DISTRIBUTED SIGNAL PROCESSING OVER LARGE-SCALE COMPLEX SYSTEMS

A Dissertation

by

DI LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Shuguang Cui
Committee Members,	Tie Liu
	I-Hong Hou
	Joel Zinn
Head of Department,	Miroslav M. Begovic

May 2017

Major Subject: Electrical Engineering

ABSTRACT

Large-scale and complex dynamical networks with high-dimension states have been emerging in the era of big data, which potentially generate massive data sets. To deal with the massive data sets, one promising method is the distributed collaboration strategy over the network. This dissertation proposes the schemes of distributed estimation and distributed quickest detection and also studies the performance of the distributed schemes with the large deviation analysis, which answers a fundamental question on how to quantify the rate at which the distributed scheme approaches the centralized performance.

First, the distributed Kalman filtering scheme with the Gossip interaction among sensors is proposed to estimate the high-dimension states at each node, where sensors exchange their filtered states (estimates and error covariance) and propagate their observations via inter-sensor communications. The conditional estimation error covariance sequence at each sensor under this scheme is proven to evolve as a random Riccati equation (RRE) with Markov modulated switching. By formulating the RRE as a random dynamical system, it is shown that the network consensus over the estimation at each node is achieved. The large deviation analysis further shows that the distributed scheme converges to the optimal centralized one at an exponentially fast rate. By considering the energy and bandwidth constrains, a Quantized Gossip-based Interactive Kalman Filtering algorithm for scalar dynamic systems is also proposed, where the sensors exchange their quantized states with neighbors via inter-sensor communications. It is shown that, in the countable infinite quantization alphabet case, the network can still achieve weak consensus with the additional information loss caused by quantization. It is also proved that, under certain conditions,

the network can also achieve weak consensus with the finite quantization alphabet, which is more restricted and practical.

Then, the distributed quickest detection scheme is proposed with multiple rounds of inter-sensor communications to propagate observations during the sampling interval. By modeling the information propagation dynamics in the network as a Markov process, the two-layer large deviation analysis is used to analyze the performance of the distributed scheme. The first layer analysis proves that the probability of false alarm decays to zero exponentially fast with the increasing of the averaged detection delay, where the Kullback-Leibler (KL) information number is established as a crucial factor. The second-layer analysis shows that the probability of the rare event that not all observations are available at a sensor decays to zero at an exponentially fast rate when the number of communications increases, where the large deviation upper and lower bounds for this rate are also derived, based on which it is shown that the performance of the distributed algorithm converges exponentially fast to that of the centralized one, by proving that the defined distributed KL information number converges to the centralized KL information number.

ACKNOWLEDGEMENTS

This thesis work summarizes my Ph.D. study and the life in the past years at College Station, and it is a milestone in my whole life. Also it incorporates many people's contribution and their memorable advices and helps. At this stage, I would like to take this great opportunity to express my sincere thanks to them.

Firstly, I would like to express my appreciation and thanks to my advisor, Prof. Shuguang Cui. I want to thank him not only for his advice and guidance in the study and research, but also his warm concern and encouragement in the life and career direction. Back to years ago, I thank him for his support and the opportunity given for me to pursue the Ph.D. degree and do the research work at Texas A&M University. I have learned from him so many whole life-beneficial things such as responsibility, focusing on details and the sharp perspective to view things. I also would like to thank my committee members, Prof. Tie Liu, Prof. I-Hong Hou and Prof. Joel Zinn. Thanks for their insightful comments and advices. It is such a wonderful and memorable experience for me to interact with my committee members. I also want to express my great thanks to my collaborators, Prof. Soummya Kar and Prof. José M. F. Moura from Carnegie Mellon University, Prof. H. Vincent Poor from Princeton University and Prof. Lifeng Lai from University of California, Davis. The experiences to collaborate with them are tremendous enjoyable and achievable, and I also learned a lot from them.

Secondly, many thanks to my colleagues and friends. Chuan Huang, Qing Zhou, Lili Zhang, Qiong Wu and Yang Zhou, they all graduated and left years ago. Thanks for their help in both the study and living during my first year here. The frequent group dinners at that time, such as hot-pot and barbecue, made life relaxed and

colorful. They are memorable. Thanks for Hang Li, Liang Ge, Yanyan Zhang and Dan Lv, my peers and labmates. It is such a pleasurable experience to share the office with them, and study and discuss problems together.

Lastly, I would like to express my thanks to my big family: my father, my mother, my father-in-law and my mother-in-law. Thank my parents for growing me up; thanks for their endless love and support; thanks for their miss when I am away from home. Thank my parents-in-law for their support and also their visiting us. Thanks for their significant help to take care our small family and let me focus on the study and work. I would like to express my special thanks to my beloved wife, Qing Wang. I am so grateful and feeling so heartwarming for her company in my Ph.D. journey. Thanks for her constant encouragement when I was in depression; thanks for her waiting me back home from office at late night; thanks for her giving-up to support me. At the end, thanks to our lovely daughter, little Annie. Her birth is such a happiness that bring so much sunshine to our family and to my life.

Thanks for everyone who helped and companied me in my Ph.D. journey!

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

Large scale and complex dynamical systems with high-dimension states have been emerging, which potentially generate the massive data sets, as the trending in the era of big data. The novel signal processing methods dealing with these large scale and complex systems are required. For estimation aspect, in research areas such as the system control, power grid and communication network, we need to resolve or estimate the states of these dynamical systems, and then further corresponding actions could be taken. Due to the high-dimension of states, it is hard to resolve the system states at one node, since only portion of high-dimension states could be observed at one node with its limited sensing capability. The nodes need to collaborate with each other in order to complete the estimation task. Since the centralized method has some disadvantages including issues such as poor scalability and robustness, and high communication and computation burdens, which are specified later, one promising method to deal with the above issues is the distributed collaboration strategy over the sensor network. This distributed philosophy is also needed in the detection problems in the large scale systems.

In this chapter, we first introduce background and motivations for the work in this thesis, including the literature review therein. Then we summarize the contributions of this thesis, with the statement of specific problems solved.

1.1 Background and Motivations

With the sensor network monitoring or sensing the large scale systems, each sensor or node is not able to observe the whole state process in the system, due to the well-known capability constrains of sensors, such as the limited battery capacity, the low computation ability, and the limited sensing range. Therefore, only a portion of the

state process in the large scale system could be possibly observed at each sensor. To resolve the dynamical states of the system, an apparent and ideal way is to collect the observations from all sensors and then implement the estimation or filtering algorithm over the whole set of observations to estimate the system states. This is the so-called centralized philosophy, requiring the existence of a fusion or control center and each sensor communicating and sending the collected observation to this fusion center, where the signal processing method is implemented over the received observations from all sensors. For this centralized method, there are some obvious disadvantages, mainly summarized into the following three aspects: 1) Scalability and robustness: The centralized method is fragile to the attack leading to the failure of the fusion center or the sensors nearby the fusion center that play the critical role in the multi-hop routing protocol to relay the observations from other sensors, and it needs to redesign the network protocol when extending the network size with additional sensor implemented. 2) Communication burden: With the large scale of the network, if designing the protocol of one-hop communication from the sensors to the fusion center, the transmission power at those faraway sensors have to be high enough to support the communication, which conflicts to the low power supplies at sensors. Usually the multi-hop routing protocol is adopted, then sensors nearby the fusion center have to relay and communicate much more frequently, which leads to heavy communication burden and easily causes death of those sensors. 3) Computation burden: The fusion center takes heavy computation burden to process the massive data sets collected from the sensors, especially in the large scale network. To solve the above concern, distributed signal processing philosophy have been taken into consideration.

To be specific, for distributed estimation in a wireless sensor network [74, 50], multiple spatially distributed sensors collaborate to estimate the system state of

interest, without the support of a central fusion center due to physical constraints such as large system size and limited communications infrastructure. Specifically, each sensor makes local partial observations and communicates with its neighbors to exchange certain information, in order to enable this collaboration. Due to its scalability for large systems and robustness to sensor failures, distributed estimation techniques find promising and wide applications including in battlefield surveillance, environment sensing, or power grid monitoring. Especially in the era of big data and large systems, which usually require overwhelming computation if implemented in centralized fashion, distributed schemes become critical since they can decompose the computational burden into local parallel procedures. A principal challenge in distributed sensing, and in distributed estimation in particular, is to design the distributed algorithm to achieve reliable and mutually agreeable estimation results across all sensors, without the help of a central fusion center. Further prior work addressing the above concerns is found in [56] and [28], with detailed surveys in [49, 17] and the literature cited therein.

This thesis studies the Modified Gossip Interactive Kalman Filtering (M-GIKF) for distributed estimation over potentially big data sets generated by a large dynamical system, in which each sensor observes only a portion of the large process, such that, if acting alone, no sensors can successfully resolve the entire system. The M-GIKF is fundamentally different from other distributed implementations of the Kalman filter, such as [53], [48], [34], and [10], which usually employ some form of averaging on the sensor observations/estimations through linear consensus or distributed optimization techniques. In [53], decentralization of the Kalman filtering algorithm is realized, where each node implements its own Kalman filter, broadcasts its estimate to every other node, and then assimilates the received information to reach certain agreement. In [48], the author proposes an approximate distributed

Kalman filtering algorithm by decomposing the central Kalman filter into n micro Kalman filters with inputs obtained by two consensus filters over the measurements and inverse covariance matrices. In [34], distributed Kalman filtering is derived for large-scale systems, where low-dimensional local Kalman filtering is achieved by spatially decomposing the large-scale system and adopting bipartite fusion graphs and consensus averaging algorithms. In [10], the authors formulate distributed Kalman filtering for a scalar system as an optimization problem to minimize the trace of the asymptotic error covariance matrix and study the interaction among the consensus matrices, the number of messages exchanged, and the Kalman gains. Single time-scale distributed approaches, i.e., in which only one round of inter-sensor message exchange is permitted per observation sampling epoch, are considered in [33], [51]. The distributed Kalman filtering algorithm in [33] involves a dynamic consensus mechanism in which at every observation sampling round each sensor updates its local estimate of the system state by combining a neighborhood consensus cooperation term (based on a single round of inter-agent message exchange) with a local innovation term (based on the new observation data sensed). The resulting distributed algorithm can track unstable dynamics with bounded mean-squared error (MSE) as long as the degree of instability of the dynamics is within a so called Network Tracking Capacity (NTC) of the agent network. A generic characterization of agent networks in which the above dynamic consensus based algorithm provides tracking with bounded MSE is provided in [18], where the authors employ structural system theoretic tools to obtain conditions on the communication topology and sensing model structure that guarantee tracking with bounded MSE. Another class of dynamic consensus type distributed observers/estimators has been proposed in [51], in which, in addition to updating their local state estimates, the agents propagate an additional *augmented* state in a distributed fashion. Conditions on local innovation

gain selection and coupling between the estimate and augmented state updates were obtained that guarantee stable tracking performance. More recently, an extension of the algorithm in [33] is proposed in [13], which performs dynamic consensus on *pseudo-innovations*, a modified version of the innovations, to improve estimation performance. A conceptually different single time-scale distributed Kalman filtering scheme was considered in [30], in which inter-agent cooperation was obtained by randomized estimate swapping among neighboring agents. Under rather weak assumptions on the detectability of the global sensing model and connectivity of the inter-agent communication network, the algorithm in [30] was shown to yield stochastically bounded estimation error at each agent. Moreover, the conditional error covariance at each agent was shown to converge to a stationary distribution of an associated random Riccati equation.

In wireless sensor networks, quantization is usually required before the data is exchanged through inter-sensor communications [44, 71, 41, 2, 20], since the limited sources, such as bandwidth and power, prevent the exchange of high-precision data (e.g., real-valued analog data) among the sensors. For quantized Kalman filtering in the literature, in [47], the innovation is quantized by either an iterative binary quantizer or a single-shot batch quantizer, and a recursive state estimator is introduced. In [63], Kalman filters based on both quantized observations and quantized innovations are proposed, and the tradeoff between energy consumption and estimation accuracy is studied. In [73], an optimal quantization method over observations and a transmit power scheduling strategy for the decentralized estimation in an inhomogeneous sensor network is proposed to minimize the total transmit power. In [40], a distributed adaptive one-bit quantization scheme over observations is proposed for distributed estimation, where each individual sensor dynamically adjusts the threshold of its quantizer.

Another distributed scheme studied in this thesis is the distributed quickest detection. Quickest change detection problems focus on detecting abrupt changes in stochastic processes as quickly as possible, with constraints to limit the detection error. Quickest change detection has wide applications in fields such as signal and image processing [35, 39, 69], computer network intrusion detection [68, 64, 9], neuroscience [12], environment and public health surveillance [21, 61], and system failure detection [54, 42]. Specifically, when quickest change detection is implemented in sensor networks [5, 45, 65], it can detect the change of statistical features, such as the mean and variance, over the observation sequences taken by sensors. For example, quickest change detection can be implemented in sensor networks for chemical industry to monitor the leakage, or to surveille the change of temperature in the field, by detecting the change in statistical patterns.

For signal processing implementation in sensor networks, essentially it can be divided into the following two categories: centralized vs. distributed algorithms. For centralized quickest change detection algorithms [67, 70, 23, 46, 75, 38, 4], a control or fusion center exists to process the data in a centralized way. Specifically, in centralized algorithms, they assume that either the raw observations from all the sensors or certain pre-processed information from the sensors (some people call this case as decentralized sensing) are available to the control or fusion center via certain communication channels; then a final centralized detection procedure is executed at the center. However, centralized algorithms have some disadvantages, such as heavy communication burden, high computation complexity, low scalability, and poor robustness. On the contrary, distributed implements do not require a control or fusion center, and the detection procedure is implemented at each sensor in a local and parallel fashion, with interactions among sensors in the neighborhood to exchange information. While centralized quickest change detection algorithms have been well-

studied, there are fewer literatures on the study of distributed algorithms for quickest change detection problems [7, 62], which become more desired in large-scale networks with a huge volume of data, in order to reduce the overall computation complexity and to enhance scalability. In [7], a distributed consensus based Page's test algorithm, using cumulative sum (CUSUM) log-likelihood of the data, was proposed, with the assumption that the change happening time is deterministic but unknown, which is called a non-Bayesian setup. In [62], a distributed change detection algorithm was proposed, to combine a global consensus scheme with the geometric moving average control charts that generate local statistics.

In both [7] and [62], non-Bayesian setups of the change happening time are considered, where the communication stage and the observation stage are interleaved, i.e., they are at the same time scale and each is executed once within one system time slot. Under such an interleaving strategy, the convergence of the test statistic is established when the system time goes to infinity. However, this type of convergence analysis over time does not fit well into quickest change detection problems, which are time-sensitive, with the goal to detect the change as quickly as possible. This is different from traditional detection problems without much consideration of the timing issue, where the convergence analysis is commonly performed as the system time goes to infinity.

1.2 Contributions

For the distributed estimation, the proposed M-GIKF achieves sensor collaboration by exchanging local estimation states and propagating observations between neighbor sensors. In M-GIKF, each sensor runs a local Kalman filter. At each signal evolution epoch, each sensor first randomly selects a neighbor with which to exchange its state (their local Kalman filter state estimate and conditional error covariance),

then propagates its observations to randomly selected neighbors, and lastly updates the estimate based on the received states and accumulated observations. This kind of collaboration through state exchange and observation propagating occurs distributedly and randomly, being controlled by the random network topology provided by an underlying gossip protocol. In M-GIKF, we assume that the communication channels among neighbors are ideal, implying that we precisely convey the sensor states and observations without distortion. The M-GIKF scheme introduced in this thesis generalizes the (GIKF) scheme introduced in the prior work [30], in which inter-sensor communication and signal evolution operate at the same time scale such that only sensor states are exchanged at each signal evolution epoch; in contrast, the M-GIKF scheme is a multi-time scale algorithm in which at each signal evolution epoch the agents cooperate through a single round communication of states exchange and the additional communication at a predefined rate $\overline{\gamma}$ (informally, $\overline{\gamma}$ denotes the average number of additional network communications per signal evolution epoch) to disseminate observations according to a randomized gossip protocol.

The GIKF proposed in the prior work [30] is a simpler version of M-GIKF without observation propagation; [30] shows that the error process is stochastically bounded and the network achieves weak consensus. The detailed characterization of this invariant measure was not established. In this thesis, we prove that the measure $\mu^{\overline{\gamma}}$ approaches the Dirac measure δ_{P^*} (P^* is the unique fixed point of the error covariance sequence in centralized Kalman filtering), and $\mu^{\overline{\gamma}}$ satisfies the Large Deviation (LD) upper and lower bounds. The LD property of $\mu^{\overline{\gamma}}$ implies that the probability of a rare event (the event of staying away from an arbitrary small neighborhood of P^*) decays exponentially; in other words, the convergence of $\mu^{\overline{\gamma}}$ to δ_{P^*} is exponentially fast in probability.

In contrast, the previous work in [31] only provides the Moderate Deviation prop-

erty of the random Riccati equation (RRE), where the RRE arises in Kalman filtering with intermittent observations, a problem discussed in [60], where the sensor observation packets, transmitted through an imperfect communication medium, are received at the estimator as a Bernoulli process with arrival probability $\gamma > 0$. In this case, the Moderate Derivation shows that the probability of a rare event decays as a power law of $(1-\gamma)$ for $\gamma \to 1$. Such setup and result are fundamentally different from those in this thesis, because Kalman filtering with intermittent observations discussed in [60] and [31] considers only the local algorithm at each sensor without inter-sensor communications.

By considering the limited sources for wireless sensor networks, such as bandwidth and power, which prevent the exchange of high-precision data (e.g., real-valued analog data) among the sensors, we further propose the quantized GIKF algorithm to involve the quantization scheme before transmitting data on the inter-sensor communication channel. The quantization procedure induces some noise to the swapped signal, such that the received state from the neighbor loses certain information. This makes the problem more challenging and different from the problem solved in Chapter 2 and 3, where we assume that the state of a sensor is perfectly transmitted to its neighbor. Then a natural question to ask is whether or not the estimation error variance sequence could still achieve weak convergence with the information loss due to quantization. To seek a positive answer, a Quantized Gossip Interactive Kalman Filtering (QGIKF) algorithm with a countable infinite quantization alphabet is first studied in the thesis, which is then extended to investigate a more restrictive and practical case with a finite quantization alphabet. In case of finite quantization alphabet, we propose a modified quantized GIKF (M-QGIKF) alphabet to solve the potential problem that quantization could saturate with finite quantization alphabet, and find the conditions under which M-QGIKF can still achieve weak consensus.

For the distributed quickest detection, different from the existing work, in this thesis we propose a distributed change detection algorithm based on a Bayesian setup of change happening time. To the best of our knowledge, this thesis is the first work discussing the distributed change detection algorithm under such a Bayesian setup. Additionally, in our proposed distributed algorithm, multiple communication steps are in between two observation instants, i.e., the communication step has a smaller time scale than that of the observation stage. In communication steps, a random point-to-point gossip based algorithm is proposed as in [37, 36]. We model the information propagation procedure governed by this communication procedure as a Makov process. We then analyze the performance of the proposed distributed change detection algorithm, with a method of two-layer large deviation analysis. Large deviation techniques [14, 8] have been used to analyze the performance of either centralized or distributed estimation and detection algorithms, for example, in [37, 3, 25, 55]. However, no existing work has utilized the technique of large deviation analysis to study the performance of the change detection algorithms, especially the distributed change detection algorithms. The most related work is [55], in which a distributed sequential detection method is proposed to solve the problem of Gaussian binary hypothesis testing. The sequential hypothesis testing problem could be considered as a special case of change detection problems, where the change happened at the initial time point [52].

The first-layer large deviation analysis shows that the relation between the conditional averaged detection delay and the probability of false alarm satisfies the large deviation principle, which implies that the probability of false alarm decays exponentially fast as the conditional averaged detection delay increases. In the first-layer analysis, the nonlinear renewal theorem is adopted, by representing the stopping time with the form of a random walk crossing a constant threshold plus a nonlinear

term. The second-layer analysis derives the large deviation upper and lower bounds for the probability of the rare event that not all observations are available at a sensor. Based on this, we further prove that the distributed Kullback-Leibler information number converges to the centralized Kullback-Leibler information number, by deriving the upper and lower bounds for the distributed form of Kullback-Leibler information numbers. We eventually show that the performance of the distributed algorithm converges exponentially fast to that of the centralized one when the averaged number of communications increases. In the analysis, the concept of hitting time in Markov chain is used to derive the large deviation upper and lower bounds.

Here I give the overview of contributions corresponding to addressing the following questions:

1. With lacking of detectability for the large scale system at a single node, how to design a distributed Kalman filtering strategy to achieve the reliable state estimate at each node for the entire system?

2. Can the estimation error covariance sequence in the proposed distributed Kalman filtering scheme achieve convergence?

3. How to quantify the rate at which the proposed distributed Kalman filtering scheme approaches the centralized performance as the inter-sensor communication rate increases?

4. With quantization involved, due to the information loss caused by quantization with infinite quantization alphabet, can the proposed distributed Kalman filtering scheme still achieve weak consensus?

5. By considering a more practical quantization method with finite quantization alphabet causing more information loss, how to design the algorithm and what's

the condition to maintain weak consensus?

6. In the distributed quickest detection scheme, what's the relation between the two performance metrics: probability of false alarm and averaged detection delay?

7. Under which condition, the distributed quickest detection scheme will approach the centralized optimal performance? What's the convergence rate?

This thesis is structured as follows. In Chapter 2, the system model is first setup, including the signal/observation model and the inter-sensor communication model. Then we propose the M-GIKF algorithm and a distributed observation dissemination protocol embedded in M-GIKF algorithm. After establishing the model for the M-GIKF, we study its conditional estimation error covariance properties. We show that the sensor network achieves weak consensus for each $\overline{\gamma} > 0$, i.e., the conditional estimation error covariance at a randomly selected sensor converges weakly (in distribution) to a unique invariant measure of an associated random Riccati equation. To prove this, we interpret the filtered state at each sensor, including state estimate and error covariance, as a stochastic particle and interpret the travelling process of filtered states among sensors as a Markov process. In particular, the sequence of travelling states or particles evolves according to a switched system of random Riccati operators, where the switching is dictated by a nonstationary Markov chain on the network graph. We formulate the corresponding random Riccati equation (RRE) as a Random Dynamical System (RDS) and establish the asymptotic distributional properties of the RRE sequence based on the properties of RDSs, where we show that the sequence of RREs converges weakly to an invariant measure.

In Chapter 3, the characterization of the converged invariant measure $\mu^{\overline{\gamma}}$ is studied with the large deviation analysis. We characterize such an invariant measure

denoted as $\mu^{\overline{\gamma}}$, which is the counterpart of the unique fixed point P^* of the error covariance sequence in centralized Kalman filtering [27]. As $\overline{\gamma} \to \infty$, we further prove that the measure $\mu^{\overline{\gamma}}$ approaches the Dirac measure δ_{P^*}, and $\mu^{\overline{\gamma}}$ satisfies the Large Deviation (LD) upper and lower bounds. The LD property of $\mu^{\overline{\gamma}}$ implies that the probability of a rare event (the event of staying away from an arbitrary small neighborhood of P^*) decays exponentially; in other words, the convergence of $\mu^{\overline{\gamma}}$ to δ_{P^*} is exponentially fast in probability. In Chapter 3, we first present the overview of large deviation principle with introducing some related definitions. Then, some preliminary results including the string theory and the Riccati equation are established to assist the large deviation analysis. Finally, we derive the upper and lower bounds of the large deviation and also present the numerical simulation results to calculate these bounds.

In Chapter 4, we first present the overview of the dithered quantization method and emphasize its property that the output is independent to the input when the Schuchman condition is satisfied. Based on this dithered quantization method, we propose the QGIKF scheme, with the state of a sensor first quantized into a infinite quantization alphabet before transmitting. With QGIKF algorithm, we further derive the corresponding estimation error variance, the format of which is quite different from that of M-GIKF algorithm. Then, an RDS formulation is established to represent the behavior of estimation error variance. Finally, we prove the weak consensus results by first giving some intermediate results including the stochastically boundedness as the perquisite conditions.

In Chapter 5, we extend the study from the infinite quantization alphabet to the finite quantization alphabet, which is a more practical and constrained case. Since the finite quantization alphabet induces more information loss and the quantization would saturate when the input data stays outside the alphabet range, we propose

the M-QGIKF algorithm to deal with these issues. We prove that under some mild conditions, the M-QGIKF algorithm can still achieve weak consensus. The simulation results further illustrate the consensus and also show the advantage of this distributed cooperation with quantization over the non-cooperation scheme.

In Chapter 6, a distributed Bayesian quickest change detection algorithm is studied. By modeling the information propagation dynamics in the network as a Markov process, two-layer large deviation analysis is presented to analyze the performance of the proposed algorithm. The first-layer analysis shows that the relation between the probability of false alarm and the conditional averaged detection delay satisfies the large deviation principle, implying that the probability of false alarm according to a rare event decays to zero at an exponentially fast rate when the conditional averaged detection decay increases. The second-layer analysis shows that the probability of the rare event that not all observations are available at a sensor decays to zero at an exponentially fast rate when the averaged number of communications increases, where the large deviation upper and lower bounds for this rate are also derived, based on which we show that the performance of the distributed algorithm converges exponentially fast to that of the centralized one.

2. DISTRIBUTED KALMAN FILTERING (M-GIKF) ALGORITHM AND WEAK CONSENSUS ANALYSIS*

In this chapter, we first setup the system model, including the signal/observation model and the communication model for the inter-sensor communications. Then, we propose the distributed Kalman Filtering (M-GIKF) algorithm, and describe the distributed observation dissemination protocol used in the M-GIKF algorithm. Finally, we present and prove the weak consensus results on the estimation error covariance over sensors.

2.1 System Model

In this section, we describe the system model, with the signal/observation and communication models.

2.1.1 Signal and Observation Model

Let $t \in \mathbb{R}_+$ denote continuous time and $\Delta > 0$ be a constant sampling interval and the global signal process $\{\mathbf{x}_{k\Delta}\}_{k\in\mathbb{N}}$ evolves as a sampled linear dynamical system:

$$\mathbf{x}_{(k+1)\Delta} = \mathcal{F}\mathbf{x}_{k\Delta} + \mathbf{w}_{k\Delta} \tag{2.1}$$

where $\mathbf{x}_{k\Delta} \in \mathbb{R}^M$ is the signal (state) vector with initial state $\mathbf{x}_0$ being distributed as a zero mean Gaussian vector with covariance $\widehat{P}_0$ and the system noise $\{\mathbf{w}_{k\Delta}\}$ is an uncorrelated zero mean Gaussian sequence independent of $\mathbf{x}_0$ with covariance $\mathcal{Q}$.

*Part of this chapter is reprinted, with permission, from [Di 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*, 61(3):1351–1372, Mar. 2015.]

The observation at the n-th sensor $\mathbf{y}^n_{k\Delta} \in \mathbb{R}^{m_n}$ at time $k\Delta$ is of the form:

$$\mathbf{y}^n_{k\Delta} = \mathcal{C}_n \mathbf{x}_{k\Delta} + \mathbf{v}^n_{k\Delta} \tag{2.2}$$

where $\mathcal{C}_n \in \mathbb{R}^{m_n \times M}$ and $\{\mathbf{v}^n_{k\Delta}\}$ is an uncorrelated zero mean Gaussian observation noise sequence with covariance $\mathcal{R}_n \gg \mathbf{0}$[1]. Also the noise sequences at different sensors are independent of each other, the system noise process and the initial system state. Because of the limited capability of the sensors, typically the dimension of $\mathbf{y}^n_k$ is much smaller than that of the signal process and the observation process at each sensor is not sufficient to make the pair $\{\mathcal{C}_n, \mathcal{F}\}$ observable[2]. We envision a totally distributed application where a reliable estimate of the entire signal process is required at each sensor. To achieve this, the sensors need collaboration via occasional communications with their neighbors, whereby they exchange their filtering states and observations. The details of the collaboration scheme will be defined precisely later.

We present the following weak assumptions on the signal/observation model:

Stabilizability: Assumption S.1 The pair $(\mathcal{F}, \mathcal{Q}^{1/2})$ is stabilizable. The non-degeneracy (positive definiteness) of $\mathcal{Q}$ guarantees this.

Weak Detectability: Assumption D.1 There exists a walk [3] $(n_1, \cdots, n_l)$ of length $l \geq 1$ covering the N nodes, such that the matrix $\sum_{i=1}^{l} (\mathcal{F}^{i-1})^T \mathcal{C}_{n_i}^T \mathcal{C}_{n_i} \mathcal{F}^{i-1}$ is invertible.

[1] The sampling interval Δ could be a function of various system parameters such as the sampling rate of the sensors and the rate of signal evolution. Thus the factor $1/\Delta$ may be viewed as the signal evolution time scale. Since Δ is fixed throughout the paper, we will drop Δ from the discrete index of sampled processes for notational convenience. Then, $\mathbf{x}_k$ will be used to denote $\mathbf{x}_{k\Delta}$ and the process $\{\mathbf{x}_{k\Delta}\}_{k\in\mathbb{N}}$ will be denoted by $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$.

[2] It is possible that some of the sensors have no observation capabilities, i.e., the corresponding C_n is a zero matrix. Thus this formulation easily carries over to networks of heterogeneous agents, which consist of 'sensors' actually sensing the field of interest and actuators implementing local control actions based on the estimated field.

[3] A walk is defined w.r.t. the graph induced by the non-zero entries of the matrix $\overline{A}$

Remark 2.1.1 Assumption (D.1) is minimal, since even in the centralized condition, where a center can access to all the sensor observations over all time, it requires the detectability for arbitrary choice of the matrix $\mathcal{F}$ governing the signal dynamics. This justifies the term of weak detectability.

2.1.2 Communication Model

Communication among sensors is constrained by several factors such as proximity, transmit power, and receiving capabilities. We model the underlying communication structure of the network in terms of an undirected graph $(V, \mathcal{E})$, where V denotes the set of N sensors and $\mathcal{E}$ is the set of edges or allowable communication links between the sensors. The notation $n \sim l$ indicates that sensors n and l can communicate, i.e., $\mathcal{E}$ contains the undirected edge (n, l). The graph can be represented by its $N \times N$ symmetric adjacency matrix $\mathcal{A}$:

$$\mathcal{A}_{nl} = \begin{cases} 1 & \text{if } (n,l) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

We assume that the diagonal elements of $\mathcal{A}$ are identically 1, indicating that a sensor n can always communicate to itself. Note that $\mathcal{E}$ is the maximal allowable set of allowable communication links in the network at any time, however, at a particular instant, each sensor may choose to communicate only to a fraction of its neighbors. The exact communication protocol is not so important for the analysis, as long as some weak connectivity assumptions are satisfied. For definiteness, we assume the following generic communication model, which subsumes the widely used gossiping protocol for real time embedded architectures ([6]) and the graph matching based communication protocols for internet architectures ([43].) Define the set $\mathcal{M}$

of symmetric 0-1 $N \times N$ matrices:

$$\mathcal{M} = \left\{ A \;\middle|\; \mathbf{1}^T A = \mathbf{1}^T, \;\; A\mathbf{1} = \mathbf{1}, \;\; A \leq \mathcal{E} \right\}, \tag{2.4}$$

where $A \leq \mathcal{E}$ is to be interpreted as component-wise. In other words, $\mathcal{M}$ is the set of adjacency matrices, where every node is incident to exactly one edge (including self edge) and allowable edges are only those included in $\mathcal{E}$.[4] Let $\mathcal{D}$ be a probability distribution on the space $\mathcal{M}$. The sequence of time-varying adjacency matrices $\{A(k)\}_{k\in\mathbb{N}}$, governing the inter-sensor communication, is then an i.i.d. sequence in $\mathcal{M}$ with distribution $\mathcal{D}$ and independent of the signal and observation processes.[5] We make the following assumption of connectivity on the average:

Assumption C.1: Define the symmetric stochastic matrix $\overline{A}$ as

$$\overline{A} = \mathbb{E}\left[A(k)\right] = \int_{\mathcal{M}} A d\mathcal{D}(A) \tag{2.5}$$

The matrix $\overline{A}$ is assumed to be irreducible and aperiodic.

Remark 2.1.2 The stochasticity of $\overline{A}$ is inherited from that of the elements in $\mathcal{M}$. Here we are not concerned with the properties of the distribution $\mathcal{D}$ as long as the *weak* connectivity assumption above is satisfied. The irreducibility of $\overline{A}$ depends both on the set of allowable edges $\mathcal{E}$ and the distribution $\mathcal{D}$. We do not detail this question here. However, to show the applicability of Assumption **C.1** and justify the notion of weak connectivity, we note that such a distribution $\mathcal{D}$ always exists if the graph $(V, \mathcal{E})$ is connected. We provide a Markov chain interpretation of the mean adjacency matrix $\overline{A}$, which is helpful for the following analysis. The matrix $\overline{A}$ can

[4]The set $\mathcal{M}$ is always non-empty, since the $N \times N$ identity matrix $I_N \in \mathcal{M}$.

[5]For convenience of presentation, we assume that $A(0) = I_N$, although communication starts at time slot $k = 1$.

be interpreted as the transition matrix of a time-homogeneous Markov chain on the state space V. Since the state space V is finite, the irreducibility of $\overline{A}$ implies the positive recurrence of the resulting Markov chain.

2.2 M-GIKF Algorithm

In this section, the M-GIKF algorithm is proposed. Before proposing the M-GIKF algorithm, we first summarize the GIKF algorithm. Lastly an example of distributed observation dissemination protocol in M-GIKF algorithm is presented in detail.

2.2.1 Overview of GIKF Algorithm

The GIKF (see [30]) assumes that inter-sensor communication rate is comparable to the signal evolution time scale and only one round of sensor communication is allowed for every epoch $[(k-1)\Delta, k\Delta)$.

We now present the algorithm GIKF (gossip based interacting Kalman filter) for distributed estimation of the signal process $\mathbf{x}_k$ over time. Let the filter at sensor n be initialized with the pair $\left(\widehat{\mathbf{x}}_{0|-1}, \widehat{P}_0\right)$, where $\widehat{\mathbf{x}}_{0|-1}$ denotes the prior estimate of $\mathbf{x}_0$ (with no observation information) and $\widehat{P}_0$ the corresponding error covariance. Also, $(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k)$ denotes the estimate at sensor n of $\mathbf{x}_k$ based on information[6] till time $k-1$ and the corresponding conditional error covariance, respectively. The pair $\left(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k\right)$ is also referred to as the state of sensor n at time $k-1$. To define the estimate update rule for the GIKF, denote by $n^{\rightarrow}_k$ the neighbor of sensor n at time k w.r.t. the adjacency matrix[7] $A(k)$. We assume that all inter-sensor communication for time k occurs at the beginning of the slot, whereby communicating sensors swap

[6]The information at sensor n till (and including) time k corresponds to the sequence of observations $\{\mathbf{y}^n_s\}_{0\le s\le k}$ obtained at the sensor and the information received by data exchange with its neighboring sensors.

[7]Note that by symmetry we have $(n^{\rightarrow}_k)^{\rightarrow}_k = n$. It is possible that $n^{\rightarrow}_k = n$, in which case $A(k)$ has a self-loop at node n.

their previous states, i.e., if at time k, $n_k^{\rightarrow} = l$, sensor n replaces its previous state $\left(\widehat{\mathbf{x}}_{k|k-1}^n, \widehat{P}_k^n\right)$ by $\left(\widehat{\mathbf{x}}_{k|k-1}^l, \widehat{P}_k^l\right)$ and sensor l replaces its previous state $\left(\widehat{\mathbf{x}}_{k|k-1}^l, \widehat{P}_k^l\right)$ by $\left(\widehat{\mathbf{x}}_{k|k-1}^n, \widehat{P}_k^n\right)$. The estimate is updated by sensor n at the end of the slot (after the communication and observation tasks have been completed) as follows:

$$\begin{aligned} \widehat{\mathbf{x}}_{k+1|k}^n &= \mathbb{E}\left[\mathbf{x}_{k+1} \;\middle|\; \widehat{\mathbf{x}}_{k|k-1}^{n_k^{\rightarrow}}, \widehat{P}_k^{n_k^{\rightarrow}}, \mathbf{y}_k^n\right] && (2.6) \\ \widehat{P}_{k+1}^n &= \mathbb{E}\Big[\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k}^n\right)\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k}^n\right)^T \\ &\qquad \Big| \; \widehat{\mathbf{x}}_{k|k-1}^{n_k^{\rightarrow}}, \widehat{P}_k^{n_k^{\rightarrow}}, \mathbf{y}_k^n\Big]. && (2.7) \end{aligned}$$

2.2.2 Proposed M-GIKF Algorithm

We start by introducing some notation. Let $\mathfrak{P}$ denote the power set of $[1, \cdots, N]$. The elements of $\mathfrak{P}$ are indexed by $\jmath \in [0, \cdots, 2^N - 1]$, with 0 denoting the null set and $2^N - 1$ the entire set. Also, for technical convenience, we will interpret the elements (sensors) in a subset $\jmath$ to be arranged in ascending order, i_1 denoting the first and $i_{|\jmath|}$ denoting the last. For each sensor n, we denote by $\{\mathcal{I}_k^n\}$, the subset valued process taking values in $\mathfrak{P}$. For a given $\jmath \in \mathfrak{P}$, by $\mathbf{y}_k^{\jmath}$ we denote the subset $[(\mathbf{y}_k^{i_1})^T \cdots (\mathbf{y}_k^{i_{|\jmath|}})^T]^T$ of observations at the k-th epoch, whereas, the matrix $\mathcal{C}_{\jmath}$ stands for the matrix $[\mathcal{C}_{i_1}^T \cdots \mathcal{C}_{i_{|\jmath|}}^T]^T$, and the matrix $\mathcal{R}_{\jmath} = \mathrm{diag}[\mathcal{R}_{i_1}, \cdots, \mathcal{R}_{i_{|\jmath|}}]$. In particular, at the k-th epoch, $\mathbf{y}_k^{\mathcal{I}_k^n}$ denotes the subset of observations available at sensor n at the end of $(k\Delta, (k+1)\Delta]$.

Suppose, in the basic GIKF scheme explained above, there is an additional step of communication. Specifically, assume that in every interval $[k\Delta, (k+1)\Delta)$ the network (as a whole) is given an opportunity for additional communication at rate $\overline{\gamma}$, i.e., additional $\overline{\gamma}$ message exchanges occur across the network in each epoch. In particular, we assume that the total number of additional sensor transmissions in

$[k\Delta, (k+1)\Delta)$ is dominated by a Poisson random variable of rate $\overline{\gamma}$[8], and that each transmission conforms to the network topology induced by the maximal adjacency matrix $\mathcal{A}$. Clearly, by exploiting this additional inter-sensor communication, the network should be able to perform a filtering task that is at least as good if not better than the basic GIKF.

A natural way to improve the performance of the GIKF is to use this additional communication to disseminate the observations across the sensors. We denote this new scheme with additional communication for disseminating the observations by Modified GIKF (M-GIKF). For each sensor n, the subset-valued process $\{\mathcal{I}_k^{\overline{\gamma},n}\}$ taking values in $\mathfrak{P}$ is used to index the subset of observations $\mathbf{y}_k^{\mathcal{I}_k^{\overline{\gamma},n}}$ available at sensor n at the end of the interval $[k\Delta, (k+1)\Delta)$, e.g., if $\{\mathcal{I}_k^{\overline{\gamma},n}\} = [m, n]$, then the observations $\mathbf{y}_k^m$ and $\mathbf{y}_k^n$ are available at sensor n by the end of the interval $[k\Delta, (k+1)\Delta)$. Also, the corresponding parameters with $\mathbf{y}_k^{\mathcal{I}_k^{\overline{\gamma},n}}$ in the observation model (2.2) are denoted by $\mathcal{C}_{\mathcal{I}_k^{\overline{\gamma},n}}$ and $\mathcal{R}_{\mathcal{I}_k^{\overline{\gamma},n}}$.

For the GIKF algorithm it is clear that

$$\mathcal{I}_k^{0,n} = \{n\},\ \forall n \in [1, \cdots, N],\ k \in \mathbb{N},$$

i.e., each sensor only has access to its own observations in each epoch. Hence, in the GIKF the only cooperation among the sensors is achieved through estimate exchanging and no explicit mixing or aggregation of instantaneous observations occur. This is in fact the key difference between the GIKF and the M-GIKF. In the M-GIKF, the sensors use the additional communication rate $\overline{\gamma}$ to exchange instantaneous observations, in addition to performing the basic estimate swapping of the GIKF.

In this work, our main focus is not on the exact nature of the instantaneous

[8]The Poisson assumption is claimed and justified at the end of this subsection.

observation dissemination protocol, as long as it is distributed (i.e., any inter-sensor exchange conforms to the network topology) and satisfies some assumptions (in Section 2.2.3 we will provide an example of such distributed protocols satisfying these assumptions). Recall $\mathcal{I}_k^{\overline{\gamma},n}$ to be the instantaneous observation set available at sensor n by the end of the interval $[k\Delta,(k+1)\Delta)$. Note, the statistics of the process $\{\mathcal{I}_k^{\overline{\gamma},n}\}$ depend on the dissemination protocol used and the operating rate $\overline{\gamma}$. Before providing details of the dissemination protocol and the assumptions on the processes $\{\mathcal{I}_k^{\overline{\gamma},n}\}$, for all n, we explain the M-GIKF scheme as follows. For the moment, the reader may assume that $\{\mathcal{I}_k^{\overline{\gamma},n}\}$ are generic set-valued processes taking values in $\mathfrak{P}$ and there exists a distributed protocol operating in the time window $[k\Delta,(k+1)\Delta)$ leading to such observation sets at the sensors by the end of the epoch. Clearly, for any protocol and $\overline{\gamma}\geq 0$[9],

$$\{n\}\subset\mathcal{I}_k^{\overline{\gamma},n},\ \forall n\in[1,\cdots,N],\ k\in\mathbb{N}.$$

Moreover, if the observation dissemination protocol is reasonable, $\mathcal{I}_k^n$ is strictly greater than $\{n\}$ with positive probability. The basic difference between the GIKF and the M-GIKF is that, in (2.6)-(2.7), instead of conditioning on $\mathbf{y}_k^n$ at sensor n, we condition on the possibly larger set $\mathbf{y}_k^{\mathcal{I}_k^n}$ of observations available at sensor n.

With this setup, now, we formally describe the M-GIKF, which generalizes the GIKF when additional inter-sensor communication at rate $\overline{\gamma}$ is allowed in every epoch $[k\Delta,(k+1)\Delta)$.

Algorithm M-GIKF: We assume that $\overline{\gamma}>0$ is given and fixed. Let the filter at sensor n be initialized with the pair $\left(\widehat{\mathbf{x}}_{0|-1},\widehat{P}_0\right)$, where $\widehat{\mathbf{x}}_{0|-1}$ denotes the prior estimate of $\mathbf{x}_0$ (with no observation information) and $\widehat{P}_0$ the corresponding error

[9]For conciseness, we will drop the superscript $\overline{\gamma}$ over the notations related to the M-GIKF with the additional communication rate $\overline{\gamma}$.

covariance. Also, by $\left(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k\right)$ we denote the estimate at sensor n of $\mathbf{x}_k$ based on information till time $k-1$ and the corresponding conditional error covariance, respectively. The pair $\left(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k\right)$ is also referred to as the state of sensor n at time $k-1$. Similar to GIKF, the M-GIKF update involves the state exchanging step (w.r.t. the adjacency matrices $\{A(k)\}$), whereby, at the beginning of the epoch $[k\Delta, (k+1)\Delta)$, sensor n exchanges its state with its neighbor $n^{\rightarrow}_k$ w.r.t. $A(k)$. This exchange is performed only once in the interval $[k\Delta, (k+1)\Delta)$. Then each sensor in M-GIKF makes its sensing observation and M-GIKF instantiates the distributed dissemination protocol before the end of the epoch. This leads observation aggregation with $\mathbf{y}_k^{\mathcal{I}^n_k}$ being the observation set available at sensor n at the end of the interval $[k\Delta, (k+1)\Delta)$. The estimate update at sensor n at the end of the slot (after the communication and observation dissemination tasks have been completed) is

$$\begin{aligned}
\widehat{\mathbf{x}}^n_{k+1|k} &= \mathbb{E}\left[\mathbf{x}_{k+1} \;\middle|\; \widehat{\mathbf{x}}^{n^{\rightarrow}_k}_{k|k-1}, \widehat{P}^{n^{\rightarrow}_k}_k, \mathbf{y}^{\mathcal{I}^n_k}_k, \mathcal{I}^n_k\right] \\
\widehat{P}^n_{k+1} &= \mathbb{E}\left[\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}^n_{k+1|k}\right)\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}^n_{k+1|k}\right)^T \right. \\
&\qquad \left. \;\middle|\; \widehat{\mathbf{x}}^{n^{\rightarrow}_k}_{k|k-1}, \widehat{P}^{n^{\rightarrow}_k}_k, \mathbf{y}^{\mathcal{I}^n_k}_k, \mathcal{I}^n_k\right].
\end{aligned}$$

Due to conditional Gaussianity, the optimal prediction steps can be implemented through the time-varying Kalman filter recursions, and it follows that the sequence $\left\{\widehat{P}^n_k\right\}$ of the conditional predicted error covariance matrices at sensor n satisfies the Riccati recursion:

$$\begin{aligned}
\widehat{P}^n_{k+1} &= \mathcal{F}\widehat{P}^{n^{\rightarrow}_k}_k \mathcal{F}^T + \mathcal{Q} - \mathcal{F}\widehat{P}^{n^{\rightarrow}_k}_k \mathcal{C}^T_{\mathcal{I}^n_k} \\
&\times \left(\mathcal{C}_{\mathcal{I}^n_k}\widehat{P}^{n^{\rightarrow}_k}_k \mathcal{C}^T_{\mathcal{I}^n_k} + \mathcal{R}_{\mathcal{I}^n_k}\right)^{-1} \mathcal{C}_{\mathcal{I}^n_k}\widehat{P}^{n^{\rightarrow}_k}_k \mathcal{F}^T. \qquad (2.8)
\end{aligned}$$

Remark 2.2.1 Note that the sequence $\left\{\widehat{P}^n_k\right\}$ is random, due to the random neigh-

borhood selection function $n_k^{\rightarrow}$. The goal of the paper is to study the asymptotic properties of the sequence of random conditional error covariance matrices $\left\{\widehat{P}_k^n\right\}$ at each sensor n and to show in what sense they reach consensus, such that, in the limit of large time, every sensor provides an equally good (stable in the sense of estimation error) estimate of the signal process.

2.2.3 Distributed Observation Dissemination Protocol in M-GIKF

We first introduce the following assumptions on the communication medium and the distributed information dissemination protocol generating the subsets $\{\mathcal{I}_k^n\}$ for all n, k.

(i) **(E.1)**: The total number of inter-sensor observation dissemination messages $\mathcal{M}(k)$ in the interval $[k\Delta, (k+1)\Delta)$, for all $k \in \mathbb{T}_+$ follows a Poisson distribution with mean $\overline{\gamma}$.

(ii) **(E.2)**: For each n, the process $\{\mathcal{I}_k^n\}$ is (conditionally) i.i.d. For each k, the protocol initiates at the beginning of the interval $[k\Delta, (k+1)\Delta)$ and operates on the most recent observations $\{\mathbf{y}_k^n\}_{1\leq n\leq N}$. The protocol terminates at the end of the epoch. For observation dissemination in the next epoch $[(k+1)\Delta, (k+2)\Delta)$, the protocol is re-initiated and acts on the new observation data $\{\mathbf{y}_{k+1}^n\}_{1\leq n\leq N}$, independent of its status in the previous epoch. Necessarily, the sequence is (conditionally) i.i.d.. We define

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{M}(i) = \overline{\gamma}, \text{ a.s.}, \tag{2.9}$$

i.e., the average number of dissemination messages per epoch is $\overline{\gamma}$.

(iii) **(E.3)**: Recall the notations $\jmath$ and $\{i_1, \cdots, i_{|\mathfrak{P}_\jmath|}\}$ at the beginning of this section.

For each $\jmath \in [0, \cdots, 2^N - 1]$, define

$$\mathbb{P}\left(\mathcal{I}_k^n = \{i_1, \cdots, i_{|\mathfrak{P}_\jmath|}\}\right) = q_n(\jmath),\ \forall n, k. \tag{2.10}$$

We assume that for all $\overline{\gamma} > 0$

$$\mathbb{P}\left(\mathcal{I}_k^n = \{1, 2, \cdots, N\}\right) = q_n(2^N - 1) > 0,\ \forall n, k. \tag{2.11}$$

(iv) **(E.4)**: For each $\jmath \neq 2^N - 1$, define

$$\begin{aligned} -\underline{q}_n(\jmath) &\leq \liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln\left(q_n(\jmath)\right) \\ &\leq \limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln\left(q_n(\jmath)\right) \leq -\overline{q}_n(\jmath). \end{aligned} \tag{2.12}$$

We assume that, for $\jmath \neq 2^N - 1$, $\overline{q}_n(\jmath) > 0,\ \forall n$. Since $\{n\} \subset \mathcal{I}_k^n$ for all n, necessarily for all $\jmath$, such that $n \notin \{i_1, \cdots, i_{|\mathfrak{P}_\jmath|}\}$, $\overline{q}_n(\jmath) = \infty$.

Remark 2.2.2 We now comment on the assumptions and justify their applicability under reasonable conditions (an example of distributed observation dissemination protocol with rate constraints is provided in the sequel):

(i) Assumption **(E.1)** essentially means that the waiting times between successive transmissions are i.i.d. exponential random variables with mean $1/\overline{\gamma}$. This is justified in Carrier Sense Multiple Access (CSMA) type protocols, where the back-off time is often chosen to be exponentially distributed. To be more realistic, one needs to account for packet delays and transmission/reception processing times. We ignore these in the current setting. On a more practical note, the rate $\overline{\gamma}$ may be viewed as a function of the network communication

bandwidth; the larger the bandwidth, the higher the rate of channel usages and hence $\overline{\gamma}$. In distributed network communication settings, a typical example of exponential waiting between successive transmissions is the asynchronous gossip model (see [6]).

(ii) Assumption **(E.2)** is justified for memoryless and time-invariant communication schemes. It says that the scope of an instantiation of the distributed observation dissemination protocol is confined to the interval $[k\Delta, (k+1)\Delta)$, at the end of which the protocol restarts with a new set of observations independent of its past status. Equation (2.9) is then a direct consequence of the Strong Law of Large Numbers (SLLN). This essentially means that the observation dissemination rate is $\overline{\gamma}$ times the observation acquisition or sampling rate scale.

(iii) Assumption **(E.3)** is satisfied by any reasonable distributed protocol if the network is connected. Intuitively, this is due to the fact that, if $\overline{\gamma} > 0$, the probability of having a sufficiently large number of communications in an interval of length $\Delta > 0$ is strictly greater than zero (which can be very small though, depending on the value of $\overline{\gamma}$). On the other hand, if the network is connected, it is possible by using a sufficiently large (but finite) number of communications to disseminate the observation of a sensor to every other sensor. An example of protocol satisfying **(E.3)** is provided in the sequel.

(iv) Assumption **(E.4)** is justifiable by showing that $q_n(\jmath)$ decays exponentially as $\overline{\gamma} \to \infty$. An example of protocol satisfying **(E.4)** is provided in the the sequel.

Remark 2.2.3 We claim that if random link failures are further considered in the protocol, the M-GIKF algorithm and the corresponding convergence result could still

hold with minimum modification, since link failures basically lead to no information swapping or propagation between some particular node pairs. This results in the same effect as the case where the sensors choose to communicate with themselves in our current protocol. Apparently, with random link failures, to achieve the same error performance, it would require more signal evolution epochs compared with the case without link failures.

An Example of Practical Protocol: In the following, we give an example of a gossip based distributed observation dissemination protocol. During the epoch $[k\Delta, (k+1)\Delta)$, the protocol initiates the observation dissemination at sensor n. Sensor n starts with its own current observation $\mathbf{y}_k^n$ and keeps exchanging its observation with its neighbors till the end of this epoch. The number of exchanges and the type of each exchange are determined by an asynchronous pairwise gossip protocol [6], where the inter-sensor communication occurs at successive ticks of a Poisson process with rate $\overline{\gamma}/\Delta_o$, and at each tick only one of the network links is active with uniform probability $1/\overline{M}$, where Δ_o is the time duration allocated for observation dissemination with each epoch, and $\overline{M}$ is the cardinality of the allowable communication link set $\mathcal{E}$. Equivalently, we could consider each network link activated independently of the others according to the ticks of a local Poisson clock with rate $\overline{\gamma}/\Delta\overline{M}$, where no two links will become active simultaneously due to the independence of events in the local Poisson processes. As a formal statement, the number of inter-sensor communications for observation dissemination $\mathcal{M}(k)$ in the interval $[k\Delta, (k+1)\Delta)$ follows a Poisson distribution with mean value $\overline{\gamma}$, which proves that this protocol satisfies assumption **(E.1)**. In addition, the corresponding sequence of time-varying adjacency matrices $\{A_k^o(i)\}_{i=1,\cdots,\mathcal{M}(k)}$ is an i.i.d. sequence uniformly distributed on the set $\{E^{nl}\}$, where E^{nl} is defined as a permutation matrix, such that, for each

$(n,l) \in \mathcal{E}$ and $n \neq l$, $E^{nl}_{n,l} = E^{nl}_{l,n} = 1$ and $E^{nl}_{m,m} = 1$ for $m \neq n, l$, with all other entries being 0.

Now, we establish the observation dissemination process. Let $\mathbf{s}^i_k = [s^i_k(1), \cdots, s^i_k(N)]$ with its entry $s^i_k(n) \in [1, \cdots, N]$ indexing the observation $\mathbf{y}^{s^i_k(n)}_k$ at sensor n just after the i-th exchange in the epoch $[k\Delta, (k+1)\Delta)$. Starting with $s^0_k(n) = n$ for each n means that, at the beginning of the epoch $[k\Delta, (k+1)\Delta)$ before any exchanges, each sensor n only has its own observation $\mathbf{y}^n_k$. When exchanges happen, the observations $\{\mathbf{y}^n_k\}_{1 \leq n \leq N}$ travel across the network according to

$$\mathbf{s}^i_k = A^o_k(i)\mathbf{s}^{i-1}_k, \quad i \in [1, \cdots, \mathcal{M}(k)]. \tag{2.13}$$

During this exchange process until the end of the epoch $[k\Delta, (k+1)\Delta)$, the sensors store the observations passing through them. Therefore, at the end of the epoch $[k\Delta, (k+1)\Delta)$, the set of observations available at sensor n is

$$\mathcal{I}^n_k = \bigcup_{i=0}^{\mathcal{M}(k)} \{s^i_k(n)\}. \tag{2.14}$$

Finally, the observation dissemination for the epoch $[k\Delta, (k+1)\Delta)$ terminates at the end of this epoch, right before the sensor starts the next epoch $[(k+1)\Delta, (k+2)\Delta)$. Then similarly the observation dissemination repeats during the epoch $[(k+1)\Delta, (k+2)\Delta)$ independent of its prior state. Therefore, the sequence $\{\mathcal{I}^n_k\}$ as the set of observation indices available at sensor n at the end of each epoch is a temporally i.i.d. process, which satisfies assumption **(E.2)**. Moreover, this observation dissemination process is assumed to be independent of the estimate exchange process.

Remark 2.2.4 It is readily seen that the above observation dissemination protocol conforms to the preassigned gossip network structure. In fact, to execute the above

protocol, each sensor needs to know its local communication neighborhood only, no global topology information is required. Secondly, note that, at each communication, a sensor forwards a single observation $\mathbf{y}_k^{s_k^i(n)}$ to a neighboring sensor. Since, the sensor observations are typically low-dimensional, the data overhead of each communication is modest. Finally, since the above protocol is fully randomized (neighbors are chosen independently uniformly), it is likely that a sensor will receive multiple copies of the same observation (possibly through different neighbors), i.e., some communications might end up being redundant.

To prove that this protocol satisfies assumptions **(E.3)** and **(E.4)**, we have the following analysis employing the hitting time concept of Markov chains. For each $\jmath \neq 2^N - 1$, without loss of generality, we assume that $\jmath$ corresponds to the sensor subset $\{n_1, n_2, ..., n_m\}$, with $\{n'_1, n'_2, ..., n'_{N-m}\}$ denoting the complementary subset. As explained by the interacting particle representation in the next section, the link formation process following the sequence $\{A_k^o(i)\}$ for the observation dissemination can be represented as N particles moving on the graph as identical Markov chains. We use T_i to denote the hitting time starting from sensor i to another sensor n in the Markov chain, with the transition probability matrix as the mean adjacency matrix $\overline{A^o}$, which is irreducible and defined in a similar way as (2.5). Then, we have

$$\begin{aligned} q_n(\jmath) &= P\Big(T_{n'_1} > \mathcal{M}(k), \cdots, T_{n'_{N-m}} > \mathcal{M}(k), \\ &\qquad T_{n_1} \leq \mathcal{M}(k), \cdots, T_{n_m} \leq \mathcal{M}(k)) \\ &\leq P\left(T_{n'_1} > \mathcal{M}(k), \cdots, T_{n'_{N-m}} > \mathcal{M}(k)\right) \\ &\leq \min_{1 \leq i \leq N-m} P\left(T_{n'_i} > \mathcal{M}(k)\right). \end{aligned} \tag{2.15}$$

From Theorem 7.26 in [19], since the transition matrix $\overline{A^o}$ is irreducible, there exist constants $0 < \alpha < 1$ and $0 < L < \infty$ such that $P(T_i > L) \leq \alpha, \forall i$, and more generally,

$$P(T_i > kL) \leq \alpha^k, \ k = 0, 1, 2, \cdots . \tag{2.16}$$

Also, there exists constant $0 < \beta < 1$ such that $P(T_i > L) \geq \beta, \forall i$, and more generally,

$$P(T_i > kL) \geq \beta^k, \ k = 0, 1, 2, \cdots . \tag{2.17}$$

Then, following (2.15), we have

$$\begin{aligned}
&\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln\left(q_n(\jmath)\right) \\
&\quad \leq \limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln\left(\min_{1\leq i\leq N-m} P(T_{n_i'} > \mathcal{M}(k))\right) \\
&\quad \leq \limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln\left(\alpha^{\lfloor \frac{\mathcal{M}(k)}{L} \rfloor}\right) = \frac{\ln\, \alpha}{L}
\end{aligned} \tag{2.18}$$

where the last equation is obtained since $\lim_{\overline{\gamma}\to\infty} \frac{\mathcal{M}(k)}{\overline{\gamma}} = 1$.

We also have

$$\begin{aligned}
q_n(\jmath) = P\Big(&T_{n_1'} > \mathcal{M}(k), \cdots, T_{n_{N-m}'} > \mathcal{M}(k), \\
&T_{n_1} \leq \mathcal{M}(k), \cdots, T_{n_m} \leq \mathcal{M}(k)) \\
\geq\ & P\left(T_{n_1'} > \mathcal{M}(k)\right) \cdots P\left(T_{n_{N-m}'} > \mathcal{M}(k)\right) \\
& P\left(T_{n_1} \leq \mathcal{M}(k)\right) \cdots P\left(T_{n_m} \leq \mathcal{M}(k)\right).
\end{aligned} \tag{2.19}$$

Then, from (6.52) and (6.53), we have

$$\begin{aligned}
&\liminf_{\overline{\gamma}\to\infty}\frac{1}{\overline{\gamma}}\ln\left(q_n(\jmath)\right)\\
&\quad\geq \liminf_{\overline{\gamma}\to\infty}\frac{1}{\overline{\gamma}}\ln\left[\left(\beta^{\lceil\frac{\mathcal{M}(k)}{L}\rceil}\right)^{N-m}\left(1-\alpha^{\lfloor\frac{\mathcal{M}(k)}{L}\rfloor}\right)^{m}\right]\\
&\quad= (N-m)\frac{\ln\beta}{L}
\end{aligned}\tag{2.20}$$

where the last equation is obtained since $\lim_{\overline{\gamma}\to\infty}\frac{\mathcal{M}(k)}{\overline{\gamma}}=1$ and $0<\alpha<1$.

Therefore, from (6.56) and (6.58), we have $\underline{q}_n(\jmath)$ and $\overline{q}_n(\jmath)$ in (2.12) well defined as

$$\underline{q}_n(\jmath)=(m-N)\frac{\ln\beta}{L},\quad \overline{q}_n(\jmath)=-\frac{\ln\,\alpha}{L}.\tag{2.21}$$

Since $\overline{q}_n(\jmath)=-\frac{\ln\alpha}{L}$ and $\alpha<1$, clearly we see that, for $\jmath\neq 2^N-1$, $\overline{q}_n(\jmath)>0$. Therefore, we have completed the proof that assumption **(E.4)** holds.

To establish assumption **(E.3)**, we denote $T_m=\max\{T_1,...,T_N\}$, i.e., T_m is the longest time among all hitting times to sensor n from other sensors. Then, $q_n(2^N-1)=P(T_m\leq\mathcal{M}(k))=1-P(T_m>\mathcal{M}(k))$, which is greater than zero according to (6.52). This access to all the observations at the end of an epoch may be arbitrarily small but strictly greater than zero.

2.3 Weak Consensus Analysis for M-GIKF Algorithm

In this section, we study the weak consensus over the network implementing M-GIKF algorithm. To this end, an interacting particular representation strategy is first proposed. Then, we formulate the estimation error covariance sequences as a random dynamical system (RDS). By adopting the properties in the RDS, we prove the weak consensus result.

2.3.1 Interacting Particular Representation

To simplify the notation in (2.8), we define the functions of $f_\jmath : \mathbb{S}_+^M \longmapsto \mathbb{S}_+^M$ for $\jmath \in [0, \cdots, 2^N - 1]$ denoting the respective subset Riccati operators[10]:

$$f_\jmath(X) = \mathcal{F}X\mathcal{F}^T + \mathcal{Q} - \mathcal{F}X\mathcal{C}_\jmath^T \left(\mathcal{C}_\jmath X \mathcal{C}_\jmath^T + \mathcal{R}_\jmath\right)^{-1} \mathcal{C}_\jmath X \mathcal{F}^T. \tag{2.22}$$

Recall the sequence $n_k^{\rightarrow}$ of neighbors of sensor n. The sequence of conditional error covariance matrices $\{P_k^n\}$ at sensor n then evolves according to

$$\widehat{P}_{k+1}^n = f_{\jmath\left(\mathcal{I}_k^n\right)} \left(\widehat{P}_k^{n_k^{\rightarrow}}\right) \tag{2.23}$$

where $\jmath\left(\mathcal{I}_k^n\right)$ denotes the index of $\mathcal{I}_k^n$ in the set $\mathfrak{P}$. The above sequence $\left\{\widehat{P}_k^n\right\}$ is non-Markovian (and is not even semi-Markov given the random adjacency matrix sequence $\{A(k)\}$), as $\widehat{P}_k^n$ at time k is a random functional of the conditional error covariance of sensor $n_k^{\rightarrow}$ at time $k-1$, which, in general, is different from that of sensor n. This makes the evolution of the sequence $\left\{\widehat{P}_k^n\right\}$ difficult to track. To overcome this, we give the following interacting particle interpretation of the conditional error covariance evolution, from which we can completely characterize the evolution of the desired covariance sequences $\left\{\widehat{P}_k^n\right\}$ for $n = 1, \cdots, N$.

To this end, we note that the link formation process given by the sequence $\{A(k)\}$ can be represented by N particles moving on the graph as identical Markov chains. The state of the n-th particle is denoted by $z_n(k)$, and the sequence $\{z_n(k)\}_{k\in\mathbb{N}}$ takes values in $[1, \cdots, N]$. The evolution of the n-th particle is given as follows:

$$z_n(k) = z_n(k-1)_k^{\rightarrow},\ z_n(0) = n. \tag{2.24}$$

[10]For $\jmath = 0$, the corresponding Riccati operator f_0 in (2.22) reduces to the Lyapunov operator, see [30].

Recall the (random) neighborhood selection $n_k^{\rightarrow}$. Thus, the n-th particle can be viewed as originating from node n at time 0 and then traveling on the graph (possibly changing its location at each time) according to the link formation process $\{A(k)\}$. The following proposition establishes important statistical properties of the sequence $\{z_n(k)\}$:

Proposition 2.3.1

(i) For each n, the process $\{z_n(k)\}$ is a Markov chain on $V = [1, \cdots, N]$ with the transition probability matrix $\overline{A}$.

(ii) The Markov chain $\{z_n(k)\}$ is ergodic with the uniform distribution on V being the attracting invariant measure.

For each of the Markov chains $\{z_n(k)\}$, we define a sequence of switched Riccati iterates $\{P_n(k)\}$:

$$P_n(k+1) = f_{\jmath(\mathcal{I}^k_{z_n(k)})}(P_n(k)). \tag{2.25}$$

The sequence $\{P_n(k)\}$ can be viewed as an iterated system of Riccati maps, in which the random switching sequence is governed by the Markov chain $\{z_n(k)\}$. A more intuitive explanation comes from the particle interpretation; precisely the n-th sequence may be viewed as a particle originating at node n and hopping around the network as a Markov chain with transition probability $\overline{A}$ whose instantaneous state $P_n(k)$ evolves via the Riccati operator at its current location. In particular, in contrast to the sequence $\left\{\widehat{P}_k^n\right\}$ of the original conditional error covariances at sensor n, the sequence $\{P_n(k)\}$ does not correspond to the evolution of the error covariance at a particular sensor. The following proposition establishes the relation between $\{P_n(k)\}$ and the sequence $\left\{\widehat{P}_k^n\right\}$ of interest.

Proposition 2.3.2 Consider the sequence of random permutations $\{\pi_k\}$ on V, given by

$$(\pi_{k+1}(1),\cdots,\pi_{k+1}(N))=(\pi_k(1)_k^{\rightarrow},\cdots,\pi_k(N)_k^{\rightarrow}) \tag{2.26}$$

with initial condition

$$(\pi_0(1),\cdots,\pi_0(N))=(1,\cdots,N)\,. \tag{2.27}$$

Note that $\pi_k(n) = z_n(k)$ for every n, where $z_n(k)$ is defined in (2.24). Then, for $k \in \mathbb{N}$,

$$(P_1(k),\cdots,P_N(k))=\left(\widehat{P}_k^{\pi_k(1)},\cdots,\widehat{P}_k^{\pi_k(N)}\right). \tag{2.28}$$

The above proposition suggests that the asymptotics of the desired sequence $\left\{\widehat{P}_k^n\right\}$ for every n can be obtained by studying the asymptotics for the sequences $\{P_n(k)\}$. Hence, in the subsequent sections, we will focus on $\{P_n(k)\}$, rather than working directly with the sequences $\left\{\widehat{P}_k^n\right\}$ of interest, which involve a much more complicated statistical dependence.

2.3.2 An Auxiliary Sequence

Since the switching Markov chains $\{z_n(k)\}$ are non-stationary, in order to analyze the processes $\{P_n(k)\}$ for $n = 1, ..., N$ under the scope of iterated random systems [16] or RDSs [1], we propose an auxiliary process $\left\{\widetilde{P}(k)\right\}$ evolving with similar random Riccati iterates, but for which the corresponding switching Markov chain $\{\widetilde{z}(k)\}$ is stationary, i.e., $\{\widetilde{z}(k)\}$ is initialized by the uniform invariant measure on V. Then, we can analyze the asymptotic properties of the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$ by formulating it as an RDS on the space $\mathbb{S}_+^N$ and derive the asymptotics of the sequence $\{P_n(k)\}$ for $n = 1, ..., N$. The auxiliary sequence $\left\{\widetilde{P}(k)\right\}$ is formally defined as follows, which follows the concept proposed in [30], but with necessary and non-

trivial modifications to take into account observation dissemination.

Consider a Markov chain $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$ on the graph V, with transition matrix $\overline{A}$ and uniform initial distribution as follows:

$$\mathbb{P}[\widetilde{z}(0)=n]=\frac{1}{N},\quad n=1,...,N. \tag{2.29}$$

By proposition 2.3.1, the Markov chain $\{\widetilde{z}(k)\}$ is stationary.

Now we can define the auxiliary process $\left\{\widetilde{P}(k)\right\}$ with similar random Riccati iterates as

$$\widetilde{P}(k+1)=f_{\jmath(\mathcal{I}^k_{\widetilde{z}(k)})}\left(\widetilde{P}(k)\right) \tag{2.30}$$

with (possibly random) initial condition $\widetilde{P}(0)$[11].

2.3.3 RDS Formulation

In order to proceed with the asymptotic analysis of the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$, we construct an RDS (θ,φ) on $\mathbb{S}^N_+$, equivalent to the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$ in the sense of distribution. To achieve this, we construct the Markov chain $\{\widetilde{z}(k)\}$ on a canonical path space. Let $\widetilde{\Omega}$ denote the set $\{1,...,N\}$ with $\widetilde{\mathcal{F}}$ as the corresponding Borel algebra on $\widetilde{\Omega}$, thus $\widetilde{\mathcal{F}}$ is the power set of $\{1,...,N\}$. Denote $\Omega^R=\otimes_{k=-\infty}^{\infty}\widetilde{\Omega}$, which is the two-sided infinite product of sets $\widetilde{\Omega}$, i.e., Ω^R is the space of two-sided sequences of entries in $\{1,...,N\}$,

$$\Omega^R=\{w=(\cdots,w_{-1},w_0,w_1,\cdots)|w_t\in\{1,\cdots,N\},\quad\forall t\in\mathbb{T}\}. \tag{2.31}$$

[11]Note that the sequences $\{P_n(k)\}$ of interest have deterministic initial conditions, but it is required for technical reasons to allow random initial states $\widetilde{P}(0)$ to study the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$.

Equip Ω^R with the corresponding product Borel algebra $\mathcal{F}^R = \otimes_{k=-\infty}^{\infty}\widetilde{\mathcal{F}}$. Note that $\{w_k\}_{k\in\mathbb{T}_+}$ for all $w \in \Omega^R$ denotes the canonical path space of the Markov chain $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$. Consider the unique probability measure $\mathbb{P}^R$ on $\mathcal{F}^R$, under which the two-sided stochastic process $\{w_k\}_{k\in\mathbb{T}}$ is a stationary Markov chain on the finite state space $\{1,\cdots,N\}$ with transition probability matrix $\overline{A}$. By the assumption of stationarity and Proposition 2.3.1, the distribution of w_k for each $k \in \mathbb{T}$ is necessarily the uniform distribution on $\{1,\cdots,N\}$. We note that the stochastic processes $\{\widetilde{z}(k)\}$ and $\{w_k\}$ are equivalent in terms of the distribution induced on the path space. We define the family of transformations $\{\theta_k\}_{k\in\mathbb{T}}$ on Ω as the family of left-shifts, i.e.,

$$\theta_k w(\cdot) = w(k+\cdot), \quad \forall k \in \mathbb{T}. \tag{2.32}$$

Then, the space $(\Omega, \mathcal{F}, \mathbb{P}, \{\theta_k^R, k \in \mathbb{T}\})$ becomes the canonical path space of a two-sided stationary sequence equipped with the left-shift operator, satisfying the properties in definition of RDS in [30] to be a metric dynamical system and in fact, also ergodic.

We now define the cocycle φ over $\mathbb{S}_+^N$, which constructs the RDS of interest. We define $\varphi:\ \mathbb{T}_+ \times \Omega^R \times \mathbb{S}_+^N \mapsto \mathbb{S}_+^N$ by:

$$\varphi(0,w,X) = X,\ \forall w, X \tag{2.33}$$

$$\varphi(1,w,X) = f_{\mathcal{I}^0_{w(0)}}(X),\ \forall w, X \tag{2.34}$$

$$\begin{aligned}\varphi(k,w,X) &= f_{\mathcal{I}^{k-1}_{\theta_{k-1}w(0)}}(\varphi(k-1,w,X)) \\ &= f_{\mathcal{I}^{k-1}_{w(k-1)}}(\varphi(k-1,w,X)),\ \forall k>1, w, X.\end{aligned} \tag{2.35}$$

The equality in (2.35) comes from the property of the left-shit θ, i.e., $\theta_{k-1}w(0) = w(k-1)$. The cocycle $\varphi^{\overline{\gamma},R}$ satisfies the assumptions of measurability seen in its

arguments, and the continuity of the map $\varphi^{\overline{\gamma},R}(k,w,\cdot):\mathbb{S}_+^N\mapsto\mathbb{S}_+^N$ w.r.t. the phase variable X for each fixed k,w follows from the continuity of the corresponding Riccati operator. Therefore, the pair (θ,φ) forms a well-defined RDS on the phase space $\mathbb{S}_+^N$. Now we consider a random variable sequence $\left\{\varphi(k,w,\widetilde{P}(0))\right\}_{k\in\mathbb{T}_+}$ (the randomness is induced by w), which can be considered as successive random iterates of the RDS $\{\theta,\varphi\}$ with the initial state $\widetilde{P}(0)$. From the construction of $\{\theta,\varphi\}$, the sequence $\left\{\varphi(k,w,\widetilde{P}(0))\right\}_{k\in\mathbb{T}_+}$ is distributionally equivalent to the sequence $\widetilde{P}(k)_{k\in\mathbb{T}_+}$, i.e.,

$$\varphi(k,w,\widetilde{P}(0))\overset{d}{=}\widetilde{P}(k),\ \forall k\in\mathbb{T}_+. \tag{2.36}$$

Therefore, analyzing the asymptotic distribution properties of the sequence $\left\{\widetilde{P}(k)\right\}$ equals to studying the sequence $\{\varphi(k,w,P_n(0))\}$, which we will analyze in the sequel.

We first establish some properties of the RDS (θ,φ) that represents the sequence $\left\{\widetilde{P}(k)\right\}$.

Lemma 2.3.3

(i) The RDS (θ,φ) is conditionally compact.

(ii) The RDS (θ,φ) is order preserving.

(iii) If in addition $\mathcal{Q}$ is positive definite, i.e., $\mathcal{Q}\gg 0$, the RDS (θ,φ) is strongly sublinear.

The proof of Lemma 2.3.3 and the concepts including conditionally compact, order preserving, and sublinearity, are discussed in prior work [30].

2.3.4 Weak Consensus of Error Covariance over Network

We fix a $\overline{\gamma}>0$. First, we present the asymptotic properties of the auxiliary sequences $\left\{\widetilde{P}(k)\right\}$.

Theorem 2.3.4 Under the assumptions **C.1**, **S.1**, and **D.1**, there exists a unique invariant probability measure $\mu^{\overline{\gamma}}$ on the space of positive semidefinite matrices $\mathbb{S}_+^N$, such that the sequence $\left\{\widetilde{P}(k)\right\}$ converges weakly (in distribution) to $\mu^{\overline{\gamma}}$ from every initial condition $P_n(0)$ for each $n \in [1, \cdots, N]$, i.e.,

$$\left\{\widetilde{P}(k)\right\} \Rightarrow \mu^{\overline{\gamma}}. \tag{2.37}$$

Theorem 2.3.4 implies that the sequence $\left\{\widetilde{P}(k)\right\}$ reaches consensus in the weak sense to the same invariant measure $\mu^{\overline{\gamma}}$ irrespective of the initial states, since $\mu^{\overline{\gamma}}$ does not depend on the index n and on the initial state $\widetilde{P}(0)$ of the sequence $\left\{\widetilde{P}(k)\right\}$.

Based on Theorem 2.3.4, we can deduce Theorem 2.3.5, which does not directly touch the sequences $\left\{\widehat{P}_k^n\right\}$ for $n = 1, \cdots, N$, but sets the stage for showing the key result regarding the convergence of these sequences.

Theorem 2.3.5 As defined in Section 2.3.3, $\{\widetilde{z}(k)\}$ is a stationary Markov chain on V with transition probability matrix $\overline{A}$, i.e., $\widetilde{z}(0)$ is distributed uniformly on V. Let ν be a probability measure on $\mathbb{S}_+^M$; and the process $\left\{\widetilde{P}(k)\right\}$ is given by

$$\widetilde{P}(k+1) = f_{\jmath(\mathcal{I}^k_{\widetilde{z}(k)})}\left(\widetilde{P}(k)\right),\ k \in \mathbb{T}_+ \tag{2.38}$$

where $\widetilde{P}(0)$ is distributed as ν, independent of the Markov chain $\{\widetilde{z}(k)\}$ and the processes $\{\mathcal{I}_n^k\}$ for all n. Then, there exists a unique probability measure $\mu^{\overline{\gamma}}$ such that, for every ν, the process $\left\{\widetilde{P}(k)\right\}$ constructed above converges weakly to $\mu^{\overline{\gamma}}$ as $k \to \infty$, i.e.,

$$f_{\jmath(\mathcal{I}^k_{\widetilde{z}(k)})} \circ f_{\jmath(\mathcal{I}^{k-1}_{\widetilde{z}(k-1)})} \cdots \circ f_{\jmath(\mathcal{I}^0_{\widetilde{z}(0)})}\left(\widetilde{P}(0)\right) \Longrightarrow \mu^{\overline{\gamma}}. \tag{2.39}$$

We now state the theorem characterizing the convergence properties of the sequences $\left\{\widehat{P}_k^n\right\}$.

Theorem 2.3.6 Let q be a uniformly distributed random variable on V, independent of the sequence of adjacency matrices $\{A(k)\}$ and the processes $\{\mathcal{I}_n^k\}$. Then, the sequence $\left\{\widehat{P}_k^q\right\}$ converges weakly to $\mu^{\overline{\gamma}}$ defined in Theorem 2.3.5, i.e.,

$$\widehat{P}_k^q \Longrightarrow \mu^{\overline{\gamma}}. \tag{2.40}$$

In other words, the conditional error covariance $\left\{\widehat{P}_k^q\right\}$ of a randomly selected sensor converges in distribution to $\mu^{\overline{\gamma}}$.

Remark 2.3.7 Theorem 2.3.6 reinforces the weak consensus achieved by the M-GIKF algorithm, i.e., the conditional error covariance at a randomly selected sensor converges in distribution to an invariant measure $\mu^{\overline{\gamma}}$. In other words, it provides an estimate $\{\widehat{\mathbf{x}}_q(k)\}$ for the entire signal $\mathbf{x}$, where $\{\widehat{\mathbf{x}}_q(k)\}$ is obtained by uniformly selecting a sensor q independent with the random gossip protocol $\{A(k)\}$ and the process $\{\mathcal{I}_n^k\}$ and using its estimate $\{\widehat{\mathbf{x}}_q(k)\}$ for all time k. Also note that the results here pertain to the limiting distribution of the conditional error covariance and, hence, the pathwise filtering error, which is a much stronger result than just providing the moment estimates of the conditional error covariance, which does not provide much insight into the pathwise instantiation of the filter. In the following subsection, we provide the the analytical characterizations of the invariant measure $\mu^{\overline{\gamma}}$ by showing it satisfies the Large Deviation lower and upper bounds as $\mu^{\overline{\gamma}} \to \infty$.

The proofs of these Theorems are presented in Appendix of this chapter.

2.4 Appendices

2.4.1 Proof of Theorem 2.3.4

We take some steps to prove Theorem 2.3.4. First, we have the following lemma.

Lemma 2.4.1 Recall Assumption **D.1**, we assume there exists a walk on the graph induced by the non-zero entries of the matrix $\overline{A}$, $w_0 = \{n_1, \cdots, n_l\}$ covering the N nodes, such that the following Gramian matrix

$$G_{w_0}^{\overline{\gamma}} = \sum_{i=1}^{l} (\mathcal{F}^{i-1})^T \mathcal{C}_{n_i}^T \mathcal{C}_{n_i} \mathcal{F}^{i-1} \tag{2.41}$$

is invertible.

We define the function $g_{w_0}^{\overline{\gamma}} : \mathbb{S}_+^N \mapsto \mathbb{S}_+^N$ by

$$g_{w_0}^{\overline{\gamma}}(X) = f_{\mathcal{I}_l^{\overline{\gamma},n_l}} \circ f_{\mathcal{I}_{l-1}^{\overline{\gamma},n_{l-1}}} \circ \cdots \circ f_{\mathcal{I}_1^{\overline{\gamma},n_1}}(X), \tag{2.42}$$

where $\mathcal{I}_i^{\overline{\gamma},n_i} = \{n_1, \cdots, n_{|\jmath_i|}\}$.

Then, there exists a constant $\alpha_0 > 0$ such that the following uniformity condition holds,

$$g_{w_0}^{\overline{\gamma}}(X) \preceq \alpha_0 I, \quad \forall X \in \mathcal{S}_+^N, \tag{2.43}$$

i.e., the iterates $g_{w_0}^{\overline{\gamma}}(\cdot)$ is uniformly bounded irrespective of the initial value.

Proof 1 Recall (3.18),

$$f_\jmath(X) = \mathcal{F} X \mathcal{F}^T + \mathcal{Q} - \sum_{j=1}^{|\jmath|} \mathcal{F} X \mathcal{C}_{i_j}^T \left(\mathcal{C}_{i_j} X \mathcal{C}_{i_j}^T + \mathcal{R}_{i_j} \right)^{-1} \mathcal{C}_{i_j} X \mathcal{F}^T. \tag{2.44}$$

Clearly, $n_i \in \mathcal{I}_i^{\overline{\gamma},n_i}$ as stated in Section 2.2.2, thus we have

$$f_{\mathcal{I}_i^{\overline{\gamma},n_i}}(X) \preceq f_{n_i}(X). \tag{2.45}$$

In the Lemma 15 of [30], we have shown that there exists a constant $\alpha_0 > 0$ such that

$$f_{n_l} \circ f_{n_{l-1}} \circ \cdots \circ f_{n_1}(X) \preceq \alpha_0 I, \quad \forall X \in \mathcal{S}_+^N. \tag{2.46}$$

Combining (2.45) and (2.46), we conclude that

$$g_{w_0}^{\overline{\gamma}}(X) \preceq \alpha_0 I, \quad \forall X \in \mathcal{S}_+^N. \tag{2.47}$$

The following lemma establishes asymptotic boundedness of $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$.

Lemma 2.4.2 The sequence $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$ is stochastically bounded for each n under the assumptions of Theorem 2.3.4,

$$\lim_{J\to\infty} \sup_{k\in\mathbb{T}_+} \mathbb{P}\left(\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| > J\right) = 0. \tag{2.48}$$

Proof 2 In case that $\mathcal{F}$ is stable, the result is obvious, since the suboptimal estimate of 0 at each node for all time is stochastically bounded. Therefore, in the following we consider the case that $\mathcal{F}$ is unstable.

The proof mainly uses the uniform boundedness of the composition of Riccati operators in Lemma 2.4.1 and the ergodicity of the underlying switching Markov chain $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$. From Lemma 2.4.1, we see that a successive application of l Riccati maps (in the composition order of $f_{\mathcal{I}_l^{\overline{\gamma},n_l}} \circ \cdots \circ f_{\mathcal{I}_1^{\overline{\gamma},n_1}}$) constrains the iterate in the conic interval $[0, \alpha_0 I]$ irrespective of its initial value. Our approach is to relate the probability of large exceeding of $\widetilde{P}_{\overline{\gamma}}(k)$ to the hitting time statistic of a modified

Markov chain. We show in detail as follows.

First, we note that the regularity of the distribution of $\widetilde{P}_{\overline{\gamma}}(k)$ for every k implies that it suffices to show that

$$\lim_{J\to\infty}\sup_{k\geq k_0}\mathbb{P}\left(\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\|>J\right)=0 \tag{2.49}$$

for some arbitrary selected large $k_0\in\mathbb{T}_+$. For every n, the Riccati operator is upper bounded by the Lyapunov operator,

$$f_n(X)\preceq\mathcal{F}X\mathcal{F}^T+\mathcal{Q},\quad\forall X\in\mathbb{S}_+^N. \tag{2.50}$$

For sufficiently large $J>0$, we define $k(J)$ as follows, which will be used later.

$$k(J)=\max_k\left\{k\in\mathbb{T}_+\left|\alpha^{2k}\alpha_0+\frac{\alpha^{2k}-1}{\alpha^2-1}\|\mathcal{Q}\|\leq J\right.\right\}, \tag{2.51}$$

where $\alpha=\|\mathcal{F}\|$. Since $\mathcal{F}$ is unstable, i.e., $\alpha>1$, we have that $k(J)\to\infty$ when $J\to\infty$.

We introduce another notation. For integers $k_0,k_1\geq l$, the phrase "there exists a $(n_1,\cdots,n_l)$ cycle in the interval $[k_0,k_1]$" indicates the existence of an integer $k_0+l-1\leq k'\leq k_1$ such that,

$$\widetilde{z}(k'-l+s)=n_s,\quad 1\leq s\leq l, \tag{2.52}$$

where $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$ is the switching Markov chain.

We now make the following claim to relate the probability of interest for suffi-

ciently large J, then we prove it.

$$\mathbb{P}\left(\left\|\widetilde{P}_{\bar{\gamma}}(k)\right\| > J\right) \leq \mathbb{P}(\text{no}(n_1, \cdots, n_l)\text{exists in}[k-k(J), k]). \tag{2.53}$$

Indeed, we assume on the contrary that a $(n_1, \cdots, n_l)$ cycle exists in the interval $[k-k(J), k]$. Then, from (2.52), it means that there exists $k' \in [k-k(J), k]$ such that $\widetilde{z}(k'-l+s) = n_s, \quad 1 \leq s \leq l$, which implies that

$$\widetilde{P}_{\bar{\gamma}}(k') = f_{\mathcal{I}_l^{\bar{\gamma}, n_l}} \circ \cdots \circ f_{\mathcal{I}_1^{\bar{\gamma}, n_1}} \left(\widetilde{P}_{\bar{\gamma}}(k'-l+1)\right), \tag{2.54}$$

hence by Lemma 2.4.1, we have

$$\widetilde{P}_{\bar{\gamma}}(k') \preceq \alpha_0 I, \tag{2.55}$$

which holds irrespective of the value of $\widetilde{P}_{\bar{\gamma}}(k'-l+1)$. By (2.50), we see that

$$\widetilde{P}_{\bar{\gamma}}(s) \preceq \mathcal{F}\widetilde{P}_{\bar{\gamma}}(s-1)\mathcal{F}^T + \mathcal{Q}, \quad \forall s. \tag{2.56}$$

Continuing the recursion from k' and with the fact $\widetilde{P}_{\bar{\gamma}}(k') \preceq \alpha_0 I$ and $\alpha = \|\mathcal{F}\|$, we have

$$\left\|\widetilde{P}_{\bar{\gamma}}(k)\right\| \leq \alpha^{2(k-k')} \left\|\widetilde{P}_{\bar{\gamma}}(k')\right\| + \frac{\alpha^{2(k-k')}-1}{\alpha^2-1}\|\mathcal{Q}\| \leq \alpha^{2(k-k')}\alpha_0 + \frac{\alpha^{2(k-k')}-1}{\alpha^2-1}\|\mathcal{Q}\|. \tag{2.57}$$

Since $(k-k') \leq k(J)$, with the definition of $k(J)$ (2.51), we have from the above that

$$\left\|\widetilde{P}_{\bar{\gamma}}(k)\right\| \leq \alpha^{2(k-k')}\alpha_0 + \frac{\alpha^{2(k-k')}-1}{\alpha^2-1}\|\mathcal{Q}\| \leq \alpha^{2k(J)}\alpha_0 + \frac{\alpha^{2k(J)}-1}{\alpha^2-1}\|\mathcal{Q}\| \leq J. \tag{2.58}$$

Thus we note that the existence of a $(n_1, \cdots, n_l)$ cycle in the interval $[k-k(J), k]$ implies $\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| \leq J$, i.e., we have the following event inclusion,

$$\{\text{there exists a } (n_1, \cdots, n_l) \text{ cycle in } [k-k(J), k]\} \subset \left\{\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| \leq J\right\}, \tag{2.59}$$

from which the claim in (2.53) is established. Therefore estimating the probability of $\mathbb{P}\left(\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| > J\right)$ can be reduced to estimating the R.H.S. of (2.53). To this end, we construct another Markov chain $\{z'(k)\}_{k\geq l}$, with the sate space $\mathcal{Z}$ is a subset of V^l given by

$$\mathcal{Z} = \left\{z' = (i_1, \cdots, i_l) | \overline{A}_{i_j, i_{j+1}} > 0, 1 \leq j < l\right\}. \tag{2.60}$$

The dynamic of the Markov chain $\{z'(k)\}_{k\geq l}$ is given in terms of the Markov chain $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$ as follows,

$$z'(k) = (\widetilde{z}(k-l+1), \cdots, \widetilde{z}(k)). \tag{2.61}$$

From the dynamics of $\{\widetilde{z}(k)\}_{k\in\mathbb{T}_+}$, it follows that $\{z'(k)\}_{k\geq l}$ is a Markov chain with the transition probability $\overline{A}_{nm}$ between allowable states $(i_1, i_2, \cdots, i_{l-1}, n)$ and $(i_2, \cdots, i_{l-1}, n, m)$. With state space $\mathcal{Z}$, the Markov chain $\{z'(k)\}$ inherits the irreducibility and aperiodicity from that of $\{\widetilde{z}(k)\}$. And $\{z'(k)\}$ is also stationary from the stationarity of $\{\widetilde{z}(k)\}$ with invariant distribution,

$$\mathbb{P}(z'(k) = (i_1, \cdots, i_l)) = \frac{1}{N}\prod_{j=1}^{l-1}\overline{A}_{i_j, i_{j+1}}, \quad (i_1, \cdots, i_l) \in \mathcal{Z},\ k \geq l,\ k \in \mathbb{T}_+. \tag{2.62}$$

Denote the hitting time τ_0 of $\{z'(k)\}$ to the state $(n_1, \cdots, n_l)$ as

$$\tau_0 = \min\{k > l | z'(k) = (n_1, \cdots, n_l)\} \tag{2.63}$$

and for all $z' \in \mathcal{Z}$, we define

$$\mathbb{P}_{z'}(\tau_0 > s) = \mathbb{P}(\tau_0 > s | z'(l) = z'). \tag{2.64}$$

Also for each $k \geq l$ and J sufficiently large, we define the stopping times as

$$\tau_k^J = \min\{k \geq k - k(J) | z'(k) = (n_1, \cdots, n_l)\}. \tag{2.65}$$

From the Markovian property then it follows

$$\mathbb{P}(\tau_k^J > k | z'(k - k(J) - 1) = z') = \mathbb{P}_{z'}(\tau_0 > k(J) + 1). \tag{2.66}$$

Then it follows successively that

$$\begin{aligned}
&\mathbb{P}(\text{no } (n_1, \cdots, n_l) \text{ exists in } [k - k(J), k]) \\
&= \mathbb{P}(\tau_k^J > k) \\
&= \sum_{z' \in \mathcal{Z}} \left[\mathbb{P}(z'(k - k(J) - 1) = z')\mathbb{P}(\tau_k^J > k | z'(k - k(J) - 1) = z')\right] \\
&= \sum_{z' \in \mathcal{Z}} \mathbb{P}(z'(k - k(J) - 1) = z')\mathbb{P}_{z'}(\tau_0 > k(J) + 1).
\end{aligned} \tag{2.67}$$

Since the above result holds for all $k \geq k_0$ for some sufficiently large k_0, we conclude from (2.53)

$$\sup_{k \geq k_0} \mathbb{P}\left(\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| > J\right) \leq \sum_{z' \in \mathcal{Z}} \mathbb{P}(z'(k - k(J) - 1) = z')\mathbb{P}_{z'}(\tau_0 > k(J) + 1). \tag{2.68}$$

The recurrence of the finite state Markov chain $\{z'(k)\}$ and the fact that $k(J) \to \infty$

as $J \to \infty$ imply, for all $z' \in \mathcal{Z}$,

$$\lim_{J\to\infty} \mathbb{P}_{z'}(\tau_0 > k(J) + 1) = 0. \tag{2.69}$$

Since $\mathcal{Z}$ is finite, letting $J \to \infty$, we have from (2.68) that

$$\lim_{J\to\infty} \sup_{k\geq k_0} \mathbb{P}\left(\left\|\widetilde{P}_{\overline{\gamma}}(k)\right\| > J\right) = 0. \tag{2.70}$$

Then the lemma holds.

We now prepare to prove Theorem 2.3.4.

From Lemma 2.3.3, $(\theta^{\overline{\gamma},R}, \varphi^{\overline{\gamma},R})$ is conditionally compact, order preserving, and strongly sublinear. And the cone $\mathbb{S}_+^N$ satisfies the conditions required in the assumptions of Theorem 27 in [30]. We also have for $k > 0$,

$$\varphi^{\overline{\gamma},R}(k, w, 0) = f_{\mathcal{I}_{w_{k-1}}^{k-1} w(0)}(\varphi^{\overline{\gamma},R}(k-1, w, 0)) \succeq \mathcal{Q} \gg 0. \tag{2.71}$$

Therefore, the conditions in Theorem 27 in [30] are all satisfied, and exactly one of the claims **a)** and **b)** holds. By an argument similar to Lemma 6.1 in [32], the claim **a)** can not hold in the case of stochastically boundedness of $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$ from Lemma 2.4.2. Therefore, **b)** holds, then as a direct conclusion of Theorem 27 in [30], we can establish the existence of a unique almost equilibrium $u^{\overline{\gamma}}(w) \gg 0$ defined on a θ^R invariant set $\Omega^* \in \mathcal{F}^R$ with $\mathbb{P}(\Omega^*) = 1$, such that for random variable $v(w)$ possessing the property $0 \preceq v(w) \preceq \alpha u^{\overline{\gamma}}(w)$ for all $w \in \Omega^*$ and deterministic $\alpha > 0$, the following result holds,

$$\lim_{k\to\infty} \varphi(k, \theta_{-k}w, v(\theta_{-k}w)) = u^{\overline{\gamma}}(w), \quad w \in \Omega^*. \tag{2.72}$$

Since the distribution of the pull-back and forward orbits are equal, from Lemma 24 in [30], $u^{\overline{\gamma}}(w)$ is also the converged unique equilibrium measure for the sequence $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$. However, to show that the measure induced by $u^{\overline{\gamma}}$ on $\mathbb{S}_+^N$ is attracting for $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$, (2.72) must hold for all initial v. We state the following results to extend to the general initial conditions.

Lemma 2.4.3 Under the assumption of Theorem 2.3.4, let $u^{\overline{\gamma}}$ be the unique almost equilibrium of the RDS $(\theta^{\overline{\gamma},R}, \varphi^{\overline{\gamma},R})$. Then

$$\mathbb{P}\left(w : u^{\overline{\gamma}}(w) \succeq \mathcal{Q}\right) = 1. \tag{2.73}$$

Proof 3 The proof uses the fact that for all n, $f_{\mathcal{I}_k^n} \succeq \mathcal{Q}$, and follows the development in Lemma 6.2 of [32].

Now we start the proof of Theorem 2.3.4. The proof logic is first finding a suitable modification $\widetilde{X}(w)$ of an arbitrary initial condition P_0, such that $\widetilde{X}(w) = P_0$ a.s. and a deterministic $\alpha > 0$ satisfying $0 \preceq \widetilde{X}(w) \preceq \alpha u^{\overline{\gamma}}(w)$. Then, from (2.72), we can establish the weak convergence of the sequence $\left\{\varphi^{\overline{\gamma},R}(k, w, \widetilde{X}(w))\right\}$ with initial condition $\widetilde{X}(w)$ to $\mu^{\overline{\gamma}}$. Finally, since $\widetilde{X}(w)$ is a.s. equal to P_0, we can deduce the weak convergence of the desired sequence $\left\{\varphi^{\overline{\gamma},R}(k, w, P_0)\right\}$. We state the proof of Theorem 2.3.4 as follows.

Proof 4 Denote $\mu^{\overline{\gamma}}$ as the distribution of the unique almost equilibrium in (2.72). With Lemma 2.4.3, we have $\mu^{\overline{\gamma}}(\mathbb{S}_{++}^N) = 1$. Let $P_0 \in \mathbb{S}_+^N$ be an arbitrary initial condition. By construction of the RDS $(\theta^{\overline{\gamma},R}, \varphi^{\overline{\gamma},R})$, the sequences $\{P_k\}$ and $\left\{\varphi^{\overline{\gamma},R}(k, w, P_0)\right\}$ are distributional equivalent, i.e., $P_t \overset{d}{=} \varphi^{\overline{\gamma},R}(k, w, P_0)$. Recall Ω^* as the θ^R-invariant set with $\mathbb{P}(\Omega^*) = 1$ in (2.72) on which the almost equilibrium $u^{\overline{\gamma}}$ is

defined. By Lemma 2.4.3, there exists $\Omega_1 \in \Omega^*$ with $\mathbb{P}(\Omega_1) = 1$, such that

$$u^{\overline{\gamma}}(w) \succeq \mathcal{Q}, \quad \forall w \in \Omega_1. \tag{2.74}$$

We define the random variable $\widetilde{X}(w)$ by

$$\widetilde{X}(w) = \begin{cases} P_0, & w \in \Omega_1, \\ 0, & w \in \Omega_1^c. \end{cases}$$

Now choose $\alpha > 0$ large enough, such that $P_0 \preceq \alpha\mathcal{Q}$, which is possible since $\mathcal{Q} \gg 0$. Then, we have

$$0 \preceq P_0 = \widetilde{X}(w) \preceq \alpha\mathcal{Q} \preceq \alpha u^{\overline{\gamma}}(w), \ w \in \Omega_1, \tag{2.75}$$

$$0 = \widetilde{X}(w) \preceq \alpha u^{\overline{\gamma}}(w), \ w \in \Omega_1^c. \tag{2.76}$$

Therefore,

$$0 \preceq \widetilde{X}(w) \preceq \alpha u^{\overline{\gamma}}(w), \ w \in \Omega^*. \tag{2.77}$$

Then, with (2.72), we have

$$\lim_{k\to\infty} \varphi^{\overline{\gamma},R}(k, \theta_{-k}w, \widetilde{X}(\theta_{-k}w)) = u^{\overline{\gamma}}(w), \quad w \in \Omega^*. \tag{2.78}$$

Since convergence a.s. implies convergence in distribution, we have

$$\varphi^{\overline{\gamma},R}(k, \theta_{-k}w, \widetilde{X}(\theta_{-k}w)) \Rightarrow \mu^{\overline{\gamma}}, \tag{2.79}$$

as $k \to \infty$. By Lemma 24 in [30], the sequence $\left\{\varphi^{\overline{\gamma},R}(k, w, \widetilde{X}(w))\right\}$ also converges

in distribution to the unique distribution $\mu^{\overline{\gamma}}$, i.e, as $k \to \infty$

$$\varphi^{\overline{\gamma},R}(k, w, \widetilde{X}(w)) \Rightarrow \mu^{\overline{\gamma}}. \tag{2.80}$$

Now, since $\mathbb{P}(\Omega_1) = 1$, by (2.75),

$$\varphi^{\overline{\gamma},R}(k, w, P_0) = \varphi^{\overline{\gamma},R}(k, w, \widetilde{X}(w)),\ \mathbb{P}\ a.s., \tag{2.81}$$

which shows

$$\varphi^{\overline{\gamma},R}(k, w, P_0) \stackrel{d}{=} \varphi^{\overline{\gamma},R}(k, w, \widetilde{X}(w)). \tag{2.82}$$

From (2.80) and (2.82), we have $\varphi^{\overline{\gamma},R}(k, w, P_0) \Rightarrow \mu^{\overline{\gamma}}$. Since $P_t \stackrel{d}{=} \varphi^{\overline{\gamma},R}(k, w, P_0)$, finally we have $P_t \Rightarrow \mu^{\overline{\gamma}}$, as $k \to \infty$. The proof is completed.

2.4.2 Proof of Theorem 2.3.5

By Theorem 2.3.4, we have obtained that such a sequence $\left\{\widetilde{P}_{\overline{\gamma}}(k)\right\}$ converges weakly to $\mu_{\overline{\gamma}}$ when stated from any deterministic initial condition. In the case that $\widetilde{P}_{\overline{\gamma}}(0)$ is distributed as ν, by the independence of $\widetilde{P}_{\overline{\gamma}}(0)$ with the Markov chain $\{\widetilde{z}(k)\}$ and the processes $\{\mathcal{I}_n^k\}$ for all n, we have for $g \in \mathcal{C}(\mathbb{S}_+^N)$,

$$\mathbb{E}\left[g\left(\widetilde{P}_{\overline{\gamma}}(k)\right)\right] = \int_{\mathbb{S}_+^N} \mathbb{E}\left[g\left(\widetilde{P}_{\overline{\gamma}}(k)\right) | \widetilde{P}_{\overline{\gamma}}(0) = X\right] d\nu(X). \tag{2.83}$$

Now the distribution of the sequence $\{\widetilde{P}_{\overline{\gamma}}(k)\}$ conditioned on the event $\widetilde{P}_{\overline{\gamma}}(0) = X$ is the same as that when the sequence starts with the deterministic initial condition X (this is true because $\widetilde{P}_{\overline{\gamma}}(0)$ is independent of $\{\widetilde{z}(k)\}$). Hence by Theorem 2.3.4,

$$\lim_{k\to\infty} \mathbb{E}\left[g\left(\widetilde{P}_{\overline{\gamma}}(k)\right) | \widetilde{P}_{\overline{\gamma}}(0) = X\right] = \int_{\mathbb{S}_+^N} g(y) d\mu_{\overline{\gamma}}(Y) \tag{2.84}$$

for all X. Since g is bounded, the dominated convergence theorem and (2.83) result in

$$\lim_{k\to\infty} \mathbb{E}\left[g\left(\widetilde{P}_{\overline{\gamma}}(k)\right)\right] = \int_{\mathbb{S}_+^N} g(y) d\mu_{\overline{\gamma}}(Y) \tag{2.85}$$

for all $g \in \mathcal{C}(\mathbb{S}_+^N)$. Hence the required convergence in distribution follows.

3. LARGE DEVIATION ANALYSIS FOR M-GIKF ALGORITHM*

In this chapter, we first give the overview of large deviation principle. Then, we present some intermediate results including the definitions about string and the properties of Riccati equation. At last, as the main part in this chapter, we present the large deviation results for the invariant measure $\mu^{\overline{\gamma}}$, by deriving the large deviation lower and upper bounds, respectively.

3.1 Overview of Large Deviation Principle

Let $\{\mu^{\overline{\gamma}}\}$ be a family of probability measures on the complete separable metric space $(\mathcal{X}, d_{\mathcal{X}})$ indexed by the real-valued parameter $\overline{\gamma}$ taking values in $\mathbb{R}_+$. Let $\overline{I} : \mathcal{X} \longmapsto \overline{\mathbb{R}}_+$ be an extended-valued lower semicontinuous function. The family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy a large deviations upper bound with rate function $\overline{I}(\cdot)$ if the following holds:

$$\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(\mathcal{F}) \leq -\inf_{X\in\mathcal{F}} \overline{I}(X), \text{ for every closed set } \mathcal{F}\in\mathcal{X}. \tag{3.1}$$

Similarly, for an extended-valued lower semicontinuous function $\underline{I} : \mathcal{X} \longmapsto \overline{\mathbb{R}}_+$, the family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy a large deviations lower bound with rate function $\underline{I}(\cdot)$, if

$$\liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(\mathcal{O}) \geq -\inf_{X\in\mathcal{O}} \underline{I}(X), \text{ for every open set } \mathcal{O}\in\mathcal{X}. \tag{3.2}$$

*Part of this chapter is reprinted, with permission, from [Di 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*, 61(3):1351–1372, Mar. 2015.]

In addition, if the functions $\overline{I}$ and $\underline{I}$ coincide, i.e., $\overline{I} = \underline{I} = I$, the family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy a large deviations principle (LDP) with rate function $I(\cdot)$ (see [15]). The lower semicontinuity implies that the level sets of $\overline{I}(\cdot)$ (or $\underline{I}(\cdot)$), i.e., sets of the form $\{X \in \mathcal{X} \mid \overline{I}(X) \leq \alpha\}$ (or $\underline{I}(\cdot)$) for every $\alpha \in \mathbb{R}_+$, are closed. If, in addition, the levels sets are compact (for every α), $\overline{I}(\cdot)$ (or $\underline{I}(\cdot)$) is said to be a good rate function.

Before interpreting the consequences of the LD upper and lower bounds as defined above, we consider the notion of a rare event, which is the central motivation to all large deviations:

Definition 3.1.1 (Rare Event) A set $\Gamma \subset \mathcal{B}(\mathcal{X})$ is called a rare event with respect to (w.r.t.) the family $\{\mu^{\overline{\gamma}}\}$ of probability measures, if $\lim_{\overline{\gamma}\to\infty} \mu^{\overline{\gamma}}(\Gamma) = 0$. In other words, the event Γ becomes increasingly difficult to observe (i.e., it becomes rare) as $\overline{\gamma} \to \infty$.

Once a rare event Γ is identified, the next natural question is the rate at which its probability goes to zero under $\mu^{\overline{\gamma}}$ as $\overline{\gamma} \to \infty$. This is answered by the LD upper and lower bounds, which also characterize the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \to \infty$. Indeed, it is not hard to see that, if the family $\{\mu^{\overline{\gamma}}\}$ satisfies the LD upper and lower bounds, we have for every measurable set $\Gamma \in \mathcal{X}$:

$$\mu^{\overline{\gamma}}(\Gamma) \leq e^{-\overline{\gamma}(\inf_{X\in\overline{\Gamma}} \overline{I}(X)+o(1))} \tag{3.3}$$

$$\mu^{\overline{\gamma}}(\Gamma) \geq e^{-\overline{\gamma}(\inf_{X\in\Gamma^\circ} \underline{I}(X)+o(1))}, \tag{3.4}$$

where $o(1)$ is the little-o notation. Now assume $\inf_{X\in\overline{\Gamma}} \overline{I}(X) > 0$. Then, from (3.3) it is clear that Γ is a rare event and, in fact, we conclude that the probability of Γ decays exponentially with a LD exponent greater than or equal to $\inf_{X\in\overline{\Gamma}} \overline{I}(X)$. Similarly, $\inf_{X\in\Gamma^\circ} \underline{I}(X) > 0$ suggests that the LD decay exponent is not arbitrary

and cannot be larger than $\inf_{X\in\Gamma^\circ} \underline{I}(X)$. In addition, if the rate functions $\overline{I}$ and $\underline{I}$ close to each other, the estimate of the exact decay exponent is tight.

3.2 Some Intermediate Results

In this section, some preliminary results on string and approximation results regarding Riccati equation are presented.

3.2.1 Preliminary Results on String

The RRE sequence is an iterated function system (see, e.g., [16]) comprising of random compositions of Riccati operators. Understanding the system requires studying the behavior of such random function compositions, where not only the numerical value of the composition is important, but also the composition pattern is relevant. To formalize this study, we start with the following definitions.

Definition 3.2.1 (String) Let $P_0 \in \mathbb{S}_+^M$. A string $\mathcal{R}$ with initial state P_0 and length $r \in \mathbb{N}$ is a $(r+1)$-tuple of the form:

$$\mathcal{R} = \left(f_{\jmath_r}, f_{\jmath_{r-1}}, \cdots f_{\jmath_1}, P_0\right), \quad \jmath_1, \cdots, \jmath_r \in \mathfrak{P} \tag{3.5}$$

where $f_\jmath$ corresponds to the Riccati operator defined in (2.22). The length of a string $\mathcal{R}$ is denoted by $\text{len}(\mathcal{R})$. The set of all possible strings is denoted by $\overline{\mathcal{S}}$.

Fix $\overline{\gamma} > 0$. A string $\mathcal{R}$ of the form

$$\mathcal{R} = \left(f_{\jmath_r}, f_{\jmath_{r-1}}, \cdots f_{\jmath_1}, P_0\right), \quad \jmath_1, \cdots, \jmath_r \in \mathfrak{P}$$

is called $\overline{\gamma}$-feasible, if there exists a path[1] $(n_r, n_{r-1}, \cdots, n_1)$ of length r w.r.t. $\overline{A}$, such that $q_{n_i}(\jmath_i) > 0$ (recall $q_{n_i}(\jmath_i)$ defined in (2.10)) for all $1 \le i \le r$. The set of all

[1] A sequence of nodes $(n_r, n_{r-1}, \cdots, n_1)$ is called a path w.r.t. $\overline{A}$ if $\overline{A}_{n_i, n_{i+1}} > 0$ for all $1 \le i < r$.

$\overline{\gamma}$-feasible strings is further denoted by $\overline{\mathcal{S}}_{\overline{\gamma}}$.

Remark 3.2.2 Note that a string $\mathcal{R}$ can be of length 0; then it is represented as a 1-tuple, consisting of only the initial condition.

Let $r_1, r_2, \cdots, r_l$ be non-negative integers, such that $\sum_{i=1}^{l} r_i = r$ and $\jmath_i^k \in \mathfrak{P}$ for $1 \leq i \leq r_k$ and $1 \leq k \leq l$, where for all k, we have $\jmath_i^k = \jmath_1^k,\ 1 \leq i \leq r_k$. Let $\mathcal{R}$ be a string of length r of the following form:

$$\mathcal{R} = \left(f_{\jmath_1^1}, \cdots, f_{\jmath_{r_1}^1}, \cdots, f_{\jmath_1^2}, \cdots, f_{\jmath_{r_2}^2}, \cdots, f_{\jmath_1^l}, \cdots, f_{\jmath_{r_l}^l}, P_0 \right). \tag{3.6}$$

For brevity, we write $\mathcal{R}$ as

$$\mathcal{R} = \left(f_{\jmath_1^1}^{r_1}, f_{\jmath_1^2}^{r_2}, \cdots, f_{\jmath_1^l}^{r_l}, P_0 \right). \tag{3.7}$$

For example, the string $(f_1, f_2, f_2, f_2, f_1, f_1, P_0)$ could be written concisely as (f_1, f_2^3, f_1^2, P_0).

Definition 3.2.3 (Numerical Value of a String) Every string $\mathcal{R}$ is associated with its numerical value, denoted by $\mathcal{N}(\mathcal{R})$, which is the numerical evaluation of the function composition on the initial state P_0; i.e., for $\mathcal{R}$ of the form

$$\mathcal{R} = \left(f_{\jmath_r}, f_{\jmath_{r-1}}, \cdots f_{\jmath_1}, P_0 \right), \quad \jmath_1, \cdots, \jmath_r \in \mathfrak{P},$$

we have

$$\mathcal{N}(\mathcal{R}) = f_{\jmath_r} \circ f_{\jmath_{r-1}} \circ \cdots \circ f_{\jmath_1}(P_0). \tag{3.8}$$

Thus[2], the numerical value can be viewed as a function $\mathcal{N}(\cdot)$ mapping from the space $\overline{\mathcal{S}}$ of strings to $\mathbb{S}_+^M$. We abuse notation by denoting $\mathcal{N}(\overline{\mathcal{S}})$ as the set of numerical

[2]For function compositions, we adopt a similar notation to that of strings; for example, we denote the composition $f_1 \circ f_2 \circ f_2 \circ f_2 \circ f_1 \circ f_1(P_0)$ by $f_1 \circ f_2^3 \circ f_1^2(P_0)$.

values attainable, i.e.,

$$\mathcal{N}(\overline{\mathcal{S}}) = \left\{\mathcal{N}(\mathcal{R}) \mid \mathcal{R} \in \overline{\mathcal{S}}\right\}. \tag{3.9}$$

Similarly, by $\mathcal{N}(\overline{\mathcal{S}}_{\overline{\gamma}})$ we denote the subset of numerical values associated to the $\overline{\gamma}$-feasible strings $\overline{\mathcal{S}}_{\overline{\gamma}}$.

Remark 3.2.4 Note the difference between a string and its numerical value. Two strings are equal if and only if they comprise the same order of function compositions applied to the same initial state. In particular, two strings can be different, even if they are evaluated with the same numerical value.

For fixed $P_0 \in \mathbb{S}_+^M$ and $r \in \mathbb{N}$, the subset of strings of length r and initial condition P_0 is denoted by $\mathcal{S}_r^{P_0}$. The corresponding set of numerical values is denoted by $\mathcal{N}(\mathcal{S}_r^{P_0})$. Finally, for $X \in \mathbb{S}_+^M$, the set $\mathcal{S}_r^{P_0}(X) \subset \mathcal{S}_r^{P_0}$ consists of all strings with numerical value X, i.e.,

$$\mathcal{S}_r^{P_0}(X) = \left\{\mathcal{R} \in \mathcal{S}_r^{P_0} \mid \mathcal{N}(\mathcal{R}) = X\right\}. \tag{3.10}$$

In the following, we present some important properties of strings to be used later. Recall from [30] that $\mathbb{S}_{++}^M$ is the cone of positive definite matrices.

Proposition 3.2.5

(i) For $r_1 \leq r_2 \in \mathbb{N}$, we have $\mathcal{N}\left(\mathcal{S}_{r_1}^{P^*}\right) \subset \mathcal{N}\left(\mathcal{S}_{r_2}^{P^*}\right)$, where $P^* \in \mathbb{S}_{++}^M$ denotes the unique fixed point of the Riccati operator f_{2^N-1}. In particular, if for some $X \in \mathbb{S}_+^M$, $r_0 \in \mathbb{N}$, and $\jmath_{r_0}, \cdots, \jmath_1 \in \mathfrak{P}$, the string $\mathcal{R} = \left(f_{\jmath_{r_0}}, \cdots, f_{\jmath_1}, P^*\right)$ belongs to $\mathcal{S}_{r_0}^{P^*}(X)$, we have

$$\left(f_{\jmath_{r_0}}, \cdots, f_{\jmath_1}, f_{2^N-1}^{r-r_0}, P^*\right) \in \mathcal{S}_r^{P^*}(X) \subset \mathcal{S}^{P^*}(X), \ \forall r \geq r_0. \tag{3.11}$$

(ii) Let $r \in \mathbb{N}$ and $\mathcal{R} \in \mathcal{S}_r^{P_0} = (f_{\jmath_r}, \cdots, f_{\jmath_1}, P_0)$ be a string. Define the function $\pi(\cdot)$ by

$$\pi(\mathcal{R}) = \begin{cases} \sum_{i=1}^{r} \left(1 - \mathbb{I}_{\{2^N-1\}}(\jmath_i)\right), & \text{if } r \geq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

i.e., $\pi(\mathcal{R})$ counts the number of occurrences of the non-centralized Riccati operator f_{2^N-1} in $\mathcal{R}$.

Also denote $\widehat{\mathcal{R}} = (f_{\hat{\jmath}_{\pi(\mathcal{R})}}, f_{\hat{\jmath}_{\pi(\mathcal{R})-1}}, \cdots, f_{\hat{\jmath}_1}, P_0)$, which represents the string of length $\pi(\mathcal{R})$ obtained by removing the occurrences of f_{2^N-1} from $\mathcal{R}$[3].

Then, there exists $\alpha_{P_0} \in \mathbb{R}_+$, depending on P_0 only, such that

$$f_{\hat{\jmath}_{\pi(\mathcal{R})}} \circ f_{\hat{\jmath}_{\pi(\mathcal{R})-1}} \cdots \circ f_{\hat{\jmath}_1}(\alpha_{P_0} I) \succeq \mathcal{N}(\mathcal{R}). \tag{3.13}$$

Proof 5 The proof is a straightforward generalization of Proposition 3.6 in [31] and is omitted.

3.2.2 Riccati Equation

In this section, we present several approximation results needed in the sequel. We discuss generic properties, like uniform convergence of the Riccati operator, which will be used in the sequel to obtain various tightness estimates required for establishing the LD results.

Proposition 3.2.6

(i) For every $X \in \mathbb{S}_+^M$ and $\jmath \in [0, \cdots, 2^N - 1]$, we have

$$f_\jmath(X) \succeq f_{2^N-1}(X). \tag{3.14}$$

[3]For example, if $\mathcal{R} = (f_1, f_{2^N-1}, f_3, f_{2^N-1}, f_2, P_0)$, $\widehat{\mathcal{R}} = (f_1, f_3, f_2, P_0)$.

(ii) For every $\varepsilon > 0$, there exists $r_\varepsilon \geq M$, such that, for every $X \in \mathbb{S}_+^M$, with $X \succeq P^*$ (P^* is the unique fixed point of the centralized Riccati operator f_{2^N-1}), we have

$$\left\| f_{2^N-1}^r(X) - P^* \right\| \leq \varepsilon, \quad r \geq r_\varepsilon. \tag{3.15}$$

Note, in particular, that r_ε can be chosen independent of the initial state X.

(iii) For a fixed $r \in \mathbb{N}$ and $\jmath_r, \cdots, \jmath_1 \in \mathfrak{P}$, define the function $g : \mathbb{S}_+^M \longmapsto \mathbb{S}_+^M$ by

$$g(X) = f_{\jmath_r} \circ \cdots \circ f_{\jmath_1}(X), \quad X \in \mathbb{S}_+^M. \tag{3.16}$$

Then $g(\cdot)$ is Lipschitz continuous with some constant $K_g > 0$. Also, for every $\varepsilon_2 > 0$, there exists r_{ε_2}, such that the function $f_{2^N-1}^{r_{\varepsilon_2}}(\cdot)$ is Lipschitz continuous with constant $K_{f_{2^N-1}^{r_{\varepsilon_2}}} < \varepsilon_2$.

Proof 6 The second and third assertions follow from Lemmas 3.1 and 3.2 in [31]. For the the first assertion, note that by (2.22) we have

$$f_\jmath(X) = \mathcal{F} X \mathcal{F}^T + \mathcal{Q} - \mathcal{F} X \mathcal{C}_\jmath^T \left(\mathcal{C}_\jmath X \mathcal{C}_\jmath^T + \mathcal{R}_\jmath \right)^{-1} \mathcal{C}_\jmath X \mathcal{F}^T \tag{3.17}$$

where $\mathcal{C}_\jmath = [\mathcal{C}_{i_1}^T \cdots \mathcal{C}_{i_{|\jmath|}}^T]^T$. Hence, we can rewrite the above equation as

$$f_\jmath(X) = \mathcal{F} X \mathcal{F}^T + \mathcal{Q} - \sum_{j=1}^{|\jmath|} \mathcal{F} X \mathcal{C}_{i_j}^T \left(\mathcal{C}_{i_j} X \mathcal{C}_{i_j}^T + \mathcal{R}_{i_j} \right)^{-1} \mathcal{C}_{i_j} X \mathcal{F}^T \tag{3.18}$$

that is obtained due to the fact that $\mathcal{C}_\jmath X \mathcal{C}_\jmath^T + \mathcal{R}_\jmath$ is block diagonal, from which it

follows that

$$\begin{aligned}
&\left(\mathcal{C}_\jmath X \mathcal{C}_\jmath^T + \mathcal{R}_\jmath\right)^{-1} \\
&= \begin{pmatrix} \left(\mathcal{C}_{i_1} X \mathcal{C}_{i_1}^T + \mathcal{R}_{i_1}\right)^{-1} & & \\ & \ddots & \\ & & \left(\mathcal{C}_{i_{|\jmath|}} X \mathcal{C}_{i_{|\jmath|}}^T + \mathcal{R}_{i_{|\jmath|}}\right)^{-1} \end{pmatrix}. \qquad (3.19)
\end{aligned}$$

Since $\jmath = 2^N - 1$ corresponds to the entire set of nodes $\{1, \cdots, N\}$, we have $|\jmath| \leq N$ and hence

$$\begin{aligned}
&f_{2^N-1}(X) = \mathcal{F} X \mathcal{F}^T + \mathcal{Q} \\
&- \sum_{j=1}^{N} \mathcal{F} X \mathcal{C}_{i_j}^T \left(\mathcal{C}_{i_j} X \mathcal{C}_{i_j}^T + \mathcal{R}_{i_j}\right)^{-1} \mathcal{C}_{i_j} X \mathcal{F}^T. \qquad (3.20)
\end{aligned}$$

Therefore, $f_\jmath(X) \succeq f_{2^N-1}(X)$.

3.3 Large Deviation Analysis

In this section, we first present the main results on large deviation analysis. The rest of this section is devoted to prove these main results. The upper and lower rate functions are defined, followed by their properties. Then, the large deviation lower and upper bounds are derived sequentially, which together complete the large deviation analysis.

3.3.1 Main Results

We characterize the invariant measure $\mu^{\overline{\gamma}}$ governing the asymptotics of the conditional sensor error covariance process $\{\widehat{P}_k^n\}$, $n = 1, \cdots, N$. The following result is a first step to understanding the behavior of the invariant distribution $\mu^{\overline{\gamma}}$ family.

Theorem 3.3.1 The family of invariant distributions $\mu^{\overline{\gamma}}$ converges weakly, as $\overline{\gamma} \to$

∞, to the Dirac measure δ_{P^*} corresponding to the performance of the centralized estimator (recall, P^* is the unique fixed point of the centralized Riccati operator f_{2^N-1}).

Remark 3.3.2 Theorem 3.3.1 states that the family $\{\mu^{\overline{\gamma}}\}$ converges weakly to the Dirac measure δ_{P^*} concentrated at P^*, as $\overline{\gamma} \to \infty$, which is intuitive, since with $\overline{\gamma} \to \infty$, the distributed M-GIKF filtering process reduces to classical Kalman filtering with all the observations available at a fusion center, i.e., centralized filtering, where P^* is the unique fixed point of this centralized filtering. Therefore, with $\overline{\gamma} \to \infty$, we expect the M-GIKF to perform more and more similarly to the centralized case, which leads to the weak convergence of the measure $\mu^{\overline{\gamma}}$ to δ_{P^*} as $\overline{\gamma} \to \infty$. An immediate consequence of Theorem 3.3.1 is

$$\lim_{\overline{\gamma}\to\infty} \mu^{\overline{\gamma}}(\Gamma) = 0, \quad \forall \overline{\Gamma} \cap P^* = \emptyset \tag{3.21}$$

which means, w.r.t. $\{\mu^{\overline{\gamma}}\}$, every event Γ with $P^* \notin \overline{\Gamma}$ is a rare event. This is intuitively correct, since as $\overline{\gamma} \to \infty$, the measures $\{\mu^{\overline{\gamma}}\}$ become more and more concentrated on an arbitrarily small neighborhood of P^*, resulting in the event Γ becoming very difficult to observe.

The proof of this theorem is presented in Appendix 3.5.1. In the sequel, we establish the LD upper and lower bounds for the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \to \infty$, which completely characterizes the behavior of $\{\mu^{\overline{\gamma}}\}$.

Recall the set of strings $\overline{\mathcal{S}}$ in Definition 3.2.1. For an integer $r \geq 1$, let $\mathcal{P}_r$ denote the set of all paths of length r in the sensor graph w.r.t. the adjacency matrix $\mathcal{A}$,

i.e.,

$$\begin{aligned}\mathcal{P}_r &= \{(n_r, \cdots, n_1) \mid n_i \in [1, \cdots, N],\ \forall 1 \le i \le r \text{ and} \\ &\qquad \mathcal{A}_{n_i, n_{i+1}} > 0,\ \forall 1 \le i < r\big\}. \end{aligned} \tag{3.22}$$

To each string $\mathcal{R} = (f_{\jmath_r}, f_{\jmath_{r-1}}, \cdots, f_{\jmath_1}, P^*) \in \mathcal{S}_r^{P^*}$ of length r, defined in Section 3.2, we assign its upper and lower weights respectively as,

$$\overline{w}(\mathcal{R}) = \min_{(n_r, \cdots, n_1) \in \mathcal{P}_r} \sum_{i=1}^{r} \mathbb{I}_{\jmath_i \neq 2^N - 1} \overline{q}_{n_i}(\jmath_i) \tag{3.23}$$

$$\underline{w}(\mathcal{R}) = \min_{(n_r, \cdots, n_1) \in \mathcal{P}_r} \sum_{i=1}^{r} \mathbb{I}_{\jmath_i \neq 2^N - 1} \underline{q}_{n_i}(\jmath_i). \tag{3.24}$$

We set $\overline{w}(\mathcal{R}) = \underline{w}(\mathcal{R}) = 0$, if $r = 0$ in the above.

Note that $|\mathcal{P}_r| < \infty$ for each $r \in \mathbb{N}$; hence $\overline{w}(\cdot)$ and $\underline{w}(\cdot)$ are well-defined extended valued functions mapping from $\mathcal{S}^{P^*}$ to $\mathbb{R}_+$ (we adopt the convention that the minimum of an empty set is ∞).

Finally, define the upper and lower rate functions, $\overline{I}, \underline{I} : \mathbb{S}_+^M \longmapsto \mathbb{R}_+$ by

$$\underline{I}(X) = \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \underline{w}(\mathcal{R}),\ \overline{I}(X) = \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \overline{w}(\mathcal{R}). \tag{3.25}$$

We then have the following large deviation results for the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \to \infty$.

Theorem 3.3.3 Assume that **(C.1)**, **(S.1)**, **(D.1)**, and **(E.4)** hold. Then, as $\overline{\gamma} \to \infty$, the family $\mu^{\overline{\gamma}}$ satisfies the LD upper and lower bounds with rate functions $\overline{I}$ and $\underline{I}$, i.e.,

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(\mathcal{F}) \le - \inf_{X \in \mathcal{F}} \overline{I}(X),\ \text{for every closed set } \mathcal{F} \in \mathcal{X} \tag{3.26}$$

$$\liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(\mathcal{O}) \geq - \inf_{X\in\mathcal{O}} \underline{I}(X), \text{ for every open set } \mathcal{O} \in \mathcal{X}. \tag{3.27}$$

Remark 3.3.4 Theorem 3.3.3 characterizes the invariant measure $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \to \infty$. It establishes the important qualitative behavior of $\{\mu^{\overline{\gamma}}\}$ that rare events decay exponentially when $\overline{\gamma} \to \infty$. For a rare event Γ, from (3.26) and (3.27), we have

$$e^{-\overline{\gamma}(\inf_{X\in\Gamma^\circ} \underline{I}(X))} \leq \mu^{\overline{\gamma}}(\Gamma) \leq e^{-\overline{\gamma}(\inf_{X\in\overline{\Gamma}} \overline{I}(X))}. \tag{3.28}$$

The exact exponent of the exponential decay is bounded within

$$[\overline{\gamma}(\inf_{X\in\overline{\Gamma}} \overline{I}(X)), \overline{\gamma}(\inf_{X\in\Gamma^\circ} \underline{I}(X))].$$

The result suggests how the system designer could trade off estimation accuracy with the communication rate $\overline{\gamma}$. For instance, given a tolerance $\varepsilon > 0$, in order to guarantee the probability of estimation errors lying outside the $\varepsilon-$neighborhood of the optimal centralized estimation error P^* is less than some $\delta > 0$, $\overline{\gamma}$ should be selected according to

$$e^{-\overline{\gamma}\left(\inf_{X\in B_\varepsilon^C(P^*)^\circ} \underline{I}(X)\right)} \leq \mu^{\overline{\gamma}}(B_\varepsilon^C(P^*)) \leq e^{-\overline{\gamma}\left(\inf_{X\in\overline{B_\varepsilon^C(P^*)}} \overline{I}(X)\right)}$$

where $B_\varepsilon^C(P^*)$ is the complement of the open ball $B_\varepsilon(P^*)$. By computing $\inf_{X\in B_\varepsilon^C(P^*)^\circ} \underline{I}(X)$ and $\inf_{X\in\overline{B_\varepsilon^C(P^*)}} \overline{I}(X)$, the designer obtains an estimate of the communication rate $\overline{\gamma}$ required to maintain the probability of outlying errors less than δ.

3.3.2 The Upper and Lower Rate Functions

We define

$$I(X) = \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \pi(\mathcal{R}), \ \forall X \in \mathbb{S}_+^M. \tag{3.29}$$

Recall that $\overline{I}, \underline{I} : \mathbb{S}_+^M \longmapsto \overline{\mathbb{R}}_+$ are defined as

$$\underline{I}(X) = \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \underline{w}(\mathcal{R}),\ \overline{I}(X) = \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \overline{w}(\mathcal{R}),\ \forall X \in \mathbb{S}_+^M. \tag{3.30}$$

The functions $\overline{I}, \underline{I}$ are not generally lower semicontinuous and hence do not qualify as rate functions. However, candidate rate functions for the family of invariant distributions can be the lower semicontinuous regularizations of $\overline{I}, \underline{I}$, which are defined as

$$\begin{aligned} \overline{I}_L(X) &= \lim_{\varepsilon \to \infty} \inf_{Y \in B_\varepsilon(X)} \overline{I}(Y),\ \forall X \in \mathbb{S}_+^M \\ \underline{I}_L(X) &= \lim_{\varepsilon \to \infty} \inf_{Y \in B_\varepsilon(X)} \underline{I}(Y),\ \forall X \in \mathbb{S}_+^M. \end{aligned} \tag{3.31}$$

The following proposition gives some readily verifiable properties of $\overline{I}_L(X)$, whose proof may be obtained from Proposition 6.1 of [31].The semicontinuous regularization $\underline{I}_L(X)$ also has similar properties.

Proposition 3.3.5

(i) The function $\overline{I}_L(X)$ is a good rate function on $\mathbb{S}_+^M$.

(ii) For every $X \in \mathbb{S}_+^M$, $\overline{I}_L(X) = \lim_{\varepsilon \to 0} \inf_{Y \in \overline{B_\varepsilon(X)}} \overline{I}(Y)$.

(iii) For every non-empty set $\Gamma \in \mathcal{B}(\mathbb{S}_+^M)$, $\inf_{X \in \Gamma} \overline{I}_L(X) \leq \inf_{X \in \Gamma} \overline{I}(X)$. In addition, if Γ is open, the reverse inequality holds and thus $\inf_{X \in \Gamma} \overline{I}_L(X) = \inf_{X \in \Gamma} \overline{I}(X)$.

(iv) Let $K \subset \mathbb{S}_+^M$ be a non-empty compact set; then we have $\lim_{\varepsilon \to 0} \inf_{Y \in \overline{K_\varepsilon}} \overline{I}_L(Y) = \inf_{Y \in K} \overline{I}_L(Y)$.

3.3.3 Large Deviation Lower Bound

The following lemma establishes the LD lower bound for the sequence $\{\mu^{\overline{\gamma}}\}$ of invariant distributions as $\overline{\gamma} \to \infty$.

Lemma 3.3.6 Let $\Gamma \in \mathcal{B}(\mathbb{S}_+^M)$; then the following lower bound holds:

$$\liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(\Gamma^\circ) \geq - \inf_{X\in\Gamma^\circ} \underline{I}_L(X). \tag{3.32}$$

Proof 7 Since the sequence $\{P_n(k)\}$ converges weakly (in distribution) to $\mu^{\bar{\gamma}}$, we have

$$\limsup_{k\to\infty} \mathbb{P}(P_n(k) \in F) \leq \mu^{\overline{\gamma}}(F), \; \forall \text{ closed set space } F \subset \mathbb{S}_+^M. \tag{3.33}$$

Consider a measurable set $\Gamma \in \mathcal{B}(\mathbb{S}_+^M)$. Note that if Γ has an empty interior Γ°, the assertion in (3.32) holds trivially since the right-hand side becomes $-\infty$. We thus consider the non-trivial case in which $\Gamma^\circ \neq \emptyset$. Let $X \in \Gamma^\circ \cap \mathcal{D}_{\underline{I}}$, with $\mathcal{D}_{\underline{I}}$ as the effective domain of $\underline{I}(\cdot)$, i.e., the set on which $\underline{I}(\cdot)$ is finite. There exists a small enough $\varepsilon > 0$, such that the closed ball $\overline{B}_\varepsilon(X) \in \Gamma^\circ$. Then, from (3.33), we have

$$\mu^{\overline{\gamma}}(\Gamma^\circ) \geq \mu^{\overline{\gamma}}\left(\overline{B}_\varepsilon(X)\right) \geq \limsup_{t\to\infty} \mathbb{P}\left(P_n(k) \in \overline{B}_\varepsilon(X)\right). \tag{3.34}$$

Now we calculate the right-hand side of (3.34). The set $\mathcal{S}^{P^*}(X)$ is non-empty, due to the fact that $X \in \mathcal{D}_{\underline{I}}$ implying that $\underline{I}(X)$ is finite and $\mathcal{S}^{P^*}(X)$ is non-empty. Hence, for some $r_0 \in \mathbb{T}_+$ and $\jmath_1, ..., \jmath_{r_0} \in \mathfrak{P}$, we have a string $\mathcal{R} = (f_{\jmath_1}, ..., f_{\jmath_{r_0}}, P^*) \in \mathcal{S}^{P^*}(X)$. Define the function g: $\mathbb{S}_+^M \mapsto \mathbb{S}_+^M$ by $g(Y) = f_{\jmath_1} \circ ... \circ f_{\jmath_{r_0}}(Y)$. Since g is continuous, there exists $\varepsilon_1 > 0$ such that

$$\|g(Y) - g(P^*)\| \leq \varepsilon, \forall\, Y \in \overline{B}_{\varepsilon_1}(P^*). \tag{3.35}$$

With Proposition 3.2.6 (ii), for $\varepsilon_1 > 0$, there exists r_{ε_1} such that

$$\left\| f^r_{2^N-1}(Y) - P^* \right\| \leq \varepsilon_1, \forall\ r \geq r_{\varepsilon_1},\ Y \in \mathbb{S}^M_+. \tag{3.36}$$

For any $r \in \mathbb{T}_+$ such that $r \geq r_0 + r_{\varepsilon_1}$ and any string $\mathcal{R}_1 \in \mathcal{S}^{P_0}_r$ of the form

$$\mathcal{R}_1 = \left\{ f_{\jmath_1}, ..., f_{\jmath_{r_0}}, f^{r_{\varepsilon_1}}_{2^N-1}, f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}}, P_0 \right\}$$

where $f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}} \in \mathfrak{P}$, it follows that

$$\begin{aligned} &\|\mathcal{N}(\mathcal{R}_1) - X\| = \|\mathcal{N}(\mathcal{R}_1) - \mathcal{N}(\mathcal{R})\| \\ &= \left\| g\left(f^{r_{\varepsilon_1}}_{2^N-1}(f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}}(P_0)) \right) - g(P^*) \right\| \leq \varepsilon, \end{aligned}$$

which is derived from the fact that

$$\left\| f^{r_{\varepsilon_1}}_{2^N-1}(f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}}(P_0)) - P^* \right\| \leq \varepsilon_1.$$

Therefore,

$$\mathcal{N}(\mathcal{R}_1) \in \overline{B}_\varepsilon(X).$$

For $r \geq r_0 + r_{\varepsilon_1}$, define the set of strings

$$\begin{aligned} \mathcal{R}_t = \Big\{ &\left(f_{\jmath_1}, ..., f_{\jmath_{r_0}}, f^{r_{\varepsilon_1}}_{2^N-1}, f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}}, P_0 \right) \Big| \\ &f_{i_1}, ..., f_{i_{r-r_0-r_{\varepsilon_1}}} \in \mathfrak{P} \Big\}. \end{aligned} \tag{3.37}$$

Then, it follows that $\mathcal{N}(\mathcal{R}_2) \in \overline{B}_\varepsilon(X), \forall\ \mathcal{R}_2 \in \mathcal{R}_t$. Thus, for $r \geq r_0 + r_{\varepsilon_1}$, we have

$$\begin{aligned}
&\mathbb{P}\left(P_n(k) \in \overline{B}_\varepsilon(X)\right) \geq \mathbb{P}\left(P_n(k) \in \mathcal{N}(\mathcal{R}_t)\right) \\
&= \sum_{i_1,\ldots,i_{r-r_0-r_{\varepsilon_1}} \in \mathfrak{P}} \left[\prod_{k=1}^{r_0} q_{n_k}(\jmath_k)\right] \left[\prod_{k=r_0+1}^{r_0+r_{\varepsilon_1}} q_{n_k}(2^N-1)\right] \\
&\quad \left[\prod_{k'=1}^{r-r_0-r_{\varepsilon_1}} q_{n_{k'}}(i_{k'})\right] \\
&= \prod_{k=1}^{r_0} q_n(\jmath_k) \prod_{k=r_0+1}^{r_0+r_{\varepsilon_1}} q_{n_k}(2^N-1). \qquad (3.38)
\end{aligned}$$

From (3.34) and (3.38), there exists

$$\mu^{\overline{\gamma}}(\Gamma^\circ) \geq \prod_{k=1}^{r_0} q_n(\jmath_k) \prod_{k=r_0+1}^{r_0+r_{\varepsilon_1}} q_{n_k}(2^N-1) \qquad (3.39)$$

and hence

$$\begin{aligned}
\ln \mu^{\overline{\gamma}}(\Gamma^\circ) \geq &\sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k \neq 2^N-1} \ln q_{n_k}(\jmath_k) + \sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k = 2^N-1} \ln q_{n_k}(\jmath_k) \\
&+ \sum_{k=r_0+1}^{r_0+r_{\varepsilon_1}} \ln q_{n_k}(2^N-1). \qquad (3.40)
\end{aligned}$$

Since $\lim_{\overline{\gamma}\to\infty} q_{n_k}(2^N-1) = 1$, i.e., the probability of each sensor obtaining the full set of observations through the observation dissemination protocol approaches 1 as

the communication rate $\overline{\gamma} \to \infty$, we have

$$\begin{aligned}\liminf_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(\Gamma^\circ)}{\overline{\gamma}} &\geq \liminf_{\overline{\gamma}\to\infty} \sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k \neq 2^N-1} \frac{1}{\overline{\gamma}} \ln q_{n_k}(\jmath_k) \\ &\geq \sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k \neq 2^N-1} \liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln q_{n_k}(\jmath_k) \\ &\geq -\sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k \neq 2^N-1} \underline{q}_{n_k}(\jmath_k) \end{aligned} \tag{3.41}$$

where the last inequality follows from the fact that $\liminf_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln q_{n_k}(\jmath_k) \geq \underline{q}_{n_k}(\jmath_k)$.

Since the above holds for all $(n_{r_0}, \cdots, n_1) \in \mathcal{P}_{r_0}$, we have

$$\begin{aligned}\liminf_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(\Gamma^\circ)}{\overline{\gamma}} &\geq \max_{(n_{r_0},\cdots,n_1)\in\mathcal{P}_{r_0}} \left\{ -\sum_{k=1}^{r_0} \mathbb{I}_{\jmath_k \neq 2^N-1} \underline{q}_{n_k}(\jmath_k) \right\} \\ &= -\underline{w}(\mathcal{R}) \end{aligned} \tag{3.42}$$

with $\underline{w}(\mathcal{R}) = \min_{(n_{r_0},\cdots,n_1)\in\mathcal{P}_{r_0}} \sum_{i=1}^{r_0} \mathbb{I}_{\jmath_i \neq 2^N-1} \underline{q}_{n_i}(\jmath_i)$.

Given that the above holds for all $\mathcal{R} \in \mathcal{S}^{P^*}(X)$, we have

$$\begin{aligned}\liminf_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(\Gamma^\circ)}{\overline{\gamma}} &\geq \sup_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} (-\underline{w}(\mathcal{R})) \\ &= -\inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \underline{w}(\mathcal{R}) = -\underline{I}(X). \end{aligned} \tag{3.43}$$

Finally, from the fact that for $X \notin D_{\underline{I}}$, $\underline{I}(X) = \infty$, we have

$$\liminf_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(\Gamma^\circ)}{\overline{\gamma}} \geq -\inf_{X\in\Gamma^\circ\cap\mathcal{D}_{\underline{I}}} \underline{I}(X) = -\inf_{X\in\Gamma^\circ} \underline{I}(X).$$

Since Γ° is open, from Proposition 3.3.5 (iii), we have

$$-\inf_{X\in\Gamma^\circ} \underline{I}_L(X) = -\inf_{X\in\Gamma^\circ} \underline{I}(X). \tag{3.44}$$

Thus, the proof is completed.

3.3.4 Large Deviation Upper Bound

In this subsection, we establish the LD upper bound for the family of invariant distributions as $\overline{\gamma} \to \infty$. The proof is divided into three steps. First, we establish the upper bound on compact sets. Then, we derive a tightness result for the family of invariant distributions. Finally, we establish the LD upper bound on the required closed sets.

First, we provide some basic results on the topological properties of strings.

Definition 3.3.7 (Truncated String) Let the string $\mathcal{R}$ be given as $\mathcal{R} = (f_{\jmath_1}, \cdots, f_{\jmath_r}, P_0)$ where $r \in \mathbb{T}_+, \jmath_1, \cdots, \jmath_r \in \mathfrak{P}$. Then for $s \leq r$, the truncated string $\mathcal{R}^s$ of length s is defined as

$$\mathcal{R}^s = (f_{\jmath_1}, \cdots, f_{\jmath_s}, P_0). \tag{3.45}$$

Lemma 3.3.8 Define the set of strings $\mathcal{U} \subset \mathcal{S}^{P^*}$ and the quantities $l(F)$, for a closed set $F \in \mathbb{S}_+^M$, as

$$\mathcal{U}(F) = \left\{\mathcal{R} \in \mathcal{S}^{P^*} | \mathcal{N}(\mathcal{R}) \in F\right\} \tag{3.46}$$

$$l(F) = \inf_{\mathcal{R} \in \mathcal{U}(F)} \pi(\mathcal{R}) \tag{3.47}$$

$$l^{'}(F) = \inf_{\mathcal{R} \in \mathcal{U}(F)} \overline{w}(\mathcal{R}) \tag{3.48}$$

where

$$\overline{w}(\mathcal{R}) = \min_{(n_r, \cdots, n_1) \in \mathcal{P}_r} \sum_{i=1}^{r} \mathbb{I}_{\jmath_i \neq 2^N - 1} \overline{q}_{n_i}(\jmath_i). \tag{3.49}$$

Then, if $l(F) < \infty$ and $l'(F) < \infty$, there exists $r_F \in \mathbb{T}_+$ large enough, such that for all $\mathcal{R} \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}) \geq r_F$, we have $\pi(\mathcal{R}^{r_F}) \geq l(F)$ and $\overline{w}(\mathcal{R}^{r_F}) \geq l'(F)$.

In the statement of Lemma 3.3.8, we assume that the infimum of an empty set is ∞. The proof of Lemma 3.3.8 is provided in Appendix 3.5.2.

From the definition of $\mathcal{U}(F)$, we see that

$$\mathcal{U}(F) = \bigcup_{X \in F} \mathcal{S}^{P^*}(X) \tag{3.50}$$

and hence

$$l(F) = \inf_{X \in F} \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \pi(\mathcal{R}) = \inf_{X \in F} I(X) \tag{3.51}$$

$$l^{'}(F) = \inf_{X \in F} \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \overline{w}(\mathcal{R}) = \inf_{X \in F} \overline{I}(X). \tag{3.52}$$

If $l(F) < \infty$, i.e., the set $\mathcal{U}(F)$ is non-empty, the infimum is attained. That is, there exists $\mathcal{R}^* \in \mathcal{U}(F)$ such that $l(F) = \pi(\mathcal{R}^*)$.

Now we prove the LD upper bound for the family of $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \to \infty$ over compact sets.

Lemma 3.3.9 Let $K \in \mathcal{B}(\mathbb{S}_+^M)$ be a compact set. Then the following upper bound holds:

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(K) \leq - \inf_{X \in K} \overline{I}_L(X). \tag{3.53}$$

The proof is presented in Appendix 3.5.3.

We use the following tightness result to extend the upper bound from compact sets to arbitrary closed sets.

Lemma 3.3.10 The family of invariant distributions $\{\mu^{\overline{\gamma}}\}$ satisfies the following tightness property: For every $a > 0$, there exists a compact set $K_a \subset \mathbb{S}_+^M$ such that,

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}\left(K_a^C\right) \leq -W(a) \tag{3.54}$$

where

$$W(a) = \min_{\substack{(n_1, \cdots, n_{\text{len}(\mathcal{R})}) \in \mathcal{P}_{\text{len}(\mathcal{R})} \\ \mathcal{R} \,:\, \pi(\mathcal{R}) = \lfloor a \rfloor}} \sum_{i=1}^{\text{len}(\mathcal{R})} \mathbb{I}_{\jmath_i \neq 2^N - 1} \overline{q}_{n_i}(\jmath_i). \tag{3.55}$$

The proof is presented in Appendix 3.5.4.

Now we can complete the proof of the LD upper bound for arbitrary closed sets by using the upper bound on compact sets in Lemma 3.3.9 and the tightness result in Lemma 3.3.10.

Lemma 3.3.11 For a closed set $F \in \mathcal{B}(\mathbb{S}_+^M)$, the following upper bound holds:

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F) \leq - \inf_{X \in F} \overline{I}_L(X). \tag{3.56}$$

Proof 8 Let $a > 0$ be arbitrary. By the tightness estimate in Lemma 3.3.10, there exists a compact set $K_a \subset \mathbb{S}_+^M$ such that

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}\left(K_a^C\right) \leq -W(a). \tag{3.57}$$

The set $F \cap K_a$, as the intersection of a closed and a compact set, is compact. Then the LD upper bound in Lemma 3.3.9 holds, and we have

$$\limsup_{\overline{\gamma} \to \infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F \cap K_a) \leq - \inf_{X \in F \cap K_a} \overline{I}_L(X). \tag{3.58}$$

To estimate the probability $\mu^{\overline{\gamma}}(F)$, we use the following decomposition:

$$\mu^{\overline{\gamma}}(F) = \mu^{\overline{\gamma}}(F \cap K_a) + \mu^{\overline{\gamma}}(F \cap K_a^C) \leq \mu^{\overline{\gamma}}(F \cap K_a) + \mu^{\overline{\gamma}}(K_a^C). \tag{3.59}$$

From the results on the limits of real number sequences (see Lemma 1.2.15 of [14]), we have

$$\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F) \leq$$
$$\max\left(\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F \cap K_a), \limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}\left(K_a^C\right)\right).$$

From (3.57) and (3.58), we have

$$\begin{aligned}\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F) &\leq \max\left(-\inf_{X\in F\cap K_a} \overline{I}_L(X), -W(a)\right) \\ &\leq \max\left(-\inf_{X\in F} \overline{I}_L(X), -W(a)\right) \\ &= -\min\left(\inf_{X\in F} \overline{I}_L(X), W(a)\right).\end{aligned}$$

Since the above inequality holds for an arbitrary $a > 0$, taking the limit as $a \to \infty$ on both sides together with $W(a) \to \infty$, we have

$$\begin{aligned}\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(F) &\leq \max\left(-\inf_{X\in F\cap K_a} \overline{I}_L(X), -W(a)\right) \\ &\leq -\inf_{X\in F} \overline{I}_L(X). \end{aligned} \tag{3.60}$$

3.4 Simulation Results

In this section, we simulate the M-GIKF to estimate a 10-dimensional state-unknown system[4] with a network of 5 sensors. The matrices $\mathcal{F}$, $\mathcal{C}_n$, and $\mathcal{Q}$ satisfy Assumptions S.1 and D.1. The simulation is based on the example of distributed observation dissemination protocol discussed in Section 2.2.3, in which this protocol does not use knowledge of global topology and is the simplest random walk on the

[4]We acknowledge that this is not a large system size; it just illustrates the concept.

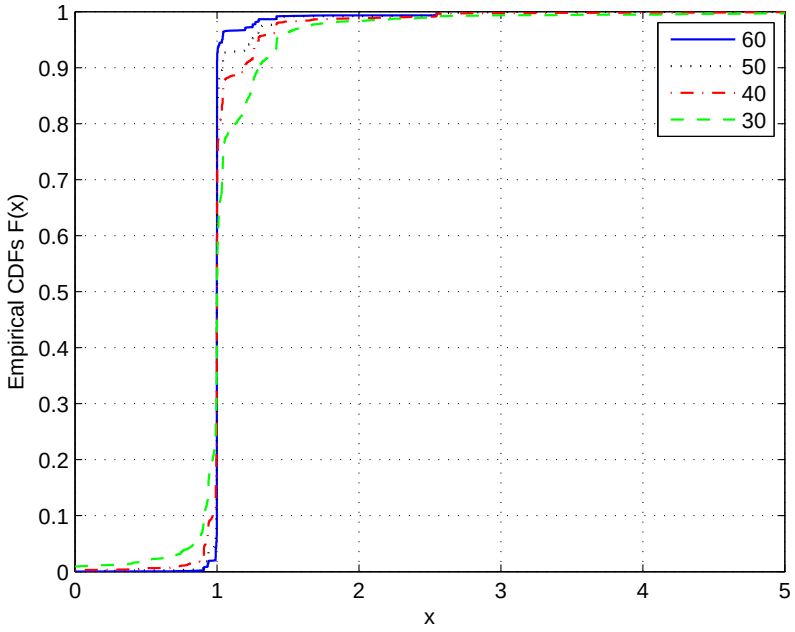

Figure 3.1: CDF of the normalized largest eigenvalue from $\mu^{\overline{\gamma}}$ for varying $\overline{\gamma} = 30, 40, 50, 60$.

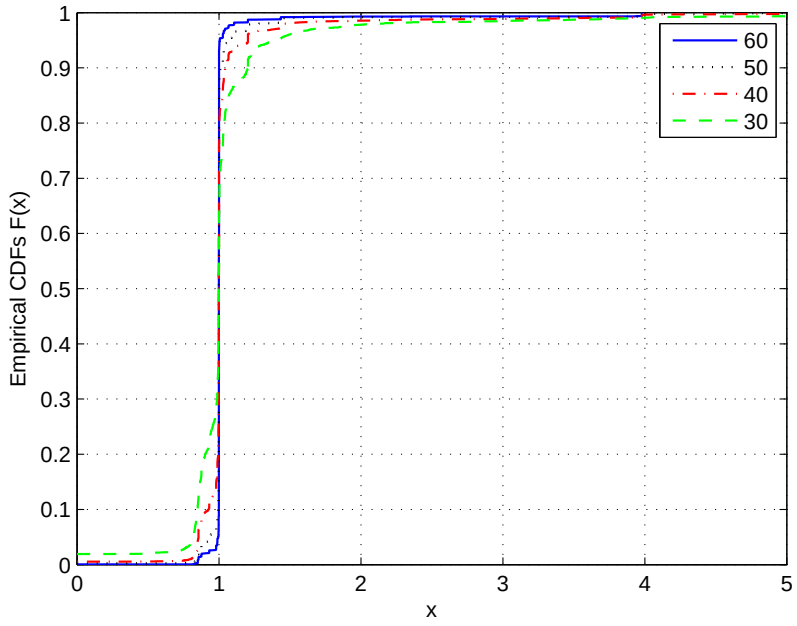

Figure 3.2: CDF of the normalized trace from $\mu^{\overline{\gamma}}$ for varying $\overline{\gamma} = 30, 40, 50, 60$

graph with uniform (unoptimized) neighbor selection. By tuning the link selection probabilities (using full knowledge of global topology), it could be possible to perform better. The protocol in Section 2.2.3 is just an example of possible protocols, while the theoretical analysis in our paper is protocol independent.

We study the behavior of $\mu^{\overline{\gamma}}$ for different values of $\overline{\gamma}$. We iterate the RRE 10^4 times to ensure the error covariance sequence at a randomly selected sensor converged in distribution to $\mu^{\overline{\gamma}}$ as shown in Theorems 2.3.5 and 2.3.6, where we simulate $5,000$ samples for each $\overline{\gamma}$. In order to graphically present the distribution for the covariance matrix, we focus on its largest eigenvalue and trace here. The resulting empirical Cumulative Distribution Functions (CDFs) of the normalized largest eigenvalue $\lambda_{\max}(\cdot)$ and the normalized trace $\mathrm{Tr}(\cdot)$ (which is the conditional mean-squared error) of the error covariance matrices are plotted in Fig. 3.1 and

Fig. 3.2, where the x-axis is $\lambda_{\max}(\cdot)/\lambda_{\max}(P^*)$ and $\mathrm{Tr}(\cdot)/\mathrm{Tr}(P^*)$, respectively. As $\overline{\gamma}$ increases, we see that the empirical measure $\mu^{\overline{\gamma}}$ converges in distribution to the Dirac measure δ_{P^*} of P^*.

Then we simulate the LD decay exponent of the rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$ with $\varepsilon = \mathrm{Tr}(P^*)/2$, and the LD decay exponent of the rare event $B_{\varepsilon'}^C(\lambda_{\max}(P^*))$ with $\varepsilon' = \lambda_{\max}(P^*)/2$. For each $\overline{\gamma}$, we estimate the LD decay exponents $\frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(B_\varepsilon^C(\lambda_{\max}(P^*)))$ and $\frac{1}{\overline{\gamma}} \ln \mu^{\overline{\gamma}}(B_\varepsilon^C(\mathrm{Tr}(P^*)))$ by using the samples obtained above for calculating the empirical CDFs. Then we take effort to numerically calculate the LD lower and upper bounds. From Theorem 3.3.3, the LD upper bound for the rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$ can be obtained as the negative infimum of $\overline{I}(\cdot)$ over the set of rare events, and the LD lower bound for the rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$ can be obtained as the negative infimum of $\underline{I}(\cdot)$ over the set of rare events. Recall (3.23), (3.25), and (3.26), we present the LD upper bound for the rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$ as

$$- \inf_{\mathrm{Tr}(X) \in \overline{B_\varepsilon^C(\mathrm{Tr}(P^*))}} \inf_{\mathcal{R} \in \mathcal{S}^{P^*}(X)} \min_{(n_{\mathrm{len}(\mathcal{R})}, \cdots, n_1) \in \mathcal{P}_{\mathrm{len}(\mathcal{R})}} \sum_{i=1}^{\mathrm{len}(\mathcal{R})} \mathbb{I}_{\jmath_i \neq 2^N - 1} \overline{q}_{n_i}(\jmath_i), \ X \in \mathbb{S}_+^M, \tag{3.61}$$

where $\overline{q}_{n_i}(\jmath_i)$ is defined in (2.21). Now we present one way to set α in (2.21). Recall Section 2.2.3, where T_i is the hitting time starting from sensor i to another particular sensor n in the Markov chain with transition matrix $\overline{A^o}$. Then we have

$$P(T_i > L) = \sum_{n_1, \cdots, n_L \neq n} \overline{A^o}_{n_1 n_2} \overline{A^o}_{n_2 n_3} \cdots \overline{A^o}_{n_{L-1} n_L}, \tag{3.62}$$

and α can be selected as $\alpha = \max_i P(T_i > L)$.

For the LD lower bound of the rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$, recall (3.24), (3.25), and

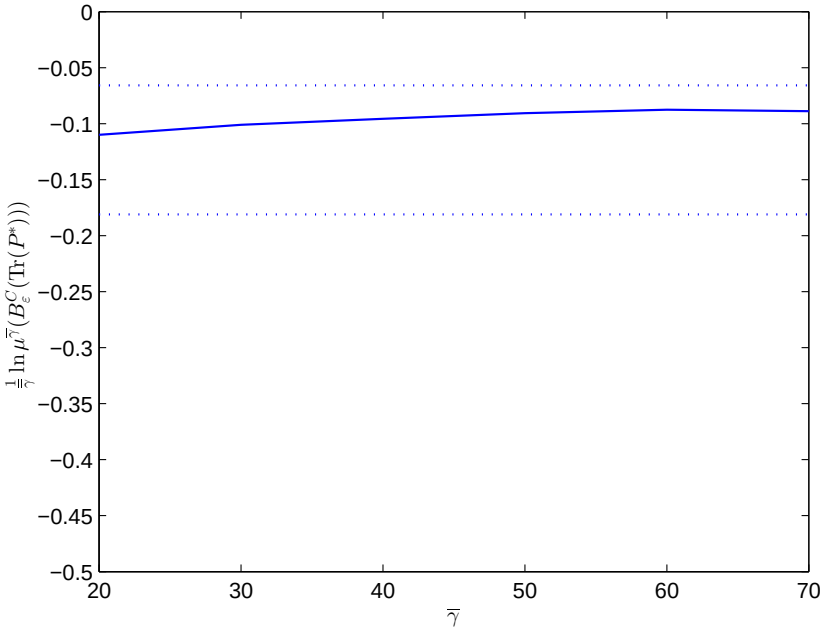

Figure 3.3: LD decay exponent for probability of rare event $B_\varepsilon^C(\mathrm{Tr}(P^*))$ and the LD upper and lower bounds.

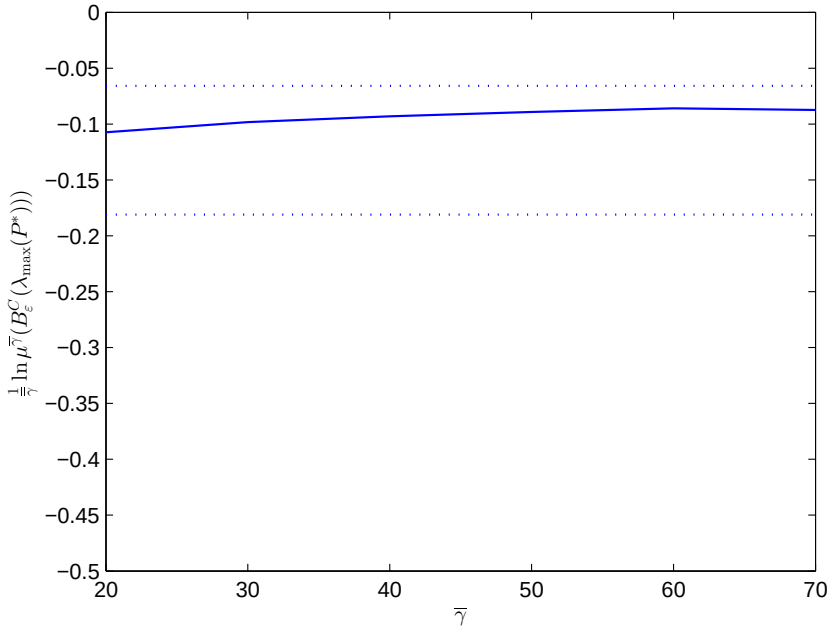

Figure 3.4: LD decay exponent for probability of rare event $B_{\varepsilon'}^{C}(\lambda_{\max}(P^*))$ and the LD upper and lower bounds.

(3.27), we present the LD lower bound as

$$-\inf_{\mathrm{Tr}(X)\in B_{\varepsilon}^{C}(\mathrm{Tr}(P^*))^o}\ \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)}\ \min_{(n_{\mathrm{len}(\mathcal{R})},\cdots,n_1)\in\mathcal{P}_{\mathrm{len}(\mathcal{R})}} \sum_{i=1}^{\mathrm{len}(\mathcal{R})} \mathbb{I}_{\jmath_i\neq 2^N-1}\underline{q}_{n_i}(\jmath_i),\ X\in\mathbb{S}_{+}^{M}, \tag{3.63}$$

where $\underline{q}_{n_i}(\jmath_i)$ is defined in (2.21). We could set β in (2.21) as $\beta=\min_i P(T_i>L)$.

Then the problems of computing the LD upper and lower bounds could be converted to solving the optimization problems in (3.61) and (3.63) respectively. We could then apply some search method to numerically solve those problems to obtain the LD upper and lower bounds. The same analysis could be applied for the case of $B_{\varepsilon}^{C}(\lambda_{\max}(P^*))$.

Fig. 3.3 and Fig. 3.4 display the estimated LD decay exponents of the rare events of $B_{\varepsilon}^{C}(\mathrm{Tr}(P^*))$ and $B_{\varepsilon'}^{C}(\lambda_{\max}(P^*))$ for different values of $\overline{\gamma}$, and the corresponding LD upper and lower bounds of the decay exponents, respectively. The empirically estimated decay exponents in these two rare events perform quite similar, which is due to the fact that ε and ε' have the same relative factor 0.5 for the maximum eigenvalue and the trace of P^*, respectively.

Finally, note that the convergence rate with respect to $\overline{\gamma}$ (i.e., the large deviation exponent) may be improved by considering a more sophisticated observation dissemination protocol. For instance, the neighbor selection probabilities in the observation dissemination protocol from Section 2.2.3 may be optimized for a given communication network structure. This could lead to a faster mixing Markov chain governing the observation dissemination and, hence, for the same rate $\overline{\gamma}$, a sensor could more likely receive the observations of more sensors in each epoch (also see Remark 2.2.4), leading to faster convergence of $\{\mu^{\overline{\gamma}}\}$ to δ_{P^*}.

3.5 Appendices

3.5.1 Proof of Theorem 3.3.1

First define the following class of sets:

$$C = \{\mathcal{F} | \mathcal{F} \text{ is closed and } P^* \in \mathcal{F}\}. \tag{3.64}$$

Then proving this theorem is equivalent to proving the following:

$$\lim_{\overline{\gamma}\to\infty} d_P(\mu^{\overline{\gamma}}, \delta_{P^*}) = 0 \tag{3.65}$$

where the Prohorov metric $d_P(\mu^{\overline{\gamma}}, \delta_{P^*}) = \inf\{\varepsilon > 0 \mid \mu^{\overline{\gamma}}(\mathcal{F}_\varepsilon) + \varepsilon \geq 1,\ \forall \mathcal{F} \in C\}$ is defined in [24].

Consider $0 < \varepsilon < 1$ small enough. Then there exists a $\varepsilon_0 > 0$ such that, for every $\mathcal{F} \in C$, we have $B_{\varepsilon_0}(P^*) \subset \mathcal{F}_\varepsilon$. The numerical value of the string $\mathcal{R} = P^*$ belongs to $B_{\varepsilon_0}(P^*)$, and hence by (3.39), there exists an integer r^* such that

$$\begin{aligned}\mu^{\overline{\gamma}}(B_{\varepsilon_0}(P^*)) &\geq \prod_{k=1}^{r_0} q_n(\jmath_k) \prod_{k=r_0+1}^{r_0+r_{\varepsilon_1}} q_{n_k}(2^N - 1) \\ &= \prod_{k=1}^{r^*} q_{n_k}(2^N - 1), \quad \text{where } r^* = r_0 + r_{\varepsilon_1}.\end{aligned}$$

Thus, for all $\mathcal{F} \in C$, we have

$$\mu^{\overline{\gamma}}(\mathcal{F}_\varepsilon) \geq \mu^{\overline{\gamma}}(B_{\varepsilon_0}(P^*)) \geq \prod_{k=1}^{r^*} q_{n_k}(2^N - 1). \tag{3.66}$$

Since $q_{n_k}(2^N-1) \to 1$ as $\overline{\gamma} \to \infty$, we have for $\overline{\gamma} \to \infty$

$$\mu^{\overline{\gamma}}(\mathcal{F}_\varepsilon) + \varepsilon \geq \prod_{k=1}^{r^*} q_{n_k}(2^N-1) + \varepsilon \geq 1. \tag{3.67}$$

Then, following the definition of $d_P(\mu^{\overline{\gamma}}, \delta_{P^*})$, when $\overline{\gamma} \to \infty$, we have

$$d_P(\mu^{\overline{\gamma}}, \delta_{P^*}) \leq \varepsilon, \ \overline{\gamma} \to \infty. \tag{3.68}$$

Hence,

$$\lim_{\overline{\gamma}\to\infty} d_P(\mu^{\overline{\gamma}}, \delta_{P^*}) \leq \varepsilon. \tag{3.69}$$

Since $\varepsilon > 0$ is arbitrary, by considering the limit as $\varepsilon \to 0$, we conclude that $\lim_{\overline{\gamma}\to\infty} d_P(\mu^{\overline{\gamma}}, \delta_{P^*}) = 0$.

3.5.2 Proof of Lemma 3.3.8

We first prove, if $l'(F) < \infty$, there exists $r_F \in \mathbb{T}_+$ large enough, such that for all $\mathcal{R} \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}) \geq r_F$, we have $\overline{w}(\mathcal{R}^{r_F}) \geq l'(F)$. Then the proof for the case of $l(F)$ naturally follows.

The case $l'(F) = 0$ is trivial, by choosing an arbitrary positive r_F. Consider the case $l'(F) \geq \overline{q}$, where $\overline{q} = \min_{1\leq n\leq N, \jmath\in\mathfrak{P}} \overline{q}_n(\jmath)$. Using an inductive argument, it suffices to show that for every $\overline{q} \leq i \leq l'(F)$, there exists a positive $r_F^i \in \mathbb{T}_+$ such that, for $\mathcal{R} \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}) \geq r_F^i$, we have

$$\overline{w}\left(\mathcal{R}^{r_F^i}\right) \geq i. \tag{3.70}$$

First we consider the case $i = \overline{q}$. We assume on the contrary that there is no such $r_F^{\overline{q}} \in \mathbb{T}_+$ for which the above property holds. Since $\mathcal{U}(F)$ is not empty, by

Proposition 3.2.5 (i), there exists $r_0 \in \mathbb{T}_+$ such that

$$\mathcal{S}_r^{P^*} \cap \mathcal{U}(F) \neq \emptyset, \ \forall r \geq r_0. \tag{3.71}$$

Thus, the non-existence of $r_F^{\overline{q}}$ implies that, for every $r \geq r_0$, there exists a string $\mathcal{R}_r \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}_r) \geq r$, such that $\overline{w}(\mathcal{R}_r^r) = 0$. Therefore, such $\mathcal{R}_r$ is of the form

$$\mathcal{R}_r = \left(f_{2^N-1}^r, f_{\jmath_1}, \cdots, f_{\jmath_{\text{len}(\mathcal{R}_r)-r}}, P^*\right) \tag{3.72}$$

where $\jmath_1, \cdots, \jmath_{\text{len}(\mathcal{R}_r)-r} \in \mathfrak{P}$. Thus, by denoting

$$X_r = f_{\jmath_1} \circ \cdots \circ f_{\jmath_{\text{len}(\mathcal{R}_r)-r}}(P^*), \tag{3.73}$$

we have $\mathcal{N}(\mathcal{R}_r) = f_{2^N-1}^r(X_r)$. By Proposition 3.2.6 (ii), the uniform convergence of the Riccati iterates implies that, for an arbitrary $\varepsilon > 0$, there exists $r_\varepsilon \geq M$, such that, for every $X \in \mathbb{S}_+^M$,

$$\left\| f_{2^N-1}^r(X) - P^* \right\| \leq \varepsilon, \ r \geq r_\varepsilon \tag{3.74}$$

where the constant r_ε can be chosen independently of X. Then, by defining $r'_\varepsilon = \max(r_0, r_\varepsilon)$, we have

$$\|\mathcal{N}(\mathcal{R}_r) - P^*\| = \left\| f_{2^N-1}^r(X_r) - P^* \right\| \leq \varepsilon, \ r \geq r'_\varepsilon. \tag{3.75}$$

Since ε is arbitrary, the above result shows that the sequence $\{\mathcal{N}(\mathcal{R}_r)\}_{r \geq r'_\varepsilon}$ of numerical results converges to P^* as $r \to \infty$. By construction, the sequence $\{\mathcal{N}(\mathcal{R}_r)\}_{r \geq r'_\varepsilon}$ belongs to the set F, and we conclude that P^* is a limit point of the set F. Since F

is closed, we have $P^* \in F$, which implies

$$\left\{\mathcal{R} \in \mathcal{S}^{P^*} | \mathcal{N}(\mathcal{R}) = P^*\right\} \subset \mathcal{U}(F). \tag{3.76}$$

Hence, specifically, $(f_{2^N-1}, P^*) \in \mathcal{U}(F)$. Thus the fact that $\overline{w}\left((f_{2^N-1}, P^*)\right) = 0$ contradicts the hypothesis $l'(F) \geq \overline{q}$.

Therefore, we establish that, if $l'(F) \geq \overline{q}$, there exists $r_F^{\overline{q}}$ satisfying the property in (3.70) for $i = \overline{q}$. Note here that, if $l'(F) = \overline{q}$, this step has completed the proof of the lemma. In the general case, to establish (3.70) for all $\overline{q} \leq i \leq l'(F)$, we need the following additional steps.

Let us now assume $l'(F) \geq 2\overline{q}$. We further assume on the contrary that the claim in (3.70) does not hold for any $\overline{q} \leq i \leq l'(F)$. By the previous step, clearly the claim holds for $i = \overline{q}$. Then, let k, $\overline{q} \leq k < l'(F)$, be the largest number such that the claim in (3.70) holds for all $\overline{q} \leq i \leq k$, which implies that there exists no $r_F^{k+\overline{q}} \in \mathbb{T}_+$ satisfying the claim in (3.70) for $i = k + \overline{q}$. Since the claim holds for $i = k$, there exists $r_F^k \in \mathbb{T}_+$ such that, for all $\mathcal{R} \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}) \geq r_F^k$, we have $\overline{w}(\mathcal{R}^{r_F^k}) \geq k$. The non-existence of $r_F^{k+\overline{q}}$ and (3.71) imply that, for every $r \geq r_0$, there exists a string $\mathcal{R}_r \in \mathcal{U}(F)$ with $\text{len}(\mathcal{R}_r) \geq r$ such that $\overline{w}(\mathcal{R}_r^r) < k + \overline{q}$.

Define $r_0' = \max(r_0, r_F^k)$, then by the existence of r_F^k and $\overline{w}(\mathcal{R}_r^r) < k + \overline{q}$, we have $\overline{w}(\mathcal{R}_r^r) = k$ for $r \geq r_0'$. Therefore, for $r \geq r_0'$, $\mathcal{R}_r$ is necessarily of the form

$$\mathcal{R}_r = \left(f_{\jmath_1}, \cdots, f_{\jmath_{r_F^k}}, f_{2^N-1}^{r-r_F^k}, f_{i_1}, \cdots, f_{i_{\text{len}(\mathcal{R}_r)-r}}, P^*\right)$$

where $\jmath_1, \cdots, \jmath_{r_f^k} \in \mathfrak{P}$ such that $\overline{w}(\mathcal{R}_r^{r_F^k}) = k$ and $i_1, \cdots, i_{\text{len}(\mathcal{R}_r)-r} \in \mathfrak{P}$.

Now consider the sequence $\{\mathcal{R}_r\}_{r \geq r_0'}$. Define the set $\mathcal{J}$ as $\mathcal{J} = \{\mathcal{R}_r,\ r \geq r_0'\}$, and also define the set $\mathcal{J}_1$ as $\mathcal{J}_1 = \{\mathcal{R} \in \mathcal{S}_{r_F^k}^{P^*} | \overline{w}(\mathcal{R}) = k\}$. Consider the mapping

$\Theta^{r_F^k} : \mathcal{J} \mapsto \mathcal{J}_1$ by

$$\Theta^{r_F^k}(\mathcal{R}) = \mathcal{R}^{r_F^k},\ \forall \mathcal{R} \in \mathcal{J}. \tag{3.77}$$

Since the cardinality of the set $\mathcal{J}_1$ is finite and the set $\mathcal{J}$ is countably infinite, for a specific $\mathcal{R}' \in \mathcal{J}_1$, the set $\left(\Theta^{r_F^k}\right)^{-1}(\mathcal{R}')$ is countably infinite. This in turn implies that we can extract a subsequence $\{\mathcal{R}_{r_m}\}_{m\geq 0}$ from the sequence $\{\mathcal{R}_r\}_{r\geq r_0'}$, such that

$$\mathcal{R}_{r_m}^{r_F^k} = \mathcal{R}',\ \forall m \geq 0. \tag{3.78}$$

In other words, if $\mathcal{R}'$ is represented by $\mathcal{R}' = \left(f_{\jmath_1'}, \cdots, f_{\jmath'_{r_F^k}}, P^*\right)$ for some fixed $\jmath_1', \cdots, \jmath'_{r_F^k} \in \mathfrak{P}$, for each m the string $\mathcal{R}_{r_m}$ is of the form

$$\mathcal{R}_{r_m} = \left(f_{\jmath_1'}, \cdots, f_{\jmath'_{r_F^k}}, f_{2^N-1}^{r_m - r_F^k}, f_{i_1}, \cdots, f_{i_{\text{len}(\mathcal{R}_{r_m})-r_m}}, P^*\right)$$

where $i_1, \cdots, i_{\text{len}(\mathcal{R}_{r_m})-r_m} \in \mathfrak{P}$ are arbitrary. We denote by

$$X_m = f_{i_1} \circ \cdots \circ f_{i_{\text{len}(\mathcal{R}_{r_m})-r_m}}(P^*),\ \forall m, \tag{3.79}$$

and we have

$$\mathcal{N}(\mathcal{R}_{r_m}) = f_{\jmath_1'} \circ \cdots \circ f_{\jmath'_{r_F^k}}\left(f_{2^N-1}^{r_m - r_F^k}(X_m)\right). \tag{3.80}$$

Since $r_m \to \infty$ as $m \to \infty$, by Proposition 14 (ii), we have

$$\lim_{m\to\infty} f_{2^N-1}^{r_m - r_F^k}(X_m) = P^*. \tag{3.81}$$

Note that the function $f_{\jmath_1'} \circ \cdots \circ f_{\jmath'_{r_F^k}} : \mathbb{S}_+^M \mapsto \mathbb{S}_+^M$, being the finite composition of

continuous functions, is continuous. We then have

$$\begin{aligned}\lim_{m\to\infty}\mathcal{N}(\mathcal{R}_{r_m}) &= \lim_{m\to\infty} f_{J'_1}\circ\cdots\circ f_{J'_{r_F^k}}\left(f_{2^N-1}^{r_m-r_F^k}(X_m)\right)\\ &= f_{J'_1}\circ\cdots\circ f_{J'_{r_F^k}}\left(\lim_{m\to\infty} f_{2^N-1}^{r_m-r_F^k}(X_m)\right)\\ &= f_{J'_1}\circ\cdots\circ f_{J'_{r_F^k}}(P^*)\\ &= \mathcal{N}(\mathcal{R}'). \end{aligned} \tag{3.82}$$

Therefore, the sequence $\{\mathcal{N}(\mathcal{R}_{r_m})\}_{m\geq 0}$ in F converges to $\mathcal{N}(\mathcal{R}')$ as $m\to\infty$. Hence $\mathcal{N}(\mathcal{R}')$ is a limit point in F, and $\mathcal{N}(\mathcal{R}')\in F$ as F is closed. This implies that $\mathcal{R}'\in\mathcal{U}_F$. Since $\overline{w}(\mathcal{R}')=k$ and $\mathcal{R}'\in\mathcal{U}_F$, this contradicts the hypothesis that $k<l'(F)$ and thus the claim in (3.70) holds for all $\overline{q}\leq i\leq l'(F)$.

To prove, if $l(F)<\infty$, there exists $r_F\in\mathbb{T}_+$ large enough, such that for all $\mathcal{R}\in\mathcal{U}(F)$ with $\text{len}(\mathcal{R})\geq r_F$, we have $\pi(\mathcal{R}^{r_F})\geq l(F)$, the method is the same as above, where $l(F)$ becomes a non-negative integer. We choose r_F as the maximum one in these two cases, then the Lemma is proved.

3.5.3 Proof of Lemma 3.3.9

For $\varepsilon>0$, define K_ε as the ε-neighborhood of K and $\overline{K_\epsilon}$ as its ε-closure, i.e.,

$$K_\varepsilon=\left\{X\in\mathbb{S}_+^M|\inf_{Y\in K}\|X-Y\|<\varepsilon\right\} \tag{3.83}$$

$$\overline{K_\varepsilon}=\left\{X\in\mathbb{S}_+^M|\inf_{Y\in K}\|X-Y\|\leq\varepsilon\right\}. \tag{3.84}$$

Since K_ε is open, by the weak convergence of the sequence $\{P_n(r)\}$ to $\mu^{\overline{\gamma}}$, we have

$$\liminf_{r\to\infty}\mathbb{P}\left(P_n(r)\in K_\varepsilon\right)\geq\mu^{\overline{\gamma}}(K_\varepsilon), \tag{3.85}$$

which implies that

$$\liminf_{r\to\infty} \mathbb{P}\left(P_n(r) \in \overline{K_\varepsilon}\right) \geq \mu^{\overline{\gamma}}(K). \tag{3.86}$$

Now we calculate the left-hand side of (3.86). Since $\overline{K_\varepsilon}$ is closed, the results of Lemma 3.3.8 apply. Recall the definition of $\mathcal{U}(F)$. Also, for every $r \in \mathbb{T}_+$ and the closed set F, we define

$$\mathcal{U}^r(F) = \mathcal{U}(F) \cap \mathcal{S}_r^{P^*}. \tag{3.87}$$

We consider first $l(K) < \infty$ and $l'(K) < \infty$ (i.e., $\mathcal{U}(K)$ is non-empty). We then have

$$\mathbb{P}\left(P_n(r) \in \overline{K_\varepsilon}\right) = \mathbb{P}\left(P_n(r) \in \mathcal{N}\left(\mathcal{U}^r\left(\overline{K_\varepsilon}\right)\right)\right). \tag{3.88}$$

Since $K \subset \overline{K_\varepsilon}$ and $l(K) < \infty$, we have $l(\overline{K_\varepsilon}) < \infty$. Thus, since $\overline{K_\varepsilon}$ is closed, Lemma 3.3.8 shows that there exists $r_{\overline{K_\varepsilon}} \in \mathbb{T}_+$, such that, for any string $\mathcal{R} \in \mathcal{U}(\overline{K_\varepsilon})$ with $\text{len}(\mathcal{R}) \geq r_{\overline{K_\varepsilon}}$, we have $\pi\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l(\overline{K_\varepsilon})$ and $\overline{w}\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l'(\overline{K_\varepsilon})$. In other words, for all $r \geq r_{\overline{K_\varepsilon}}$, we have

$$\pi\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l(\overline{K_\varepsilon}),\ \overline{w}\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l'(\overline{K_\varepsilon}),\ \forall\ \mathcal{R} \in \mathcal{U}^r(\overline{K_\varepsilon}). \tag{3.89}$$

Now consider $r \geq r_{\overline{K_\varepsilon}}$ and define $\mathcal{J}_r^{P^*}$ as the set of strings

$$\mathcal{J}_r^{P^*} = \left\{\mathcal{R} \in \mathcal{S}_r^{P^*} | \pi\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l(\overline{K_\varepsilon}),\ \overline{w}\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right) \geq l'(\overline{K_\varepsilon})\right\}. \tag{3.90}$$

The set $\mathcal{J}_r^{P^*}$ consists of all strings $\mathcal{R}$ with length r such that we have at least $l(\overline{K_\varepsilon})$ occurrences of non-f_{2^N-1} and the value of $\overline{w}$ no less than $l'(\overline{K_\varepsilon})$ in the truncated string $\mathcal{R}^{r_{\overline{K_\varepsilon}}}$.

For $r \geq r_{\overline{K_\varepsilon}}$, it is obvious that the following holds:

$$\mathcal{U}^r(\overline{K_\varepsilon}) \subset \mathcal{J}_r^{P^*} \subset \mathcal{S}_r^{P^*}. \tag{3.91}$$

Clearly, we have that, for $r \geq r_{\overline{K_\varepsilon}}$,

$$\begin{aligned}
&\mathbb{P}\left(P_n(r) \in \mathcal{N}(\mathcal{J}_r^{P^*})\right) \\
&= \sum_{\mathcal{R} \in \mathcal{J}_r^{P^*}} \prod_{k=1}^{r} q_{n_k}(\jmath_k) \leq (2^N - 1)^{\underline{l}(\overline{K_\varepsilon})} \begin{pmatrix} r_{\overline{K_\varepsilon}} \\ \underline{l}(\overline{K_\varepsilon}) \end{pmatrix} \\
&\quad \times \max_{\substack{(n_1, \cdots, n_{r_{\overline{K_\varepsilon}}}) \in \mathcal{P}_{r_{\overline{K_\varepsilon}}} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_{\overline{K_\varepsilon}}}) = \underline{l}(\overline{K_\varepsilon})}} \prod_{k=1}^{r_{\overline{K_\varepsilon}}} q_{n_k}(\jmath_k | \jmath_k \neq 2^N - 1),
\end{aligned} \tag{3.92}$$

where $\underline{l}(\overline{K_\varepsilon}) = \min_{\mathcal{R} \in \mathcal{J}_r^{P^*}} \pi\left(\mathcal{R}^{r_{\overline{K_\varepsilon}}}\right)$. Then, from (3.88) and (3.91), we have

$$\begin{aligned}
\mu^{\overline{\gamma}}(K) &\leq \liminf_{r \to \infty} \mathbb{P}\left(P_n(r) \in \mathcal{N}\left(\mathcal{U}^r(\overline{K_\varepsilon})\right)\right) \\
&\leq \liminf_{r \to \infty} \mathbb{P}\left(P_n(r) \in \mathcal{N}(\mathcal{J}_r^{P^*})\right) \\
&\leq (2^N - 1)^{\underline{l}(\overline{K_\varepsilon})} \begin{pmatrix} r_{\overline{K_\varepsilon}} \\ \underline{l}(\overline{K_\varepsilon}) \end{pmatrix} \\
&\qquad \max_{\substack{(n_1, \cdots, n_{r_{\overline{K_\varepsilon}}}) \in \mathcal{P}_{r_{\overline{K_\varepsilon}}} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_{\overline{K_\varepsilon}}}) = \underline{l}(\overline{K_\varepsilon})}} \prod_{k=1}^{r_{\overline{K_\varepsilon}}} q_{n_k}(\jmath_k | \jmath_k \neq 2^N - 1).
\end{aligned}$$

Taking the logarithm, dividing by $\overline{\gamma}$ on both sides, and taking the limits, we have

$$\begin{aligned}
&\limsup_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(K)}{\overline{\gamma}} \\
&\leq \limsup_{\overline{\gamma}\to\infty} \max_{\substack{(n_1,\cdots,n_{r_{\overline{K_\varepsilon}}})\in \mathcal{P}_{r_{\overline{K_\varepsilon}}} \\ \mathcal{R}\,:\,\pi(\mathcal{R}^{r_{\overline{K_\varepsilon}}})=\underline{l}(\overline{K_\varepsilon})}} \sum_{k=1}^{r_{\overline{K_\varepsilon}}} \mathbb{I}_{\jmath_k\neq 2^N-1} \frac{1}{\overline{\gamma}} \ln q_{n_k}(\jmath_k) \\
&\leq \max_{\substack{(n_1,\cdots,n_{r_{\overline{K_\varepsilon}}})\in \mathcal{P}_{r_{\overline{K_\varepsilon}}} \\ \mathcal{R}\,:\,\pi(\mathcal{R}^{r_{\overline{K_\varepsilon}}})=\underline{l}(\overline{K_\varepsilon})}} \sum_{k=1}^{r_{\overline{K_\varepsilon}}} \mathbb{I}_{\jmath_k\neq 2^N-1} \left(\limsup_{\overline{\gamma}\to\infty} \frac{1}{\overline{\gamma}} \ln q_{n_k}(\jmath_k) \right) \\
&\leq - \min_{\substack{(n_1,\cdots,n_{r_{\overline{K_\varepsilon}}})\in \mathcal{P}_{r_{\overline{K_\varepsilon}}} \\ \mathcal{R}\,:\,\pi(\mathcal{R}^{r_{\overline{K_\varepsilon}}})=\underline{l}(\overline{K_\varepsilon})}} \sum_{k=1}^{r_{\overline{K_\varepsilon}}} \mathbb{I}_{\jmath_k\neq 2^N-1} \overline{q}_{n_k}(\jmath_k) \leq -l^{'}(\overline{K_\varepsilon}).
\end{aligned}$$

Then, taking the limit as $\varepsilon \to 0$ on both sides leads to

$$\limsup_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(K)}{\overline{\gamma}} \leq -\lim_{\varepsilon\to 0} l'(\overline{K_\varepsilon}). \tag{3.93}$$

From Proposition 3.3.5 (iii), we have

$$l'(\overline{K_\varepsilon}) = \inf_{X\in\overline{K_\varepsilon}} \overline{I}(X) \geq \inf_{X\in\overline{K_\varepsilon}} \overline{I}_L(X) \tag{3.94}$$

where $\overline{I}(X) = \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \overline{w}(\mathcal{R})$.

Again, taking the limit as $\varepsilon \to 0$ and from Proposition 3.3.5 (iv), we have

$$\lim_{\varepsilon\to 0} l'(\overline{K_\varepsilon}) \geq \lim_{\varepsilon\to 0} \inf_{X\in\overline{K_\varepsilon}} \overline{I}_L(X) = \inf_{X\in K} \overline{I}_L(X). \tag{3.95}$$

The lemma then follows from (3.93) and (3.95).

3.5.4 Proof of Lemma 3.3.10

Let $a > 0$ be arbitrary and choose $z \in \mathbb{N}$ such that $z \geq a$. From Proposition 3.2.5 (ii), there exists $\alpha_{P^*} \in \mathbb{R}_+$ depending on P^* only, such that

$$f_{\hat{\jmath}_{\pi(\mathcal{R})}} \circ f_{\hat{\jmath}_{\pi(\mathcal{R})-1}} \cdots \circ f_{\hat{\jmath}_1} (\alpha_{P^*} I) \succeq \mathcal{N}(\mathcal{R}), \ \forall \mathcal{R} \in \mathcal{S}^{P^*}.$$

We define $b \in \mathbb{R}_+$ such that $\| f_{\hat{\jmath}_z} \circ f_{\hat{\jmath}_{z-1}} \cdots \circ f_{\hat{\jmath}_1} (\alpha_{P^*} I) \| < b$. Consider the compact set $K_a = \{X \in \mathbb{S}_+^M | \|X\| \leq b\}$, and also define the closed set $F_b = \{X \in \mathbb{S}_+^M | \|X\| \geq b\}$. From Lemma 3.3.8, define the set $\mathcal{U}(F_b)$ as

$$\mathcal{U}(F_b) = \left\{ \mathcal{R} \in \mathcal{S}^{P^*} | \mathcal{N}(\mathcal{R}) \in F_b \right\}. \tag{3.96}$$

Then, we have the following inclusion:

$$\mathcal{U}(F_b) \subset \left\{ \mathcal{R} \in \mathcal{S}^{P^*} | \pi(\mathcal{R}) \geq z \right\}. \tag{3.97}$$

Hence, $l(F_b) = \inf_{\mathcal{R} \in \mathcal{U}(F_b)} \pi(\mathcal{R}) \geq z$. Since F_b is closed, by Lemma 3.3.8, there exists $r_{F_b} \in \mathbb{T}_+$ such that

$$\pi(\mathcal{R}^{r_{F_b}}) \geq z, \ \forall \mathcal{R} \in \mathcal{U}(F_b). \tag{3.98}$$

To estimate the probability $\mu^{\overline{\gamma}}(K_a^C)$, we follow the method in Lemma 3.3.9. First, we have the following by weak convergence:

$$\mu^{\overline{\gamma}}(K_a^C) \leq \liminf_{r \to \infty} \mathbb{P}\big(P_n(r) \in K_a^C\big) \leq \liminf_{r \to \infty} \mathbb{P}(P_n(r) \in F_b).$$

For $r \in \mathbb{T}_+$, denote the set $\mathcal{J}_r^{P^*} = \mathcal{S}_r^{P^*} \cap \mathcal{U}(F_b)$. For $r \geq r_{F_b}$, similar to (3.92), we have

$$\begin{aligned}
\mathbb{P}\left(P_n(r) \in F_b\right) &= \sum_{\mathcal{R} \in \mathcal{J}_r^{P^*}} \prod_{k=1}^{r} q_{n_k}(\jmath_k) \\
&\leq (2^N - 1)^{\underline{l}(F_b)} \begin{pmatrix} r_{F_b} \\ \underline{l}(F_b) \end{pmatrix} \\
&\quad \max_{\substack{(n_1, \cdots, n_{r_{F_b}}) \in \mathcal{P}_{r_{F_b}} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_{F_b}}) = \underline{l}(F_b)}} \prod_{k=1}^{r_{F_b}} q_{n_k}(\jmath_k | \jmath_k \neq 2^N - 1) \\
&\leq (2^N - 1)^{\underline{l}(F_b)} \begin{pmatrix} r_{F_b} \\ \underline{l}(F_b) \end{pmatrix} \\
&\quad \max_{\substack{(n_1, \cdots, n_{r_z}) \in \mathcal{P}_{r_z} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_z}) = z}} \prod_{k=1}^{r_z} q_{n_k}(\jmath_k | \jmath_k \neq 2^N - 1).
\end{aligned}$$

Arguments similar to those in Lemma 3.3.9 lead to

$$\begin{aligned}
\mu^{\overline{\gamma}}(K_a^C) &\leq (2^N - 1)^{\underline{l}(F_b)} \begin{pmatrix} r_{F_b} \\ \underline{l}(F_b) \end{pmatrix} \\
&\quad \max_{\substack{(n_1, \cdots, n_{r_z}) \in \mathcal{P}_{r_z} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_z}) = z}} \prod_{k=1}^{r_z} q_{n_k}(\jmath_k | \jmath_k \neq 2^N - 1),
\end{aligned}$$

from which we obtain,

$$\begin{aligned}
&\limsup_{\overline{\gamma} \to \infty} \frac{\ln \mu^{\overline{\gamma}}(K_a^C)}{\overline{\gamma}} \leq \\
&- \min_{\substack{(n_1, \cdots, n_{r_z}) \in \mathcal{P}_{r_z} \\ \mathcal{R} \,:\, \pi(\mathcal{R}^{r_z}) = z}} \sum_{i=1}^{r_z} \mathbb{I}_{\jmath_i \neq 2^N - 1} \overline{q}_{n_i}(\jmath_i) \leq -W(z),
\end{aligned}$$

where $W(z)$ is defined as

$$W(z) = \min_{\substack{(n_1,\cdots,n_{\mathrm{len}(\mathcal{R})}) \in \mathcal{P}_{\mathrm{len}(\mathcal{R})} \\ \mathcal{R}\,:\,\pi(\mathcal{R})=z}} \sum_{i=1}^{\mathrm{len}(\mathcal{R})} \mathbb{I}_{\jmath_i \neq 2^N-1} \overline{q}_{n_i}(\jmath_i). \tag{3.99}$$

Obviously, $W(z) \geq W(a)$ follows from $z \geq \lfloor a \rfloor$. Then we have $\limsup_{\overline{\gamma}\to\infty} \frac{\ln \mu^{\overline{\gamma}}(K_a^C)}{\overline{\gamma}} \leq -W(a)$.

3.5.5 Proof of Theorem 3.3.3

We are now ready to complete the proof of Theorem 3.3.3.

Proof 9 Lemma 3.3.6 and Lemma 3.3.11 have established that the family of $\{\mu^{\overline{\gamma}}\}$ satisfies the LD lower and upper bounds at scale $\overline{\gamma}$ with rate functions $\underline{I}_L$ and $\overline{I}_L$, respectively, as $\overline{\gamma} \to \infty$. To complete the proof of Theorem 3.3.3 it suffices to show that $\overline{I}_L(\cdot) = \overline{I}(\cdot)$ and $\underline{I}_L(\cdot) = \underline{I}(\cdot)$, i.e., $\overline{I}(\cdot)$ and $\underline{I}(\cdot)$ are lower semicontinuous. We first prove $\overline{I}_L(\cdot) = \overline{I}(\cdot)$, and it takes the same method to prove $\underline{I}_L(\cdot) = \underline{I}(\cdot)$.

If $\overline{I}_L(X) = \infty$, from Proposition 3.3.5 (iii), clearly $\overline{I}_L(X) = \overline{I}(X),\ \forall X \in \mathbb{S}_+^M$. Then, we consider the case $\overline{I}_L(X) < \infty$. From the definition

$$\overline{I}_L(X) = \lim_{\varepsilon\to\infty} \inf_{Y\in B_\varepsilon(X)} \overline{I}(Y), \tag{3.100}$$

we know the discrete quantity $\inf_{Y\in B_\varepsilon(X)} \overline{I}(Y)$ is non-decreasing w.r.t. ε; then there exists $\varepsilon_0 > 0$ such that

$$\overline{I}_L(X) = \inf_{Y\in B_\varepsilon(X)} \overline{I}(Y),\ \forall \varepsilon \leq \varepsilon_0. \tag{3.101}$$

The infimum above is achieved for every $\varepsilon_0 > 0$, and we conclude that there exists a

sequence $\{X_n\}_{n\in\mathbb{N}}$ such that

$$X_n \in \overline{B_{\varepsilon_0}}(X),\ \lim_{n\to\infty} X_n = X,\ \overline{I}(X_n) = \overline{I}_L(X). \tag{3.102}$$

Recall the set of strings

$$\mathcal{U}(\overline{B_{\varepsilon_0}}(X)) = \{\mathcal{R} \in \mathcal{S}^{P^*} | \mathcal{N}(\mathcal{R}) \in \overline{B_{\varepsilon_0}}(X)\}. \tag{3.103}$$

Then we have

$$l'(\overline{B_{\varepsilon_0}}(X)) = \inf_{Y\in\overline{B_{\varepsilon_0}}(X)} \overline{I}(Y) = \overline{I}_L(X). \tag{3.104}$$

Since $\overline{B_{\varepsilon_0}}(X)$ is closed, by Lemma 3.3.8, there exists $r_0 \in \mathbb{T}_+$ such that for $\mathcal{R} \in \mathcal{U}(\overline{B_{\varepsilon_0}}(X))$ with $\mathrm{len}(\mathcal{R}) \geq r_0$,

$$\overline{w}(\mathcal{R}^{r_0}) \geq l'(\overline{B_{\varepsilon_0}}(X)) = \overline{I}_L(X). \tag{3.105}$$

By the existence of $\{X_n\}$, there exists a sequence $\{\mathcal{R}_n\}$ of strings in $\mathcal{U}(\overline{B_{\varepsilon_0}}(X))$ such that

$$\mathcal{N}(\mathcal{R}_n) = X_n,\ \overline{w}(\mathcal{R}_n) = \overline{I}_L(X). \tag{3.106}$$

Without loss of generality, we assume that $\mathrm{len}(\mathcal{R}_n) = r_0$ for all n. Indeed, if $\mathrm{len}(\mathcal{R}_n) < r_0$, we can modify $\mathcal{R}_n$ by appending the requisite number of f_{2^N-1} at the right end, which still satisfies (3.106). On the other hand, if $\mathrm{len}(\mathcal{R}_n) > r_0$, we note that $\mathcal{R}_n$ must be of the form

$$\mathcal{R}_n = \left(f_{\jmath_1}, \cdots, f_{\jmath_{r_0}}, f_{2^N-1}^{\mathrm{len}(\mathcal{R}_n)-r_0}, P^*\right) \tag{3.107}$$

where the truncated string (defined in Definition 3.3.7) $\mathcal{R}_n^{r_0}$ satisfies $\mathcal{N}(\mathcal{R}_n^{r_0}) = X_n$

and $\overline{w}(\mathcal{R}_n^{r_0}) = \overline{I}_L(X)$. Hence, if $\text{len}(\mathcal{R}_n) > r_0$, we may consider the truncated string $\mathcal{R}_n^{r_0}$ instead, which also satisfies (3.106). We thus assume that the sequence $\{\mathcal{R}_n\}$ with the properties in (3.106) further satisfies $\text{len}(\mathcal{R}_n) = r_0$ for all n.

The number of distinct strings in the sequence $\{\mathcal{R}_n\}$ is at most $(2^N-1)^{r_0}$; in fact, it should be less than $(2^N-1)^{r_0}$ due to the constraint $\overline{w}(\mathcal{R}_n) = \overline{I}_L(X)$. Hence, at least one pattern is repeated infinitely often in the sequence $\{\mathcal{R}_n\}$, i.e., there exists a string $\mathcal{R}^*$ such that we have $\text{len}(\mathcal{R}^*) = r_0,\ \overline{w}(\mathcal{R}^*) = \overline{I}_L(X)$, and a subsequence $\{\mathcal{R}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mathcal{R}_n\}$ with $\mathcal{R}_{n_k} = \mathcal{R}^*$.

The corresponding subsequence $\{X_{n_k}\}$ of numerical values then satisfies

$$X_{n_k} = \mathcal{N}(\mathcal{R}_{n_k}) = \mathcal{N}(\mathcal{R}^*),\ \forall k \in \mathbb{N}, \tag{3.108}$$

and hence we have

$$X = \lim_{k\to\infty} X_{n_k} = \mathcal{N}(\mathcal{R}^*). \tag{3.109}$$

Therefore, we have the string $\mathcal{R}^* \in \mathcal{S}^{P^*}(X)$ and

$$\overline{I}(X) = \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \overline{w}(\mathcal{R}) \le \overline{w}(\mathcal{R}^*) = \overline{I}_L(X). \tag{3.110}$$

With the fact that $\overline{I}(X) \ge \overline{I}_L(X)$, we have the final conclusion:

$$\overline{I}(X) = \overline{I}_L(X). \tag{3.111}$$

4. QUANTIZATION BASED ALGORITHM (QGIKF) WITH INFINITE QUANTIZATION ALPHABET*

In this chapter, we propose and study a quantized GIKF (QGIKF) algorithm with infinite quantization alphabet in a dynamic scalar large scale system. Due to the limited sources, such as the bandwidth and power, the exact analog signal is prevented to transmit in the wireless sensor network. It is required to first quantize the data before transmitting it. In previous chapters, we have assumed the state of a sensor can be ideally transmitted to its neighbors. In this chapter, we extend the previous study to investigate the performance of the GIKF algorithm with quantization to meet the constrains of limited sources. Since quantization could cause the information loss on the transmitted data, an natural question is raised that whether or not the QGIKF can still achieve weak consensus with this type of information loss. In the sequel, we will find the answer of this question. First, we describe the quantization scheme used in this chapter, named dithered quantization. Then, we propose the QGIKF algorithm and derive the iteration formulation of estimation error variance with QGIKF. Finally, the analysis over weak consensus is presented.

4.1 System Setup

4.1.1 System Model

In this chapter and Chapter 5, we consider a discrete-time linear Gaussian dynamic scalar system observed by a network of N sensors. The system model is given

*Part of this chapter is reprinted, with permission, from [Di Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui, "Distributed Kalman filtering with quantized sensing state." *IEEE Transactions on Signal Processing*, 63(19):5180–5193, Oct. 2015.]

as

$$x_{k+1} = Fx_k + w_{k+1}, \tag{4.1}$$

where F is the system parameter, x_k is the system state with initial state x_0 distributed as a Gaussian value with $x_0 \sim N(\widehat{x}_{0|-1}, \widehat{P}_{0|-1})$ ($\widehat{x}_{0|-1}$ denotes the prior estimate of x_0 (with no observation information), and $\widehat{P}_{0|-1}$ denotes the corresponding error variance), and the system noise $\{w_k\}$ is an uncorrelated zero mean Gaussian sequence with variance Q, independent of x_0.

The observation signal at sensor n is

$$y_k^n = C_n x_k + v_k^n, \tag{4.2}$$

where C_n is the observation parameter and the observation noise $\{v_k^n\}$ is another uncorrelated zero mean Gaussian sequence with variance R_n. These noise sequences at different sensors are independent of each other, and independent of the system noise sequence $\{w_k\}$ and the initial system state x_0.

The inter-sensor communication model is same as the communication model described in Section 2.1.2, which is omitted here.

4.1.2 Quantization Scheme

The quantization scheme adopted in this paper is the dithered quantization [22, 57, 2], where a controlled noise or dither is added to randomize the value before quantization, where a uniform quantizer is applied. The detailed quantization process is stated as follows. The dither v as a random variable is first added to the value x to be quantized. We then adopt the uniform quantizer with a quantization step Δ

and the countably infinite quantization alphabet [29] given by

$$\mathcal{Q} = \{k\Delta | k \in \mathbb{Z}\}. \tag{4.3}$$

With $\mathcal{Q}$ in (4.3), the quantizing function $q(\cdot) : \mathbb{R} \to \mathcal{Q}$ is given as

$$q(x) = \arg\min_{k\Delta} |k\Delta - (x + v)|.$$

In other words,

$$q(x) = k\Delta, \ \forall \ k\Delta - \frac{\Delta}{2} \leq x + v < k\Delta + \frac{\Delta}{2}.$$

Then the quantization noise ε is defined as

$$\varepsilon = q(x) - x, \tag{4.4}$$

while the quantization error e is

$$e = q(x) - (x + v). \tag{4.5}$$

We see that $\varepsilon = e + v$. Here we adopt the non-subtractive dithered quantization [22], which is more practical compared to the subtractive dithered quantization where the receiver is assumed to know the dither signal and subtract it from the reconstructed quantizer value such that $\varepsilon = e$.

If the dither v satisfies the Schuchman conditions [57], the quantization error e is i.i.d. uniformly distributed on $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ and independent of the input value x. A sufficient condition for v to satisfy the Schuchman conditions is that v is i.i.d. uniformly distributed on $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ and independent of the input value x. In the sequel, we assume that this sufficient conditions holds.

If the above sufficient condition holds, the dithered quantization scheme is equivalent to the probabilistic quantization [73] [2], where x is quantized in a probabilistic fashion as

$$P\{q(x) = (k+1)\Delta\} = (x - k\Delta)/\Delta,$$
$$P\{q(x) = k\Delta\} = 1 - (x - k\Delta)/\Delta,$$

where $q(x)$ is an unbiased estimator as $E[q(x)] = x$.

4.2 QGIKF Scheme

In this section, the procedure of QGIKF algorithm is first described in detail, then the iteration of the corresponding estimation error variance is derived.

4.2.1 QGIKF Algorithm

With the quantization scheme described in Section 4.1.2, we now introduce the quantized gossip-based interacting Kalman filtering (QGIKF) scheme for distributed estimation of the state process x_k over time. Let the filter at sensor n be initialized by the pair $\left(\widehat{x}_{0|-1}, \widehat{P}_{0|-1}\right)$, where $\widehat{x}_{0|-1}$ denotes the prior estimate of x_0 (with no observation information) and $\widehat{P}_{0|-1}$ is the corresponding error variance. Also, $\left(\widehat{x}^n_{k|k-1}, \widehat{P}^n_{k|k-1}\right)$ denote the prediction of x_k at sensor n based on information till time $k-1$ and the corresponding conditional error variance, respectively. The pair $\left(\widehat{x}^n_{k|k-1}, \widehat{P}^n_{k|k-1}\right)$ is also referred as the state of sensor n at time $k-1$. To define the estimate update rule for the QGIKF, let $n_k^{\rightarrow}$ be the communication neighbor of sensor n at time k w.r.t. the adjacency matrix[1] $A(k)$. We assume that all inter-sensor communications of time k occur at the beginning of the slot, after the state $\left(\widehat{x}^n_{k|k-1}, \widehat{P}^n_{k|k-1}\right)$ is quantized according to the dithered quantization scheme with output $q\left(\widehat{x}^n_{k|k-1}, \widehat{P}^n_{k|k-1}\right)$.

[1]Note that by symmetry we have $(n_k^{\rightarrow})_k^{\rightarrow} = n$. It is possible that $n_k^{\rightarrow} = n$, where the graph corresponding to $A(k)$ has a self-loop at node n.

The paired communicating sensors receive the quantized states from each other and swap out their previous states, i.e., if at time k, $n_k^{\rightarrow} = l,\ l \neq n$, sensor n replaces its previous state $\left(\widehat{x}_{k|k-1}^n, \widehat{P}_{k|k-1}^n\right)$ by $q\left(\widehat{x}_{k|k-1}^l, \widehat{P}_{k|k-1}^l\right)$ and sensor l replaces its previous state $\left(\widehat{x}_{k|k-1}^l, \widehat{P}_{k|k-1}^l\right)$ by $q\left(\widehat{x}_{k|k-1}^n, \widehat{P}_{k|k-1}^n\right)$. If at time k, $n_k^{\rightarrow} = n$, i.e., sensor n communicates to itself (no inter-sensor communication occurs), sensor n keeps its previous state $\left(\widehat{x}_{k|k-1}^n, \widehat{P}_{k|k-1}^n\right)$.

4.2.2 Estimation Error Variance Iteration

After the above communication procedure is over and a new observation is made, by the recursion algorithm of Kalman filtering, when the inter-sensor communication occurs as $n_k^{\rightarrow} \neq n$, the estimate update at sensor n at the end of the slot k executes as

$$\widehat{x}_{k+1|k}^n = F\widehat{x}_{k|k}^n, \tag{4.6}$$

where

$$\widehat{x}_{k|k}^n = q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right) + K_k^n\left[y_k^n - C_n q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right], \tag{4.7}$$

and K_k^n is the Kalman filtering gain. Then for the estimation error variance, we have

$$\begin{aligned}\widehat{P}_{k+1|k}^n =& \mathbb{E}\left[\left(x_{k+1} - \widehat{x}_{k+1|k}^n\right)^2 \middle| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right), q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right), n_k^{\rightarrow}, y_k^n\right] \\ =& \mathbb{E}'\Big\{\mathbb{E}\left[\left(x_{k+1} - \widehat{x}_{k+1|k}^n\right)^2 \middle| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right), q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right), \right. \\ & \left. n_k^{\rightarrow}, y_k^n, \varepsilon_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}, \varepsilon_{\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}}\right]\Big\},\end{aligned} \tag{4.8}$$

where $\varepsilon_{\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}}$ and $\varepsilon_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}$ are the quantization noises, and the expectation $\mathbb{E}'$ is over $\varepsilon_{\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}}$ and $\varepsilon_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}$. For concision, in the following we denote the combined operations of all expectations as $\mathbb{E}' \circ \mathbb{E}$. To calculate (4.8), first we have the following correction

step to derive $\widehat{P}^n_{k|k}$:

$$\begin{aligned}
\widehat{P}^n_{k|k} &= \mathbb{E}' \circ \mathbb{E}\left[\left(x_k - \widehat{x}^n_{k|k}\right)^2 \middle| \ \cdot \ \right] \\
&= \mathbb{E}' \circ \mathbb{E}\left[\left(x_k - q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right)^2 \middle| \ \cdot \ \right] \\
&+ (K^n_k)^2 \mathbb{E}' \circ \mathbb{E}\left[\left(y^n_k - C_n q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right)^2 \middle| \ \cdot \ \right] \\
&- 2K^n_k \mathbb{E}' \circ \mathbb{E}\left[\left(x_k - q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right)\left(y^n_k - C_n q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right) \middle| \cdot \right], \qquad (4.9)
\end{aligned}$$

where the condition '$\cdot$' denotes the condition in (4.8). The first term in (4.9) is calculated as

$$\begin{aligned}
&\mathbb{E}' \circ \mathbb{E}\left[\left(x_k - q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right)^2 \middle| \ \cdot \ \right] \\
&= \mathbb{E}' \circ \mathbb{E}\left[\left(x_k - \widehat{x}^{n\rightarrow_k}_{k|k-1}\right)^2 \middle| \ \cdot \ \right] + \mathbb{E}'\left[\varepsilon^2_{\widehat{x}^{n\rightarrow_k}_{k|k-1}} \middle| q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right] \\
&= q\left(\widehat{P}^{n\rightarrow_{k|k-1}}_k\right) - \mathbb{E}'\left[\varepsilon_{\widehat{P}^{n\rightarrow_k}_{k|k-1}} \middle| q\left(\widehat{P}^{n\rightarrow_k}_{k|k-1}\right)\right] \\
&\quad + \mathbb{E}'\left[\varepsilon^2_{\widehat{x}^{n\rightarrow_k}_{k|k-1}} \middle| q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right] \\
&= q\left(\widehat{P}^{n\rightarrow_k}_{k|k-1}\right) + \frac{\Delta^2}{6}, \qquad (4.10)
\end{aligned}$$

where the first equation follows from $\varepsilon_{\widehat{x}^{n\rightarrow_k}_{k|k-1}} = e_{\widehat{x},k} + v$, in which $e_{\widehat{x},k}$ and v are independent of x_k and $\widehat{x}^{n\rightarrow_k}_{k|k-1}$, and the last equation is derived in Appendix 4.4.

The third term of (4.9) equals to

$$\begin{aligned}
&2K^n_k \mathbb{E}' \circ \mathbb{E}\left[\left(x_k - \widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\left(y^n_k - C_n\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right) \middle| \ \cdot \ \right] \\
&+ 2\mathbb{E}'\left[\varepsilon^2_{\widehat{x}^{n\rightarrow_k}_{k|k-1}} \middle| q\left(\widehat{x}^{n\rightarrow_k}_{k|k-1}\right)\right] C_n K^n_k \\
&= 2q\left(\widehat{P}^{n\rightarrow_k}_{k|k-1}\right) C_n K^n_k + \frac{\Delta^2}{3} C_n K^n_k.
\end{aligned}$$

By (4.2), the second term of (4.9) is derived as

$$\begin{aligned}
&(K_k^n)^2\mathbb{E}'\circ\mathbb{E}\left[\left(y_k^n - C_n q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right)^2\middle|\ \cdot\ \right] \\
&= (K_k^n)^2\mathbb{E}'\circ\mathbb{E}\left[\left(y_k^n - C_n \widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)^2\middle|\ \cdot\ \right] \\
&\quad + (K_k^n C_n)^2\mathbb{E}'\left[\varepsilon^2_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}\middle| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right] \\
&= (K_k^n)^2\left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right] + \frac{\Delta^2}{6}(K_k^n C_n)^2.
\end{aligned}$$

Then, we can rewrite (4.9) as

$$\begin{aligned}
\widehat{P}_{k|k}^n = {} & q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) - 2q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) C_n K_k^n \\
& + (K_k^n)^2\left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right] + Z_k^n,
\end{aligned} \tag{4.11}$$

where $Z_k^n = \frac{\Delta^2}{6}\left(1 - K_k^n C_n\right)^2$.

Here we derive the optimal K_k^n by minimizing $\widehat{P}_{k|k}^n$. Let

$$\frac{\partial \widehat{P}_{k|k}^n}{\partial K_k^n} = \left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + C_n^2\frac{\Delta^2}{6} + R_n\right] K_k^n - q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) C_n - \frac{\Delta^2}{6} C_n = 0;$$

the solution leads to the optimal K_k^n as

$$K_k^{n*} = \frac{\left[q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + \frac{\Delta^2}{6}\right] C_n}{C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + C_n^2\frac{\Delta^2}{6} + R_n}. \tag{4.12}$$

Then at the update step, we have

$$\widehat{P}_{k+1|k}^n = F^2 \widehat{P}_{k|k}^n + Q. \tag{4.13}$$

Equs (4.6), (4.7), (4.11), (4.12) and (4.13) construct an optimal distributed and

quantized Kalman filtering algorithm. To establish the property that the estimation error variance at a randomly selected sensor converges in distribution to a unique invariant distribution, we first study the following suboptimal algorithm; later we will show that the convergence property is automatically verified in the optimal case based on its proof for the suboptimal algorithm.

For the suboptimal algorithm, instead of adopting the optimal K_k^{n*} in (4.12), we choose the gain K_k^n as

$$K_k^n = q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) C_n \left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right]^{-1}. \tag{4.14}$$

By plugging (4.14) into (4.11), we have

$$\widehat{P}_{k|k}^n = (1 - K_k^n C_n) q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + Z_k^n. \tag{4.15}$$

And then according to (4.13), we deduce

$$\begin{aligned}\widehat{P}_{k+1|k}^n &= F^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + Q + F^2 Z_k^n \\ &- F^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right)^2 C_n^2 \left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right]^{-1}.\end{aligned} \tag{4.16}$$

When sensor n communicates to itself (no inter-sensor communication occurs) as $n_k^{\rightarrow} = n$, the problem degenerates to the classic Kalman filtering problem [27] with $\widehat{x}_{k+1|k}^n$ and $\widehat{P}_{k+1|k}^n$ respectively iterated as

$$\widehat{x}_{k+1|k}^n = F\widehat{x}_{k|k-1}^n + FK_k^n\left(y_k^n - C_n\widehat{x}_{k|k-1}^n\right), \tag{4.17}$$

$$\widehat{P}_{k+1|k}^n = F^2\widehat{P}_{k|k-1}^n + Q - \left(F\widehat{P}_{k|k-1}^n C_n\right)^2 \left[C_n^2\widehat{P}_{k|k-1}^n + R_n\right]^{-1}, \tag{4.18}$$

where

$$K_k^n = \widehat{P}_{k|k-1}^n C_n \left(C_n^2 \widehat{P}_{k|k-1}^n + R_n \right)^{-1}. \tag{4.19}$$

In the sequel, we will study the asymptotic property of the error variance sequence $\left\{ \widehat{P}_{k+1|k}^n \right\}$ iterated as (4.16) or (4.18) to show that the network achieves weak consensus.

4.3 Weak Consensus Analysis

In this section, we first present the interacting particle representation for the estimation error variance sequence. Then an RDS formulation is setup by considering an auxiliary sequence. Finally, the weak consensus result is proved.

4.3.1 Interacting Particle Representation

First, to simplify the notation in (4.16), we define the function $f_{1,n}$ as

$$\begin{aligned} f_{1,n}(X) = F^2 q(X) + \mathcal{Q} - F^2 q(X)^2 \mathcal{C}_n^2 \left[\mathcal{C}_n^2 q(X) + R_n \right]^{-1} \\ + \frac{\Delta^2}{6} F^2 \left(1 - \mathcal{C}_n^2 q(X) \left[\mathcal{C}_n^2 q(X) + R_n \right]^{-1} \right)^2, \end{aligned} \tag{4.20}$$

and to simplify the notation in (4.18), we define the function $f_{2,n}$ as

$$f_{2,n}(X) = F^2 X + Q - (FXC_n)^2 \left[C_n^2 X + R_n \right]^{-1}. \tag{4.21}$$

Then the sequence of error variance $\widehat{P}_{k+1|k}^n$ at sensor n iterates according to $\widehat{P}_{k+1|k}^n = f_n \left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}} \right)$, where $f_n = f_{1,n}$, if $n_k^{\rightarrow} \neq n$; and $f_n = f_{2,n}$, if $n_k^{\rightarrow} = n$.

To track the sequence $\left\{ \widehat{P}_{k|k-1}^n \right\}$, we adopt the following interacting particle process to represent it. We will show that by the interacting particle representation, we can completely characterize and track the evolution of the sequence $\left\{ \widehat{P}_{k|k-1}^n \right\}$ for $n = 1, ..., N$.

Note that the inter-sensor communication link formation process controlled by the sequence $\{A(k)\}$ can be represented by N particles moving on the graph as a Markov chain. The state of the n-th particle at time k is denoted by $p_n(k)$, where $p_n(k)$ takes values from the state space $[1, ..., N]$, and the transition of the n-th particle is given by

$$p_n(k) = (p_n(k-1))_k^{\rightarrow},\ p_n(0) = n.$$

Therefore, the n-th particle can be considered as originating from node n and then travelling on the graph according to the link formation process $\{A(k)\}$.

For each n, the process $\{p_n(k)\}$ is a Markov chain on $V = [1, \cdots, N]$ with the transition probability matrix $\bar{A}$. For each of the Markov chains $\{p_n(k)\}$, we define a sequence of iteration $P_n(k)$ as

$$P_n(k+1) = f_{p_n(k)}\left(P_n(k)\right), \tag{4.22}$$

where $f_{p_n(k)} = f_{1,p_n(k)}$ if $p_n(k) \neq p_n(k+1)$; and $f_{p_n(k)} = f_{2,p_n(k)}$ if $p_n(k) = p_n(k+1)$.

Note that the sequence $\{P_n(k)\}$ is governed by the Markov chain $\{p_n(k)\}$, and from the perspective of the particle, $\{P_n(k)\}$ can be considered as a particle originating at sensor n and hopping around the network as a Markov chain with transition probability matrix $\bar{A}$, whose state $P_n(k)$ evolves according to function (4.22). In contrast to the sequence $\left\{\widehat{P}^n_{k|k-1}\right\}$ of the conditional error variance at a particular sensor n, the sequence $\{P_n(k)\}$ does not correspond to the error variance evolution at a particular sensor. With the Markov chain $\{p_n(k)\}$, the relation between $\left\{\widehat{P}^n_{k|k-1}\right\}$ and $\{P_n(k)\}$ could be shown as

$$(P_1(k), \cdots, P_N(k)) = \left(\widehat{P}^{p_1(k)}_{k|k-1}, \cdots, \widehat{P}^{p_N(k)}_{k|k-1}\right), \tag{4.23}$$

from which we see that the properties of the sequence of interest $\left\{\widehat{P}^n_{k|k-1}\right\}$ could be obtained by studying the corresponding sequence $\{P_n(k)\}$. Hence, in the sequel, we will first study the sequence $\{P_n(k)\}$ to show its weak convergence.

4.3.2 An Auxiliary Sequence with RDS Formulation

Since the Markov chain $\{p_n(k)\}$ is not with a stationary distribution, in order to perform the analysis based on RDS [1] [26] to analyze $\{P_n(k)\}$, we need an auxiliary sequence $\left\{\widetilde{P}(k)\right\}$ based on a Markov chain $\{\widetilde{p}(k)\}$ with a stationary distribution, which is a Markov chain with the transition matrix $\bar{A}$ and an uniform initial distribution $\mathbb{P}[\widetilde{p}(0) = n] = 1/N,\ n = 1, \cdots, N$. For the corresponding $\left\{\widetilde{P}(k)\right\}$, with random initial condition $\widetilde{P}(0)$, it is defined as

$$\widetilde{P}(k+1) = f_{\widetilde{p}(k)}\left(\widetilde{P}(k)\right), \tag{4.24}$$

where $f_{\widetilde{p}(k)} = f_{1,\widetilde{p}(k)}$ if $\widetilde{p}(k) \neq \widetilde{p}(k+1)$; and $f_{\widetilde{p}(k)} = f_{2,\widetilde{p}(k)}$ if $\widetilde{p}(k) = \widetilde{p}(k+1)$.

Now in order to proceed the asymptotic analysis of the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$, we can construct an RDS (θ^R, φ^R) equivalent to the auxiliary sequence $\left\{\widetilde{P}(k)\right\}$ in the sense of distribution. The construction process is similar to that in our previous paper [30], for which the details are skipped here. Briefly, denote $\theta^R = (\Omega^R, \mathcal{F}^R, \mathbb{P}^R, \{\theta^R_k, k \in \mathbb{T}\})$ as a metric dynamic system, where $(\Omega^R, \mathcal{F}^R, \mathbb{P}^R)$ is a probability space and the family of transformations $\{\theta^R_k\}$ on Ω^R is the family of left-shifts, i.e., $\theta^R_k w = w(k+\cdot),\ \forall k \in \mathbb{T},\ w \in \Omega^R$; The cocycle $\varphi^R :\ \mathbb{T}_+ \times \Omega^R \times \mathbb{R}_+ \mapsto \mathbb{R}_+$

is defined as: $\forall k > 1, w, X$,

$$\begin{aligned}
\varphi^R(0, w, X) &= X, \\
\varphi^R(1, w, X) &= f_{w_0}(X), \\
\varphi^R(k, w, X) &= f_{\theta^R_{k-1} w(0)} \left(\varphi^R(k-1, w, X) \right) \\
&= f_{w_{k-1}} \left(\varphi^R(k-1, w, X) \right).
\end{aligned}$$

Based on the construction of (θ^R, φ^R), the sequence $\left\{ \varphi^R \left(k, w, P_n(0) \right) \right\}$ is distributionally equivalent to the sequence $\left\{ \widetilde{P}(k) \right\}$. At this stage, we can analyze the asymptotic distributional properties of the sequence $\left\{ \widetilde{P}(k) \right\}$ by utilizing the properties of an RDS, which is presented in the next subsection.

4.3.3 Some Intermediate Results

The following proposition states two boundedness properties to be used later for proving the lemmas.

Proposition 4.3.1

(i) Denote $w = (n_1, \cdots, n_l)$ as a walk on the graph $(V, \mathcal{E})$. Define an auxiliary sequence $\{P'_w(k)\}_{1 \le k \le l}$ with initial condition $P'_w(1) = X$ and the iteration as

$$\begin{aligned}
P'_w(k+1) &= F^2 P'_w(k) + Q \\
&\quad - F^2 C^2_{n_k} P'_w(k)^2 (C^2_{n_k} P'_w(k) + R_{n_k})^{-1} + \frac{\Delta^2}{6} F^2.
\end{aligned} \tag{4.25}$$

Define another auxiliary sequence $\{P''_w(k)\}_{1 \le k \le l}$ with the same initial condition $P''_w(1) =$

X and the iteration as

$$P''_w(k+1) = F^2\left(P''_w(k)+\Delta\right) + Q - F^2 C_{n_k}^2 \left(P''_w(k)+\Delta\right)^2 \times \left[C_{n_k}^2\left(P''_w(k)+\Delta\right) + R_{n_k}\right]^{-1} + \frac{\Delta^2}{6}F^2. \tag{4.26}$$

Then, we have recursively

$$P''_w(l+1) < P'_w(l+1) + \frac{F^2(F^{2l}-1)\Delta}{(F^2-1)}. \tag{4.27}$$

(ii) Recall $f_n(X)$ defined at the beginning of Section 4.3.1, we have that $f_n(X)$ is upper-bounded as

$$f_n(X) < F^2[X + Y(\Delta)] + Q, \tag{4.28}$$

where $Y(\Delta) = \frac{\Delta^2}{6} + \Delta$.

Proof 10 For part (i), we define a function $h(X)$: $h(X) = F^2X + Q - F^2C_{n_k}^2X^2(C_{n_k}^2X + R_{n_k})^{-1} + \frac{\Delta^2}{6}F^2$, and it is easy to verify that $h(X)$ is monotonically non-decreasing. Then compare the structures of $P'_w(l+1)$ and $P''_w(l+1)$, we can deduce that

$$P''_w(l+1) < P'_w(l+1) + \frac{F^2(F^{2l}-1)\Delta}{(F^2-1)}.$$

For part (ii), $f_n(X)$ is either $f_{1,n}(X)$ or $f_{2,n}(X)$. We have

$$f_{1,n}(X) \le F^2 q(X) + Q + \frac{\Delta^2}{6}F^2.$$

Additionally, since $X - \Delta \leq q(X) < X + \Delta$, we have

$$\begin{aligned} f_{1,n}(X) &< F^2(X + \Delta) + Q + \frac{\Delta^2}{6}F^2 \\ &= F^2[X + Y(\Delta)] + Q. \end{aligned}$$

We also have

$$f_{2,n}(X) \leq F^2 X + Q < F^2[X + Y(\Delta)] + Q.$$

Thus, we conclude that

$$f_n(X) < F^2[X + Y(\Delta)] + Q.$$

Lemma 4.3.2 Let $w = (n_1, \cdots, n_l)$ as a walk on the graph $(V, \mathcal{E})$ such that, there exists at least one C_{n_i}, $i = 1, \cdots, l$, be non-zero. If we define the function g_w as

$$g_w(X) = f_{n_l} \circ f_{n_{l-1}} \circ \cdots \circ f_{n_1}(X), \tag{4.29}$$

there exists a constant $\alpha > 0$ such that the following result holds:

$$g_w(X) \leq \alpha, \ \forall X \geq 0. \tag{4.30}$$

Proof 11 With the walk $w = (n_1, \cdots, n_l)$, we construct an error variance sequence $\{P_w(k)\}_{1 \leq k \leq l}$ satisfying the iteration:

$$P_w(k+1) = f_{n_k}(P_w(k)) \tag{4.31}$$

with initial condition $P_w(1) = X$, where the function f_{n_k} is defined at the beginning

of Section 4.3.1. Then we have

$$P_w(l+1) = f_{n_l} \circ f_{n_{l-1}} \circ \cdots \circ f_{n_1}(X), \tag{4.32}$$

which implies that $g_w(X) = P_w(l+1)$.

If $n_{k+1} \neq n_k$, recall (4.20); we then have

$$\begin{aligned}
&P_w(k+1) = F^2 q[P_w(k)] + Q - F^2 C_{n_k}^2 q[P_w(k)]^2 (C_{n_k}^2 q[P_w(k)] + R_{n_k})^{-1} \\
&+ \frac{\Delta^2}{6} F^2 \left\{ q[P_w(k)] C_{n_k}^2 (C_{n_k}^2 q[P_w(k)] + R_{n_k})^{-1} - 1 \right\}^2 \\
&\overset{a}{\leq} F^2 q[P_w(k)] + Q - F^2 C_{n_k}^2 q[P_w(k)]^2 \\
&(C_{n_k}^2 q[P_w(k)] + R_{n_k})^{-1} + \frac{\Delta^2}{6} F^2 \\
&\overset{b}{\leq} F^2 (P_w(k) + \Delta) + Q - F^2 C_{n_k}^2 (P_w(k) + \Delta)^2 \\
&(C_{n_k}^2 (P_w(k) + \Delta) + R_{n_k})^{-1} + \frac{\Delta^2}{6} F^2
\end{aligned} \tag{4.33}$$

where inequality a is due to the fact that $0 \leq q[P_w(k)] C_{n_k}^2 (C_{n_k}^2 q[P_w(k)] + R_{n_k})^{-1} < 1$, and inequality b is due to the fact that $P_w(k) - \Delta \leq q[P_w(k)] < P_w(k) + \Delta$ and the monotonically non-decreasing property of the function $h(X) = F^2 X + Q - F^2 C_{n_k}^2 X^2 (C_{n_k}^2 X + R_{n_k})^{-1} + \frac{\Delta^2}{6} F^2$.

If $n_{k+1} = n_k$, recall (4.21); we then have

$$\begin{aligned}
&P_w(k+1) = \\
&F^2 P_w(k) + Q - F^2 C_{n_k}^2 P_w(k)^2 (C_{n_k}^2 P_w(k) + R_{n_k})^{-1} \\
&\leq F^2 (P_w(k) + \Delta) + Q - F^2 C_{n_k}^2 (P_w(k) + \Delta)^2 \\
&(C_{n_k}^2 (P_w(k) + \Delta) + R_{n_k})^{-1} + \frac{\Delta^2}{6} F^2
\end{aligned} \tag{4.34}$$

which is due to the fact that the function $f_{2,n}(X) = F^2X+Q-(FXC_n)^2\left[C_n^2X+R_n\right]^{-1}$ is monotonically non-decreasing.

Recall the auxiliary sequence $P_w''(k)$ defined in (4.26). Due to the facts shown in (4.33) and (4.34), $P_w(1) = P_w''(1) = X$, and the monotonically non-decreasing property of the function $h(X) = F^2X + Q - F^2C_{n_k}^2X^2(C_{n_k}^2X + R_{n_k})^{-1} + \frac{\Delta^2}{6}F^2$, we have $P_w(l+1) < P_w''(l+1)$.

For the auxiliary sequence $\{P_w'(l+1)\}$ defined in (4.25), from Lemma 15 in [30], and with the weak detectability assumption, we have $P_w'(l+1) \leq \alpha'$, where α' is a positive constant. Since $P_w'(l+1) \leq \alpha'$, according to (4.27), $P_w''(l+1)$ is also bounded. With $P_w(l+1) < P_w''(l+1)$ obtained above, we conclude that $P_w(l+1)$ could be bounded by a constant α.

Lemma 4.3.3 The sequence $\{\widetilde{P}(k)\}$ is stochastically bounded, i.e.,

$$\lim_{J\to\infty}\sup_{k\in\mathbb{T}_+}\mathbb{P}\left(\widetilde{P}(k) > J\right) = 0. \tag{4.35}$$

Proof 12 First, in the case of a stable F, i.e., $|F| \leq 1$, this claim is obvious, due to the fact that the suboptimal estimate of 0 at each sensor for all time is stochastically bounded. Next, we consider the case with $|F| > 1$.

Note that the regularity of the distribution of $\widetilde{P}(k)$ for each k implies that to prove (4.35) it suffices to show

$$\lim_{J\to\infty}\sup_{k\geq k_0}\mathbb{P}\left(\widetilde{P}(k) > J\right) = 0 \tag{4.36}$$

for some arbitrary large $k_0 \in \mathbb{T}_+$.

Recall $Y(\Delta)$ in (4.28) and α in Lemma 4.3.2, for a sufficiently large $J > 0$, we

define,

$$k(J) = \max_k \left\{ k \in \mathbb{T}_+ \,\middle|\, F^{2k}\alpha + \frac{F^{2k}-1}{F^2-1}(Q + F^2 Y(\Delta)) \le J \right\}. \tag{4.37}$$

Here we introduce some new notations. For integers $k_0, k_1 \ge l$, the statement "there exists a $(n_1, n_2, \cdots, n_l)$ walk in the time interval $[k_0, k_1]$" indicates the existence of an integer $k_0 + l - 1 \le k' \le k_1$ such that

$$\widetilde{p}(k' - l + s) = n_s, \quad 1 \le s \le l, \tag{4.38}$$

where $\{\widetilde{p}(k)\}_{k\in\mathbb{T}_+}$ is a Markov chain.

Now we make the following claim regarding the probability of interest:

$$\mathbb{P}\left(\widetilde{P}(k) > J\right) \le \mathbb{P}(\text{no } (n_1, n_2, \cdots, n_l) \text{ exists in } [k - k(J), k]). \tag{4.39}$$

In the sequel, we justify the correctness of this claim.

First, we assume the contrary that a walk $(n_1, n_2, \cdots, n_l)$ exists in the interval $[k-k(J), k]$. Then there exists $k' \in [k-k(J), k]$ such that $\widetilde{p}(k'-l+s) = n_s$, $1 \le s \le l$, which implies that $\widetilde{P}(k') = f_{n_l} \circ \cdots \circ f_{n_1}(\widetilde{P}(k' - l + 1))$. Hence, by Lemma 4.3.2, we have $\widetilde{P}(k') \le \alpha$. By (4.28), $\widetilde{P}(s) < F^2[\widetilde{P}(s-1) + Y(\Delta)] + Q$, $\forall s$; and run the recursion, we have

$$\begin{aligned}
\widetilde{P}(k) &< F^{2(k-k')}\widetilde{P}(k') + \frac{F^{2(k-k')}-1}{F^2-1}(Q + F^2 Y(\Delta)) \\
&\le F^{2(k-k')}\alpha + \frac{F^{2(k-k')}-1}{F^2-1}(Q + F^2 Y(\Delta)) \\
&\le F^{2k(J)}\alpha + \frac{F^{2(k(J))}-1}{F^2-1}(Q + F^2 Y(\Delta)) \le J,
\end{aligned}$$

where the second inequality is based on $\widetilde{P}(k') \leq \alpha$ by Lemma 4.3.2, the third inequality follows from $k - k' \leq k(J)$ by the fact that $k' \in [k - k(J), k]$, and the last inequality follows from (4.37).

Thus we can conclude that the existence of a walk $(n_1, n_2, \cdots, n_l)$ in the interval $[k–k(J), k]$ implies that $\widetilde{P}(k) \leq J$, i.e., the event $\{$Existence of a walk $(n_1, n_2, \cdots, n_l)$ in $[k - k(J), k]\} \subset \left\{\widetilde{P}(k) \leq J\right\}$, which implies the claim in (4.39). Therefore, characterizing $\mathbb{P}(\widetilde{P}(k) > J)$ on the L.H.S. of (4.39) can be reduced to studying the probability of a $(n_1, n_2, \cdots, n_l)$ walk existing in $[k - k(J), k]$. The remaining procedure is to construct another Markov chain, and relate the probability of a $(n_1, n_2, \cdots, n_l)$ walk existing in $[k - k(J), k]$ to the hitting time statistics of this Markov chain.

We construct a Markov chain $\{z(k)\}_{k \geq l}$. Recall the stochastic matrix $\bar{A}$ in Section 2.1.2. The state space $\mathcal{Z}$ is the subset of V^l given as

$$\mathcal{Z} = \{z = (i_1, \cdots, i_l) \,|\, \bar{A}_{i_j, i_{j+1}} > 0, 1 \leq j < l\}. \tag{4.40}$$

The dynamic of $\{z(k)\}_{k \geq l}$ is could be established by considering the Markov chain $\{\widetilde{p}(k)\}_{k \in \mathbb{T}_+}$ defined in Section 4.3.2 as

$$z(k) = (\widetilde{p}(k - l + 1), \widetilde{p}(k - l + 2), \cdots, \widetilde{p}(k)). \tag{4.41}$$

From the dynamic of $\{\widetilde{p}(k)\}_{k \in \mathbb{T}_+}$, it implies that $\{z(k)\}_{k \geq l}$ is a Markov chain with transition probability $\bar{A}_{nm}$ between available states $(i_1, \cdots, i_{l-1}, n)$ and $(i_2, \cdots, i_{l-1}, n, m)$. With the state space $\mathcal{Z}$, the Markov chain $\{z(k)\}_{k \geq l}$ inherits irreducibility, aperiodicity, and stationarity from that of the Markov chain $\{\widetilde{p}(k)\}_{k \in \mathbb{T}_+}$.

Denote the hitting time τ_0 of $\{z(k)\}$ to the state $(n_1, \cdots, n_l)$ as

$$\tau_0 = \min\{k > l \,|\, z(k) = (n_1, \cdots, n_l)\}.$$

For each $k \geq l$ and a sufficiently large J, we define the stopping time τ^J as

$$\tau^J = \min\{k \geq k - k(J) \,|\, z(k) = (n_1, \cdots, n_l)\}.$$

For all $z \in \mathcal{Z}$, define

$$\mathbb{P}_z(\tau_0 > h) = \mathbb{P}(\tau_0 > h \,|\, z(l) = z).$$

Then from the Markovian property, it follows that

$$\mathbb{P}(\tau^J > k \,|\, z(k - k(J) - 1) = z) = \mathbb{P}_z(\tau_0 > k(J) + 1).$$

Therefore, we have the following results:

$$\begin{aligned}
&\mathbb{P}(\text{no } (n_1, \cdots, n_l) \text{ exists in } [k - k(J), k]) = \mathbb{P}(\tau^J > k) \\
&= \sum_{z \in \mathcal{Z}} \big[\mathbb{P}(\tau^J > k \,|\, z(k - k(J) - 1) = z) \\
&\qquad \mathbb{P}(z(k - k(J) - 1) = z)] \\
&= \sum_{z \in \mathcal{Z}} [\mathbb{P}_z(\tau_0 > k(J) + 1)\mathbb{P}(z(k - k(J) - 1) = z)] \qquad (4.42)
\end{aligned}$$

Since the above result is established for all $k \geq k_0$ with some sufficient large k_0, we

have from (4.39) that

$$\begin{aligned}&\sup_{k\geq k_0}\mathbb{P}\left(\widetilde{P}(k)>J\right)\\&\leq\sum_{z\in\mathcal{Z}}\left[\mathbb{P}_z(\tau_0>k(J)+1)\mathbb{P}(z(k-k(J)-1)=z)\right]\end{aligned}\tag{4.43}$$

Due to the positive recurrence of the finite state Markov chain $\{z(k)\}$ and the fact that $k(J)\to\infty$ as $J\to\infty$, for all $z\in\mathcal{Z}$, we have

$$\lim_{J\to\infty}\mathbb{P}_z(\tau_0>k(J)+1)=0.\tag{4.44}$$

Since $\mathcal{Z}$ is finite, by combining (4.43) and (4.44) and using the dominated convergence theorem, we have

$$\lim_{J\to\infty}\sup_{k\geq k_0}\mathbb{P}\left(\widetilde{P}(k)>J\right)=0.\tag{4.45}$$

By the regularity of the distribution of $\widetilde{P}(k)$ for each k, (4.45) implies that

$$\lim_{J\to\infty}\sup_{k\in\mathbb{T}_+}\mathbb{P}\left(\widetilde{P}(k)>J\right)=0.$$

4.3.4 Main Results on Weak Consensus

With the property of stochastic boundedness for the sequence $\left\{\widetilde{P}(k)\right\}$, Lemma 6.1 in [32], and Theorem 27 in [30], we can conclude that only claim b) in Theorem 27 of [30] holds, i.e., there exists a unique almost equilibrium $u^{\bar{A}}(w)$ defined on a θ^R-invariant set $\Omega^*\in\mathcal{F}^R$ with $\mathbb{P}(\Omega^*)=1$, such that for any random variable $v(w)$ possessing the property $0\leq v(w)\leq\eta u^{\bar{A}}(w)$ for all $w\in\Omega^*$ and deterministic η, the

following holds:

$$\lim_{k\to\infty} \varphi(k, \theta^R_{-k} w, v(\theta^R_{-k} w)) = u^{\bar{A}}(w), \quad w \in \Omega^*. \tag{4.46}$$

Further incorporating Lemma 17 in [30], we have the following theorem regarding the weak convergence of the sequence $\left\{\widetilde{P}(k)\right\}$:

Theorem 4.3.4 Under the assumption of connectivity, there exists a unique probability measure $\mu^{\bar{A}}$ (functional of the stochastic matrix $\bar{A}$), such that for each $n \in \{1, \cdots, N\}$, the sequence $\left\{\widetilde{P}(k)\right\}$ converges weakly to $\mu^{\bar{A}}$ from every initial condition $P_n(0)$:

$$\widetilde{P}(k) \Rightarrow \mu^{\bar{A}}, \quad \forall n \in \{1, \cdots, N\}. \tag{4.47}$$

After establishing Theorem 4.3.4, following the logic in the proof for Theorem 10 in [30] (so the proof is skipped here), we now present the key result characterizing the convergence property of the sequence $\left\{\widehat{P}^n_{k|k-1}\right\}$.

Theorem 4.3.5 Under the assumption of connectivity, denote n as the index of the sensor (uniformly) randomly selected from the whole set of sensors $\{1, \cdots, N\}$. Then the sequence $\left\{\widehat{P}^n_{k|k-1}\right\}$ converges weakly to the unique probability measure $\mu^{\bar{A}}$ as in Theorem 4.3.4, i.e.,

$$\widehat{P}^n_{k|k-1} \Rightarrow \mu^{\bar{A}}. \tag{4.48}$$

For the optimal algorithm with the error variance sequence $\left\{\widehat{P}^n_{k|k-1}\right\}$ in (4.13) taking the optimal gain K^{n*}_k in (4.12), the convergence or consensus property over the network can be easily established based on the above analysis on the suboptimal case. Specifically, we could define the corresponding functions for the optimal algorithm, in the same way how f_n is defined at the beginning of Section 4.3.1. Then with

the interacting particle representation, we can construct the RDS formulation for the corresponding auxiliary sequence. Since we have established the stochastically boundedness in the suboptimal case, the stochastically boundedness of $\left\{\widehat{P}^n_{k|k-1}\right\}$ in the optimal algorithm is automatically established. Finally, with the properties of the constructed RDS, we can prove the convergence property for the optimal algorithm.

4.4 Appendix: Derivation of (4.10)

We have

$$\begin{aligned}
&\mathbb{E}'\left[\varepsilon_{\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}} \left| q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right)\right.\right] \\
&= \mathbb{E}^*\mathbb{E}'\left[\varepsilon_{\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}} \left| q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right), \widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right.\right] \\
&= \mathbb{E}^*\left[q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right) - \widehat{P}^{n_k^{\rightarrow}}_{k|k-1} \left| q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right), \widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right.\right],
\end{aligned} \tag{4.49}$$

where $\mathbb{E}^*$ is over $\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}$. From the nature of dithered quantization, conditioned on $q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right)$, $\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}$ is within the interval $\left(q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right) - \Delta, q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right) + \Delta\right)$. To be concise, in the sequel we write $q\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}\right)$ as $q(\cdot)$. Then we have (4.49) equal to

$$\int_{q(\cdot)-\Delta}^{q(\cdot)+\Delta} (q(\cdot) - p)\, f_{\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}}(p|q(\cdot))dp, \tag{4.50}$$

where $f_{\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}}(p|q(\cdot))$ is obtained from the Bayesian rule as

$$\begin{aligned}
f_{\widehat{P}^{n_k^{\rightarrow}}_{k|k-1}}(p|q(\cdot)) &= \frac{P\left(q(\cdot)|\widehat{P}^{n_k^{\rightarrow}}_{k|k-1} = p\right) P\left(\widehat{P}^{n_k^{\rightarrow}}_{k|k-1} = p\right)}{P(q(\cdot))} \\
&= \begin{cases} \frac{1}{\Delta}\left(1 - \frac{q(\cdot)-p}{\Delta}\right), & \text{if } q(\cdot) - \Delta < p \le q(\cdot) \\ \frac{1}{\Delta}\left(1 - \frac{p-q(\cdot)}{\Delta}\right), & \text{if } q(\cdot) < p < q(\cdot) + \Delta, \end{cases}
\end{aligned} \tag{4.51}$$

with the last equation based on the dithered quantization rule and the uninform prior distribution for $\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}$ from the non-informative perspective. Then it is easy to derive that (4.50) equals to zero, i.e.,

$$\mathbb{E}'\left[\varepsilon_{\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}} \left| q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right)\right.\right] = 0. \tag{4.52}$$

We also have

$$\begin{aligned}
&\mathbb{E}'\left[\varepsilon^2_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}} \left| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right.\right] \\
&= \mathbb{E}^*\left[\left(q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right) - \widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)^2 \left| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right), \widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right.\right], \\
&= \int_{q(\cdot)-\Delta}^{q(\cdot)+\Delta} \left(q(\cdot) - x\right)^2 f_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}(x|q(\cdot))dx,
\end{aligned} \tag{4.53}$$

where $\mathbb{E}^*$ is over $\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}$ and $q(\cdot)$ denotes $q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)$. Similarly as solving (4.50), we could derive $f_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}}(x|q(\cdot))$ and further calculate (4.53) as

$$\mathbb{E}'\left[\varepsilon^2_{\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}} \left| q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right.\right] = \frac{\Delta^2}{6}. \tag{4.54}$$

5. QUANTIZATION BASED ALGORITHM (QGIKF) WITH FINITE QUANTIZATION ALPHABET*

In this chapter, we extend the analysis with infinite quantization alphabet in Chapter 4 to the quantization with finite quantization alphabet, which is more practical and constrained. With the finite quantization alphabet, more information loss would occur with limiting the quantization output within the finite range. It is an interesting question to study whether or not there exists certain (possible modified) quantized GIKF algorithm can still achieve weak consensus. To seek a positive answer, we will propose a modified quantized GIKF (M-QGIKF) scheme and study its convergence performance.

5.1 M-QGIKF Scheme

In this section, we study the case with a finite quantization alphabet, i.e., each inter-sensor communication channel adopts a uniform quantizer with $\lceil \log_2(2L+1) \rceil$ bits and a step size Δ, where L is a positive integer. In other words, the quantization alphabet is

$$\widetilde{\mathcal{Q}} = \{l\Delta | l = 0, \pm 1, \cdots, \pm L\}. \tag{5.1}$$

We claim that this quantizer saturates if the quantization input exceeds the range $[-(L+\frac{1}{2})\Delta, (L+\frac{1}{2})\Delta)$. The dither v is defined the same as the case of countable infinite alphabet, i.e., v is a i.i.d. uniformly distributed random variable over $[-\frac{\Delta}{2}, \frac{\Delta}{2})$.

The QGIKF algorithm with the finite quantization alphabet could be modified

*Part of this chapter is reprinted, with permission, from [Di Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui, "Distributed Kalman filtering with quantized sensing state." *IEEE Transactions on Signal Processing*, 63(19):5180–5193, Oct. 2015.]

as follows. Operate the algorithm in the same way as QGIKF with an infinite quantization alphabet, when the quantizer does not saturate for the communication pair of sensors n and l , i.e., when the following two conditions at time k are both satisfied:

$$\left|\widehat{x}^i_{k|k-1} + v\right| < (L + 0.5)\Delta,\ i = n, l, \tag{5.2}$$

$$\left|\widehat{P}^i_{k|k-1} + v\right| < (L + 0.5)\Delta,\ i = n, l. \tag{5.3}$$

Whenever either one of above conditions is violated, i.e., the quantizer saturates, a message indicating the status of quantization saturation will be shared between the communication pair. Then at the end of slot k, sensor n acts like no information received from its communication counterpart and updates its state $\widehat{P}^n_{k+1|k}$ with its own previous state $\widehat{P}^n_{k|k-1}$ as in (4.18); similar rules apply to sensor l. Note that when the quantizer does not saturate, each sensor behaves the same as for the QGIKF with an infinite quantization alphabet, by updating its state $\widehat{P}^n_{k+1|k}$ as (4.16) with the swapped state from its communication counterpart.

In the sequel, we will study the asymptotic property of the error variance sequence $\left\{\widehat{P}^n_{k+1|k}\right\}$ described above, to show that the network still achieves weak consensus with the finite quantization alphabet under certain conditions.

5.2 Weak Consensus Analysis

In this section, we first present the RDS formulation for the estimation error variance sequence of the M-QGIKF. Then, we prove the weak consenus results.

5.2.1 RDS Formulation

Recall the interacting particle representation of $\widehat{P}^n_{k|k-1}$ for QGIKF with an infinite quantization alphabet in Section 4.3.2, where a sequence of iteration for $P_n(k)$ is

defined as

$$P_n(k+1) = f_{p_n(k)}\left(P_n(k)\right),$$

where $f_{p_n(k)} = f_{1,p_n(k)}$ if $p_n(k) \neq p_n(k+1)$, $f_{p_n(k)} = f_{2,p_n(k)}$ if $p_n(k) = p_n(k+1)$, and the process $\{p_n(k)\}$ is a Markov chain on $V = [1, \cdots, N]$ with transition probability matrix $\bar{A}$. Then, with the Markov chain $\{p_n(k)\}$, the relation between $\left\{\widehat{P}^n_{k|k-1}\right\}$ and $\{P_n(k)\}$ could be shown as

$$\left(P_1(k), \cdots, P_N(k)\right) = \left(\widehat{P}^{p_1(k)}_{k|k-1}, \cdots, \widehat{P}^{p_N(k)}_{k|k-1}\right). \tag{5.4}$$

In this section, we will represent $\widehat{P}^n_{k|k-1}$ for QGIKF with a finite quantization alphabet by another interacting particle process. With the condition in (5.2) and (5.3), the probability of saturation $P^{n,l}_{\text{sat},k}$ for the communication pair of sensors n and l at time slot k is defined as

$$\begin{aligned} P^{n,l}_{\text{sat},k} =& \mathbb{P}\left[\left(\left|\widehat{x}^n_{k|k-1} + v\right| \geq (L+0.5)\Delta\right)\right. \\ & \bigcup \left(\left|\widehat{P}^n_{k|k-1} + v\right| \geq (L+0.5)\Delta\right) \\ & \bigcup \left(\left|\widehat{x}^l_{k|k-1} + v\right| \geq (L+0.5)\Delta\right) \\ & \left. \bigcup \left(\left|\widehat{P}^l_{k|k-1} + v\right| \geq (L+0.5)\Delta\right)\right]. \end{aligned} \tag{5.5}$$

Recall $\bar{A}$ in Section 2.1.2. We define another symmetric stochastic matrix [1] $\bar{A}^F(k)$ at time slot k with the element $\bar{A}^F(k)_{nl}$ as

$$\bar{A}^F(k)_{nl} = \bar{A}_{nl}(1 - P^{n,l}_{\text{sat},k}),\ \forall n, l \in \{1, \cdots, N\},\ n \neq l,$$

[1] To distinguish from the terms used in the case of an infinite quantization alphabet, we add the superscript F over terms in this section.

$$\bar{A}^F(k)_{nn} = \bar{A}_{nn} + \sum_{i=1,\cdots,N,i\neq n} \bar{A}_{ni} P^{n,i}_{\text{sat},k},\ \forall n \in \{1,\cdots,N\}. \tag{5.6}$$

Note that, with a finite quantization alphabet, the formation process of inter-sensor communication links for state swapping can still be represented by N particles moving on the graph as a Markov chain. The state of the n-th particle at time k is denoted by $p_n^F(k)$, where $p_n^F(k)$ takes value in the state space $[1, ..., N]$, and the transition of the n-th particle is given by

$$p_n^F(k) = (p_n^F(k-1))^{\rightarrow}_k,\ p_n^F(0) = n. \tag{5.7}$$

For each n, the process $\left\{p_n^F(k)\right\}$ is a Markov chain on $V = [1,\cdots,N]$ with a symmetric transition probability matrix $\bar{A}^F(k)$. We define a sequence of iteration $P_n^F(k)$ as

$$P_n^F(k+1) = f_{p_n^F(k)}\left(P_n^F(k)\right),$$

where $f_{p_n^F(k)} = f_{1,p_n^F(k)}$ if $p_n^F(k) \neq p_n^F(k+1)$; and $f_{p_n^F(k)} = f_{2,p_n^F(k)}$ if $p_n^F(k) = p_n^F(k+1)$.

Then, similar to (5.4), with the Markov chain $\{p_n^F(k)\}$, the relation between $\left\{\widehat{P}^n_{k|k-1}\right\}$ and $\{P_n^F(k)\}$ could be shown as

$$\left(P_1^F(k),\cdots,P_N^F(k)\right) = \left(\widehat{P}^{p_1^F(k)}_{k|k-1},\cdots,\widehat{P}^{p_N^F(k)}_{k|k-1}\right),$$

from which we see that the properties of the sequence of interest $\left\{\widehat{P}^n_{k|k-1}\right\}$ could be obtained by studying the corresponding sequence $\{P_n^F(k)\}$.

Similar to Section 4.3.2, in order to perform the analysis based on RDS to analyze $\{P_n^F(k)\}$, we need construct an auxiliary sequence $\left\{\widetilde{P}^F(k)\right\}$ with a Markov chain $\left\{\widetilde{p}^F(k)\right\}$, which has the transition matrix $\bar{A}^F(k)$ and an uniform initial distribution $\mathbb{P}[\widetilde{p}^F(0) = n] = 1/N,\ n = 1,\cdots,N$. Since the transition matrix $\bar{A}^F(k)$ is symmetric,

$\left\{\widetilde{p}^F(k)\right\}$ is a Markov chain with a stationary distribution as the uniform distribution. For $\left\{\widetilde{P}^F(k)\right\}$, with a random initial condition $\widetilde{P}(0)$, it is defined as

$$\widetilde{P}^F(k+1) = f_{\widetilde{p}(k)}(\widetilde{P}^F(k)), \tag{5.8}$$

where $f_{\widetilde{p}^F(k)} = f_{1,\widetilde{p}^F(k)}$ if $\widetilde{p}^F(k) \neq \widetilde{p}^F(k+1)$; and $f_{\widetilde{p}^F(k)} = f_{2,\widetilde{p}^F(k)}$ if $\widetilde{p}^F(k) = \widetilde{p}^F(k+1)$.

Now in order to execute the asymptotic analysis of the auxiliary sequence $\left\{\widetilde{P}^F(k)\right\}$, we can construct a RDS $(\theta^{R,F}, \varphi^{R,F})$ equivalent to the auxiliary sequence $\left\{\widetilde{P}^F(k)\right\}$ in the sense of distribution. The construction process is similar to that in our previous paper [30] and that in Section 4.3.2, such that the details are skipped here. Briefly, we restate the definitions as follows. Denote $\theta^{R,F} = (\Omega^{R,F}, \mathcal{F}^{R,F}, \mathbb{P}^{R,F}, \{\theta_k^{R,F}, k \in \mathbb{T}\})$ as a metric dynamic system, where $(\Omega^{R,F}, \mathcal{F}^{R,F}, \mathbb{P}^{R,F})$ is a probability space and the family of transformations $\{\theta_k^{R,F}\}$ on $\Omega^{R,F}$ is the family of left-shifts, i.e., $\theta_k^{R,F} w = w(k+\cdot),\ \forall k \in \mathbb{T},\ w \in \Omega^{R,F}$; the cocycle $\varphi^{R,F} :\ \mathbb{T}_+ \times \Omega^{R,F} \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ is defined as: $\forall k > 1, w, X$,

$$\begin{aligned}
\varphi^{R,F}(0, w, X) &= X, \\
\varphi^{R,F}(1, w, X) &= f_{w_0}(X), \\
\varphi^{R,F}(k, w, X) &= f_{\theta_{k-1}^{R,F} w(0)}\left(\varphi^{R,F}(k-1, w, X)\right) \\
&= f_{w_{k-1}}\left(\varphi^{R,F}(k-1, w, X)\right).
\end{aligned}$$

Based on the above construction of $(\theta^{R,F}, \varphi^{R,F})$, the sequence $\left\{\varphi^{R,F}\left(k, w, P_n^F(0)\right)\right\}$ is distributionally equivalent to the sequence $\left\{\widetilde{P}^F(k)\right\}$. We now analyze the asymptotic distributional properties of the sequence $\left\{\widetilde{P}^F(k)\right\}$ by utilizing the properties in RDS, which is presented in the next subsection.

5.2.2 Main Results on Weak Consensus

Lemma 5.2.1 If the system shown in (4.1) is stable, i.e., $|F| < 1$, and the quantization parameter $L\Delta$ is sufficiently large such that $L\Delta > \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}$, in the asymptotic case as $k \to \infty$, we have the probability of saturation $P_{\text{sat},k}^{n,l}$ strictly less than 1, and $P_{\text{sat},k}^{n,l} \to 0$, as $L\Delta \to \infty$.

The proof of Lemma 5.2.1 is presented in Appendix 5.4.1.

The following theorem shows that for the QGIKF algorithm with a finite quantization alphabet, the network can still achieve weak consensus under certain conditions.

Theorem 5.2.2 For the QGIKF algorithm with a finite quantization alphabet, under the assumption of connectivity, denote n as the index of the sensor (uniformly) randomly selected from the whole set of sensors $\{1, \cdots, N\}$. If the system shown in (4.1) is stable, i.e., $|F| < 1$, and the quantization parameter $L\Delta$ is sufficiently large such that $L\Delta > \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}$, the sequence $\left\{\widehat{P}_{k|k-1}^{n}\right\}$ converges weakly to a unique probability measure $\mu^{\bar{A},L\Delta}$ (functional of the matrix $\bar{A}$ and the quantization parameter $L\Delta$) as $k \to \infty$, i.e.,

$$\widehat{P}_{k|k-1}^{n} \Rightarrow \mu^{\bar{A},L\Delta}.$$

Proof 13 According to (5.6), $\bar{A}^F(k)_{nl}$ is positive whenever $\bar{A}_{nl}$ defined in Section 2.1.2 is positive in the asymptotic case as $k \to \infty$, which is implied by the fact that $P_{\text{sat},k}^{n,l}$ is strictly less than 1 from Lemma 5.2.1. Then the stochastic matrix $\bar{A}^F(k)$ inherits the irreducibility and aperiodicity from that of $\bar{A}$, which implies the connectivity of the communication network for the QGIKF with the finite quantization alphabet in the asymptotic case as $k \to \infty$.

According to (5.16), it is easy to show that the sequence $\left\{\widetilde{P}^F(k)\right\}$ is stochastically

bounded, in a similar way as for Lemma 4.3.3. With the property of stochastic boundedness for the sequence $\{\widetilde{P}^F(k)\}$, Lemma 6.1 in [32], and Theorem 27 in [30], we again could conclude that only claim b) in Theorem 27 of [30] holds, i.e., there exists a unique almost equilibrium $u^{\bar{A},L\Delta}(w)$ defined on a $\theta^{R,F}$-invariant set $\Omega^* \in \mathcal{F}^{R,F}$ with $\mathbb{P}(\Omega^*) = 1$, such that for any random variable $v(w)$ possessing the property $0 \leq v(w) \leq \eta u^{\bar{A},L\Delta}(w)$ for all $w \in \Omega^*$ and deterministic η, the following holds:

$$\lim_{k\to\infty} \varphi(k, \theta^{R,F}_{-k} w, v(\theta^{R,F}_{-k} w)) = u^{\bar{A},L\Delta}(w), \ w \in \Omega^*.$$

Further incorporating Lemma 17 in [30], we have the following statement regarding the weak convergence of the sequence $\left\{\widetilde{P}^F(k)\right\}$:

Under the assumption of connectivity, there exists a unique probability measure $\mu^{\bar{A},L\Delta}$ (functional of the stochastic matrix $\bar{A}$ and the quantization parameter $L\Delta$), such that for each $n \in \{1, \cdots, N\}$, the sequence $\left\{\widetilde{P}^F(k)\right\}$ converges weakly to $\mu^{\bar{A},L\Delta}$ from every initial condition $P^F_n(0)$:

$$\widetilde{P}^F(k) \Rightarrow \mu^{\bar{A},L\Delta}, \quad \forall n \in \{1, \cdots, N\}.$$

Then, following the logic in the proof for Theorem 10 in [30], we can prove Theorem 5.2.2 to characterize the convergence property of the sequence $\left\{\widehat{P}^n_{k|k-1}\right\}$ for the QGIKF algorithm with a finite quantization alphabet.

5.3 Simulation Results

The simulation is based on a network with five sensors and an adjacency matrix satisfying the connectivity requirement of the network. The parameters C_n and R_n in the observation model (4.2) are selected differently for various sensors, resulting in different estimation error variances when each sensor running its own local Kalman

filter without cooperation. Denote $\mathbf{C} = [C_1, \cdots, C_5]$ and $\mathbf{R} = [R_1, \cdots, R_5]$, where for our simulation, we set $\mathbf{C} = [1\ 0\ 2\ 1.5\ 1]$, $\mathbf{R} = [2\ 3\ 4\ 8\ 5]$, $F = 0.9$, and $Q = 1$. We run the simulation for the QGIKF with a finite quantization alphabet $\widetilde{\mathcal{Q}}$ in (5.1). We set the quantization step $\Delta = 1$ and $L = 14$, to satisfy the condition $L\Delta > \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2} = 13.52$. According to the fact that the quantizer is with $\lceil \log_2(2L+1) \rceil$ bits, we have a 5-bit quantizer here. The QGIKF algorithm runs with $1,000$ iterations to ensure the convergence. We simulate the optimal estimation error variance of the QGIKF algorithm with gain K_k^{n*} in (4.12) for $5,000$ times and calculate the corresponding empirical cumulative distribution function (CDF).

In Fig. 5.1, we show the comparison between the empirical CDFs for the convergence measure of the QGIKF and that of the non-quantized GIKF in [30]. Since QGIKF involves more error or information loss due to quantization, the performance of the estimation error variance with QGIKF is worse than (but close to) that of the non-quantized GIKF. In Fig. 2, we show the performance of the QGIKF versus the non-cooperation scheme, i.e., each sensor runs its own local Kalman filter such that there is no information exchange among the sensors. The probability histograms in Fig. 5.2 illustrate the statistic of the convergence measure obtained with QGIKF and the statistic of the error variance obtained with the non-cooperation scheme, by uniformly selecting a sensor index. Compared with the non-cooperation scheme, the proposed QGIKF provides much more chances to achieve a lower estimation error variance, which demonstrates the advantage of cooperation even with quantization incorporated in the inter-sensor communications.

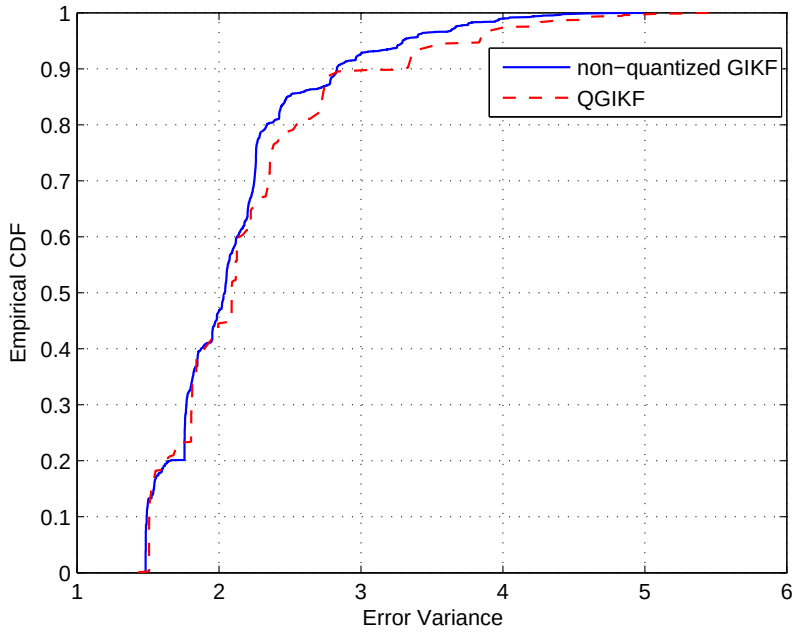

Figure 5.1: Empirical CDFs for the measure $\mu^{\bar{A}}$ in the QGIKF and the non-quantized GIKF.

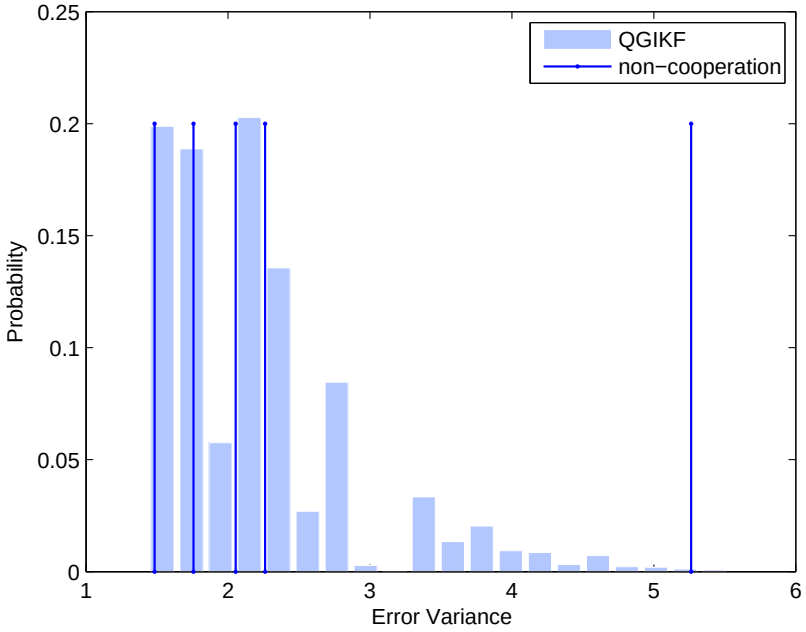

Figure 5.2: Histogram of the measure $\mu^{\bar{A}}$ in the QGIKF and the error variance distribution in the non-cooperation algorithm.

5.4 Appendices

5.4.1 Proof of Lemma 5.2.1

Recall Proposition 4.3.1 (ii):

$$f_n(X) < F^2[X + Y(\Delta)] + Q. \tag{5.9}$$

For a sufficiently large $J > 0$, we define an auxiliary value $k(J)$ as

$$k(J) = \max_k \left\{ k \in \mathbb{T}_+ \,\middle|\, \alpha + \frac{1 - F^{2k}}{1 - F^2}(Q + F^2 Y(\Delta)) \leq J \right\}, \tag{5.10}$$

where α is the bound in Lemma 4.3.2.

Note that the first part below is very similar to the proof for Lemma 4.3.3; we restate them here for self-completeness. For integers $k_0, k_1 \geq l$, the statement "there exists a $(n_1, n_2, \cdots, n_l)$ walk in the time interval $[k_0, k_1]$" indicates the existence of an integer $k_0 + l - 1 \leq k' \leq k_1$ such that

$$\widetilde{p}^F(k' - l + s) = n_s, \quad 1 \leq s \leq l, \tag{5.11}$$

where $\{\widetilde{p}^F(k)\}_{k \in \mathbb{T}_+}$ is a Markov chain.

Now we make the following claim regarding the probability $\mathbb{P}(\widetilde{P}^F(k) > J)$:

$$\begin{aligned} &\mathbb{P}(\widetilde{P}^F(k) > J) \\ &\leq \mathbb{P}(\text{no } (n_1, n_2, \cdots, n_l) \text{ exists over } [k - k(J), k]). \end{aligned} \tag{5.12}$$

In the sequel, we justify the correctness of this claim.

First assume the contrary that a walk $(n_1, n_2, \cdots, n_l)$ exists in the interval $[k -$

$k(J), k]$. Then there exists $k' \in [k-k(J), k]$ such that $\widetilde{p}^F(k'-l+s) = n_s,\ 1 \le s \le l$, which implies that $\widetilde{P}^F(k') = f_{n_l} \circ \cdots \circ f_{n_1}(\widetilde{P}^F(k'-l+1))$. Hence, by Lemma 4.3.2, we have $\widetilde{P}^F(k') \le \alpha$. By (5.9), $\widetilde{P}^F(s) < F^2[\widetilde{P}^F(s-1) + Y(\Delta)] + Q,\ \forall s$, and run the iterations, we have

$$\begin{aligned}\widetilde{P}^F(k) &< F^{2(k-k')}\widetilde{P}^F(k') + \frac{1-F^{2(k-k')}}{1-F^2}(Q + F^2Y(\Delta)) \\ &< \alpha + \frac{1-F^{2(k-k')}}{1-F^2}(Q + F^2Y(\Delta)) \\ &\le \alpha + \frac{1-F^{2(k(J))}}{1-F^2}(Q + F^2Y(\Delta)) \le J,\end{aligned}$$

where the second inequality is based on $\widetilde{P}^F(k') \le \alpha$ and $F < 1$, the third inequality follows from $k - k' \le k(J)$ by the fact that $k' \in [k-k(J), k]$, and the last inequality follows from (5.10).

Thus we can conclude that the existence of a walk $(n_1, n_2, \cdots, n_l)$ in the interval $[k–k(J), k]$ implies that $\widetilde{P}^F(k) < J$, i.e., the event $\{$Existence of a walk $(n_1, n_2, \cdots, n_l)$ over $[k-k(J), k]\} \subset \left\{\widetilde{P}^F(k) < J\right\}$, which implies the claim in (5.12). Now we study the probability of a $(n_1, \cdots, n_l)$ walk existing over $[k-k(J), k]$ in (5.12). The remaining procedure is to construct another Markov chain, and relate the probability of a $(n_1, \cdots, n_l)$ walk existing over $[k-k(J), k]$ to the hitting time statistics of this Markov chain.

We construct another Markov chain $\{z(k)\}_{k \ge l}$ with the state space $\mathcal{Z}$, as a subset of V^l, given by

$$\mathcal{Z} = \{z = (i_1, i_2, \cdots, i_l) | \bar{A}_{i_j i_{j+1}}, 1 \le j < l\}, \tag{5.13}$$

and the dynamics is given in terms of the Markov chain $\left\{\widetilde{p}^F(k)\right\}_{k \in \mathbb{T}_+}$ as:

$$z(k) = (\widetilde{p}^F(k-l+1), \cdots, \widetilde{p}^F(k)).$$

From the dynamics of $\left\{\widetilde{p}^F(k)\right\}_{k\in\mathbb{T}_+}$, we have that $\{z(k)\}_{k\geq l}$ is a Markov chain with transition probability $\bar{A}^F(k)_{nl}$ between allowable states $(i_1, i_2, \cdots, i_{l-1}, n)$ and $(i_2, \cdots, i_{l-1}, n, l)$.

Denote the hitting time τ_0 of $z(k)$ to the state $(n_1, \cdots, n_l)$ as

$$\tau_0 = \min\{k > l | z(k) = (n_1, \cdots, n_l)\}.$$

For each $k \geq l$ and a sufficiently large J, we define the stopping time τ^J as

$$\tau^J = \min\{k \geq k - k(J) \,|\, z(k) = (n_1, \cdots, n_l)\}.$$

For all $z \in \mathcal{Z}$, define

$$\mathbb{P}_z(\tau_0 > h) = \mathbb{P}(\tau_0 > h \,|\, z(l) = z).$$

Then from the Markovian property, it follows that

$$\mathbb{P}(\tau^J > k \,|\, z(k - k(J) - 1) = z) = \mathbb{P}_z(\tau_0 > k(J) + 1).$$

Therefore, we have the following results:

$$\begin{aligned}
&\mathbb{P}(\text{no } (n_1, \cdots, n_l) \text{ exists in } [k - k(J), k]) = \mathbb{P}(\tau^J > k) \\
&= \sum_{z\in\mathcal{Z}} \big[\mathbb{P}(\tau^J > k \,|\, z(k - k(J) - 1) = z) \\
&\qquad \mathbb{P}(z(k - k(J) - 1) = z)] \\
&= \sum_{z\in\mathcal{Z}} [\mathbb{P}_z(\tau_0 > k(J) + 1)\mathbb{P}(z(k - k(J) - 1) = z)] \qquad (5.14)
\end{aligned}$$

Since the above result is established for all $k \geq k_0$ with some sufficient large k_0, we

have from (5.12) that

$$\begin{aligned}
&\sup_{k\geq k_0} \mathbb{P}\left(\widetilde{P}^F(k) > J\right) \\
&\leq \sum_{z\in\mathcal{Z}} [\mathbb{P}_z(\tau_0 > k(J)+1)\mathbb{P}(z(k-k(J)-1)=z)]
\end{aligned} \tag{5.15}$$

From (5.10), we see that, if $J \geq \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}$, $k(J) \to \infty$. Then, with $J \geq \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}$, due to the recurrence of the finite state Markov chain $\{z(k)\}$ and the fact that $k(J) \to \infty$, we have

$$\mathbb{P}_z(\tau_0 > k(J)+1) = 0.$$

Since $\mathcal{Z}$ is finite, (5.15) implies that

$$\sup_{k\geq k_0} \mathbb{P}\left(\widetilde{P}^F(k) > J\right) = 0, \ \forall J \geq \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}. \tag{5.16}$$

From (5.5), we have

$$\begin{aligned}
P^{n,l}_{\text{sat},k} \leq& \mathbb{P}\left[\left(\left|\widehat{x}^n_{k|k-1}+v\right| \geq (L+0.5)\Delta\right)\right] \\
&+ \mathbb{P}\left[\left(\left|\widehat{P}^n_{k|k-1}+v\right| \geq (L+0.5)\Delta\right)\right] \\
&+ \mathbb{P}\left[\left(\left|\widehat{x}^l_{k|k-1}+v\right| \geq (L+0.5)\Delta\right)\right] \\
&+ \mathbb{P}\left[\left(\left|\widehat{P}^l_{k|k-1}+v\right| \geq (L+0.5)\Delta\right)\right] \\
\leq& \mathbb{P}\left(\left|\widehat{x}^n_{k|k-1}\right| \geq L\Delta\right) + \mathbb{P}\left(\widehat{P}^n_{k|k-1} \geq L\Delta\right) \\
&+ \mathbb{P}\left(\left|\widehat{x}^l_{k|k-1}\right| \geq L\Delta\right) + \mathbb{P}\left(\widehat{P}^l_{k|k-1} \geq L\Delta\right)
\end{aligned} \tag{5.17}$$

where the last equality is due to the fact that $v \in [-\frac{\Delta}{2}, \frac{\Delta}{2})$.

Recall the construction process of $\widehat{P}^n_{k|k-1}$ from (5.7) to (5.8), and the given condition $L\Delta > \alpha + \frac{Q+F^2Y(\Delta)}{1-F^2}$. From (5.16), we have

$$\mathbb{P}\left(\widehat{P}^n_{k|k-1} \geq L\Delta\right) = 0, \ \forall k \geq k_0, \ \forall n \in \{1, \cdots, N\}. \tag{5.18}$$

Recall the system model in (4.1), we have

$$x_{k+1} = F^{k+1}x_0 + (F^k + F^{k-1} + \cdots + 1)w, \tag{5.19}$$

where x_0 is the initial state and w is a zero mean Gaussian distributed random variable with variance Q.

In the asymptotic case as $k \to \infty$, we have that x_k converges to a Gaussian distribution as $x_k \sim N\left(0, Q/(1-F)^2\right)$. Due to the fact that the dithered quantizer generates the output as an unbiased estimate of the input [2], and the Kalman filter is an unbiased estimator of the unknown state [27], with (4.6), we have

$$\mathbb{E}(\widehat{x}^n_{k|k-1}) = \mathbb{E}(x_k).$$

Then, in the asymptotic case as $k \to \infty$, $\mathbb{E}(\widehat{x}^n_{k|k-1}) = \mathbb{E}(x_k) = 0$. From (4.8) and

(5.18), we have that

$$\begin{aligned}
&\mathbb{E}\left[\left(x_k - \widehat{x}^n_{k|k-1}\right)^2\right] \\
&= \mathbb{E}\left[\left(x_k - \mathbb{E}(x_k) + \mathbb{E}(\widehat{x}^n_{k|k-1}) - \widehat{x}^n_{k|k-1}\right)^2\right] \\
&= \mathbb{E}\left[(x_k - \mathbb{E}(x_k))^2\right] + \mathbb{E}\left[\left(\widehat{x}^n_{k|k-1} - \mathbb{E}(\widehat{x}^n_{k|k-1})\right)^2\right] \\
&\quad + 2\mathbb{E}\left[(x_k - \mathbb{E}(x_k))\left(\widehat{x}^n_{k|k-1} - \mathbb{E}(\widehat{x}^n_{k|k-1})\right)\right] \\
&< L\Delta
\end{aligned} \tag{5.20}$$

Then, in the asymptotic case as $k \to \infty$, since $x_k \sim N\left(0, Q/(1-F)^2\right)$ and $\mathbb{E}(\widehat{x}^n_{k|k-1}) = \mathbb{E}(x_k)$, we have the variance of $\widehat{x}^n_{k|k-1}$ as

$$\begin{aligned}
\mathrm{Var}\left(\widehat{x}^n_{k|k-1}\right) &= \mathbb{E}\left[\left(\widehat{x}^n_{k|k-1} - \mathbb{E}(\widehat{x}^n_{k|k-1})\right)^2\right] \\
&< L\Delta - \mathbb{E}\left[(x_k - \mathbb{E}(x_k))^2\right] - 2\mathbb{E}\left(x_k \widehat{x}^n_{k|k-1}\right) \\
&= L\Delta - \frac{Q}{(1-F)^2} - 2\mathbb{E}\left(x_k \widehat{x}^n_{k|k-1}\right) \\
&\leq L\Delta - \frac{Q}{(1-F)^2},
\end{aligned} \tag{5.21}$$

where the last inequality is due to the fact $\mathbb{E}\left(x_k \widehat{x}^n_{k|k-1}\right) \geq 0$, which is proved in Appendix 5.4.2.

According to the Chebyshev's inequality, in the asymptotic case as $k \to \infty$ with $\mathbb{E}(\widehat{x}^n_{k|k-1}) = 0$, we have

$$\mathbb{P}\left(\left|\widehat{x}^n_{k|k-1}\right| \geq L\Delta\right) \leq \frac{\mathrm{Var}\left(\widehat{x}^n_{k|k-1}\right)}{(L\Delta)^2} < \frac{L\Delta - \frac{Q}{(1-F)^2}}{(L\Delta)^2}, \tag{5.22}$$

where the last inequality is from (5.21).

By combining (5.17), (5.18), and (5.22), as $k \to \infty$, we conclude that

$$P_{\text{sat},k}^{n,l} < \frac{2L\Delta - \frac{2Q}{(1-F)^2}}{(L\Delta)^2} \tag{5.23}$$

We see that with a sufficiently large $L\Delta$, the probability of saturation $P_{\text{sat},k}^{n,l}$ could be maintained at a small value, and $P_{\text{sat},k}^{n,l} \to 0$, as $L\Delta \to \infty$.

5.4.2 Proof of $\mathbb{E}\left(x_k \widehat{x}_{k|k-1}^n\right) \geq 0$ *in* (5.21)

In the case of $n_k^{\rightarrow} \neq n$, recalling (4.6) and (4.7), we have

$$\widehat{x}_{k+1|k}^n = Fq\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right) + K_k^n \left[Fy_k^n - FC_n q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right]. \tag{5.24}$$

Since $y_k^n = C_n x_k + v_k^n$, we have

$$Fy_k^n = C_n F x_k + F v_k^n. \tag{5.25}$$

By plugging (5.25) into (5.24), we obtain

$$\begin{aligned}
\widehat{x}_{k+1|k}^n &= Fq\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right) + K_k^n \left[C_n F x_k + F v_k^n - FC_n q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right)\right]. \\
&= FR_n \left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right]^{-1} q\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}\right) \\
&\quad + K_k^n \left[C_n F x_k + F v_k^n\right],
\end{aligned} \tag{5.26}$$

where the last equality is due to the definition of K_k^n in (4.14). In the sequel, we write $FR_n \left[C_n^2 q\left(\widehat{P}_{k|k-1}^{n_k^{\rightarrow}}\right) + R_n\right]^{-1}$ as $\widetilde{K}_k^n$.

Then,

$$\begin{aligned}
&\mathbb{E}\left(\widehat{x}^n_{k+1|k}x_{k+1}\right) = \widetilde{K}^n_k\mathbb{E}\left(q\left(\widehat{x}^{n^{\rightarrow}_k}_{k|k-1}\right)x_{k+1}\right) + K^n_kC_nF\mathbb{E}[x_kx_{k+1}] \\
&= \widetilde{K}^n_k\mathbb{E}\left(q\left(\widehat{x}^{n^{\rightarrow}_k}_{k|k-1}\right)(Fx_k + w_{k+1})\right) \\
&\quad + K^n_kC_nF\mathbb{E}[x_k(Fx_k + w_{k+1})] \\
&= \widetilde{K}^n_kF\mathbb{E}\left(q\left(\widehat{x}^{n^{\rightarrow}_k}_{k|k-1}\right)x_k\right) + K^n_kC_nF^2\mathbb{E}[x^2_k] \\
&= \widetilde{K}^n_kF\mathbb{E}\left[\left(\widehat{x}^{n^{\rightarrow}_k}_{k|k-1} + e + v\right)x_k\right] + K^n_kC_nF^2\mathbb{E}[x^2_k] \\
&= \widetilde{K}^n_kF\mathbb{E}\left(\widehat{x}^{n^{\rightarrow}_k}_{k|k-1}x_k\right) + K^n_kC_nF^2\mathbb{E}[x^2_k], \qquad (5.27)
\end{aligned}$$

where we have the non-negative factors of $\widetilde{K}^n_kF = F^2R_n\left[C^2_nq\left(\widehat{P}^{n^{\rightarrow}_k}_{k|k-1}\right) + R_n\right]^{-1}$, $K^n_kC_n = C^2_nq\left(\widehat{P}^{n^{\rightarrow}_k}_{k|k-1}\right)\left[C^2_nq\left(\widehat{P}^{n^{\rightarrow}_k}_{k|k-1}\right) + R_n\right]^{-1}$, and $\mathbb{E}[x^2_k]$.

In the case of $n^{\rightarrow}_k = n$, recalling (4.17), we have

$$\widehat{x}^n_{k+1|k} = F\widehat{x}^n_{k|k-1} + FK^n_k\left(y^n_k - C_n\widehat{x}^n_{k|k-1}\right). \qquad (5.28)$$

By plugging (5.25) into (5.28), we obtain

$$\begin{aligned}
\widehat{x}^n_{k+1|k} &= F\widehat{x}^n_{k|k-1} + K^n_k\left(C_nFx_k + Fv^n_k - FC_n\widehat{x}^n_{k|k-1}\right). \\
&= FR_n\left(C^2_n\widehat{P}^n_{k|k-1} + R_n\right)^{-1}\widehat{x}^n_{k|k-1} \\
&\quad + K^n_k\left(C_nFx_k + Fv^n_k\right), \qquad (5.29)
\end{aligned}$$

where the last equality is due to the definition of K^n_k in (4.19). In the sequel, we

write $FR_n\left(C_n^2\widehat{P}_{k|k-1}^n+R_n\right)^{-1}$ as $\widetilde{K}_k^n$. Then,

$$\begin{aligned}
&\mathbb{E}\left(\widehat{x}_{k+1|k}^n x_{k+1}\right)=\widetilde{K}_k^n\mathbb{E}\left(\widehat{x}_{k|k-1}^n\right)x_{k+1}+K_k^nC_nF\mathbb{E}[x_kx_{k+1}]\\
&=\widetilde{K}_k^n\mathbb{E}\left(\widehat{x}_{k|k-1}^n(Fx_k+w_{k+1})\right)\\
&\quad+K_k^nC_nF\mathbb{E}[x_k(Fx_k+w_{k+1})]\\
&=\widetilde{K}_k^nF\mathbb{E}\left(\widehat{x}_{k|k-1}^nx_k\right)+K_k^nC_nF^2\mathbb{E}[x_k^2],
\end{aligned}\tag{5.30}$$

in which $\widetilde{K}_k^nF=F^2R_n\left(C_n^2\widehat{P}_{k|k-1}^n+R_n\right)^{-1}$, $K_k^nC_n=C_n^2\widehat{P}_{k|k-1}^n\left[C_n^2\widehat{P}_{k|k-1}^n+R_n\right]^{-1}$ and $\mathbb{E}[x_k^2]$ are non-negative.

Considering (5.27) and (5.30), we have that, in either case of $n_k^{\rightarrow}\neq n$ or $n_k^{\rightarrow}=n$, in order to prove $\mathbb{E}\left(\widehat{x}_{k+1|k}^nx_{k+1}\right)\geq 0$, we only need to show $\mathbb{E}\left(\widehat{x}_{k|k-1}^{n_k^{\rightarrow}}x_k\right)\geq 0$. To this end, by following the iteration (5.30), it suffices to prove $\mathbb{E}\left(\widehat{x}_{0|-1}^nx_0\right)\geq 0$, which is established as follows. Since the initial state $x_0\sim(\widehat{x}_{0|-1},\widehat{P}_{0|-1})$, for each n, we have

$$\mathbb{E}\left(\widehat{x}_{0|-1}x_0\right)=\widehat{x}_{0|-1}^2\geq 0.\tag{5.31}$$

Thus, according to (5.27) or (5.30), recursively we have that $\mathbb{E}\left(\widehat{x}_{k+1|k}^nx_{k+1}\right)\geq 0$.

6. DISTRIBUTED BAYESIAN QUICKEST CHANGE DETECTION

In this chapter, we turn to the distributed detection problem and propose a distributed Bayesian quickest change detection algorithm for sensor networks, based on a random gossip inter-sensor communication structure. Without a control or fusion center, each sensor executes its local change detection procedure in a parallel and distributed fashion, interacting with its neighbor sensors via random inter-sensor communications to propagate information. By modeling the information propagation dynamics in the network as a Markov process, two-layer large deviation analysis is presented to analyze the performance of the proposed algorithm.

6.1 System Setup

Consider a network with N nodes. Assume that a change happens at time $\lambda = k$. Then conditioned on $\lambda = k$, independent and identically distributed (i.i.d.) observations $X_1^i, \cdots, X_{k-1}^i$ at sensor i follow a distribution with density function $f_0^i(x)$; observations $X_k^i, X_{k+1}^i \cdots$ follow another distribution with density function $f_1^i(x)$. We assume that observations at different sensors are independent of each other and the various densities are absolutely continuous with respect to the Lebesgue measure. Denote $\mathbf{X}_n^i = [X_1^i, \cdots, X_n^i]$ as observations up to time n at node i. Let $\mathbb{P}_k$ be the probability measure of $\mathbf{X}_n^i$ when the change occurs at time k, and $\mathbb{E}_k$ be the corresponding expectation operator. We need to design a sequential on-line detection algorithm (with a stopping criterion) over the observation sequence to detect the change.

Consider a Bayesian setup, and assume the prior distribution for the change-point time λ as

$$\pi_k = \mathbb{P}(\lambda = k).$$

Let $\mathbb{P}^\pi$ denote the probability measure, defined as $\mathbb{P}^\pi(\cdot) = \sum_{k=1}^{\infty} \pi_k \mathbb{P}_k(\cdot)$, and let $\mathbb{E}^\pi$ denote the expectation operator with respect to the measure $\mathbb{P}^\pi$.

The change detection problem can be converted to the hypothesis testing problem with hypotheses "$H_0 : \lambda > n$" and "$H_1 : \lambda \leq n$", i.e., to sequentially decide which hypothesis is true at each time n. If H_0 is decided, it indicates that the change hasn't happened; if H_1 is decided, it claims that the change has happened.

6.1.1 Centralized Scheme

First we discuss the centralized change detection algorithm, which means that observations from all sensors are available at a control center, where the detection algorithm is performed. Denote $\mathbf{X}_n = [\mathbf{X}_n^1, \cdots, \mathbf{X}_n^N]$ as observations up to time n from all sensors; denote the likelihood ratio for "$H_1 : \lambda \leq n$" vs. "$H_0 : \lambda > n$" averaged over the change point (see [66]) as:

$$\begin{aligned} \Lambda_n &= \frac{\mathbb{P}(\mathbf{X}_n|\lambda \leq n)\mathbb{P}(\lambda \leq n)}{\mathbb{P}(\mathbf{X}_n|\lambda > n)\mathbb{P}(\lambda > n)} \\ &= \frac{\sum_{k=1}^{n} \left[\pi_k \prod_{i=1}^{N} \prod_{j=k}^{n} f_1^i(X_j^i) \prod_{j=1}^{k-1} f_0^i(X_j^i)\right]}{\sum_{k=n+1}^{\infty} \pi_k \prod_{i=1}^{N} \prod_{j=1}^{n} f_0^i(X_j^i)}. \end{aligned} \tag{6.1}$$

Assume the prior distribution is geometric [52], i.e.,

$$\pi_k = \rho(1-\rho)^{k-1}, \text{ with } \rho \text{ in } (0,1).$$

Then, we have

$$\Lambda_n = \frac{1}{(1-\rho)^n} \sum_{k=1}^{n} \pi_k \prod_{j=k}^{n} \prod_{i=1}^{N} \frac{f_1^i(X_j^i)}{f_0^i(X_j^i)}. \tag{6.2}$$

We further have the following recursive form as

$$\Lambda_n = \frac{1}{1-\rho}(\Lambda_{n-1} + \rho) \prod_{i=1}^{N} \frac{f_1^i(X_n^i)}{f_0^i(X_n^i)}, \tag{6.3}$$

with the initial state $\Lambda_0 = 0$. Taking logarithms on both sides, we have

$$\log \Lambda_n = \log \frac{1}{1-\rho} + \log(\Lambda_{n-1} + \rho) + \sum_{i=1}^{N} \log \frac{f_1^i(X_n^i)}{f_0^i(X_n^i)}. \tag{6.4}$$

Let $\mathcal{F}_n^X = \sigma(\mathbf{X}_n)$ be the $\sigma-$algebra generated by the observations $\mathbf{X}_n$, and we denote

$$p_n = \mathbb{P}\{\lambda \leq n | \mathcal{F}_n^X\} \tag{6.5}$$

as the posterior probability that the change has occurred before time n. It follows that $\Lambda_n = p_n/(1-p_n)$.

We intend to detect the change as soon as possible, with a constraint on the detection error. Thus, the change detection problem can be formulated as the following optimization problem over certain decision rules:

$$\begin{aligned} &\inf_{\tau \in \Delta(\alpha)} \mathrm{ADD}(\tau) \\ &\text{s. t. } \Delta(\alpha) = \{\tau : \mathrm{PFA}(\tau) \leq \alpha\}, \end{aligned} \tag{6.6}$$

where the Averaged Detection Delay (ADD) is

$$\mathrm{ADD}(\tau) = \mathbb{E}^{\pi}(\tau - \lambda | \tau \geq \lambda),$$

the Probability of False Alarm (PFA) is

$$\text{PFA}(\tau) = \mathbb{P}^{\pi}(\tau < \lambda) = \sum_{k=1}^{\infty} \pi_k \mathbb{P}_k(\tau < k),$$

with $\mathbb{E}^{\pi}$ and $\mathbb{P}^{\pi}$ defined at the beginning of this section, and α the upper limit of PFA.

The optimal solution to this problem is given by the Shiryaev test (see [58, 59]), where the detection strategy corresponds to claiming a change when the likelihood ratio Λ_n exceeds a threshold, i.e., the optimal stopping time τ^* is

$$\tau^*(A) = \inf\{n \geq 1 : \Lambda_n \geq A\}, \tag{6.7}$$

where A is chosen such that $\text{PFA}(\tau^*(A)) = \alpha$. It is difficult to set a threshold A exactly matching the above condition. We could set $A = (1-\alpha)/\alpha$ guaranteeing that $\text{PFA}(\tau^*(A)) \leq \alpha$, which is due to the fact that $\mathbb{P}^{\pi}(\tau^*(A) < \lambda) = \mathbb{E}^{\pi}\left(1 - p_{\tau^*(A)}\right)$ and $1-p_{\tau^*(A)} \leq 1/(1+A)$ with $p_{\tau^*(A)}$ defined in (6.5), such that $\text{PFA}(\tau^*(A)) \leq 1/(1+A)$. Therefore, setting $A = (1-\alpha)/\alpha$ guarantees $\text{PFA}(\tau^*(A)) \leq \alpha$ [66].

6.1.2 Isolated Scheme

If there is no control center and each sensor implements the local change detection algorithm purely based on its own observations, the log-likelihood ratio for hypotheses "$H_0 : \lambda \leq n$" vs. "$H_1 : \lambda > n$" of sensor i at time n is derived as

$$\log \Lambda_n^i = \log \frac{1}{1-\rho} + \log(\Lambda_{n-1}^i + \rho) + \log \frac{f_1^i(X_n^i)}{f_0^i(X_n^i)}, \tag{6.8}$$

with the initial state $\Lambda_0^i = 0$.

Then, to solve the optimization problem in (6.6) at sensor i, the Shiryaev test

with test statistic in (6.8) is the optimal solution [58, 59], with the optimal stopping time τ^{i*} at sensor i as

$$\tau^{i*}(A) = \inf\{n \geq 1 : \Lambda_n^i \geq A\}, \tag{6.9}$$

where A is chosen such that $\text{PFA}(\tau^{i*}(A)) = \alpha$. Since this detection strategy is exclusively based on local observations at each sensor, it is called the isolated scheme.

Intuitively, the larger the difference between densities $f_1^i(x)$ and $f_0^i(x)$ is, the faster the change can be detected. To quantify the difference between densities $f_1^i(x)$ and $f_0^i(x)$, the Kullback-Leibler information number is defined as

$$D(f_1^i, f_0^i) = \int \log\left\{\frac{f_1^i(x)}{f_0^i(x)}\right\} f_1^i(x)dx, \tag{6.10}$$

which is also called divergence or KL distance between densities $f_1^i(x)$ and $f_0^i(x)$. We assume a mild condition that $0 < D(f_1^i, f_0^i) < \infty$ and $0 < D(f_0^i, f_1^i) < \infty$, for each i.

In the sequel, we will show that the Kullback-Leibler information number is a crucial factor in analyzing the performance of the change detection algorithms.

6.2 Large Deviation Analysis for Centralized and Isolated Algorithms

Large deviation studies the asymptotic behavior of a rare event. Generally, for a rare event satisfying the large deviation principle, the probability of this rare event occurring decays to zero at an exponentially fast rate in the asymptotic sense over certain quantity. In this section, we analyze the performance of the centralized algorithm, by quantifying the relation between the conditional ADD and the PFA via large deviation analysis, showing that the event of false alarm can be considered as a rare event and the corresponding PFA decays to zero exponentially fast, when the conditional ADD increases. The results in this section will set the background for analyzing the distributed case in the next section.

Since ADD might be difficult to characterize, following [66], we instead analyze the conditional ADD (CADD). The CADD is defined as $\mathrm{CADD}_k(\tau) = \mathbb{E}_k(\tau - k | \tau \geq k),\ k = 1, 2, \cdots$. The relation between ADD and CADD is described as follows:

$$\begin{aligned}\mathrm{ADD}(\tau) &= \mathbb{E}^{\pi}(\tau - \lambda | \tau \geq \lambda) \\ &= \frac{\sum_{k=1}^{\infty} \pi_k \mathbb{P}_k(\tau \geq k) \mathbb{E}_k(\tau - k | \tau \geq k)}{\mathbb{P}^{\pi}\{\tau \geq \lambda\}} \\ &= \frac{\sum_{k=1}^{\infty} \pi_k \mathbb{P}_k(\tau \geq k) \mathrm{CADD}_k(\tau)}{\mathbb{P}^{\pi}\{\tau \geq \lambda\}}. \end{aligned} \tag{6.11}$$

According to the optimal stopping rule (6.7) and the test statistic (6.3), we find $\mathrm{CADD}_1(\tau^*) \geq \mathrm{CADD}_k(\tau^*)$, for $k \geq 2$, which is explained as follows. For $k = 1$ (which means that the change happens at time 1), by investigating (6.3), Λ_1 is updated based on the initial state $\Lambda_0 = 0$. For $k \geq 2$, by investigating (6.3), Λ_k is updated based on Λ_{k-1}, where $0 \leq \Lambda_{k-1} < A$ according to the optimal stopping rule (6.7) and the condition $\tau^* \geq k$. Thus, we have $\Lambda_{k-1} \geq \Lambda_0$. According to the optimal stopping rule (6.7), the spent time of crossing the threshold after the change happens (detection delay) in the case of $k \geq 2$ is less than that in the case of $k = 1$ on average. Therefore, we have $\mathrm{CADD}_1(\tau^*) \geq \mathrm{CADD}_k(\tau^*)$. Additionally, the difference between $\mathrm{CADD}_1(\tau^*)$ and $\mathrm{CADD}_k(\tau^*)$ could be treated as a constant for large A, which approximately equals $\mathbb{E}_{\infty}(\log \Lambda_{k-1}), k \geq 2$ [66]. Therefore, in the sequel, we focus on the use of $\mathrm{CADD}_1(\tau^*)$, which could be also considered as the worst-case study.

The relation between $\mathrm{CADD}_1(\tau^*)$ and $\mathrm{PFA}(\tau^*)$, for the centralized scheme, is presented in the following theorem.

Theorem 6.2.1 The probability of false alarm ($\mathrm{PFA}(\tau^*)$), with the optimal stopping rule (6.7), satisfies the large deviation principle, in the asymptotic sense with respect

to the increasing conditional ADD ($\mathrm{CADD}_1(\tau^*)$), i.e.,

$$\lim_{\mathrm{CADD}_{1(\tau^*)}\to\infty} \frac{1}{\mathrm{CADD}_1(\tau^*)} \log[\mathrm{PFA}(\tau^*)] = -(\mathcal{D} + |\log(1-\rho)|), \tag{6.12}$$

where $\mathcal{D}$ is the sum of the Kullback-Leibler information numbers across all sensors, i.e., $\mathcal{D} = \sum_{i=1}^{N} D(f_1^i, f_0^i)$, and $\mathcal{D} + |\log(1-\rho)|$ is the large deviation decay rate, quantifying how fast the probability of false alarm decays to zero over the increasing conditional ADD.

Proof 14 Recall Theorem 5 in [66], which establishes the following results:

$$\mathrm{PFA}(\tau^*) = \frac{\zeta(\rho, \mathcal{D})}{A}(1 + o(1)), \text{ as } A \to \infty; \tag{6.13}$$

$$\mathbb{E}_1(\tau^*) = \frac{1}{\mathcal{D} + |\log(1-\rho)|}\left[\log\frac{A}{\rho} - \xi(\rho, \mathcal{D})\right] + o(1), \text{ as } A \to \infty, \tag{6.14}$$

where $\mathcal{D} = \sum_{i=1}^{N} D(f_1^i, f_0^i)$, and both $\zeta(\rho, \mathcal{D})$ and $\xi(\rho, \mathcal{D})$ are functions of ρ and $\mathcal{D}$. Since ρ and $\mathcal{D}$ are constants once the system parameters are set, $\zeta(\rho, \mathcal{D})$ and $\xi(\rho, \mathcal{D})$ are also system constants.

Since $\mathrm{CADD}_1(\tau^*) = \mathbb{E}_1(\tau^* - 1) = \mathbb{E}_1(\tau^*) - 1$, by combining (6.13) and (6.14), we have

$$\log\frac{\mathrm{PFA}(\tau^*)\rho}{\zeta(\rho, \mathcal{D})(1 + o(1))} = -\mathrm{CADD}_1(\tau^*)(\mathcal{D} + |\log(1-\rho)|) - \xi(\rho, \mathcal{D}) + (o(1) - 1)(\mathcal{D} + |\log(1-\rho)|). \tag{6.15}$$

Then, after dividing the left-hand and right-hand sides of (6.15) by $\mathrm{CADD}_1(\tau^*)$ and

taking the limit as $\mathrm{CADD}_1(\tau^*) \to \infty$, we have

$$\lim_{\mathrm{CADD}_1(\tau^*)\to\infty} \frac{1}{\mathrm{CADD}_1(\tau^*)} \log \mathrm{PFA}(\tau^*) = -(\mathcal{D} + |\log(1-\rho)|). \tag{6.16}$$

The above theorem quantifies the tradeoff between two performance metrics: PFA and CADD_1, in the defined change detection problems, i.e., as CADD_1 increases, PFA decays to zero exponentially fast and the decay rate is $\mathcal{D} + |\log(1-\rho)|$.

For the isolated scheme, at each node i, the relation between $\mathrm{PFA}(\tau^{i*})$ and $\mathrm{CADD}_1(\tau^{i*})$ has a similar format to that in the centralized case shown in Theorem 6.2.1. We give the following corollary.

Corollary 6.2.2 The probability of false alarm ($\mathrm{PFA}(\tau^{i*})$), with the optimal stopping rule (6.9), satisfies the large deviation principle, in the asymptotic sense with respect to the increasing conditional ADD ($\mathrm{CADD}_1(\tau^{i*})$), i.e.,

$$\lim_{\mathrm{CADD}_1(\tau^{i*})\to\infty} \frac{1}{\mathrm{CADD}_1(\tau^{i*})} \log[\mathrm{PFA}(\tau^{i*})] = -(D(f_1^i, f_0^i) + |\log(1-\rho)|), \tag{6.17}$$

which implies that the large deviation decay rate of the PFA is $D(f_1^i, f_0^i) + |\log(1-\rho)|$.

Theorem 6.2.1 and Corollary 6.2.2 imply that the Kullback-Leibler information number is a crucial factor that determines the performance of change detection algorithms. Specifically, Corollary 6.2.2 shows that, for different sensors with different pairs of densities $f_1^i(x)$ and $f_0^i(x)$, the sensor associated with a density pair bearing a larger Kullback-Leibler information number asymptotically leads to a smaller PFA, under the same CADD performance. Compared with the isolated scheme,

Theorem 6.2.1 shows that, in the centralized scheme, the sum of Kullback-Leibler information numbers $\mathcal{D}$ is used to quantify the relation between PFA and CADD, which can be intuitively explained as follows.

In the next section, we propose a distributed change detection scheme and analyze its performance. Due to the information propagation among sensors, we show that the distributed scheme will outperform the isolated one, and the outperforming is reflected by the averaged partial sum over individual Kullback-Leibler information numbers.

6.3 Distributed Change Detection Scheme

In this section, a random gossip based distributed change detection algorithm is first introduced. Then, we model the information propagation in this distributed scheme as a Markov process. Finally, two-layer large deviation analysis is presented to analyze the performance of the proposed distributed algorithm.

First, we interpret the network as a non-directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes with $|\mathcal{V}| = N$ and $\mathcal{E}$ is the set of edges. If node i is connected to node j, then we have that edge $(i, j) \in \mathcal{E}$. The connection in graph $\mathcal{G}$ is represented by the following $N \times N$ symmetric adjacency matrix $\mathcal{A}$ with each element $\mathcal{A}_{ij}$ as:

$$\mathcal{A}_{ij} = \begin{cases} 1, & (i,j) \in \mathcal{E} \text{ or } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{6.18}$$

We assume that the network is connected, i.e., each node has a path to any other node.

We propose a random gossip based distributed change detection algorithm, where a random gossip algorithm, as the inter-sensor communication structure, is used to propagate information among sensors within the neighborhood.

Communication among sensors is constrained by factors such as proximity, transmitting power, and receiving capabilities. We model the communication structure in terms of the non-directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which is defined at the beginning of this section. If node i can communicate with node j, there is an edge existing between i and j, i.e., the set of edges $\mathcal{E}$ contains the edge (i, j). We assume that the diagonal elements in adjacency matrix $\mathcal{A}$ are identically 1, which indicates that a node can always communicate with itself. The set $\mathcal{E}$ is the maximal set of allowable communication links in the network at any time; however, at a particular instant, only a fraction of the allowable communication links are active, for example, to avoid strong interference among communications. The exact communication protocol is not that important for the theoretical analysis, as long as the connectivity of network is satisfied. For definiteness, we assume the following generic communication model, which subsumes the widely used gossip protocol for real-time embedded architectures [6] and the graph matching based communication protocols for internet architectures [43]. Define the set $\mathcal{M}$ of binary symmetric $N \times N$ matrices as follows:

$$\mathcal{M} = \left\{ \mathbf{A} | \mathbf{1}^T \mathbf{A} = \mathbf{1}^T,\ \mathbf{A}\mathbf{1} = \mathbf{1},\ \mathbf{A} \leq \mathcal{A} \right\} \tag{6.19}$$

where $\mathbf{A} \leq \mathcal{A}$ is interpreted as component-wise. In other words, $\mathcal{M}$ is the set of adjacency matrices, where each node is incident to exactly one edge, which is included in the edge set $\mathcal{E}$. Let $\mathfrak{D}$ denote a probability distribution on the space $\mathcal{M}$. We define a sequence of time-varying matrices $\mathbf{A}(m)$, $m = 1, 2, \cdots$, as an independent and identically distributed sequence in $\mathcal{M}$ with distribution $\mathfrak{D}$. Define the averaged matrix $\bar{\mathbf{A}}$ as

$$\bar{\mathbf{A}} = \int_{\mathcal{M}} \mathbf{A} d\mathfrak{D}(\mathbf{A}). \tag{6.20}$$

According to the definition of $\mathcal{M}$ in (6.19), $\bar{\mathbf{A}}$ is a symmetric stochastic matrix. We assume $\bar{\mathbf{A}}$ to be irreducible and aperiodic. This assumption depends on the allowable edges $\mathcal{E}$ and the distribution $\mathfrak{D}$. Such a distribution $\mathfrak{D}$ making this assumption valid always exists if the graph $(\mathcal{V}, \mathcal{E})$ is connected, e.g., the uniform distribution. In addition, $\bar{\mathbf{A}}$ could be interpreted as the transition matrix of a Markov chain, which we will discuss later.

Assume that the sampling time interval for taking observations is Δ, within which there are M rounds of inter-sensor communications, where M is a Poisson random variable with mean γ [6]. At the m-th ($m \in \{1, \cdots, M\}$) round, a node randomly selects another node from its neighborhood to construct a two-way communication pair to exchange the observations between them. At each sampling time interval, this communication structure is modeled by the sequence of matrices $\mathbf{A}(m)$, $m = 1, 2, \cdots, M$, i.e., the establishment of a communication link between node i and node j indicates that nodes i and j are neighbors with respect to the time varying adjacency matrix $\mathbf{A}(m)$. Note that there may exist multiple communication links or pairs simultaneously in the network, but only one communication link is associated with one given node in each round, which is also implied by the mathematical model in (6.19).

Now we model the communication link formation process from the perspective of Markov process. To this end, the communication link process governed by the time varying adjacency matrix sequence $\{\mathbf{A}(m)\}$ can be represented by N particles traveling on the graph [37]. We denote the state of the i-th particle as $z_i(m)$, where $z_i(m)$ indicates the index of node that the i-th particle travels to at time m, with $z_i(m) \in \{1, \cdots, N\}$. The evolution of the i-th particle is given as follows:

$$z_i(m) = [z_i(m-1)]_m^{\rightarrow}, \; z_i(0) = i, \tag{6.21}$$

where the notation $[i]_m^{\rightarrow}$ denotes the neighbor of node i at time m with respect to the adjacency matrix $\mathbf{A}(m)$, i.e., a communication link is established between $i_m^{\rightarrow}$ and i at time m. Thus, the travelling process of the i-th particle can be viewed as originating from node i initially and then traveling on the graph according to the link formation process $\{\mathbf{A}(m)\}$ (possibly changing its location at each step). For each i, the process $\{z_i(m)\}$ is a Markov chain on the state space $\{1, \cdots, N\}$ with the transition probability matrix $\bar{\mathbf{A}}$ [37].

After M rounds of inter-sensor communications, each node accumulates some observations from other nodes, with which the local test statistic at each node is updated. Denote O_n^i as the set of nodes whose observations are available at node i after inter-sensor communications at the end of the observation time period n. We will describe the accumulation process to obtain O_n^i later. Then, the distributed test statistic $\Lambda_{n,D}^i$ is updated as

$$\Lambda_{n,D}^i = \frac{1}{1-\rho}(\Lambda_{n-1}^i + \rho) \prod_{j \in O_n^i} \frac{f_1^j(X_n^j)}{f_0^j(X_n^j)}. \tag{6.22}$$

With this test statistic updating rule, at each sensor i, the distributed change detection scheme is executed with the following stopping time τ_D^i:

$$\tau_D^i(A) = \inf\{n \geq 1 : \Lambda_{n,D}^i \geq A\}, \tag{6.23}$$

where A is chosen as $A = (1-\alpha)/\alpha$ such that $\mathrm{PFA}(\tau_D^i(A)) \leq \alpha$.

Now we describe the observation accumulation process to obtain O_n^i. Let $\mathbf{s}_n^m = [s_n^m(1), \cdots, s_n^m(N)]$, with element $s_n^m(i) \in \{1, \cdots, N\}$ indexing the observation $X_n^{s_n^m(i)}$ at sensor i just after the m-th round of communication in the observation time period n. The initial state is $s_n^0(i) = i$ at each sensor i, which means that at

the beginning of the time slot n, each sensor i only has its own observation X_n^i. When the communication starts, by following the communication model $\mathbf{A}(m)$, the observations $\{X_n^i\}_{i\in\{1,\cdots,N\}}$ travel across the network in the following way:

$$\mathbf{s}_n^m = \mathbf{A}(m)\mathbf{s}_n^{m-1},\ 1 \le m \le M. \tag{6.24}$$

During these M rounds of inter-sensor communications until the end of the time period n, each sensor stores observations exchanged from other sensors. Then, at the end of the time period n, observations from other sensors are accumulated at sensor i, and the set of sensors whose observations are available at sensor i is denoted by

$$O_n^i = \bigcup_{m=0}^{M} \{s_n^m(i)\}. \tag{6.25}$$

This observation accumulation process terminates at the end of the time period n. Then, a similar observation accumulation process repeats during the time period $n+1$, which is independent of the previous process. Therefore, the sequence $\{O_n^i\}$, as the set denoting observation indices which are available at sensor i at the end of the n-th period, is an i.i.d. process.

To better describe our work in the sequel, we introduce some notations here. Let Ψ denote the power set of node indices $\{1,\cdots,N\}$, where elements of Ψ are indexed by ν, with $\nu \in \{0,1,\cdots,2^N-1\}$. We use Ψ_0 to denote the null set and Ψ_{2^N-1} to denote the whole set of node indices. For technical convenience, we interpret sensors in the set Ψ_ν indexed by ν to be arranged in an ascending order with j_1 denoting the first one and $j_{|\Psi_\nu|}$ denoting the last one, i.e., $\Psi_\nu = \{j_1,\cdots,j_{|\Psi_\nu|}\}$. Therefore, the set O_n^i, denoting nodes whose observations are available at node i after the observation accumulation process, is a random variable taking values from Ψ. We denote the

following probability as

$$\Pr(O_n^i = \Psi_\nu) = q_n^i(\nu),\ \nu \in \{0, 1, \cdots, 2^N - 1\}. \tag{6.26}$$

6.4 First-layer Large Deviation Analysis

To perform large deviation analysis, we first need to interpret the stopping time $\tau_D^i(A)$ as a form of random walk crossing a threshold plus a nonlinear term [66]. To this end, the stopping time $\tau_D^i(A)$ could be rewritten as:

$$\tau_D^i(A) = \inf\left\{n \geq 1 : W_n(\rho) + l_n\ \geq \log(A/\rho)\right\}, \tag{6.27}$$

where $W_n(\rho) = Z_n + n|\log(1-\rho)|$ is a random walk with

$$Z_n = \sum_{k=1}^{n} \sum_{j \in O_k^i} \log \frac{f_0^j(X_k^j)}{f_1^j(X_k^j)}, \tag{6.28}$$

and

$$l_n = \log\left\{1 + \sum_{k=1}^{n-1} (1-\rho)^k \prod_{s=1}^{k} \prod_{j \in O_s^i} \frac{f_0^j(X_s^j)}{f_1^j(X_s^j)}\right\}. \tag{6.29}$$

Specifically, $W_n(\rho)$ is a random walk with mean

$$\mathbb{E}_1\{W_n(\rho)\} = n \sum_{\nu=1}^{2^N-1} \bar{q}_\gamma^i(\nu) \sum_{j \in \Psi_\nu} D(f_1^j, f_0^j) + n|\log(1-\rho)|, \tag{6.30}$$

where $\bar{q}_\gamma^i(\nu)$ is the probability defined as

$$\bar{q}_\gamma^i(\nu) = \Pr(O_\gamma^i = \Psi_\nu),\ \nu \in \{0, 1, \cdots, 2^N - 1\}, \tag{6.31}$$

in which O_γ^i, a random variable taking values from Ψ, denotes the set of nodes whose observations are available at node i after γ rounds of communications, and γ is the mean value of the number of communication rounds. Then, based on the above random walk interpretation for the stopping time, we have the following theorem regarding the relation between PFA and CADD in the proposed distributed change detection scheme.

Theorem 6.4.1 The probability of false alarm ($\mathrm{PFA}(\tau_D^i)$), with the stopping rule (6.23) in the distributed change detection algorithm with the parameter γ as the averaged number of inter-sensor communications, satisfies the large deviation principle in the asymptotic sense with respect to increasing conditional ADD ($\mathrm{CADD}_1(\tau_D^i)$), i.e.,

$$\lim_{\mathrm{CADD}_1(\tau_D^i)\to\infty} \frac{1}{\mathrm{CADD}_1(\tau_D^i)} \log[\mathrm{PFA}(\tau_D^i)] = -(\mathcal{D}_\gamma^i + |\log(1-\rho)|), \tag{6.32}$$

where $\mathcal{D}_\gamma^i = \sum_{\nu=1}^{2^N-1} \bar{q}_\gamma^i(\nu) \sum_{j\in\Psi_\nu} D(f_1^j, f_0^j)$, and $\mathcal{D}_\gamma^i + |\log(1-\rho)|$ is the large deviation decay rate of PFA. We call $\mathcal{D}_\gamma^i$ as the distributed Kullback-Leibler information number.

Theorem 6.4.1 shows that $\mathcal{D}_\gamma^i$, whose function is similar to $\mathcal{D}$ in the centralized scheme and $D(f_1^i, f_0^i)$ in the isolated scheme, is a crucial factor determining the performance of the distributed change detection algorithm. The physical meaning of $\mathcal{D}_\gamma^i$ is explained as follows. Due to the observation propagation process, observations and the corresponding log-likelihood ratios from other sensors are available at each sensor; to some extent, $\mathcal{D}_\gamma^i$ can be considered as an accumulated form of these information. In particular, $\mathcal{D}_\gamma^i$ is an averaged partial sum of the Kullback-Leibler

information numbers $D(f_1^i, f_0^i),\ i = 1, \cdots, N$, compared to $\mathcal{D}$ as the total sum. Also, from the mathematical form of $\mathcal{D}_\gamma^i$, we see that $D(f_1^i, f_0^i) \leq \mathcal{D}_\gamma^i \leq \mathcal{D}$, and the case of $\bar{q}_\gamma^i(1) = 1$ corresponds to the lower bound $\mathcal{D}_\gamma^i = D(f_1^i, f_0^i)$, while the case of $\bar{q}_\gamma^i(2^N - 1) = 1$ corresponds to the upper bound $\mathcal{D}_\gamma^i = \mathcal{D}$. Since $D(f_1^i, f_0^i) \leq \mathcal{D}_\gamma^i \leq \mathcal{D}$ and $\mathcal{D}_\gamma^i$ determines the performance of the change detection algorithm, the above analysis proves that the distributed algorithm outperforms the isolated algorithm, but falls behind the centralized algorithm.

We present the proof for the above theorem as follows.

Proof 15 The proof adopts the relevant results from the nonlinear renewal theory in [72]. To complete the proof, we first present two preliminary results, regarding the proposed distributed algorithm, as follows:

$$\mathrm{PFA}(\tau_D^i) = \frac{\zeta(\rho, \mathcal{D}_\gamma^i)}{A}(1 + o(1)),\ \text{as } A \to \infty, \tag{6.33}$$

$$\mathbb{E}_1(\tau_D^i) = \frac{1}{\mathcal{D}_\gamma^i + |\log(1-\rho)|}\left[\log\frac{A}{\rho} - \xi(\rho, \mathcal{D}_\gamma^i)\right] + o(1), \quad \text{as } A \to \infty, \tag{6.34}$$

where $\mathcal{D}_\gamma^i$ is defined below (6.32), denoting the averaged value of the Kullback-Leibler information number in the distributed algorithm, and $\zeta(\rho, \mathcal{D}_\gamma^i)$ and $\xi(\rho, \mathcal{D}_\gamma^i)$ are functions of parameters ρ and $\mathcal{D}_\gamma^i$.

Note that the above results for the distributed algorithm is similar to Theorem 5 in [66], which is related to the performance of the isolated algorithm. The difference is that the averaged partial sum of the Kullback-Leibler numbers is involved in the distributed algorithm, due to the observation accumulation at each node. In the

sequel, we provide the proof flow for these two results.

First, we verify (6.33). By recalling p_n defined in (6.5) and $\Lambda_n = p_n/(1-p_n)$, we have

$$\begin{aligned}\mathrm{PFA}(\tau_D^i) &= \mathbb{E}^\pi(1-p_{\tau_D^i}) = \mathbb{E}^\pi(1+\Lambda_{\tau_D^i})^{-1} \\ &= \mathbb{E}^\pi\left(1+A\frac{\Lambda_{\tau_D^i}}{A}\right)^{-1} \\ &= \frac{1}{A}\mathbb{E}^\pi\left(e^{-\omega_a}\right)(1+o(1)),\ A\to\infty, \end{aligned} \tag{6.35}$$

where $\omega_a = \log\Lambda_{\tau_D^i} - a$ and $a = \log(A/\rho)$. For $\mathbb{E}^\pi\left(e^{-\omega_a}\right)$, we have

$$\begin{aligned}\mathbb{E}^\pi\left(e^{-\omega_a}\right) &= \mathbb{E}^\pi\left(e^{-\omega_a}|\tau_D^i \geq \lambda\right)(1-\mathrm{PFA}(\tau_D^i)) \\ &\quad + \mathbb{E}^\pi\left(e^{-\omega_a}|\tau_D^i < \lambda\right)\mathrm{PFA}(\tau_D^i) \\ &= \mathbb{E}^\pi\left(e^{-\omega_a}|\tau_D^i \geq \lambda\right) + O(A^{-1}),\ A\to\infty, \end{aligned} \tag{6.36}$$

which is due to $\mathrm{PFA}(\tau_D^i) \leq 1/(1+A) < 1/A$.

Thus, we turn to study $\mathbb{E}^\pi\left(e^{-\omega_a}|\tau_D^i \geq \lambda\right)$ as

$$\begin{aligned}&\mathbb{E}^\pi\left(e^{-\omega_a}|\tau_D^i \geq \lambda\right) \\ &= \sum_{k=1}^{\infty}\mathbb{E}_k\left(e^{-\omega_a}|\tau_D^i \geq k\right)\mathbb{P}(\lambda = k|\tau_D^i \geq k). \end{aligned} \tag{6.37}$$

For any $1 \leq k < \infty$, we have

$$\tau_D^i = \inf\left\{n \geq 1 : W_{n,k}(\rho) + l_{n,k} \geq a\right\}, \tag{6.38}$$

where $W_{n,k}(\rho) = Z_{n,k} + (n-k+1)|\log(1-\rho)|$, $n \geq k$, is a random walk with $\mathbb{E}_k\left[W_{n,k}(\rho)\right] = (n-k+1)(\mathcal{D}_\gamma^i + |\log(1-\rho)|)$ and $l_{n,k}$ is a nonlinear term. In $W_{n,k}(\rho)$,

we have

$$Z_{n,k} = \sum_{t=k}^{n} \sum_{j \in O_t^i} \log \frac{f_0^j(X_t^j)}{f_1^j(X_t^j)}. \tag{6.39}$$

Then, by applying Theorem 4.1 in [72], we obtain

$$\lim_{A \to \infty} \mathbb{E}_k \left(e^{-\omega_a} | \tau_D^i \geq k \right) = \zeta(\rho, \mathcal{D}_\gamma^i), \tag{6.40}$$

where $\zeta(\rho, \mathcal{D}_\gamma^i)$ is a function of parameters ρ and $\mathcal{D}_\gamma^i$.

We also have

$$\lim_{A \to \infty} \mathbb{P}(\lambda = k | \tau_D^i \geq k) = \lim_{A \to \infty} \frac{\pi_k \mathbb{P}_k(\tau_D^i \geq k | \lambda = k)}{\mathbb{P}^\pi(\tau_D^i \geq k)} = \pi_k. \tag{6.41}$$

Therefore, by plugging (6.40) and (6.41) into (6.37), we have

$$\lim_{A \to \infty} \mathbb{E}^\pi \left(e^{-\omega_a} | \tau_D^i \geq \lambda \right) = \zeta(\rho, \mathcal{D}_\gamma^i). \tag{6.42}$$

Finally, by combining (6.35), (6.36), and (6.42), we prove (6.33).

The proof of (6.34) depends on Theorem 4.5 in [72]. In order to use this theorem, the validity of the following three conditions needs to be checked:

$$\sum_{n=1}^{\infty} \mathbb{P}_1\{l_n \leq -\theta n\} < \infty, \text{ for some } 0 < \theta < \mathcal{D}_D^i;$$

$$\max_{0 \leq k \leq n} |l_{n+k}|, \ n \geq 1, \text{ are } \mathbb{P}_1 \text{ uniformly integrable;}$$

$$\lim_{A \to \infty} a \mathbb{P}_1\{\tau_D^i(A) \leq \varepsilon a (\mathcal{D}_D^i + |\log(1-\rho)|)^{-1}\} = 0,$$

$$\text{for some } 0 < \varepsilon < 1, \text{ where } a = \log(A/\rho),$$

where l_n is defined in (6.29).

It is easy to check that the first condition is valid, as $l_n \geq 0$. For the second condition, we have $\max_{0\leq k\leq n} |l_{n+k}| = l_{2n}$, since l_n, $n = 1, 2, \cdots$, are non-decreasing. Thus, to check that the second condition is valid, we only need to show that l_n, $n = 1, 2, \cdots$, are $\mathbb{P}_1$ uniformly integrable. To this end, we have that l_n converges almost surely, as $n \to \infty$, to the following random variable

$$l = \log \left\{ 1 + \sum_{k=1}^{\infty} (1-\rho)^k \prod_{s=1}^{k} \prod_{j\in O_s^i} \frac{f_0^j(X_s^j)}{f_1^j(X_s^j)} \right\}. \tag{6.43}$$

By taking the expectation, we have

$$\mathbb{E}_1(l) \leq \log \left\{ 1 + \sum_{k=1}^{\infty} (1-\rho)^k \right\} = \log \frac{1}{\rho}. \tag{6.44}$$

Since l_n, $n = 1, 2, \cdots$, are non-decreasing, we have $l_n \leq l$. Then, we have $\mathbb{E}_1(l_n) < \infty$, implying the uniform integrability. Therefore, the second condition is satisfied.

Now we intend to show the validity of the third condition. According to Lemma 1 in [66], we have

$$\mathbb{P}_1 \left\{ \tau_D^i(A) \leq 1 + (1-\epsilon)L_a \right\} \leq e^{-\phi_\epsilon a} + \beta(\epsilon, A), \tag{6.45}$$

where $L_a = a(\mathcal{D}_D^i + |\log(1-\rho)|)^{-1}$, $\phi_\epsilon > 0$ for all $0 < \epsilon < 1$, and $\beta(\epsilon, A) = \mathbb{P}_1 \left\{ \max_{1\leq n<K_{\epsilon,A}} Z_n \geq (1+\epsilon)\mathcal{D}_D^i K_{\epsilon,A} \right\}$, in which $K_{\epsilon,A} = (1-\epsilon)L_a$ and Z_n is defined in (6.28). The term $e^{-\phi_\epsilon a}$ on the right-hand side is $o(1/a)$. Thus, in order to show

$$\lim_{A\to\infty} a\mathbb{P}_1 \left\{ \tau_D^i(A) \leq 1 + (1-\epsilon)L_a \right\} = 0, \tag{6.46}$$

we only need to prove that the other term $\beta(\epsilon, A)$ is also $o(1/a)$, since $a = \log(A/\rho)$. To this end, by applying Theorem 1 of [11], for $\nu > 0$ and $r \geq 0$, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbb{P}_1 \left\{ \max_{1 \leq k \leq n} (Z_k - \mathcal{D}_D^i k) \geq \nu n \right\} \\
&\leq C_r \left\{ \mathbb{E}_1 [(Z_1 - \mathcal{D}_D^i)^+]^{r+1} + [\mathbb{E}_1 (Z_1 - \mathcal{D}_D^i)^2]^r \right\}, \qquad (6.47)
\end{aligned}$$

where C_r is a constant. When $r = 1$, the finiteness of the right-hand side of the above inequality implies that the left-hand side is also finite. Thus, we obtain $\mathbb{P}_1 \{\max_{1 \leq k \leq n} (Z_k - \mathcal{D}_D^i k) \geq \nu n\} = o(1/n)$.

Then, with the fact that

$$\beta(\epsilon, A) \leq \mathbb{P}_1 \left\{ \max_{1 \leq n < K_{\epsilon, A}} (Z_n - \mathcal{D}_D^i n) \geq \epsilon \mathcal{D}_D^i K_{\epsilon, A} \right\}, \qquad (6.48)$$

we have $\beta(\epsilon, A) = o(1/a)$. Therefore,

$$\lim_{A \to \infty} a \mathbb{P}_1 \left\{ \tau_D^i(A) \leq 1 + (1 - \epsilon) L_a \right\} = 0. \qquad (6.49)$$

By taking $\varepsilon = 1 - \epsilon$, finally we have

$$\begin{aligned}
&\lim_{A \to \infty} a \mathbb{P}_1 \{ \tau_D^i(A) \leq \varepsilon L_a \} \\
&\leq \lim_{A \to \infty} a \mathbb{P}_1 \left\{ \tau_D^i(A) \leq 1 + (1 - \epsilon) L_a \right\} \\
&= 0 \qquad (6.50)
\end{aligned}$$

Hence the third condition is satisfied. Therefore, the conditions of Theorem 4.5 in [72] are satisfied. This theorem shows that (6.34) is valid.

Then, with (6.33) and (6.34), by taking the same proof method of Theorem 6.2.1,

we have

$$\lim_{\mathrm{CADD}_1(\tau_D^i)\to\infty} \frac{1}{\mathrm{CADD}_1(\tau_D^i)} \log[\mathrm{PFA}(\tau_D^i)]$$
$$= -(\mathcal{D}_\gamma^i + |\log(1-\rho)|). \tag{6.51}$$

6.5 Second-layer Large Deviation Analysis

Since $\mathcal{D}_\gamma^i$ has been shown as a crucial factor in the large deviation analysis of last subsection, in this subsection, we focus on studying the behavior of $\mathcal{D}_\gamma^i$. As we still stay in the scope of large deviation analysis as we did in the last subsection, we call it as the second-layer large deviation analysis, where the analysis in the last subsection is called the first-layer large deviation analysis.

As we cannot obtain the closed-form for $\mathcal{D}_\gamma^i$ due to the complicated probabilities incorporated, we discuss its asymptotic behavior when $\gamma \to \infty$. To this end, we first study the behavior of $\vec{q}_\gamma^i(\nu)$, defined below (6.30), when $\gamma \to \infty$, by employing the concept of hitting times in Markov chains.

For each $\nu \neq 2^N - 1$, without loss of generality, we assume that ν corresponds to the index of the sensor subset $\{i_1, i_2, \cdots, i_m\}$, with $\{i'_1, i'_2, \cdots, i'_{N-m}\}$ as the complementary subset, where $m \geq 1$ due to the fact that at least its own observation is available at each sensor. Let T_j denote the hitting time, starting from state (index of sensor) j to hit another specific state i in the Markov chain, whose transition probability matrix is $\bar{\mathbf{A}}$ defined in (6.20). From Theorem 7.26 in [19], since the transition probability matrix $\bar{\mathbf{A}}$ is irreducible, there exists constants $0 < \alpha < 1$ and $0 < L < \infty$ such that $P(T_j > L) \leq \alpha, \forall j$, and more generally,

$$P(T_j > kL) \leq \alpha^k, \ k = 0, 1, 2, \cdots. \tag{6.52}$$

Also, there exists a constant $0 < \beta < 1$ such that $P(T_j > L) \geq \beta, \forall j$, and more generally,

$$P(T_j > kL) \geq \beta^k, \ k = 0, 1, 2, \cdots . \tag{6.53}$$

Based on the above results of hitting times in Markov chains, we first present the following large deviation related theorem on the asymptotic behavior of $\sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v)$, as $\gamma \to \infty$. Since $\nu \in \{0, 1, \cdots, 2^N - 1\}$ according to (6.31), we have $\sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v) = 1 - \bar{q}_\gamma^i(2^N - 1)$, where $\bar{q}_\gamma^i(2^N - 1)$ denotes the probability that the observations from all sensors are available at sensor i, i.e., $\sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v)$ is the probability of the event that not all observations are available at sensor i.

Theorem 6.5.1 As $\gamma \to \infty$, the probability $\sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v)$ has the large deviation upper and lower bounds as follows,

$$\frac{\ln \beta}{L} \leq \lim_{\gamma \to \infty} \frac{1}{\gamma} \ln \sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v) \leq \frac{\ln \alpha}{L}, \tag{6.54}$$

where α, β and L are parameters in (6.52) and (6.53).

Since $\sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v)$ presents the probability of the event that not all observations are available at sensor i, Theorem 6.5.1 implies that this event is a rare event and its probability decays exponentially fast to zero as $\gamma \to \infty$.

The proof is presented as follows.

Proof 16 Recall that ν corresponds to the index of the sensor subset $\{i_1, i_2, \cdots, i_m\}$, with $\{i'_1, i'_2, \cdots, i'_{N-m}\}$ as the complementary subset, and T_j denotes the hitting time, starting from state (index of sensor) j to hit another specific state i in the Markov

chain. Then, the probability $\bar{q}_\gamma^i(\nu)$ could be represented as

$$\begin{aligned}\bar{q}_\gamma^i(\nu) &= \Pr(T_{i'_1} > \gamma, \cdots, T_{i'_{N-m}} > \gamma, T_{i_1} \leq \gamma, \cdots, T_{i_m} \leq \gamma) \\ &\leq \Pr(T_{i'_1} > \gamma, \cdots, T_{i'_{N-m}} > \gamma) \\ &\leq \min_{1 \leq n \leq N-m} \Pr(T_{i'_n} > \gamma). \end{aligned} \tag{6.55}$$

Thus, we have

$$\begin{aligned}\lim_{\gamma\to\infty} \frac{1}{\gamma} \ln\left(\bar{q}_\gamma^i(\nu)\right) &\leq \lim_{\gamma\to\infty} \frac{1}{\gamma} \ln\left(\min_{1 \leq n \leq N-m} P(T_{i'_n} > \gamma)\right) \\ &\leq \lim_{\gamma\to\infty} \frac{1}{\gamma} \ln\left(\alpha^{\lfloor \gamma/L \rfloor}\right) = \frac{\ln\, \alpha}{L} \end{aligned} \tag{6.56}$$

where the second inequality is due to (6.52).

For $\bar{q}_\gamma^i(\nu)$, we also have

$$\begin{aligned}\bar{q}_\gamma^i(\nu) &= \Pr(T_{i'_1} > \gamma, \cdots, T_{i'_{N-m}} > \gamma, T_{i_1} \leq \gamma, \cdots, T_{i_m} \leq \gamma) \\ &\geq \Pr(T_{i'_1} > \gamma) \cdots \Pr(T_{i'_{N-m}} > \gamma) \\ &\quad \Pr(T_{i_1} \leq \gamma) \cdots \Pr(T_{i_m} \leq \gamma). \end{aligned} \tag{6.57}$$

This leads to

$$\begin{aligned}&\lim_{\gamma\to\infty} \frac{1}{\gamma} \ln\left(\bar{q}_\gamma^i(\nu)\right) \\ &\quad \geq \lim_{\gamma\to\infty} \frac{1}{\gamma} \ln\left[\left(\beta^{\lceil \gamma/L \rceil}\right)^{N-m} \left(1 - \alpha^{\lfloor \gamma/L \rfloor}\right)^m\right] \\ &\quad = (N-m)\frac{\ln \beta}{L} \end{aligned} \tag{6.58}$$

where the first inequality is due to (6.52) and (6.53), and the last equality is derived with $0 < \alpha < 1$.

By combining (6.56) and (6.58), we have

$$(N-m)\frac{\ln\beta}{L} \leq \lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\left(\bar{q}_\gamma^i(\nu)\right) \leq \frac{\ln\,\alpha}{L}. \tag{6.59}$$

Then, we obtain

$$\begin{aligned}
&\lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\sum_{v=0}^{2^N-2}\bar{q}_\gamma^i(v)\\
&\leq \lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\left[(2^N-1)\max_v(\bar{q}_\gamma^i(v))\right]\\
&= \lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\left[\max_v(\bar{q}_\gamma^i(v))\right]\\
&\leq \frac{\ln\,\alpha}{L},
\end{aligned} \tag{6.60}$$

where the last inequality is due to (6.59).

We also have

$$\begin{aligned}
\lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\sum_{v=0}^{2^N-2}\bar{q}_\gamma^i(v) &\overset{a}{\geq} \lim_{\gamma\to\infty}\frac{1}{\gamma}\ln(\bar{q}_\gamma^i(v_p))\\
&\overset{b}{=} \frac{\ln\beta}{L},
\end{aligned} \tag{6.61}$$

where v_p on the right-hand side of inequality a denotes a particular index of the subset of sensors such that $m = N-1$, i.e., v_p is the index of the sensor subset $\{i_1, i_2, \cdots, i_{N-1}\}$, recalling the notations defined at the beginning of this section. Since for $v_p \in \{0, 2^N-2\}$, we have $\sum_{v=0}^{2^N-2}\bar{q}_\gamma^i(v) \geq \bar{q}_\gamma^i(v_p)$, implying the establishment of inequality a. According to (6.59) and taking $m = N-1$, we derive the equation b in (6.61).

By combining (6.60) and (6.61), we conclude that

$$\frac{\ln \beta}{L} \leq \lim_{\gamma \to \infty} \frac{1}{\gamma} \ln \sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v) \leq \frac{\ln \alpha}{L} \tag{6.62}$$

Based on Theorem 6.5.1, we further have the following theorem regarding the behavior of the distributed Kullback-Leibler information number $\mathcal{D}_\gamma^i$ defined in Theorem 6.4.1.

Theorem 6.5.2 As $\gamma \to \infty$, we have the following upper and lower bounds for the value of $\mathcal{D}_\gamma^i$,

$$\begin{aligned} \mathcal{D} - \left[\max_{j \in \{1,\cdots,N\} \backslash i} D(f_1^j, f_0^j) \right] e^{\frac{\ln \alpha}{L} \gamma} &\leq \mathcal{D}_\gamma^i \\ &\leq \mathcal{D} - \left[\min_{j \in \{1,\cdots,N\} \backslash i} D(f_1^j, f_0^j) \right] e^{\frac{\ln \beta}{L} \gamma}, \end{aligned} \tag{6.63}$$

where $D(f_1^j, f_0^j)$ is the Kullback-Leibler information number defined in (6.10) and $\mathcal{D}$ is the centralized Kullback-Leibler information number defined in Theorem 6.2.1, and $\ln \alpha / L$ and $\ln \beta / L$ are the upper and lower bounds derived in Theorem 6.5.1.

Theorem 6.5.2 implies that $\mathcal{D}_\gamma^i$ converges to $\mathcal{D}$ exponentially fast, as $\gamma \to \infty$. Since $\mathcal{D}_\gamma^i$ and $\mathcal{D}$ determine the performance of the distributed and centralized algorithms respectively, this theorem also implies that the performance of the proposed distributed algorithm converges to that of the centralized one at an exponentially fast rate.

Proof 17 Recall $\mathcal{D}_\gamma^i = \sum_{\nu=1}^{2^N-1} \bar{q}_\gamma^i(\nu) \sum_{j \in \Psi_\nu} D(f_1^j, f_0^j)$ and $\mathcal{D} = \sum_{i=1}^N D(f_1^i, f_0^i)$. We

have

$$\begin{aligned}\mathcal{D}_\gamma^i &\stackrel{a}{=} \bar{q}_\gamma^i(2^N-1)\mathcal{D} + \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \sum_{j\in\Psi_\nu} D(f_1^j, f_0^j) \\ &\stackrel{b}{=} \left(1 - \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu)\right)\mathcal{D} + \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \sum_{j\in\Psi_\nu} D(f_1^j, f_0^j), \end{aligned} \tag{6.64}$$

where equation a is due to the fact that $\Psi_\nu = \{1, \cdots, N\}$ with $\nu = 2^N - 1$, i.e., Ψ_{2^N-1} denotes the set of indices of all sensors, and equation b is based on $\sum_{\nu=1}^{2^N-1} \bar{q}_\gamma^i(\nu) = 1$.

Then, from (6.64), we have

$$\begin{aligned}\mathcal{D}_\gamma^i &\leq \left(1 - \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu)\right)\mathcal{D} \\ &\quad + \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \max_{1\leq\nu\leq 2^N-2} \sum_{j\in\Psi_\nu} D(f_1^j, f_0^j) \\ &= \mathcal{D} - \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \min_{j\in\{1,\cdots,N\}\setminus i} D(f_1^j, f_0^j). \end{aligned} \tag{6.65}$$

We could also obtain

$$\begin{aligned}\mathcal{D}_\gamma^i &\geq \left(1 - \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu)\right)\mathcal{D} \\ &\quad + \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \min_{1\leq\nu\leq 2^N-2} \sum_{j\in\Psi_\nu} D(f_1^j, f_0^j) \\ &= \mathcal{D} - \sum_{\nu=1}^{2^N-2} \bar{q}_\gamma^i(\nu) \max_{j\in\{1,\cdots,N\}\setminus i} D(f_1^j, f_0^j). \end{aligned} \tag{6.66}$$

According to Theorem 6.5.1, as $\gamma \to \infty$, we have

$$e^{\frac{\ln \beta}{L}\gamma} \leq \sum_{v=0}^{2^N-2} \bar{q}_\gamma^i(v) \leq e^{\frac{\ln \alpha}{L}\gamma}. \tag{6.67}$$

Then, by combining (6.65), (6.66) and (6.67), as $\gamma \to \infty$, we derive

$$\begin{aligned} \mathcal{D} - \left[\max_{j\in\{1,\cdots,N\}\setminus i} D(f_1^j, f_0^j)\right] e^{\frac{\ln \alpha}{L}\gamma} &\leq \mathcal{D}_\gamma^i \\ &\leq \mathcal{D} - \left[\min_{j\in\{1,\cdots,N\}\setminus i} D(f_1^j, f_0^j)\right] e^{\frac{\ln \beta}{L}\gamma} \end{aligned} \tag{6.68}$$

6.6 Simulation Results

In this section, we simulate the proposed distributed algorithm with a network of 5 nodes taking observations. We consider a Bayesian setup, and set the prior distribution of the change-point time as a geometric distribution with parameter $\rho = 0.1$. Before the change happens, we consider that the observation at each node follows a Gaussian distribution $N(0, 1)$; after the change happens, the observation at node $i,\ i = 1, \cdots, 5$, turns to follow another Gaussian distribution $N(0.1 \times i, 1)$. Note that here we consider a setup that observations at different nodes have different post-change distributions, which is to mimic the more general situation that different nodes could suffer different levels of impact from the same physical change. For example, certain physical event, such as the leakage of chemical gas or the abrupt increasing of temperature, would lead to different degrees of impacts in different nodes, due to their various locations. The nodes near the origin of the physical event could suffer from a more serious influence, which is reflected by a larger mean in the post-distribution; the nodes faraway the origin could suffer from a less serious influence, which is reflected by a smaller mean in the post-distribution.

In Fig. 6.1, we show the simulated and analytical results corresponding to the first-layer large deviation analysis, and also compare the performance of the distributed scheme versus the centralized and isolated ones. In the simulation, we set γ as 6, recalling that γ is the mean value for number of communication rounds within each sampling time period. In Fig. 6.1, the dashed curves denote the simulated decay rate, and the solid lines present the analytical decay rates in Theorem 6.2.1 for the centralized scheme, Corollary 6.2.2 for the isolated scheme, and Theorem 6.4.1 for the distributed scheme, respectively. A higher decay rate implies a lower PFA under the same conditional ADD, which means that the performance is better. Therefore, from the simulation results of the decay rates, we see that the the distributed scheme outperforms the isolated one, but performs worse than the centralized one, which conforms to the analytical result from Theorem 6.4.1.

In Fig. 6.2, we show the simulation results of the value $\frac{1}{\gamma}\ln\sum_{v=0}^{2^N-2}\bar{q}_\gamma^i(v)$, denoting the decay rate of the rare event that not all observations are available at sensor i, as the parameter γ increases, which is the second-layer large deviation analysis shown in Theorem 6.5.1. We also present the large deviation lower and upper bounds in Fig. 6.2, from which we see that the simulated decay rate locates between the large deviation lower and upper bounds, and the bounds are relatively tight, which verifies the analytical result in Theorem 6.5.1. Here we also present the lower bound $\ln\beta/L$ and the upper bound $\ln\alpha/L$ of Theorem 6.5.1, which are shown in Fig. 6.2. Recall that T_j denotes the hitting time, starting from state (index of sensor) j to hit another specific state i in the Markov chain with the transition probability matrix $\bar{\mathbf{A}}$. Then we have

$$P(T_j > L) = \sum_{i_1,\cdots,i_L\neq i} \bar{\mathbf{A}}_{ji_1}\bar{\mathbf{A}}_{i_1i_2}\bar{\mathbf{A}}_{i_2i_3}\cdots\bar{\mathbf{A}}_{i_{L-1}i_L}. \tag{6.69}$$

Recall that we intend to find α such that $P(T_j > L) \leq \alpha,\ \forall j$. Thus, we can set

$\alpha = \max_j P(T_j > L)$. In order to find β such that $P(T_j > L) \geq \beta$, $\forall j$, we can set $\beta = \min_j P(T_j > L)$. Then, we are ready to calculate $\ln \alpha/L$ and $\ln \beta/L$. To this end, the selection of L is a critical step, as both α and β are calculated based on the selection of L. Here we show the calculation of $\ln \alpha/L$ and $\ln \beta/L$ with different L values in Fig. 6.3. From Fig. 6.3, we see a very interesting phenomenon that these two bounds look converging as L increases, although here we will not provide the mathematical proof of this result. This observation could imply some potential properties for hitting time in Markov chains. The further exploration with analytical analysis based on this observation will be left for our future work. Note that the upper and lower bounds in Fig. 6.2 are set as the values calculated with $L = 15$.

In Fig. 6.4, we show the simulation results for the distributed Kullback-Leibler information $\mathcal{D}_\gamma^i$, the value of the centralized Kullback-Leibler information $\mathcal{D}$, and the calculation results for the upper and lower bounds presented in Theorem 6.5.2. From Fig. 6.4, we see that the upper bound is a very tight bound, while the lower bound is relatively looser. However, the range of y-axis in this figure is very small from 0.3765 to 0.3810; so both the lower and upper bounds are tight bounds in this sense. We also see that the distributed Kullback-Leibler information $\mathcal{D}_\gamma^i$ converges to the centralized Kullback-Leibler information $\mathcal{D}$, as γ increases, which implies that the performance of the distributed change detection scheme converges to that of the centralized one, since $\mathcal{D}_\gamma^i$ and $\mathcal{D}$ determine the performance of the distributed and centralized schemes, respectively.

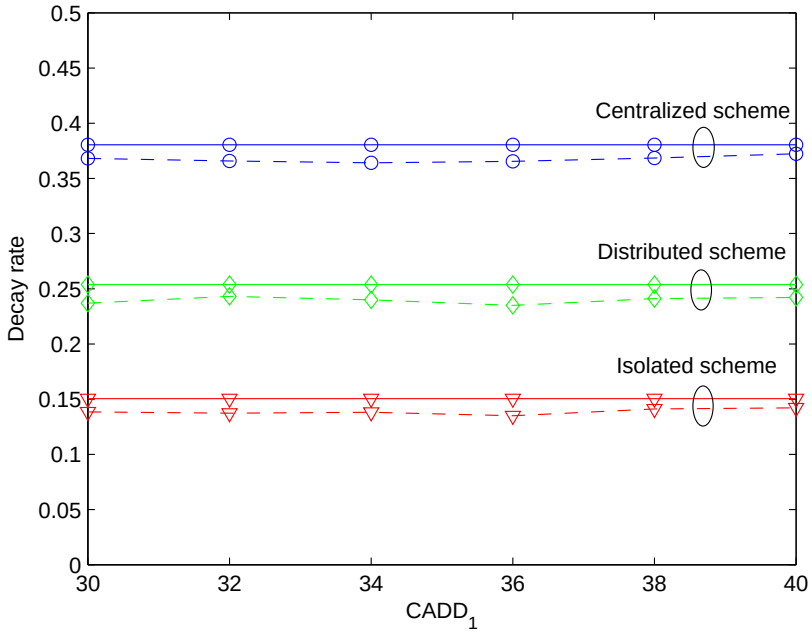

Figure 6.1: First-layer large deviation analysis: comparison of decay rates in distributed, centralized, and isolated schemes with simulation (dash curve) vs. analytical results (solid line).

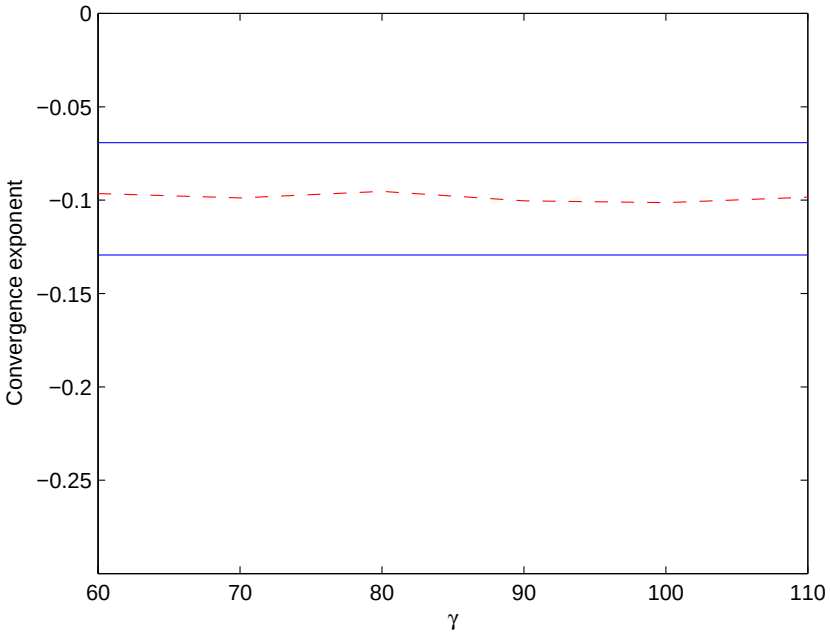

Figure 6.2: Second-layer large deviation analysis in Theorem 6.5.1: simulated decay rate (dash curve) of the probability of the rare event that not all observations are available at a sensor, and the corresponding large deviation upper and lower bounds (solid lines)

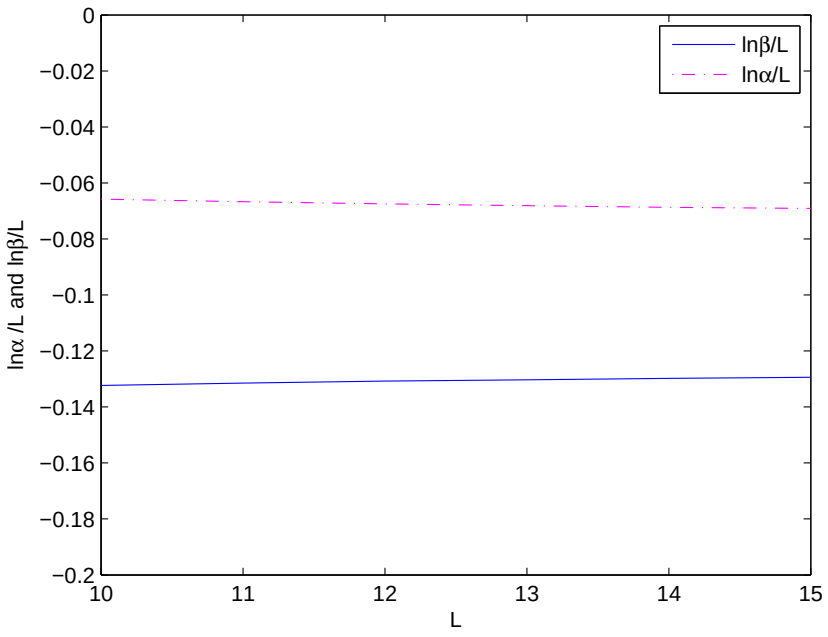

Figure 6.3: Calculation of the lower bound $\ln \beta/L$ and the upper bound $\ln \alpha/L$ in Theorem 6.5.1 with varying L.

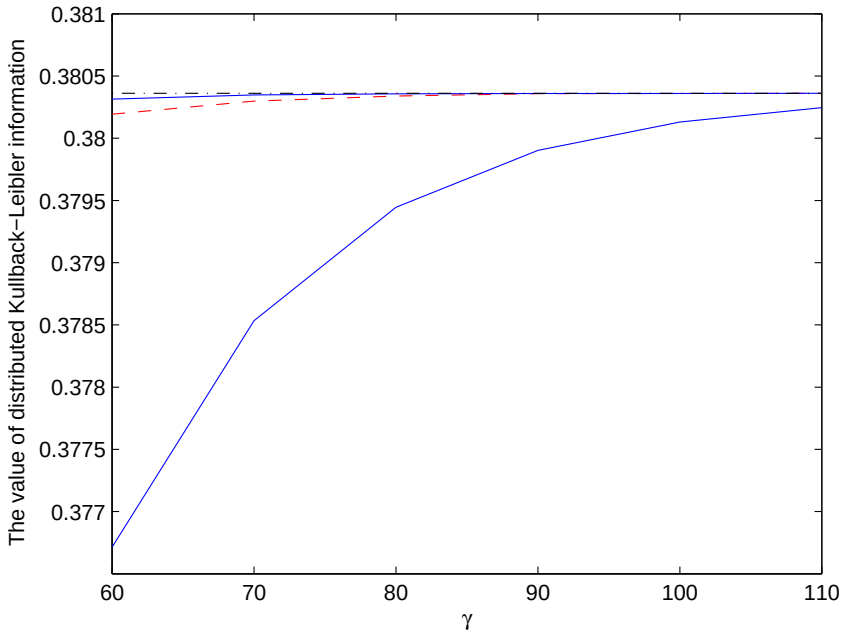

Figure 6.4: Simulated distributed Kullback-Leibler information (dash curve), centralized Kullback-Leibler information (dash-dot line) and the corresponding analytical upper and lower bounds (solid curve) in Theorem 6.5.2.

7. CONCLUSIONS

This thesis has studied the distributed signal processing methods in large scale and complex systems, by proposing a distributed Kalman filtering scheme to estimate the high-dimension states and a distributed quickest change detection scheme to detect the change happening. A random gossip based scheme called Modified Gossip Interactive Kalman filtering (M-GIKF) has been first studied, where the inter-sensor communications among neighbor sensors are used to exchange information (estimates and error covariances) among sensors. It is shown that the conditional estimation error covariance sequence at each sensor evolves as a random Riccati equation (RRE) with Markov modulated switching. With the idea of the random dynamic system (RDS), it is proven that the network achieves weak consensus, i.e., the conditional estimation error covariance at a randomly selected sensor converges weakly (in distribution) to a unique invariant measure. Further, it is proved that as the number of communications goes to infinity, this invariant measure has the Large Deviation (LD) upper and lower bounds, implying that this measure converges exponentially fast (in probability) to the Dirac measure δ_{P^*}, where P^* is the stable error covariance in the centralized Kalman filtering setup. The LD results have answered a fundamental question on how to quantify the rate at which the distributed scheme approaches the centralized performance as the inter-sensor communication rate increases. Moreover, the quantization method used before communications in the proposed distributed scheme is also investigated, which is called the dithered quantization. Although the network consensus is sensitive to the information loss caused by the quantization, it is shown that the network can still achieve weak consensus with the dithered quantization in the case of countable infinite quantization alphabet. Then, a Bayesian

quickest change detection scheme is proposed, where multiple communication rounds happen during the observation sampling interval to propagate the observations. The two-layer large deviation analysis is used to analyze the performance of the distributed scheme. The first layer analysis proves that the probability of false alarm decays to zero exponentially fast with the increasing of the averaged detection delay. The second-layer analysis shows that the probability of the rare event that not all observations are available at a sensor decays to zero at an exponentially fast rate when the number of communications increases, based on which it is shown that the performance of the distributed algorithm converges exponentially fast to that of the centralized one.

REFERENCES

[1] L. Arnold. *Random Dynamical Systems*. New York: Springer, 1998.

[2] T. C. Aysal, M. J. Coates, and M. G. Rabbat. Distributed average consensus with dithered quantization. *IEEE Transactions on Signal Processing*, 56(10):4905–4918, Oct. 2008.

[3] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura. Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis. *IEEE Transactions on Signal Processing*, 59(9):4381–4396, Sep. 2011.

[4] T. Banerjee, V. Sharma, V. Kavitha, and A. K. JayaPrakasam. Generalized analysis of a distributed energy efficient algorithm for change detection. *IEEE Transactions on Wireless Communications*, 10(1):91–101, Jan. 2011.

[5] T. Banerjee and V. V. Veeravalli. Energy-efficient quickest change detection in sensor networks. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 636–639, Ann Arbor, USA, Aug. 2012.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, Jun. 2006.

[7] P. Braca, S. Marano, V. Matta, and P. Willett. Consensus-based Page's test in sensor networks. *Signal Processing*, 91:919–930, Apr. 2011.

[8] J. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley, New York, 1990.

[9] A. A. Cardenas, S. Radosavac, and J. S. Baras. Evaluation of detection algorithms for mac layer misbehavior: Theory and experiments. *IEEE/ACM Transactions on Networking*, 17(2):605–617, Apr. 2009.

[10] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering using consensus strategies. *IEEE Journal on Selected Areas of Communications*, 26(4):622–633, Sep. 2008.

[11] Y. S. Chow. Some convergence theorems for independent random variables. *Ann. Math. Statist.*, 37:1482–1493, 1966.

[12] D. Commenges, J. Seal, and F. Pinatel. Inference about a change point in experimental neurophysiology. *Mathematical Biosciences*, 80(1):81–108, July 1986.

[13] S. Das and J. M. F. Moura. Distributed Kalman filtering and network tracking capacity. In *47th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2013, (Invited paper).

[14] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 1998.

[15] J. D. Deuschel and D. W. Stroock. *Large Deviations*. Academic Press, Inc., London, 1984.

[16] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Review*, 41(1):45–76, 1999.

[17] 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*, 98(11):1847–1864, Nov. 2010.

[18] 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*, 7(2):195–204, Apr. 2013.

[19] B. K. Driver. Introduction to stochastic processes II. [Online]. Available: http://www.math.ucsd.edu/~bdriver/math180C_S2011/Lecture%20Notes/180Lec6b.pdf.

[20] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri. Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19(16):1787–1816, Nov. 2009.

[21] M. Frisen. Optimal sequential surveillance for finance, public health, and other areas. *Sequential Analysis*, 28(3):310–337, July 2009.

[22] R. M. Gray and T. G. Stockham. Dithered quantizers. *IEEE Transactions on Information Theory*, 39(3):805–812, May 1993.

[23] O. Hadjiliadis, H. Zhang, and H. V. Poor. One shot schemes for decentralized quickest change detection. *IEEE Transactions on Information Theory*, 55(7):3346–3359, July 2009.

[24] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin Heidelberg, 1987.

[25] D. Jakovetić, J. M. F. Moura, and J. Xavier. Distributed detection over noisy networks: Large deviations analysis. *IEEE Transactions on Signal Processing*, 60(8):4306–4320, Aug. 2012.

[26] O. Kallenberg. *Foundations of Modern Probability*. Springer Series in Statistics, New York, 2002.

[27] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering on Automatic Control*, 82(D):35–45, Mar. 1960.

[28] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise. *IEEE Transactions on Signal Processing*, 57(1):355–369, Jan. 2009.

[29] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: Quantized data and random link failures. *IEEE Transactions on Signal Processing*, 58(3):1383–1400, Mar. 2010.

[30] S. Kar and J. M. F. Moura. Gossip and distributed Kalman filtering: Weak consensus under weak detectability. *IEEE Transactions on Signal Processing*, 59(4):1766–1784, Apr. 2011.

[31] S. Kar and J. M. F. Moura. Moderate deviations of a random Riccati equation. *IEEE Transactions on Automatic Control*, 57(9):2250–2265, Sep. 2012.

[32] S. Kar, B. Sinopoli, and J. M. F. Moura. Kalman filtering with intermittent observations: Weak convergence to a stationary distribution. *IEEE Transactions on Automatic Control*, 57(2):405–420, Feb. 2012.

[33] 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 (CDC)*, pages 6639–6644, Atlanta, GA, Dec. 2010.

[34] U. A. Khan and J. M. F. Moura. Distributing the Kalman filter for large-scale systems. *IEEE Transactions on Signal Processing*, 56(10):4919–4935, Oct. 2008.

[35] L. Lai, Y. Fan, and H. V. Poor. Quickest detection in cognitive radio: A sequential change detection framework. In *2008 IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–5, New Orleans, USA, Nov. 2008.

[36] D. Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui. Distributed Kalman filtering with quantized sensing state. *IEEE Transactions on Signal Processing*,

63(19):5180–5193, Oct. 2015.

[37] 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*, 61(3):1351–1372, Mar. 2015.

[38] D. Li, L. Lai, and S. Cui. Quickest change detection and identification across a sensor array. In *2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 145–148, Austin, USA, Dec. 2013.

[39] H. Li, H. Dai, and C. Li. Collaborative quickest spectrum sensing via random broadcast in cognitive radio systems. *IEEE Transactions on Wireless Communications*, 9(7):2338–2348, July 2010.

[40] H. Li and J. Fang. Distributed adaptive quantization and estimation for wireless sensor networks. *IEEE Signal Processing Letters*, 14(10):669–672, Oct. 2007.

[41] J. Li and G. AlRegib. Distributed estimation in energy-constrained wireless sensor networks. *IEEE Transactions on Signal Processing*, 57(10):3746–3758, Oct. 2009.

[42] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson. Wireless sensor networks for habitat monitoring. In *1st ACM International Workshop on Wireless Sensor Networks and Applications*, pages 88–97, Atlanta, USA, Sep. 2002.

[43] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand. Achieving 100 % throughput in an input-queued switch. *IEEE Transactions on Communications*, 47(8):1260–1267, Aug. 1999.

[44] V. Megalooikonomou and Y. Yesha. Quantizer design for distributed estimation with communication constraints and unknown observation statistics. *IEEE Transactions on Communications*, 48(2):181–184, Feb. 2000.

[45] Y. Mei. Quickest detection in censoring sensor networks. In *2011 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 2148–2152, St. Petersburg, Russia, July 2011.

[46] G. V. Moustakides. Decentralized CUSUM change detection. In *9th International Conference on Information Fusion*, pages 1–6, Florence, Italy, July 2006.

[47] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis. Decentralized quantized Kalman filtering with scalable communication cost. *IEEE Transactions on Signal Processing*, 56(8):3727–3741, Aug. 2008.

[48] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In *44th IEEE Conference on Decision and Control (CDC) and 2005 European Control Conference (ECC)*, pages 8179–8184, Seville, Spain, Dec. 2005.

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

[50] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, Sep. 2004.

[51] S. Park and N. C. Martins. An augmented observer for the distributed estimation problem for LTI systems. In *2012 American Control Conference (ACC)*, pages 6775–6780, Montreal, QC, Jun. 2012.

[52] H. V. Poor and O. Hadjiliadis. *Quickest Detection*. Cambridge University Press, Cambridge, UK, 2008.

[53] B. Rao and H. Durrant-Whyte. Fully decentralized algorithm for multisensor Kalman filtering. *Inst. Electr. Eng.-Control Theory Appl.*, 138:413–420, Sep. 1991.

[54] J. A. Rice, K. Mechitov, S. Sim, T. Nagayama, S. Jang, R. Kim, B. F. Spencer, G. Agha, and Y. Fujino. Flexible smart sensor framework for autonomous structural health monitoring. *Smart Structures and Systems*, 6(5):423–438, May 2010.

[55] A. K. Sahu and S. Kar. Distributed sequential detection for gaussian binary hypothesis testing. *IEEE Transactions on Signal Processing*, to appear.

[56] 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*, 56(4):1650–1666, Apr. 2008.

[57] L. Schuchman. Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, 12(4):162–165, Dec. 1964.

[58] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, 8(1):22–46, 1963.

[59] A. N. Shiryaev. *Optimal Stopping Rules*. Springer-Verlag, New York, 1978.

[60] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions of Automatic Control*, 49(9):1453–1464, Sep. 2004.

[61] C. Sonesson and D. Bock. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society*, 166(1):5–21, July 2003.

[62] S. S. Stankovic, N. Ilic, M. S. Stankovic, and K. H. Johansson. Distributed change detection based on a consensus algorithm. *IEEE Transactions on Signal Processing*, 59(12):5686–5697, Dec. 2011.

[63] S. Sun, J. Lin, L. Xie, and W. Xiao. Quantized Kalman filtering. In *the IEEE 22nd International Symposium on Intelligent Control*, pages 7–12, Singapore, Oct. 2007.

[64] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, 54(9):3372–3382, Sep. 2006.

[65] A. G. Tartakovsky and V. V. Veeravalli. Quickest change detection in distributed sensor systems. In *6th IEEE International Conference on Information Fusion*, pages 756–763, Cairns, Australia, July 2003.

[66] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *Theory of Probability and Its Applications*, 49(3):458–497, 2005.

[67] A. G. Tartakovsky and V. V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis*, 27(4):441–475, Oct. 2008.

[68] M. Thottan and C. Ji. Anomaly detection in IP networks. *IEEE Transactions on Signal Processing*, 51(8):2191–2204, Aug. 2003.

[69] S. Trivedi and R. Chandramouli. Secret key estimation in sequential steganography. *IEEE Transactions on Signal Processing*, 53(2):746–757, Feb. 2005.

[70] V. V. Veeravalli. Decentralized quickest change detection. *IEEE Transactions on Information Theory*, 47(4):1657–1665, May 2001.

[71] P. Venkitasubramaniam, G. Mergen, Lang Tong, and A. Swami. Quantization for distributed estimation in large scale sensor networks. In *Third International Conference on Intelligent Sensing and Information Processing*, pages 121–127, Bangalore, Dec. 2005.

[72] M. Woodroofe. *Nonlinear Renewable Theory in Sequential Analysis*. SIAM, Philadelphia, PA, 1982.

[73] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith. Power scheduling of universal decentralized estimation in sensor networks. *IEEE Transactions on Signal Processing*, 54(2):413–422, Feb. 2006.

[74] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 52:65–78, 2004.

[75] L. Zacharias and R. Sundaresan. Decentralized sequential change detection using physical layer fusion. *IEEE Transactions on Wireless Communications*, 7(12):4999–5008, Dec. 2008.