 & 0 & 0 & (I - C)
\end{bmatrix} \qquad (3.28)
$$

$$
\eta(i) = \begin{bmatrix}
\delta(i) \\
(I \otimes A)K(i)\Big(I - C\Big)P_{_{NM}}\nu(i-1) \\
(I \otimes A)K(i)(I - P_{_{NM}})\nu(i) \\
(I - P_{_{NM}})\nu(i) \\
\Big(I - C\Big)P_{_{NM}}\nu(i-1)
\end{bmatrix} \qquad (3.29)
$$

Note that all the elements in the noise process $\eta(i)$ are zero-mean. Therefore the stability of the PIKF depends on the error matrix $\tilde{A}(i)$. For the error processes to be stable, the spectral radius of $\tilde{A}(i)$ should be less than unity. Note that $\tilde{A}(i)$ is a block triangular matrix. So the design parameters: consensus weight matrix, estimator gain matrices, and the pseudo-innovations gain matrices are required to be designed such that the spectral radii of each of the diagonal blocks of $\tilde{A}(i)$ are less than unity.

3.4 Estimator Design

We design the estimator gain matrices $K_n(i)$, the pseudo-innovations gain matrices C_n and the consensus weight matrix W such that the spectral radii of each of the diagonal blocks of $\tilde{A}(i)$ are strictly less than one.

3.4.1 Gain Matrices

In this problem we assume that the dynamical system $\theta(i)$ is distributedly observable, i.e., the matrix G in (4.5) is invertible. This assumption ensures that each of the state variables of the system $\theta(i)$ is being observed by at least one of the agents. Under this assumption we design the local estimator gain matrix $K_n(i)$ at agent n as:

$$K_n(i) = K_c(i) = G^{-1}, \qquad \forall n. \tag{3.30}$$

Equation (3.30) makes the spectral radii of the matrices A_{11} and A_{22} strictly less than one. Similarly we design the pseudo-innovations gain matrices C_n as:

$$C_n \;=\; \frac{1}{\|G^{-1}\|_2} G^{-1}, \qquad \forall n \tag{3.31}$$

which makes the spectral radius of A_{55} to be strictly less than one. The matrix G, and hence G^{-1}, is symmetric positive definite. Therefore the eigenvalues of C_n lie in the range $(0, 1]$. Now we discuss another way of designing the gain matrices $K_n(i)$. The gain matrices premultiplies the pseudo-innovation estimates of each sensor. At each sensor the pseudo-innovation estimate initializes with the local pseudo-innovation and then eventually converges to the global average of the pseudo-innovations. To maintain the parity between the sparsity of the gain matrix and the pseudo-innovation estimate at each sensor, we design the initial values and the update equations for

the gain matrices $K_n(i)$ at each sensor n as:

$$K_n(0) = \left(H_n^T R_n^{-1} H_n\right)^+ \tag{3.32}$$

$$K_n(i+1) = \left(\sum_{l\in\Omega_n} w_{nl} K_l^+(i)\right)^+ \tag{3.33}$$

where F^+ is the Moore-Penrose pseudo-inverse of the matrix F as defined in [52]. Note that the gain matrices at each sensor converges to G^{-1} as time progresses. Also note that all the gain matrices can be pre-computed and saved at each agent. This will reduce the computational complexity while running the algorithm.

Now we discuss the design of the consensus weight matrix W. Here we consider W to be uniform-weight, i.e.,

$$W = I_N - \beta L \tag{3.34}$$

where, β is the consensus weight. For stable systems, the choice of β is straight-forward. However, for unstable dynamical systems there may not always exist a β such that the PIKF converges with bounded MSE solution. In the next Subsection 3.4.2, we study the conditions under which the PIKF converges and the corresponding design of the consensus weight β.

3.4.2 Network Tracking Capacity

The design of the consensus weight depends on the agent communication network. The properties of the network are represented by its graph Laplacian matrix L. Since the graph Laplacian L is a positive semi-definite matrix with $\lambda_2(L) > 0$ for connected networks, we define the diffusion rate of the network γ as:

$$\gamma = \frac{\lambda_2(L)}{\lambda_N(L)} \tag{3.35}$$

Let a denote the spectral norm $||A||_2$ of the system matrix A, which is a measure of the degree of stability or instability of the dynamical system (3.1). To be able to design the consensus weight matrix W such that the spectral radius of A_{33} is less than one, we require an upper bound on the degree of instability of the system which results in the following theorem.

Theorem 3.4.1. *If the spectral norm of the system matrix (A) is:*

$$a = \left\|A\right\|_2 < \frac{1+\gamma}{1-\gamma} \tag{3.36}$$

then there exists a consensus weight

$$\beta \in \Big(\frac{a-1}{a\lambda_2(L)}, \frac{a+1}{a\lambda_N(L)}\Big) \tag{3.37}$$

such that

$$\rho(A_{33}) = \rho\Big((I_N \otimes A)(W \otimes I_M - P_{NM})\Big) < 1 \tag{3.38}$$

where, $\rho(F)$ denotes the spectral radius of F.

Proof. Start with the left-hand side of (3.38),

$$\begin{aligned}
\rho\Big((I_N \otimes A)(W \otimes I_M - P_{NM})\Big) &\leq \Big\|(I_N \otimes A)(W \otimes I_M - P_{NM})\Big\|_2 \\
&\leq \Big\|I_N \otimes A\Big\|_2 \Big\|W \otimes I_M - P_{NM}\Big\|_2 \\
&= \Big\|A\Big\|_2 \Big\|I_N - \beta L - J\Big\|_2 \quad \text{where,} \quad J = \frac{1}{N} 1_N 1_N^T \\
&= a.\max\{|1-\beta\lambda_2(L)|, |\beta\lambda_N(L) - 1|\}
\end{aligned} \tag{3.39}$$

Replacing γ in (3.36) by (3.35), we have $\frac{a-1}{a\lambda_2(L)} < \frac{a+1}{a\lambda_N(L)}$. Now choose β such that,

$$\frac{a-1}{a\lambda_2(L)} < \beta < \frac{a+1}{a\lambda_N(L)} \tag{3.40}$$

Use (3.40) to analyze the first absolute term in (3.39),

$$\begin{aligned}
a.|1-\beta\lambda_2(L)| &= a.\max\big\{(1-\beta\lambda_2(L)), (\beta\lambda_2(L)-1)\big\} \\
&< a.\max\left\{\Big(1 - \frac{a-1}{a\lambda_2(L)}\lambda_2(L)\Big), \Big(\frac{a+1}{a\lambda_N(L)}\lambda_2(L) - 1\Big)\right\} \\
&< a.\max\left\{\frac{1}{a}, \Big(\frac{a+1}{a\lambda_N(L)}\lambda_N(L) - 1\Big)\right\} \\
&= 1
\end{aligned} \tag{3.41}$$

Similarly, analyze the second absolute term in (3.39),

$$\begin{aligned}
a.|\beta\lambda_N(L) - 1| &= a.\max\big\{(\beta\lambda_N(L)-1), (1-\beta\lambda_N(L))\big\} \\
&< a.\max\left\{\Big(\frac{a+1}{a\lambda_N(L)}\lambda_N(L) - 1\Big), \Big(1 - \frac{a-1}{a\lambda_2(L)}\lambda_N(L)\Big)\right\} \\
&< a.\max\left\{\frac{1}{a}, \Big(1 - \frac{a-1}{a\lambda_2(L)}\lambda_2(L)\Big)\right\}
\end{aligned}$$

$$= \quad 1 \tag{3.42}$$

(3.39), (3.41) and (3.42) implies that if (3.36) holds true, then there exist β satisfying (3.37) such that

$$\rho\Big((I_N \otimes A)(W \otimes I_M - P_{NM})\Big) < 1.$$

□

Thus (3.38) along with (4.39) make the spectral radii of A_{33} and A_{55} less than one. To achieve this we have a upper bound on the degree of instability of the dynamical system as in (3.36). This bound is the Network Tracking Capacity of the distributed Kalman filter. Note that the NTC, which is a function of the diffusion rate of the agent communication network, depends only on the network topology and is independent of the local observation patterns in contrast to [30] and [31] where it depends on both. Also, regarding the choice of the consensus weight note that (3.40) is sufficient, but the optimal choice of β is:

$$\begin{aligned} 1 - \beta^* \lambda_2(L) &= \beta^* \lambda_N(L) - 1 \\ \text{i.e., } \beta^* &= \frac{2}{\lambda_2(L) + \lambda_N(L)}. \end{aligned} \tag{3.43}$$

3.4.3 Network Models

Now we discuss about designing the network topology such that (3.36) is satisfied. From (3.36), we see that the lower bound on the diffusion rate of the network depends on the degree of instability of the dynamical system (3.1) as:

$$\gamma > \frac{1-a}{1+a} \qquad \text{where,} \quad a = \Big\|A\Big\|_2 \tag{3.44}$$

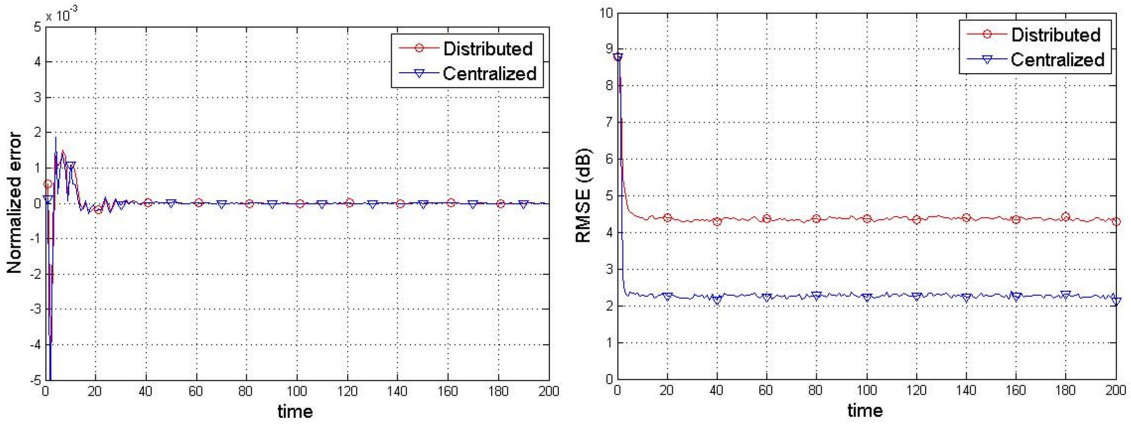
For a given dynamical system, if a particular network topology fails to satisfy (3.44) then rewire the network connections to increase the diffusion rate γ so that it satisfies (3.44). For example, a lattice graph has a lower diffusion rate as compared to Erdös-Renýi or Watts-Strogatz graphs with equal number of nodes and edges. However, in [53] it is proved that for a fixed number of edges in the network there exists a class of expander graphs, called Ramanujan Graphs, which maximizes the eigenratio γ. In the following section, we simulate the distributed Kalman filter algorithm to compare the MSE for different models and see how the NTC affect the performance of the algorithm.

3.5 Simulation Results

We simulate, using Matlab, a discrete-time linear dynamical system with $M = 10$ number of state variables. The system matrix A is generated randomly. A group of $N = 10$ agents observe the dynamical system, where each agent observes only two state variables, i.e., $M_n = 2$. The covariance matrices of the system noise V, the observation noises R and the initial condition Σ_0 are randomly generated and then squared to ensure that they are positive definite. The mean of the initial condition $\bar{\theta}_0$ is also randomly generated. The initial condition of the field vector is chosen from $\theta(0) \sim \mathcal{N}(\bar{\theta}_0, \Sigma_0)$, the evolution noise at each time index are chosen from $v(i) \sim \mathcal{N}(0, V)$ and the global measurement noise $r(i)$ is produced from $\mathcal{N}(0, R)$. From the global measurement noise, the local measurement noises $r_n(i)$ are fed into each sensor to obtain the noise corrupted measurements. The system and observation model is such that it is not locally observable but it is globally observable and distributedly observable, i.e., G is invertible.

3.5.1 Unbiasedness and Bounded MSE Convergence

First, we evaluate the convergence properties of the Pseudo-Innovations Kalman Filter (PIKF) and benchmark its performance with respect to the centralized filter. The sensor network is considered to be a regular lattice graph with neighborhood size $= 4$. The diffusion rate of the network is $\gamma = 0.0802$. The values of γ and $\|A\|$ satisfy the assumption (3.36). We chose the consensus weight $\beta = 0.3135$, that satisfies (3.37). We compare the distributed results with the centralized estimator where the gain matrices are designed using the principles from the Kalman filter. These estimators with the design parameters, as stated above, are then simulated up to time index $(i = 200)$. We have done Monte-Carlo simulations using Matlab by repeating the algorithm 1000 times to obtain the estimation error. The simulation plots are shown in Fig. 3.1. We see that the PIKF is asymptotically unbiased since the expected value of the normalized error converges to zero asymptotically. Also we see that the MSE of the PIKF is bounded. The distributed case shows a loss of approximately 2dB with respect to the centralized estimator, in steady state condition, due to the fact that the centralized estimator has access to all the random time-varying pseudo-innovations, whereas the PIKF relies only on its own pseudo-innovations and cooperation from its neighbors.



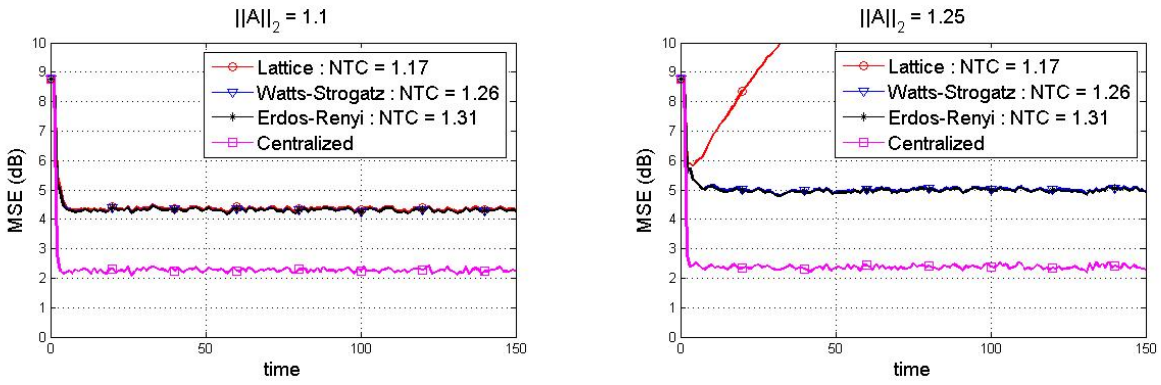
a Unbiased estimates

b Bounded MSE estimates

Figure 3.1: Normalized error and root mean-squared error (RMSE in dB) plots of distributed and centralized estimators.

3.5.2 Network dependencies

First we consider a symmetric system matrix with $||A||_2 = 1.1$, which is obtained by scaling the singular values of the previously randomly generated matrix A, using singular value decomposition. Then we generate a lattice graph, an Erdös-Renýi graph and a Watts-Strogatz graph with $N = 10$ nodes and $E = 17$ edges. The NTC, computed using (3.36), of these three graphs are 1.17, 1.26 and 1.31 respectively. We simulate the centralized Kalman filter and the distributed Kalman filter with these three network models upto 200 time iterations. The mean-squared errors (MSE) in dB are computed in each of these cases by 1000 Monte-Carlo simulations as shown in Fig. 3.2(a). Then we repeated the above simulation after just increasing the spectral norm to $||A||_2 = 1.25$ and the plots are shown in Fig. 3.2(b). Fig. 3.2(a) shows that the distributed Kalman filters, with network models having NTC greater than the spectral norm of the system dynamics, have bounded MSE with a performance gap of nearly 2dB as compared to the centralized filter. When we increase the degree of instability of the system dynamics such that it is greater than the NTC of the lattice network but smaller than the NTC of the Erdös-Renýi and Watts-Strogatz networks, we see that the MSE is unbounded for the lattice network but are bounded for the other two networks as depicted in Fig. 3.2(b). Hence for a given unstable dynamical system, if a network model fails to satisfy the NTC condition then it can be rewired to increase the diffusion rate of the network such that it satisfies NTC condition to obtain a distributed Kalman filter with bounded MSE.



a Below NTC b Above NTC

Figure 3.2: MSE of centralized and distributed Kalman filters with different graphical models when: (a). $||A||_2 = 1.1$, and, (b). $||A||_2 = 1.25$

3.6 Conclusions

In this chapter, we proposed a distributed estimator, Pseudo-Innovations Kalman Filter (PIKF), for estimation of a linear discrete-time field dynamics. We have considered a *consensus+innovations* type algorithm as introduced in [27] and [28]. This is a single time-scale update algorithm, i.e., between successive evolutions of the system dynamics the sensors in the distributed network can communicate among themselves only once. Our results are based on two important assumptions: the model is globally observable, not necessarily locally observable, and the degree of instability of the dynamics is upper bounded by a function of the diffusion rate of the network. The degree of instability of the dynamics is independent of the specifics of the measurement models. We empirically demonstrated that for a given set of agents with a fixed number of communication links, the NTC of the network can be increased by rewiring the edges of the network. We showed theoretically and illustrated by simulation that the PIKF is unbiased and its mean-squared error remains bounded.

In the next chapter, we develop a single time scale distributed information Kalman filter (DIKF). For the DIKF, we obtain an analytical NTC based on the eigenvalues of the network Laplacian and the local observation matrices, whereas in [30, 31] the NTC is computed by solving a large optimization problem. In contrast to consensus and gain matrices designed for the scalar case, in DIKF, we present a generalized analytical design of the estimator gain matrices following the first principles of Kalman filter and develop a distributed version of the algebraic Riccati equation. The

proposed DIKF accomplishes this by operating a companion consensus step on a modified local observation, the pseudo-observation. The DIKF is asymptotically unbiased and improves MSE performance by 2dB over the PIKF.

CHAPTER 4

Distributed Information Kalman Filter

4.1 Introduction

In Chapter 2, we proposed a single time-scale *consensus+innovations* distributed estimator, Pseudo-Innovations Kalman Filter (PIKF), that performs consensus on the state estimates and introduced an additional consensus on a modified version of the innovations. References [49–51, 54] showed that the distributed estimator is unbiased with bounded MSE, but, in contrast to [30], their NTC depends only on the degree of connectivity of the sensor network and is independent of the observation models.

In this chapter, we propose a single time scale distributed information Kalman filter (DIKF) to obtain unbiased estimates with bounded MSE for unstable systems. In contrast to [30, 31], we derive the NTC based on the eigenvalues of the network Laplacian and the local observation matrices. We develop a general analytical design of the estimator gain matrices. We achieve this by introducing a novel *consensus+innovations* type dynamic averaging step that runs in parallel with the DIKF. The DIKF achieves better asymptotic MSE performance than the PIKF proposed in Chapter 2. We confirm the theoretical results by extensive Monte-Carlo simulation. We draw further insights on how the performance of the proposed solution is affected by varying model parameters, noise statistics, gain constants, and network structure.

4.2 Preliminaries

The key component of the DIKF is the role of the global average of the pseudo-observations in the distributed estimation and prediction of time-varying fields. In this section, we introduce

pseudo-observations and discuss the additional conditions required for the convergence of the DIKF.

4.2.1 Pseudo-observations

The dimensions M_n of the local observations, z_i^n, in (2.2) are different for different agents. We introduce the pseudo-observations, y_i^n, all of which have the same dimension M,

$$y_i^n = H_n^T R_n^{-1} z_i^n, \quad n = 1, \ldots, N. \tag{4.1}$$

Note that the pseudo-observations y_i^n are linearly transformed (by H_n^T) and normalized (by R_n^{-1}) versions of the observations z_i^n. Pseudo-observations are the observations expressed in information filter form, the information vectors. Let, $y_i = [(y_i^1)^T, \ldots, (y_i^N)^T]^T$. Then, in vector form,

$$y_i = D_H^T R^{-1} z_i \tag{4.2}$$

where $D_H = \text{blockdiag}\{H_1, \ldots, H_N\}$. The vector $y_i \in \mathbb{R}^{MN}$. The averaged pseudo-observations, $\overline{y}_i$, are

$$\overline{y}_i = \frac{1}{N} \sum_{n=1}^{N} y_i^n = \frac{1}{N} \sum_{n=1}^{N} H_n^T R_n^{-1} z_i^n \tag{4.3}$$

$$= Gx_i + \frac{1}{N} H^T R^{-1} r_i \tag{4.4}$$

where G is:

$$G = \frac{1}{N} \sum_{n=1}^{N} H_n^T R_n^{-1} H_n = \frac{1}{N} H^T R^{-1} H. \tag{4.5}$$

In a centralized solution, at every time iteration i, all the agents transmit their observations, $z_i^n, n = 1, \ldots, N$, to a fusion center that can compute the averaged pseudo-observations $\overline{y}_i$. But in the distributed setup, each agent has its own observation z_i^n (or pseudo-observation y_i^n) and, at each time iteration i, it can communicate only once with its neighbors. Our distributed estimator computes distributed estimates, $\widehat{y}_i^n$, of the averaged pseudo-observations $\overline{y}_i$ through dynamic consensus on the pseudo-observations, y_i^n. This is akin to the challenge of distributed dynamic averaging. Although distributed averaging is well-studied, [35], very limited literature is available when the inputs are time-varying [38–41]. In this paper, we propose a *consensus+innovations* approach, the *Dynamic Consensus on Pseudo-Observations*, for dynamic averaging.

4.2.2 Distributed Observability

For the DIKF, in addition to the Assumptions 4-5 of global detectability and connected network, we make an additional assumption of distributed observability, as stated below.

Assumption 6 (Distributed observability)**.** *The system is distributedly observable, i.e., the matrix G is full rank.*

Assumption 6 ensures that each of the state variables of the time-varying field $\boldsymbol{x}_i$ is being observed by at least one of the sensors in the agent network. This extends weak detectability defined in [24] and is similar to the distributed observability in [27] and [28]. This assumption is crucial for the convergence of the DIKF with bounded MSE.

Before we proceed further, we define the following matrices:

$$A_{_G} = GAG^{-1} \tag{4.6}$$

$$H_{_G}^n = H_n^T R_n^{-1} H_n G^{-1} \tag{4.7}$$

$$\overline{D}_H = D_H^T R^{-1} D_H \tag{4.8}$$

$$J = 1_N 1_N^T \tag{4.9}$$

where: A is the system matrix; G and D_H are in (4.5) and (4.2), respectively; and the N-dimensional vector $1_N = [1, \ldots, 1]^T$. In the following Section 4.3, we present the DIKF algorithm and discuss its memory, communication and computational costs.

4.3 Distributed Estimator

In this section we describe the novel single time scale distributed information Kalman filter (DIKF) presented in [55]. The distributed information Kalman filter (DIKF) is a *consensus+innovations* algorithm [29] operating at single time scale. Each agent aims to compute a distributed estimate of the averaged pseudo-observations $\overline{y}_i$ for distributed filtering. Each agent runs two companion sub-routines: 1) dynamic consensus on the pseudo-observations (DCPO) to estimate $\overline{y}_i$; and 2) distributed filtering to estimate the states x_i.

4.3.1 Notation

The n^{th} agent's distributed filter and prediction estimates of x_i are $\widehat{x}^n_{i|i}$ and $\widehat{x}^n_{i+1|i}$, respectively. The distributed filter error $\epsilon^n_{i|i}$ and the distributed prediction error $\epsilon^n_{i+1|i}$ are:

$$\begin{aligned} \epsilon^n_{i|i} &= \widehat{x}^n_{i|i} - x_i, \\ \text{and,} \qquad \epsilon^n_{i+1|i} &= \widehat{x}^n_{i+1|i} - x_{i+1}. \end{aligned}$$

The corresponding distributed filter and prediction error covariance matrices are $\Sigma^n_{i|i}$ and $\Sigma^n_{i+1|i}$, respectively. We define the distributed pseudo-observation error q^n_i as

$$q^n_i = \widehat{y}^n_i - Gx_i.$$

The distributed pseudo-observations error covariance matrix is Q^n_i. The distributed prediction estimate $\widehat{x}^n_{i|i-1}$ and pseudo-observations estimate $\widehat{y}^n_i$ are statistically dependent. Let Π^n_i be the covariance between the error processes $\epsilon^n_{i|i-1}$ and q^n_i. We collect all the distributed estimates in a vector and define

$$\widehat{x}_{i|i} = \begin{bmatrix} \widehat{x}^1_{i|i} \\ \vdots \\ \widehat{x}^N_{i|i} \end{bmatrix}, \ \widehat{x}_{i+1|i} = \begin{bmatrix} \widehat{x}^1_{i+1|i} \\ \vdots \\ \widehat{x}^N_{i+1|i} \end{bmatrix}, \ \widehat{y}_i = \begin{bmatrix} \widehat{y}^1_i \\ \vdots \\ \widehat{y}^N_i \end{bmatrix}.$$

Using (4.10)-(4.10), define the error processes in vector form

$$\begin{aligned} \epsilon_{i|i} &= \widehat{x}_{i|i} - 1_N \otimes x_i \\ \epsilon_{i+1|i} &= \widehat{x}_{i+1|i} - 1_N \otimes x_{i+1} \\ q_i &= \widehat{y}_i - 1_N \otimes (Gx_i) \end{aligned}$$

where $\otimes$ denotes Kronecker product of vectors and matrices. The corresponding global filter, prediction, and pseudo-observations error covariance matrices are $\Sigma_{i|i}$, $\Sigma_{i+1|i}$, and Q_i respectively. The covariance between $\epsilon_{i|i-1}$ and q_i is Π_i. The error matrices $\Sigma_{i|i}$, $\Sigma_{i+1|i}$, Q_i, and Pi_i are block matrices consisting of $N \times N$ blocks, where each block is $M \times M$.

4.3.2 Distributed Information Kalman Filter Algorithm

At time i, agents communicate their pseudo-observations estimates $\widehat{y}^n_{i-1}$ to neighbors and implement the steps:

Step 1: *Dynamic consensus on pseudo-observations (DCPO)*

$$\widehat{y}_i^n = A_{_G} \sum_{l \in \Omega_n} w_{nl} \widehat{y}_{i-1}^l + B \left(y_i^n - H_{_G}^n A_{_G} \widehat{y}_{i-1}^n \right) \tag{4.10}$$

$$Q_{i+1} = F Q_i F^T + \Psi \tag{4.11}$$

Step 2: *Filtering*

$$K_i^n = \left(\left(Q_i^n + (\Pi_i^n)^T G \right) \left(\Sigma_{i|i-1}^n G + \Pi_i^n \right)^{-1} + G \right)^{-1} \tag{4.12}$$

$$\widehat{x}_{i|i}^n = \widehat{x}_{i|i-1}^n + K_i^n (\widehat{y}_i^n - G \widehat{x}_{i|i-1}^n) \tag{4.13}$$

$$\begin{aligned} \Sigma_{i|i} = {} & \left(I_{_{MN}} - K_i (I_{_N} \otimes G) \right) \Sigma_{_{i|i-1}} \left(I_{_{MN}} - (I_{_N} \otimes G) K_i^T \right) + K_i Q_i K_i^T + \left(I_{_{MN}} - K_i (I_{_N} \otimes G) \right) \Pi_i K_i^T \\ & + K_i \Pi_i^T \left(I_{_{MN}} - (I_{_N} \otimes G) K_i^T \right) \end{aligned} \tag{4.14}$$

Step 3: *Prediction*

$$\widehat{x}_{i+1|i}^n = A \widehat{x}_{i|i}^n \tag{4.15}$$

$$\Sigma_{i+1|i} = (I_N \otimes A) \, \Sigma_{i|i} \left(I_N \otimes A^T \right) + J \otimes V \tag{4.16}$$

$$\begin{aligned} \Pi_{i+1} = {} & \left(I_{_N} \otimes A \right) \left(I_{_{MN}} - K_i (I_{_N} \otimes G) \right) \Pi_i F^T + (I_N \otimes A) \, K_i Q_i F^T \\ & + (J \otimes V) \left(I_N \otimes G - \overline{D}_H \left(I_N \otimes B^T \right) \right) \end{aligned} \tag{4.17}$$

where: where: $W = [w_{nl}]$ is a stochastic consensus weight matrix; K_i^n are the estimator gain matrices; $K_i = \text{blockdiag}\{K_i^1, \ldots, K_i^N\}$; and B is the pseudo-observations gain matrix. The error matrix F and the noise covariance matrix Ψ are:

$$\begin{aligned} F &= W \otimes A_{_G} - (I_N \otimes B) \overline{D}_H (I_N \otimes A G^{-1}) \\ \Psi &= \left((I_N \otimes B) \overline{D}_H - I_N \otimes G \right) (J \otimes V) \left(\overline{D}_H (I_N \otimes B^T) - I_N \otimes G \right) + (I_N \otimes B) \overline{D}_H (I_N \otimes B^T), \end{aligned}$$

and the initial conditions are:

$$\begin{aligned} \Sigma_{0|-1} &= J \otimes \Sigma_0 \\ Q_0 &= \left((I_N \otimes B) \overline{D}_H - I_N \otimes G \right) (J \otimes \Sigma_0) \left(\overline{D}_H (I_N \otimes B^T) - I_N \otimes G \right) + (I_N \otimes B) \overline{D}_H (I_N \otimes B^T) \\ \Pi_0 &= (J \otimes \Sigma_0) \left(I_N \otimes G - \overline{D}_H \left(I_N \otimes B^T \right) \right), \end{aligned}$$

We term the distributed dynamic averaging (4.10) as Dynamic consensus on pseudo-observations in [55]. The initial conditions are:

$$\widehat{x}_{0|-1}^n = \overline{x}_0$$

Algorithm 2 Distributed Information Kalman Filter

Input: Model parameters A, V, H, R, G, L, $\overline{\boldsymbol{x}}_0$, Σ_0.
Initialize: $\widehat{\boldsymbol{x}}^n_{0|-1} = \overline{\boldsymbol{x}}_0$, $A_G\widehat{y}^n_{-1} = G\overline{x}_0$.
Pre-compute: Gain matrices W, B_i and K_i using Algorithm 3.
while $i \geq 0$ **do**
Communications:
Broadcast $\widehat{\boldsymbol{y}}^n_{i-1}$ to all neighbors $l \in \Omega_n$.
Receive $\{\widehat{\boldsymbol{y}}^l_{i-1}\}_{l\in\Omega_n}$ from neighbors.
Observation:
Make measurement $\boldsymbol{z}^n_i$ of the state $\boldsymbol{x}_i$.
Transform $\boldsymbol{z}^n_i$ in pseudo-observation $\boldsymbol{y}^n_i$ using (4.1).
Filter updates:
Compute the estimate $\widehat{\boldsymbol{y}}^n_i$ of $\overline{\boldsymbol{y}}_i$ using (4.10).
Compute the estimate $\widehat{\boldsymbol{x}}^n_{i|i}$ of the state $\boldsymbol{x}_i$ using (4.13).
Prediction updates:
Predict the estimate $\widehat{\boldsymbol{x}}^n_{i+1|i}$ of the state $\boldsymbol{x}_{i+1}$ using (4.15).
end while

$$
\begin{aligned}
A_G\widehat{y}^n_{-1} &= G\overline{x}_0 \\
\Sigma_{0|-1} &= J\otimes\Sigma_0 && (4.18)\\
Q_0 &= \Big((I_N\otimes B)\overline{D}_H - I_N\otimes G\Big)(J\otimes\Sigma_0)\Big(\overline{D}_H(I_N\otimes B^T) - I_N\otimes G\Big) + (I_N\otimes B)\overline{D}_H(I_N\otimes B^T) && (4.19)\\
\Pi_0 &= (J\otimes\Sigma_0)\left(I_N\otimes G - \overline{D}_H\left(I_N\otimes B^T\right)\right), && (4.20)
\end{aligned}
$$

Note that here we chose the same gain matrix B at all agents. Designing different gain matrices at different agents, such that the MSE is further reduced, is discussed in Chapter 5.

The filtering (4.10)-(4.13), and prediction (4.15) steps of the DIKF, in vector form, are:

$$\widehat{y}_i = (I_N\otimes A_G)(W\otimes I_M)\widehat{y}_{i-1} + (I_N\otimes B)\left(y_i - \overline{D}_H(I_N\otimes A)(I_N\otimes G^{-1})\widehat{y}_{i-1}\right) \quad (4.21)$$

$$\widehat{x}_{i|i} = \widehat{x}_{i|i-1} + K_i(\widehat{y}_i - (I_N\otimes G)\,\widehat{x}_{i|i-1}) \quad (4.22)$$

$$\widehat{x}_{i+1|i} = (I_N\otimes A)\,\widehat{x}_{i|i}. \quad (4.23)$$

In Algorithm 2, we state the steps that each sensor implements to estimate the time varying random field. We will show that the distributed estimates $\widehat{y}^n_i$ are unbiased and converge to the averaged pseudo-observations $\overline{y}_i$ with bounded MSE under Assumptions 1-7. Assumption 7 stated in Subsection 4.4.3 assumes the asymptotic stability of the matrix F defined in (4.18). However, we provide in Subsection (4.4.3), a construction on the system parameters that shows how to

guarantee that the matrix F is actually asymptotically stable. We will discuss the design of the pseudo-observations gain matrix B and its relation with the NTC of the DIKF. We will also prove that, given the pseudo-observation estimates $\widehat{y}_i^n$, the estimator gain matrices K_i^n, in (4.12), provide minimum MSE distributed estimates $\widehat{x}_{i|i-1}^n$ of the field x_i.

4.3.3 Memory, Communication and Computational Cost

The error covariance matrices Q_i, $\Sigma_{i|i}$, $\Sigma_{i+1|i}$, and Π_i can be pre-computed by (4.11), (4.14), (4.16), and (4.17). Using these matrices and (4.12), at each agent we compute and store the gain matrices K_i^n for all iterations to reduce the cost of the online DIKF. Offline computation of the gain matrices is explained in [56]. In Section **??**, we propose an alternative time-invariant gain matrix $K_i^n = \alpha G^{-1},\ \forall\ n,\ \forall\ i$, where α is a parameter. Choosing a time-invariant gain over time-varying gain matrices significantly reduces at each agent: (a) the computation cost, since computing the equations (4.11), (4.12), (4.14), (4.16), and (4.17) are no longer required; and (b) the memory usage, since the time-varying, K_i^n for each iteration, are not required to be stored. In Sections **??** and **??**, we discuss in detail the time-invariant gain and design of α.

We discuss the memory, communication and computational costs of the DIKF.

1. Memory: Each agent n stores a local copy of the dynamics model, A and V, the statistics of the initial condition, $\bar{x}_0$ and Σ_0, the observations model, H and R, and the network graph Laplacian, L. The consensus weight matrix, W and the gain matrices B and $K_i^n \left(= \alpha G^{-1}\right)$ are pre-computed and stored at each agent. Note that, in practice A, H and L are sparse, R is block diagonal, and, Σ_0 and B are diagonal matrices. In addition, at time iteration i, each agent stores the estimates $\widehat{y}_{i-1}^n$ and $\widehat{x}_{i|i-1}^n$ computed at the previous iteration $i-1$. The maximum number of elements that need to be stored in memory is on the order of

$$3M^2+5M+\frac{1}{2}\big(N^2+N\big)+\sum_{n=1}^{N}M_n^2+(M+0.5)\sum_{n=1}^{N}M_n.$$

   For example, if $M = N = 20$ and if $\forall n, M_n = 4$, then the maximum number of elements that need to be stored is 3470, which is equivalent to 27 kilobytes of memory usage (considering a real number in single-precision floating-point format occupies 4 bytes of computer memory).

2. Communication: At iteration i, each agent n transmits its pseudo-observation estimate $\widehat{y}_{i-1}^n$, containing M scalar elements, to its neighbors and receives the estimates $\widehat{y}_{i-1}^l,\ l \in \Omega_n$, from

its neighbors. The total number of such transmissions across the entire network at each iteration is twice the total number of edges in the network.

3. Computation: At iteration i, each agent n observes the new measurement z_i^n and computes the estimates $\widehat{y}_i^n$ and $\widehat{x}_{i+1|i}^n$, using equations (4.10), (4.13), and (4.15). These computations involve multiplication and addition operations between matrices and vectors. The computation requirement per iteration at each sensor is approximately

$$(\eta+4)M^2+\left(d_{\text{avg}}+1\right)M \text{ multiplications, and,}$$

$$(\eta+5)M^2+\left(d_{\text{avg}}-3-\eta\right)M \text{ additions}$$

where η is the sparsity of the field dynamics matrix, A, and d_{avg} is the average degree of the sensor communication network. For example, when $M=20$, $\eta=0.25$ and $d_{\text{avg}}=6$, then the number of multiplications and additions required per iteration at each agent are about 1840 and 2155, respectively. In this work we consider floating point operations and do not take into account quantization effects.

4.3.4 DIKF vs CKF

We compare the DIKF with the information of the CKF presented in Subsection 2.4.2.

1. CKF accesses all the pseudo-observations, $y_i^n, n=1,\ldots,N$, and computes (2.17), the averaged pseudo-observation sequence $\overline{y}_i$. In DIKF, each agent accesses only its own pseudo-observation, and it communicates only with its neighbors. They compute (4.10)-(4.11), distributed estimates $\widehat{y}_i^n$ and use them in place of $\overline{y}_i$.

2. The error covariance matrices of $\overline{y}_i$ and $\widehat{y}_i$ are $\frac{1}{N}G$ and Q_i. In the filtering step, if we replace $Q_i=\frac{1}{N}\left(I_N\otimes G\right)$ and the cross error covariance $\Pi_i=0$ in (4.12)-(4.14), then the distributed gains K_i^n and filter error covariances $\Sigma_{i|i}^n$ become equal to the centralized gain K_i^c and filter error covariance matrix $\Sigma_{i|i}^c$ in (2.18)-(2.20).

3. The prediction steps of the DIKF (4.15)-(4.16) and CKF (2.21)-(2.22) are similar.

The rest of this chapter derives the steps of the DIKF, analyzes its MSE, compares with the CKF, and finally validates these results through simulation.

4.4 DCPO: Convergence Analysis

The pseudo-observations, y_i^n, are time-varying. Computing the distributed estimates, $\widehat{y}_i^n$, of the averaged pseudo-observation $\overline{y}_i$ with DCPO is a *dynamic* consensus problem. This Section analyzes the DCPO step (4.10) of the DIKF.

4.4.1 Dynamics of the pseudo-observations

Before the analysis of the dynamic consensus step, we derive the dynamics of the averaged pseudo-observation, $\overline{y}_i$. We express the pseudo-observations, y_i^n, as the local measurements of $\overline{y}_i$. We also characterize the noises both in the dynamics and the observations. We start from the system, observation, and noise equations, (2.1)-(2.2). The detailed analysis of the $\overline{y}_i$ and y_i^n leads to the following proposition.

Proposition 2. *Consider (2.1)-(2.2), assumption 6, and definitions (4.1) and (2.17).*

(i) Dynamics: *The averaged pseudo-observation $\overline{y}_i$ follows*

$$\overline{y}_{i+1} = A_{\scriptscriptstyle G}\overline{y}_i + \xi_i \tag{4.24}$$

where the dynamics matrix is $A_{\scriptscriptstyle G}$ and the noise $\xi_i \sim \mathcal{N}(0, \Xi)$, with

$$\Xi = GVG + \frac{1}{N}G + \frac{1}{N}A_{\scriptscriptstyle G}GA_{\scriptscriptstyle G}^T.$$

The initial condition is $\overline{y}_0 \sim \mathcal{N}(G\overline{x}_0, \Sigma_{\overline{y}_0})$ with

$$\Sigma_{\overline{y}_0} = G\Sigma_0 G + \frac{1}{N}G.$$

(ii) Observations: *The pseudo-observation y_i^n at agent n is*

$$y_i^n = H_{\scriptscriptstyle G}^n\overline{y}_i + \delta_i^n, \qquad n = 1, \dots, N \tag{4.25}$$

where the observation matrix is $H_{\scriptscriptstyle G}^n = H_n^T R_n^{-1} H_n G^{-1}$ and the noise $\delta_i^n \sim \mathcal{N}(0, \Delta_n)$, with

$$\Delta_n = H_n^T R_n^{-1} H_n - \frac{1}{N}H_n^T R_n^{-1} H_n G^{-1} H_n^T R_n^{-1} H_n.$$

In vector notation,

$$y_i = \left(\overline{D}_H \left(I_N \otimes G^{-1}\right)\right)(1_N \otimes \overline{y}_i) + \delta_i$$

where the noise $\delta_i \sim \mathcal{N}(0, \Delta)$ *with*

$$\Delta = \overline{D}_H - \frac{1}{N}\overline{D}_H \left(J \otimes G^{-1}\right) \overline{D}_H.$$

(iii) Uncorrelated noises: *The dynamics noise,* ξ_i*, the observation noise,* δ_i*, and the initial condition* $\overline{y}_0$ *are uncorrelated Gaussian random vectors.* □

Proof. Proposition 1 consists of three key formulations: dynamics, observations, and independence of noises.

(i) *Dynamics*: Consider the averaged pseudo-observations sequence $\overline{y}_{i+1}$ in (4.3), substitute z_{i+1}^n and then x_{i+1} using (2.2) and (2.1) respectively.

$$\begin{aligned}\overline{y}_{i+1} &= \frac{1}{N}\sum_{n=1}^{N} H_n^T R_n^{-1}\left(H_n\left(Ax_i + v_i\right) + r_{i+1}^n\right) \\ &= GAx_i + Gv_i + \frac{1}{N}H^T R^{-1} r_{i+1}\end{aligned}$$

Replace GA with $A_{_G}G$ $(= GAG^{-1}G)$ and substitute G using (4.5). Then add and subtract a term with r_i^n leading to

$$\begin{aligned}\overline{y}_{i+1} &= A_{_G}\frac{1}{N}\sum_{n=1}^{N} H_n^T R_n^{-1}(H_n x_i + r_i^n - r_i^n) + Gv_i + \frac{1}{N}H^T R^{-1} r_{i+1} \\ &= A_{_G}\overline{y}_i + \underbrace{Gv_i - \frac{1}{N}A_{_G}H^T R^{-1} r_i + \frac{1}{N}H^T R^{-1} r_{i+1}}_{\xi_i}.\end{aligned}$$

The equation above shows that the dynamics of $\overline{y}_i$ follows (4.24). The noises v_i, r_i and r_{i+1} are zero-mean Gaussian. Hence the noise ξ_i is zero-mean Gaussian with covariance

$$\begin{aligned}\mathbb{E}[\xi_i \xi_i^T] &= G\mathbb{E}[v_i v_i^T]G + \frac{1}{N^2}A_{_G}H^T R^{-1}\mathbb{E}[r_i r_i^T]R^{-1}HA_{_G}^T + \frac{1}{N^2}H^T R^{^{-1}}\mathbb{E}[r_{_{i+1}} r_{i+1}^T]R^{-1}H \\ &= GVG + \frac{1}{N}G + \frac{1}{N}A_{_G}GA_{_G}^T \;=\; \Xi\end{aligned}$$

satisfying (4.25). Now using (4.4) we compute the Gaussian parameters of the initial condition $\overline{y}_0$ and prove (4.25).

$$\begin{aligned}\mathbb{E}[\overline{y}_0] &= \mathbb{E}[Gx_0 + \frac{1}{N}H^T R^{-1} r_0] = G\mathbb{E}[x_0] = G\overline{x}_0 \\ \Sigma_{\overline{y}_0} &= \mathbb{E}[(\overline{y}_0 - \mathbb{E}[\overline{y}_0])(\overline{y}_0 - \mathbb{E}[\overline{y}_0])^T] = G\Sigma_0 G + \frac{1}{N}G\end{aligned}$$

(ii) *Observations*: Pre-multiply both sides of (4.4) with G^{-1}

$$x_i = G^{-1}\overline{y}_i - \frac{1}{N}G^{-1}H^TR^{-1}r_i$$

and substitute this x_i in (4.1) to obtain the observations y_i^n

$$\begin{aligned}
y_i^n &= H_n^TR_n^{-1}z_i^n = H_n^TR_n^{-1}H_nx_i + H_n^TR_n^{-1}r_i^n \\
&= H_n^TR_n^{-1}H_n\Big(G^{-1}\overline{y}_i - \frac{1}{N}G^{-1}H^TR^{-1}r_i\Big) + H_n^TR_n^{-1}r_i^n \\
&= \underbrace{H_n^TR_n^{-1}H_nG^{-1}}_{H_G^n}\overline{y}_i + \underbrace{H_n^TR_n^{-1}\Big(r_i^n - \frac{1}{N}H_nG^{-1}H^TR^{-1}r_i\Big)}_{\delta_i^n}
\end{aligned}$$

as in (4.25). Since r_i is zero-mean Gaussian, δ_i^n is zero mean Gaussian with covariance (Δ_n) and cross-covariance $(\Delta_{n,l})$

$$\begin{aligned}
\Delta_n &= \mathbb{E}[\delta_i^n(\delta_i^n)^T] = H_n^TR_n^{-1}\Big(\mathbb{E}[r_i^n(r_i^n)^T] - \frac{1}{N^2}H_nG^{-1}H^TR^{-1}\mathbb{E}[r_ir_i^T]R^{-1}HG^{-1}H_n^T\Big)R_n^{-1}H_n \\
&= H_n^TR_n^{-1}H_n - \frac{1}{N}H_n^TR_n^{-1}H_nG^{-1}H_n^TR_n^{-1}H_n \\
\Delta_{n,l} &= \mathbb{E}[\delta_i^n(\delta_i^l)^T] = -\frac{1}{N}H_n^TR_n^{-1}H_nG^{-1}H_l^TR_l^{-1}H_l
\end{aligned}$$

The covariance Δ in (4.26) contains Δ_n and $\Delta_{n,l}$ as blocks.

(iii) *Uncorrelated noises*: The initial condition y_0 consists of x_0 and r_0, whereas the noises ξ_i and δ_i consists of v_i, r_i and r_{i+1}. The noises v_i and r_i are independent of x_0 and r_0 (for $i > 0$). Hence y_0 is independent (zero covariance) of both ξ_i and δ_i. The covariance between ξ_i and δ_i is

$$\begin{aligned}
\mathbb{E}[\xi_i\delta_i^T] = \mathbb{E}\Bigg[&\Big(Gv_i + \frac{1}{N}H^TR^{-1}r_{i+1} - \frac{1}{N}A_GH^TR^{-1}r_i\Big)\Big(D_H^TR^{-1}r_i \\
&\qquad - \frac{1}{N}\overline{D}_H(I_N\otimes G^{-1})(I_N\otimes(H^TR^{-1}r_i))\Big)^T\Bigg] \\
= &-\frac{1}{N}A_GH^TR^{-1}D_H + \frac{1}{N}A_G\Big(\frac{1}{N}H^TR^{-1}H\Big)G^{-1}\overline{D}_H \\
= &-\frac{1}{N}A_G\overline{D}_H + \frac{1}{N}A_G\overline{D}_H = 0.
\end{aligned}$$

The pair-wise zero covariances between y_0, ξ_i and δ_i ensures that they are uncorrelated. □

Thus, we formulated distributed dynamic consensus as the distributed estimation of a linear dynamical system (4.24), with observations (4.25), and uncorrelated Gaussian noises. Before we study the

error process of DCPO step (4.21), we express y_{i+1} in terms of $\overline{y}_i$ by applying (4.24) on (4.26),

$$y_{i+1} = \overline{D}_H(I_N \otimes G^{-1})(I_N \otimes A_{_G})(1_N \otimes \overline{y}_i) + \overline{D}_H(I_N \otimes G^{-1})(1_N \otimes \xi_i) + \delta_{i+1}. \tag{4.26}$$

We use the formulation in Proposition 2 to analyze the convergence of dynamic consensus. From this point on, we write all equations in vector form for ease of notation.

4.4.2 Error Analysis

To show convergence and to analyze the performance of the dynamic consensus step (4.21), we analyze the error process e_i

$$e_i = \widehat{y}_i - 1_N \otimes \overline{y}_i. \tag{4.27}$$

Now, we study the dynamics of e_i. Equations (4.21) and (4.24) along with (4.27) imply

$$\begin{aligned} e_{i+1} &= \widehat{y}_{i+1} - 1_N \otimes \overline{y}_{i+1} \\ &= (I_N \otimes A_{_G})(W \otimes I_M)\widehat{y}_i - (I_N \otimes A_{_G})(1_N \otimes \overline{y}_i) + (I_{_N} \otimes B)\left(y_{_{i+1}} - \overline{D}_{_H}(I_{_N} \otimes A)(I_{_N} \otimes G^{-1})\widehat{y}_i\right) - (1_{_N} \otimes \xi_i). \end{aligned}$$

Substitute y_{i+1} from (4.26),

$$\begin{aligned} e_{i+1} &= (I_N \otimes A_{_G})(W \otimes I_M)\,(\widehat{y}_i - 1_N \otimes \overline{y}_i) + (I_N \otimes B)\delta_{i+1} - (I_N \otimes B)\overline{D}_H(I_N \otimes AG^{-1})\,(\widehat{y}_i - 1_N \otimes \overline{y}_i) \\ &\quad + \Big((I_N \otimes B)\overline{D}_H(I_N \otimes G^{-1}) - I_{MN}\Big)(1_N \otimes \xi_i) \\ &= \Big(W \otimes A_{_G} - (I_N \otimes B)\overline{D}_H(I_N \otimes AG^{-1})\Big)e_i + \overline{\phi}_i \end{aligned}$$

where the noise $\overline{\phi}_i$ is defined in (4.29). The error process, e_i, evolves with time i as:

$$e_{i+1} = Fe_i + \overline{\phi}_i \tag{4.28}$$

where, the error matrix F is defined in (4.18) and the noise $\overline{\phi}_i$,

$$\overline{\phi}_i = \Big((I_N \otimes B)\overline{D}_H(I_N \otimes G^{-1}) - I_{MN}\Big)(1_N \otimes \xi_i) + (I_N \otimes B)\delta_{i+1}. \tag{4.29}$$

The noise $\overline{\phi}_i$ in (4.28) is a linear combination of the noises ξ_i and δ_{i+1}. By Proposition 2, ξ_i and δ_i are uncorrelated zero-mean Gaussian. Therefore, $\overline{\phi}_i \sim \mathcal{N}(0, \overline{\Phi})$, with

$$\begin{aligned} \overline{\Phi} = \Big((I_N \otimes B)\overline{D}_H(I_N \otimes G^{-1}) \;-\; I_{MN}\Big)(J \otimes \Xi)\Big((I_N \otimes G^{-1})\overline{D}_H(I_N \otimes B^T) \\ - I_{MN}\Big) + (I_N \otimes B)\Delta(I_N \otimes B^T). \end{aligned} \tag{4.30}$$

Lemma 4.4.1. *The error process e_i and the noise sequence $\overline{\phi}_i$ are correlated and their covariance is*

$$\widetilde{\Phi} = \mathbb{E}[e_i \overline{\phi}_i^T] = -\frac{1}{N}\Big((I_N \otimes B)\overline{D}_H(I_N \otimes G^{-1}) - I_{MN}\Big) \times (J \otimes A^T G)\Big((I_N \otimes G^{-1})\overline{D}_H(I_N \otimes B^T) - I_{MN}\Big). \quad (4.31)$$

Proof. The noise, $\overline{\phi}_i$ in (4.29), and the error process, e_i in (4.47), are both zero-mean. The covariance between e_i and $\overline{\phi}_i$ is

$$\mathbb{E}[e_i \overline{\phi}_i^T] = \mathbb{E}\left[\left(F e_{i-1} + \overline{\phi}_{i-1}\right)\overline{\phi}_i^T\right].$$

Note that $\overline{\phi}_i$ is composed of ξ_i and δ_{i+1}. The error e_{i-1} contains the noise $\overline{\phi}_{i-2}$, which is composed of ξ_{i-2} and δ_{i-1}. The noises ξ_{i-2} and δ_{i-1} are independent of ξ_i and δ_{i+1}. Hence, $\mathbb{E}\left[e_{i-1}\overline{\phi}_i^T\right] = 0$. Using (4.29) expand $\overline{\phi}_{i-1}$ and $\overline{\phi}_i$,

$$\begin{aligned}\mathbb{E}[e_i \overline{\phi}_i^T] &= \mathbb{E}\left[\overline{\phi}_{i-1}\overline{\phi}_i^T\right] \\ &= \Big((I_N \otimes B)\overline{D}_H(I_N \otimes G^{-1}) - I_{MN}\Big)\left(J \otimes \mathbb{E}\left[\xi_{i-1}\xi_i^T\right]\right) \\ &\quad \times \Big((I_N \otimes G^{-1})\overline{D}_H(I_N \otimes B^T) - I_{MN}\Big) \end{aligned} \quad (4.32)$$

where, the expectation of all the cross terms are zero except ξ_{i-1} and ξ_i, whose covariance

$$\begin{aligned}\mathbb{E}\left[\xi_{i-1}\xi_i^T\right] &= -\frac{1}{N^2}H^T R^{-1}\mathbb{E}\left[r_i r_i^T\right]R^{-1}H A_G^T \\ &= -\frac{1}{N}GG^{-1}A^T G = -\frac{1}{N}A^T G\end{aligned}$$

when substituted in (4.32) proves the lemma. □

The MSE of $\widehat{y}_i$ is bounded if and only if the dynamics of the error process, e_i, are asymptotically stable, i.e., the spectral radius $\rho(F) < 1$. With dynamic consensus step, (4.21), the consensus weight matrix, W, and the pseudo-observation gain matrix, B are the two design parameters. Next, we provide an example construction of W and B such that $\rho(F) < 1$.

4.4.3 Estimator Design:

The agents communicate their distributed estimates to their neighbors. The graph Laplacian L of the undirected agent network is symmetric. In this example design, we consider a stochastic

constant-weight consensus matrix W [35], i.e.,

$$W = I_M - \beta_1 L \tag{4.33}$$

where, β_1 is the consensus weight. The agents weigh their neighbors' estimates with β_1 to update their distributed estimates $\widehat{y}_i^n$. For the purpose of illustration, we choose a scalar pseudo-observation gain matrix B, i.e.,

$$B = \beta_2 I_M. \tag{4.34}$$

Note that the parameters β_1 and β_2 are positive, i.e., $\beta_1 > 0$ and $\beta_2 > 0$. We design β_1 and β_2 and find the conditions such that $\rho(F) < 1$. Substitute W and B in the error matrix F (4.18) using equations (4.33) and (4.34),

$$\begin{aligned} F &= (I_N \otimes G)\Big(I_{MN} - \Big(\beta_1 \left(L \otimes I_M\right) + \beta_2 \left(I_N \otimes G^{-1}\right) \overline{D}_H\Big)\Big)(I_N \otimes A)(I_N \otimes G^{-1}) \qquad (4.35)\\ &= (I_N \otimes G) F_2 (I_N \otimes A)(I_N \otimes G^{-1}), \end{aligned}$$

where, F_2 denotes the part of F dependent on the network topology L and observation model $\overline{D}_H$,

$$\begin{aligned} F_2 &= I_{MN} - \Big(\beta_1 \left(L \otimes I_M\right) + \beta_2 \left(I_N \otimes G^{-1}\right) \overline{D}_H\Big) \qquad (4.36)\\ &= I_{MN} - \beta_1 F_1, \qquad \text{with,} \end{aligned}$$

$$F_1 = L \otimes I_M + \overline{\beta}_2 (I_N \otimes G^{-1}) \overline{D}_H, \qquad \overline{\beta}_2 = \frac{\beta_2}{\beta_1} > 0. \tag{4.37}$$

The eigenvalues of F_1 play a key role in the design of β_1 and β_2 such that the spectral radius of the error matrix $\rho(F) < 1$.

Lemma 4.4.2. *The matrix F_1 has positive eigenvalues.* □

Proof. The matrices L and $\overline{D}_H$ are symmetric positive semi-definite (PSD) matrices, whereas, G and hence G^{-1} are symmetric positive definite (PD) matrices[1]. Note that F_1 is not necessarily symmetric since $I_N \otimes G^{-1}$ and $\overline{D}_H$ may not commute. So we decompose F_1 as follows:

$$F_1 = (I_N \otimes G^{-\frac{1}{2}}) \underbrace{\left(L \otimes I_M + \overline{\beta}_2 (I_N \otimes G^{-\frac{1}{2}}) \overline{D}_H (I_N \otimes G^{-\frac{1}{2}})\right)}_{\overline{F}_1} (I_N \otimes G^{\frac{1}{2}}) \tag{4.38}$$

[1]The symmetricity, definiteness and semi-definiteness of matrices remain conserved after Kronecker product with identity matrices.

where, $G^{-\frac{1}{2}}$ denotes the square-root[2] of G. Since $G^{\frac{1}{2}}$ is a symmetric positive definite matrix, equation (4.36) implies that F_1 and $\overline{F}_1$ are similar matrices[3]. Therefore, F_1 and $\overline{F}_1$ share the same eigenvalues. So it is sufficient to show that the eigenvalues of $\overline{F}_1$ are positive.

Note that the matrices $L \otimes I_M$ and $(I_N \otimes G^{-\frac{1}{2}})\overline{D}_H(I_N \otimes G^{-\frac{1}{2}})$ are symmetric PSD, and, $\overline{\beta}_2 > 0$. Therefore $\overline{F}_1$ is also symmetric PSD and its eigenvalues are non-negative. Now we show that zero is not an eigenvalue of $\overline{F}_1$ by proving it PD.

Take any $u \in \mathbb{R}^{NM}, u \neq 0$. Similar to Lemma 6 in [27], orthogonally decompose u into $u_{\mathcal{C}}$and $u_{\mathcal{C}^\perp}$, the projections into consensus subspace $\mathcal{C}$ and orthogonal subspace $\mathcal{C}^\perp$

$$\begin{aligned} u &= u_{\mathcal{C}} + u_{\mathcal{C}^\perp} \\ \text{where,} \qquad \mathcal{C} &= \left\{ v \in \mathbb{R}^{NM} | v = 1_N \otimes s, s \in \mathbb{R}^M \right\}. \end{aligned}$$

The orthogonal projections $u_{\mathcal{C}}$ and $u_{\mathcal{C}^\perp}$ satisfy

$$\begin{aligned} u_{\mathcal{C}}^T \left(L \otimes I_M\right) = \left(L \otimes I_M\right) u_{\mathcal{C}} &= 0 \qquad \forall\, u_{\mathcal{C}}, \quad \text{and} \\ u_{\mathcal{C}^\perp}^T \left(L \otimes I_M\right) u_{\mathcal{C}^\perp} &> 0 \qquad \forall\, u_{\mathcal{C}^\perp} \neq 0. \end{aligned}$$

The quadratic form of $\overline{F}_1$ with u results in

$$\begin{aligned} u^T\overline{F}_1 u = &\left(u_{\mathcal{C}} + u_{\mathcal{C}^\perp}\right)^T \left(L \otimes I_M + \overline{\beta}_2 (I_N \otimes G^{-\frac{1}{2}})\overline{D}_H (I_N \otimes G^{-\frac{1}{2}})\right) \left(u_{\mathcal{C}} + u_{\mathcal{C}^\perp}\right) \\ = &\overline{\beta}_2 \left(u_{\mathcal{C}} + u_{\mathcal{C}^\perp}\right)^T \left((I_N \otimes G^{-\frac{1}{2}})\overline{D}_H (I_N \otimes G^{-\frac{1}{2}})\right) \left(u_{\mathcal{C}} + u_{\mathcal{C}^\perp}\right) + u_{\mathcal{C}^\perp}^T \left(L \otimes I_M\right) u_{\mathcal{C}^\perp}. \end{aligned} \tag{4.39}$$

Both quadratic terms are non-negative due to positive semi-definiteness of the matrices. Since $u \neq 0$, then at least one of the orthogonal components $u_{\mathcal{C}}$ and $u_{\mathcal{C}^\perp}$ must be non-zero. If $u_{\mathcal{C}^\perp} \neq 0$, then for all $u_{\mathcal{C}}$

$$u^T\overline{F}_1 u \geq u_{\mathcal{C}^\perp}^T \left(L \otimes I_M\right) u_{\mathcal{C}^\perp} > 0.$$

To complete the proof, now we need to show that $u^T\overline{F}_1 u > 0$ when $u_{\mathcal{C}^\perp} = 0$ and $u_{\mathcal{C}} \neq 0$. We use the form $u_{\mathcal{C}} = 1_N \otimes s$ where, $s \in \mathbb{R}^M$ and $s \neq 0$. Then,

$$u^T\overline{F}_1 u = \overline{\beta}_2 u_{\mathcal{C}}^T \left(I_N \otimes G^{-\frac{1}{2}}\right) \overline{D}_H \left(I_N \otimes G^{-\frac{1}{2}}\right) u_{\mathcal{C}}$$

[2]The square-root exists since the symmetric PD matrix G is diagonalizable. Note $G^{-\frac{1}{2}}$ retains the symmetric positive definiteness property of G.

[3]Refer to [48] for properties of similar matrices.

$$
\begin{aligned}
&= \overline{\beta}_2 \sum_{n=1}^{N} s^T G^{-\frac{1}{2}} H_n^T R_n^{-1} H_n G^{-\frac{1}{2}} s \\
&= \overline{\beta}_2 s^T G^{-\frac{1}{2}} G G^{-\frac{1}{2}} s \\
&= \overline{\beta}_2 ||s||^2 \quad > \ 0.
\end{aligned}
$$

□

Denote the positive eigenvalues of F_1 by $0 < \lambda_1 \leq \ldots \leq \lambda_N$. Given the local observation model (4.26), define the diffusion rate, γ, of the agent communication network and $\gamma_{\scriptscriptstyle G}$ as

$$\gamma = \max_{\overline{\beta}_2} \frac{\lambda_1}{\lambda_N} = \frac{\overline{\lambda}_1}{\overline{\lambda}_N} \tag{4.40}$$

$$\gamma_{\scriptscriptstyle G} = \frac{\lambda_{\min}(G)}{\lambda_{\max}(G)}. \tag{4.41}$$

Note that $\gamma, \gamma_{\scriptscriptstyle G} \in (0, 1]$. Let, $\overline{\beta}_2^*$ denote the maximizer in (4.40). Recall the error matrix F in (4.35). The next lemma derives the conditions and the design of β_1, β_2 such that $\rho(F) < 1$.

Lemma 4.4.3. *If the spectral norm of the system matrix, A, is upper bounded as:*

$$a = ||A||_2 < \sqrt{\gamma_{\scriptscriptstyle G}} \frac{1+\gamma}{1-\gamma} \tag{4.42}$$

then there exists

$$\beta_1 \in \left(\frac{a - \sqrt{\gamma_{\scriptscriptstyle G}}}{a\overline{\lambda}_1}, \frac{a + \sqrt{\gamma_{\scriptscriptstyle G}}}{a\overline{\lambda}_N} \right) \tag{4.43}$$

such that $\rho(F) < 1$. □

Proof. We write the error matrix F in (4.35) as

$$F = (I_N \otimes G)\, F_3 \left(I_N \otimes G^{-1}\right)$$

where, $F_3 = (I_{MN} - \beta_1 F_1)\,(I_N \otimes A)$.

Note that F and F_3 are similar matrices.

$$
\begin{aligned}
\rho(F) &= \rho(F_3) \\
&\leq \Big\| (I_{MN} - \beta_1 F_1)\,(I_N \otimes A) \Big\|_2 \\
&\leq \Big\| I_{MN} - \beta_1 F_1 \Big\|_2 \Big\| A \Big\|_2
\end{aligned}
$$

$$
\begin{aligned}
&\leq a \left\| (I_N \otimes G^{-\frac{1}{2}}) \left(I_{MN} - \beta_1 \overline{F}_1 \right) (I_N \otimes G^{\frac{1}{2}}) \right\|_2 \\
&\leq a \left\| G^{-\frac{1}{2}} \right\|_2 \left\| G^{\frac{1}{2}} \right\|_2 \left\| I_{MN} - \beta_1 \overline{F}_1 \right\|_2 \\
&= \frac{a}{\sqrt{\gamma_G}} \, \rho \left(I_{MN} - \beta_1 \overline{F}_1 \right) \\
&= \frac{a}{\sqrt{\gamma_G}} \max \Big\{ \underbrace{\left(1 - \beta_1 \overline{\lambda}_1 \right)}_{(i)}, \underbrace{\left(\beta_1 \overline{\lambda}_N - 1 \right)}_{(ii)} \Big\}. \qquad (4.44)
\end{aligned}
$$

Before we analyze the terms (i) and (ii) of (4.44), we comment on the choice of β_1. Substitute γ in (4.42) by (4.40) to get

$$
\frac{a - \sqrt{\gamma_G}}{\overline{\lambda}_1} < \frac{a + \sqrt{\gamma_G}}{\overline{\lambda}_N}.
$$

Divide both sides by a and choose β_1 such that

$$
\frac{a - \sqrt{\gamma_G}}{a\overline{\lambda}_1} < \beta_1 < \frac{a + \sqrt{\gamma_G}}{a\overline{\lambda}_N},
$$

which is in accord with (4.43). Substitute these minimum and maximum values of β_1 into the terms (i) and (ii) of (4.44),

$$
\begin{aligned}
\rho(F) &\leq \frac{a}{\sqrt{\gamma_G}} \max \left\{ \left(1 - \beta_1 \overline{\lambda}_1 \right), \left(\beta_1 \overline{\lambda}_N - 1 \right) \right\} \\
&< \frac{a}{\sqrt{\gamma_G}} \max \left\{ \left(1 - \frac{a - \sqrt{\gamma_G}}{a\overline{\lambda}_1} \overline{\lambda}_1 \right), \left(\frac{a + \sqrt{\gamma_G}}{a\overline{\lambda}_N} \overline{\lambda}_N - 1 \right) \right\} \\
&\leq \frac{a}{\sqrt{\gamma_G}} \max \left\{ \frac{\sqrt{\gamma_G}}{a}, \frac{\sqrt{\gamma_G}}{a} \right\} = 1.
\end{aligned}
$$

□

Network Tracking Capacity: We choose β_1 satisfying (4.43) of Lemma 4.4.3 and construct the consensus weight matrix W using (4.33). Using this β_1 and the maximizer $\overline{\beta}_2^*$ from (4.40), we obtain $\beta_2 = \overline{\beta}_2^* \beta_1$ and subsequently the gain matrix B using (4.34). Such design of W and B enforces asymptotic stability of the error process, e_i in (4.28), i.e., $\rho(F) < 1$. To obtain such asymptotic stability, Lemma 4.4.3 states that the degree of instability of the system dynamics (2.1), characterized by the spectral norm of the system matrix A, should be within a threshold determined by (4.42). The corresponding upper bound on the dynamics of the averaged pseudo-observation, $\overline{y}_i$ in (4.24), given by spectral norm of the dynamics matrix $A_{\scriptscriptstyle G}$ is

$$
\|A_{\scriptscriptstyle G}\|_2 = \|GAG^{-1}\| \leq \|G\|\|G^{-1}\|\|A\|
$$

$$< \quad \frac{1}{\gamma_G}\sqrt{\gamma_G}\frac{1+\gamma}{1-\gamma} = \frac{1}{\sqrt{\gamma_G}}\frac{1+\gamma}{1-\gamma}. \tag{4.45}$$

Inequality (4.45) defines the dynamic consensus step's Network Tracking Capacity (NTC). With this design, we can choose any β_1 that satisfies (4.43), but the β_1 that minimizes $\rho(F_2)$ is

$$\beta_1^* = \frac{2}{\overline{\lambda}_1 + \overline{\lambda}_N}. \tag{4.46}$$

Note that β_1^* lies in the acceptable range given by (4.43). Since the design of the B and W matrices is not the primary focus of this paper, we stop here by providing an example design. Any other design of B and W matrices is also acceptable as long as it satisfies the $\rho(F) < 1$assumption stated below:

Assumption 7 (Stability)**.** *The pseudo-observations error matrix F, defined in* (4.18)*, has $\rho(F) < 1$.*

From now onwards, we assume that given (4.24) and (4.25), we design the matrices W and B such that $\rho(F) < 1$. With this assumption, we discuss the convergence and performance of the DCPO step in Subsection 4.4.4. The design of a generic gain B and consensus-weights W to improve the overall performance of the DIKF, is discussed in Chapter 5. For a given A, L and D_H, there may not always exist B and W such that $\rho(F) < 1$. In these cases, we should rewire the agent communication network (i.e., the network topology, L) and/or modify the sensor placements (i.e., observations, D_H), so that we can design B and W satisfying $\rho(F) < 1$.

4.4.4 Unbiasedness and bounded MSE:

This Subsection analyzes the convergence and performance of DCPO, (4.10) of Step 1 of DIKF. We consider the DCPO formulation of Subsection 4.4.1 with the design matrices as constructed in Subsection 4.4.3 and study the error equations from Subsection 4.4.2, which results in the following theorem.

Theorem 4.4.1. *At each agent, the distributed pseudo-observation estimates, $\widehat{y}_i^n$, of the averaged pseudo-observation, $\overline{y}_i$, are unbiased at all time indices, i.e.,*

$$\mathbb{E}[e_i^n] = \mathbb{E}[\widehat{y}_i^n - \overline{y}_i] = 0, \qquad \forall i, \ \ \forall n$$

and their MSE are asymptotically bounded

$$\lim_{i\to\infty} \frac{1}{M}\mathbb{E}[(\widehat{y}_i^n - \overline{y}_i)^T(\widehat{y}_i^n - \overline{y}_i)] < \infty, \qquad \forall n.$$

The covariance P_i of the error process e_i satisfies the Lyapunov iteration

$$P_{i+1} \;=\; FP_iF^T + \Phi$$

where, Φ, defined using (4.30). □

Proof. Consider the error process e_i in (4.27) and compute $\mathbb{E}[e_0]$, the bias of the distributed pseudo-observation estimates $\widehat{y}_i$ at $i = 0$. Substitute $\widehat{y}_0$, the initial condition of $\overline{y}_0$, and initial pseudo-observation y_0 using (4.21), (4.25) and (4.26) respectively,

$$\begin{aligned}
\mathbb{E}[e_0] &= \mathbb{E}[\widehat{y}_0 - 1_N \otimes \overline{y}_0] \\
&= \Big((I_N \otimes G) - (I_N \otimes B)\overline{D}_H\Big)(1_N \otimes \overline{x}_0) + (I_N \otimes B) \\
&\qquad \times \mathbb{E}[\overline{D}_H(I_N \otimes G^{-1})(1_N \otimes \overline{y}_0) + \delta_i] - 1_N \otimes \mathbb{E}[\overline{y}_0] \\
&= 1_N \otimes (G\overline{x}_0) - (I_N \otimes B)\overline{D}_H(1_N \otimes \overline{x}_0) + (I_N \otimes B) \\
&\qquad \times \overline{D}_H(I_N \otimes G^{-1})\big(1_N \otimes (G\overline{x}_0)\big) - 1_N \otimes (G\overline{x}_0) = 0.
\end{aligned}$$

Recall the dynamics of e_i in (4.28) and the unbiasedness of the pseudo-observation estimates $\widehat{y}_i$ at any time i follows:

$$\begin{aligned}
\mathbb{E}[e_i] &= F\mathbb{E}[e_{i-1}] + \mathbb{E}[\overline{\phi}_{i-1}] \\
&= F\mathbb{E}[e_{i-1}] \;=\; F^i\mathbb{E}[e_0] \;=\; 0
\end{aligned} \tag{4.47}$$

For the second part of the theorem, denote the covariance of e_i by P_i. Expand P_{i+1} using (4.28), (4.30) and Lemma 4.4.1,

$$\begin{aligned}
P_{i+1} &= \mathbb{E}[e_{i+1}e_{i+1}^T] \\
&= F\mathbb{E}[e_ie_i^T]F^T + \mathbb{E}[\overline{\phi}_i\overline{\phi}_i^T] + F\mathbb{E}[e_i\overline{\phi}_i^T] + \mathbb{E}[\overline{\phi}_ie_i^T]F^T \\
&= FP_iF^T + \underbrace{\overline{\Phi} + F\widetilde{\Phi} + \widetilde{\Phi}^TF^T}_{\Phi} \\
&= F^iP_0(F^T)^i + \sum_{j=0}^{i-1} F^j\Phi(F^T)^j.
\end{aligned} \tag{4.48}$$

Taking the limit as $i \to \infty$ and noting that $\rho(F) < 1$,

$$\lim_{i\to\infty} P_i \;=\; \lim_{i\to\infty} F^iP_0(F^T)^i + \lim_{i\to\infty}\sum_{j=0}^{i-1} F^j\Phi(F^T)^j$$

$$P_\infty \;=\; \lim_{i\to\infty}\sum_{j=0}^{\infty} F^j\Phi(F^T)^j.$$

The infinite sum $\sum_{j=0}^{\infty} F^j\Phi(F^T)^j$ converges to finite P_∞, where P_∞ is the solution of the following equation:

$$P_\infty = FP_\infty F^T + \Phi.$$

Thus the asymptotic MSE is bounded,

$$\lim_{i\to\infty}\frac{1}{NM}\mathbb{E}[e_i^T e_i] = \frac{1}{NM}\text{trace}(P_\infty) < \infty. \qquad \square$$

In summary, dynamic consensus on pseudo-observation enables each agent to obtain an unbiased distributed pseudo-observation estimate, $\widehat{y}_i^n$, of the averaged pseudo-observation, $\overline{y}_i$, based on its own observations and by communicating with its neighbors. We proved that the MSE of the distributed estimates are bounded. Based on this, if we compare the Step 1 of the DIKF in Subsection 4.3.2 with the Step 1 of the centralized Kalman filter in Subsection 2.4.2, we see it is justified to substitute $\overline{y}_i$ by their estimates $\widehat{y}_i^n$ at each agent.

4.4.5 Intermediate Results

This Subsection analyzes an intermediate step that connects the convergence results of the dynamic consensus on pseudo-observation to the convergence analysis of the DIKF, namely, the impact of substituting the global average $\overline{y}_i$ by their distributed estimates $\widehat{y}_i^n$ in DIKF. From (2.17), we infer that $\overline{y}_i$ is Gaussian,

$$\overline{y}_i \sim \mathcal{N}\left(Gx_i, \frac{1}{N}G\right). \tag{4.49}$$

Now we derive the dynamics of the distributed estimates $\widehat{y}_i^n$ with respect to the field dynamics x_i. To obtain the dynamics, we analyze the corresponding error process q_i in (4.10) and its covariance Q_i. The next theorem summarizes the dynamics and the distribution of q_i.

Theorem 4.4.2. *The error $q_i \sim \mathcal{N}(0, Q_i)$. The covariance Q_i is asymptotically bounded and it satisfies the Lyapunov iteration*

$$Q_{i+1} \;=\; FQ_iF^T + \Psi$$

where Ψ *and the initial covariance* Q_0 *are defined in equation* (4.18) *and Subsection* **??** *respectively.*

□

Proof. Begin with the error q_i in (4.10), then replace $\widehat{y}_{i+1}$ by the DCPO (4.21)

$$\begin{aligned}
q_{i+1} &= \widehat{y}_{i+1} - 1_N \otimes (Gx_{i+1}) \\
&= (I_N \otimes A_{_G})(W \otimes I_M)\widehat{y}_i - 1_N \otimes (Gx_{i+1}) \\
&\quad + (I_N \otimes B)\left(y_{i+1} - \overline{D}_H(I_N \otimes A)(I_N \otimes G^{-1})\widehat{y}_i\right) \\
&= \Big(\underbrace{W \otimes A_{_G} - (I_N \otimes B)\overline{D}_H(I_N \otimes AG^{-1})}_{F}\Big)\widehat{y}_i \\
&\quad + (I_N \otimes B)y_{i+1} - 1_N \otimes (Gx_{i+1})
\end{aligned}$$

where, we use the Kronecker product property,

$$(I_N \otimes A_{_G})(W \otimes I_M) = (W \otimes I_M)(I_N \otimes A_{_G}) = (W \otimes A_{_G}).$$

Substitute x_{i+1} and y_{i+1} by (2.1) and (4.26) , respectively,

$$\begin{aligned}
q_{i+1} &= F\widehat{y}_i + (I_N \otimes B)\Big(\overline{D}_H(I_N \otimes G^{-1})(I_N \otimes A_{_G})(1_N \otimes \overline{y}_i) \\
&\quad + \overline{D}_H(I_N \otimes G^{-1})(1_N \otimes \xi_i) + \delta_{i+1}\Big) \\
&\quad - 1_N \otimes \big(G(Ax_i + v_i)\big) \\
&= F\widehat{y}_i + (I_N \otimes B)\overline{D}_H(I_N \otimes AG^{-1})(1_N \otimes \overline{y}_i) \\
&\quad - (I_N \otimes A_{_G})(1_N \otimes Gx_i) - 1_N \otimes Gv_i \\
&\quad + (I_N \otimes B)\Big(\overline{D}_H(I_N \otimes G^{-1})(1_N \otimes \xi_i) + \delta_{i+1}\Big).
\end{aligned}$$

Replace ξ_i and δ_{i+1} using their definitions from the proof of Proposition 2,

$$\begin{aligned}
q_{i+1} &= F(\widehat{y}_i - 1_N \otimes Gx_i) - 1_N \otimes Gv_i \\
&\quad + (I_N \otimes B)\overline{D}_H\left(I_N \otimes AG^{-1}\right)\left(1_N \otimes \left(\frac{1}{N}H^T R^{-1} r_i\right)\right) \\
&\quad + (I_N \otimes B)\overline{D}_H\left(I_N \otimes G^{-1}\right)\left(1_N \otimes \left(\frac{1}{N}H^T R^{-1} r_{i+1}\right.\right. \\
&\quad \left.\left. + Gv_i - \frac{1}{N}A_{_G}H^T R^{-1} r_i\right)\right) + (I_N \otimes B)\Big(D_H^T R^{-1} r_{i+1} \\
&\quad - \frac{1}{N}\overline{D}_H(I_N \otimes G^{-1})(I_N \otimes (H^T R^{-1} r_{i+1}))\Big)
\end{aligned}$$

$$= Fq_i + \underbrace{\Big((I_N \otimes B)\overline{D}_H - I_N \otimes G\Big)\,(1_N \otimes v_i) + (I_N \otimes B)D_H^T R^{-1} r_{i+1}}_{\psi_i}$$
$$= Fq_i + \psi_i. \tag{4.50}$$

Note that the noise process ψ_i is a linear combination of statistically independent zero-mean Gaussian noises, v_i, and r_{i+1}. Therefore, $\psi_i \sim \mathcal{N}(0, \Psi)$, with

$$\Psi = \Big((I_N \otimes B)\overline{D}_H - I_N \otimes G\Big)(J \otimes V)\Big(\overline{D}_H(I_N \otimes B^T) - I_N \otimes G\Big) + (I_N \otimes B)\overline{D}_H(I_N \otimes B^T).$$

Similarly ψ_{i-1} consists of v_{i-1}, and r_i, which are independent of v_i, and r_{i+1}. So ψ_i is statistically independent over time. Note that q_i contains all the noises from ψ_o to ψ_{i-1}. Hence, ψ_i and q_i are statistically independent. Therefore the error covariance Q_i follows the Lyapunov type iteration as computed in the DCPO step (4.11) of the DIKF algorithm.

Now to derive the initial condition Q_0, we compute q_0 using (4.21), (4.26), (4.4) and $\widehat{y}_{-1}^n$,

$$\begin{aligned} q_0 &= \widehat{y}_0 - 1_N \otimes (Gx_0) \\ &= \Big((I_N \otimes B)\overline{D}_H - I_N \otimes G\Big)\,(1_N \otimes x_0 - 1_N \otimes \overline{x}_0) \\ &\quad + (I_N \otimes B)D_H^T R^{-1} r_0 \\ \mathbb{E}[q_0] &= 0 \qquad (4.51) \\ Q_0 &= \Big((I_N \otimes B)\overline{D}_H - I_N \otimes G\Big)(J \otimes \Sigma_0)\Big(\overline{D}_H(I_N \otimes B^T) \\ &\quad - I_N \otimes G\Big) + (I_N \otimes B)\overline{D}_H(I_N \otimes B^T). \qquad (4.52) \end{aligned}$$

Equations (4.11), (4.51), and (4.52) together with Assumption 7, $\rho(F) < 1$, establish that q_i is zero mean and the covariance Q_i is asymptotically bounded. □

The error matrix Q_i is a block matrix with N^2 blocks of sizes $M \times M$ each. Let Q_i^n denote the n^{th} diagonal block of Q_i.

Corollary 4.4.3. *The dynamics of the distributed estimates $\widehat{y}_i^n$ with respect to the system dynamics x_i is*

$$\begin{aligned} \widehat{y}_i^n &= Gx_i + q_i^n, \qquad \forall n \qquad (4.53) \\ \textit{where,} \qquad q_i^n &\sim \mathcal{N}(0, Q_i^n), \end{aligned}$$

$$\textit{then,} \qquad \widehat{y}_i^n \;\sim\; \mathcal{N}\left(Gx_i, Q_i^n\right), \quad \forall n. \tag{4.54}$$

Proof. Follows from equation (4.10) and Theorem 4.4.2. □

In vector form, the dynamics of $\widehat{y}_i$ is

$$\widehat{y}_i \;=\; 1_N \otimes (Gx_i) + q_i. \tag{4.55}$$

The key results are (4.53) and (4.54) of Corollary 4.4.3. Compare (4.54) with its centralized counter-part (4.49). We use (4.54) to design the distributed estimator gains K_i^n and analyze the convergence and performance of the DIKF next.

4.5 DIKF: Convergence Analysis

This section studies the DIKF filtering and prediction steps, (4.13)-(4.16), and compares them with the centralized filtering and prediction steps, (2.19)-(2.22). First, we analyze the filtering and prediction error processes and then show that the distributed prediction estimates $\widehat{x}_{i|i-1}^n$ of the dynamical system x_i are unbiased and their MSE are bounded. Given the distributed pseudo-observation estimates $\widehat{y}_i^n$ from the DCPO, we design, later in this section, the estimator gains K_i^n that minimize the MSE of the DIKF and hence improve its performance.

4.5.1 Error Analysis:

Given (2.1) and the pseudo-observation estimate (4.53), we aim to obtain at each agent the following prediction estimate,

$$\widehat{x}_{i|i-1}^n \;=\; \mathbb{E}[x_i \mid \widehat{y}_{0,i-1}^n]$$

where, $\widehat{y}_{0,i-1}^n = [(\widehat{y}_0^n)^T, \ldots, (\widehat{y}_{i-1}^n)^T]^T$. Another estimate of interest is the filtering estimate,

$$\widehat{x}_{i|i}^n \;=\; \mathbb{E}[x_i \mid \widehat{y}_{0,i}^n].$$

Recall the vector form of the distributed filter error $\epsilon_{i|i}$ in (4.10) and the distributed prediction error $\epsilon_{i+1|i}$ in (4.10) of the DIKF. Substitute (4.22) and (4.55) in (4.10), and (4.23) and (**??**) in (4.10),

$$\epsilon_{i|i} \;=\; \left(I_{MN} - K_i(I_N \otimes G)\right)\epsilon_{i|i-1} + K_i q_i \tag{4.56}$$

$$\epsilon_{i+1|i} \;=\; (I_N \otimes A)\left(I_{MN} - K_i(I_N \otimes G)\right)\epsilon_{i|i-1} + (I_N \otimes A)K_i q_i - (1_n \otimes v_i). \tag{4.57}$$

In the prediction error dynamics (4.57), v_i is independent of $\epsilon_{i|i-1}$ and q_i. We summarize the joint statistics of $\epsilon_{i|i-1}$ and q_i in the next lemma 4.5.1.

Lemma 4.5.1. *The error and noise sequences $\epsilon_{i|i-1}$ and q_i are statistically dependent and their time-varying covariance Π_i*

$$\begin{aligned}\Pi_i &= \mathbb{E}[\epsilon_{i|i-1} q_i^T] \qquad (4.58)\\ &= \left(I_N \otimes A\right)\left(I_{MN} - K_i(I_N \otimes G)\right)\Pi_{i-1}F^T + \left(I_N \otimes A\right)K_i Q_i F^T \\ &\qquad + (J \otimes V)\left(I_N \otimes G - \overline{D}_H\left(I_N \otimes B^T\right)\right)\end{aligned}$$

Proof. Equations (4.61) and (4.53) imply that $\epsilon_{i|i-1}$ and q_i are zero-mean. Write the covariance matrix of $\epsilon_{i|i-1}$ and $\overline{\phi}_i$ and expand $\epsilon_{i|i-1}$ and q_i using (4.57) and (4.50) respectively,

$$\begin{aligned}\mathbb{E}[\epsilon_{i|i-1} q_i^T] = \mathbb{E}\Bigg[\Bigg(&(I_N \otimes A)\Big(I_{MN} - K_{i-1}(I_N \otimes G)\Big)\epsilon_{i-1|i-2} \\ &+ (I_N \otimes A)K_{i-1}q_{i-1} - 1_n \otimes v_{i-1}\Bigg)\Big(Fq_{i-1} + \psi_{i-1}\Big)^T\Bigg] \\ = &(I_N \otimes A)\Big(I_{MN} - K_{i-1}(I_N \otimes G)\Big)\Pi_{i-1}F^T \\ &+ (I_N \otimes A)K_{i-1}Q_{i-1}F^T - \mathbb{E}\left[(1_N \otimes v_{i-1})\psi_{i-1}^T\right]. \qquad (4.59)\end{aligned}$$

From (4.50), we see that ψ_{i-1} consists of v_{i-1} and r_i. Therefore, expectations of the rest of the cross terms in (4.59) are zeros. Using (4.50) compute the only cross-covariance term in (4.59),

$$\mathbb{E}\left[(1_N \otimes v_{i-1})\psi_{i-1}^T\right] = (J \otimes V)\left(\overline{D}_H\left(I_N \otimes B^T\right) - I_N \otimes G\right)$$

and substitute it back in (4.59) to prove Lemma 4.5.1. □

So far we have formulated the statistics of all the noise sequences. Now we compute the mean of the error processes.

$$\begin{aligned}\mathbb{E}[\epsilon_{i|i}] &= \mathbb{E}[\widehat{x}_{i|i} - 1_N \otimes x_i] \\ &= \mathbb{E}\left[\mathbb{E}[\widehat{x}_{i|i} - 1_N \otimes x_i \mid \widehat{y}_{0,i}]\right] \\ &= \mathbb{E}\left[\widehat{x}_{i|i} - \widehat{x}_{i|i}\right] = 0 \qquad (4.60)\end{aligned}$$

$$\text{and similarly,} \qquad \mathbb{E}[\epsilon_{i+1|i}] = 0. \qquad (4.61)$$

The error covariances then follow from (4.56)-(4.61),

$$\begin{aligned}
\Sigma_{i|i} &= \mathbb{E}[\epsilon_{i|i}(\epsilon_{i|i})^T] \\
&= \left(I_{MN} - K_i(I_N \otimes G)\right) \Sigma_{i|i-1} \left(I_{MN} - (I_N \otimes G)K_i^T\right) \; + \; K_i Q_i K_i^T \\
&\quad + K_i \Pi_i^T \left(I_{MN} - (I_N \otimes G)K_i^T\right) + \left(I_{MN} - K_i(I_N \otimes G)\right) \Pi_i K_i^T \\
\Sigma_{i+1|i} &= \mathbb{E}[\epsilon_{i+1|i}(\epsilon_{i+1|i})^T] \\
&= (I_N \otimes A) \, \Sigma_{i|i} \left(I_N \otimes A^T\right) + J \otimes V. \qquad (4.62)
\end{aligned}$$

From equations (4.62) and (4.62), we see that the covariance of the error processes, and hence the performance of the DIKF, depends on the design of the gain matrices K_i^n.

4.5.2 Unbiasedness and minimum MSE:

This subsection analyzes the convergence of the proposed DIKF. We start with the design of the time-invariant distributed gain. The following theorem states and proves the asymptotic properties and convergence of the DIKF algorithm.

Theorem 4.5.1. *Given the pseudo-observation estimates $\widehat{y}_i^n$, the distributed prediction estimates $\widehat{x}_{i+1|i}^n$ of the dynamical system x_{i+1} are unbiased and their MSE are asymptotically bounded at each agent for all time-invariant gains K_i^n satisfying,*

$$\begin{aligned}
K_i^n &= \alpha G^{-1}, \qquad (4.63) \\
\textit{where,} \qquad \frac{\|A\|_2 - 1}{\|A\|_2} &< \alpha < \frac{\|A\|_2 + 1}{\|A\|_2}. \qquad \square
\end{aligned}$$

Proof. The zero mean of the predictor error $\epsilon_{i+1|i}^n$ in (4.61) shows that the distributed estimates $\widehat{x}_{i+1|i}^n$ are unbiased at each agent. Now consider the predictor error equation (4.57)

$$\epsilon_{i+1|i}^n = \underbrace{A(I_M - K_i^n G)}_{\overline{A}_i^n} \epsilon_{i|i-1}^n + \underbrace{A K_i^n q_i^n - v_i}_{s_i^n} \qquad (4.64)$$

where $\overline{A}_i^n$ is the error matrix and s_i^n is the noise process. Consider $K_i^n = \alpha G^{-1}$, then the spectral radius of $\overline{A}_i^n$ is,

$$\begin{aligned}
\rho(\overline{A}_i^n) &\leq \left\| A(I_M - \alpha G^{-1} G) \right\|_2 \\
&\leq \left\| A \right\|_2 \max\{|1-\alpha|, |\alpha - 1|\}
\end{aligned}$$

$$< \quad 1 \qquad\qquad \forall\ \frac{\|A\|_2 - 1}{\|A\|_2} < \alpha < \frac{\|A\|_2 + 1}{\|A\|_2},$$

and the noise $s_i^n = \alpha A G^{-1} q_i^n - v_i$ is zero-mean Gaussian, i.e., $s_i^n \sim \mathcal{N}(0, S_i^n)$ with

$$S_i^n = \alpha^2 A G^{-1} Q_i^n G^{-1} A^T + V. \tag{4.65}$$

Since the error covariance matrix Q_i^n is asymptotically bounded, the noise covariance S_i^n in (4.65) is also asymptotically bounded. Bounded S_i^n together with $\rho\left(\overline{A}_i^n\right) < 1,\ \forall\, i$, proves that the error process (4.64) is stable and the MSE of the distributed estimates $\widehat{x}_{i+1|i}^n$ are asymptotically bounded. □

Now we design the time-varying gain matrices K_i^n that provide minimum MSE distributed estimates $\widehat{x}_{i+1|i}^n$ of the dynamic system x_i. First consider the distributed filter (4.13) and define the distributed pseudo-innovations ν_i^n

$$\nu_i^n = \widehat{y}_i^n - G\widehat{x}_{i|i-1}^n. \tag{4.66}$$

Lemma 4.5.2 outlines the statistical properties of ν_i^n.

Lemma 4.5.2. *The pseudo-innovations ν_i^n are an uncorrelated sequence of zero-mean Gaussian random vectors.* □

Proof. The pseudo-innovations ν_i^n in (4.66) are a sequence of Gaussian random vectors. We compute its mean by substituting $\widehat{y}_i^n$ in (4.66) from (4.53) and then using (4.61),

$$\begin{aligned}\mathbb{E}\left[\nu_i^n\right] &= \mathbb{E}\left[Gx_i + q_i^n - G\widehat{x}_{i|i-1}^n\right] \\ &= G\mathbb{E}\left[\epsilon_{i|i-1}^n\right] + \mathbb{E}\left[q_i^n\right] = 0\end{aligned}$$

Now we compute the cross-covariances. Without loss of generality, consider $i > j$ and apply iterated law of expectations

$$\begin{aligned}\mathbb{E}\left[\nu_i^n (\nu_j^n)^T\right] &= \mathbb{E}\left[\left(G\epsilon_{i|i-1}^n + q_i^n\right)\left(G\epsilon_{j|j-1}^n + q_j^n\right)^T\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(G\epsilon_{i|i-1}^n + q_i^n\right)\left(G\epsilon_{j|j-1}^n + q_j^n\right)^T \mid \widehat{y}_{0,i-1}^n\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(G\epsilon_{i|i-1}^n + q_i^n\right) \mid \widehat{y}_{0,i-1}^n\right]\left(G\epsilon_{j|j-1}^n + q_j^n\right)^T\right]\end{aligned}$$

$$= \mathbb{E}\left[\left(G\underbrace{\mathbb{E}\left[\epsilon^n_{i|i-1} \mid \widehat{y}^n_{0,i-1}\right]}_{=0} + \underbrace{\mathbb{E}\left[q^n_i\right]}_{=0}\right)\left(G\epsilon^n_{j|j-1} + q^n_j\right)^T\right]$$
$$= 0.$$

□

Substitute (4.66) in (4.13), then the distributed filter reduces to

$$\widehat{x}^n_{i|i} \;=\; \widehat{x}^n_{i|i-1} + K^n_i \nu^n_i. \tag{4.67}$$

By Lemma 4.5.2, the pseudo-innovations ν^n_i are zero-mean, Gaussian and uncorrelated. We use the principles of the CKF and state the design of the gain K^n_i in the following lemma.

Lemma 4.5.3. *Given the distributed filter (4.67), the gain matrices that minimize the MSE are*

$$K^n_i = \Sigma_{x_i\nu^n_i}\Sigma^{-1}_{\nu^n_i}, \qquad \forall n \tag{4.68}$$

where, $\Sigma_{\nu^n_i}$ is the covariance of the pseudo-innovations ν^n_i and $\Sigma_{x_i\nu^n_i}$ is the cross-covariance between x_i and ν^n_i. □

Refer to [47] for the proof of the lemma. In the following theorem, we present the design of the time-varying gain matrices K^n_i that minimize the asymptotic MSE of the DIKF algorithm.

Theorem 4.5.2. *The gain matrices, K^n_i that provide the minimum asymptotic MSE distributed estimates of the field x_i are*

$$K^n_i = \left(\left(Q^n_i + \left(\Pi^n_i\right)^T G\right)\left(\Sigma^n_{i|i-1}G + \Pi^n_i\right)^{-1} + G\right)^{-1} \quad \forall n,$$

and the Lyapunov-type iteration of the error covariance is

$$\begin{aligned}\Sigma_{i+1|i} = (I_N \otimes A)\Big(&\left(I_{MN} - K_i(I_N \otimes G)\right)\Sigma_{i|i-1}\left(I_{MN} - (I_N \otimes G)K_i^T\right)\\ &+ K_i Q_i K_i^T + \left(I_{MN} - K_i(I_N \otimes G)\right)\Pi_i K_i^T\\ &+ K_i\Pi_i^T\left(I_{MN} - (I_N \otimes G)K_i^T\right)\Big)(I_N \otimes A^T) + V\end{aligned} \tag{4.69}$$

with the initial condition, $\Sigma_{0|-1} = J \otimes \Sigma_0$. □

Proof. Compute the matrices $\Sigma_{\nu_i^n}$ and $\Sigma_{x_i\nu_i^n}$,

$$\begin{aligned}
\Sigma_{\nu_i^n} &= \mathbb{E}[\nu_i^n(\nu_i^n)^T] = \mathbb{E}[(G\epsilon_{i|i-1}^n + q_i^n)(G\epsilon_{i|i-1}^n + q_i^n)^T] \\
&= G\Sigma_{i|i-1}^n G + Q_i^n + G\Pi_i^n + (\Pi_i^n)^T G \\
\Sigma_{x_i\nu_i^n} &= \mathbb{E}[(x_i - \overline{x}_i)(\nu_i^n)^T] \\
&= \mathbb{E}[(\epsilon_{i|i-1}^n + (\widehat{x}_i^n - \overline{x}_i))(G\epsilon_{i|i-1}^n + q_i^n)^T] \\
&= \Sigma_{i|i-1}^n G + \Pi_i^n.
\end{aligned}$$

The remaining terms in the above equations are all zero due to statistical independence of the noises. Now, these covariances along with (4.68) provide the optimal gain matrices

$$\begin{aligned}
K_i^n &= \left(\Sigma_{i|i-1}^n G + \Pi_i^n\right)\left(G\Sigma_{i|i-1}^n G + Q_i^n + G\Pi_i^n + \left(\Pi_i^n\right)^T G\right)^{-1} \\
&= \left(\left(Q_i^n + \left(\Pi_i^n\right)^T G\right)\left(\Sigma_{i|i-1}^n G + \Pi_i^n\right)^{-1} + G\right)^{-1}.
\end{aligned}$$

Define the block gain matrix $K = \text{blockdiag}\{K_i^1, \ldots, K_i^N\}$. Substitute K_i and the Lyapunov-type filter covariance $\Sigma_{i|i}$ equation (4.62) into (4.62) to obtain the Lyapunov-type predictor covariance (4.69). The equation (4.69) along with (4.58) and (4.11) is the distributed version of the algebraic Riccati equation. □

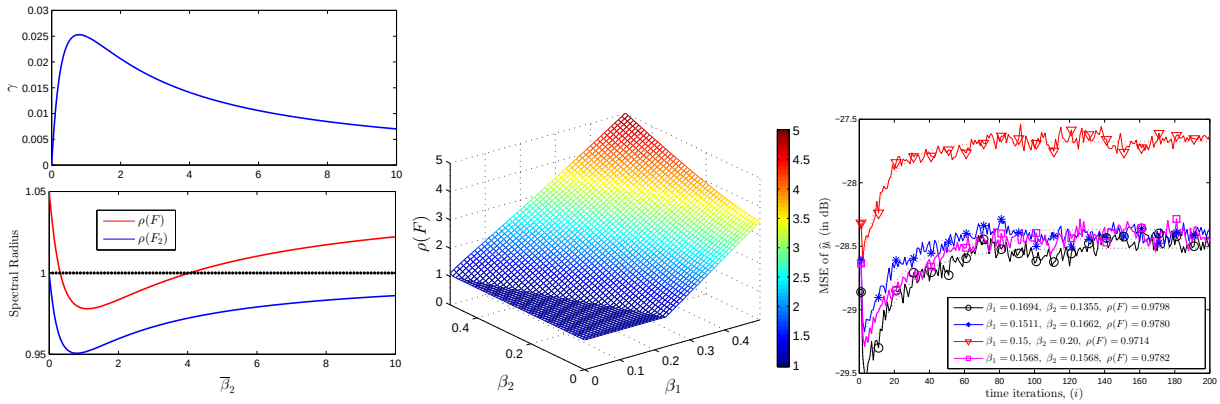
The algebraic Riccati equation (4.69) quantifies the performance of the DIKF algorithm. The trace of the covariance matrix gives the MSE of the distributed estimates $\widehat{x}_i^n$. It is the minimum MSE solution of the problem given the distributed pseudo-observation estimates $\widehat{y}_i^n$, from the DCPO algorithm. Note that these time-varying K_i^n are computationally more expensive than the time invariant gain matrices stated in Theorem 4.5.1, but provide better asymptotic performance.

Algorithm 3 Gain Design of DIKF

Input: Model parameters A, V, H, R, G, L, Σ_0.
Dynamic Averaging: Compute consensus weight β_1 and pseudo-observation weight β_2 using (4.34)-(4.46).
Field Estimation: Compute the gain K_i^n.
if Scalar Gain **then**
  Compute the scalar gain $K_i^n = \alpha G^{-1}$ using (4.63).
else
  Compute the optimized filter gain matrix K_i^n using the following iterative approach.
  Initialize: $\Sigma_{0|-1}$, Q_0 and Π_0 using (4.18)-(4.20).
  while $i \geq 0$ **do**
   Filter gain:
    Compute K_i^n using (4.67)-(4.69).
   Prediction error covariance updates:
    Update $P_{i+1}, \Sigma_{i+1|i}, \Pi_{i+1}$ using (4.11), (4.16), (4.17).
  end while
end if

4.6 Experimental Evaluations

We simulate, using Matlab, a discrete-time dynamical system, A, with $M = 20$ state variables and unstable time-invariant dynamics, i.e., $||A||_2 = 1.05$. Note that time-varying A requires a completely different set of analysis, which is beyond the scope of this paper. For simplicity, we chose A to be symmetric. This is not required and all the results hold for any other choice of non-symmetric dynamics. A group of $N = 20$ agents observes the dynamical system, where each agent observes only four (possibly different) state variables, i.e., $M_n = 4$. The covariance matrices of the input noise V, of the observation noises R, and of the initial condition Σ_0 are randomly generated and then squared to make them positive definite. The spectral norm of the covariance matrices are $\|V\|_2 = 2$, $\|R\|_2 = 16$ and $\|\Sigma_0\|_2 = 8$. The mean of the initial condition, $\overline{x}_0$, is randomly generated. The system and observation model is not locally observable but is globally detectable (Assumption 4) and distributedly observable (Assumption 7), i.e., G in (4.5) is invertible. We consider a lattice graph [57] with 20 nodes and 60 edges for the agent communication network. This network is connected (Assumption 5) with a positive algebraic connectivity, $\lambda_2(L) = 1.3$, of the Laplacian matrix L of the lattice graph. Based on this experimental setup, we evaluate the performance of dynamic consensus on pseudo-observation and Distributed Information Kalman



a γ,$\rho(F)$, $\rho(F_2)$ Vs $\overline{\beta}_2$ b $\rho(F)$ Vs β_1, β_2 c MSE of DCPO Step

Figure 4.1: (4.1a) Plot of γ and $\rho(F), \rho(F_2)$ with respect to $\overline{\beta}_2$. (4.1b) Plot of $\rho(F)$ with respect to β_1 and β_2. (4.1c) MSE (in dB) plots of distributed estimates $\widehat{y}_i$ with time for the sets of β_1 and β_2 shown in Table 4.1.

Filter (DIKF) by Monte-Carlo simulations.

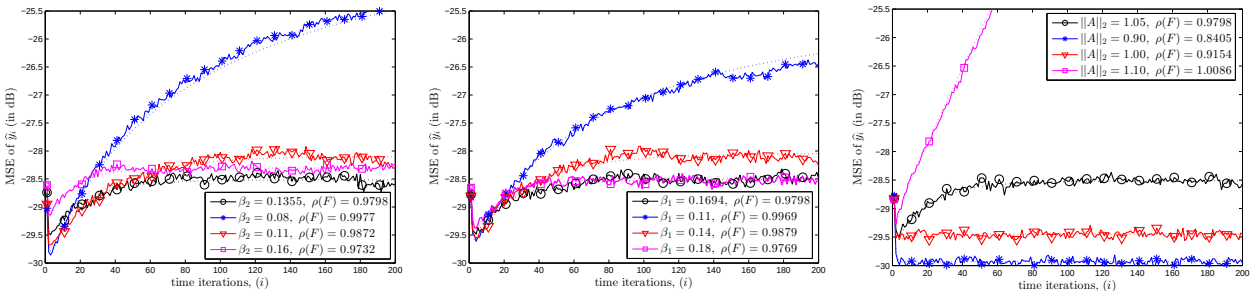
4.6.1 DCPO: Convergence and Performance Results

The first step for the dynamic consensus on the pseudo-observations is to design the constant weight consensus matrix, W, and the pseudo-observation gain matrix, B, i.e., to design the two parameters β_1 and β_2. Using equations (4.35)-(4.40), we plot the diffusion rate, $\frac{\lambda_1}{\lambda_N}$, and the corresponding spectral radii, $\rho(F)$ and $\rho(F_2)$, with varying $\overline{\beta}_2 = \frac{\beta_2}{\beta_1}$ in Fig. 4.1a. For each $\overline{\beta}_2$, we compute β_1 using (4.46) and $\beta_2 = \overline{\beta}_2\beta_1$. We select the maximizer $\overline{\beta}_2^*$ that maximizes the diffusion rate and minimizes $\rho(F_2)$, and compute one set (Case I in Table 4.1) of parameters (β_1, β_2) that yields $\rho(F) < 1$. Note that the $\overline{\beta}_2^*$ that minimizes $\rho(F_2)$ may not yield the minimum $\rho(F)$ as evident from Fig. 4.1a. We consider the set (β_1, β_2) satisfying the $\overline{\beta}_2^*$ that yields minimum $\rho(F)$ as another acceptable set of parameters (Case II in Table 4.1). Another way of analyzing $\rho(F)$ is by

	$\overline{\beta}_2$	β_1	β_2	$\rho(F)$
Case I	0.8	0.1694	0.1355	0.9798
Case II	1.1	0.1511	0.1662	0.9780
Case III	1.33	0.15	0.20	0.9714
Case IV	1.0	0.1568	0.1568	0.9782

Table 4.1: Sets of β_1 and β_2 used in Fig. 4.1c.

varying β_1 and β_2 over randomly selected ranges and then computing $\rho(F)$ using (4.35). We vary



a MSE with varying β_2 b MSE with varying β_1 c MSE with varying $||A||_2$

Figure 4.2: MSE (in dB) plots of distributed estimates $\widehat{y}_i$ with time for different: (4.2a) β_2; (4.2b) β_1; and (4.2c) $||A||_2$.

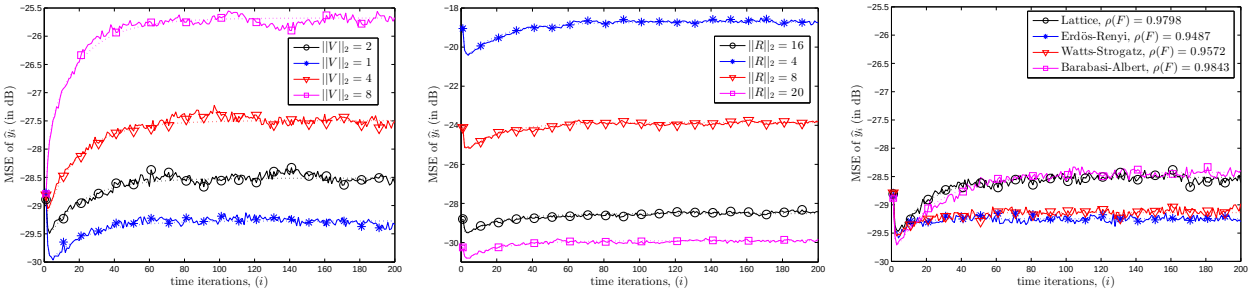
both β_1 and β_2 in the range $[0,\ 5]$ and the corresponding plot of $\rho(F)$ is shown in Fig 4.1b. The set of parameters (β_1, β_2) that minimizes $\rho(F)$ is Case III in Table 4.1. In Case IV, we consider a special case $\beta_1 = \beta_2$ and obtain the optimal set of parameters using (4.46) that yields $\rho(F) < 1$.

In Table 4.1, we accumulated four acceptable designs of the matrices W and B, i.e., β_1 and β_2, that satisfy Assumption 4. With these four sets of β_1 and β_2, we simulate the dynamic averaging algorithm of the distributed pseudo-innovations estimate $\widehat{y}_i^n$ up to 200 time iterations. Then we compute the MSE (in dB) of the distributed estimates by 500 Monte-Carlo simulations. From Fig. 4.1c, we see that the dynamic consensus step (4.10) enables each agent to obtain the unbiased distributed estimates $\widehat{y}_i^n$ of the global average $\overline{y}_i^n$ with bounded MSE. Note that the Monte-Carlo simulated results match the theoretical MSE obtained from the Lyapunov type iteration (4.48) of the pseudo-observation error covariance matrix P_i. For the remaining simulations, we consider Case I in Table 4.1 as the reference plot (black line with circle markers), since it yields the lowest asymptotic MSE amongst these four cases.

We now analyze the performance of dynamic consensus step by varying different parameters. First we plot the performance of dynamic consensus step with fixed β_1 and varying β_2 in Fig 4.2a. Similarly, we repeat the Monte-Carlo simulations with fixed β_2 and varying β_1, as shown in Fig 4.2b. In all these simulations, we see that the reference Case I yields the lowest asymptotic MSE. Next we study the performance of dynamic consensus step for different dynamics matrix A, input noise covariance V and measurement noise covariance R. We randomly generate (as described earlier) different A, V, and R with specific spectral norms. We plot the MSE of the distributed estimates $\widehat{y}_i^n$ with varying $||A||_2$, $||V||_2$, and $||R||_2$ in Fig 4.2c, Fig 4.3a, and Fig 4.3b, respectively. From Fig 4.2c,

	λ_2	β_1	β_2	$\rho(F)$
Lattice	1.3043	0.1694	0.1355	0.9798
Erdös-Renýi	2.3293	0.1079	0.1511	0.9847
Watts-Strogatz	2.0193	0.1197	0.1436	0.9572
Barabasi-Albert	1.8376	0.0708	0.1133	0.9843

Table 4.2: Connectivity and gain parameters used in Fig. 4.3c.



a MSE with varying $||V||_2$ b MSE with varying $||R||_2$ c MSE for different network models

Figure 4.3: MSE (in dB) plots of distributed estimates $\widehat{y}_i$ with time for different: (4.3a) $||V||_2$; (4.3b) $||R||_2$; and (4.3c) network models.

we see that the asymptotically stable dynamics with $||A||_2 = 0.90$ yield the lowest asymptotic MSE. When $||A||_2 = 1.10$ the error process is unstable as $||A||_2$ is greater than the NTC (1.0519). If the degree of instability $||A||_2$ of the field dynamics is greater than the NTC, then it is not possible to design β_1 and β_2, such that $\rho(F) < 1$. From Fig 4.3a, we see that the asymptotic MSE of dynamic consensus step increases with increasing $||V||_2$, whereas it decreases with $||R||_2$ as shown in Fig 4.3b.

Finally in Fig 4.3c, we plot the performance of the dynamic consensus step for varying agent communication network models. We consider Erdös-Renýi [57], Watts-Strogatz [58], and Barabasi-Albert [59] graphs with the same number of nodes $(N = 20)$ and edges $(E = 60)$ as the Lattice graph. Following a similar approach, we compute the gain parameters, β_1 and β_2, for each network. The algebraic connectivity, λ_2, and the gain parameters, β_1 and β_2, are listed in Table 4.2. From Fig 4.3c, we see that the Lattice model with the lowest λ_2 has the maximum asymptotic MSE, whereas the Erdös-Renýi model with the highest λ_2 has the minimum asymptotic MSE. The asymptotic MSE is sensitive to the algebraic connectivity λ_2 of the network. We will use this Erdös-Renýi network for the DIKF analysis in Subsection 4.6.2. From the simulation plots we see that the performance of the DCPO step depends on a combination of multiple factors including the design parameters, noise

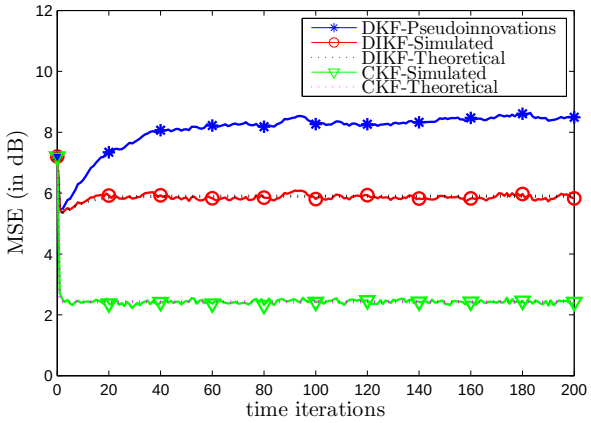
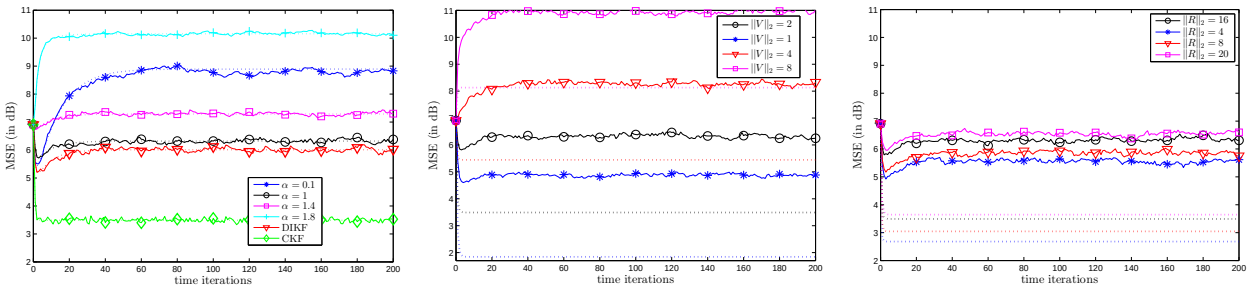


Figure 4.4: MSE of DIKF - theoretical and simulated, CKF, and, DKF-Pseudoinnovations (PIKF in Chapter 3)

covariances, field dynamics and the agent communication network models. There is no monotonic relationship between $\rho(F)$ and asymptotic MSE. In Fig 4.1a, due to varying gain parameters, higher $\rho(F)$ lead to lower asymptotic MSE, whereas in Fig 4.3c, lower $\rho(F)$ lead to lower asymptotic MSE due to the substantially increased connectivity caused by the network models. In this context, note that we do not expect any monotonic or dramatic changes in the performance of the distributed estimates with varying M_n as long as the field dynamics, A, is locally undetectable at each agent but globally detectable (Assumption 1) and distributedly observable (Assumption 3). Other observation models, i.e., sensor placements, may yield different asymptotic MSE performance, as is to be expected like in centralized Kalman filtering.

4.6.2 DIKF: Convergence and Performance Results

Now, we use the distributed pseudo-observations estimate $\widehat{y}_i^n$ in the DIKF algorithm (4.13)-(4.16). Then we compute the MSE (in dB) of the distributed state estimates by 500 Monte-Carlo simulations. We also compute the theoretical MSE of DIKF using the Riccati equation (4.69). Note that the distributed gain matrices K_i^n are computationally expensive. So we also Monte-Carlo simulate the DIKF with time invariant gain matrices $K^n = \alpha G^{-1}$ and compute the corresponding theoretical MSE asymptotics for different α. We compare these simulated and theoretical MSE of DIKF with the centralized Kalman filter and the PIKF presented in Chapter 3. The plot shown in Fig. 4.4 shows that DIKF improves 2dB over the PIKF, while its MSE performance gap from the CKF is 3dB. Fig. 4.5a shows that the DIKF is capable of distributed estimation of the unstable dynamical system



a MSE of DIKF b MSE with varying $||V||_2$ c MSE with varying $||R||_2$

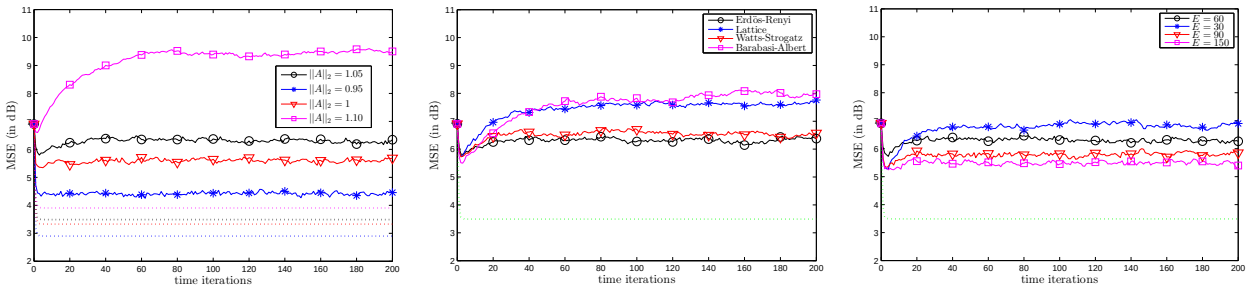
Figure 4.5: Comparison of MSE between centralized filter, the proposed DIKF with time-varying K_i^n, and, time-invariant $K_i^n = \alpha G^{-1}$ for different α. MSE (in dB) plots of distributed estimates $\widehat{x}_i$ with time for: (4.5b) different $||V||_2$; and, (4.5c) different $||R||_2$.

with bounded MSE as expected from the results in Theorems 4.5.1 and 4.5.2. The simulated MSE of the DIKF has a performance gap of about 2.5 dB when compared to the centralized Kalman filter. On the other hand, with varying α, the best performance is obtained for $\alpha = 1$. Choosing among time-varying and time-invariant gain matrices is a tradeoff between computation complexity and asymptotic MSE performance. For the remaining simulations, we consider time-invariant gain matrices with $\alpha = 1$ as the reference case (black line with circle markers). Now, we study the

	λ_2	β_1	β_2	$\rho(F)$
E = 60	2.3293	0.1079	0.1511	0.9487
E = 30	0.3127	0.2240	0.1120	0.9694
E = 90	3.6303	0.0802	0.1604	0.9353
E = 150	11.8390	0.0559	0.1788	0.9104

Table 4.3: Connectivity and gain parameters used in Fig. 4.6c.

performance of DIKF for different input noise covariance V, measurement noise covariance R, and dynamics matrices A. We randomly generate (similar to dynamic consensus) different V, R, and A with specific spectral norms. We plot the MSE of the distributed estimates $\widehat{x}_i^n$ with varying $||V||_2$, $||R||_2$ and A in Fig 4.5b, Fig 4.5c, and Fig 4.6a respectively. From Fig 4.5b and 4.5c, we see that with increasing $||V||_2$ and $||R||_2$ the asymptotic MSE of both DIKF and centralized Kalman filter (corresponding dotted lines) increase proportionally. But both DIKF and centralized Kalman filter are more robust to the increase in measurement noise level as compared to the input noise level. In contrast, increasing the degree of instability of the dynamics matrix $||A||_2$ affects the performance of the DIKF significantly more than the centralized Kalman filter, as evident from Fig



a MSE with varying $||A||_2$ b MSE for different network models c MSE with varying E

Figure 4.6: Mean-squared error (MSE in dB) plots of distributed estimates $\widehat{x}_i$ with time for: (4.6a) different $||A||_2$; (4.6b) different network models; and, (4.6c) varying number of edges E in the network.

4.6a. That is why $||A||_2$ plays a key role in the convergence of the DIKF. Finally in Fig 4.6b and Fig 4.6c, we analyze the performance of the DIKF with varying agent communication networks, considering both different network models and different number of edges. Similar to dynamic consensus, we consider Erdös-Renýi, Watts-Strogatz, and Barabasi-Albert graphs with the same number of nodes $(N = 20)$ and edges $(E = 60)$ that yield the parameters listed in Table 4.2. From Fig 4.6b, we see that the Barabasi-Albert model results in maximum asymptotic MSE, whereas the Erdös-Renýi model achieves minimum asymptotic MSE. On the other hand, when we consider an Erdös-Renýi network model and vary the number of edges, E, in the network, we obtain the set of parameters shown in Table 4.3. Fig 4.6c illustrates that, increasing the number of edges in the network, increases the algebraic connectivity, (λ_2), and hence results in lower asymptotic MSE of DIKF. The performance gap between the DIKF and centralized Kalman filter depends on the design of the gain matrices, the input and the measurement noises, the system dynamics, and the agent network model and its connectivity.

4.7 Conclusions

This chapter proposes a novel distributed filter, DIKF, for the estimation of unstable time-varying random fields over sensor networks to achieve unbiased distributed estimates with bounded MSE. The distributed filter includes a new *consensus+innovations* type dynamic consensus on pseudo-observations (DCPO) algorithm. We proved that both distributed filter and dynamic consensus converge asymptotically. Given the pseudo-observations estimates from the DCPO, we designed

the gain matrices of the distributed filter algorithm such that the estimation MSE is minimized.

The main contributions of this chapter are: (i) an explicit expression for the Network Tracking Capacity of the distributed estimation based on the eigenvalues of the network Laplacian and the local observation matrices; (ii) analytical design of the general estimator gain matrices and subsequent characterization of a distributed version of the algebraic Riccati equation; and (iii) experimental evaluation of the sensitivity of model parameters, noise statistics, gain variations, and network models on the performance of the distributed estimator.

CHAPTER 5

Consensus+Innovations Kalman Filter

5.1 Introduction

In Chapter 4, we proposed the *Distributed Information Kalman Filter* (DIKF) that is a distributed estimator of time-varying random fields consisting of two substructures. The first is the *Dynamic Consensus on Pseudo-Observations* (DCPO), a distributed estimator of the global average of the pseudo-observations (modified versions of the observations) of the agents. The second substructure uses these average estimates of pseudo-observations to estimate the time-varying random field. In this chapter, we develop a distributed Kalman filter like estimator, the *Consensus+Innovations* Kalman Filter (CIKF), that instead of using the pseudo-observations uses distributed estimates of the pseudo-state (modified version of a state) to estimate the field. We show how to design optimally the gain matrices of the CIKF. We prove that the CIKF converges in the mean-squared error (MSE) sense when the degree of instability of the dynamics of the random field is within the network tracking capacity [30], a threshold determined by the cyber network connectivity and the local observation models. Numerical simulations show that the proposed CIKF improves the performance by 3dB over the DIKF, reducing by half the gap to the centralized (optimal) Kalman filter, while showing a faster convergence rate than the DIKF. These improvements significantly distinguish the CIKF from the DIKF.

The rest of the chapter is organized as follows. Section 5.2 introduces the pseudo-state and presents the proposed optimal gain distributed Kalman filter (CIKF). In Section 5.3, we analyze the dynamics of the error processes and their covariances. Section 5.4 includes the analysis of the tracking capacity of the proposed CIKF. We design the optimal gain matrices to obtain distributed

estimates in Section 5.5. Numerical simulations are in Section 5.6. We present the concluding remarks in Section 5.7.

5.2 Distributed filtering and prediction

This section considers our single time-scale distributed solution. We start with the introduction and derivation of the key components of our distributed estimator and then present our distributed field estimator.

5.2.1 Pseudo-state model

In a centralized information filter [47], all the observations are converted into pseudo-observations [55] to obtain the optimal estimates. Following (2.2), the pseudo-observation $\widetilde{\boldsymbol{z}}_i^n$ of agent n is

$$\widetilde{\boldsymbol{z}}_i^n = H_n^T R_n^{-1} \boldsymbol{z}_i^n = \overline{H}_n \boldsymbol{x}_i + H_n^T R_n^{-1} \boldsymbol{r}_i^n, \tag{5.1}$$

$$\text{where,} \qquad \overline{H}_n = H_n^T R_n^{-1} H_n. \tag{5.2}$$

The centralized information filter computes the sum, $\overline{\boldsymbol{z}}_i$ of all the pseudo-observations

$$\overline{\boldsymbol{z}}_i = \sum_{n=1}^{N} \widetilde{\boldsymbol{z}}_i^n = G\boldsymbol{x}_i + H^T R^{-1} \boldsymbol{r}_i \tag{5.3}$$

$$\text{where,} \qquad G = \sum_{n=1}^{N} H_n^T R_n^{-1} H_n = \sum_{n=1}^{N} \overline{H}_n. \tag{5.4}$$

Note that the definition of G in (5.4) is different from the G defined in (4.5). Here G is the sum of all the pseudo-observations, whereas in (4.5) it is the average of the pseudo-observations. The aggregated pseudo-observation, $\overline{\boldsymbol{z}}_i$, is the key term in the centralized filter. It provides the innovations term in the filter updates enabling the filter to converge with minimum MSE estimates. However, in the distributed solution, each agent n does not have access to all the pseudo-observations; instead it can only communicate with its neighbors. To address this issue, in [55] we introduced a dynamic consensus algorithm to compute the distributed estimates of the averaged pseudo-observations, $\overline{\boldsymbol{z}}_i$, at each agent. In (5.3), we note that the crucial term is $G\boldsymbol{x}_i$ which carries the information of the dynamic state, $\boldsymbol{x}_i$; and the second term in $\overline{\boldsymbol{z}}_i$ in (5.3) is noise. We refer to it as the pseudo-state, $\boldsymbol{y}_i$,

$$\boldsymbol{y}_i = G\boldsymbol{x}_i. \tag{5.5}$$

The pseudo-state, $\boldsymbol{y}_i$, is also a random field whose time dynamics can be represented by a discrete-time linear dynamical system. The pseudo-observations $\widetilde{\boldsymbol{z}}_i^n$ are its linear measurements. We summarize the state-space model for the pseudo-state in the following proposition.

Proposition 3. *The dynamics and observations of the pseudo-state $\boldsymbol{y}_i$ are:*

$$\boldsymbol{y}_{i+1} = \widetilde{A}\boldsymbol{y}_i + G\boldsymbol{v}_i + \check{A}\boldsymbol{x}_i \tag{5.6}$$

$$\widetilde{\boldsymbol{z}}_i^n = \widetilde{H}_n\boldsymbol{y}_i + H_n^T R_n^{-1}\boldsymbol{r}_i^n + \check{H}_n\boldsymbol{x}_i. \tag{5.7}$$

The pseudo-dynamics matrix $\widetilde{A}$, pseudo-observations matrix $\widetilde{H}_n$, and the matrices $\check{A}, \check{H}_n$, and $\widetilde{I}$ at agent n are:

$$\widetilde{A} = GAG^\dagger \tag{5.8}$$

$$\widetilde{H}_n = H_n^T R_n^{-1} H_n G^\dagger = \overline{H}_n G^\dagger \tag{5.9}$$

$$\check{A} = GA\widetilde{I} \tag{5.10}$$

$$\check{H}_n = H_n^T R_n^{-1} H_n \widetilde{I} \tag{5.11}$$

$$\widetilde{I} = I - G^\dagger G, \tag{5.12}$$

where, $G^\dagger$ denotes the Moore-Penrose pseudo-inverse of G.

Proof. First we derive the dynamics (5.6) of the pseudo-state $\boldsymbol{y}_i$. Using (5.5) and (2.1),

$$\begin{aligned}
\boldsymbol{y}_{i+1} &= G\boldsymbol{x}_{i+1} \\
&= G\left(A\boldsymbol{x}_i + \boldsymbol{v}_i\right) \\
&= GA\left(G^\dagger G + \widetilde{I}\right)\boldsymbol{x}_i + G\boldsymbol{v}_i, \quad \left[\text{by (5.12), } I = G^\dagger G + \widetilde{I}\right] \\
&= GAG^\dagger\boldsymbol{y}_i + G\boldsymbol{v}_i + GA\widetilde{I}\boldsymbol{x}_i \\
&= \widetilde{A}\boldsymbol{y}_i + G\boldsymbol{v}_i + \check{A}\boldsymbol{x}_i, \qquad \text{[by (5.5)]}\,.
\end{aligned}$$

Now, we derive the observations (5.7) of the pseudo-observations $\widetilde{\boldsymbol{z}}_i^n$. Using (5.1) and (2.2),

$$\begin{aligned}
\widetilde{\boldsymbol{z}}_i^n &= H_n^T R_n^{-1} H_n \boldsymbol{x}_i + H_n^T R_n^{-1}\boldsymbol{r}_i^n \\
&= H_n^T R_n^{-1} H_n \left(G^\dagger G + \widetilde{I}\right)\boldsymbol{x}_i + H_n^T R_n^{-1}\boldsymbol{r}_i^n \\
&= H_n^T R_n^{-1} H_n G^\dagger G\boldsymbol{x}_i + H_n^T R_n^{-1}\boldsymbol{r}_i^n + H_n^T R_n^{-1} H_n \widetilde{I}\boldsymbol{x}_i
\end{aligned}$$

$$= \widetilde{H}\boldsymbol{y}_i + H_n^T R_n^{-1}\boldsymbol{r}_i^n + \check{H}_n\boldsymbol{x}_i.$$

□

In [55], the distributed information Kalman filter (DIKF) assumes distributed observability, i.e., it considers the case where G is invertible. Under this assumption, $G^\dagger = G^{-1}$ and $\widetilde{I} = 0$. In this chapter we relax the requirement of invertibility of G, proposing a distributed estimator for general dynamics-observation models under the assumption of global detectability. In most cases $\widetilde{I}$ is low-rank. In (5.6), the term $\Big(G\boldsymbol{v}_i \ + \ \check{A}\boldsymbol{x}_i = \xi_i, \text{say}\Big)$ can be interpreted as the pseudo-state input noise, which follows Gauss dynamics

$$\xi_i \sim \mathcal{N}\left(\check{A}\overline{\boldsymbol{x}}_i \ , \ GVG + \check{A}\Sigma_i\check{A}^T\right)$$
$$\text{where,} \qquad \Sigma_i = \mathbb{E}\left[(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)^T\right].$$

Similarly, in (5.7), the term $(\delta_i^n = H_n^T R_n^{-1}\boldsymbol{r}_i^n + \check{H}_n\boldsymbol{x}_i)$ is the pseudo-state observation noise at agent n, which is Gaussian

$$\delta_i^n \sim \mathcal{N}\Big(\check{H}_n\overline{\boldsymbol{x}}_i, \overline{H}_n + \check{H}_n\Sigma_i\check{H}_n\Big).$$

For the ease of analysis, we express the pseudo-state observation model in vector form by $\widetilde{\boldsymbol{z}}_i \in \mathbb{R}^{\sum_{n=1}^N M_n}$, the aggregate of the noisy local temperature measurements, $\widetilde{\boldsymbol{z}}_i^1, \cdots \widetilde{\boldsymbol{z}}_i^N$, of all the agents,

$$\underbrace{\begin{bmatrix}\widetilde{\boldsymbol{z}}_i^1\\ \vdots \\ \widetilde{\boldsymbol{z}}_i^N\end{bmatrix}}_{\widetilde{\boldsymbol{z}}_i} = \underbrace{\begin{bmatrix}\widetilde{H}_1\\ \vdots \\ \widetilde{H}_N\end{bmatrix}}_{\widetilde{H}}\boldsymbol{y}_i + D_H^T R^{-1}\boldsymbol{r}_i + \underbrace{\begin{bmatrix}\check{H}_1\\ \vdots \\ \check{H}_N\end{bmatrix}}_{\check{H}}\boldsymbol{x}_i, \tag{5.13}$$

where, the matrices $\widetilde{H}, \check{H} \in \mathbb{R}^{NM\times M}$, and, $D_H = \text{blockdiag}\{H_1, \cdots, H_N\}$.

We have established in (5.6)-(5.7) the state-space dynamics and observations model of the pseudo-state, $\boldsymbol{y}_i$. The structure of (5.6)-(5.7) is similar to the dynamics and observations model (2.1)-(2.2) of the random field, $\boldsymbol{x}_i$. In the following subsection, we develop a distributed estimator with optimized gains to obtain unbiased estimates of the pseudo-state, $\boldsymbol{y}_i$, and of the state, $\boldsymbol{x}_i$, at each agent with minimized mean-squared error.

5.2.2 *Consensus*+*Innovations* Kalman Filter (CIKF)

At time i, denote the n^{th} agent's distributed filter and prediction estimates of the state $\boldsymbol{x}_i$ by $\widehat{\boldsymbol{x}}^n_{i|i}$ and $\widehat{\boldsymbol{x}}^n_{i+1|i}$ respectively. Similarly, its distributed filter and prediction estimates of the pseudo-state $\boldsymbol{y}_i$ are denoted by $\widehat{\boldsymbol{y}}^n_{i|i}$ and $\widehat{\boldsymbol{y}}^n_{i+1|i}$. At any time i, each agent n has access to its own pseudo-observation $\widetilde{\boldsymbol{z}}^n_i$ and receives the prediction state pseudo-state estimates, $\widehat{\boldsymbol{y}}^l_{i|i-1}$, $l \in \Omega_n$, of its neighbors at the previous time $i-1$. Under this setup, the minimized MSE filter and prediction estimates are the conditional means,

$$\widehat{\boldsymbol{y}}^n_{i|i} = \mathbb{E}\Big[\boldsymbol{y}_i \mid \widetilde{\boldsymbol{z}}^n_i, \{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\overline{\Omega}_n}\Big] \tag{5.14}$$

$$\widehat{\boldsymbol{x}}^n_{i|i} = \mathbb{E}\Big[\boldsymbol{x}_i \mid \widehat{\boldsymbol{y}}^n_{i|i}\Big] \tag{5.15}$$

$$\widehat{\boldsymbol{y}}^n_{i+1|i} = \mathbb{E}\Big[\boldsymbol{y}_{i+1} \mid \widetilde{\boldsymbol{z}}^n_i, \{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\overline{\Omega}_n}\Big] \tag{5.16}$$

$$\widehat{\boldsymbol{x}}^n_{i+1|i} = \mathbb{E}\Big[\boldsymbol{x}_{i+1} \mid \widehat{\boldsymbol{y}}^n_{i|i}\Big]. \tag{5.17}$$

In (5.14), $\widehat{\boldsymbol{y}}^n_{i|i}$ is the filtered estimate of the pseudo-state $\boldsymbol{y}_i$ given all the pseudo-observations available at agent n up to time i including those of its neighbors. By the principle of recursive linear estimation, instead of storing all the pseudo-observations $\big\{\{\widetilde{\boldsymbol{z}}^n_t\}_{t=0,\cdots,i}, \{\widetilde{\boldsymbol{z}}^{n_1}_t\}^{n_1\in\Omega_n}_{t=0,\cdots,i-1}, \{\widetilde{\boldsymbol{z}}^{n_2}_t\}^{n_2\in\Omega_{n1},\forall n_1}_{t=0,\cdots,i-2}, \cdots\big\}$, we need only the current pseudo-observation $\widetilde{\boldsymbol{z}}^n_i$ and the pseudo-state estimates (including those of its neighbors) from the previous time instant $\{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\Omega_n}$. The filtered estimate $\widehat{\boldsymbol{x}}^n_{i|i}$ of $\boldsymbol{x}_i$ in (5.15) depends on the current pseudo-state filtered estimate $\widehat{\boldsymbol{y}}^n_{i|i}$. Similarly, the prediction estimates $\widehat{\boldsymbol{y}}^n_{i+1|i}$ and $\widehat{\boldsymbol{x}}^n_{i+1|i}$ of $\boldsymbol{y}_{i+1}$ and $\boldsymbol{x}_{i+1}$ in (5.16)-(5.17), respectively, are conditioned on the corresponding available quantities up to time i.

Theorem 5.2.1. *The iterative updates to compute the distributed filtered estimates with optimized gains in* (5.14)-(5.15) *are:*

$$\widehat{\boldsymbol{y}}^n_{i|i} = \widehat{\boldsymbol{y}}^n_{i|i-1} + \sum_{l\in\Omega_n} \underbrace{B^{nl}_i\Big(\widehat{\boldsymbol{y}}^l_{i|i-1} - \widehat{\boldsymbol{y}}^n_{i|i-1}\Big)}_{\textit{Consensus}} + B^{nn}_i \underbrace{\Big(\widetilde{\boldsymbol{z}}^n_i - \Big(\widetilde{H}_n\widehat{\boldsymbol{y}}^n_{i|i-1} + \breve{H}_n\widehat{\boldsymbol{x}}^n_{i|i-1}\Big)\Big)}_{\textit{Innovations}}, \tag{5.18}$$

$$\widehat{\boldsymbol{x}}^n_{i|i} = \widehat{\boldsymbol{x}}^n_{i|i-1} + K^n_i \underbrace{\Big(\widehat{\boldsymbol{y}}^n_{i|i} - G\widehat{\boldsymbol{x}}^n_{i|i-1}\Big)}_{\textit{Innovations}} \tag{5.19}$$

where, pseudo-state gain block matrix, $B_i \in \mathbb{R}^{MN\times MN}$, *is* $B_i = \Big[B^{nl}_i\Big]^{l=1,\cdots,N}_{n=1,\cdots,N}$, $B^{nl}_i = 0$ *if* $l \notin \Omega_n$. *The state gain block-diagonal matrix is* $K_i = \textit{blockdiag}\{K^1_i, \cdots, K^N_i\}$, $K^n_i \in \mathbb{R}^{M\times M}$. *The*

optimized MSE prediction estimates in (5.16)-(5.17) *are:*

$$\widehat{\boldsymbol{y}}_{i+1|i}^n = \widetilde{A}\widehat{\boldsymbol{y}}_{i|i}^n + \check{A}\widehat{\boldsymbol{x}}_{i|i}^n, \tag{5.20}$$

$$\widehat{\boldsymbol{x}}_{i+1|i}^n = A\widehat{\boldsymbol{x}}_{i|i}^n. \tag{5.21}$$

Proof. In Lemma 5.5.1, we showed that $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{\nu}_i^n$ are independent Gaussian sequences. By the Innovations Property [47], there are $1-1$ correspondence between $\{\widetilde{\boldsymbol{z}}_i^n, \{\widehat{\boldsymbol{y}}_{i|i-1}^l\}_{l\in\overline{\Omega}_n}\}$ and $\widetilde{\boldsymbol{\nu}}_i^n$, and between $\{\widehat{\boldsymbol{y}}_{i|i}^n\}$ and $\boldsymbol{\nu}_i^n$. The Innovations Property guarantees that there exists a unique way to get one from the other.

$$\widehat{\boldsymbol{y}}_{i|i}^n = \mathbb{E}\Big[\boldsymbol{y}_i \mid \widetilde{\boldsymbol{z}}_i^n, \{\widehat{\boldsymbol{y}}_{i|i-1}^l\}_{l\in\overline{\Omega}_n}\Big] \iff \widehat{\boldsymbol{y}}_{i|i}^n = \mathbb{E}\big[\boldsymbol{y}_i \mid \widetilde{\boldsymbol{\nu}}_i^n\big]$$
$$\widehat{\boldsymbol{x}}_{i|i}^n = \mathbb{E}\left[\boldsymbol{x}_i \mid \widehat{\boldsymbol{y}}_{i|i}^n\right] \iff \widehat{\boldsymbol{x}}_{i|i}^n = \mathbb{E}\left[\boldsymbol{x}_i \mid \boldsymbol{\nu}_i^n\right]$$

By the Gauss-Markov principle,

$$\widehat{\boldsymbol{y}}_{i|i}^n = \widehat{\boldsymbol{y}}_{i|i-1}^n + \widehat{B}_i^n\widetilde{\boldsymbol{\nu}}_i^n$$
$$\widehat{\boldsymbol{x}}_{i|i}^n = \widehat{\boldsymbol{x}}_{i|i-1}^n + K_i^n\boldsymbol{\nu}_i^n$$

where $\widehat{B}_i^n$ are the non-zero blocks of the n^{th} row of B_i. Now we expand the term $\widehat{B}_i^n\widetilde{\boldsymbol{\nu}}_i^n$ by multiplying the gain blocks B_i^{nl} with the corresponding $(\widehat{\boldsymbol{y}}_{i|i-1}^l - \widehat{\boldsymbol{y}}_{i|i-1}^n)$ and the gain block B_i^{nn} with $\Big(\widetilde{\boldsymbol{z}}_i^n - \widetilde{H}_n\widehat{\boldsymbol{y}}_{i|i-1}^n - \check{H}_n\widehat{\boldsymbol{x}}_{i|i-1}^n\Big)$. This gives us the *consensus+innovations* filtering pseudo-state update (5.18).

The pseudo-state and state prediction updates are

$$\begin{aligned}
\widehat{\boldsymbol{y}}_{i+1|i}^n &= \mathbb{E}\left[\boldsymbol{y}_{i+1} \mid \widetilde{\boldsymbol{z}}_i^n, \{\widehat{\boldsymbol{y}}_{i|i-1}^l\}_{l\in\overline{\Omega}_n}\right] \\
&= \mathbb{E}\left[\widetilde{A}\boldsymbol{y}_i + G\boldsymbol{v}_i + \check{A}\boldsymbol{x}_i \mid \widetilde{\boldsymbol{z}}_i^n, \{\widehat{\boldsymbol{y}}_{i|i-1}^l\}_{l\in\overline{\Omega}_n}\right] \\
&= \widetilde{A}\widehat{\boldsymbol{y}}_{i|i}^n + \check{A}\widehat{\boldsymbol{x}}_{i|i}^n \\
\widehat{\boldsymbol{x}}_{i+1|i}^n &= \mathbb{E}\left[\boldsymbol{x}_{i+1} \mid \widehat{\boldsymbol{y}}_{i|i}^n\right] = \mathbb{E}\left[A\boldsymbol{x}_i + \boldsymbol{v}_i \mid \widehat{\boldsymbol{y}}_{i|i}^n\right] = A\widehat{\boldsymbol{x}}_{i|i}^n.
\end{aligned}$$

□

The update equations reflect the computation tasks of each agent n at each time index i. The gain matrices, B_i and K_i, in (5.18)-(5.21) are deterministic and can be pre-computed and saved at each agent. We discuss the details of the design of these optimal gain matrices in Section 5.5. With

these optimal gain matrices, the Kalman type *Consensus+Innovations* filter and prediction updates (5.18)-(5.21) provide the minimized MSE distributed estimates of the dynamic states, and hence we term our solution as *Consensus+Innovations* Kalman Filter (CIKF).

5.2.3 CIKF: Assumptions

The *Consensus+Innovations* Kalman Filter (CIKF) achieves convergence given the Assumptions 4-5 of global detectability and connected network hold true. By Assumption 4, the state-observation model (2.1)-(2.3) is globally detectable but not necessarily locally detectable, i.e., $(A, H_n), \forall n$, are not necessarily detectable. Note that these two are minimal assumptions. Assumption 4 is mandatory even for a centralized system, and Assumption 5 is required for consensus algorithms to converge. Further, note that in this chapter we do not consider distributed observability (invertibility of G) of the model setup, which is the strong and restrictive assumption taken in [55], [27], [28] and [60] and similar to weak detectability presented in [24].

5.2.4 CIKF: Update algorithm

In this subsection, we present the step-by-step tasks executed by each agent n in the cyber layer to implement the *Consensus+Innovations* Kalman Filter (CIKF) and thereby obtain the unbiased minimized MSE distributed estimates of the dynamic state $\boldsymbol{x}_i$. Each agent n runs Algorithm 4 locally.

Later in Section 5.6, we analyze and compare the performance of CIKF with that of the distributed information Kalman filter (DKF) [55] and of the centralized Kalman filter (CKF). The centralized filter collects measurements from all the agents in the cyber layer. Before going into the numerical evaluation, we do theoretical error analysis of CIKF, derive the conditions for convergence guarantees, and design the optimal consensus and innovations gains in the following sections.

Algorithm 4 *Consensus+Innovations* Kalman Filter

Input: Model parameters A, V, H, R, G, L, $\overline{\boldsymbol{x}}_0$, Σ_0.
Initialize: $\widehat{\boldsymbol{x}}^n_{0|-1} = \overline{\boldsymbol{x}}_0$, $\widehat{\boldsymbol{y}}^n_{0|-1} = G\overline{\boldsymbol{x}}_0$.
Pre-compute: Gain matrices B_i and K_i using Algorithm 5.
while $i \geq 0$ **do**
Communications:
Broadcast $\widehat{\boldsymbol{y}}^n_{i|i-1}$ to all neighbors $l \in \Omega_n$.
Receive $\{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\Omega_n}$ from neighbors.
Observation:
Make measurement $\boldsymbol{z}^n_i$ of the state $\boldsymbol{x}_i$.
Transform $\boldsymbol{z}^n_i$ in pseudo-observation $\widetilde{\boldsymbol{z}}^n_i$ using (5.2).
Filter updates:
Compute the estimate $\widehat{\boldsymbol{y}}^n_{i|i}$ of $\boldsymbol{y}_i$ using (5.18).
Compute the estimate $\widehat{\boldsymbol{x}}^n_{i|i}$ of the state $\boldsymbol{x}_i$ using (5.19).
Prediction updates:
Predict the estimate $\widehat{\boldsymbol{y}}^n_{i+1|i}$ of $\boldsymbol{y}_{i+1}$ using (5.20).
Predict the estimate $\widehat{\boldsymbol{x}}^n_{i+1|i}$ of the state $\boldsymbol{x}_{i+1}$ using (5.21).
end while

5.3 Error Analysis

We analyze the MSE performance of the CIKF and derive its error covariance matrices. First, we define the different error processes and determine their dynamics. Denote the filtering error processes $\boldsymbol{e}^n_{i|i}$ and $\boldsymbol{\epsilon}^n_{i|i}$ of the pseudo-state and of the state at agent n by

$$\boldsymbol{e}^n_{i|i} = \boldsymbol{y}_i - \widehat{\boldsymbol{y}}^n_{i|i}, \tag{5.22}$$

$$\boldsymbol{\epsilon}^n_{i|i} = \boldsymbol{x}_i - \widehat{\boldsymbol{x}}^n_{i|i}. \tag{5.23}$$

Similarly, represent the prediction error processes $\boldsymbol{e}^n_{i+1|i}$ and $\boldsymbol{\epsilon}^n_{i+1|i}$ of the pseudo-state and the state at agent n by

$$\boldsymbol{e}^n_{i+1|i} = \boldsymbol{y}_{i+1} - \widehat{\boldsymbol{y}}^n_{i+1|i}, \tag{5.24}$$

$$\boldsymbol{\epsilon}^n_{i+1|i} = \boldsymbol{x}_{i+1} - \widehat{\boldsymbol{x}}^n_{i+1|i}. \tag{5.25}$$

We establish that the CIKF provides unbiased estimates of the state and pseudo-state in the following lemma.

Lemma 5.3.1. *The distributed filter and prediction estimates, $\widehat{\boldsymbol{y}}^n_{i|i}$, $\widehat{\boldsymbol{x}}^n_{i|i}$, $\widehat{\boldsymbol{y}}^n_{i+1|i}$ and $\widehat{\boldsymbol{x}}^n_{i+1|i}$, of the pseudo-state and state are unbiased, i.e., error processes, $\boldsymbol{e}^n_{i|i}$, $\boldsymbol{e}^n_{i+1|i}$, $\boldsymbol{\epsilon}^n_{i|i}$, and $\boldsymbol{\epsilon}^n_{i+1|i}$ are zero-mean at all agents n:*

$$\mathbb{E}[\boldsymbol{e}^n_{i|i}] = 0,\ \mathbb{E}[\boldsymbol{e}^n_{i+1|i}] = 0,\ \mathbb{E}[\boldsymbol{\epsilon}^n_{i|i}] = 0,\ \mathbb{E}[\boldsymbol{\epsilon}^n_{i+1|i}] = 0. \tag{5.26}$$

Proof. Consider the filtering error definitions (5.22)-(5.23). We take expectations on both sides,

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{e}^n_{i|i}\right] &= \mathbb{E}\left[\boldsymbol{y}_i - \widehat{\boldsymbol{y}}^n_{i|i}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{y}_i - \widehat{\boldsymbol{y}}^n_{i|i} \mid \widetilde{\boldsymbol{z}}^n_i, \{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\Omega_n}\right]\right] \\
&= \mathbb{E}\left[\widehat{\boldsymbol{y}}^n_{i|i} - \widehat{\boldsymbol{y}}^n_{i|i}\right] = 0 \qquad \text{[by(5.14)]} \\
\mathbb{E}\left[\boldsymbol{\epsilon}^n_{i|i}\right] &= \mathbb{E}\left[\boldsymbol{x}_i - \widehat{\boldsymbol{x}}^n_{i|i}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{x}_i - \widehat{\boldsymbol{x}}^n_{i|i} \mid \widehat{\boldsymbol{y}}^n_{i|i}\right]\right] \\
&= \mathbb{E}\left[\widehat{\boldsymbol{x}}^n_{i|i} - \widehat{\boldsymbol{x}}^n_{i|i}\right] = 0 \qquad \text{[by(5.23)]}\,.
\end{aligned}$$

Similarly, taking expectations on prediction errors (5.24)-(5.15),

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{e}^n_{i+1|i}\right] &= \mathbb{E}\left[\boldsymbol{y}_{i+1} - \widehat{\boldsymbol{y}}^n_{i+1|i}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{y}_{i+1} - \widehat{\boldsymbol{y}}^n_{i+1|i} \mid \widetilde{\boldsymbol{z}}^n_i, \{\widehat{\boldsymbol{y}}^l_{i|i-1}\}_{l\in\Omega_n}\right]\right] \\
&= \mathbb{E}\left[\widehat{\boldsymbol{y}}^n_{i+1|i} - \widehat{\boldsymbol{y}}^n_{i+1|i}\right] = 0 \qquad \text{[by(5.16)]} \\
\mathbb{E}\left[\boldsymbol{\epsilon}^n_{i+1|i}\right] &= \mathbb{E}\left[\boldsymbol{x}_{i+1} - \widehat{\boldsymbol{x}}^n_{i+1|i}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{x}_{i+1} - \widehat{\boldsymbol{x}}^n_{i+1|i} \mid \widehat{\boldsymbol{y}}^n_{i|i}\right]\right] \\
&= \mathbb{E}\left[\widehat{\boldsymbol{x}}^n_{i+1|i} - \widehat{\boldsymbol{x}}^n_{i+1|i}\right] = 0 \qquad \text{[by(5.25)]}\,.
\end{aligned}$$

□

Each agent exchanges their estimates with their neighbors, hence their error processes are correlated. It is not feasible to analyze the error process of each agent separately as they depend on each other. To analyze all of them together, we stack the estimates and the errors of all the agents as we have done earlier for observations (2.3) and pseudo-observations (5.13):

$$\widehat{\boldsymbol{x}}_{i|i}=\begin{bmatrix}\widehat{\boldsymbol{x}}^1_{i|i}\\ \vdots \\ \widehat{\boldsymbol{x}}^N_{i|i}\end{bmatrix} \quad ,\widehat{\boldsymbol{x}}_{i+1|i}=\begin{bmatrix}\widehat{\boldsymbol{x}}^1_{i+1|i}\\ \vdots \\ \widehat{\boldsymbol{x}}^N_{i+1|i}\end{bmatrix} \quad ,\widehat{\boldsymbol{y}}_{i|i}=\begin{bmatrix}\widehat{\boldsymbol{y}}^1_{i|i}\\ \vdots \\ \widehat{\boldsymbol{y}}^N_{i|i}\end{bmatrix} \quad ,\widehat{\boldsymbol{y}}_{i+1|i}=\begin{bmatrix}\widehat{\boldsymbol{y}}^1_{i+1|i}\\ \vdots \\ \widehat{\boldsymbol{y}}^N_{i+1|i}\end{bmatrix} \quad ,$$

$$\boldsymbol{\epsilon}_{i|i}=\begin{bmatrix}\boldsymbol{\epsilon}^1_{i|i}\\ \vdots \\ \boldsymbol{\epsilon}^N_{i|i}\end{bmatrix} \quad ,\boldsymbol{\epsilon}_{i+1|i}=\begin{bmatrix}\boldsymbol{\epsilon}^1_{i+1|i}\\ \vdots \\ \boldsymbol{\epsilon}^N_{i+1|i}\end{bmatrix} \quad ,\boldsymbol{e}_{i|i}=\begin{bmatrix}\boldsymbol{e}^1_{i|i}\\ \vdots \\ \boldsymbol{e}^N_{i|i}\end{bmatrix} \quad ,\boldsymbol{e}_{i+1|i}=\begin{bmatrix}\boldsymbol{e}^1_{i+1|i}\\ \vdots \\ \boldsymbol{e}^N_{i+1|i}\end{bmatrix} \quad .$$

We summarize the dynamics of the error processes in the following lemma.

Lemma 5.3.2. *The error processes, $\boldsymbol{e}_{i|i}$, $\boldsymbol{e}_{i+1|i}$, $\boldsymbol{\epsilon}_{i|i}$, and $\boldsymbol{\epsilon}_{i|i}$ are Gaussian and their dynamics are:*

$$\boldsymbol{e}_{i|i}=\left(I_{MN}-B_i^{\mathcal{C}}-B_i^{\mathcal{I}}\widetilde{D}_H\right)\boldsymbol{e}_{i|i-1}-B_i^{\mathcal{I}}\check{D}_H\boldsymbol{\epsilon}_{i|i-1}-B_i^{\mathcal{I}}D_H^TR^{-1}r_i, \tag{5.27}$$

$$\boldsymbol{\epsilon}_{i|i}=\left(I_{MN}-K_i\left(I_N\otimes G\right)\right)\boldsymbol{\epsilon}_{i|i-1}+K_i\boldsymbol{e}_{i|i}, \tag{5.28}$$

$$\boldsymbol{e}_{i+1|i}=\left(I_N\otimes\widetilde{A}\right)\boldsymbol{e}_{i|i}+\left(I_N\otimes\check{A}\right)\boldsymbol{\epsilon}_{i|i}+1_N\otimes\left(G\boldsymbol{v}_i\right), \tag{5.29}$$

$$\boldsymbol{\epsilon}_{i+1|i}=\left(I_N\otimes A\right)\boldsymbol{\epsilon}_{i|i}+1_N\otimes\boldsymbol{v}_i, \tag{5.30}$$

where, $B_i^{\mathcal{C}}$ is the consensus gain matrix and $B_i^{\mathcal{I}}$, K_i are the innovations gain matrices for the pseudo-state and state estimation, respectively. The block diagonal matrices are $\widetilde{D}_H=$ blockdiag$\{\widetilde{H}_1,\cdots,\widetilde{H}_N\}$ and $\check{D}_H=$ blockdiag$\{\check{H}_1,\cdots,\check{H}_N\}$.

Proof. We write the pseudo-state filtering update (5.18) in vector form,

$$\widehat{\boldsymbol{y}}_{i|i}=\widehat{\boldsymbol{y}}_{i|i-1}-B_i^{\mathcal{C}}\widehat{\boldsymbol{y}}_{i|i-1}+B_i^{\mathcal{I}}\left(\widetilde{\boldsymbol{z}}_i-\left(\widetilde{D}_H\widehat{\boldsymbol{y}}_{i|i-1}+\check{D}_H\widehat{\boldsymbol{x}}_{i|i-1}\right)\right),$$

where, $\left[B_i^{\mathcal{C}}\right]_{nl}=-B_i^{n,l},\ \forall n\neq l,\ \left[B_i^{\mathcal{C}}\right]_{nn}=\sum_{l\in\Omega_n}B_i^{n,l},\ \left[B_i^{\mathcal{I}}\right]_{nn}=B_i^{n,n},\ \left[B_i^{\mathcal{I}}\right]_{nl}=0,\ \forall n\neq l.$ The block-diagonal matrices: $\widetilde{D}_H=$ blockdiag$\{\widetilde{H}_1,\cdots,\widetilde{H}_N\},\ \check{D}_H=$ blockdiag$\{\check{H}_1,\cdots,\check{H}_N\}$. Note that $B_i^{\mathcal{C}}\left(1_N\otimes\boldsymbol{y}_i\right)=\mathbf{0}$. Using this relation and the vector form of $\widehat{\boldsymbol{y}}_{i|i}$, we expand the pseudo-state filter error process $\boldsymbol{e}_{i|i}$,

$$\begin{aligned}\boldsymbol{e}_{i|i}&=1_N\otimes\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i|i}\\&=\left(1_N\otimes\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i|i-1}\right)-B_i^{\mathcal{C}}\left(1_N\otimes\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i|i-1}\right)-B_i^{\mathcal{I}}D_H^TR^{-1}\boldsymbol{r}_i\\&\quad-B_i^{\mathcal{I}}\widetilde{D}_H\left(1_N\otimes\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i|i-1}\right)-B_i^{\mathcal{I}}\check{D}_H\left(1_N\otimes\boldsymbol{x}_i-\widehat{\boldsymbol{x}}_{i|i-1}\right)\end{aligned}$$

$$= \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right)\boldsymbol{e}_{i|i-1} - B_i^{\mathcal{I}}\check{D}_H\boldsymbol{\epsilon}_{i|i-1} - B_i^{\mathcal{I}}D_H^T R^{-1} r_i.$$

The state filtering update (5.19), in vector form, is

$$\widehat{\boldsymbol{x}}_{i|i} = \widehat{\boldsymbol{x}}_{i|i-1} + K_i\left(\widehat{\boldsymbol{y}}_{i|i} - (I_N \otimes G)\,\widehat{\boldsymbol{x}}_{i|i-1}\right),$$

where, $K_i = \text{blockdiag}\{K_i^1, \cdots, K_i^N\}$. Using the relation $\widehat{\boldsymbol{y}}_{i|i} = (1_N \otimes \boldsymbol{y}_i) - \boldsymbol{e}_{i|i} = (I_N \otimes G)\,(1_N \otimes \boldsymbol{x}_i) - \boldsymbol{e}_{i|i}$, we expand the state filter error process $\boldsymbol{e}_{i|i}$,

$$\begin{aligned}
\boldsymbol{\epsilon}_{i|i} &= 1_N \otimes \boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|i} \\
&= \left(1_N \otimes \boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|i-1}\right) - K_i\,(I_N \otimes G)\left(1_N \otimes \boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|i-1}\right) + K_i\boldsymbol{e}_{i|i} \\
&= \left(I_{MN} - K_i^{\mathcal{C}} - K_i\,(I_N \otimes G)\right)\boldsymbol{\epsilon}_{i|i-1} + K_i\boldsymbol{e}_{i|i}.
\end{aligned}$$

The dynamics of the pseudo-state and state prediction errors,

$$\begin{aligned}
\boldsymbol{e}_{i+1|i} &= 1_N \otimes \boldsymbol{y}_{i+1} - \widehat{\boldsymbol{y}}_{i+1|i} \\
&= \left(I_N \otimes \widetilde{A}\right)(1_N \otimes \boldsymbol{y}_i) + \left(I_N \otimes \check{A}\right)(1_N \otimes \boldsymbol{x}_i) + 1_N \otimes (G\boldsymbol{v}_i) - \left(I_N \otimes \widetilde{A}\right)\widehat{\boldsymbol{y}}_{i|i} - \left(I_N \otimes \check{A}\right)\widehat{\boldsymbol{x}}_{i|i} \\
&= \left(I_N \otimes \widetilde{A}\right)\boldsymbol{e}_{i|i} + \left(I_N \otimes \check{A}\right)\boldsymbol{\epsilon}_{i|i} + 1_N \otimes (G\boldsymbol{v}_i) \\
\boldsymbol{\epsilon}_{i+1|i} &= 1_N \otimes \boldsymbol{x}_{i+1} - \widehat{\boldsymbol{x}}_{i+1|i} \\
&= (I_N \otimes A)\,(1_N \otimes \boldsymbol{x}_i) + 1_N \otimes \boldsymbol{v}_i - (I_N \otimes A)\,\widehat{\boldsymbol{x}}_{i|i} \\
&= (I_N \otimes A)\,\boldsymbol{\epsilon}_{i|i} + 1_N \otimes \boldsymbol{v}_i.
\end{aligned}$$

Since the state $\boldsymbol{x}_i$, pseudo-state $\boldsymbol{y}_i$, their initial condition and all the noises are Gaussian, their estimates are also Gaussian making all the filtering and prediction errors Gaussian. □

The symbol $\otimes$ denotes the Kronecker matrix product. Lemma 5.3.1 established that the error processes (5.27)-(5.30) are unbiased. It then follows that the filter and prediction error covariances of the pseudo-state and state are simply:

$$P_{i|i} = \mathbb{E}\left[\boldsymbol{e}_{i|i}\boldsymbol{e}_{i|i}^T\right] \tag{5.31}$$

$$P_{i+1|i} = \mathbb{E}\left[\boldsymbol{e}_{i+1|i}\boldsymbol{e}_{i+1|i}^T\right] \tag{5.32}$$

$$\Sigma_{i|i} = \mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}\boldsymbol{\epsilon}_{i|i}^T\right] \tag{5.33}$$

$$\Sigma_{i+1|i} = \mathbb{E}\left[\boldsymbol{\epsilon}_{i+1|i}\boldsymbol{\epsilon}_{i+1|i}^T\right] \tag{5.34}$$

Note that the state estimates $\widehat{\boldsymbol{x}}_{i|i}$, $\widehat{\boldsymbol{x}}_{i+1|i}$ depend on the pseudo-state estimates $\widehat{\boldsymbol{y}}_{i|i}$, $\widehat{\boldsymbol{y}}_{i+1|i}$. Hence the error process (5.27)-(5.30) are not uncorrelated. The filter and prediction cross-covariances are:

$$\Pi_{i|i} = \mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}\boldsymbol{e}_{i|i}^T\right] \tag{5.35}$$

$$\Pi_{i+1|i} = \mathbb{E}\left[\boldsymbol{\epsilon}_{i+1|i}\boldsymbol{e}_{i+1|i}^T\right] \tag{5.36}$$

$$\Gamma_i = \mathbb{E}\left[\boldsymbol{\epsilon}_{i|i-1}\boldsymbol{e}_{i|i}^T\right] \tag{5.37}$$

In the following theorem, we define and derive the evolution of the state, pseudo-state, and cross error covariances.

Theorem 5.3.1. *The filter error covariances, $P_{i|i}$, $\Sigma_{i|i}$, $\Pi_{i|i}$, and the predictor error covariances, $P_{i+1|i}$, $\Sigma_{i+1|i}$, $\Pi_{i+1|i}$, follow Lyapunov-type iterations:*

$$\Gamma_i = \Pi_{i|i-1}\left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right)^T - \Sigma_{i|i-1}\check{D}_H^T B_i^{\mathcal{I}^T} \tag{5.38}$$

$$\begin{aligned} P_{i|i} = {} & \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right) P_{i|i-1}\left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right)^T + B_i^{\mathcal{I}}\overline{D}_H B_i^{\mathcal{I}^T} \\ & - \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right)\Pi_{i|i-1}^T\check{D}_H B_i^{\mathcal{I}^T} - B_i^{\mathcal{I}}\check{D}_H\Gamma_i \end{aligned} \tag{5.39}$$

$$\begin{aligned} \Sigma_{i|i} = {} & \left(I_{MN} - K_i(I_N\otimes G)\right)\Sigma_{i|i-1}\left(I_{MN} - K_i\left(I_N\otimes G\right)\right) + K_i P_{i|i} K_i^T + \left(I_{MN} - K_i(I_N\otimes G)\right)\Pi_{i|i-1}K_i^T \\ & + K_i\Pi_{i|i-1}^T\left(I_{MN} - K_i(I_N\otimes G)\right)^T \end{aligned} \tag{5.40}$$

$$\Pi_{i|i} = \left(I_{MN} - K_i(I_N\otimes G)\right)\Gamma_i + K_i P_{i|i}^T \tag{5.41}$$

$$\begin{aligned} P_{i+1|i} = {} & \left(I_N\otimes\widetilde{A}\right) P_{i|i}\left(I_N\otimes\widetilde{A}^T\right) + \left(I_N\otimes\check{A}\right)\Sigma_{i|i}\left(I_N\otimes\check{A}^T\right) + \left(I_N\otimes\check{A}\right)\Pi_{i|i}\left(I_N\otimes\widetilde{A}^T\right) \\ & + J\otimes(GVG) + \left(I_N\otimes\widetilde{A}\right)\Pi_{i|i}^T\left(I_N\otimes\check{A}^T\right) \end{aligned} \tag{5.42}$$

$$\Sigma_{i+1|i} = (I_N\otimes A)\,\Sigma_{i|i}\left(I_N\otimes A^T\right) + J\otimes V \tag{5.43}$$

$$\Pi_{i+1|i} = (I_N\otimes A)\,\Pi_{i|i}\left(I_N\otimes\widetilde{A}^T\right) + (I_N\otimes A)\,\Sigma_{i|i}\left(I_N\otimes\check{A}^T\right) + J\otimes(VG) \tag{5.44}$$

where, $J = (\mathbf{1}_N\mathbf{1}_N^T)\otimes I_M$ and the initial conditions are $\Sigma_{0|-1} = J\otimes\Sigma_0$, $P_{0|-1} = J\otimes(G\Sigma_0 G)$, $\Pi_{0|-1} = J\otimes(\Sigma_0 G)$.

Proof. By Lemma 5.3.1 and Lemma 5.3.2, the error processes, $\boldsymbol{e}_{i|i}$, $\boldsymbol{e}_{i+1|i}$, $\boldsymbol{\epsilon}_{i|i}$, and $\boldsymbol{\epsilon}_{i+1|i}$ are zero-mean Gaussian. The Lyapunov-type iterations (5.39)-(5.44) of the filter and predictor error covariances, $P_{i|i}$, $\Sigma_{i|i}$, $\Pi_{i|i}$, $P_{i+1|i}$, $\Sigma_{i+1|i}$, and $\Pi_{i+1|i}$, follow directly from the definitions (5.31)-(5.36) and error dynamics (5.27)-(5.30) by algebraic manipulations. □

The iterations (5.40) and (5.43) combined together constitute the distributed version of the discrete algebraic Riccati equation. The MSE of the proposed CIKF is the trace of the error covariance, $\Sigma_{i+1|i}$ in (5.43). The optimal design of the gain matrices, B_i and K_i, such that the CIKF yields minimized MSE estimates, is discussed in Section 5.5. Before that in Section 5.4 we derive the conditions under which the CIKF converges, in other words, the MSE given by the trace of $\Sigma_{i+1|i}$ is bounded.

5.4 Tracking Capacity

The convergence properties of the CIKF is determined by the dynamics of the pseudo-state and state error processes, $\boldsymbol{e}_{i+1|i}$ and $\boldsymbol{\epsilon}_{i+1|i}$. If the error dynamics are asymptotically stable, then the error processes have asymptotically bounded error covariances that in turn guarantee the convergence of the CIKF. Note that if the dynamics of the prediction error processes, $\boldsymbol{e}_{i+1|i}$, $\boldsymbol{\epsilon}_{i+1|i}$ are asymptotically stable, then the dynamics of the filter error processes, $\boldsymbol{e}_{i|i}$, $\boldsymbol{\epsilon}_{i|i}$ are also asymptotically stable. That is why we study the dynamics of only one of the error processes and in this chapter we consider the prediction error processes.

5.4.1 Asymptotic stability of error processes

To analyze the stability of the error processes, we first write the evolution of the prediction error processes, combining (5.27)-(5.30),

$$\boldsymbol{e}_{i+1|i} = \underbrace{\left(I_N \otimes \widetilde{A}\right)\left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right)}_{\widetilde{F}} \boldsymbol{e}_{i|i-1} + \widetilde{\boldsymbol{\phi}}_i, \tag{5.45}$$

$$\boldsymbol{\epsilon}_{i+1|i} = \underbrace{(I_N \otimes A)\left(I_{MN} - K_i\left(I_N \otimes G\right)\right)}_{F} \boldsymbol{\epsilon}_{i|i-1} + \boldsymbol{\phi}_i, \tag{5.46}$$

where, the noise processes $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$ are

$$\widetilde{\boldsymbol{\phi}}_i = \left(I_N \otimes \check{A}\right)\boldsymbol{\epsilon}_{i|i} - \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}}\check{D}_H\boldsymbol{\epsilon}_{i|i-1} + 1_N \otimes (G\boldsymbol{v}_i) - \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}}D_H^T R^{-1} r_i, \tag{5.47}$$

$$\boldsymbol{\phi}_i = (I_N \otimes A)K_i\boldsymbol{e}_{i|i} + 1_N \otimes \boldsymbol{v}_i. \tag{5.48}$$

The statistical properties of the noises, $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$, of the error processes, $\boldsymbol{e}_{i+1|i}$ and $\boldsymbol{\epsilon}_{i+1|i}$, are stated in the following Lemma.

Lemma 5.4.1. *The noise sequences $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$ are zero-mean Gaussian that follow $\widetilde{\boldsymbol{\phi}}_i \in \mathcal{N}\left(\mathbf{0}, \widetilde{\Phi}_i\right)$ and $\boldsymbol{\phi}_i \in \mathcal{N}\left(\mathbf{0}, \Phi_i\right)$.*

Proof. Lemma 5.3.2 and Assumption 1 guarantee that $\boldsymbol{\epsilon}_{i|i}, \boldsymbol{\epsilon}_{i+1|i}, \boldsymbol{e}_{i|i}, \boldsymbol{v}_i, \boldsymbol{r}_i$ are Gaussian. The error noises $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$ are therefore Gaussian as they are linear combinations of the error processes and the model noises $\boldsymbol{\epsilon}_{i|i}, \boldsymbol{\epsilon}_{i+1|i}, \boldsymbol{e}_{i|i}, \boldsymbol{v}_i, \boldsymbol{r}_i$. By Lemma 5.3.1, we have $\mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}\right] = \mathbb{E}\left[\boldsymbol{\epsilon}_{i+1|i}\right] = \mathbb{E}\left[\boldsymbol{e}_{i|i}\right] = \mathbf{0}$. From Assumption 1, we know $\mathbb{E}\left[\boldsymbol{v}_i\right] = \mathbb{E}\left[\boldsymbol{r}_i\right] = \mathbf{0}$. We take expectation on both sides of (5.47)-(5.48) and apply these relations

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\boldsymbol{\phi}}_i\right] &= \left(I_N \otimes \check{A}\right)\mathbb{E}\left[\boldsymbol{\epsilon}_{i|i}\right] - \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}} D_H^T R^{-1}\mathbb{E}\left[r_i\right] \\
&\quad + 1_N \otimes \left(G\mathbb{E}\left[\boldsymbol{v}_i\right]\right) - \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}}\check{D}_H\mathbb{E}\left[\boldsymbol{\epsilon}_{i|i-1}\right] = \mathbf{0}, \\
\mathbb{E}\left[\boldsymbol{\phi}_i\right] &= (I_N \otimes A)K_i\mathbb{E}\left[\boldsymbol{e}_{i|i}\right] + 1_N \otimes \mathbb{E}\left[\boldsymbol{v}_i\right] = \mathbf{0}.
\end{aligned}$$

Combining (5.28), (5.29) and (5.47), we have

$$\widetilde{\boldsymbol{\phi}}_i = F_1\boldsymbol{\epsilon}_{i|i-1} + F_2\boldsymbol{e}_{i|i-1} - F_3 D_H^T R^{-1} r_i + 1_N \otimes (G\boldsymbol{v}_i)$$

where,

$$F_1 = \left(I_N \otimes \check{A}\right)\left(I_{MN} - K_i(I_N \otimes G) - K_i B_i^{\mathcal{I}}\check{D}_H\right) - \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}}\check{D}_H \tag{5.49}$$

$$F_2 = \left(I_N \otimes \check{A}\right)K_i\left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}}\widetilde{D}_H\right) \tag{5.50}$$

$$F_3 = \left(I_N \otimes \widetilde{A}\right)B_i^{\mathcal{I}}. \tag{5.51}$$

Since $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$ are zero-mean, the noise covariances are,

$$\widetilde{\Phi}_i = \mathbb{E}\left[\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_i^T\right] = F_1\Sigma_{i|i-1}F_1^T + F_2P_{i|i-1}F_2^T + F_3\overline{D}_H F_3^T + J \otimes (GVG) + F_1\Pi_{i|i-1}F_2^T + F_2\Pi_{i|i-1}^T F_1^T, \tag{5.52}$$

$$\Phi_i = \mathbb{E}\left[\boldsymbol{\phi}_i\boldsymbol{\phi}_i^T\right] = (I_N \otimes A)K_i\Sigma_{i|i}K_i^T\left(I_N \otimes A^T\right) + J \otimes V, \tag{5.53}$$

where, F_1, F_2 and F_3 are defined in (5.49)-(5.51). □

The dynamics of the error processes are characterized by (5.45)-(5.46) and Lemma 5.4.1. Let $\rho(.)$ and $\|.\|_2$ denote the spectral radius and the spectral norm of a matrix, respectively. The error

processes are asymptotically stable if and only if the spectral radii of $\widetilde{F}$, F are less than one, i.e.,

$$\rho\left(\widetilde{F}\right) < 1, \quad \rho\left(F\right) < 1 \tag{5.54}$$

and the noise covariances, $\widetilde{\Phi}_i$, Φ_i are bounded, i.e., $\|\widetilde{\Phi}_i\|_2 < \infty$, $\|\Phi_i\|_2 < \infty$, $\forall i$. Now if (5.54) holds, then the prediction error covariances $P_{i+1|i}$, $\Sigma_{i+1|i}$ are bounded; this ensures the filter error covariances $P_{i|i}$, $\Sigma_{i|i}$ are also bounded. Further, the model noise covariances V and R are bounded. Then, by (**??**)-(5.53), the noise covariances $\widetilde{\Phi}_i$ and Φ_i are bounded if the spectral radii are less than one. Thus, (5.54) are the necessary and sufficient conditions for the convergence of the CIKF algorithm.

5.4.2 Tracking capacity for unstable systems

The stability of the underlying dynamical system (2.1) in the physical layer is determined by the dynamics matrix A. If the system is asymptotically stable, i.e., $\rho(A) < 1$, then there always exist gain matrices B_i, K_i such that (5.54) holds true. Hence for stable systems, the CIKF always converges with a bounded MSE solution. In contrast, for an unstable dynamical system (2.1), $\rho(A) > 1$, it may not always be possible to find gain matrices B_i, K_i satisfying (5.54) conditions. There exists an upper threshold on the degree of instability of the system dynamics, A, that guarantees the convergence of the proposed CIKF. The threshold, similar to Network Tracking Capacity in [30], is the tracking capacity of the CIKF algorithm, and it depends on the agent communication network and observation models, as summarized in the following theorem.

Theorem 5.4.1. *The tracking capacity of the CIKF is, C,*

$$C = \max_{B_i^{\mathcal{C}}, B_i^{\mathcal{I}}} \frac{\lambda_1}{\lambda_m \left\| I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H \right\|_2} \tag{5.55}$$

where, $B_i^{\mathcal{C}}$ has the same block sparsity pattern as the graph Laplacian L, $B_i^{\mathcal{I}}$ is a block diagonal matrix, and, λ_1 and λ_m are the minimum and maximum non-zero eigenvalues of G, $0 < \lambda_1 \leq \cdots \leq \lambda_m$. If $\|A\|_2 < C$, then there exists B_i, K_i such that the CIKF (5.18)-(5.21) *converges with bounded MSE.*

Proof. For any square matrix, $\rho(\widetilde{F}) \leq \|\widetilde{F}\|$. Hence if $\|\widetilde{F}\|< 1$, then it implies that $\rho(\widetilde{F}) < 1$. We derive the tracking capacity with the sufficient condition, $\|\widetilde{F}\|< 1$,

$$\begin{aligned}
\|\widetilde{F}\| &= \left\| \left(I_N \otimes \widetilde{A}\right) \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\| \\
&\leq \left\| I_N \otimes \widetilde{A} \right\| \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\| \\
&\leq \left\| GAG^{\dagger} \right\| \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\| \\
&\leq \frac{\lambda_m}{\lambda_1} \left\| A \right\| \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\|,
\end{aligned}$$

where, $\|G\|_2 = \lambda_m$ and $\|G^{\dagger}\|_2 = \frac{1}{\lambda_1}$. Since G is a symmetric positive semi-definite matrix, its spectral norm is its largest eigenvalue (λ_m) and the spectral norm of its pseudo-inverse, $G^{\dagger}$, is the inverse of its smallest non-zero eigenvalue (λ_1). If there exists $B_i^{\mathcal{C}}$ and $B_i^{\mathcal{I}}$ such that

$$\frac{\lambda_m}{\lambda_1} \left\| A \right\| \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\| < 1,$$

then $\|\widetilde{F}\|< 1$ and also $\rho(\widetilde{F}) < 1$. The bound on the spectral norm of A is,

$$\begin{aligned}
\|A\| &< \frac{\lambda_1}{\lambda_m \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\|} \\
&\leq \max_{B_i^{\mathcal{C}}, B_i^{\mathcal{I}}} \frac{\lambda_1}{\lambda_m \left\| \left(I_{MN} - B_i^{\mathcal{C}} - B_i^{\mathcal{I}} \widetilde{D}_H\right) \right\|} = C.
\end{aligned}$$

Thus as long as $\|A\|_2< C$, there exists $B_i^{\mathcal{C}}$ and $B_i^{\mathcal{I}}$ such that $\rho(\widetilde{F}) < 1$. By global detectability Assumption 4, there exists K_i such that $\rho(F) < 1$. Refer to [47], for the convergence conditions of the centralized information filters. Further, by Lemma 5.4.1 the Gaussian noises processes $\widetilde{\boldsymbol{\phi}}_i$ and $\boldsymbol{\phi}_i$ have bounded noise covariances. Thus, if $\|A\|_2< C$, then from (5.45)-(5.46) we conclude that the CIKF (5.18)-(5.21) converges with bounded MSE. □

In the above theorem, the structural constraints on the gain matrices $B_i^{\mathcal{C}}, B_i^{\mathcal{I}}$ ensure that each agent combines its neighbors' estimates for consensus and its own pseudo-observations for the innovation part of the CIKF. The block sparsity pattern of $B_i^{\mathcal{C}}$ being similar to that of L implies that the tracking capacity is dependent on the connectivity of the communication network. Similarly, since $\widetilde{D}_H$ is a block-diagonal matrix containing the observation matrices H_n we conclude that the tracking capacity is also a function of the observation models.

The tracking capacity increases with the increase in communication graph connectivity and observation density. For instance, the tracking capacity is infinity if all agents are connected with everyone else (complete graph) or all the agents observe the entire dynamical system (local observability). Given the tracking capacity is satisfied for the system, observation, and communication models (2.1)-(2.4), the question remains how to design the gain matrices B_i and K_i to minimize the MSE of the CIKF, which we discuss in the following section.

5.5 Optimal Gain Design

The asymptotic stability of the error dynamics guarantees convergence of CIKF and bounded MSE, but here we discuss how to design the B_i and K_i such that the MSE is not only bounded but also minimum.

5.5.1 New uncorrelated information

In CIKF Algorithm 4, at any time i each agent n makes pseudo-observation $\widetilde{\boldsymbol{z}}_i^n$ of the state and receives prior estimates $\{\widehat{\boldsymbol{y}}_{i|i-1}^l\}_{l\in\Omega_n}$ from its neighbors. The CIKF algorithm employs this new information to compute the distributed filter estimates of the pseudo-state and state. Denote the new information for the pseudo-state and state filtering by $\widetilde{\boldsymbol{\theta}}_i^n$ and $\boldsymbol{\theta}_i^n$, respectively,

$$\widetilde{\boldsymbol{\theta}}_i^n = \begin{bmatrix} \widehat{\boldsymbol{y}}_{i|i-1}^{l_1} \\ \vdots \\ \widehat{\boldsymbol{y}}_{i|i-1}^{l_{d_n}} \\ \widetilde{\boldsymbol{z}}_i^n \end{bmatrix}, \quad \boldsymbol{\theta}_i^n = \widehat{\boldsymbol{y}}_{i|i}^n \tag{5.56}$$

where, $\{l_1, \cdots, l_{d_n}\} = \Omega_n$ and $d_n = |\Omega_n|$ is the degree of agent n. Note the new information, $\widetilde{\boldsymbol{\theta}}_i^n$ and $\boldsymbol{\theta}_i^n$, are Gaussian since they are linear combinations of Gaussian sequences. However, $\widetilde{\boldsymbol{\theta}}_i^n$ and $\boldsymbol{\theta}_i^n$ are correlated with the previous estimates $\widehat{\boldsymbol{y}}_{i|i-1}^n$ and $\widehat{\boldsymbol{x}}_{i|i-1}^n$. So, we transform them into uncorrelated new information and then combine the uncorrelated information with the previous estimates $\widehat{\boldsymbol{y}}_{i|i-1}^n$ and $\widehat{\boldsymbol{x}}_{i|i-1}^n$ to compute the current filtered estimates.

Lemma 5.5.1. *The new uncorrelated information $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{\nu}_i^n$ for filtering update at agent n are,*

$$\widetilde{\boldsymbol{\nu}}_i^n = \widetilde{\boldsymbol{\theta}}_i^n - \overline{\widetilde{\boldsymbol{\theta}}}_i^n, \qquad \overline{\widetilde{\boldsymbol{\theta}}}_i^n = \mathbb{E}\left[\widetilde{\boldsymbol{\theta}}_i^n | \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \tag{5.57}$$

$$\boldsymbol{\nu}_i^n = \boldsymbol{\theta}_i^n - \overline{\boldsymbol{\theta}}_i^n, \qquad \overline{\boldsymbol{\theta}}_i^n = \mathbb{E}\left[\boldsymbol{\theta}_i^n | \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] \tag{5.58}$$

that expands to

$$\widetilde{\boldsymbol{\nu}}_i^n = \begin{bmatrix} \widehat{\boldsymbol{y}}_{i|i-1}^{l_1} - \widehat{\boldsymbol{y}}_{i|i-1}^n \\ \vdots \\ \widehat{\boldsymbol{y}}_{i|i-1}^{l_{d_n}} - \widehat{\boldsymbol{y}}_{i|i-1}^n \\ \widetilde{\boldsymbol{z}}_i^n - \widetilde{H}_n \widehat{\boldsymbol{y}}_{i|i-1}^n - \check{H}_n \widetilde{I} \widehat{\boldsymbol{x}}_{i|i-1}^n \end{bmatrix}, \quad \boldsymbol{\nu}_i^n = \widehat{\boldsymbol{y}}_{i|i}^n - G\widehat{\boldsymbol{x}}_{i|i-1}^n. \tag{5.59}$$

The uncorrelated sequences $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{\nu}_i^n$ are zero-mean Gaussian random vectors. Hence $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{\nu}_i^n$ are independent sequences.

Proof. We first compute the conditional means $\overline{\widetilde{\boldsymbol{\theta}}}_i^n$ and $\overline{\boldsymbol{\theta}}_i^n$ of the new information $\widetilde{\boldsymbol{\theta}}_i^n$ and $\boldsymbol{\theta}_i^n$ from (5.56). The means $\overline{\widetilde{\boldsymbol{\theta}}}_i^n$ and $\overline{\boldsymbol{\theta}}_i^n$ depend on the conditional means of $\widehat{\boldsymbol{y}}_{i|i-1}^l$, $\widetilde{\boldsymbol{z}}_{i-1}^n$ and $\widehat{\boldsymbol{y}}_{i|i}^n$.

$$\begin{aligned}
&\mathbb{E}\left[\widehat{\boldsymbol{y}}_{i|i-1}^l | \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{y}_i | \widetilde{\boldsymbol{z}}_{i-1}^l, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^k\}_{k\in\Omega_l}\right] \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{y}_i | \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \mid \widetilde{\boldsymbol{z}}_{i-1}^l, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^k\}_{k\in\Omega_l}\right] \\
&\quad = \mathbb{E}\left[\widehat{\boldsymbol{y}}_{i|i-1}^n \mid \widetilde{\boldsymbol{z}}_{i-1}^l, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^k\}_{k\in\Omega_l}\right] = \widehat{\boldsymbol{y}}_{i|i-1}^n, \quad \forall\, l \in \Omega_n. \\
&\mathbb{E}\left[\widetilde{\boldsymbol{z}}_i^n \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \mathbb{E}\left[\widetilde{H}_n \boldsymbol{y}_i + H_n^T R_n^{-1} \boldsymbol{r}_i^n + \check{H}_n \boldsymbol{x}_i \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \widetilde{H}_n \mathbb{E}\left[\boldsymbol{y}_i \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] + \check{H}_n \mathbb{E}\left[\boldsymbol{x}_i \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \widetilde{H}_n \widehat{\boldsymbol{y}}_{i|i-1}^n + \check{H}_n \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{x}_i | \widehat{\boldsymbol{y}}_{i|i}^n, \widehat{\boldsymbol{x}}_{i-1|i-2}^l, \widetilde{\boldsymbol{z}}_{i-1}^n, \widehat{\boldsymbol{y}}_{i-1|i-2}^l, \in \Omega_n\right]\right] \\
&\quad = \widetilde{H}_n \widehat{\boldsymbol{y}}_{i|i-1}^n + \check{H}_n \mathbb{E}\left[\widehat{\boldsymbol{x}}_{i|i-1}^n \mid \widetilde{\boldsymbol{z}}_{i-1}^n, \{\widehat{\boldsymbol{y}}_{i-1|i-2}^l\}_{l\in\Omega_n}\right] \\
&\quad = \widetilde{H}_n \widehat{\boldsymbol{y}}_{i|i-1}^n + \check{H}_n \widehat{\boldsymbol{x}}_{i|i-1}^n. \\
&\mathbb{E}\left[\widehat{\boldsymbol{y}}_{i|i}^n | \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] = \mathbb{E}\left[\boldsymbol{y}_i - \boldsymbol{e}_{i|i}^n \mid \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] \\
&\quad = \mathbb{E}\left[\boldsymbol{y}_i \mid \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] - \mathbb{E}\left[\boldsymbol{e}_{i|i}^n \mid \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] \\
&\quad = G\mathbb{E}\left[\boldsymbol{x}_i \mid \widehat{\boldsymbol{y}}_{i-1|i-1}^n\right] = G\widehat{\boldsymbol{x}}_{i|i-1}^n.
\end{aligned}$$

The conditional means of $\widehat{\boldsymbol{y}}^l_{i|i-1}, \widetilde{\boldsymbol{z}}^n_{i-1}$ and $\widehat{\boldsymbol{y}}^n_{i|i}$ shows that the new uncorrelated information defined in (5.57)-(5.58) expands to the vectors $\widetilde{\boldsymbol{\nu}}^n_i$ and $\boldsymbol{\nu}^n_i$ in (5.59). Note that $\widetilde{\boldsymbol{\nu}}^n_i$ and $\boldsymbol{\nu}^n_i$ are zero mean, by definition, and, are Gaussian since they are linear combination of Gaussian vectors. Now to prove $\widetilde{\boldsymbol{\nu}}^n_i$ and $\boldsymbol{\nu}^n_i$ are sequence of uncorrelated vectors, we have to show:

$$\mathbb{E}\left[\widetilde{\boldsymbol{\nu}}^n_i \widetilde{\boldsymbol{\nu}}^{n^T}_j\right] = \mathbb{E}\left[\boldsymbol{\nu}^n_i \boldsymbol{\nu}^{n^T}_j\right] = \mathbf{0},\ \forall\, i \neq j,\ \forall\, n.$$

First, we write $\widetilde{\boldsymbol{\nu}}^n_i$ and $\boldsymbol{\nu}^n_i$ in terms of the filtering and prediction error processes using (5.7), (5.22)-(5.25),

$$\widetilde{\boldsymbol{\nu}}^n_i = \begin{bmatrix} \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_1}_{i|i-1} \\ \vdots \\ \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_{d_n}}_{i|i-1} \\ \widetilde{H}_n \boldsymbol{e}^n_{i|i-1} + \check{H}_n \boldsymbol{\epsilon}^n_{i|i-1} + H^T_n R^{-1}_n \boldsymbol{r}^n_i \end{bmatrix}, \quad \boldsymbol{\nu}^n_i = G\boldsymbol{\epsilon}^n_{i|i-1} - \boldsymbol{e}^n_{i|i}.$$

Without loss of generality, we consider $i > j$. The rest of the proof is similar to the proof of Lemma 5 in [55]. Here, the only difference is that we should condition on $\{\widetilde{\boldsymbol{z}}^l_{i-1}\}_{l\in\Omega_n}, \{\widehat{\boldsymbol{y}}^l_{i-1|i-1}\}_{l\in\Omega_n}$. □

We write the CIKF filter updates (5.18) and (5.19) in terms of the new uncorrelated information $\widetilde{\boldsymbol{\nu}}^n_i$ and $\boldsymbol{\nu}^n_i$ from (5.59),

$$\widehat{\boldsymbol{y}}^n_{i|i} = \widehat{\boldsymbol{y}}^n_{i|i-1} + \widehat{B}^n_i \widetilde{\boldsymbol{\nu}}^n_i \tag{5.60}$$

$$\widehat{\boldsymbol{x}}^n_{i|i} = \widehat{\boldsymbol{x}}^n_{i|i-1} + K^n_i \boldsymbol{\nu}^n_i \tag{5.61}$$

where, $\widehat{B}^n_i$ are the building blocks of the pseudo-state gain matrix B_i.

5.5.2 Consensus and innovation gains

Here, we present the methods to: (a) design the matrices $\widehat{B}^n_i$ and K^n_i; and (b) obtain the optimal gains B_i and K_i from them. These optimal gains provide the distributed minimized MSE estimates of the field. At agent n, we define the matrix $\widehat{B}^n_i$ as

$$\widehat{B}^n_i = \left[B^{nl_1}_i, \cdots, B^{nl_{d_n}}_i, B^{nn}_i\right] \tag{5.62}$$

where, $\{l_1, \cdots, l_{d_n}\} = \Omega_n$. The gain matrix B_i is a linear combination of $B^{\mathcal{C}}_i$ and $B^{\mathcal{I}}_i$, where $B^{\mathcal{C}}_i$ has the same block structure as the graph Laplacian L and $B^{\mathcal{I}}_i$ is a block diagonal. The $(n,l)^{\text{th}}$ blocks of

the n^{th} row of $B_i^{\mathcal{C}}$ are:

$$\left[B_i^{\mathcal{C}}\right]_{nl} = \begin{cases} -B_i^{nl}, & \text{if } l \in \Omega_n \\ \sum_{j=1}^{d_n} B_i^{nl_j}, & \text{if } l = n \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{5.63}$$

The $\{n,n\}^{\text{th}}$ blocks of the diagonal block matrices $B_i^{\mathcal{I}}$ and K_i are:

$$\left[B_i^{\mathcal{I}}\right]_{nn} = B_i^{nn}, \tag{5.64}$$

$$[K_i]_{nn} = K_i^n, \tag{5.65}$$

Hence, once we design the matrices $\widehat{B}_i^n$ and K_i^n, it will provide the optimal gain matrices $B_i^{\mathcal{C}}$, $B_i^{\mathcal{I}}$, and K_i.

Theorem 5.5.1. *The optimal gains for the CIKF algorithm are*

$$\widehat{B}_i^n = \Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n}\left(\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}\right)^{-1},$$
$$K_i^n = \Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n}\left(\Sigma_{\boldsymbol{\nu}_i^n}\right)^{-1}$$

where, $\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}$, $\Sigma_{\boldsymbol{\nu}_i^n}$ are the covariances of the new uncorrelated information $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{\nu}_i^n$; and, $\Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n}$, $\Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n}$ are cross-covariances between $\boldsymbol{y}_i$, $\widetilde{\boldsymbol{\nu}}_i^n$ and $\boldsymbol{x}_i$, $\boldsymbol{\nu}_i^n$, respectively. These covariance and cross-covariance matrices are related to the error covariance matrices, $P_{i|i-1}, P_{i|i}, \Sigma_{i|i-1}, \Pi_{i|i-1}, \Gamma_i$, by the following functions:

$$\Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n} = \left[P_{i|i-1}^{nn} - P_{i|i-1}^{nl_1} \;\vdots\; \cdots \;\vdots\; P_{i|i-1}^{nn} - P_{i|i-1}^{nl_{d_n}} \;\vdots\; P_{i|i-1}^{nn}\widetilde{H}_n^T + \Pi_{i|i-1}^{nn^T}\check{H}_n^T\right]$$

$$\Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n} = \Sigma_{i|i-1}^{nn}G - \Gamma_i^{nn}$$

$$\left[\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}\right]_{qs} = \begin{cases} P_{i|i-1}^{nn} - P_{i|i-1}^{nl_s} - P_{i|i-1}^{l_q n} + P_{i|i-1}^{l_q l_s}, & \text{if } q \le s \le d_n \\ \left(P_{i|i-1}^{nn} - P_{i|i-1}^{l_q n}\right)\widetilde{H}_n^T + \left(\Pi_{i|i-1}^{nn} - \Pi_{i|i-1}^{l_q n}\right)^T\check{H}_n^T, & \text{if } q < s = d_n{+}1 \\ \widetilde{H}_n P_{i|i-1}^{nn}\widetilde{H}_n^T + \widetilde{H}_n\Pi_{i|i-1}^{nn^T}\check{H}_n^T + \check{H}_n\Pi_{i|i-1}^{nn}\widetilde{H}_n^T & \\ \qquad + \check{H}_n\Sigma_{i|i-1}^{nn}\check{H}_n^T + \overline{H}_n, & \text{if } q = s = d_n{+}1 \\ \left[\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}\right]_{sq}^T, & \text{if } q > s \end{cases}$$

$$\left[\Sigma_{\boldsymbol{\nu}_i^n}\right]_{qs} = G\Sigma_{i|i-1}^{nn}G - G\Gamma_i^{nn} - \Gamma_i^{nn^T}G + P_{i|i}^{nn}$$

where, $\left[\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}\right]_{qs}$ denotes the $\{q,s\}^{th}$ block of the $(d_n{+}1) \times (d_n{+}1)$ block matrix $\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}$.

Proof. By the Innovations Property and the Gauss-Markov principle [47], the optimal gains $\widehat{B}_i^n$ and K_i^n in (5.60)-(5.61) are:

$$\widehat{B}_i^n = \Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n}\left(\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}\right)^{-1},$$
$$K_i^n = \Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n}\left(\Sigma_{\boldsymbol{\nu}_i^n}\right)^{-1}$$

that yield minimized MSE estimates $\widehat{\boldsymbol{y}}_{i|i}^n$ and $\widehat{\boldsymbol{x}}_{i|i}^n$ of the pseudo-state $\boldsymbol{y}_i$ and of the field $\boldsymbol{x}_i$, respectively, at each agent n. The cross-covariances $\Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n}$ and $\Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n}$ are

$$\begin{aligned}
\Sigma_{\boldsymbol{y}_i\widetilde{\boldsymbol{\nu}}_i^n} &= \mathbb{E}\left[\left(\boldsymbol{y}_i - \overline{\boldsymbol{y}}_i\right)\widetilde{\boldsymbol{\nu}}_i^{n^T}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{e}_{i|i-1}^n + \left(\widehat{\boldsymbol{y}}_{i|i-1}^n - \overline{\boldsymbol{y}}_i\right)\right)\begin{bmatrix} \boldsymbol{e}_{i|i-1}^n - \boldsymbol{e}_{i|i-1}^{l_1} \\ \vdots \\ \boldsymbol{e}_{i|i-1}^n - \boldsymbol{e}_{i|i-1}^{l_{d_n}} \\ \widetilde{H}_n\boldsymbol{e}_{i|i-1}^n + \check{H}_n\boldsymbol{\epsilon}_{i|i-1}^n + H_n^T R_n^{-1}\boldsymbol{r}_i^n \end{bmatrix}^T\right] \\
&= \mathbb{E}\left[\boldsymbol{e}_{i|i-1}^n\left[\boldsymbol{e}_{i|i-1}^{n^T} - \boldsymbol{e}_{i|i-1}^{l_1^T} \;\middle|\; \cdots \;\middle|\; \boldsymbol{e}_{i|i-1}^{n^T} - \boldsymbol{e}_{i|i-1}^{l_{d_n}^T} \;\middle|\; \widetilde{H}_n\boldsymbol{e}_{i|i-1}^{n^T} + \check{H}_n\boldsymbol{\epsilon}_{i|i-1}^{n^T}\right]\right] \\
&= \left[P_{i|i-1}^{nn} - P_{i|i-1}^{nl_1} \;\middle|\; \cdots \;\middle|\; P_{i|i-1}^{nn} - P_{i|i-1}^{nl_{d_n}} \;\middle|\; P_{i|i-1}^{nn}\widetilde{H}_n^T + \Pi_{i|i-1}^{nn^T}\check{H}_n^T\right], \\
\Sigma_{\boldsymbol{x}_i\boldsymbol{\nu}_i^n} &= \mathbb{E}\left[\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i\right)\boldsymbol{\nu}_i^{n^T}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{i|i-1}^n + \left(\widehat{\boldsymbol{x}}_{i|i-1}^n - \overline{\boldsymbol{x}}_i\right)\right)\left(G\boldsymbol{\epsilon}_{i|i-1}^n - \boldsymbol{e}_{i|i}^n\right)^T\right] \\
&= \mathbb{E}\left[\boldsymbol{\epsilon}_{i|i-1}^n\left(G\boldsymbol{\epsilon}_{i|i-1}^n - \boldsymbol{e}_{i|i}^n\right)^T\right] = \Sigma_{i|i-1}^{nn}G - \Gamma_i^{nn}.
\end{aligned}$$

In the above derivation, using the iterated law of expectation it can be shown that:

$$\mathbb{E}\left[\left(\widehat{\boldsymbol{y}}_{i|i-1}^n - \overline{\boldsymbol{y}}_i\right)\widetilde{\boldsymbol{\nu}}_i^{n^T}\right] = \mathbb{E}\left[\left(\widehat{\boldsymbol{x}}_{i|i-1}^n - \overline{\boldsymbol{x}}_i\right)\boldsymbol{\nu}_i^{n^T}\right] = \mathbf{0}$$

Also, $\mathbb{E}\left[\boldsymbol{e}_{i|i-1}^n\boldsymbol{r}_i^{n^T}\right] = \mathbf{0}$ due to the statistical independence of the noise sequences. Similarly, the covariances matrices $\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n}$ and $\Sigma_{\boldsymbol{\nu}_i^n}$ are:

$$\Sigma_{\widetilde{\boldsymbol{\nu}}_i^n} = \mathbb{E}\left[\widetilde{\boldsymbol{\nu}}_i^n\widetilde{\boldsymbol{\nu}}_i^{n^T}\right]$$

$$= \mathbb{E}\left[\begin{bmatrix} \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_1}_{i|i-1} \\ \vdots \\ \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_{d_n}}_{i|i-1} \\ \widetilde{H}_n \boldsymbol{e}^n_{i|i-1} + \check{H}_n \boldsymbol{\epsilon}^n_{i|i-1} \\ + H_n^T R_n^{-1} \boldsymbol{r}^n_i \end{bmatrix} \begin{bmatrix} \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_1}_{i|i-1} \\ \vdots \\ \boldsymbol{e}^n_{i|i-1} - \boldsymbol{e}^{l_{d_n}}_{i|i-1} \\ \widetilde{H}_n \boldsymbol{e}^n_{i|i-1} + \check{H}_n \boldsymbol{\epsilon}^n_{i|i-1} \\ + H_n^T R_n^{-1} \boldsymbol{r}^n_i \end{bmatrix}^T\right]$$

$$\Sigma_{\boldsymbol{\nu}^n_i} = \mathbb{E}\left[\boldsymbol{\nu}^n_i \boldsymbol{\nu}^{n^T}_i\right] = \mathbb{E}\left[\left(G\boldsymbol{\epsilon}^n_{i|i-1} - \boldsymbol{e}^n_{i|i}\right)\left(G\boldsymbol{\epsilon}^n_{i|i-1} - \boldsymbol{e}^n_{i|i}\right)^T\right].$$

The rest of the derivation of $\Sigma_{\widetilde{\boldsymbol{\nu}}^n_i}$ and $\Sigma_{\boldsymbol{\nu}^n_i}$ in terms of the error covariance matrices is by block-by-block multiplication of the above expressions. □

By the Gauss-Markov theorem, the CIKF algorithm, along with this design of the consensus and innovation gain matrices, as stated in Theorem 5.5.1, results in the minimized MSE distributed estimates of the dynamic random field $\boldsymbol{x}_i$. The gain matrices are deterministic. Hence they can be precomputed offline and saved for online implementation. In Algorithm 5, we state the steps that each agent n runs to compute the optimal gain matrices.

Algorithm 5 Gain Design of CIKF

Input: Model parameters A, V, H, R, G, L, Σ_0.
Initialize: $\Sigma_{0|-1} = J \otimes \Sigma_0,\ P_{0|-1} = J \otimes (G\Sigma_0 G)\,,\ \Pi_{0|-1} = J \otimes (\Sigma_0 G)$.
while $i \geq 0$ **do**
Optimal gains:
Compute $\widehat{B}^n_i$ and K^n_i using Theorem 5.5.1.
Using (5.63)-(5.65), obtain $B^{\mathcal{C}}_i$, $B^{\mathcal{I}}_i$ and K_i from $\widehat{B}^n_i$ and K^n_i.
Prediction error covariance updates:
Update $P_{i+1|i}, \Sigma_{i+1|i}, \Pi_{i+1|i}$ using (5.39)-(5.44).
end while

The offline Algorithm 5 along with the online Algorithm 4 completes our proposed distributed solution to obtain minimized MSE estimates of the dynamic field $\boldsymbol{x}_i$ at each agent in the cyber network.

5.6 Numerical Evaluation

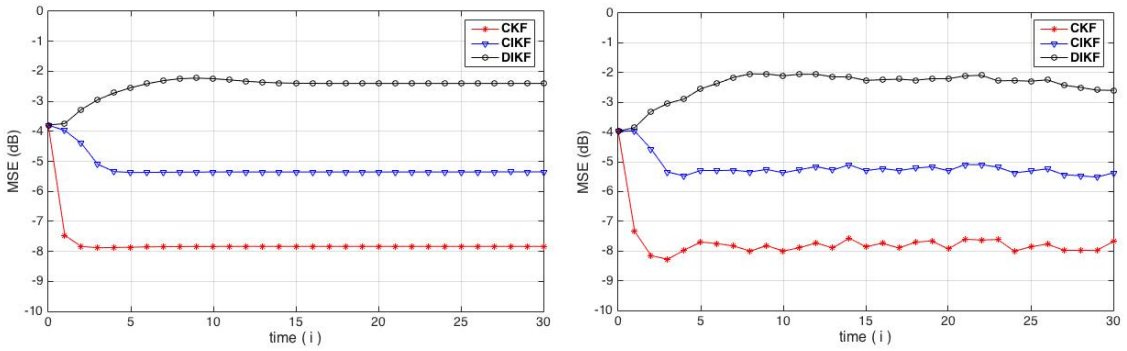
We numerically evaluate the MSE performance of the CIKF and compare it against the centralized Kalman filter (CKF) and the distributed information Kalman filter (DIKF) in [55]. To this

objective, we build a time-varying random system, observation and network model that satisfies the Assumptions 1-5. The Algorithms 4-5 run on these model parameters. First the Algorithm 5 computes and save the gain matrices and the error covariance matrices. The traces of the error covariance matrices provide the theoretical MSE trajectory of the CIKF with time. Then, we Monte-Carlo simulate Algorithm 4 to compute the numerical MSE of the distributed estimators, CIKF and DIKF, and the centralized estimator CKF.

5.6.1 Model specifications

Here, we consider a time-varying field, $\boldsymbol{x}_i$, with dimension $M = 50$. The physical layer, consisting of $M = 50$ sites, is monitored by a cyber layer consisting of $N = 50$ agents. Each agent in the cyber layer observes $M_n = 2$ sites of the physical layer. We build the field dynamics matrix A to be sparse and distributed. The dynamics A possess the structure of a Lattice graph, where the time evolution of a field variable depends on the neighboring field variables. For illustration, we consider an unstable field dynamics with $\|A\|_2 = 1.05$ to test the resilience of the algorithms under unstable conditions. The observation matrices, $H_n \in \mathbb{R}^{2\times 50}, n = 1, \cdots, 50$, are sparse $0-1$ matrices with one non-zero element at each row corresponding to the site of $\boldsymbol{x}_i$ observed by the n^{th} agent. The local observations $\boldsymbol{z}_i^n$ are 2×1 random vectors. The mean $\overline{x}_0$ of the initial state vector is generated at random. The system noise covariance V, the observation noise covariances R_n, and the initial state covariance Σ_0 are randomly generated symmetric positive definite matrices. The norms of the covariance matrices are: $\|V\|_2 = 4$, $\|R_n\|_2 = 8$, and $\|\Sigma_0\|_2 = 16$. The agents in the cyber layer communicate among themselves following a randomly generated Erdős-Rényi graph $\mathcal{G}$ with 50 nodes and $E = 138$ edges. The average degree of each node/agent is approximately 5.5. The communication network $\mathcal{G}$ is also sparse.

For Monte-Carlo simulations, we generate the noises, $\boldsymbol{v}_i, \boldsymbol{r}_i^n$, and the initial condition, $\boldsymbol{x}_0$ as Gaussian sequences, with $\boldsymbol{v}_i \sim \mathcal{N}(\bar{0}, V)$, $\boldsymbol{r}_i^n \sim \mathcal{N}(\bar{0}, R_n)$, $\boldsymbol{x}_0 \sim \mathcal{N}(\bar{\boldsymbol{x}}_0, \Sigma_0)$. The sequences $\{\{\boldsymbol{v}_i\}_i, \{\boldsymbol{r}_i\}_i, \boldsymbol{x}_0\}_{i\geq 0}$ are generated to be uncorrelated. Each agent n in the cyber layer has access to the system parameters $A, V, H, R, \bar{\boldsymbol{x}}_0, \Sigma_0$, and $\mathcal{G}$. This numerical model satisfies Assumptions 1-3. The pair (A, H) is detectable and the pairs, (A, H_n) $\forall n$, are not detectable. The agent communication graph $\mathcal{G}$ is connected with the algebraic connectivity of the Laplacian $\lambda_2(L) = 0.7 > 0$. Hence the Assumptions 4-5 hold true for this numerical system, observation, and network model.



a Theoretical MSE b Monte-Carlo MSE

Figure 5.1: Comparison of MSE performance of the proposed CIKF with CKF and DIKF.

5.6.2 Optimized gains and theoretical MSE

We run Algorithm 5 on the numerical model to obtain the gain matrices and the theoretical error covariances of the CIKF. We compute the gain matrices and the error covariances of the centralized Kalman filter (CKF) and of the distributed information Kalman filter (DIKF). In Fig. 5.1a, we plot the MSE, trace of the predictor error covariance matrices $\Sigma_{i+1|i}$, for each of these cases up to time $i = 30$. The MSE of the optimal CKF is the smallest (recall the CKF, if feasible, would be optimal) and the objective of the distributed estimators is to achieve MSE performance as close as possible to that of the CKF.

From Fig. 5.1a, we see that the MSE of the proposed CIKF is 3dB more than the CKF but is 3dB less than the DIKF. From the plot we see that the CIKF converges faster than the DIKF. Hence, the proposed CIKF provides faster convergence and 3dB MSE performance improvement over the DIKF. The performance of the CKF is 3dB better than the CIKF's due to the fact that the CKF has access to the observations of all the sensors at every time steps. In contrast in CIKF, each agent has access to its own observations and the current estimates of its neighbors only; the impact of the observations from the other agents propagate through the network with delay. As the time-varying field $\boldsymbol{x}_i$ is evolving with input noise $\boldsymbol{v}_i$, lack of access to all the observations containing the driving input $\boldsymbol{v}_i$, combined with the network diffusion delay, causes a performance gap between the CKF and the CIKF.

5.6.3 Monte-Carlo Simulations

We empirically compute the MSE of the distributed estimates given by the CIKF Algorithm 4. We implement the algorithms using Matlab in the Microsoft Azure cloud. Given the computation load because of the large system (M = 50) and network (N = 50) models, we run our simulation on Azure DS13 (8 cores, 56 GB memory) virtual machine (VM). The MSE computation for the CIKF, CKF, and DIKF algorithms with 1000 Monte-Carlo runs require approximately 30 hours in the Azure DS13 VM. Once we obtain the field prediction estimates for the three algorithms, we compute the empirical prediction error covariance matrices $\widehat{\Sigma}_{i+1|i}$ and then obtain the Monte-Carlo MSE from their trace. From the Monte-Carlo MSE plot in Fig 5.1b, we see that the the empirical plots follow closely the theoretical plots in Fig 5.1a.

Both the theoretical Fig 5.1a and Monte-Carlo simulated Fig 5.1b MSE performance confirms our CIKF analysis in Sections 5.3-5.5. The novel *Consensus+Innovations* Kalman Filter (CIKF) proposed in Section 5.2 along with the optimized gain designs in Section 5.5 provides unbiased distributed estimates with bounded and minimized MSE for the *Consensus+Innovations* distributed solution. The CIKF achieves nearly 3dB better performance than the DIKF [55].

5.7 Conclusions

In this chapter, we propose a *Consensus+Innovations* Kalman Filter (CIKF) that obtains unbiased minimized MSE distributed estimates of the pseudo-states and real-time employs them to obtain the unbiased distributed filtering and prediction estimates of the time-varying random state at each agent. The filter update iterations are of the *Consensus+Innovations* type. Using the Gauss-Markov principle, we designed the optimal gain matrices that yield approximately 3dB improvement over previous available distributed estimators like the DIKF in [55].

The three primary contributions of this chapter are: (a) introduction of the concept of pseudo-state; (b) design of a filter and corresponding gain matrices to obtain minimized MSE distributed estimates at each agent under minimal assumptions; and (c) a theoretical characterization of the tracking capacity and distributed version of the algebraic Riccati equation.

CHAPTER 6

Conclusions

The thesis develops algorithms and methodologies for distributed estimation of time-varying random fields over multi-agent networks. The random field considered in this work is assumed to be spread over a large geographical area. The sensors/agents monitoring the field are also spread over the large geographical area of the field. They make partial and local observations of the field. The sensors are low-power and inexpensive and are capable of exchanging information with their neighbors. We design distributed estimators capable of estimating and tracking the entire field at each sensor. The distributed estimation algorithms presented in this thesis are of the *consensus+innovations* Kalman filter type. The *consensus* aspect of the distributed estimators enables the agents in the network to come to a consensus on the estimate of the time-varying field. The *innovations* part ensures that the distributed estimates of the field at each agent are unbiased with bounded mean-squared error (MSE). We design consensus and innovation gain matrices such that the MSE performance is as close as possible to that of the centralized estimators, which has access to all the observations. Our solution is robust to central node failures and requires limited communication bandwidth at each sensor. The generic distributed algorithms presented in this thesis can be applied to estimate reliably and robustly large-scale systems, namely physical phenomena over large geographical areas, or beliefs in social networks, to mention two important illustrative applications.

6.1 Thesis Summary

We reiterate the structure of the three distributed estimators, Pseudo-innovations Kalman Filter (PIKF), Distributed Information Kalman Filter (DIKF), and *Consensus+Innovations* Kalman Filter

(CIKF), in Fig. 6.1. All three of them are comprised of two steps; the first is dynamic averaging and the second is field estimation. They achieve unbiased and bounded MSE distributed estimates of the entire field $\boldsymbol{x}_i,\ i \geq 0$, at each agent by employing a *consensus+innovations* algorithm.

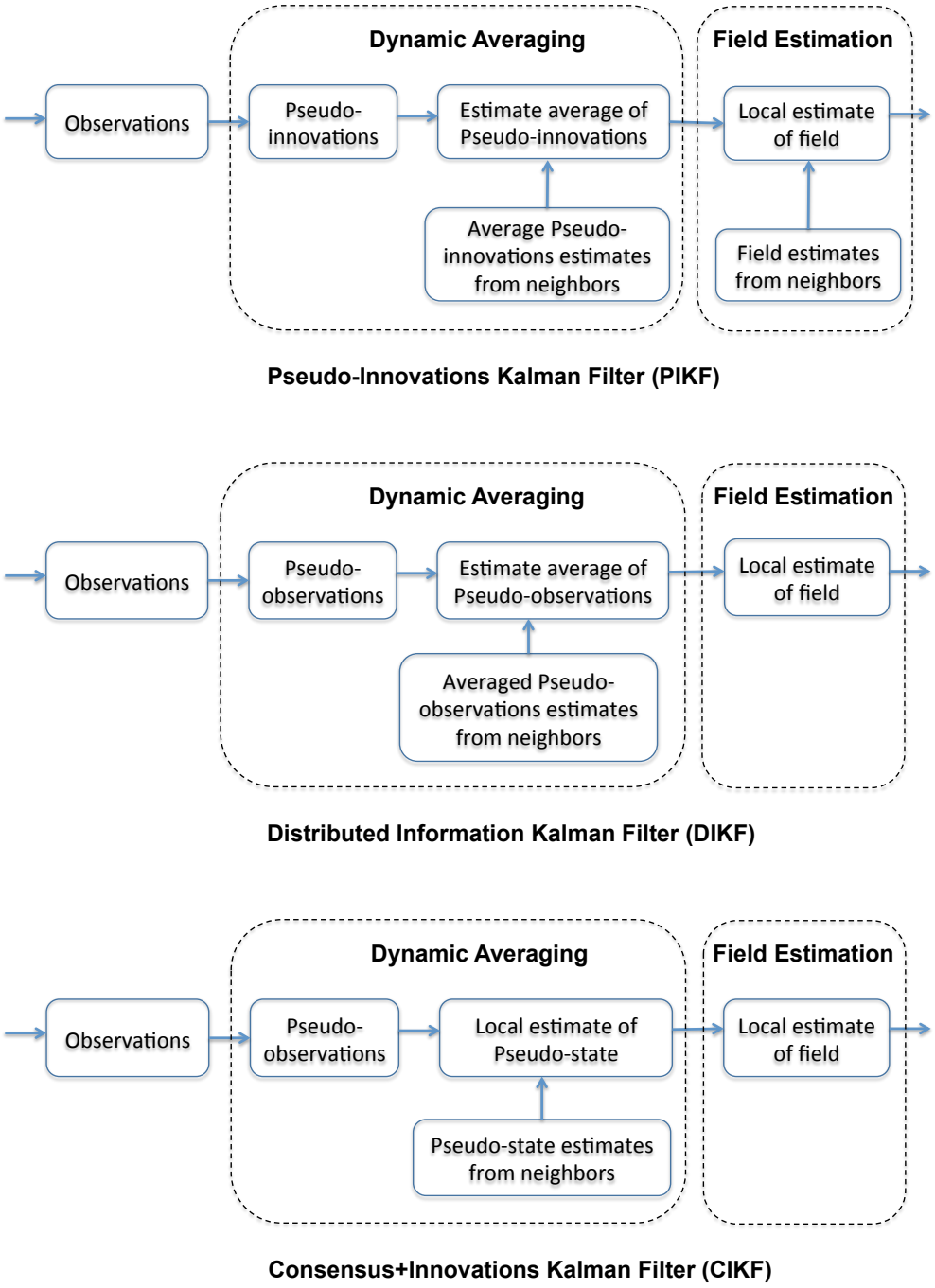


Figure 6.1: Structure of the distributed estimators: (top) PIKF, (middle) DIKF, & (bottom) CIKF.

In Chapter 3, the PIKF transforms the local observations into local pseudo-innovations, a modified version of the innovations. It employs a *consensus+innovations* step to compute the dynamic average of all the local pseudo-innovations by combining its own pseudo-innovation with that of its neighbors iteratively. In the field estimation step, the PIKF combines the dynamic averaged estimate of the pseudo-innovations with its own and neighbors' field estimates from previous time to compute the filtering and prediction estimates of the field.

In contrast to PIKF, the DIKF converts the local observations into local pseudo-observations, a linear transformation of the observations. In Chapter 4, we show that the DIKF executes a dynamic averaging step, termed Dynamic Consensus on Pseudo-observations (DCPO), to estimate the average of all the local pseudo-observations with a *consensus+innovations* iterative update. The consensus is on the local averaged pseudo-observation estimates of the neighbors. The DIKF incorporates the outcome of the dynamic averaging (DCPO) step to compute the estimate of the field. Note that, in DIKF, the agents exchange only the averaged pseudo-observations estimates and not the field estimates with their neighbors. The DIKF is a single time-scale version of the two time-scale distributed estimator proposed in [14, 19]. In [14, 19] the averaging step involves multiple communications exchanges of pseudo-observations between neighbors until they come to a consensus on the global average of all the pseudo-observations. Whereas in DIKF, each agent communicates with its neighbors only once in each update cycle and asymptotically comes to a consensus on the dynamic global average of all the pseudo-observations.

Similar to the PIKF and the DIKF, the structure of CIKF also has dynamic averaging and field estimation steps as shown in Fig. 6.1 (bottom). Here, we introduce the concept of pseudo-state, a linear transformation of the field. The pseudo-state can also be perceived as a noise-less version of the global average of all the pseudo-observations. In the dynamic averaging step, the CIKF computes the distributed estimate of the pseudo-state by applying the *consensus+innovations* approach to the pseudo-observations. Each agent then combines its own distributed pseudo-state estimate with its previous field estimate to obtain the distributed filtering and prediction estimates of the field.

For comparison, we restate the filtering and prediction update iterations of the three distributed estimators in Table 6.1. In PIKF, the local pseudo-innovation at agent n is $\boldsymbol{\nu}_i^n$. The *consensus+innovations* dynamic averaging step computes the distributed pseudo-innovations estimate $\widehat{\boldsymbol{\nu}}_i^n$ of the global average of all the pseudo-innovations. The *consensus+innovations* field

1. **PIKF**

Dynamic Averaging: $\widehat{\boldsymbol{\nu}}_i^n = \underbrace{\sum_{l\in\Omega_n} w_{nl}\widehat{\boldsymbol{\nu}}_{i-1}^l}_{\text{Consensus}} + \underbrace{\left(\boldsymbol{\nu}_i^n - B_n\sum_{l\in\Omega_n} w_{nl}\widehat{\boldsymbol{\nu}}_{i-1}^l\right)}_{\text{Innovations}}$

Field Estimation

Filtering: $\widehat{\boldsymbol{x}}_{i|i}^n = \underbrace{\sum_{l\in\Omega_n} w_{nl}\widehat{\boldsymbol{x}}_{i|i-1}^n}_{\text{Consensus}} + K_i^n \underbrace{\widehat{\boldsymbol{\nu}}_i^n}_{\text{Innovations}}$

Prediction: $\widehat{\boldsymbol{x}}_{i+1|i}^n = A\widehat{\boldsymbol{x}}_{i|i}^n$

2. **DIKF**

Dynamic Averaging: $\widehat{\boldsymbol{y}}_i^n = A_{_G}\underbrace{\sum_{l\in\Omega_n} w_{nl}\widehat{\boldsymbol{y}}_{i-1}^l}_{\text{Consensus}} + B\underbrace{\left(\boldsymbol{y}_i^n - H_{_G}^n A_{_G}\widehat{\boldsymbol{y}}_{i-1}^n\right)}_{\text{Innovations}}$

Field Estimation

Filtering: $\widehat{\boldsymbol{x}}_{i|i}^n = \widehat{\boldsymbol{x}}_{i|i-1}^n + K_i^n \underbrace{(\widehat{\boldsymbol{y}}_i^n - G\widehat{\boldsymbol{x}}_{i|i-1}^n)}_{\text{Innovations}}$

Prediction: $\widehat{\boldsymbol{x}}_{i+1|i}^n = A\widehat{\boldsymbol{x}}_{i|i}^n$

3. **CIKF**

Dynamic Averaging

Filtering: $\widehat{\boldsymbol{y}}_{i|i}^n = \widehat{\boldsymbol{y}}_{i|i-1}^n + \sum_{l\in\Omega_n}\underbrace{B_i^{nl}\left(\widehat{\boldsymbol{y}}_{i|i-1}^l - \widehat{\boldsymbol{y}}_{i|i-1}^n\right)}_{\text{Consensus}} + B_i^{nn}\underbrace{\left(\widetilde{\boldsymbol{z}}_i^n - \left(\widetilde{H}_n\widehat{\boldsymbol{y}}_{i|i-1}^n + \check{H}_n\widehat{\boldsymbol{x}}_{i|i-1}^n\right)\right)}_{\text{Innovations}}$

Prediction: $\widehat{\boldsymbol{y}}_{i+1|i}^n = \widetilde{A}\widehat{\boldsymbol{y}}_{i|i}^n + \check{A}\widehat{\boldsymbol{x}}_{i|i}^n$

Field Estimation

Filtering: $\widehat{\boldsymbol{x}}_{i|i}^n = \widehat{\boldsymbol{x}}_{i|i-1}^n + K_i^n\underbrace{\left(\widehat{\boldsymbol{y}}_{i|i}^n - G\widehat{\boldsymbol{x}}_{i|i-1}^n\right)}_{\text{Innovations}}$

Prediction: $\widehat{\boldsymbol{x}}_{i+1|i}^n = A\widehat{\boldsymbol{x}}_{i|i}^n$

Table 6.1: Dynamic averaging and field estimation updates of PIKF, DIKF and CIKF.

estimation step uses the distributed pseudo-innovations estimate $\widehat{\boldsymbol{\nu}}_i^n$ to compute the distributed filter estimate $\widehat{\boldsymbol{x}}_{i|i}^n$ and prediction estimate $\widehat{\boldsymbol{x}}_{i+1|i}^n$. In contrast, the DIKF uses the local pseudo-observation $\boldsymbol{y}_i^n$ and the averaged pseudo-observation estimates of neighbors to obtain the averaged estimate $\widehat{\boldsymbol{y}}_i^n$ of the global average of all the pseudo-observations. Using the output $\widehat{\boldsymbol{y}}_i^n$ of the

consensus+innovations dynamic averaging step, the DIKF computes the filtering and prediction estimates, $\widehat{\boldsymbol{x}}_{i|i}^n, \widehat{\boldsymbol{x}}_{i+1|i}^n$, in the field estimation step. In CIKF, the dynamic averaging step computes both the filtering and prediction distributed estimates of the pseudo-state, $\widehat{\boldsymbol{y}}_{i|i}^n, \widehat{\boldsymbol{y}}_{i+1|i}^n$, with *consensus+innovations* using local pseudo-observation $\widetilde{\boldsymbol{z}}_i^n$ and the pseudo-state estimates of the neighbors. The field estimation step then uses the filtering pseudo-state estimate $\widehat{\boldsymbol{y}}_{i|i}^n$ to compute the filtering and prediction estimates, $\widehat{\boldsymbol{x}}_{i|i}^n, \widehat{\boldsymbol{x}}_{i+1|i}^n$, of the time-varying random field $\boldsymbol{x}_i$.

The design parameters for each of the estimators are:

- PIKF: In Chapter 3, the consensus weight matrix is $W = [w]_{n,l} = I - \beta L$, where $\beta = \frac{2}{\lambda_2(L)+\lambda_N(L)}$; the pseudo-innovation gain matrix is $B_n = \frac{1}{\|G^{-1}\|_2} G^{-1}$; and the filter gain matrix is $K_n = G^{-1}$. Here, L is the graph Laplacian of the agent communication network and $G = \sum_{n=1}^N H_n^T R_n^{-1} H_n$.

- DIKF: In Chapter 4, the consensus weight matrix is $W = [w]_{n,l} = I_M - \beta_1 L$ and the pseudo-observation gain matrix is $B_n = \beta_2 I_M$, where the choice of the constants $\{\beta_1, \beta_2\}$ is discussed in (4.34)-(4.46). The optimized filter gain matrix K_i^n is computed from the distributed version of the algebraic Riccati equations (4.67)-(4.69). Alternatively, we provide a scalar and computationally simple filter gain matrix $K_i^n = \alpha G^{-1}$ in (4.63).

- CIKF: In Chapter 5, the consensus gain matrices $B_i^{n,l}$, the pseudo-state gain matrix $B_i^{n,n}$, and the filter gain matrix K_i^n are designed using the Gauss-Markov type strategy (5.60)-(5.65) and the distributed version of the algebraic Riccati equations (5.39)-(5.44). This optimal design of the gain matrices minimizes the MSE of the CIKF.

To compute the consensus weights and the gain matrices locally, each agent needs to know the field dynamics matrix A, the observation matrices $H_n,\ n = 1, \cdots, N$, the graph Laplacian L of the agent communication network, the covariance Σ_0 of the initial condition, and the noise covariances V and R. Otherwise, the computations can be done at a centralized processor that can then broadcast the local gains to the corresponding agents. The gains can be computed offline and stored to save real time implementation complexity. Instead of time-varying gain matrices, the distributed estimators can use the asymptotic gain matrices $K_\infty,\ B_\infty$. This reduces the storage requirement locally at the cost of some loss of MSE performance. Once each agent has access to the consensus

	PIKF	**DIKF**	**CIKF**
Pseudo-quantities	Pseudo-innovations	Pseudo-observations	Pseudo-state
NTC	Depends on the network connectedness (L).	Depends on both the network connectedness (L) and the local observation models (H_n).	Depends on both the network connectedness (L) and the local observation models (H_n).
Gain	Scalar consensus weight, time-invariant pseudo-innovations, and filter gain matrices.	Scalar consensus weight, scalar pseudo-observations gain, and optimized filter gain matrix.	Optimized consensus gain, pseudo-state gain, and filter gain matrices.
MSE Performance	Bounded MSE.	Bounded MSE and 2dB better performance than the PIKF.	Best MSE performance. It loses 3dB compared to centralized estimator. CIKF's MSE is 3dB better than that of DIKF.
Trade-offs	NTC can be increased by just rewiring network topology. Flexibility in estimating unstable fields.	Better MSE estimates than the PIKF and simpler to implement as compared to the CIKF.	Minimized MSE estimates with the optimized but computationally expensive gains.

Table 6.2: Comparison between the distributed estimators PIKF, DIKF and CIKF.

weights and its gain matrices, the online implementation of the distributed estimators requires each agent to know only the dynamics A and its own observation matrices H_n.

In Table 6.2, we compare and contrast the different aspects and trade-offs of the three distributed estimators developed in this thesis. In terms of MSE performance CIKF is the best among the three. The PIKF provides the additional flexibility of adapting a distributed estimator to highly unstable systems because its NTC can be increased simply by rewiring the network topology. The DIKF has computationally cheap implementation provisions compared to the CIKF at a better MSE performance than the PIKF. The three distributed estimators PIKF, DIKF, and CIKF are not linear transformations of each other, unlike the relationship between the centralized Kalman filter and the information filter. This difference is due to the dynamic averaging step where it computes the

pseudo-quantities. The comparisons are further explained in the summary and contributions of the thesis.

We outline a brief summary of the thesis.

In **Chapter 2**, we present the three key components of the distributed field estimation, namely, dynamics, observations, and communications. We assume that the input and observation noises are white Gaussian and uncorrelated among themselves and the initial conditions. The objective is to obtain unbiased distributed estimates of the entire field at each agent, where each agent makes local observations only and performs local information exchange with its neighbors. We discuss the advantages and challenges of single time-scale distributed estimators over two time-scale distributed estimators. The primary challenge to be overcome in the single time-scale solution is that diffusion of information over the agent network may be slow and that may affect the performance and the convergence of the distributed estimators.

Chapter 3 proposes Pseudo-Innovations Kalman Filter (PIKF), a distributed estimator for the time-varying random field that runs a companion algorithm to obtain the distributed estimates of the global average of all the pseudo-innovations at each agent. The averaged pseudo-innovation estimates are combined with the filter estimates to predict the field at a future time index. The prediction estimates are unbiased with bounded MSE as long as the field dynamics are upper-bounded by the Network Tracking Capacity (NTC). The NTC of the PIKF is a function of the eigenvalues of the network Laplacian and is independent of the specifics of the local observation models, assuming the dynamical system and the observations are globally observable. This provides the flexibility of increasing the NTC for unstable fields by just rewiring the agent communication network.

Chapter 4 develops a *consensus+innovations* type Dynamic Consensus on Pseudo-Observations (DCPO) algorithm that computes the distributed estimate of the global average of the pseudo-observations. The Distributed Information Kalman Filter (DIKF) presented in this chapter uses the distributed averaged pseudo-observations estimates for the distributed estimation of the dynamic random field. The MSE error performance of the DIKF is lower than the PIKF in Chapter 3. We demonstrate the bounded MSE convergence of the DIKF through extensive experimental evaluations and study the sensitivity of the DIKF with respect to model parameters, noise levels, and network parameters.

In **Chapter 5**, we present the *Consensus+Innovations* Kalman Filter (CIKF) that computes the unbiased distributed filtering and prediction estimates of the time-varying random field using unbiased minimized MSE distributed estimates of the pseudo-states. The distributed estimates of the pseudo-states reduce the noise of the innovations in the filtering step of field estimation and hence further improve the MSE performance of the distribuetd estimator. We design the optimal consensus and innovation gain matrices that provide approximately 3dB improvement over the DIKF in Chapter 4.

The main **contributions** of the thesis are:

1. Introduced the concepts of pseudo-innovations, pseudo-observations, and pseudo-state and considered their relevance in the context of distributed estimation of time-varying random fields.

2. Introduced distributed dynamic averaging as a key part of distributed field estimation. Developed *Consensus+Innovations* distributed algorithms to achieve unbiased estimates of global average of the pseudo-innovations (PIKF in Chapter 3), pseudo-observations (DIKF in Chapter 4), and pseudo-state (CIKF in Chapter 5) with bounded MSE.

3. Provided scalar consensus weights and time-invariant innovation gain matrices for dynamic averaging of pseudo-innovations in PIKF. Designed consensus and innovations scalar gains for the distributed dynamic consensus in DIKF. Derived consensus and innovations gain matrices for the distributed dynamic averaging in CIKF using the Gauss-Markov Theorem.

4. Proposed novel *Consensus+Innovations* distributed estimators, PIKF, DIKF, and CIKF of time-varying random fields over multi-agent networks. Proved that the estimators achieve unbiased estimates with bounded MSE under minimal assumptions on the local observation and network communication models.

5. Designed the filter gain matrices using the Gauss-Markov theorem, so that the estimation MSE is minimized given the distributed estimates of the averaged pseudo-observations and pseudo-state.

6. Computed explicit expressions for the Network Tracking Capacity of the distributed estimators based on the network Laplacian and the local observation matrices.

7. Analyzed the filtering and prediction error processes and derived the distributed version of the algebraic Riccati equation for the DIKF and the CIKF. The convergence rate and the MSE performance of the distributed estimators depend on the distributed algebraic Riccati equation.

8. Validated the theoretical error convergence results through numerical evaluations. Evaluated experimentally the sensitivity of the performance of the distributed estimators with respect to model parameters, noise statistics, gain variations, and network models.

6.2 Future Work

The generic nature of the results in this thesis makes them applicable in a variety of problems in distributed inference. The techniques we present can be used to design consensus gain matrices in averaging and dynamic averaging problems over large networks in order to improve the convergence rate of the distributed consensus problems. These results can help in the study of dynamic belief evolution in social networks.

Interesting extensions include developing a distributed estimator that is resilient to random sensor link and node failures. This is important in practice because the communication links between sensors or the sensors themselves can fail. For distributed parameter estimation, this has been considered in [27].

The distributed estimators proposed in the thesis converge if the degree of instability of the field dynamics is within the Network Tracking Capacity (NTC). For asymptotically stable time-varying random fields, the estimators always converge, given the assumptions of global detectability and connected network hold true. Deriving similar conditions for unstable systems is of interest to eliminate the NTC restrictions on the distributed solutions. A preliminary approach on deriving such network detectability conditions for unstable systems is presented in [24]. A complete condition would probably result from a joint optimization on a functional form of NTC and estimation MSE.

Bibliography

[1] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of basic Engineering*, vol. 83, no. 3, pp. 95–108, 1961.

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

[3] J. R. P. de Carvalho, E. D. Assad, and H. S. Pinto, "Kalman filter and correction of the temperatures estimated by precis model," *Atmospheric Research*, vol. 102, no. 1, pp. 218–226, 2011.

[4] G. Battistelli, L. Chisci, N. Forti, G. Pelosi, and S. Selleri, "Distributed finite element Kalman filter," in *European Control Conference*. IEEE, 2015, pp. 3695–3700.

[5] W. Ledsham and D. Staelin, "An extended Kalman-Bucy filter for atmospheric temperature profile retrieval with a passive microwave sounder," *Journal of Applied Meteorology*, vol. 17, pp. 1023–1033, 1978.

[6] Y. Chen, J. Xu, K. Luo, and S. Xu, "Multi-temperature and humidity data fusion algorithm based on Kalman filter," *Research Journal of Applied Sciences*, vol. 5, 2013.

[7] M. D. Ilic, L. Xie, U. A. Khan, and J. M. F. Moura, "Modeling future cyber-physical energy systems," in *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–9.

[8] P. Du, Z. Huang, Y. Sun, R. Diao, K. Kalsi, K. K. Anderson, Y. Li, and B. Lee, "Distributed dynamic state estimation with extended Kalman filter," in *North American Power Symposium (NAPS)*, 2011, pp. 1–6.

[9] S. H. Lee and J. W. Park, "Selection of optimal location and size of multiple distributed generations by using Kalman filter algorithm," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1393–1400, 2009.

[10] R. Olfati-Saber and P. Jalalkamali, "Collaborative target tracking using distributed Kalman filtering on mobile sensor networks," in *American Control Conference (ACC)*, 2011, pp. 1100–1105.

[11] H. Medeiros, J. Park, and A. Kak, "Distributed object tracking using a cluster-based Kalman filter in wireless camera networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 448–463, 2008.

[12] D. Acemoglu, A. Nedic, and A. Ozdaglar, "Convergence of rule-of-thumb learning rules in social networks," in *47th IEEE Conference on Decision and Control*, 2008, pp. 1714–1720.

[13] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *44th IEEE Conference on Decision and Control, and 8th European Control Conference*, 2005, pp. 8179–8184.

[14] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *46th IEEE Conference on Decision and Control*, 2007, pp. 5492–5498.

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

[16] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, 2008.

[17] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links - Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, 2008.

[18] A. Ribeiro, I. D. Schizas, S. Roumeliotis, and G. B. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Systems Magazine*, vol. 30, no. 2, pp. 66–86, 2010.

[19] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *48th IEEE Conference on Decision and Control, and 28th Chinese Control Conference*, 2009, pp. 7036–7042.

[20] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[21] D. W. Casbeer and R. Beard, "Distributed information filtering using consensus filters," in *American Control Conference*. IEEE, 2009, pp. 1882–1887.

[22] B. S. Rao and H. F. Durrant-Whyte, "Fully decentralised algorithm for multisensor Kalman filtering," in *IEE Proceedings D - Control Theory and Applications*, vol. 138, no. 5. IET, 1991, pp. 413–420.

[23] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4782–4795, 2006.

[24] S. Kar and J. M. F. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1766–1784, 2011.

[25] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.

[26] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman filtering over massive data sets: analysis through large deviations of random Riccati equations," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1351–1372, 2015.

[27] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[28] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, 2011.

[29] S. Kar and J. M. F. Moura, "Consensus+innovations distributed inference over networks: cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, May 2013.

[30] U. A. Khan, S. Kar, A. Jadbabaie, and J. M. F. Moura, "On connectivity, observability, and stability in distributed estimation," in *49th IEEE Conference on Decision and Control*, 2010, pp. 6639–6644.

[31] U. A. Khan and A. Jadbabaie, "Networked estimation under information constraints," *arXiv preprint arXiv:1111.4580*, 2011.

[32] U. A. Khan and A. Jadbabaie, "Coordinated networked estimation strategies using structured systems theory," in *50th IEEE Conference on Decision and Control, and 11th European Control Conference*, 2011, pp. 2112–2117.

[33] M. Doostmohammadian and U. A. Khan, "On the genericity properties in distributed estimation: Topology design and sensor placement," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 195–204, 2013.

[34] S. Park and N. C. Martins, "An augmented observer for the distributed estimation problem for LTI systems," in *American Control Conference*, 2012, pp. 6775–6780.

[35] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[36] S. A. Aldosari and J. M. F. Moura, "Distributed detection in sensor networks: Connectivity graph and small world networks," in *39th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2005, pp. 230–234.

[37] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, 2007.

[38] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Distributed sensor fusion using dynamic consensus," in *IFAC World Congress*, 2005.

[39] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[40] H. Bai, R. A. Freeman, and K. M. Lynch, "Robust dynamic average consensus of time-varying inputs," in *49th IEEE Conference on Decision and Control*, 2010, pp. 3104–3109.

[41] S. Rajagopalan and D. Shah, "Distributed averaging in dynamic networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 845–854, 2011.

[42] S. Kar and J. M. F. Moura, "Asymptotically efficient distributed estimation with exponential family statistics," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4811–4831, 2014.

[43] A. Mohammadi and A. Asif, "Distributed consensus innovation particle filtering for bearing/range tracking with communication constraints," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 620–635, 2015.

[44] F. R. Chung, *Spectral Graph Teory*. American Mathematical Society, 1997, vol. 92.

[45] O. Hlinka, F. Hlawatsch, and P. M. Djuric, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 61–81, 2013.

[46] S. S. Dias and M. G. S. Bruno, "Cooperative target tracking using decentralized particle filtering and rss sensors," *IEEE Transactions on Signal Processing*, vol. 61, no. 14, pp. 3632–3646, 2013.

[47] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Dover Publications, 2012.

[48] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.

[49] S. Das and J. M. F. Moura, "Distributed state estimation in multi-agent networks," in *38th IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 4246–4250.

[50] S. Das and J. M. F. Moura, "Distributed Kalman filtering," in *21st European Signal Processing Conference*, 2013, pp. 1–5.

[51] S. Das and J. M. F. Moura, "Distributed linear estimationof dynamic random fileds," in *51st Annual Allerton Conference on Communication, Control, and Computing*, 2013, pp. 1120–1125.

[52] A. Ben-Israel and T. Greville, *Generalized inverses: theory and applications*. Springer, 2003, vol. 15.

[53] S. Kar, S. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, 2008.

[54] S. Das and J. M. F. Moura, "Distributed Kalman filtering and network tracking capacity," in *47th Asilomar Conference on Signals, Systems, and Computers*, 2013, pp. 629–633.

[55] S. Das and J. M. F. Moura, "Distributed Kalman filtering with dynamic observations consensus," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4458–4473, 2015.

[56] P. Alriksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," in *17th International Symposium on Mathematical Theory of Networks and Systems*, 2006, pp. 2445–2450.

[57] M. E. J. Newman, *Networks: an introduction*. Oxford University Press, 2010.

[58] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[59] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[60] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed estimation of dynamic parameters: Regret analysis," *arXiv preprint arXiv:1603.00576*, 2016.