

Syracuse University

SURFACE

Dissertations - ALL

SURFACE

May 2014

Distributed Estimation and Performance Limits in Resource-constrained Wireless Sensor Networks

YUJIAO ZHENG

Syracuse University

Follow this and additional works at: <https://surface.syr.edu/etd>



Part of the [Engineering Commons](#)

Recommended Citation

ZHENG, YUJIAO, "Distributed Estimation and Performance Limits in Resource-constrained Wireless Sensor Networks" (2014). *Dissertations - ALL*. 71.

<https://surface.syr.edu/etd/71>

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Abstract

Distributed inference arising in sensor networks has been an interesting and promising discipline in recent years. The goal of this dissertation is to investigate several issues related to distributed inference in sensor networks, emphasizing parameter estimation and target tracking with resource-constrained networks.

To reduce the transmissions between sensors and the fusion center thereby saving bandwidth and energy consumption in sensor networks, a novel methodology, where each local sensor performs a censoring procedure based on the normalized innovation square (NIS), is proposed for the sequential Bayesian estimation problem in this dissertation. In this methodology, each sensor sends only the informative measurements and the fusion center fuses both missing measurements and received ones to yield more accurate inference. The new methodology is derived for both linear and nonlinear dynamic systems, and both scalar and vector measurements. The relationship between the censoring rule based on NIS and the one based on Kullback-Leibler (KL) divergence is investigated.

A probabilistic transmission model over multiple access channels (MACs) is investigated. With this model, a relationship between the sensor management and compressive sensing problems is established, based on which, the sensor management problem becomes a constrained optimization problem, where the goal is to determine the optimal values of probabilities that each sensor should transmit with such that the determinant of the Fisher information matrix (FIM) at any given time step is maximized. The performance of the proposed compressive sensing based sensor management methodology in terms of accuracy of inference is investigated.

For the Bayesian parameter estimation problem, a framework is proposed where quantized observations from local sensors are not directly fused at the fusion center, instead, an additive noise is injected independently to each quantized observation. The injected noise performs

as a low-pass filter in the characteristic function (CF) domain, and therefore, is capable of recovering the original analog data if certain conditions are satisfied. The optimal estimator based on the new framework is derived, so is the performance bound in terms of Fisher information. Moreover, a sub-optimal estimator, namely, linear minimum mean square error estimator (LMMSE) is derived, due to the fact that the proposed framework theoretically justifies the additive noise modeling of the quantization process. The bit allocation problem based on the framework is also investigated.

A source localization problem in a large-scale sensor network is explored. The maximum-likelihood (ML) estimator based on the quantized data from local sensors and its performance bound in terms of Cramér-Rao lower bound (CRLB) are derived. Since the number of sensors is large, the law of large numbers (LLN) is utilized to obtain a closed-form version of the performance bound, which clearly shows the dependence of the bound on the sensor density, *i.e.*, the Fisher information is a linearly increasing function of the sensor density. Error incurred by the LLN approximation is also theoretically analyzed. Furthermore, the design of sub-optimal local sensor quantizers based on the closed-form solution is proposed.

The problem of on-line performance evaluation for state estimation of a moving target is studied. In particular, a compact and efficient recursive conditional Posterior Cramér-Rao lower bound (PCRLB) is proposed. This bound provides theoretical justification for a heuristic one proposed by other researchers in this area. Theoretical complexity analysis is provided to show the efficiency of the proposed bound, compared to the existing bound.

Distributed Estimation and Performance Limits in Resource-constrained
Wireless Sensor Networks

By

Yujiao Zheng

M.S., Syracuse University, 2011

B.S. University of Science & Technology of China, Hefei, China, 2008

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in Electrical and Computer Engineering in the Graduate
School of Syracuse University

May 2014

©Copyright 2014

Yujiao Zheng

All Rights Reserved

Acknowledgement

Foremost, I own my deepest gratitude to my advisor Prof. Pramod Varshney for the continuous support of my study and research. His patience, enthusiasm, motivation and immense knowledge deeply impress me and will benefit me a lot in my whole life. His humor brings me amount of fun during the research. It is his kind encouragement that makes me have strong confidence and truly love the current research. I am deeply indebted to his insightful suggestions on my research problem, which always bring my understanding of the research problem to a higher level. Without him, the thesis would not have been possible.

I would like to express my appreciation to Dr. Ruixin Niu who has been always providing strong technical support and professional suggestions, especially in the first few years of my Ph.D. I really appreciate the discussion with him during the process of research towards this dissertation. I learn quite a lot from the discussion: his experience, inspiration and the talent thoughts on research.

I am also indebted to Dr. Thakshila Wimalajeewa who has provided valuable advices on part of my work in this dissertation.

I would like to show my gratitude to all my labmates in the Sensor Fusion Group. I appreciate their kind support on me in both research and life. It is these nice persons that make the whole research group like a big family. Doing research in such a family is joyful and exciting.

My sincere thanks also go to all the committee members for examining my thesis defense. I benefit tremendously from their personal insights and inspirational questions on my dissertation.

I am grateful to my whole family for their encouragement for my studying aboard, especially to my dearest father. Some of his advices on me actually change my whole life. I have

been encouraged to study and learn new things by him since I was young. He is never too strict with me but lets me do what I am really interested in. Without him, I cannot pursue the advanced education here today.

Last but not the least, I would like to thank my husband, Zhou, for his constant love and support. He has been the spiritual sustenance for me over the past years when pursuing the PhD degree.

Contents

| | |
|--|------------|
| Acknowledgement | v |
| List of tables | xi |
| List of figures | xii |
| 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Main Contributions | 5 |
| 1.3 Organization of the Dissertation | 7 |
| 2 Estimation Theory Fundamentals | 10 |
| 2.1 Estimation of A Parameter | 10 |
| 2.2 Sequential Bayesian Estimation | 11 |
| 2.2.1 Performance Bounds | 16 |
| 2.2.1.1 Cramér-Rao Lower Bound | 16 |
| 2.2.1.2 Posterior Cramér-Rao Lower Bound | 17 |
| 3 Sequential Bayesian Estimation with Censored Data | 19 |
| 3.1 Motivation | 19 |
| 3.2 Problem Formulation | 26 |

| | | |
|----------|--|-----------|
| 3.2.1 | System Model | 26 |
| 3.2.2 | Particle Filter at the FC | 28 |
| 3.2.3 | Censoring Threshold Design | 29 |
| 3.3 | Censoring and Fusion with Missing Data | 30 |
| 3.3.1 | Overview | 30 |
| 3.3.2 | The Full Likelihood Function | 31 |
| 3.3.2.1 | Feedback System | 32 |
| 3.3.2.2 | Non-feedback System | 38 |
| 3.4 | Censoring Based on an Information Theoretic Metric | 41 |
| 3.5 | The Vector Observation Case | 44 |
| 3.5.1 | Feedback System | 44 |
| 3.5.2 | Non-feedback System | 47 |
| 3.6 | Censoring and Fusion with Missing Data for Nonlinear Systems | 49 |
| 3.7 | Simulation Results | 53 |
| 3.7.1 | Linear System—the Scalar Observation Case | 54 |
| 3.7.1.1 | Feedback System | 54 |
| 3.7.1.2 | Non-feedback System | 57 |
| 3.7.2 | Linear system—the Vector Observation Case | 58 |
| 3.7.3 | Nonlinear System | 59 |
| 3.7.4 | Discussion | 62 |
| 3.8 | Summary | 63 |
| 4 | Probabilistic Sensor Management for Target Tracking via Compressive Sensing | 70 |
| 4.1 | Motivation | 70 |
| 4.2 | Problem Formulation | 72 |

| | | |
|----------|--|------------|
| 4.2.1 | System Model | 72 |
| 4.2.2 | Compressive Sensing | 73 |
| 4.2.3 | Sparsity Formulation | 74 |
| 4.3 | The Sensor Management Problem | 76 |
| 4.4 | Simulation Results | 80 |
| 4.5 | Summary | 83 |
| 5 | Fusion of Quantized Data for Bayesian Estimation Aided by Controlled Noise | 84 |
| 5.1 | Motivation | 84 |
| 5.2 | A Review of Widrow’s Statistical Theory of Quantization | 87 |
| 5.3 | Problem Formulation | 87 |
| 5.4 | Controlled Noise Aided MMSE Estimation | 89 |
| 5.4.1 | Bayesian Estimators and Fisher Information | 89 |
| 5.4.1.1 | Optimal MMSE Estimator | 90 |
| 5.4.1.2 | Sub-Optimal LMMSE Estimator | 92 |
| 5.4.2 | LPF-noise Design and Bit Allocation | 95 |
| 5.5 | Numerical Results | 96 |
| 5.5.1 | Experiment 1 | 96 |
| 5.5.2 | Experiment 2 | 96 |
| 5.6 | Summary | 100 |
| 6 | Closed-form Performance bound for Source Localization with Quantized Data in Wireless Sensor Networks | 101 |
| 6.1 | Motivation | 101 |
| 6.2 | Problem Formulation | 103 |

| | | |
|---------|---|-----|
| 6.3 | ML Location Estimator and Exact CRLB | 104 |
| 6.4 | Closed-Form Expressions for FIM and CRLB | 110 |
| 6.5 | Analysis of FIM Approximation Error | 114 |
| 6.6 | Examples | 116 |
| 6.6.1 | Example 1: Gaussian-like Decay Model | 116 |
| 6.6.1.1 | Fisher Information Matrix | 117 |
| 6.6.1.2 | Error Analysis | 118 |
| 6.6.2 | Example 2: Power Law Decay Model | 119 |
| 6.6.2.1 | Fisher Information Matrix | 119 |
| 6.6.2.2 | Error Analysis | 120 |
| 6.7 | Simulation Results | 120 |
| 6.7.1 | Quantization Thresholds | 120 |
| 6.7.2 | ML estimator | 121 |
| 6.7.3 | Closed-Form CRLB | 123 |
| 6.7.3.1 | Gaussian-like Decay Model | 124 |
| 6.7.3.2 | Power Law Decay Model | 126 |
| 6.7.4 | Suboptimal Quantizer Design Based on the Closed-Form CRLB . . . | 127 |
| 6.8 | Summary | 129 |

7 An Alternative Approach to Compute Conditional Posterior Cramér Rao

| | | |
|-----|---|------------|
| | Lower Bound for Nonlinear Filtering | 133 |
| 7.1 | Motivaton | 133 |
| 7.2 | Conditional Posterior Cramér-Rao Lower Bounds | 136 |
| 7.3 | Proposed A-CPCRLB | 137 |
| 7.4 | Computational Complexity | 143 |
| 7.5 | Numerical Results | 145 |

| | | |
|----------|--|------------|
| 7.5.1 | Highly Nonlinear Case | 146 |
| 7.5.2 | Weakly Nonlinear Case | 146 |
| 7.5.3 | Discussion | 147 |
| 7.6 | Summary | 148 |
| 8 | Conclusions and Future Directions | 151 |
| 8.1 | Conclusions | 151 |
| 8.2 | Suggestions for Future Work | 153 |
| | Bibliography | 156 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Average number of transmissions (scalar observations) | 57 |
| 3.2 | experimental average number of transmissions (vector observations) | 59 |
| 3.3 | Experimental average number of transmissions (nonlinear system) | 62 |
| 5.1 | $R = 20$, Fisher Information Comparison | 98 |
| 5.2 | $R = 12$, Fisher information comparison | 99 |

List of Figures

| | | |
|-----|--|----|
| 3.1 | RMSE comparison for the feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send. | 56 |
| 3.2 | RMSEs for the CFwMD with/without covariance feedback for different values of l | 64 |
| 3.3 | RMSE comparison for the non-feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send. | 65 |
| 3.4 | RMSE comparison for the feedback system with $N = 2$ (vector observations). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send. | 65 |
| 3.5 | RMSEs for the CFwMD with/without covariance feedback for different values of l (vector observations). | 66 |
| 3.6 | RMSE comparison for the non-feedback system with $N = 2$ (vector observation). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send. | 67 |
| 3.7 | Target trajectory and sensor deployment in the ROI | 67 |
| 3.8 | RMSE comparison for the nonlinear system with feedback. | 68 |

| | | |
|------|--|-----|
| 3.9 | RMSEs as a function of the average number of transmission at each time. . . | 68 |
| 3.10 | RMSEs for the the CFwMD scheme with/without covariance feedback for different values of l in a nonlinear system. | 69 |
| 4.1 | MSE comparison for different approaches. | 81 |
| 4.2 | \mathbf{p} at different time steps. Circle: true state; Square: sensor assigned non- negligible transmission probability. | 82 |
| 5.1 | Fusion of quantized data for Bayesian estimation. | 85 |
| 5.2 | Bayesian estimation aided by controlled noise. S: sensor; Q: quantizer; z: sensor data; u: quantized data; d: controlled noise; y: data received at fusion center. . . | 88 |
| 5.3 | Fisher information as a function of σ_d | 97 |
| 6.1 | The signal intensity contours of a target located in a sensor field. | 104 |
| 6.2 | The CRLBs as a function of number of bits M (Gaussian-like Decay Model). | 122 |
| 6.3 | The CRLBs as a function of number of bits M (Power Law Decay Model) | 123 |
| 6.4 | RMSE and CRLBs as a function of number of bits M (Gaussian-like Decay Model). SNR = 40dB. | 124 |
| 6.5 | RMSE and CRLBs as a function of number of bits M (Power Law Decay Model). SNR = 40dB. | 125 |
| 6.6 | Elements of FIM as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Exact FIM; Dashed line + triangle: Approxi- mate FIM. SNR = 40dB and $M = 1$ | 126 |
| 6.7 | Trace of the CRLB matrix as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Exact CRLB; Dashed line + triangle: Approximate CRLB. SNR = 40dB, and $M = 1$ | 127 |

| | | |
|------|---|-----|
| 6.8 | Relative FIM as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Relative FIM; Dashed line: 1σ region of the Relative FIM. SNR = 40dB, and $M = 1$ | 128 |
| 6.9 | Element of FIM as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Exact FIM; Dashed line + triangle: Approximate FIM. SNR = 40dB and $M = 4$ | 129 |
| 6.10 | Trace of the CRLB matrix as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Exact CRLB; Dashed line + triangle: Approximate CRLB. SNR = 40dB and $M = 4$ | 130 |
| 6.11 | Relative FIM as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Relative FIM; Dashed line: $1\text{-}\sigma$ region of the Relative FIM. SNR = 40dB and $M = 4$ | 131 |
| 6.12 | Performances comparison for different quantizers (Gaussian-like Decay Model). SNR = 40dB and $\lambda = 6$ | 132 |
| 6.13 | Performances comparison for different quantizers (Power Law Decay Model). SNR = 40dB and $\lambda = 4$ | 132 |
| 7.1 | Highly Nonlinear Case | 149 |
| 7.2 | Weakly Nonlinear Case | 150 |

Chapter 1

Introduction

1.1 Background

Sensor network technology has developed tremendously over the past decade due to the advances in wireless communications and digital electronics. Current sensor nodes can be made tiny in size, low-cost, low-power and consist of sensing, data processing and communication components [1]. A sensor network is composed of a large number of sensor nodes, which are either randomly deployed or placed at pre-determined locations in the region of interest (ROI). Sensor nodes can be as simple as a data collector, and in that case the acquired raw data are sent to the node responsible for data fusion, *i.e.*, for processing all received data to yield the global inference. Another option is to have a processor fitted at the sensor nodes, enabling local data processing capability, and then, instead of raw data, only required processed data are transmitted to the fusion center, that yields the global inference.

There are many different types of sensors such as seismic, thermal, visual, vibration, acoustic and radar which can form different kinds of sensor networks, and, therefore, can be employed for a wide range of applications including the following:

- **Military applications:** Sensor nodes can be deployed in the battlefield to monitor the activities of both sides, gather battle damage assessment data, and detect attacks.
- **Environmental applications:** The term Environmental Sensor Networks has evolved to cover many applications of wireless sensor networks (WSNs) to earth science research. This includes *air pollution monitoring*: WSNs are deployed to monitor the concentration of dangerous gases that are harmful for citizens; *forest fire detection*: sensor nodes can be installed in a forest to detect onset of a fire event. The nodes can be equipped with sensors to measure temperature, humidity and gases which are produced by fire in the trees or vegetation, enabling an early detection of a fire and its spread; and *water quality monitoring*: many wireless distributed sensors are deployed to monitor water properties in dams, rivers, lakes and oceans, as well as underground water reserves.
- **Passive localization and tracking:** Sensor nodes can be deployed in a ROI, and receive passive signals to localize a static target or track a moving target in the region.

In this dissertation, we are interested in distributed inference in a WSN with limited resources, and focus on two important problems: target localization and tracking. Target localization involves estimation of the position of a static target whereas target tracking deals with state estimation of a dynamic target, i.e., position and velocity. For target tracking, the Kalman filter is the optimal algorithm to infer the state of a linear Gaussian system. When the system is Gaussian but nonlinear, extended Kalman filter (EKF) provides near-optimal solutions in many cases. Due to current developments in computation power, particle filtering (PF) has become a very promising and practical tool for sequential estimation of unknown states of a general nonlinear non-Gaussian dynamical system.

Due to the small size of battery powered sensor nodes, power consumption is the most

important constraint on sensor networks. In order to prolong the lifetime of a sensor network, it is desired to reduce communications. Therefore, current solutions in the field focus on the trade-off between the performance of inference and the lifetime of the network. Another important characteristic of the sensor network is the limited bandwidth. Many researchers have devoted their research effort to solving the problem of optimization of the network performance under bandwidth constraints. These two issues also give rise to the well known problem of data selection. In the current literature, data selection is realized by selecting a subset of sensors (sensor selection problem) or, more generally, by distributing available bits to sensors (bit allocation problem) [2–7]. In this dissertation, we propose a novel distributed data selection idea for the sequential Bayesian estimation problem where each local sensor sensors its measurement first and sends it to the fusion center (FC) only if the measurement is informative enough in a certain sense. The censored measurements are treated as missing observations at the fusion center. Another novelty of our work is the proposed fusion rule at the fusion center, which fuses both missing observations and received ones, given that the censoring rule is known *a priori*. Since missing information conveyed by the missing data is exploited, the proposed algorithm yields better performance than the one that completely ignores the missing information.

An interesting observation for a large WSN is that only a few nodes have significant and informative observations. Thus, the concatenated measurement vector at the fusion center can be considered to be sparse and compressible. This interpretation motivates us to introduce the concept of compressive sensing (CS) [8,9] into the sensor management problem, and propose a novel compressive sensing based sensor management methodology for a sensor network monitoring a moving target. Our methodology utilizes a probabilistic transmission scheme over multiple access channels (MACs).

In addition, a novel idea of fusing quantized data for Bayesian estimation in sensor

networks is proposed in this dissertation, where the quantized data are not fused directly by the FC for parameter estimation, but preprocessed by injecting independent controlled noise. The basic idea was inspired by Widrow’s statistical theory of quantization [10]. The addition of noise after quantization is equivalent to low pass filtering in the characteristic function (CF) domain, such that the original analog observation can be recovered. Therefore, the whole process of quantizing and injecting controlled noise can be theoretically modeled as an additive disturbance, whose distribution is analytically derived in this dissertation. This theoretical model facilitates the derivation of the optimal minimum mean squared error (MMSE) estimator, the posterior Cramér-Rao lower bound (PCRLB), the near-optimal linear MMSE (LMMSE) estimator, and its corresponding mean squared error (MSE), all of which are in *exact* forms. Furthermore, the numerical results show that the LMMSE estimator can provide comparable performance to that of the optimal MMSE estimator while saving a lot of computation effort.

Irrespective of the type of estimator proposed, it is always desirable to have a performance assessment tool available to assess achievable estimation performance. One well known tool is the Cramér-Rao lower bound (CRLB) [11], which provides a bound on the performance of estimators of the system state for filtering problems. For a general multi-dimensional discrete-time nonlinear filtering problem, Tichavsky et al. [12] provided an elegant recursive approach for calculating the sequential PCRLB. In this dissertation, we propose two bounds. The first provides a bound on the asymptotic performance of the maximum likelihood (ML) estimator in a sensor network, where quantized data are sent to the fusion center and sensors are densely deployed in the ROI, *i.e.*, the number of sensors is large. The other provides online performance assessment for a sequential Bayesian estimator by exploiting the information provided by measurements, in contrast to the offline version of sequential PCRLB [12], which averages out the measurements.

1.2 Main Contributions

The main contributions of this dissertation are outlined as follows:

- A new framework for sequential Bayesian estimation in a sensor network is proposed, where both linear and nonlinear systems are considered. The framework consists of two processes: censoring of measurements at the local sensors and fusion of both received measurements and missing ones at the FC. In this scheme, each local sensor maintains a KF for a linear Gaussian system or an EKF for a nonlinear system and the FC runs a PF to track the system state. Informative measurements are selected for transmission by a per-sensor censoring process executed at the sensors at each time. We use an innovation based censoring rule in this work for both linear and nonlinear systems. Though the less informative measurements are not sent to the FC, their absence still conveys some information, and the proposed fusion scheme exploits this information conveyed by the missing messages. Numerical results show that, under the same bandwidth constraint, the proposed scheme outperforms the one that ignores missing data information and the one that selects sensors randomly for information transmission.
- A novel compressive sensing based sensor management methodology for a sensor network tracking a moving target is proposed. In this approach, a probabilistic transmission strategy over MACs is utilized and the sensor management problem is transformed to a constrained optimization problem where the goal is to determine the optimal probabilities with which sensors should transmit such that a desired inference performance is guaranteed. Numerical results are provided to validate the proposed methodology.
- Inspired by Widrow's statistical theory on quantization, a novel scheme for Bayesian estimation in a sensor network where the local sensor observations are quantized before

their transmission to the FC is proposed. At the FC, instead of fusing the quantized data directly, we propose to fuse the post-processed data obtained by adding an independent controlled noise to the received quantized data. The injected noise acts like a low-pass filter in the characteristic function (CF) domain such that the output is an approximation of the original raw observation. The optimal minimum mean squared error (MMSE) estimator and the posterior Cramér-Rao lower bound for this estimation problem are derived. Based on the Fisher information, the optimal controlled Gaussian noise and the optimal bit allocation are obtained. In addition, a near-optimal linear MMSE estimator is derived to significantly reduce the computational complexity.

- For a large and dense sensor network, the impact of sensor density is investigated on the performance of an ML location estimator using quantized sensor data. The ML estimator fuses quantized data transmitted from local sensors to estimate the location of a source. A general smooth, differentiable, and isotropic signal decay model is adopted to make the problem tractable. Two special cases are given as examples. The general model is suitable for situations such as passive sensors monitoring a target emitting acoustic signals. The exact Cramér-Rao lower bound (CRLB) on the estimation error is derived. In addition, an approximate closed-form CRLB by using the law of large numbers (LLN) is obtained. The closed-form results indicate that the Fisher information is a linearly increasing function of the sensor density. Even though the results are derived assuming a large number of sensors, numerical results show that the closed-form CRLB is very close to the exact CRLB for even low sensor densities. Moreover, simulation results show that the closed-form CRLB can be used as a metric to design suboptimal quantizers which achieve almost the same performance as the optimal ones designed by the exact CRLB.

- The recursive procedure to compute the posterior Cramér-Rao lower bound (PCRLB) for sequential Bayesian estimators, derived by Tichavsky *et al.*, provides an off-line performance bound for a general nonlinear filtering problem. Since the corresponding Fisher information matrix (FIM) is obtained by taking the expectation with respect to all the random variables, this PCRLB is not well suited for online adaptive resource management for dynamic systems. For online performance evaluation for a nonlinear system, the concept of conditional PCRLB was proposed by Zuo *et al.* in 2011 [13]. In this dissertation, an alternative online conditional PCRLB is proposed. Numerical examples are provided to show that the accuracy of the proposed online bound is comparable to the one proposed by Zuo *et al.* while theoretical computation complexity analysis is provided to show that it also saves much computation effort.

1.3 Organization of the Dissertation

The rest of this dissertation is organized as follows.

Some essential results in estimation theory which are frequently used in the dissertation are provided in Chapter 2.

In Chapter 3, a novel data selection and data fusion scheme for sequential Bayesian estimation problems in sensor networks is presented. The scheme consists of two procedures at any given time: firstly, each local sensor censors its measurement and sends it to the FC only if the measurement is informative enough based on the proposed censoring rule, and then the FC uses both received measurements and missing ones to infer the state of the target. A PF is maintained at the FC, and the utilization of the missing information conveyed by missing data is realized by updating the weights of particles with the full likelihood function. The advantage of the proposed scheme compared to other schemes, such as random selection

and ignoring missing data, is illustrated by numerical examples.

In Chapter 4, a novel sensor management approach for target tracking in a sensor network based on CS is investigated. We employ a multiple access channel (MAC) model with probabilistic transmissions, based on which, we can obtain an equivalent representation of our problem as the standard CS problem. With this model, the sensing matrix is completely determined by each sensor's probability of transmission, and the design of the sensing matrix is reduced to finding the optimal probability of transmission for each sensor such that a desired performance guarantee in tracking is achieved. Furthermore, due to the equivalent representation, the sensing matrix acts as a sensor management operator. Thus, the sensor management problem is solved by formulating it as a compressive sensing problem. The performance of the proposed approach is validated by numerical examples.

In Chapter 5, a novel scheme for fusing quantized data for the Bayesian estimation problem in sensor networks is discussed, where quantized data are pre-processed at the FC before fusion, namely, a controlled noise is independently injected into each received quantized measurement. The injected noise works as a low-pass filter, and therefore, under a certain condition, the original analog data can be completely recovered. Such a model also provides a theoretical justification for the additive noise modeling of the quantization procedure, and hence, facilitates the derivation of LMMSE. A bit allocation problem based on the proposed scheme is also discussed in this chapter, and illustrative examples are provided to show its performance.

Bounds on the performance of location estimation in a sensor network are considered in Chapter 6, where sensors are densely deployed in the ROI. We use a general isotropic signal attenuation model which is smooth, differentiable, as well as monotonically decreasing. The statistical model for sensor observations is assumed known. The ML location estimator based on quantized sensor data and its corresponding exact CRLB are derived. Since the

number of sensors is large, an approximation based on the law of large numbers (LLN) is used to derive the closed-form CRLB for the considered signal decay model. Theoretical analysis of the error introduced by the LLN approximation is provided. Two commonly used models in wireless communication, namely, the Gaussian-like decay model and the power law decay model, are investigated respectively as two concrete examples for the general model. Numerical examples are provided to illustrate the effectiveness of the closed-form solution that we propose.

In Chapter 7, we propose a new conditional PCRLB, which is based on the representation of the conditional PCRLB proposed in [13]. We call this bound the alternative conditional PCRLB (A-CPCRLB), since we discard the auxiliary FIM which is involved in the recursive update for the conditional PCRLB presented in [13]. Instead, an alternative approximate recursive update is proposed, which is direct, more compact and efficient than the one proposed in [13]. Furthermore, when the state dynamic model is linear and Gaussian, we show that this bound reduces to the modified PCRLB proposed in [14]. Hence, the proposed A-CPCRLB provides a generalization and theoretical justification for the one used in [14]. Numerical computation such as the sequential Monte Carlo methods is used to compute the proposed bound. Performance analysis in terms of computational complexity associated with the computation of the bound is provided. A numerical example is provided to compare the original CPCRLB [13] with our proposed bound, namely the A-CPCRLB.

Chapter 8 concludes the dissertation along with a discussion on potential directions for future research.

Chapter 2

Estimation Theory Fundamentals

2.1 Estimation of A Parameter

In parameter estimation problems, two models for the unknown parameter are employed:

(1). The parameter is not a random variable, *i.e.*, the value of the unknown parameter is fixed. (2). The parameter is a random variable with known *a priori* probability density function (pdf).

A common method to estimate a non-random variable θ is the maximum likelihood method, which yields the maximum likelihood estimate (MLE) [15]:

$$\hat{\theta}^{\text{ML}}(x) = \arg \max_{\theta} \mathcal{L}(\theta|x) \quad (2.1)$$

where the likelihood $\mathcal{L}(\theta|x) \triangleq p(x|\theta)$, the conditional pdf of the observation.

For a random parameter θ with a prior pdf $p(\theta)$, the maximum a posterior (MAP) estimator is used for estimation to incorporate both prior information and information from data [15]. That is,

$$\hat{\theta}^{\text{MAP}}(x) = \arg \max_x p(\theta|x) = \arg \max_x \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)dx} \quad (2.2)$$

The estimation error of an estimator $\hat{\theta}(x)$ is $\hat{\theta}(x) - \theta$, and the mean square error (MSE) is given by the trace of the covariance matrix, *i.e.*,

$$\text{MSE} = \text{tr} \left\{ E \{ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \} \right\} \quad (2.3)$$

The estimator which minimizes the MSE is called the minimum mean square error (MMSE) estimator and given by [16]

$$\hat{\theta}^{\text{MMSE}}(x) = E\{\theta|x\} \quad (2.4)$$

2.2 Sequential Bayesian Estimation

The sequential Bayesian estimation problem is to estimate the state of a dynamic system based on the observations on the system over time. The evolution of the unknown state sequence \mathbf{x}_k is assumed to be a first-order Markov process and modeled as

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) \quad (2.5)$$

where $f_k : \mathfrak{R}^{n_x} \times \mathfrak{R}^{n_u} \rightarrow \mathfrak{R}^{n_x}$ and \mathbf{u}_k is the independent identically distributed (i.i.d.) process noise with dimension n_u . The measurement model is given by

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{v}_k), \quad (2.6)$$

where $h_k : \mathfrak{R}^{n_x} \times \mathfrak{R}^{n_v} \rightarrow \mathfrak{R}^{n_z}$, \mathbf{v}_k is the i.i.d. measurement noise, n_z and n_v are the dimensions of the measurement and measurement noise vectors, respectively. The process and the measurement noise distributions are denoted by $p_{\mathbf{u}_k}(\mathbf{u})$ and $p_{\mathbf{v}_k}(\mathbf{v})$, respectively, and the two noises are assumed to be independent of each other.

Denote the states and measurements up to time k as $\mathbf{x}_{0:k}$ and $\mathbf{z}_{1:k}$. Then, the joint probability density function (PDF) of $(\mathbf{x}_{0:k}, \mathbf{z}_{1:k})$ can be decomposed as

$$p(\mathbf{x}_{0:k}, \mathbf{z}_{1:k}) = p(\mathbf{x}_0) \prod_{i=1}^k p(\mathbf{x}_i | \mathbf{x}_{i-1}) \prod_{j=1}^k p(\mathbf{z}_j | \mathbf{x}_j) \quad (2.7)$$

Based on (2.7), three categories of Bayesian estimation problems can be investigated [16]:

- **Prediction:** estimate the state at time $k + k'$ based on the measurements up to and including time k , where $k' > 0$, *i.e.*, it is an estimation of the state at a future time.
- **Filtering:** estimate the state at the current time k based on the measurements up to and including time k . This is the problem that this dissertation will focus on.
- **Smoothing:** estimate the state at some earlier time k'' based on the measurements up to time k ($k > k''$).

Many algorithms have been developed for the Bayesian filtering problem, among which, three frequently used filters are Kalman filter, extended Kalman filter (EKF) and particle filter.

1. Kalman Filter

The Kalman filter performs optimal Bayesian filtering for a system with linear state and measurement models and additive Gaussian noise for both state and measurement processes [16] in the MMSE sense. Specifically, the linear Gaussian system is given by

$$\mathbf{x}_{k+1} = F_k \mathbf{x}_k + \mathbf{u}_k \quad (2.8)$$

$$\mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{v}_k \quad (2.9)$$

where $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, Q_k)$, and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, R_k)$.

Given the distribution of the initial state \mathbf{x}_0 , $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_{0|0}, P_{0|0})$, the Kalman filter recursively estimates the state of the system by two steps:

Prediction step:

$$\hat{\mathbf{x}}_{k|k-1} = F_{k-1} \hat{\mathbf{x}}_{k-1|k-1} \quad (2.10)$$

$$P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1} \quad (2.11)$$

Update step:

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (2.12)$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (2.13)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k (\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k-1|k-1}) \quad (2.14)$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \quad (2.15)$$

where K_k is the Kalman gain, $\hat{\mathbf{x}}_{k|k-1}$ is defined as the estimate of the state \mathbf{x}_k conditioned on measurements up to and including time $k-1$, and $P_{k|k-1}$ is defined as the covariance matrix of the estimation error. Obviously, $\hat{\mathbf{x}}_{k|k-1}$ and $P_{k|k-1}$ are predictions, while $\hat{\mathbf{x}}_{k|k}$ and $P_{k|k}$ are updated results after incorporating the observation at time k .

Remarks: 1) The recursive computation of the covariance matrix depends only on the model parameters and time index, which means that once the model is known, one can compute the covariance matrix at each time without acquiring the observations. 2) The innovation ν_k , defined as $\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k-1|k-1}$, is white and Gaussian with mean zero and covariance S_k .

2. Extended Kalman Filter

The EKF is an extended version of the Kalman filter to solve the estimation problem for a nonlinear system by linearizing the nonlinear functions using first-order Taylor series expansion. The recursive estimation process is given as

$$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}) \quad (2.16)$$

$$P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1} \quad (2.17)$$

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (2.18)$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (2.19)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k (\mathbf{z}_k - h(\hat{\mathbf{x}}_{k|k-1})) \quad (2.20)$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \quad (2.21)$$

where

$$F_{k-1} = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}_{k-1|k-1}} \quad (2.22)$$

$$H_k = \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{\mathbf{x}_{k|k-1}} \quad (2.23)$$

Remarks: 1) At time k , the nonlinear function f is directly used to compute the predicted state. Similarly, h is used directly to compute the predicted measurement. 2) The nonlinear functions f and h are linearized to get F_{k-1} and H_k , which are involved in computing the covariance matrix. The Jacobian matrices F_{k-1} and H_k are evaluated at the estimate $\hat{\mathbf{x}}_{k-1|k-1}$ and prediction $\hat{\mathbf{x}}_{k|k-1}$ respectively.

3. Particle filter

Even though EKF can deal with nonlinear systems, it performs poorly when the system is highly nonlinear [17]. One has to resort to other approaches for a general nonlinear non-Gaussian system.

As stated before, the filtering problem is to compute the distribution $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, and this can be done recursively in two steps.

prediction step:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1}) \quad (2.24)$$

update step:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) \propto p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) \quad (2.25)$$

where $p(\mathbf{x}_k | \mathbf{z}_{1:k-1})$ can be treated as the prior, and it is updated to get the posterior after the measurement \mathbf{z}_k is available according to Bayes's rule.

The computation of (2.24) and (2.25) cannot be, in general, carried out analytically, except for some special cases (*e.g.*, for linear Gaussian system, KF is an analytical solution).

Therefore, approximate approaches are necessary, and Monte carlo based methods such as sequential importance sampling (SIS) are basic methods for this purpose.

The basic idea of SIS is to recursively update the weighted samples $\{\mathbf{x}_{0:k}^i, w_k^i\}_{i=1}^N$ used to approximate the posterior distribution $p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})$. The samples are also called particles. In fact, SIS is a sequential version of importance sampling. In important sampling, one can approximate a target distribution $p(x)$ by samples drawn from a proposal distribution $q(x)$. It is usually applicable to the case when directly drawing a sample from the target distribution is difficult. Each sample x^i is weighted by $w^i \propto \pi(x)/q(x)$, where $\pi(x) \propto p(x)$ and is able to be evaluated, to compensate for the discrepancy between the target distribution and the proposal distribution. Applying the idea to the posterior distribution $p(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1})$, one can get

$$p(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1}) = \sum_{i=1}^N w_{k-1}^i \delta(\mathbf{x}_{0:k-1}^i) \quad (2.26)$$

where $\delta(\mathbf{x}_{0:k-1}^i)$ is a delta function centered at $\mathbf{x}_{0:k-1}^i$. Then, the SIS will update the sample set $\{\mathbf{x}_{0:k-1}^i, w_{k-1}^i\}_{i=1}^N$ to approximate the posterior at the next time step. One can do this by assuming that the proposal distribution at time k can be factorized as

$$q(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k})q(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1}) \quad (2.27)$$

Then, one can augment each particle $\mathbf{x}_{0:k-1}^i$ with \mathbf{x}^i drawn from the distribution $q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k})$ to obtain the new particle $\mathbf{x}_{0:k}^i$ for the state at time k .

The weight w_k^i at time k is given by [18]

$$\begin{aligned} w_k^i &\propto \frac{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})}{q(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \\ &\propto w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{0:k-1}^i, \mathbf{z}_{1:k})} \end{aligned} \quad (2.28)$$

If we further assume that $q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}) = q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$, *i.e.*, the proposal at time k only depends on the most recent state and most recent measurement, then we can only store \mathbf{x}_{k-1}^i

and generate \mathbf{x}_k^i from the proposal $q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$. Thus, the update can be simplified as

$$\mathbf{x}_k^i \sim q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k) \quad (2.29)$$

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{0:k-1}^i, \mathbf{z}_{1:k})} \quad (2.30)$$

and the weights need to be normalized to represent a distribution.

The SIS algorithm is the most basic particle filtering approach, and there are a number of variants based on it, among which, the most commonly used one is called sequential importance resampling (SIR) particle filtering. In SIR, the update procedure is followed by a resampling procedure at each time step to avoid the degeneracy problem [18], and the transition distribution $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is chosen as the proposal distribution. Thus, in the SIR particle filtering, the update procedure is given by

$$\mathbf{x}_k^i \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}) \quad (2.31)$$

$$w_k^i \propto p(\mathbf{z}_k|\mathbf{x}_k^i) \quad (2.32)$$

Note that, the resampling process yields each sample with the same weight $1/N$, and therefore, there is no w_{k-1}^i in the weight update equation. Throughout this dissertation, we will use SIR particle filtering if a particle filter is used.

2.2.1 Performance Bounds

2.2.1.1 Cramér-Rao Lower Bound

Let \mathbf{x} be a n_x dimensional unknown deterministic parameter and \mathbf{z} be a n_z -dimensional measurements. If an unbiased estimator $\hat{\mathbf{x}}(\mathbf{z})$ exists, then a lower bound on the estimation error of this estimator is given by the Cramér-Rao lower bound (CRLB) [19], *i.e.*,

$$E\{(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})^T\} \geq J^{-1} \quad (2.33)$$

where J is the Fisher information matrix (FIM) and given by

$$\begin{aligned} J &= -E \{ \Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{z}|\mathbf{x}) \} \\ &= E \{ \nabla_{\mathbf{x}} \ln p(\mathbf{z}|\mathbf{x}) \nabla_{\mathbf{x}}^T \ln p(\mathbf{z}|\mathbf{x}) \} \end{aligned} \quad (2.34)$$

where $\nabla_{\mathbf{x}} \triangleq [\frac{\partial}{\partial x_1} \ \cdots \ \frac{\partial}{\partial x_{n_x}}]^T$, and $\Delta_{\mathbf{x}}^{\mathbf{x}} \triangleq \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^T$.

2.2.1.2 Posterior Cramér-Rao Lower Bound

If the parameter \mathbf{x} is a random variable, then the lower bound on the estimation error is named as posterior Cramér-Rao lower bound (PCRLB), and it is given by [19]

$$E\{(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})^T\} \geq J^{-1} \quad (2.35)$$

where

$$\begin{aligned} J &= -E \{ \Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{z}) \} \\ &= E \{ \nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{z}) \nabla_{\mathbf{x}}^T \ln p(\mathbf{x}, \mathbf{z}) \} \end{aligned} \quad (2.36)$$

An elegant recursive computation of the PCRLB for nonlinear filtering problems has been presented by Tichavský etc. in [20]. The results are provided here for easy reference.

Given the nonlinear system (2.5) with measurement model (2.6), the sequence of the FIM J_k for estimating state vectors \mathbf{x}_k obeys the recursion

$$J_{k+1} = D_k^{22} - D_k^{21} (J_k + D_k^{11})^{-1} D_k^{12} \quad (2.37)$$

where

$$\begin{aligned} D_k^{11} &= E \{ -\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \log p(\mathbf{x}_{k+1}|\mathbf{x}_k) \} \\ D_k^{12} &= E \{ -\Delta_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \log p(\mathbf{x}_{k+1}|\mathbf{x}_k) \} = (D_k^{21})^T \\ D_k^{22} &= E \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} [\log p(\mathbf{x}_{k+1}|\mathbf{x}_k) + \log p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})] \right\} \\ &= D_k^{22,a} + D_k^{22,b} \end{aligned}$$

These performance bounds, namely CRLB and PCRLB, provide achievable estimation performance against which performance of estimation algorithms can be assessed.

Chapter 3

Sequential Bayesian Estimation with Censored Data

3.1 Motivation

In the literature, the sequential Bayesian estimation problem has been mainly investigated for three fundamental network architectures: centralized, distributed and decentralized networks. In a centralized structure, the local sensor nodes transmit either analog [21] or quantized measurements [7, 22, 23] to a FC, where the sensor data are fused by a Bayesian filter to update the system state estimate. If all the analog sensor data are transmitted to the FC, the FC yields the optimal estimation performance, meaning that no other network architecture can deliver a better performance than the centralized architecture with analog sensor data. Further, in a centralized network, optimal information fusion can be performed in a straightforward manner at the FC. But a centralized network requires a large amount of communication between the sensors and the FC. Further, the network is vulnerable to the failure of the FC, which could compromise the whole network.

In a distributed network, each local sensor node runs a local Bayesian state estimator, and

obtains its own local state estimate based on its local measurements. These local estimates, or state posterior probability density functions (PDFs), are transmitted to a global FC, where they are fused to get a more accurate global state estimate. The distributed network has reduced communication requirements, since instead of transmitting raw sensor data at the sensor sampling rate, each sensor could transmit state estimates at a much reduced rate. Furthermore, the distributed network is much more robust, since each local sensor node maintains its own state estimate. However, one challenging problem for fusion of estimates is that all the local estimates are dependent since all the local filters are estimating the same Markov stochastic process [21]. The problem of distributed Kalman filtering has been investigated in [21,24–29]. For nonlinear filtering in distributed networks, the optimal fusion scheme was developed in [30,31] which involves the transmission of the local state posterior PDFs to the FC and high dimensional integrals at the FC.

In a decentralized network, each sensor fuses its own local state estimate with information received from its neighboring sensors, and each local sensor communicates only with its neighbors. Due to its diffusive communication strategy, this architecture does not require specialized routing, and in general avoids bottleneck in communications. It is scalable and very robust to single point of failure. However, to carry out the optimal fusion algorithm, the so-called channel filter [24,32,33], one needs to maintain all the relevant communication and fusion events history, also known as pedigree information [31], which may become a prohibitive task over time. Therefore, existing fusion algorithms in decentralized networks are typically suboptimal approaches. In decentralized networks, estimate consensus among distributed agents has drawn much attention. For the linear estimation problem in decentralized networks, algorithms have been proposed to reach a consensus among all the nodes, which include gossip algorithms [34,35], consensus algorithms [36–39], and combined approaches [40]. For the distributed nonlinear filtering problem, efforts have been made to

develop consensus particle filtering approaches [41–45].

The framework we propose in this chapter combines the advantages of both the centralized and distributed networks to achieve communication efficiency, improved estimation performance, and robustness. In this framework, each local sensor node runs its local state estimator, which facilitates censoring of its measurement so that only informative measurements are sent to the FC. Since local state estimation is performed at each local sensor, it is robust against single point of failure. Compared to the centralized network, it has reduced communication rate through sensor censoring. However, different from a typical distributed architecture but similar to a centralized architecture, only informative *raw* sensor measurements are sent to the FC in our proposed framework. In this chapter, we have developed a novel data censoring and fusion approach for such a system architecture. The approach proposed in this chapter could be extended for decentralized networks without a FC and will be discussed in our future work.

As mentioned earlier, in a sensor network with a FC, the ideal scenario is for all the sensors to send their observations to the FC for the inference task (*e.g.*, source localization and target tracking). However, in practice, due to bandwidth constraints or energy limitations in the network, it is usually desirable to have only a subset of sensors transmit their data at each time step. This gives rise to two interesting problems: 1) In a centralized sensor management framework, where local estimates are not obtained at local sensors, and the FC completely controls the selection of sensors, how to select the subset of sensors which are the most informative in either myopic (one-step ahead) manner or non-myopic (multiple steps ahead) manner, based on the accumulated information up to the current time? 2) In a distributed sensor management system, where each local sensor generates a local estimate based on local sensor measurements, and decides whether or not to send its measurement/estimate to the FC by itself, how to select the subset of measurements which are the most informative if the

current measurements are already in hand? The latter problem is sometimes referred to as the sensor censoring problem and is addressed in this chapter.

The first problem is a typical sensor management or sensor selection problem and a lot of effort has been devoted to it by different authors [2, 3, 5, 22, 46–50]. In [2], the sensor selection problem was formulated as an integer programming problem, which has been relaxed and solved through convex optimization. In [46] [5], non-myopic algorithms for sensor management for target tracking were provided. In [5], a multi-step sensor selection strategy by reformulating the Kalman filter was proposed, which is able to address different performance metrics and constraints on available resources. In the context of field estimation, the tradeoff between communication cost and estimation performance in multi-step sensor selection was considered in [47]. The finite horizon sensor scheduling problem that chooses which sensors should operate at each time-step to minimize a weighted function of the error covariance of the state estimation was addressed in [48], and algorithms were developed to solve for the optimal and suboptimal sensor schedule. In [3], a method that chooses sensors randomly according to a probability distribution such that the upper bound on the expected steady-state performance is minimized was developed. In [49], sensor selection is based on an entropy-based information measure, which is computed using the expected posterior distribution of the state to be estimated. Instead of information based metrics, in [50] [22], the recursive one-step-ahead posterior Cramér-Rao lower bound (PCRLB) on the mean squared error (MSE) of estimating the state vector has been explored as the metric to select informative sensors. More specifically, given the constraint that l sensors are selected at each time, the collection of l sensors which minimizes the cost function, the bound on the MSE of state estimate of the target, is the desired solution. Recently, a novel sensor selection approach was proposed in [51], where the Kalman gain matrix is designed through optimization and a sparsity-promoting penalty function is added to the objective function. The added term

penalizes the number of nonzero columns of the Kalman gain matrix, which corresponds to the number of active sensors. Therefore, only a few sensors send their measurements to the FC.

If quantizers are used at local sensors, then the bit allocation problem is a natural extension to the sensor selection problem [52] [7]. In [52], bit allocation among different samples was studied for a signal detection problem, while in [7], the available bandwidth R is distributed among the sensors in the sensor network in such a way that the PCRLB on the MSE is minimized for target tracking.

The second problem results in the so called censoring method in the area of distributed detection [53–56]. In [53], under a constraint on communication, an optimal censoring structure is proposed, through which, local sensors censor their likelihood ratios before sending them to the FC. Only the local likelihood ratios falling in the send region are sent to the FC for making the global decision. Later in [54], the fusion of decisions from censoring sensors transmitted over wireless fading channels was investigated, where optimal and suboptimal fusion rules were designed based on the knowledge of fading channels. Some practical issues regarding the design of censoring sensor networks including joint dependence of sensor decision rules, randomization of decision strategies, and partially known distributions of observations were further addressed in [55]. Per-sensor censoring scheme was also employed in [56], in which an ordering approach follows censoring to reduce the number of transmissions in the network, and the sensors with more informative observations transmit first. Sensor censoring has also been used to solve estimation problems [57]. The authors in [57] proposed another transmission scheme in which the sensor transmissions are ordered according to the magnitude of their measurements, and the sensors with magnitude smaller than a threshold, do not transmit.

Methods used to solve the above problems can be categorized as data selection methods

and all of them result in missing data from the viewpoint of the FC. Then, a crucial issue is whether the fact that variables are missing is related to the underlying values of the variables in the data set [58], and this categorizes the mechanism leading to missing data into three types according to [58]: i) missing completely at random (MCAR), *i.e.*, missingness does not depend on the data values; ii) missing at random (MAR), *i.e.*, missingness depends only on the observed components, not on the missing ones; iii) not missing at random (NMAR), *i.e.*, missingness depends on the missing values. Obviously, the missing data issue due to the data selection methods such as censoring belongs to the third type mentioned above. In this chapter, we focus on missing data due to the third mechanism, namely, on NMAR. Since the missing data also convey some information, they can be exploited to obtain better inference. In fact, the information conveyed by missing data due to NMAR has been considered implicitly in the distributed detection problem [53]. The parameter estimation problem that takes into account the NMAR missing data information has been considered in [59]. Nevertheless, to the best of our knowledge, for the Bayesian sequential estimation problem in the context of data selection/sensor censoring, such kind of approach has not yet been explored. A related but different work has been reported in [60] and references therein, which exploits ‘negative’ sensor evidence (expected but missing sensor data) for target tracking and data fusion. Though the work in [60] is similar to ours, it is different from our work in two major aspects: first, the missing measurements in [60] are due to the failed attempt by a radar system to detect a target, while in our work certain sensor data are missing because sensors censor their local data in a distributed manner to conserve communication bandwidth and send more informative sensor data to a FC; second, the missing information or ‘negative’ information in [60] is exploited in terms of fictitious measurements given by appropriate sensor models which is designed based on the background information on the sensor characteristics, while in our work, the missing

information is exploited in terms of the statistics of the missingness which can be computed giving the prior knowledge on the censoring rule. Hence, the two novelties of our work are: censoring measurements at local sensors to select informative measurements in a distributed manner, and fusing both received measurements and missing ones at the FC to exploit the information conveyed by the missingness of data.

The main contribution of this chapter is that we propose a scheme which provides better performance for target tracking in a sensor network when the bandwidth constraint and/or energy cost at local sensors is important to increase the lifetime of the network. In the proposed scheme, firstly, the local sensors censor their measurements in a distributed manner, and then the FC fuses both the received observations and missing ones. The proposed scheme is shown to be applicable to both linear and nonlinear systems, and both scalar and vector observations. Furthermore, we investigate the relationship between the censoring rule based on the innovation and the one based on the Kullback-Leibler (KL) divergence between the prior state distribution before the measurement is available and the posterior state distribution after the measurement is obtained. For the convenience of discussion throughout this thesis, we call the proposed scheme Censoring and Fusion with Missing Data (CFwMD), since in this scheme, a censoring method is employed at the sensors and the FC fuses data considering the information of missing data that are NMAR. We call the scheme which uses the same censoring method at the sensors but ignores the information about the missing data at the FC as Censoring and Fusion without Missing Data (CFoMD). The scheme, which does not use censoring at the sensor level but a probabilistic transmission strategy, which results in missing data that are MCAR, is called random-selection throughout this chapter. Numerical results demonstrate that CFwMD incurs less performance loss than CFoMD when compared to the all-send case (all sensors send their measurements to the FC), while they both outperform the random-selection under the same bandwidth constraint.

The rest of this chapter is organized as follows. In the next section, we formulate the problem. Then, we present the proposed CFwMD scheme for linear Gaussian systems when scalar observations are obtained at local sensors in Section 3.3, followed by the discussion on the equivalence between the censoring rule based on the innovation and the one based on the KL divergence in Section 3.4. Section 3.5 discusses the framework when vector observations are available at local sensors, and Section 3.6 generalizes the framework to nonlinear systems. We provide simulation results in Section 3.7 and summarize this chapter in Section 3.8.

3.2 Problem Formulation

3.2.1 System Model

Here, we consider a sequential Bayesian estimation problem in a sensor network with N sensors. Sensors report measurements to the FC for the inference task, *i.e.*, estimation of the system state, for example, the position and velocity of the target in the target tracking problem. Throughout this chapter, the channels between local sensors and the FC are assumed to be perfect.

The state model of the system is given as follows:

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{u}_k \quad (3.1)$$

where \mathbf{F}_k is the state transition matrix, \mathbf{x}_k is the $d \times 1$ state vector and \mathbf{u}_k is the white Gaussian process noise with zero-mean and covariance matrix \mathbf{Q}_k . Sensor i 's measurements are given by

$$\mathbf{z}_k^i = \mathbf{H}^i \mathbf{x}_k + \mathbf{n}_k^i \quad (i = 1, 2, \dots, N) \quad (3.2)$$

where \mathbf{H}^i is the observation matrix which maps the state space into the observation space and \mathbf{n}_k^i is white Gaussian measurement noise with zero-mean and covariance \mathbf{R}^i . In this

chapter, we first discuss the case in which scalar observations are obtained at local sensors, *i.e.*,

$$z_k^i = \mathbf{h}^{iT} \mathbf{x}_k + n_k^i \quad (i = 1, 2, \dots, N) \quad (3.3)$$

where \mathbf{h}^i is the measurement vector, the superscript T denotes vector/matrix transpose and n_k^i is white Gaussian noise with zero-mean and variance r^i .

In our CFwMD scheme, we design a censoring rule which measures the informativeness of the measurements at the sensor level, *i.e.*, at each time step, the i^{th} sensor first examines its measurement according to the designed censoring rule. When the measurement falls in the send region, *i.e.*, it is informative enough, the i^{th} sensor sends it to the FC. Otherwise, it is censored and not sent. For the Bayesian sequential estimation problem, we design the following measurement censoring rule based on the normalized innovation squared (NIS) [16]:

$$\nu_k^{iT} s_k^{i-1} \nu_k^i \begin{cases} \geq \eta_k, & \text{send} \\ < \eta_k, & \text{not send} \end{cases} \quad (3.4)$$

where $\nu_k^i = z_k^i - \mathbf{h}^{iT} \hat{\mathbf{x}}_{k|k-1}^i$ is the innovation [16] of the i^{th} sensor at time k , s_k^i is the variance of ν_k^i , given by $s_k^i = r^i + \mathbf{h}^{iT} \mathbf{P}_{k|k-1}^i \mathbf{h}^i$ in the KF update procedure [16] ($\mathbf{P}_{k|k-1}^i$ is the covariance of the state prediction at the i^{th} sensor), and η_k is a certain threshold that is designed based on performance requirements or bandwidth constraints. Hence, the censoring rule given by (3.4) implicitly requires that the i^{th} (for $i = 1, \dots, N$) sensor should perform a KF covariance update at each time, in order to compute the variance of its innovation. Note that Eq. (3.4) is a reasonable way to select informative measurements. Here is an intuitive justification: a larger magnitude of ν_k^i means a larger difference between the true measurement and the predicted one, which indicates that the prediction based only on the model (prior information) is not accurate enough and the corresponding measurement is

needed to improve the estimation performance.

At time k , the complete measurement vector is $\mathbf{z}_k^{1:N} \triangleq (z_k^1, z_k^2, \dots, z_k^N) \triangleq (\mathbf{z}_k^{obs}, \mathbf{z}_k^{mis})$, where \mathbf{z}_k^{obs} denotes the received measurements at the FC and \mathbf{z}_k^{mis} denotes the missing ones. For the NMAR problem induced by (3.4), we define a missing-data indicator vector $\mathbf{m}_k = (m_k^1, m_k^2, \dots, m_k^N)$ for $\mathbf{z}_k^{1:N}$, where m_k^i is the indicator variable for the i^{th} sensor, which takes value 1 if the measurement is sent to the FC and 0 otherwise. That is,

$$m_k^i = \begin{cases} 1, & \text{if sensor } i \text{ sends } z_k^i \text{ to FC at time } k; \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

Under the assumption that the channels between the local sensors and the FC are perfect, a missing sensor measurement means that it has been censored by the corresponding sensor node. Hence, \mathbf{M}_k , which contains the information on missingness, is available at the FC, and the actual observed data at the FC consist of $(\mathbf{z}_k^{obs}, \mathbf{m}_k)$. In order to exploit the information conveyed by the missing data, the corresponding likelihood function of the underlying state of the system, which is denoted as $p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k)$ should be computed by the FC, and how to compute it will be considered in Section 3.3.2.

3.2.2 Particle Filter at the FC

In the proposed CFwMD scheme, a PF is employed at the FC. The KF is known to provide the optimal solution to the Bayesian sequential estimation problem when the system is linear and Gaussian. An EKF can provide suboptimal estimation by linearizing the nonlinear state dynamics and/or nonlinear measurement equation locally in nonlinear systems. However, even for linear and Gaussian systems, when the sensor measurements are quantized, the EKF does not perform very well [61]. The censoring process defined in (3.4) can be treated as a special case of measurement quantization, since if the measurement falls in the send region,

a continuous value is sent; otherwise, no data are sent, which is equivalent to a quantization of the sensor data to the symbol “0”. Hence, the PF is a reasonable choice at the FC for Bayesian sequential estimation.

As we know, the main idea of the PF is to represent the posterior distribution $p(\mathbf{x}_k|\mathbf{Z}_{1:k})$ by a set of particles $\{\mathbf{x}_k^l\}$ with associated weights $\{w_k^l\}$. Let N_p denote the total number of particles used in the PF. The posterior distribution can be then approximated as [62]

$$p(\mathbf{x}_k|\mathbf{Z}_{1:k}) \approx \sum_{l=1}^{N_p} w_k^l \delta(\mathbf{x}_k - \mathbf{x}_k^l) \quad (3.6)$$

The missing data information can be exploited by using the full likelihood $p(\mathbf{z}_k^{obs}, \mathbf{m}_k|\mathbf{x}_k)$ instead of the simple likelihood $p(\mathbf{z}_k^{obs}|\mathbf{x}_k)$ to update the weights of particles at time k . Hence, in the CFwMD scheme, after the FC has received all the measurements sent by local sensors at time k , it computes the full likelihood and uses it to update the particle weights.

3.2.3 Censoring Threshold Design

The threshold η_k in (3.4) is designed such that on an average, l sensors send their measurements to the FC at time k . Thus, we have

$$E \left[\sum_{i=1}^N m_k^i \right] = \sum_{i=1}^N E(m_k^i) = l$$

where

$$\begin{aligned} E(m_k^i) &= p(m_k^i = 1) \\ &\stackrel{(a)}{=} p(\nu_k^i T s_k^{i-1} \nu_k^i \geq \eta_k) \end{aligned} \quad (3.7)$$

where (a) is due to the definition of m_k^i in (3.5).

Since $\nu_k^i \sim \mathcal{N}(0, s_k^i)$, we have $\nu_k^i T s_k^{i-1} \nu_k^i \sim \chi_{n_{\nu_k^i}}^2$, the chi-square distribution with degree of freedom $n_{\nu_k^i}$, and $n_{\nu_k^i}$ is the dimension of the innovation ν_k^i . Since scalar observations are obtained at local sensors, their innovations have the same dimension n_ν , which is equal to

1. Hence, $\sum_{i=1}^N E[m_k^i] = Np(\nu_k^{iT} s_k^{i-1} \nu_k^i \geq \eta_k) = l$, which implies $p(\nu_k^{iT} s_k^{i-1} \nu_k^i \geq \eta_k) = l/N$. Then, we can obtain $\eta_k = \chi_{n_\nu}^2(l/N)$, where $\chi_{n_\nu}^2(l/N)$ represents the critical value such that the probability greater than it is equal to l/N . Note that η_k completely depends on the rate of transmission l/N at time k and the dimension of the innovation n_ν . Hence, once l is set to be the same value for each time k , η_k remains constant over the entire duration of tracking, and it can be computed offline and independently by local sensors and the FC without extra transmission, *i.e.*,

$$\eta = \chi_{n_\nu}^2(l/N) \quad (3.8)$$

3.3 Censoring and Fusion with Missing Data

3.3.1 Overview

The proposed CFwMD scheme consists of two major procedures: censoring and fusion, the former is executed at each local sensor while the latter is executed at the FC. At the initial step, local sensors and the FC compute η independently according to (3.8). Then, at any given time k , each local sensor updates the covariance of its innovation S_k^i following the covariance update of the standard KF, and then determines whether its measurement at the current time is informative enough or not by the proposed innovation based censoring rule (3.4). Only if the measurement is informative, it is sent to the FC. At the FC, after it gathers all the informative measurements from the local sensors, it fuses them to infer the target state. In this paper, it is assumed that the delays in transmitting sensor measurements to the FC are all less than the sampling interval of the sensors, so that the FC can fuse the arriving measurements in time. We also assume that the FC knows the censoring rule. Since the channels in the system have been assumed to be perfect, the only cause of a missing measurement is that it is not informative enough. Then, based on the two assumptions

above, the FC can compute the statistics of the missing measurements, which we propose to incorporate in the fusion procedure for better inference performance. Note that the FC maintains a particle filter to track the target. In order to fuse both the received measurements and missing ones, we propose to use the full likelihood function, the details of which will be given in the following section, to update the particle weights.

To make the CFwMD scheme more clear to the readers, we describe one cycle of the scheme in the following algorithm:

Algorithm 1 The CFwMD scheme

Initial step: Design η by (3.8)

At time k :

At the i^{th} local sensor, ($i = 1, \dots, N$):

$$(A1.1) \quad s_k^i = r^i + \mathbf{h}^{iT} \mathbf{P}_{k|k-1}^i \mathbf{h}^i \text{ (KF update)}$$

(A1.2) Apply the censoring rule (3.4) to measurement z_k^i

At the FC: (PF with N_s particles, $l = 1, \dots, N_s$)

$$(A1.3) \quad \mathbf{x}_k^l = \mathbf{F}_{k-1} \mathbf{x}_{k-1}^l + \mathbf{u}_k^l \text{ (Propagating particles)}$$

$$(A1.4) \quad w_k^l \propto \text{full likelihood function}$$

(A1.5) Normalize weights and estimate the state

$$\text{by } \{\mathbf{x}_k^l, w_k^l\}$$

(A1.6) Resampling to get $\{\mathbf{x}_k^l, N_s^{-1}\}$

3.3.2 The Full Likelihood Function

One of the critical elements of our CFwMD scheme is the full likelihood function which includes the missing data information according to the previous section. In this section, we derive the full likelihood function at time k for two cases, *i.e.*, for a feedback system as well as for a non-feedback system, depending on whether the state prediction $\hat{\mathbf{x}}_{k|k-1}$ is a global

one or a local one.

3.3.2.1 Feedback System

The system is called a feedback system when at the beginning of time k , certain global information, such as prediction of the target state $\hat{\mathbf{x}}_{k|k-1}$, is broadcast to the local sensors by the FC.

Proposition 3.1. *For the linear Gaussian system (3.1) with measurement model (3.3), if censoring strategy (3.4) is used and the state prediction $\hat{\mathbf{x}}_{k|k-1}$ is fed back from the FC to the local sensors, then the full likelihood of the system state at time k , which is used to update the weights of particles at the FC at step (A1.4) in Algorithm 1 of the CFwMD scheme is given as*

$$\begin{aligned} & p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\ &= \prod_{i=1}^N [p(z_k^i | \mathbf{x}_k)]^{m_k^i} [Q(\xi_{k,1}^i) - Q(\xi_{k,2}^i)]^{1-m_k^i} \end{aligned} \quad (3.9)$$

where $Q(\cdot)$ is the complementary cumulative distribution function of a normal random variable with zero mean and unit variance, $\xi_{k,1}^i \triangleq \frac{-\sqrt{\eta s_k^i - \mu_k^i}}{\sqrt{r^i}}$, $\xi_{k,2}^i \triangleq \frac{\sqrt{\eta s_k^i - \mu_k^i}}{\sqrt{r^i}}$, $\mu_k^i = \mathbf{h}^{iT}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$, the conditional mean of i^{th} sensor's innovation, and m_k^i is defined in (3.5).

Proof. At time k , given $\hat{\mathbf{x}}_{k|k-1}$, the full likelihood function is $p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1})$. Let N_{obs} denote the number of received observations, and N_{mis} denote the number of missing

observations, then

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= \int p(\mathbf{m}_k, \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \int p(\mathbf{m}_k | \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis}, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \cdot \\
& \quad p(\mathbf{z}_k^{obs}, \mathbf{z}_k^{mis} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \int \prod_{i=1}^N [p(m_k^i | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) p(z_k^i | \mathbf{x}_k)] d\mathbf{z}_k^{mis} \tag{3.10}
\end{aligned}$$

The last line in (3.10) is due to the fact that local sensor observations are conditionally independent.

By decomposing the product inside the integral in (3.10) into two parts: one related to the received observations, and the other related to the missing observations, we can obtain

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= \prod_{i=1}^{N_{obs}} [p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) p(z_k^i | \mathbf{x}_k)] \cdot \\
& \quad \int \prod_{j=1}^{N_{mis}} p(m_k^j, z_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \prod_{i=1}^{N_{obs}} [p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) p(z_k^i | \mathbf{x}_k)] \cdot \\
& \quad \prod_{j=1}^{N_{mis}} p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \tag{3.11}
\end{aligned}$$

Obviously, $p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) = 1$, and

$$\begin{aligned}
& p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= p\left(\nu_k^j T s_k^{j-1} \nu_k^j < \eta \mid \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \\
&\stackrel{(b)}{=} p\left(|\nu_k^j| < \sqrt{\eta s_k^j} \mid \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}\right) \tag{3.12}
\end{aligned}$$

where (b) is due to the fact that scalar observations are obtained at local sensors. Given \mathbf{x}_k and $\hat{\mathbf{x}}_{k|k-1}$, ν_k^j is Gaussian with mean

$$\begin{aligned}
& E [\nu_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= E \left[\mathbf{h}^{jT} \mathbf{x}_k + n_k^j - \mathbf{h}^{jT} \hat{\mathbf{x}}_{k|k-1} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1} \right] \\
&= \mathbf{h}^{jT} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\
&\triangleq \mu_k^j
\end{aligned} \tag{3.13}$$

and covariance

$$Var [\nu_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] = r^j$$

Hence,

$$\begin{aligned}
& p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= Q \left(\frac{-\sqrt{\eta s_k^j} - \mu_k^j}{\sqrt{r^j}} \right) - Q \left(\frac{\sqrt{\eta s_k^j} - \mu_k^j}{\sqrt{r^j}} \right)
\end{aligned} \tag{3.14}$$

where η is given by (3.8). Thus, we can obtain (3.9) by plugging (3.14) in (3.11). ■

Remark 1: (I) We assume that the FC knows each local sensor's measurement model and it maintains a KF covariance update for each local sensor, and, therefore, the full likelihood given by (3.9) is completely computable at the FC without any extra transmissions from the local sensors. (II) It is not necessary for each sensor to run a complete KF, including the state update and the covariance update. But, at each sensor, the KF covariance update recursion is still needed to calculate its innovation covariance s_k^i , which is required to censor its measurement. (III) The threshold η is designed by assuming that local state predictions are employed to calculate the innovations, but in the feedback system, the innovations are obtained by using the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ fed back by the FC. This implies that the communication rate constraint specified in (3.7) may not be strictly satisfied in a feedback

system, which can be understood by checking the definition of innovation and its covariance right below (3.4). One can see that, in a feedback system, since the innovation is computed by the global $\hat{\mathbf{x}}_{k|k-1}$ instead of the local estimate $\hat{\mathbf{x}}_{k|k-1}^i$, it is not strictly Gaussian with covariance s_k^i which is still computed by using local $\mathbf{P}_{k|k-1}^i$. Therefore, (3.7) is not strictly true which indicates that the communication rate constraint is not strictly satisfied. Nevertheless, if the FC also feeds back $\hat{\mathbf{P}}_{k|k-1}$ which is an empirical estimate by the PF, then the bandwidth constraint can be more strictly satisfied with the cost of extra transmission, which gives us Proposition 3.2.

Proposition 3.2. *For the linear Gaussian system (3.1) with measurement model (3.3), if censoring strategy (3.4) is used and the state prediction $\hat{\mathbf{x}}_{k|k-1}$ and the related covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back from the FC to the local sensors, then the full likelihood of the system state at time k is given as*

$$\begin{aligned} & p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\ &= \prod_{i=1}^N [p(z_k^i | \mathbf{x}_k)]^{m_k^i} [Q(\xi_{k,1}^i) - Q(\xi_{k,2}^i)]^{1-m_k^i} \end{aligned} \quad (3.15)$$

where $\xi_{k,1}^i \triangleq \frac{-\sqrt{\eta s_k^i} - \mu_k^i}{\sqrt{r^i}}$, $\xi_{k,2}^i \triangleq \frac{\sqrt{\eta s_k^i} - \mu_k^i}{\sqrt{r^i}}$, $\mu_k^i = \mathbf{h}^{iT}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$, the conditional mean of i^{th} sensor's innovation, and m_k^i is defined in (3.5).

Proof. At time k , given $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{P}}_{k|k-1}$, the full likelihood function is $p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1})$.

Let N_{obs} denote the number of received observations, and N_{mis} denote the number of missing

observations, then

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\
&= \int p(\mathbf{m}_k, \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \int p(\mathbf{m}_k | \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis}, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \cdot \\
& \quad p(\mathbf{z}_k^{obs}, \mathbf{z}_k^{mis} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \int \prod_{i=1}^N \left[p(m_k^i | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) p(z_k^i | \mathbf{x}_k) \right] d\mathbf{z}_k^{mis} \tag{3.16}
\end{aligned}$$

The last line in (3.16) is due to the fact that local sensor observations are conditionally independent.

By decomposing the product inside the integral in (3.16) into two parts: one related to the received observations, and the other related to the missing observations, we can obtain

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\
&= \prod_{i=1}^{N_{obs}} \left[p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) p(z_k^i | \mathbf{x}_k) \right] \cdot \\
& \quad \int \prod_{j=1}^{N_{mis}} p(m_k^j, z_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) d\mathbf{z}_k^{mis} \\
&= \prod_{i=1}^{N_{obs}} \left[p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) p(z_k^i | \mathbf{x}_k) \right] \cdot \\
& \quad \prod_{j=1}^{N_{mis}} p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \tag{3.17}
\end{aligned}$$

Obviously, $p(m_k^i = 1 | z_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) = 1$, and

$$\begin{aligned}
& p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\
&= p \left(\nu_k^{jT} S_k'^{j-1} \nu_k^j < \eta \mid \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1} \right) \\
&\stackrel{(b)}{=} p \left(|\nu_k^j| < \sqrt{\eta S_k'^j} \mid \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1} \right) \tag{3.18}
\end{aligned}$$

where (b) is due to the fact that scalar observations are obtained at local sensors. Given \mathbf{x}_k and $\hat{\mathbf{x}}_{k|k-1}$, ν_k^j is Gaussian with mean

$$\begin{aligned}
& E [\nu_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= E \left[\mathbf{h}^{jT} \mathbf{x}_k + n_k^j - \mathbf{h}^{jT} \hat{\mathbf{x}}_{k|k-1} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1} \right] \\
&= \mathbf{h}^{jT} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\
&\triangleq \mu_k^j
\end{aligned} \tag{3.19}$$

and covariance

$$Var [\nu_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] = r^j$$

Hence,

$$\begin{aligned}
& p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= Q \left(\frac{-\sqrt{\eta} s_k^{lj} - \mu_k^j}{\sqrt{r^j}} \right) - Q \left(\frac{\sqrt{\eta} s_k^{lj} - \mu_k^j}{\sqrt{r^j}} \right)
\end{aligned} \tag{3.20}$$

where η is given by (3.8). Thus, we can obtain (3.15) by plugging (3.20) in (3.17). ■

Remark 2: (I) The superscript ‘ ν ’ in Proposition 3.2 indicates that the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$ instead of the local $\mathbf{P}_{k|k-1}^i$ is involved in the computation of the covariance of the innovation. (II) Since the global $\hat{\mathbf{P}}_{k|k-1}$ in the Proposition is an empirical estimate, Eqs. (3.18) and (3.20) involved in the proof are approximate ones.

One should keep in mind that, for the feedback system, a feedback step should be added at the beginning of the CFwMD scheme given in Algorithm 1. If only the state prediction is fed back, the remaining parts remain unchanged; if both the state prediction and related covariance are fed back, $\mathbf{P}_{k|k-1}^i$ at step (A1.1) should be replaced by the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$. We do not repeat the algorithm here for brevity.

3.3.2.2 Non-feedback System

In a non-feedback system, local sensors censor their measurements according to (3.4) using the innovations computed by their own system state prediction, which implies that each local sensor needs to run a KF. The full likelihood in the non-feedback system is derived and given as follows.

Proposition 3.3. *For the linear Gaussian system (3.1) with measurement model (3.3), if censoring strategy (3.4) is used, then the full likelihood of the target state at time k is given as*

$$p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) = \int p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) \cdot \prod_{i=1}^N [p(z_k^i | \mathbf{x}_k)]^{m_k^i} [Q(\tilde{\xi}_{k,1}^i) - Q(\tilde{\xi}_{k,2}^i)]^{1-m_k^i} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \quad (3.21)$$

where $\tilde{\xi}_{k,1}^i \triangleq \frac{-\sqrt{\eta s_k^i - \tilde{\mu}_k^i}}{\sqrt{r^i}}$, $\tilde{\xi}_{k,2}^i \triangleq \frac{\sqrt{\eta s_k^i - \tilde{\mu}_k^i}}{\sqrt{r^i}}$, and $\tilde{\mu}_k^i = \mathbf{h}^{iT} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)$, which is the conditional mean of i^{th} sensor's innovation. m_k^i is defined in (3.5) and $p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) = p(\hat{\mathbf{x}}_{k|k-1}^1, \dots, \hat{\mathbf{x}}_{k|k-1}^N | \mathbf{x}_k)$ is the joint PDF of the local sensor state predictions given the current true state, which will be given later in this section.

Proof. Let $\hat{\mathbf{x}}_{k|k-1}^{1:N} \triangleq (\hat{\mathbf{x}}_{k|k-1}^1, \dots, \hat{\mathbf{x}}_{k|k-1}^N)$ denote the local sensors' state predictions.

$$\begin{aligned} & p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) \\ &= \int \int p(\mathbf{m}_k, \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis}, \hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\mathbf{z}_k^{mis} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \\ &= \int \int p(\mathbf{m}_k | \mathbf{z}_k^{obs}, \mathbf{z}_k^{mis}, \hat{\mathbf{x}}_{k|k-1}^{1:N}, \mathbf{x}_k) \\ & \quad \cdot p(\mathbf{z}_k^{obs}, \mathbf{z}_k^{mis} | \hat{\mathbf{x}}_{k|k-1}^{1:N}, \mathbf{x}_k) p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\mathbf{z}_k^{mis} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \\ &= \int \int \prod_{i=1}^N \{p(m_k^i | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k) p(z_k^i | \mathbf{x}_k)\} \\ & \quad \cdot p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\mathbf{z}_k^{mis} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \end{aligned} \quad (3.22)$$

Similar to the feedback case, we split observed data and missing data in the inner integral, then,

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) \\
&= \int \prod_{i=1}^{N_{obs}} [p(z_k^i | \mathbf{x}_k) p(m_k^i = 1 | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k)] \cdot \\
& \int \prod_{j=1}^{N_{mis}} [p(m_k^j, z_k^j | \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{x}_k)] p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\mathbf{z}_k^{mis} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \\
&= \int \prod_{i=1}^{N_{obs}} [p(z_k^i | \mathbf{x}_k) p(m_k^i = 1 | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k)] \cdot \\
& \prod_{j=1}^{N_{mis}} [p(m_k^j = 0 | \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{x}_k)] p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\hat{\mathbf{x}}_{k|k-1}^{1:N} \tag{3.23}
\end{aligned}$$

Again, we have $p(m_k^i = 1 | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k) = 1$ in (3.23). Now, we compute $p(m_k^j = 0 | \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{x}_k)$ in (3.23) by following a similar procedure as for the feedback system:

$$\begin{aligned}
& E \left[\nu(z_k^j, \hat{\mathbf{x}}_{k|k-1}^j) | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j \right] \\
&= \mathbf{h}^{jT} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j) \\
&\triangleq \tilde{\mu}_k^j \tag{3.24}
\end{aligned}$$

and

$$Var \left[\nu(z_k^j, \hat{\mathbf{x}}_{k|k-1}^j) | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j \right] = r^j$$

Thus,

$$\begin{aligned}
& p(m_k^j | \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{x}_k) \\
&= Q \left(\frac{-\sqrt{\eta s_k^j} - \tilde{\mu}_k^j}{\sqrt{r^j}} \right) - Q \left(\frac{\sqrt{\eta s_k^j} - \tilde{\mu}_k^j}{\sqrt{r^j}} \right) \tag{3.25}
\end{aligned}$$

Hence, we can obtain (3.21) by using (3.25) in (3.23). ■

Note that the joint PDF $p(\hat{\mathbf{x}}_{k|k-1}^1, \dots, \hat{\mathbf{x}}_{k|k-1}^N | \mathbf{x}_k)$ is a multivariate normal distribution with mean $\boldsymbol{\pi}_{k|k-1}$ and covariance $\boldsymbol{\Sigma}_{k|k-1}$, where $\boldsymbol{\pi}_{k|k-1} = \tilde{\mathbf{H}} \mathbf{x}_k$, $\tilde{\mathbf{H}}$ is given by $\tilde{\mathbf{H}} = [\mathbf{I}_d \dots \mathbf{I}_d]^T$

with dimension $Nd \times d$, and \mathbf{I}_d denotes the $d \times d$ identity matrix. That is, the mean $\boldsymbol{\pi}_{k|k-1}$ is the concatenation by N true states \mathbf{x}_k . The N diagonal elements of the covariance $\boldsymbol{\Sigma}_{k|k-1}$ are filled with $\mathbf{P}_{k|k-1}^i$, the covariance of each sensor's own prediction, and the remaining terms of $\boldsymbol{\Sigma}_{k|k-1}$ are filled with $\tilde{\mathbf{P}}_{k|k-1}^{i,j}$, cross-covariance between the i^{th} sensor's prediction and the j^{th} sensor's prediction. Thus,

$$\boldsymbol{\Sigma}_{k|k-1} = \begin{bmatrix} \mathbf{P}_{k|k-1}^1 & \tilde{\mathbf{P}}_{k|k-1}^{1,2} & \cdots & \tilde{\mathbf{P}}_{k|k-1}^{1,N} \\ \tilde{\mathbf{P}}_{k|k-1}^{2,1} & \mathbf{P}_{k|k-1}^2 & \cdots & \tilde{\mathbf{P}}_{k|k-1}^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{P}}_{k|k-1}^{N,1} & \tilde{\mathbf{P}}_{k|k-1}^{N,2} & \cdots & \mathbf{P}_{k|k-1}^N \end{bmatrix} \quad (3.26)$$

For two arbitrary sensors i, j :

$$\begin{aligned} \tilde{\mathbf{P}}_{k|k-1}^{i,j} &= E \left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j)^T \right] \\ &= E \left[\mathbf{F}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i)(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^j)^T \mathbf{F}^T \right] \\ &\quad + E \left[\mathbf{u}_{k-1} \mathbf{u}_{k-1}^T \right] + E \left[\mathbf{F}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^i) \mathbf{u}_{k-1}^T \right] \\ &\quad + E \left[\mathbf{u}_{k-1} (\mathbf{F}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^j))^T \right] \\ &= \mathbf{F} \mathbf{P}_{k-1|k-1}^{i,j} \mathbf{F}^T + \mathbf{Q}_{k-1} \end{aligned} \quad (3.27)$$

where according to [21]

$$\begin{aligned} \mathbf{P}_{k-1|k-1}^{i,j} &= \\ &\left[I - \mathbf{w}_{k-1}^i \mathbf{h}^{iT} \right] \left[\mathbf{F} \mathbf{P}_{k-2|k-2}^{i,j} \mathbf{F}^T + \mathbf{Q}_{k-2} \right] \left[I - \mathbf{w}_{k-1}^j \mathbf{h}^{jT} \right]^T \end{aligned} \quad (3.28)$$

and \mathbf{w}_{k-1}^i is the Kalman gain at time $k-1$.

Note that Eq. (3.28) is recursive, and once the initialization $\mathbf{P}_{0|0}^{i,j} = \mathbf{0}$ is given, $\mathbf{P}_{k|k}^{i,j}$ at any given time step k can be computed recursively, based on which (3.26) can be evaluated.

We should point out that, for the non-feedback system, a KF state update should be added to step (A1.1) in Algorithm 1, but the remaining steps are kept the same.

It should be noted that the CFoMD follows the same procedure as the CFwMD, except that the full likelihood is replaced by the simple likelihood, *i.e.*, $p(\mathbf{z}_k^{obs}|\mathbf{x}_k)$ in step (A1.4) of Algorithm 1.

3.4 Censoring Based on an Information Theoretic Metric

In the previous sections, we proposed to use innovations in the censoring rule to select informative measurements. Though we have given an intuitive motivation for this choice, one may wonder about its optimality. In this section, we use an information theory based metric to measure the informativeness of measurements. A good metric which can measure whether or not a measurement z_k is informative enough is the KL divergence between the prior distribution $p(\mathbf{x}_k|z_{1:k-1})$ before the measurement is available and the posterior distribution $p(\mathbf{x}_k|z_{1:k})$ after the measurement is obtained. The censoring rule based on KL divergence can be expressed as

$$D_{KL}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k})) \begin{cases} \geq \zeta_k & \text{send} \\ < \zeta_k & \text{not send} \end{cases} \quad (3.29)$$

where $D_{KL}(\cdot||\cdot)$ denotes the distance between two distributions in terms of KL divergence, which is defined as

$$D_{KL}(p(y)||q(y)) = \int p(y) \ln \frac{p(y)}{q(y)} dy$$

for distributions p and q of the continuous random variable y .

We show that under certain conditions, the proposed innovation based censoring rule is equivalent to that based on the KL divergence.

Theorem 3.4. For the linear Gaussian system (3.1) with scalar measurement model (3.3), the censoring rule based on the metric $D_{KL}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k}))$ in (3.29) is equivalent to the one based on the NIS ν_k^2/s_k .

Proof. For a linear Gaussian system, we have $p(\mathbf{x}_k|z_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1})$, and $p(\mathbf{x}_k|z_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$. Then, according to [63]

$$\begin{aligned} & D_{KL}(p(\mathbf{x}_k|z_{1:k-1})||p(\mathbf{x}_k|z_{1:k})) \\ &= \frac{1}{2} \left\{ \text{tr}(\mathbf{P}_{k|k}^{-1}\mathbf{P}_{k|k-1}) - \ln \frac{|\mathbf{P}_{k|k-1}|}{|\mathbf{P}_{k|k}|} - d \right\} \\ &+ \frac{1}{2} (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{P}_{k|k}^{-1} (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) \end{aligned} \quad (3.30)$$

Since $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ are determined offline for a linear Gaussian system, and d in (3.30) is the dimension of the state \mathbf{x}_k , they are all deterministic once the system is determined. Therefore, (3.29) is equivalent to

$$(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{P}_{k|k}^{-1} (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) \begin{cases} \geq \gamma_k & \text{send} \\ < \gamma_k & \text{not send} \end{cases} \quad (3.31)$$

where $\gamma_k \triangleq 2\zeta_k + d + \ln \frac{|\mathbf{P}_{k|k-1}|}{|\mathbf{P}_{k|k}|} - \text{tr}(\mathbf{P}_{k|k}^{-1}\mathbf{P}_{k|k-1})$. Note that the censoring is performed at each local sensor which maintains a KF. Thus,

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{w}_k (z_k - \mathbf{h}^T \hat{\mathbf{x}}_{k|k-1}) \\ &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{w}_k \nu_k \end{aligned} \quad (3.32)$$

where \mathbf{w}_k is the KF gain, which is a column vector if scalar measurements are obtained. Then,

$$(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) = \mathbf{w}_k \nu_k \quad (3.33)$$

Thus, (3.31) is equivalent to

$$\nu_k^T \mathbf{w}_k^T \mathbf{P}_{k|k}^{-1} \mathbf{w}_k \nu_k \begin{cases} \geq \gamma_k & \text{send} \\ < \gamma_k & \text{not send} \end{cases} \quad (3.34)$$

When scalar measurements are obtained, both $\mathbf{w}_k^T \mathbf{P}_{k|k}^{-1} \mathbf{w}_k$ and s_k are scalars. Hence, by comparing (3.34) to (3.4), we conclude that they are equivalent when appropriate thresholds are selected. ■

Theorem 3.4 indicates that the innovation based censoring rule selects more informative measurements to send, which is intuitively pleasing. The above result can be easily extended to symmetric KL divergence.

Corollary 3.5. *For the linear Gaussian system (3.1) with scalar measurement model (3.3), the censoring rule based on the symmetric KL divergence*

$$\begin{aligned} & D_{KL}(p(\mathbf{x}_k | z_{1:k-1}) || p(\mathbf{x}_k | z_{1:k})) \\ & + D_{KL}(p(\mathbf{x}_k | z_{1:k}) || p(\mathbf{x}_k | z_{1:k-1})) \end{aligned} \quad (3.35)$$

is equivalent to that based on the NIS ν_k^2/s_k .

Proof. When symmetric KL divergence is used, the metric to select more informative data in (3.31) is changed to

$$\Lambda_k \triangleq (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1})^T (P_{k|k}^{-1} + P_{k|k-1}^{-1}) (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}) \quad (3.36)$$

Following the same manipulation on Λ_k as in the proof of Theorem 3.4, we can obtain

$$\Lambda_k = \nu_k^T \mathbf{w}_k^T (P_{k|k}^{-1} + P_{k|k-1}^{-1}) \mathbf{w}_k \nu_k \quad (3.37)$$

Again, since scalar measurements are obtained, $\mathbf{w}_k^T (P_{k|k}^{-1} + P_{k|k-1}^{-1}) \mathbf{w}_k$ is a scalar, so is s_k .

Therefore, we have

$$\mathbf{w}_k^T (P_{k|k}^{-1} + P_{k|k-1}^{-1}) \mathbf{w}_k \propto s_k^{-1} \quad (3.38)$$

■

3.5 The Vector Observation Case

So far, our discussion was limited to the scalar observation case. When vector observations are obtained at the local sensors, *i.e.*, the measurement model (3.2) is used, we still propose to use NIS based censoring rule, *i.e.*,

$$\boldsymbol{\nu}_k^i T \mathbf{S}_k^{i-1} \boldsymbol{\nu}_k^i \begin{cases} \geq \tilde{\eta}_k, & \text{send} \\ < \tilde{\eta}_k, & \text{not send} \end{cases} \quad (3.39)$$

Again, we use m_k^i as the indicator variable for the i^{th} sensor, which takes the value 1 if the vector measurement of sensor i is sent to the FC and 0 otherwise.

As in the scalar measurement case, we design $\tilde{\eta}_k$ such that, at time k , there are only l sensors that are active. Without loss of generality, we assume that local sensors' innovations have the same dimension n_ν . If l is set to be the same value at every time and the dimension of the innovation n_ν remains unchanged over time, *i.e.*, the measurement model (3.2) remains unchanged, then we still have

$$\tilde{\eta}_k = \tilde{\eta} = \chi_{n_\nu}^2(l/N) \quad (3.40)$$

According to the discussion above, Algorithm 1 can be straightforwardly applied to the vector observation case by replacing $s_k^i = r^i + \mathbf{h}^{iT} \mathbf{P}_{k|k-1}^i \mathbf{h}^i$ by $\mathbf{S}_k^i = \mathbf{R}^i + \mathbf{H}^i \mathbf{P}_{k|k-1}^i \mathbf{H}^{iT}$. Then, the main concern now is to compute the corresponding full likelihood for the vector observation case which is discussed in the following sub-sections.

3.5.1 Feedback System

Proposition 3.6. *For the linear Gaussian system (3.1) with vector measurement (3.2), when the global state estimate feedback from the FC is available, and the censoring strategy*

(3.39) is used, the full likelihood of the system state at time k is given as

$$\begin{aligned}
& p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= \prod_{i=1}^N [p(\mathbf{z}_k^i | \mathbf{x}_k)]^{m_k^i} \left[p(q^{iT} q^i < \tilde{\eta}) \right]^{1-m_k^i}
\end{aligned} \tag{3.41}$$

where $q^i \sim \mathcal{N}(\mathbf{S}_k^i^{-\frac{1}{2}} \mathbf{H}^i (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^i^{-\frac{1}{2}} \mathbf{R}^i \mathbf{S}_k^i^{-\frac{1}{2}})$.

Proof. Following a similar procedure as in Proposition 3.1, we can obtain

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) \\
&= \prod_{i=1}^{N_{obs}} [p(m_k^i = 1 | \mathbf{z}_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) p(\mathbf{z}_k^i | \mathbf{x}_k)] \cdot \\
& \quad \prod_{j=1}^{N_{mis}} p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1})
\end{aligned} \tag{3.42}$$

where

$$p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}) = p(\boldsymbol{\nu}_k^{jT} \mathbf{S}_k^{j-1} \boldsymbol{\nu}_k^j < \tilde{\eta} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1})$$

Denoting $q^j \triangleq \mathbf{S}_k^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j$, we have

$$\begin{aligned}
E [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] &= \mathbf{S}_k^{j-\frac{1}{2}} E [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})
\end{aligned} \tag{3.43}$$

$$\begin{aligned}
& Cov [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= E [q^j q^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
& \quad - E [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] E^T [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= E \left[\mathbf{S}_k^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} \mathbf{S}_k^{j-\frac{1}{2}} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1} \right] \\
& \quad - E [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] E^T [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}]
\end{aligned} \tag{3.44}$$

Since

$$\begin{aligned}
& E \left[\boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1} \right] \\
&= Cov[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] + E[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] E^T[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] \\
&= \mathbf{R}^j + \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{H}^{jT}
\end{aligned} \tag{3.45}$$

we can obtain

$$Cov [q^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] = \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k^{j-\frac{1}{2}} \tag{3.46}$$

Therefore, $q^j \sim \mathcal{N}(\mathbf{S}_k^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k^{j-\frac{1}{2}})$. ■

Following a similar discussion as that in Remark 1 (III), we provide the following result.

Proposition 3.7. *For the linear Gaussian system (3.1) with vector measurements (3.2), when the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ and its covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back from the FC to the sensors, the full likelihood of the system state at time k is given as*

$$\begin{aligned}
& p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\
&= \prod_{i=1}^N [p(\mathbf{z}_k^i | \mathbf{x}_k)]^{m_k^i} \left[p(q^{iT} q^i < \tilde{\eta}) \right]^{1-m_k^i}
\end{aligned} \tag{3.47}$$

where $q^i \sim \mathcal{N}(\mathbf{S}_k^{i-\frac{1}{2}} \mathbf{H}^i (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{i-\frac{1}{2}} \mathbf{R}^i \mathbf{S}_k^{i-\frac{1}{2}})$, and \mathbf{S}_k^i is computed using the global state prediction covariance $\hat{\mathbf{P}}_{k|k-1}$ instead of the local one.

Proof. Following a similar procedure as in Proposition 3.1, we can obtain

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) \\
&= \prod_{i=1}^{N_{obs}} \left[p(m_k^i = 1 | \mathbf{z}_k^i, \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) p(\mathbf{z}_k^i | \mathbf{x}_k) \right] \cdot \\
& \prod_{j=1}^{N_{mis}} p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1})
\end{aligned} \tag{3.48}$$

where

$$p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}) = p(\boldsymbol{\nu}_k^{jT} \mathbf{S}_k'^{j-1} \boldsymbol{\nu}_k^j < \tilde{\eta} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1})$$

Denoting $q'^j \triangleq \mathbf{S}_k'^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j$, we have

$$\begin{aligned} E [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] &= \mathbf{S}_k'^{j-\frac{1}{2}} E [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &= \mathbf{S}_k'^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \end{aligned} \quad (3.49)$$

$$\begin{aligned} &Cov [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &= E [q'^j q'^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &\quad - E [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] E^T [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &= E [\mathbf{S}_k'^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} \mathbf{S}_k'^{j-\frac{1}{2}} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &\quad - E [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] E^T [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \end{aligned} \quad (3.50)$$

Since

$$\begin{aligned} &E [\boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &= Cov [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] + E [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] E^T [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}] \\ &= \mathbf{R}^j + \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{H}^{jT} \end{aligned} \quad (3.51)$$

we can obtain

$$Cov [q'^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}] = \mathbf{S}_k'^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k'^{j-\frac{1}{2}} \quad (3.52)$$

Therefore, $q'^j \sim \mathcal{N}(\mathbf{S}_k'^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k'^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k'^{j-\frac{1}{2}})$. ■

3.5.2 Non-feedback System

Proposition 3.8. *For the linear Gaussian system (3.1) with vector measurements, when global estimate feedback from the FC is not available, if censoring strategy (3.39) is used, then the full likelihood of the system state at time k is given as*

$$p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) = \int p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) \prod_{i=1}^N [p(\mathbf{z}_k^i | \mathbf{x}_k)]^{m_k^i} \left[p(t^{iT} t^i < \tilde{\eta}_k) \right]^{1-m_k^i} d\hat{\mathbf{x}}_{k|k-1}^{1:N} \quad (3.53)$$

where $t^i \sim \mathcal{N}(\mathbf{S}_k^{i-\frac{1}{2}} \mathbf{H}^i (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^i), \mathbf{S}_k^{i-\frac{1}{2}} \mathbf{R}^i \mathbf{S}_k^{i-\frac{1}{2}})$.

Proof. Following a similar procedure as in the proof of Proposition 3.3, we have

$$\begin{aligned} & p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) \\ &= \int \prod_{i=1}^{N_{obs}} [p(z_k^i | \mathbf{x}_k) p(m_k^i = 1 | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k)] \cdot \\ & \quad \prod_{j=1}^{N_{mis}} [p(m_k^j = 0 | \hat{\mathbf{x}}_{k|k-1}^j, \mathbf{x}_k)] p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) d\hat{\mathbf{x}}_{k|k-1}^{1:N} \end{aligned} \quad (3.54)$$

where $p(m_k^i = 1 | z_k^i, \hat{\mathbf{x}}_{k|k-1}^i, \mathbf{x}_k) = 1$ and

$$p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j) = p(\boldsymbol{\nu}_k^{jT} \mathbf{S}_k^{j-1} \boldsymbol{\nu}_k^j < \tilde{\eta} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j)$$

Denoting $t^j \triangleq \mathbf{S}_k^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j$, we have

$$\begin{aligned} E [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] &= \mathbf{S}_k^{j-\frac{1}{2}} E [\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\ &= \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j) \end{aligned} \quad (3.55)$$

$$\begin{aligned} & Cov [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\ &= E [t^j t^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\ & \quad - E [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] E^T [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\ &= E [\mathbf{S}_k^{j-\frac{1}{2}} \boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} \mathbf{S}_k^{j-\frac{1}{2}} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\ & \quad - E [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] E^T [t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \end{aligned} \quad (3.56)$$

Since

$$\begin{aligned}
& E \left[\boldsymbol{\nu}_k^j \boldsymbol{\nu}_k^{jT} | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j \right] \\
&= Cov[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] + E[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] E^T[\boldsymbol{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j] \\
&= \mathbf{R}^j + \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j) (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j)^T \mathbf{H}^{jT}
\end{aligned} \tag{3.57}$$

we can obtain

$$Cov \left[t^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k|k-1}^j \right] = \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k^{j-\frac{1}{2}} \tag{3.58}$$

Therefore, $t^j \sim \mathcal{N}(\mathbf{S}_k^{j-\frac{1}{2}} \mathbf{H}^j (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^j), \mathbf{S}_k^{j-\frac{1}{2}} \mathbf{R}^j \mathbf{S}_k^{j-\frac{1}{2}})$, and

$$p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k) = \int p(\hat{\mathbf{x}}_{k|k-1}^{1:N} | \mathbf{x}_k) \prod_{i=1}^N [p(\mathbf{z}_k^i | \mathbf{x}_k)]^{m_k^i} \left[p(t^{iT} t^i < \tilde{\eta}_k) \right]^{1-m_k^i} d\hat{\mathbf{x}}_{k|k-1}^{1:N}$$

■

3.6 Censoring and Fusion with Missing Data for Non-linear Systems

In the previous sections, we have discussed the proposed CFwMD scheme for linear Gaussian systems. To make it more general, we extend the scheme to a general nonlinear system in this section. Consider the following nonlinear state-space model

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{u}_k \tag{3.59}$$

and measurement model for the i^{th} sensor

$$z_k^i = h^i(\mathbf{x}_k) + n_k^i \tag{3.60}$$

where $\mathbf{u}_k \sim \mathcal{N}(0, \mathbf{Q})$ is the state process noise, and $n_k^i \sim \mathcal{N}(0, r^i)$ is the measurement noise. We first consider the scalar observation case. Note that, due to the nonlinearity

of the system, when the CFwMD scheme is used in the considered nonlinear system, each local sensor maintains an EKF and the FC uses a particle filter to infer the target state. We should point out that the nonlinearity of the system makes it different from the linear Gaussian system in several aspects:

1) The innovation $\tilde{\nu}_k^i$ is no longer distributed as Gaussian with zero mean and variance s_k^i , but we approximate it as $\mathcal{N}(0, s_k^i)$.

2) Since $s_k^i = \mathbf{g}_k^i \mathbf{P}_{k|k-1}^i \mathbf{g}_k^{iT} + r_k^i$, where $\mathbf{g}_k^i = \frac{\partial h}{\partial \mathbf{x}}|_{\hat{\mathbf{x}}_{k|k-1}^i}$, $\mathbf{P}_{k|k-1}^i = \mathbf{F}_{k-1}^i \mathbf{P}_{k-1|k-1}^i \mathbf{F}_{k-1}^{iT} + \mathbf{Q}$, and $\mathbf{F}_{k-1}^i = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}|_{\hat{\mathbf{x}}_{k-1|k-1}^i}$, s_k^i cannot be evaluated offline as in the case of linear systems.

Inspired by the linear Gaussian system we have discussed earlier, we propose that, for a nonlinear system, the i^{th} sensor again censors its measurement based on the NIS, *i.e.*, $(\tilde{\nu}_k^i)^T s_k^{i-1} \tilde{\nu}_k^i$ at time k , where $\tilde{\nu}_k^i \triangleq z_k^i - \hat{z}_{k|k-1}^i$, and it is approximated as a Gaussian random variable with zero mean and covariance s_k^i .

The censoring threshold can also be designed by the bandwidth constraint as in linear Gaussian systems, given the approximation that $\tilde{\nu}_k^i \sim \mathcal{N}(0, s_k^i)$. Following a similar procedure as in Section 3.2.3, we have

$$\tilde{\eta} = \chi_{n_{\tilde{\nu}}}^2(l/N) \quad (3.61)$$

where $n_{\tilde{\nu}} = 1$, since scalar observations are obtained.

For the considered nonlinear system, if the global state estimate is fed back from the FC to the local sensors, the full likelihood function in the CFwMD scheme is provided in the following proposition.

Proposition 3.9. *For a general nonlinear system given by (3.59)-(3.60), if innovation based censoring strategy is used with threshold given by (3.61) and the global estimate of the state $\hat{\mathbf{x}}_{k-1|k-1}$ is fed back to the local sensors, then the full likelihood of the system state at time k*

of the CFwMD scheme is given as

$$\begin{aligned}
& p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}) \\
&= \prod_{i=1}^N [p(z_k^i | \mathbf{x}_k)]^{m_k^i} [Q(\check{\xi}_{k,1}^i) - Q(\check{\xi}_{k,2}^i)]^{1-m_k^i}
\end{aligned} \tag{3.62}$$

where $\check{\xi}_{k,1}^i \triangleq \frac{-\sqrt{\check{\eta}_{s_k^i} - \check{\mu}_k^i}}{\sqrt{r^i}}$, $\check{\xi}_{k,2}^i \triangleq \frac{\sqrt{\check{\eta}_{s_k^i} - \check{\mu}_k^i}}{\sqrt{r^i}}$, m_k^i is defined in (3.5), $\check{\mu}_k^i = h^j(\mathbf{x}_k) - h(\hat{\mathbf{x}}_{k|k-1})$, the conditional mean of i^{th} sensor's innovation, and $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1})$.

Proof. Following a procedure similar to that in Proposition 3.1, we can obtain

$$\begin{aligned}
& p(\mathbf{m}_k, \mathbf{z}_k^{obs} | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}) \\
&= \prod_{i=1}^{N_{obs}} [p(z_k^i | \mathbf{x}_k)] \cdot \prod_{j=1}^{N_{mis}} p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1})
\end{aligned} \tag{3.63}$$

where

$$\begin{aligned}
& p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}) \\
&= p\left(|\check{\nu}_k^j| < \sqrt{\eta_{s_k^j}} \mid \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}\right)
\end{aligned} \tag{3.64}$$

Given \mathbf{x}_k and $\hat{\mathbf{x}}_{k-1|k-1}$, $\check{\nu}_k^j$ is Gaussian distributed with mean

$$\begin{aligned}
& E[\check{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}] \\
&= E[h^j(\mathbf{x}_k) + n_k^j - h^j(\hat{\mathbf{x}}_{k|k-1}) | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}] \\
&= h^j(\mathbf{x}_k) - h^j(\hat{\mathbf{x}}_{k|k-1}) \triangleq \check{\mu}_k^j
\end{aligned} \tag{3.65}$$

and covariance

$$\text{var}[\check{\nu}_k^j | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}] = r^j$$

where $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1})$.

Therefore, $p(m_k^j = 0 | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}) = Q(\check{\xi}_{k,1}^j) - Q(\check{\xi}_{k,2}^j)$ with $\check{\xi}_{k,1}^j \triangleq \frac{-\sqrt{\check{\eta}_{s_k^j} - \check{\mu}_k^j}}{\sqrt{r^j}}$ and $\check{\xi}_{k,2}^j \triangleq \frac{\sqrt{\check{\eta}_{s_k^j} - \check{\mu}_k^j}}{\sqrt{r^j}}$. ■

Remark 3: (I) As in the linear Gaussian system, we assume that the FC knows each local sensor's measurement model and it performs an EKF covariance update for each local sensor. Note that an EKF is also maintained at each local sensor, and each local sensor computes the linearized state transition matrix F_k^i and measurement matrix (vector) \mathbf{g}_k^i using the global state estimate $\hat{\mathbf{x}}_{k-1|k-1}$ fed back from the FC. Also, since the local sensors use the global feedback $\hat{\mathbf{x}}_{k-1|k-1}$ in its censoring process, and the FC maintains an EKF covariance update for each local sensor, the FC is able to compute s_k^i involved in $\check{\xi}_{k,1}^i$ and $\check{\xi}_{k,2}^i$ in the proposition above, and therefore, Eq. (3.62) is completely computable by the FC without requiring extra information from local sensors. (II) In addition to the state estimate $\hat{\mathbf{x}}_{k-1|k-1}$, the FC can also feed back the covariance $\hat{\mathbf{P}}_{k-1|k-1}$ to local sensors as in the linear Gaussian system. Note that, due to the nonlinearity, $\check{\nu}_k^i$ is approximated as a Gaussian random variable with a distribution of $\mathcal{N}(0, s_k^i)$. Nevertheless, if the FC also feeds back the global covariance $\hat{\mathbf{P}}_{k-1|k-1}$, and then, $\check{\nu}_k^i$ can be approximated as $\mathcal{N}(0, s_k^i)$ (the global $\hat{\mathbf{P}}_{k-1|k-1}$ contributes to the computation of s_k^i), which is more accurate than the previous approximation.

Proposition 3.10. *For a general nonlinear system given by (3.59)-(3.60), if innovation based censoring strategy is used with the threshold given by (3.61) and both the global estimate of the state $\hat{\mathbf{x}}_{k-1|k-1}$ and the related covariance $\hat{\mathbf{P}}_{k-1|k-1}$ are fed back to local sensors, then the full likelihood of the system state at time k of the CFwMD scheme is given as*

$$\begin{aligned} & p(\mathbf{z}_k^{obs}, \mathbf{m}_k | \mathbf{x}_k, \hat{\mathbf{x}}_{k-1|k-1}, \hat{\mathbf{P}}_{k-1|k-1}) \\ &= \prod_{i=1}^N [p(z_k^i | \mathbf{x}_k)]^{m_k^i} [Q(\check{\xi}_{k,1}^i) - Q(\check{\xi}_{k,2}^i)]^{1-m_k^i} \end{aligned} \quad (3.66)$$

where $\check{\xi}_{k,1}^i \triangleq \frac{-\sqrt{\tilde{\eta}s_k^i} - \check{\mu}_k^i}{\sqrt{r^i}}$, $\check{\xi}_{k,2}^i \triangleq \frac{\sqrt{\tilde{\eta}s_k^i} - \check{\mu}_k^i}{\sqrt{r^i}}$, m_k^i is defined in (3.5), $\check{\mu}_k^i = h^j(\mathbf{x}_k) - h(\hat{\mathbf{x}}_{k|k-1})$, the conditional mean of i^{th} sensor's innovation, and $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1})$.

Proof. Following a procedure similar to that in Proposition 3.9 and the discussion in **Remark 3** (II), we can obtain (3.66) in a straightforward manner. ■

Remark 4: (I) The global $\hat{\mathbf{P}}_{k-1|k-1}$ contributes to the computation of s_k^i . (II) If vector observations are obtained by local sensors, one can follow a similar procedure as in Section 3.5 to get the corresponding full likelihood for the nonlinear system with feedback (feedback consists of state estimate with/without covariance), which will be not discussed here. (III) For the considered nonlinear system without feedback, one may expect to get a similar result as in Proposition 3.3. But, this is not true. The reason is as follows: consider the joint PDF $p(\hat{\mathbf{x}}_{k|k-1}^{1:N}|\mathbf{x}_k)$ in the nonlinear system. Let us approximate it as Gaussian with mean $\boldsymbol{\pi}_{k|k-1}$ and covariance $\check{\boldsymbol{\Sigma}}_{k|k-1}$, which has the same structure as (3.26). However, it can be easily found that the diagonal element $\mathbf{P}_{k|k-1}^i$ in $\check{\boldsymbol{\Sigma}}_{k|k-1}$ depends on the state estimate $\hat{\mathbf{x}}_{k-1|k-1}^i$, and the off-diagonal element $\tilde{\mathbf{P}}_{k|k-1}^{i,j}$ depends on the state estimate $\hat{\mathbf{x}}_{k-1|k-1}^i$ and $\hat{\mathbf{x}}_{k-1|k-1}^j$, which prevents us from obtaining a similar result to that in Proposition 3.3.

3.7 Simulation Results

In this section, we show the advantage of the proposed CFwMD scheme for both linear and nonlinear systems via simulation. For linear systems, we show that, for a certain threshold, the CFwMD scheme has less performance loss than CFoMD, while saving the same amount of communication resources compared to the all-send case. We also show that among the three schemes, *i.e.*, CFwMD, CFoMD and the random-selection method, the proposed CFwMD scheme performs the best, under the same bandwidth constraint. We explore the performance comparison for both feedback and non-feedback scenarios. For nonlinear systems, the advantage of the proposed CFwMD scheme over CFoMD and the random-selection schemes is shown by simulation when feedback is included in the system.

3.7.1 Linear System—the Scalar Observation Case

A one-dimensional target tracking system is considered in this scenario, with state vector $\mathbf{x}_k = [x_k \dot{x}_k]^T$, state transition matrix

$$\mathbf{F} = \begin{bmatrix} 1 & \mathcal{D} \\ 0 & 1 \end{bmatrix}$$

and observation matrix $\mathbf{h}^T = [1 \ 0]$, where $\mathcal{D} = 1$ second, which is the sampling interval.

Without loss of generality, in this example, we use N identical sensors to track the target which moves only along the x-axis following the white noise acceleration model. The state process noise covariance is set as

$$\mathbf{Q} = \sigma^2 \begin{bmatrix} \mathcal{D}^4/4 & \mathcal{D}^3/2 \\ \mathcal{D}^3/2 & \mathcal{D}^2 \end{bmatrix}$$

where $\sigma^2 = 4$. The measurement noise variance is set as $r^i = 1$ for $i = 1, 2$. The initial state of the target \mathbf{x}_0 is chosen to be $[0 \ 10]$. We observe the target for 20 seconds, namely, we track the target over $T_S = 20$ time steps for each Monte-Carlo trial. The number of particles used in the particle filter at the FC is $N_s = 10^3$.

3.7.1.1 Feedback System

In this example, at the beginning of each time step in a trial, the FC broadcasts the global state prediction to local sensors. We compare the RMSEs, averaged over 5000 Monte-Carlo trials at each time, for the random-selection, CFwMD, CFoMD and all-send cases. To perform the comparison under the same bandwidth constraint, we set the censoring threshold η for the CFwMD and CFoMD schemes at the value such that the average number of active sensors is $l = 1$ at any given time, and we let each sensor send its measurement to the FC with a probability $1/N$ for the random-selection scheme.

In Fig. 3.1, there are $N = 2$ sensors. Since we set the censoring threshold η to constrain the average number of active sensors as $l = 1$ at any given time, both the CFwMD and CFoMD schemes save 50% transmissions, compared to the all-send case. However, the CFoMD scheme incurs a larger performance loss compared to the proposed CFwMD scheme according to Fig. 3.1. The reason is that the missing data due to censoring process in the CFoMD and CFwMD schemes depends on the underlying values of the data, *i.e.*, is NMAR and non-negligible [58]. Ignoring the data as in the CFoMD scheme will certainly result in some information loss. For the random-selection scheme, each sensor has probability of $1/2$ to send its observation, and therefore, it also saves 50% transmissions on an average, compared to the all-send case. But, it performs the worst among the four schemes as expected, since the per-sensor censoring process in the CFwMD and CFoMD schemes select more informative data than random selection.

In Fig. 3.2, we compare the RMSEs of two feedback cases with different values of l , *i.e.*, $l = 1, 1.2, 1.5, 2, 3$, when the total number of sensors is increased to $N = 4$. The CFwMD curves in the figure correspond to the case when only the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ is fed back, while the CFwMD2 curves correspond to the case when both the global state prediction $\hat{\mathbf{x}}_{k|k-1}$ and its covariance $\hat{\mathbf{P}}_{k|k-1}$ are fed back at any given time k . We can observe that, when $l > 1$, the CFwMD2 scheme performs better than the CFwMD scheme, due to the extra feedback from the FC. However, when $l = 3$, the CFwMD2 scheme does not provide much performance improvement. This is because $l = 3$ is already large enough to give good performance, and therefore, the extra feedback does not contribute much. On the other hand, it can be observed that the performance of the CFwMD scheme is better than that of the CFwMD2 scheme when $l = 1$. The reason is as follows: when $l = 1$, the probability that at a particular time none of the sensors sends data, which is $(3/4)^4 = 0.32$, is much greater than that when $l = 2$, which is $(1/2)^4 = 0.06$. If at a certain time step, no

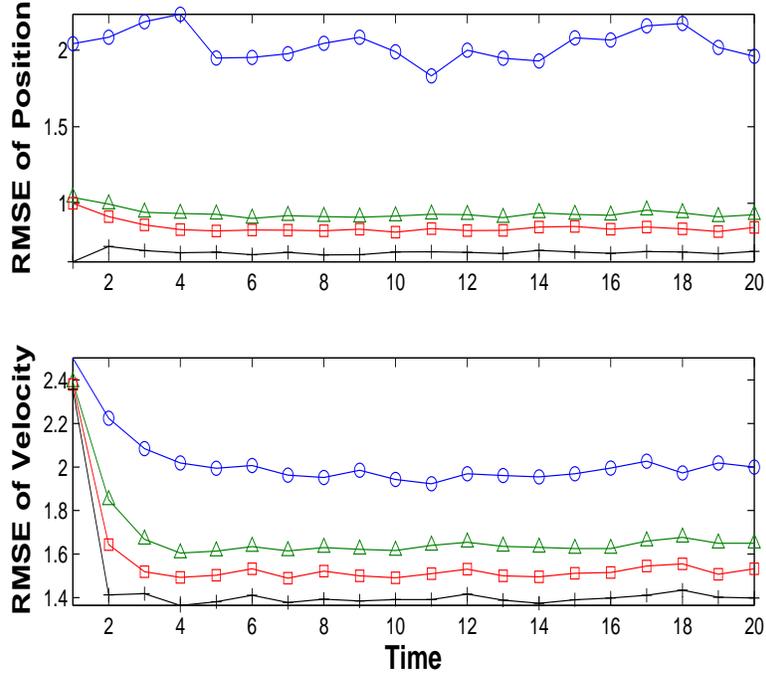


Figure 3.1: RMSE comparison for the feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

data are sent to the FC, it would be more likely that at the next time step no sensor data are sent to the FC either. This is because if no data are available for the FC to update its state estimate at time $k - 1$, both $\hat{\mathbf{P}}_{k-1|k-1}$ and $\hat{\mathbf{P}}_{k|k-1}$ will increase significantly. A larger $\hat{\mathbf{P}}_{k|k-1}$, which is fed back to local sensors in the CFwMD2 scheme, results in a larger s_k^i and makes it more difficult for the sensor data to pass the censoring rule defined in (3.4) at time k , while in the CFwMD scheme, $\mathbf{P}_{k|k-1}^i$, which completely depends on the local system model, and is not affected by the global estimation process at all. Hence, the probability that no data are sent for several consecutive time steps is much larger for the CFwMD2 scheme when $l = 1$. This has been verified by Monte-Carlo simulations, where we observed more instances of no sensor data being sent over several consecutive time steps in the case of

the CFwMD2 scheme than those in the CFwMD scheme when $l = 1$. Indeed, in Table 3.1, one can observe that, when $l = 1$, the experimental average number of transmissions of the CFwMD2 scheme is smaller than that of the CFwMD scheme. We did not observe similar phenomena for the cases when the state process noise σ is smaller or when the observation is a vector consisting of both position and velocity observations, the latter of which will be given later in the following sections. This is because $\hat{\mathbf{P}}_{k|k-1}$ is smaller in either of these two cases.

Another observation from Table 3.1 is that, for each l , the average number of transmissions of the CFwMD2 scheme is closer to the theoretical value than that of the CFwMD scheme, which verifies our expectation that the bandwidth constraint is more strictly satisfied by the CFwMD2 scheme than the CFwMD scheme.

Table 3.1: Average number of transmissions (scalar observations)

| Theoretical l | \bar{l} (CFwMD) | \bar{l} (CFwMD2) |
|-----------------|-------------------|--------------------|
| 1 | 1.0625 | 1.0530 |
| 1.2 | 1.1709 | 1.2230 |
| 1.5 | 1.3536 | 1.4992 |
| 2 | 1.7399 | 2.0002 |
| 3 | 2.7838 | 3.0095 |

3.7.1.2 Non-feedback System

For a non-feedback system, again the RMSEs of the four schemes, *i.e.*, random-selection, CFwMD, CFoMD, and all-send, are compared. In Fig. 3.3, the results for a system with $N =$

2 sensors are presented. As in the feedback system, it is obvious that CFoMD outperforms random-selection, and CFwMD performs the best among the three schemes, *i.e.*, random-selection, CFoMD and CFwMD. By observing Fig. 3.3, we can also conclude that, though the random-selection saves 50% transmissions when $N = 2$, it incurs a large loss of performance as expected.

3.7.2 Linear system—the Vector Observation Case

In this example, the same one-dimensional moving target is tracked as that in Section 3.7.1. But, the observation matrix is set as $H = \mathbf{I}_2$, an identity matrix with dimension 2×2 . Thus, both the position and the velocity of the target can be observed by local sensors. Again, $N = 2$ identical sensors are used, and the measurement covariance is set as $\mathbf{R}^i = \text{diag}[2 \ 4]$ for $i = 1, 2$. As in Section 3.7.1, we design the censoring threshold $\tilde{\eta}$ such that there is only one active sensor, *i.e.*, $l = 1$, at any given time on the average. The target is tracked for 20 seconds for each Monte-Carlo trial and 5000 Monte-Carlo trials are performed. We compare the RMSEs for the random-selection, CFwMD, CFoMD and all-send cases. In Fig. 3.4, the results for the feedback system with vector observations are presented. Obviously, similar conclusion as that in Section 3.7.1 can be drawn here.

In Fig. 3.5, as in the scalar observation case, the position and velocity RMSEs of the CFwMD scheme with only global state feedback and the CFwMD2 scheme with both the global state and covariance feedback are compared for different values of l , and the total number of sensors is again set as $N = 4$. Obviously, the CFwMD2 scheme outperforms the CFwMD scheme for each l , which is due to the extra feedback. On the other hand, the experimental average number of transmissions \bar{l} of the CFwMD2 scheme for each l , especially when $l > 1$, provided in Table 3.2 is closer to the theoretical value than that of the CFwMD scheme, which again verifies that the bandwidth constraint is more strictly satisfied by the

CFwMD2 scheme due to the feedback of the global covariance.

Table 3.2: experimental average number of transmissions (vector observations)

| Theoretical l | \bar{l} (CFwMD) | \bar{l} (CFwMD2) |
|-----------------|-------------------|--------------------|
| 1 | 0.9296 | 1.0069 |
| 2 | 1.7573 | 2.0035 |
| 3 | 2.8029 | 3.0151 |

The results for the non-feedback system with vector observations are provided in Fig. 3.6. Obviously, we can draw similar conclusions as that in Section 3.7.1.2.

We should point out that a simulation based approach has been used to compute the probability $p(q^{iT} q^i < \tilde{\eta}_k)$ to get the full likelihood function (3.41) when using the CFwMD scheme for a feedback system. That is, we first draw N_q samples from the normal distribution $\mathcal{N}(\mathbf{S}_k^{-\frac{1}{2}} \mathbf{H}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}), \mathbf{S}_k^{-\frac{1}{2}} \mathbf{R} \mathbf{S}_k^{-\frac{1}{2}})$, and then count the number of samples which satisfy the condition $q^T q < \tilde{\eta}_k$, denoted as n_q . Then, the probability can be approximated by n_q/N_q . The same approach is also used to compute the probability $p(t^{iT} t^i < \tilde{\eta}_k)$ involved in (3.53) for a non-feedback system.

3.7.3 Nonlinear System

In this experiment, we assume $N = 9$ sensors are grid deployed in a $b^2 = 20m \times 20m$ surveillance area, and an acoustic or an electromagnetic source is moving in this region, as shown in Fig. 3.7. Target motion is defined by the white noise acceleration model (3.1) with state vector $\mathbf{x} = [x \ \dot{x} \ y \ \dot{y}]$, where the state transition matrix F and the state noise covariance \mathbf{Q} are given as follows

$$\mathbf{F} = \begin{bmatrix} 1 & \mathcal{D} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathcal{D} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{Q} = \sigma^2 \begin{bmatrix} \mathcal{D}^4/4 & \mathcal{D}^3/2 & 0 & 0 \\ \mathcal{D}^3/2 & \mathcal{D}^2 & 0 & 0 \\ 0 & 0 & \mathcal{D}^4/4 & \mathcal{D}^3/2 \\ 0 & 0 & \mathcal{D}^3/2 & \mathcal{D}^2 \end{bmatrix}$$

At time k , the signal power received at the i^{th} sensor is given as $z_k^i = \sqrt{\frac{P_0}{1+\alpha(d_k^i)^n}} + n_k^i$, where P_0 denotes the signal power of the target, d_k^i is the distance between the target and the i^{th} sensor at time k , α and n are model parameters, and n_k^i is Gaussian noise with zero mean and variance r^i . Without loss of generality, local sensors are set up with the same measurement noise variance $r^i = 1$ ($i = 1, \dots, N$) in this example. We set $P_0 = 10^3$, $\alpha = 1$, and $n = 2$. The target's initial state \mathbf{x}_0 is assumed to be Gaussian with mean $\boldsymbol{\mu}_0 = [-8 \ 2 \ -8 \ 2]$ and covariance $\boldsymbol{\Sigma}_0 = \text{diag}[9 \ 4 \ 9 \ 4]$ (*i.e.*, a poor prior on the initial state). The state process noise parameter σ^2 is set as 0.1, indicating that the target trajectory has relatively large uncertainty. Measurements are assumed to be taken at regular intervals of $\mathcal{D} = 0.5$ seconds and the tracking length is 10 seconds, namely, we track the target over $T_S = 20$ time steps for each Monte-Carlo trial. 200 Monte-Carlo trials are performed in this experiment. The number of particles used in the particle filter at the FC is $N_s = 10^4$.

As in linear systems, the RMSEs, averaged over the Monte-Carlo trials at each time, for the random-selection, CFwMD, CFoMD and all-send cases are compared. The average number of transmission at any given time in this experiment is constrained as $l = 2$.

In Fig. 3.8, the RMSE comparison results are shown. Note that only the state estimate is fed back to obtain the results shown in this figure. It can be observed that the proposed CFwMD scheme outperforms the CFoMD and random-selection scheme under the same bandwidth constraint. On the other hand, compared to the all-send case, the CFwMD scheme does not lose much performance but saves 78% transmissions. One may observe that

RMSEs increase with time at later time steps in Fig. 3.8. This is because the target is moving out of the region of interest (ROI) monitored by the sensors, so there is less and less information available for the estimator.

In Fig. 3.9, the RMSEs of the four schemes, namely, the random-selection, CFoMD, CFwMD, and all-send schemes, are plotted as a function of the average number of transmissions at one time step. One can observe that, when the allowed number of transmissions is small, the proposed CFwMD scheme has significant advantage over both CFoMD and random-selection schemes. It incurs a little bit performance loss compared to the all-send case. As we increase the allowed number of transmissions, the RMSEs of the four schemes approach each other, especially when l is close to the total number of sensors $N = 9$ in the network. This is intuitively reasonable, since when the number of transmissions is large enough, the received observations can already provide enough information for good inference performance, and then neither the censoring procedure nor the information conveyed by the missing data can improve the performance much.

For the nonlinear system, we are also interested in the performance comparison between the two feedback scenarios: 1) only global state estimate feedback is available; 2) the feedback consists of both the global state estimate and its covariance, and the results are provided in Fig. 3.10 for $l = 2, 4, 6$ (the total number of sensors in the ROI is $N = 9$). It can be observed that, as in the linear Gaussian system, the CFwMD2 scheme performs better than the CFwMD scheme as time goes along for each l , since extra global information is fed back to local sensors by the FC. Again, the experimental average number of transmissions over 200 Monte-carlo trials provided in Table 3.3 indicates that the bandwidth constraint is more strictly satisfied by the CFwMD2 than the CFwMD scheme.

Table 3.3: Experimental average number of transmissions (nonlinear system)

| Theoretical l | \bar{l} (CFwMD) | \bar{l} (CFwMD2) |
|-----------------|-------------------|--------------------|
| 2 | 1.5978 | 1.8970 |
| 4 | 3.3480 | 3.8775 |
| 6 | 5.4835 | 5.9533 |

3.7.4 Discussion

It should be noted that the models used in the simulations have relatively low dimension and the network size is rather small. However, such scenarios are frequently used in the target tracking literature [16] [21] [7], and have been found to be appropriate to illustrate the effectiveness of the proposed algorithm. We would like to point out that the proposed methodology can also be applied to moderately high-dimensional systems without requiring large computation effort if feedback is available from the fusion center to local sensors. This is clear if one checks Eqs. (3.9) (3.15) (3.41) and (3.47) for linear systems, and Eqs. (3.62) and (3.66) for nonlinear systems. For a non-feedback system, if the dimension of the dynamic system is high and/or the number of sensors is large, the proposed methodology involves computationally intensive multiple integrals in (3.21) and (3.53). However, if the fusion center is very powerful, the proposed methodology can still be applicable relying on efficient numerical integration approaches, such as those that based on Monte Carlo integration techniques [64]. Note that in this chapter, we have implicitly assumed that identical dynamical model is observed at each sensor. However, this may not be true in some realistic scenarios such as very large-scale dynamical systems [65] [66], and this will be addressed in future work.

3.8 Summary

In this chapter, we have proposed a new methodology to solve linear Bayesian sequential estimation problems by combining the censoring procedure at local sensors and the fusion procedure which fuses both received observations and missing ones, due to the censoring process, at the FC. Both scalar observation and vector observation cases have been discussed in this work. In addition, for the scalar observation case, it has been shown that the proposed innovation based censoring rule is equivalent to that based on the KL divergence between the prior state PDF and the posterior state PDF. Then, we extended the proposed CFwMD scheme to a general nonlinear filtering problem when feedback is available. Numerical results showed that, for both linear and nonlinear filtering problems considered in this chapter, the CFwMD scheme incurs less performance loss than the CFoMD scheme, while both save the same amount of transmissions, compared to the all-send case. In addition, under the same bandwidth constraint, the proposed CFwMD scheme is shown to perform the best among the three schemes, *i.e.*, CFwMD, CFoMD and random-selection schemes. Future work can include a theoretical analysis of the performance of the proposed CFwMD scheme. In the work presented in this chapter, the channels between the local sensors and the FC were assumed to be perfect. Taking a fading channel into consideration is another interesting future direction.

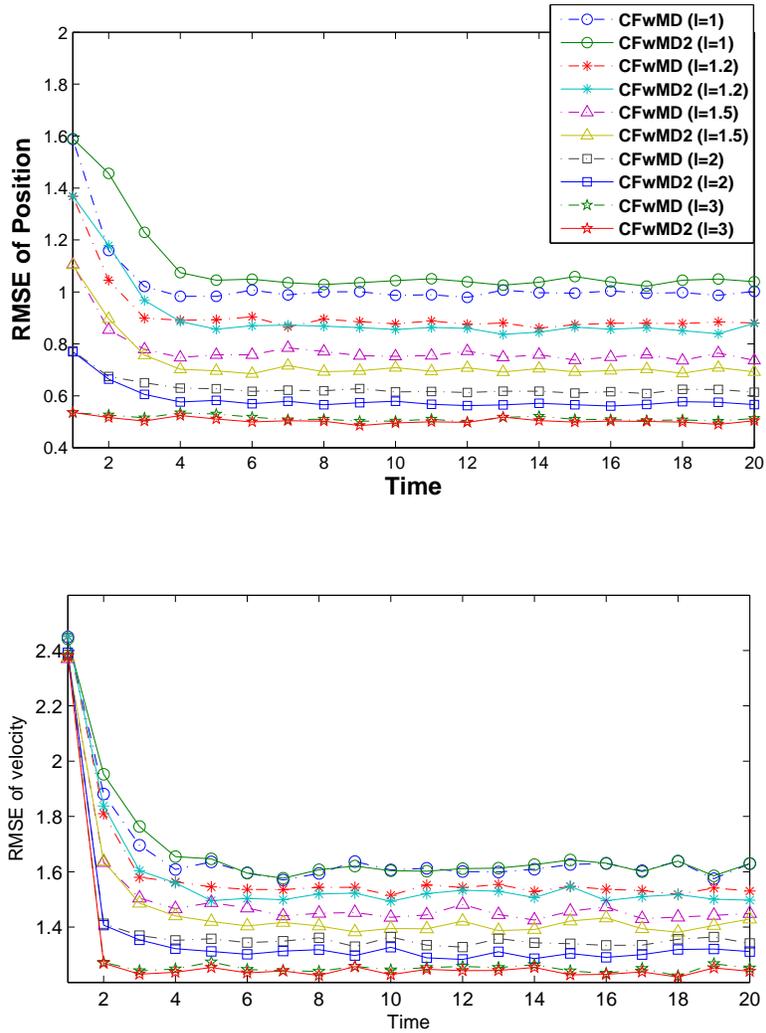


Figure 3.2: RMSEs for the CFwMD with/without covariance feedback for different values of l .

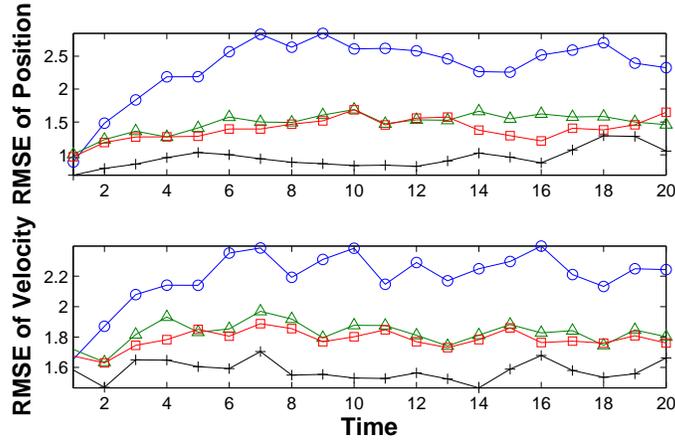


Figure 3.3: RMSE comparison for the non-feedback system with $N = 2$. Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

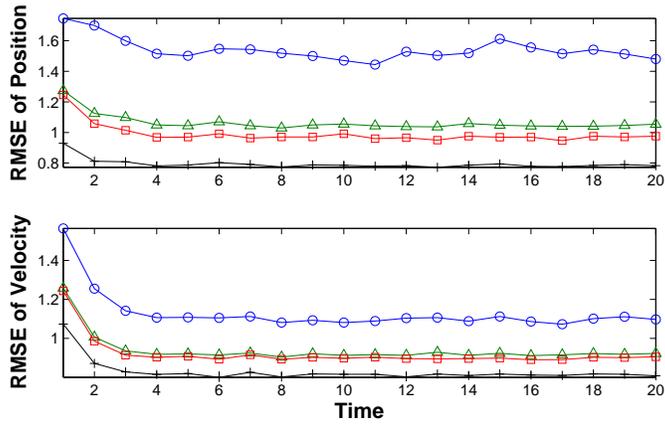


Figure 3.4: RMSE comparison for the feedback system with $N = 2$ (vector observations). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

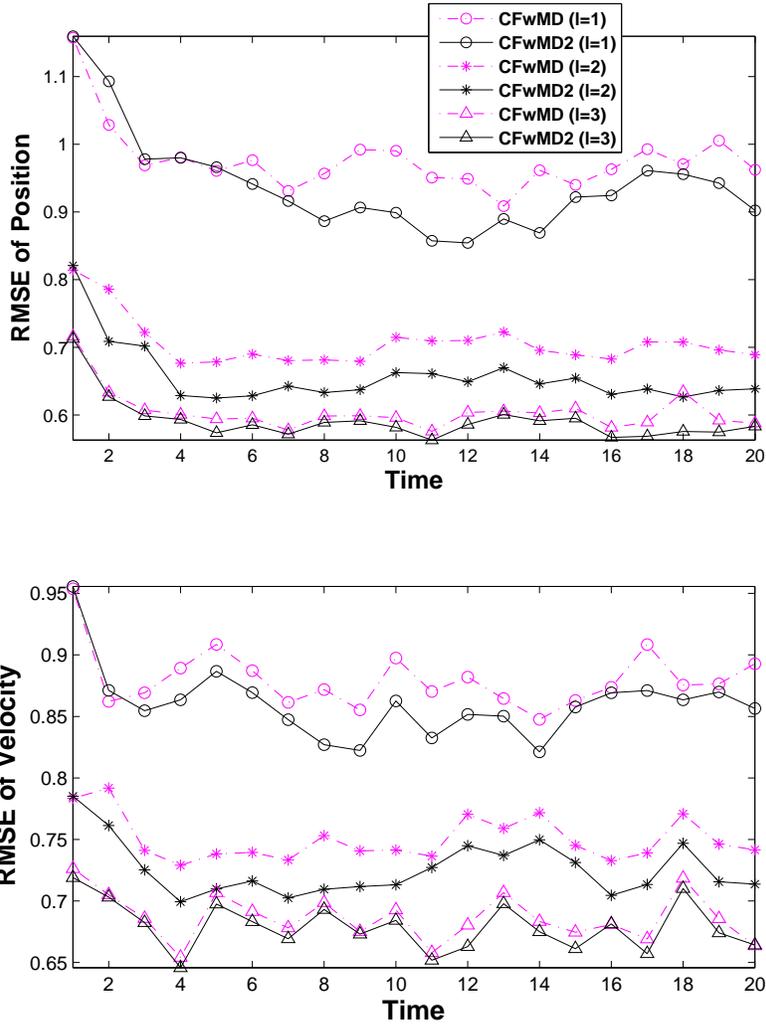


Figure 3.5: RMSEs for the CFwMD with/without covariance feedback for different values of l (vector observations).

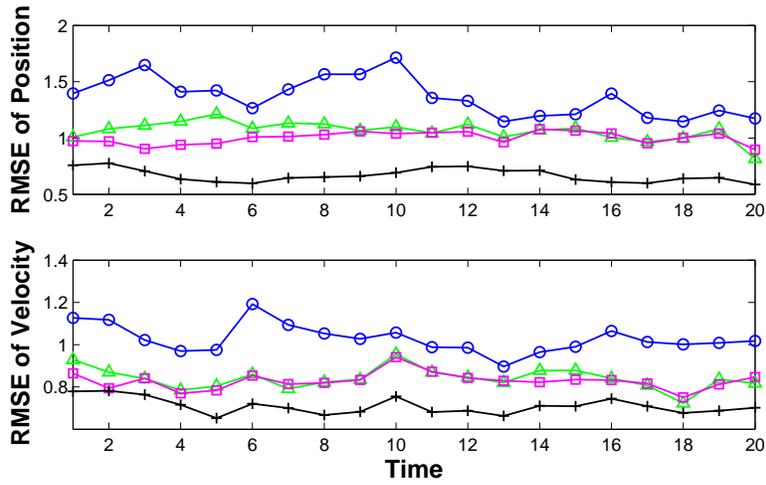


Figure 3.6: RMSE comparison for the non-feedback system with $N = 2$ (vector observation). Solid line with circle: random-selection, solid line with triangle: CFoMD, solid line with square: CFwMD, solid line with plus: all-send.

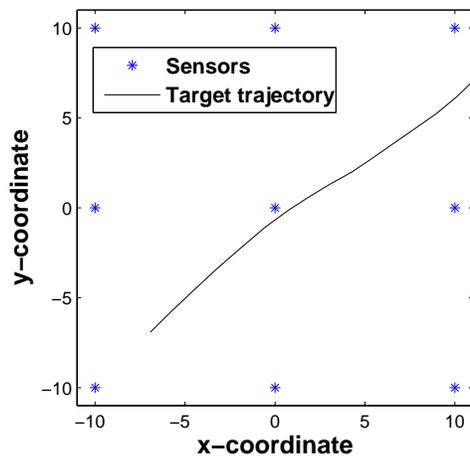


Figure 3.7: Target trajectory and sensor deployment in the ROI

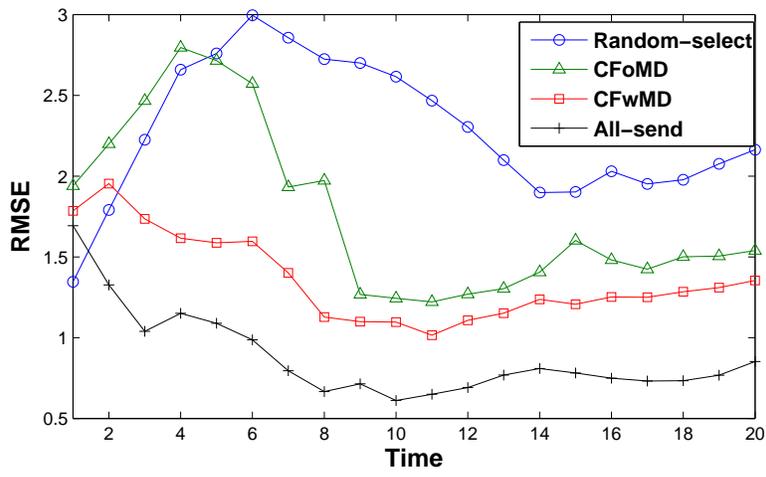


Figure 3.8: RMSE comparison for the nonlinear system with feedback.

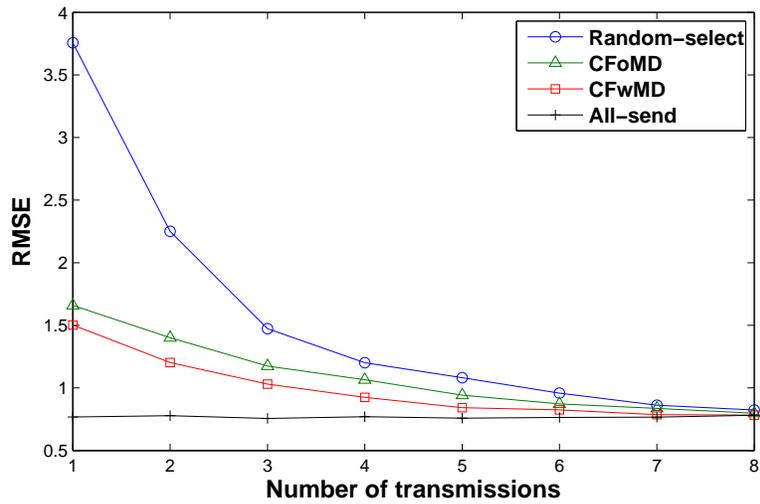


Figure 3.9: RMSEs as a function of the average number of transmission at each time.

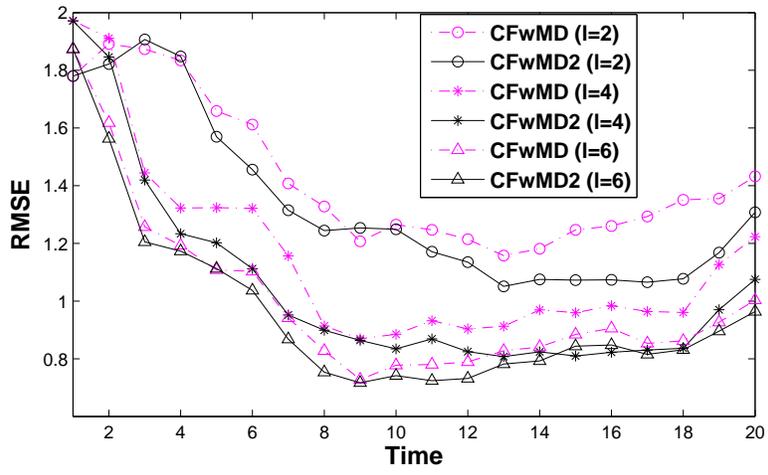


Figure 3.10: RMSEs for the the CFwMD scheme with/without covariance feedback for different values of l in a nonlinear system.

Chapter 4

Probabilistic Sensor Management for Target Tracking via Compressive Sensing

4.1 Motivation

Sensor management is an important problem in resource constrained wireless sensor networks (WSNs). Different approaches have been proposed to solve this problem in the literature for various inference tasks. To name a few, in [2], the sensor selection problem was formulated as an integer programming problem, which has been relaxed and solved through convex optimization. In [5], a multi-step sensor selection strategy by reformulating the Kalman filter was proposed, which is able to address different performance metrics and constraints on available resources. In [49], a sensor selection scheme based on an entropy-based information measure is proposed. Instead of information based metrics, in [50] [22], the recursive one-step-ahead posterior Cramér-Rao lower bound (PCRLB) on the mean squared error (MSE)

of estimating the state vector has been explored as the metric to select informative sensors.

For a WSN with a fusion center, due to the fact that only a few nodes have significant and informative observations, the concatenated measurement vector at the fusion center can be considered to be sparse and compressible. This interpretation naturally brings the concept of compressive sensing (CS) [8] [9] into sensor management problem. The first attempt to solve the sensor management problem by CS was reported in [67], in which the sensor selection decision is considered as a sparse signal, and the sensor selection problem is solved in terms of recovering the sparse signal by l_1 norm minimization.

In this work, we propose a novel CS based sensor management approach. To get compressed measurements at the fusion center, we employ a multiple access channel (MAC) model with probabilistic transmissions, based on which, we can obtain an equivalent representation of our problem as the standard CS problem. After establishing this equivalence, the sensor management problem can be solved by designing the sensing matrix. Since the sensing matrix is completely determined by the probability of transmission by each sensor, design of the sensing matrix is equivalent to finding the optimal probability of transmission for each sensor such that the compressed measurements can yield the best inference performance in a certain sense with limited resources. Numerical results show that the proposed scheme loses a little performance compared to the case where all sensors measurements are completely known at the fusion center but saves a lot of energy. On the other hand, under the same energy constraint, the proposed scheme outperforms the random selection method significantly.

There are several major differences between the work in [67] and the one presented in this chapter: 1) In [67], a subset of sensors is selected and selected sensors send their measurements to the fusion center over parallel channels. In this work, a subset of sensors is probabilistically chosen and different combinations of weighted measurements are sent to the

fusion center over M MACs. 2) In [67], the sensor selection decision is considered as a sparse signal and the sensor selection problem is solved by recovering the sparse signal by l_1 norm optimization. However, in this work, the concatenated measurement vector is considered to be sparse due to non-informative measurements, and the sensing matrix is designed such that a desired tracking performance is achieved with compressed measurements. Thus, there is no recovery of signal, but the compressed signal is used directly for state inference; 3) In [67], the sensing matrix is deterministic or is made semi-random by adding some random disturbance, while in this work, the sensing matrix is random, and each of its element is a random variable whose distribution is related to a certain sensor's probability of transmitting.

4.2 Problem Formulation

4.2.1 System Model

We focus on a target tracking problem, where a moving target is tracked by a WSN with N uniformly deployed sensors in the region of interest (ROI). The dynamical model of an acoustic or electromagnetic target is assumed to be

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + \mathbf{w}_k \quad (4.1)$$

where $\mathbf{x}_k \in \mathcal{R}^d$ is the state vector of the target at time instant k , $F \in \mathcal{R}^{d \times d}$ is the state transition model and \mathbf{w}_k is the process noise which is assumed to be Gaussian with zero mean and covariance matrix $Q \in \mathcal{R}^{d \times d}$.

At time k , the measurement model at each sensor is

$$s_{i,k} = a_{i,k} + v_{i,k}, \quad i = 1, 2, \dots, N \quad (4.2)$$

where $a_{i,k} = \sqrt{\frac{P_0}{1+d_{i,k}^n}}$, P_0 is the signal power of the source, n is the signal decay parameter, $d_{i,k}$ denotes the distance between the target and the i^{th} sensor at time k , *i.e.*,

$d_{i,k} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$, where (x_i, y_i) is the location of the i^{th} sensor, and $v_{i,k}$ is the measurement noise, which is assumed to be Gaussian with mean zero and variance r and mutually independent for different sensors.

4.2.2 Compressive Sensing

CS is a recently developed signal processing technique for acquiring and reconstructing a sparse signal with a small number of measurements compared to the original signal dimension. Consider a signal $\mathbf{f} \in \mathcal{R}^L$, that can be expressed in an orthonormal basis $\Psi = [\Psi_1 \ \Psi_2 \ \dots \ \Psi_L]$ as

$$\mathbf{f} = \sum_{l=1}^L b_l \Psi_l \text{ or } \mathbf{f} = \Psi \mathbf{b} \quad (4.3)$$

where b_l is the coefficient of the signal projected on Ψ_l and $\mathbf{b} = [b_1, \dots, b_L]^T$. The signal \mathbf{f} is said to be K -sparse, if only K coefficients in \mathbf{b} are significant and all others are zeros or are negligible.

To obtain a compressed signal, the sparse signal \mathbf{f} is projected to a lower dimension via a sensing matrix Φ with dimensions $M \times L$, where $M \ll L$, *i.e.*,

$$\mathbf{y} = \Phi \mathbf{f} = \Phi \Psi \mathbf{b}. \quad (4.4)$$

The standard CS problem is to recover \mathbf{b} from only $M \ll L$ measurements \mathbf{y} . The reconstruction capability is determined by the properties of the sensing matrix Φ in addition to the sparsity index and the number of compressed measurements [8]. Several such properties including restricted isometry property (RIP) and mutual coherence of the sensing matrix, and recovery algorithms are developed in [68–70].

4.2.3 Sparsity Formulation

For the sensor management problem considered in this chapter, we can convert it to a CS problem, and hence solve it efficiently with a small number of measurements.

At any given time k , let the measurement vector corresponding to the WSN be $\mathbf{s}_k = [s_{1,k}, \dots, s_{N,k}]^T$ at time k , where $(\cdot)^T$ denotes the matrix or vector transpose. We consider a relatively large distributed network. Based on the observation model (4.2), it is seen that the signal amplitude received at a given node at a given time becomes smaller and eventually negligible as the distance between that particular node and the true target location increases. Therefore, at time k , $\mathbf{a}_k = [a_{1,k}, \dots, a_{N,k}]^T$ can be considered to contain only few significant values.

To obtain the compressed observation model, we consider the following transmission scheme as considered in [71]. Let the j^{th} sensor transmit its measurement after multiplying it by $\phi_{ij,k}$ (to be defined later) via a MAC, so that after M transmissions, the received signal at the fusion center is given by

$$z_{i,k} = \sum_{j=1}^N \phi_{ij,k} s_{j,k} + e_{i,k}, \quad i = 1, \dots, M \quad (4.5)$$

where $e_{i,k}$ is the receiver noise, which is assumed to be white and Gaussian with zero mean and variance ϵ . Note that (4.5) can be written in a vector form as

$$\mathbf{z}_k = \Phi \mathbf{s}_k + \mathbf{e}_k \quad (4.6)$$

where $\mathbf{z}_k = [z_{1,k}, \dots, z_{M,k}]^T$ and \mathbf{e}_k is the receiver noise, which is assumed to be white, zero-mean and Gaussian with covariance $\Sigma_{\mathbf{e}} = \epsilon I_{M \times M}$, where $I_{M \times M}$ is an identity matrix of size $M \times M$.

We consider each $\phi_{ij,k}$ to be a random variable so that

$$\phi_{ij,k} = \begin{cases} 1, & \text{with probability } \frac{1}{2}p_{j,k} \\ 0, & \text{with probability } 1 - p_{j,k} \\ -1, & \text{with probability } \frac{1}{2}p_{j,k} \end{cases} \quad i = 1, \dots, M \quad j = 1, \dots, N \quad (4.7)$$

where $p_{j,k}$ is the probability of transmission from j^{th} sensor at time instant k .

Based on how Φ_k is constructed, it is obvious that, though the elements in a given column in Φ_k are independent and identically distributed (*i.i.d.*), elements in different columns are independent but not identically distributed. Therefore, Φ_k does not follow the RIP as the one with the same isometry constant which has *i.i.d.* random elements. Further, it is noted that, the matrix Φ_k can be very sparse when only a small number of sensors decide to transmit with a high probability. With this sensing matrix, we show numerically that compressed observations in (4.6) provide us with a comparable performance in target tracking to that with (4.2) with relatively small M .

In the context of sensor management, Φ_k plays the role of a sensor management entity that divides the sensors into M sub-sets such that sensors in the same sub-set send their weighted measurements over the same MAC (there are a total of M MACs) if FDMA is used or in the same time slot (there are total M time slots) if TDMA is used. Note that, the weight could be '0', which means that the associated sensor does not send its measurement. For example, if there are $N = 5$ sensors and $M = 3$, one realization of $\Phi_{3 \times 5}$ is given by

$$\Phi_{3 \times 5} = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 \\ -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 \end{bmatrix}.$$

Then, sensors 1, 3, and 5 will send their measurements over all the 3 MACs but with different weights, while sensors 2 and 4 do not send their measurements over any MAC. Therefore,

the problem of managing sensors is equivalent to the design of the sensing matrix Φ_k or the probability vector $\mathbf{p}_k = [p_{1,k}, p_{2,k}, \dots, p_{N,k}]^T$, such that a certain objective function is optimized. In the next section, we propose to use the Fisher information matrix (FIM), the inverse of which is the lower bound of the estimation performance, as the objective function, and details will be given therein.

4.3 The Sensor Management Problem

We find the probability vector \mathbf{p}_k such that, the determinant of the FIM of the system averaged over the sensing matrix Φ is maximized at time k . For the target tracking problem under consideration, a nice recursive computation of the FIM is proposed in [20], which is given as follows

$$J_{k+1} = D_k^{22} - D_k^{21}(J_k + D_k^{11})^{-1}D_k^{12} \quad (4.8)$$

where $D_k^{11} = E \{ -\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \log p(\mathbf{x}_{k+1}|\mathbf{x}_k) \}$

$$D_k^{12} = E \{ -\Delta_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \log p(\mathbf{x}_{k+1}|\mathbf{x}_k) \} = (D_k^{21})^T$$

$$\begin{aligned} D_k^{22} &= E \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} [\log p(\mathbf{x}_{k+1}|\mathbf{x}_k) + \log p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})] \right\} \\ &= D_k^{22,a} + D_k^{22,b}. \end{aligned}$$

For the problem considered in this chapter, we have $D_k^{11} = F^T Q^{-1} F$, $D_k^{12} = -F^T Q^{-1}$, $D_k^{22,a} = Q^{-1}$ and

$$D_k^{22,b} = -E \left\{ \Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \log p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}, \Phi_{k+1}) \right\} \quad (4.9)$$

where the expectation is with respect to \mathbf{z}_{k+1} , \mathbf{x}_{k+1} and Φ_{k+1} . Hence,

$$D_k^{22,b} = -E_{p(\Phi_{k+1})} E_{p(\mathbf{x}_{k+1})} \{ J^D \} \quad (4.10)$$

where

$$J^D = E_{p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}, \Phi_{k+1})} \left\{ \Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \log p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}, \Phi_{k+1}) \right\}. \quad (4.11)$$

We can write (4.6) as

$$\mathbf{z}_k = \Phi_k \mathbf{a}_k + \Phi_k \mathbf{v}_k + \mathbf{e}_k \quad (4.12)$$

where $\mathbf{v}_k = [v_{1,k}, v_{2,k}, \dots, v_{N,k}]^T$, *i.e.*, it is the concatenation of the measurement noises of N sensors. Since the measurement noises are mutually independent, $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, rI_{N \times N})$.

Given \mathbf{x}_{k+1} and Φ_{k+1} , based on Eq. (4.12), one can get $\mathbf{z}_{k+1} \sim \mathcal{N}(\Phi_{k+1} \mathbf{a}_{k+1}, R_{k+1})$, where $R_{k+1} = r\Phi_{k+1}\Phi_{k+1}^T + \Sigma_e$. Then

$$\begin{aligned} & \log p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}, \Phi_{k+1}) \\ &= -\frac{1}{2}(\mathbf{z}_{k+1} - \Phi_{k+1} \mathbf{a}_{k+1})^T R_{k+1}^{-1} (\mathbf{z}_{k+1} - \Phi_{k+1} \mathbf{a}_{k+1}). \end{aligned} \quad (4.13)$$

Therefore,

$$J^D = -\nabla_{\mathbf{x}_{k+1}} (\Phi_{k+1} \mathbf{a}_{k+1}) R_{k+1}^{-1} \nabla_{\mathbf{x}_{k+1}}^T (\Phi_{k+1} \mathbf{a}_{k+1}), \quad (4.14)$$

$$\begin{aligned} & D_k^{22,b} \\ &= E_{\mathbf{x}_{k+1}} \left\{ \nabla_{\mathbf{x}_{k+1}} \mathbf{a}_{k+1} E_{\Phi} \left\{ \Phi_{k+1}^T R_{k+1}^{-1} \Phi_{k+1} \right\} \nabla_{\mathbf{x}_{k+1}}^T \mathbf{a}_{k+1} \right\} \end{aligned} \quad (4.15)$$

where

$$\nabla_{\mathbf{x}_{k+1}} \mathbf{a}_{k+1} = \begin{bmatrix} \frac{\partial a_{1,k+1}}{\partial x_{k+1}} & \frac{\partial a_{2,k+1}}{\partial x_{k+1}} & \dots & \frac{\partial a_{N,k+1}}{\partial x_{k+1}} \\ 0 & 0 & \dots & 0 \\ \frac{\partial a_{1,k+1}}{\partial y_{k+1}} & \frac{\partial a_{2,k+1}}{\partial y_{k+1}} & \dots & \frac{\partial a_{N,k+1}}{\partial y_{k+1}} \\ 0 & 0 & \dots & 0 \end{bmatrix}_{d \times N} \quad (4.16)$$

and

$$\frac{\partial a_{i,k+1}}{\partial x_{k+1}} = \frac{P_0 n d_{i,k+1}^{n-2}}{2a_{i,k+1}(1 + d_{i,k+1}^n)^2} (x_i - x_{k+1}), \quad (4.17)$$

$$\frac{\partial a_{i,k+1}}{\partial y_{k+1}} = \frac{P_0 n d_{k+1,i}^{n-2}}{2a_{i,k+1}(1 + d_{k+1,i}^n)^2} (y_i - y_{k+1}) \quad (4.18)$$

for $i = 1, 2, \dots, N$. Note that (x_i, y_i) is the location of the i^{th} sensor and (x_{k+1}, y_{k+1}) represents the location of the target at time $k + 1$.

Up to this point, we have not observed the explicit relationship between the FIM at time $k + 1$ and \mathbf{p}_{k+1} , due to the complexity of $D_k^{22,b}$. The following result can be used to simplify the mathematical representation.

Let $\Gamma_{k+1} \triangleq E_{\Phi} \{ \Phi_{k+1}^T R_{k+1}^{-1} \Phi_{k+1} \}$. If N is large, then we have the following proposition.

Proposition 4.1. *If the number of sensors in the WSN N is large, then, at any given time $k + 1$, we may approximate*

$$\Gamma_{k+1} \approx \frac{M}{r \sum_{j=1}^N p_{j,k+1} + \epsilon} \text{diag}(\mathbf{p}_{k+1}) \quad (4.19)$$

where $\text{diag}(\mathbf{p})$ denotes a diagonal matrix, with \mathbf{p} on the main diagonal.

Proof. Let $\Theta \triangleq \Phi \Phi^T$. The time index is omitted in the proof for the sake of simplicity.

Diagonal elements of Θ are given by

$$\Theta_{i,i} = \sum_{j=1}^N \phi_{ij}^2 \quad i = 1, 2, \dots, M \quad (4.20)$$

and off-diagonal elements are

$$\Theta_{i,j} = \sum_{c=1}^N \phi_{ic} \phi_{jc} \quad i, j = 1, 2, \dots, M \text{ and } i \neq j. \quad (4.21)$$

It is straightforward to get $E\{\phi_{ij}^2\} = p_j$, $\text{var}\{\phi_{ij}^2\} = \sum (\phi_{ij}^2 - p_j)^2 p(\phi_{ij}) = p_j(1 - p_j)$, $E\{\phi_{ic} \phi_{jc}\} = 0$, and $\text{var}\{\phi_{ic} \phi_{jc}\} = p_c^2$.

Therefore, according to the law of large number (LLN) for independent and non-identical random variables, we get

$$\Theta_{i,i} \approx \sum_{j=1}^N E\{\phi_{ij}^2\} = \sum_{j=1}^N p_j, \quad \Theta_{i,j} \approx \sum_{c=1}^N E\{\phi_{ic} \phi_{jc}\} = 0. \quad (4.22)$$

Hence, $\Theta = \left(\sum_{j=1}^N p_j\right) I_{M \times M}$ and $R = r\Theta + \Sigma_{\mathbf{e}} = (r \sum_{j=1}^N p_j + \epsilon) I_{M \times M}$.

Then,

$$\Gamma \approx \left(r \sum_{j=1}^N p_j + \epsilon\right)^{-1} E_{\Phi} \{\Phi^T \Phi\}. \quad (4.23)$$

Diagonal elements of Γ are given by

$$\Gamma_{i,i} = \left(r \sum_{j=1}^N p_j + \epsilon\right)^{-1} M p_i \quad (i = 1, 2, \dots, N)$$

and off-diagonal elements are

$$\begin{aligned} \Gamma_{i,j} &\approx \left(r \sum_{j=1}^N p_j + \epsilon\right)^{-1} E \left\{ \sum_{k=1}^M \phi_{ki} \phi_{kj} \right\} \\ &= 0 \quad (i, j = 1, 2, \dots, N \text{ and } i \neq j). \end{aligned}$$

Therefore,

$$\Gamma \approx \frac{M}{r \sum_{j=1}^N p_j + \epsilon} \text{diag}(\mathbf{p}) \quad (4.24)$$

completing the proof. ■

The goal is to solve the resource management problem in a WSN, and the limited resource that we focus on here is the energy consumption in the network. For simplicity, we assume that each transmission from a local sensor to the fusion center consumes unit power. Finding the optimal values for transmitting power at sensor nodes while achieving a desired performance is another interesting aspect which will be studied in the future. We aim to solve the following optimization problem:

$$\max_{\mathbf{p}_k} \det(J_k(\mathbf{p}_k)) \quad (4.25)$$

$$\text{s.t.} \quad M \sum_{j=1}^N p_{j,k} \leq E_k \quad (4.26)$$

where E_k is the total energy constraint at time k .

Remark: (1) The fusion center maintains a particle filter to track the target, since the system model considered here is nonlinear. (2) At time step k , the fusion center first solves

the optimization problem in (4.25) to get the optimal \mathbf{p}_k before measurements at this time are available. Then, it generates the sensing matrix Φ_k using \mathbf{p}_k , and, according to which, sends control messages to local sensors. Based on these control messages, local sensors will send their measurements over assigned MACs to the fusion center.

4.4 Simulation Results

In this section, we illustrate the performance of the proposed sensor management algorithm by numerical examples. The MATLAB function ‘fmincon’ is used to solve the constrained optimization problem (4.25). We compare the mean square error (MSE) of CS based sensor management method to that of the random selection method under the same energy constraint. Both methods are compared to the all-send case where all sensor measurements are available at the fusion center via a set of parallel channels. The effect of the number of MACs, *i.e.*, M of the sensing matrix Φ on the inference performance is also studied.

We consider a WSN, consisting of $N = 25$ sensors grid deployed in a $20m \times 20m$ surveillance area. The dynamical model of the target is given by (4.1) with state vector $\mathbf{x}_k = [x_k \dot{x}_k y_k \dot{y}_k]^T$. The state transition model F and the covariance of the process noise Q are given as follows

$$F = \begin{bmatrix} 1 & \mathcal{D} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathcal{D} \\ 0 & 0 & 0 & 1 \end{bmatrix}, Q = \rho \begin{bmatrix} \frac{\mathcal{D}^3}{3} & \frac{\mathcal{D}^2}{2} & 0 & 0 \\ \frac{\mathcal{D}^2}{2} & \mathcal{D} & 0 & 0 \\ 0 & 0 & \frac{\mathcal{D}^3}{3} & \frac{\mathcal{D}^2}{2} \\ 0 & 0 & \frac{\mathcal{D}^2}{2} & \mathcal{D} \end{bmatrix}$$

where $\mathcal{D} = 0.5$ seconds is the time interval and $\rho = 0.1$ is the process noise parameter. The parameters of the observation model (4.2) are set as $P_0 = 10^3$ and $r = 1$. The initial state of the target \mathbf{x}_0 is assumed to be Gaussian with mean $\mu_0 = [-13 \ 2 \ -13 \ 2]^T$ and covariance

$\Sigma_0 = \text{diag}([4 \ 1 \ 4 \ 1])$. We perform target tracking over $T_s = 15$ time steps for each Monte-Carlo trial, and set $N_s = 5000$ particles for the particle filter. The total energy available in the WSN at any given time is assumed to be $E = 6$. The MSE of the estimation at each time is averaged over $MC = 100$ Monte-Carlo trials.

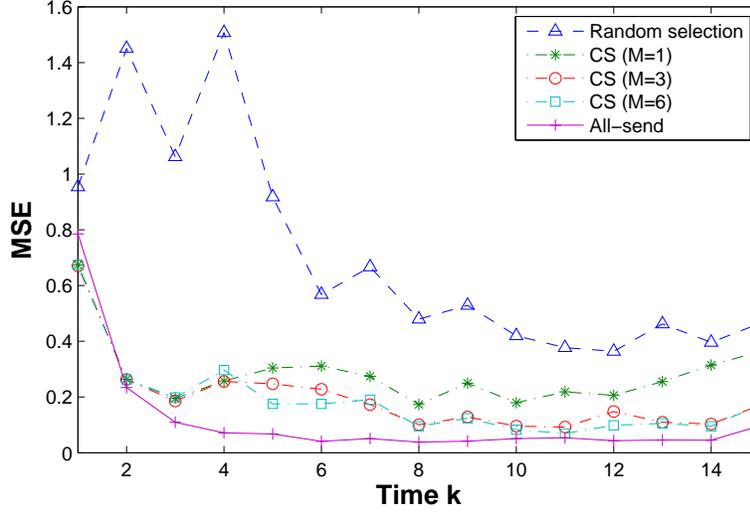


Figure 4.1: MSE comparison for different approaches.

In Figure 4.1, the MSEs of the CS based approach with different values of M are compared to that of the random selection method. One can observe that the former one outperforms the latter under the same energy constraint. The MSEs for $M = 1, 3, 6$ show that the proposed approach achieves a better performance as M increases. This reasonable result can be justified based on the following two reasons: 1) If M is interpreted as the number of MACs, then more channels definitely yield better performance; 2) In the context of CS, M is the number of compressed measurements. Then, a larger M should have higher probability to recover the original signal, and therefore, should yield better performance. Note that, according to the theory of CS, it is not necessary to choose a large M , if it already attains a threshold which guarantee the recovery of the original signal with overwhelming probability

[9]. In Figure 4.1, we also compare the performance of the CS based method with the all-send case. Compared to the all-send case where a total of 25 units of energy are consumed at each time since $N = 25$ parallel transmissions are necessary, the proposed approach loses only a little performance, especially when $M = 6$, but is energy efficient in the sense that it consumes only $E = 6$ units of energy on an average at any given time. Note that, the compressed measurements are used directly for state estimation in the considered target tracking problem, which is different from the traditional CS problems where the goal is to recover a sparse signal. As M increases, we expect that the tracking performance of the proposed scheme will be close to that of the all-send case where there is no compression.

Since the proposed CS based sensor management scheme employs a probabilistic transmission strategy, intuitively, the optimal solution \mathbf{p}_k to (4.25) should assign significant probabilities to those sensors which can obtain more informative observations. To show this, we investigate one Monte-Carlo trial of the tracking trajectory and observe the optimal probability vector \mathbf{p}_k at any given time step k . In Figure 4.2, the optimal \mathbf{p} at time steps $k = 4$

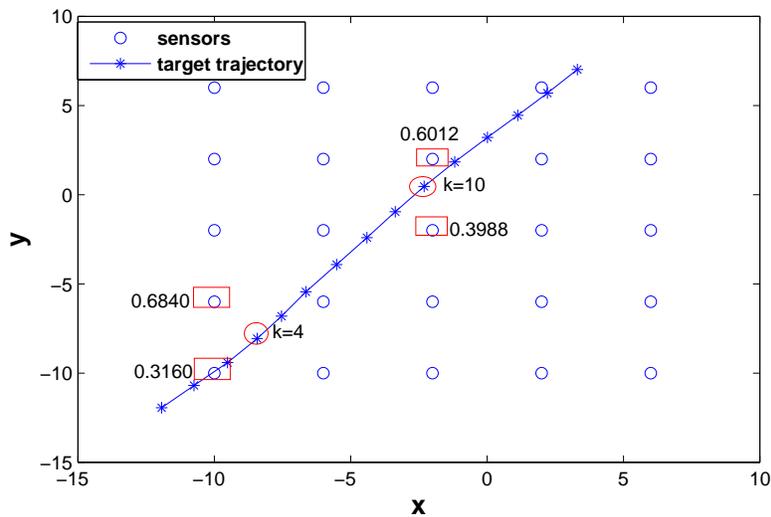


Figure 4.2: \mathbf{p} at different time steps. Circle: true state; Square: sensor assigned non-negligible transmission probability.

and $k = 10$ are marked. We can observe that, at time step $k = 4$, two sensors close to the target are assigned non-negligible probabilities, *i.e.*, one is 0.6840 and the other is 0.3160, while others are assigned almost zero probability. This is because, the two sensors are close to the target from two very different locations relative to the target. Similar observations can be made for time step $k = 10$.

4.5 Summary

In this chapter, we proposed a novel probabilistic sensor management approach for target tracking in sensor networks based on compressed observations. With this model, the sensor management problem becomes a constrained optimization problem, where the goal is to determine the optimal values of probabilities that each sensor should transmit with such that the determinant of the FIM at any given time step is maximized. Numerical results show that the proposed approach saves a lot of energy with a little performance loss compared to the optimal scenario in which all sensor observations are transmitted to the fusion center via parallel channels. Under the same energy constraint, it outperforms the random selection approach significantly. Future work will focus on the theoretical analysis on the RIP of the sensing matrix constructed in this chapter. Another interesting future work is to take the channel statistics into consideration.

Chapter 5

Fusion of Quantized Data for Bayesian Estimation Aided by Controlled Noise

5.1 Motivation

For a sensor network (SN) with limited resources (bandwidth and/or energy), it is important to limit the communication within the network. Therefore, transmission of binary or multi-bit quantized data is a desirable solution. For a centralized sensor network architecture with quantized data illustrated in Figure 5.1, each sensor node sends its quantized data to a FC, where all the quantized sensor data are fused to perform parameter estimation (e.g. target localization). Previous work [72–74] has focused on target localization with quantized sensor data using static quantizers when no prior on the target’s location is available. In [72, 73], target localization methods based on quantized sensor data have been developed assuming perfect communication channels between the sensors and the FC, while in [74], wireless channel statistics are taken into consideration. In this chapter, instead of estimation of deterministic target location, we are interested in estimating a random variable (RV) with known prior based on quantized data collected at the FC, keeping the assumption of perfect

communication channels.

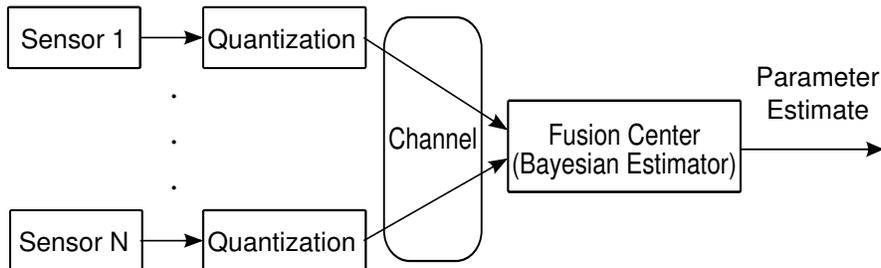


Figure 5.1: Fusion of quantized data for Bayesian estimation.

The novelty of the idea proposed in this chapter is that the quantized data are not fused directly by the FC for parameter estimation, but preprocessed by injecting independent controlled noise. The basic idea was inspired by Widrow’s statistical theory of quantization [10]. The addition of noise after quantization is equivalent to low pass filtering in the characteristic function (CF) domain, such that the original analog observation can be recovered. A similar idea has been applied to solve a distributed detection problem in [75], where the approach of adding external noise reduces the computational complexity. Promising results in [75] motivate us to think about the possible application of this approach to the parameter estimation problem. Therefore, the major contribution of this chapter is the derivation of a computationally efficient estimator by applying Widrow’s quantization theory, and the elegant result relies on the fact that the whole process of quantizing and injecting controlled noise can be theoretically modeled as an additive disturbance, whose distribution is analytically derived. This theoretical model facilitates the derivation of the near-optimal linear MMSE (LMMSE) estimator, and its corresponding mean squared error (MSE) in *exact* forms. The derivation of the optimal minimum mean squared error (MMSE) estimator and the posterior Cramér-Rao lower bound (PCRLB) based on the theoretical model are also provided for performance comparison. Numerical results show that the LMMSE estimator can provide comparable

performance to that of the optimal MMSE estimator while saves a lot of computation effort. A related but different work is documented in [76] where the problem of estimating a deterministic parameter in noise using quantized observations has been discussed. However, in [76], the authors proposed to add the dither noise before quantization at local sensors which amounts to anti-alias filtering [76], while we propose to add it post quantization, which is performed by the FC.

Since quantizers are involved in the problem, the issue of bit allocation naturally arises, which has been formulated as an optimization problem in several publications [77,78]. In this chapter, we will also address the bit allocation problem. There are two major differences between our work and that in [78]: 1) In this chapter, the probability density function (PDF) of the equivalent additive disturbance, which models the whole quantization and noise injection process, is derived. Based on this, as discussed earlier, *exact* solutions for estimators and their performance measures are derived. In [78], a quasi-MMSE estimator that fuses quantized data directly was proposed in an ad-hoc manner, by simply replacing the analog data with the quantized ones in the MMSE estimator designed for analog data. This may incur large estimation error and severe sub-optimality in many cases, as clearly shown in the numerical examples in [78]. 2) In [78], the bit allocation problem was solved by minimizing either an upper bound on the MSE of the quasi-MMSE, or an approximated difference between the Fisher information of the analog data and that of the quantized data, the latter of which is based on the strong assumption that the quantization interval approaches to zero. In contrast, in this chapter, the bit allocation problem is solved based on the *exact* Fisher information.

5.2 A Review of Widrow's Statistical Theory of Quantization

In [10], the uniform quantization of a RV is interpreted as sampling of its PDF, and it was shown that the PDF of the quantized RV is the convolution of the input RV's PDF with the PDF of a uniform distribution followed by conventional sampling. Thus, at the i th sensor, the PDF of the quantizer output, u_i , is

$$p_{U_i}(u) = [p_{W_i}(u) * p_{Z_i}(u)] \cdot \sum_{k \in \mathcal{Z}} q_i \delta \left(u - kq_i - \frac{q_i}{2} \right), \quad (5.1)$$

where $p_{Z_i}(z)$ is the PDF of the input RV z_i , $p_{W_i}(w)$ denotes the PDF of a uniform distribution over $[-q_i/2, q_i/2]$, and q_i is the quantization step-size of the uniform quantizer. Thus, uniform quantization introduces two types of distortions or errors: (a) the additive noise w_i , and (b) the aliasing error due to sampling. However, if the input PDF is bandlimited so that its CF $\phi_{Z_i}(v) = 0$ for $|v| > \frac{\pi}{q_i}$, the aliasing error can be avoided and, in principle, the original PDF can be reconstructed from the knowledge of p_{U_i} . This is Widrow's first quantization theorem as proved in [10]:

Theorem 5.1. *If the CF of the input variable Z_i is bandlimited, i.e.,*

$$\phi_{Z_i}(v) = 0, \quad |v| > \frac{\pi}{q_i} \quad (5.2)$$

then the replicas of $\phi_{U_i}(v)$ do not overlap, and in principle, the original PDF p_{Z_i} can be recovered from p_{U_i} .

5.3 Problem Formulation

Let us consider the estimation of a RV θ in a wireless sensor network, where $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$. It is also assumed that N sensors are observing the parameter θ , and each local sensor's

observation is corrupted by independent additive Gaussian noise, *i.e.*, observation model for sensor i is

$$z_i = \theta + n_i, \quad i = 1, 2, \dots, N \quad (5.3)$$

where $n_i \sim \mathcal{N}(0, \sigma_{n_i}^2)$.

Each sensor performs uniform quantization before transmission and the step-size of quantizer i is set as q_i . Denote the quantized data as u_i . In [75], for hypothesis testing problems, the fusion process is simplified by adding *controlled noise* to the observations received at the FC. For the Bayesian estimation problem considered in this chapter, we propose a similar fusion system as shown in Fig. 5.2. An externally generated noise (d_i) with a band-limited CF, is added to the quantized observations from the i th sensor to filter out the repeated and phase-shifted CF side lobes in the CF of u_i . This is analogous to low pass filtering in signal processing. We, therefore, call the noise d_i , the LPF-noise.

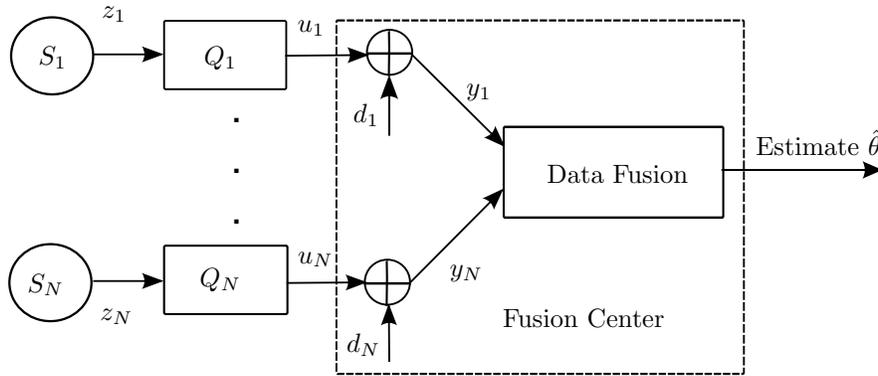


Figure 5.2: Bayesian estimation aided by controlled noise. S: sensor; Q: quantizer; z: sensor data; u: quantized data; d: controlled noise; y: data received at fusion center.

Then, the received data at the FC is given as

$$y_i = u_i + d_i \quad (5.4)$$

Note that an ideal noise source would be one with a rectangular CF in the pass-band, $-\frac{\pi}{q_i} \leq v \leq \frac{\pi}{q_i}$, such that there is no distortion incurred when filtering the signal. However, a rectangular function in the CF domain corresponds to a PDF whose shape corresponds to a sinc function, which is obviously an invalid PDF. Therefore, we limit our consideration to only Gaussian noise in this chapter. That is, $d_i \sim \mathcal{N}(0, \sigma_{d_i}^2)$, and the variance $\sigma_{d_i}^2$ controls the bandwidth of the filter.

Note that once (5.2) is satisfied, we have

$$y_i = u_i + d_i = z_i + w_i + d_i \quad (5.5)$$

where $w_i \sim U(-\frac{q_i}{2}, \frac{q_i}{2})$. One needs to carefully design the PDF of d_i so that it causes minimal distortion while transforming the discrete-valued RV, u_i , into a continuous variable, y_i .

5.4 Controlled Noise Aided MMSE Estimation

In this section, the design of the controlled noise and allocation of bits across the network will be solved jointly, such that the estimation performance of the system is optimized. Since the posterior Cramér Rao lower bound (PCRLB) is the lower bound on the MSE, it is used as the metric in the chapter for optimization.

5.4.1 Bayesian Estimators and Fisher Information

Let us denote the received data vector at the FC as $\mathbf{y} = [y_1, \dots, y_N]^T$. Since z_i and d_i are Gaussian RV respectively, we have $p_{(z_i+d_i)|\theta} = \mathcal{N}(\theta, \sigma_{n_i}^2 + \sigma_{d_i}^2)$. Then, using (5.5), we can

express the likelihood as

$$\begin{aligned}
p(y_i|\theta) &= p(z_i+d_i|\theta) * p_{w_i} \\
&= \mathcal{N}(\theta, \sigma_{n_i}^2 + \sigma_{d_i}^2) * \mathcal{U}\left[-\frac{q_i}{2}, \frac{q_i}{2}\right] \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_{n_i}^2 + \sigma_{d_i}^2)}} e^{-\frac{(\tau-\theta)^2}{2(\sigma_{n_i}^2 + \sigma_{d_i}^2)}} \cdot p_{w_i}(y_i - \tau) d\tau \\
&= \frac{1}{q_i} \int_{y_i - q_i/2}^{y_i + q_i/2} \frac{1}{\sqrt{2\pi(\sigma_{n_i}^2 + \sigma_{d_i}^2)}} e^{-\frac{(\tau-\theta)^2}{2(\sigma_{n_i}^2 + \sigma_{d_i}^2)}} d\tau \\
&= \frac{1}{q_i} \left[\Phi\left(\frac{y_i - \theta + q_i/2}{\sqrt{\sigma_{n_i}^2 + \sigma_{d_i}^2}}\right) - \Phi\left(\frac{y_i - \theta - q_i/2}{\sqrt{\sigma_{n_i}^2 + \sigma_{d_i}^2}}\right) \right] \tag{5.6}
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a Gaussian RV with zero mean and unit variance. Since sensors' observations are conditionally independent, we have

$$p(\mathbf{y}|\theta) = \prod_{i=1}^N p(y_i|\theta) \tag{5.7}$$

5.4.1.1 Optimal MMSE Estimator

The optimal MMSE estimator, *i.e.*, the posterior conditional mean is given as follows

$$\hat{\theta}_{\text{MMSE}} = \int \theta \frac{p(\mathbf{y}|\theta)p(\theta)}{\int p(\mathbf{y}|\theta)p(\theta)d\theta} d\theta \tag{5.8}$$

For any Bayesian estimator, its MSE is bounded below by the PCRLB, which is the inverse of the Bayesian Fisher information. The Bayesian Fisher information is derived and provided in the following theorem.

Theorem 5.2. *For a sensor network with N sensors, and the observation model given by (5.3), if z_i and q_i , for $i = 1, 2, \dots, N$, satisfy the condition specified in (5.2), and the controlled noise d_i is Gaussian, then the Fisher information is given as*

$$J = \sum_{i=1}^N \tilde{J}_i + \sigma_{\theta}^{-2} \tag{5.9}$$

where

$$\tilde{J}_i = \int_{\theta} p(\theta) \int_{y_i} \frac{1}{q_i 2\pi(\sigma_{n_i}^2 + \sigma_{d_i}^2)} \cdot \frac{\left[e^{-\frac{1}{2}\xi_{i,1}^2} - e^{-\frac{1}{2}\xi_{i,2}^2} \right]^2}{\left[\Phi(\xi_{i,1}) - \Phi(\xi_{i,2}) \right]} dy_i d\theta \quad (5.10)$$

$$\text{and } \xi_{i,1} = \frac{y_i - \theta + q_i/2}{\sqrt{(\sigma_{n_i}^2 + \sigma_{d_i}^2)}}, \quad \xi_{i,2} = \frac{y_i - \theta - q_i/2}{\sqrt{(\sigma_{n_i}^2 + \sigma_{d_i}^2)}}.$$

Proof. The Fisher information is given as

$$\begin{aligned} J &= -E \left[\frac{\partial^2}{\partial \theta^2} \log p(\mathbf{y}|\theta) p(\theta) \right] \\ &= \sum_{i=1}^N \left\{ -E \left[\frac{\partial^2}{\partial \theta^2} \log p(y_i|\theta) \right] \right\} + \sigma_{\theta}^{-2} \end{aligned} \quad (5.11)$$

Let $\tilde{J}_i \triangleq -E_{p(y_i, \theta)} \left[\frac{\partial^2}{\partial \theta^2} \log p(y_i|\theta) \right]$, then

$$J = \sum_{i=1}^N \tilde{J}_i + \sigma_{\theta}^{-2}$$

Alternatively,

$$\tilde{J}_i = E_{p(y_i, \theta)} \left[\left(\frac{\partial}{\partial \theta} \log p(y_i|\theta) \right)^2 \right] \quad (5.12)$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta} \log p(y_i|\theta) &= \frac{\frac{\partial}{\partial \theta} p(y_i|\theta)}{p(y_i|\theta)} \\ &= \frac{\frac{1}{q_i \sqrt{(2\pi(\sigma_{n_i}^2 + \sigma_{d_i}^2))}} \left(e^{-\frac{1}{2}\xi_{i,1}^2} - e^{-\frac{1}{2}\xi_{i,2}^2} \right)}{p(y_i|\theta)} \end{aligned} \quad (5.13)$$

and $\xi_{i,1} = \frac{y_i - \theta + q_i/2}{\sqrt{(\sigma_{n_i}^2 + \sigma_{d_i}^2)}}$ and $\xi_{i,2} = \frac{y_i - \theta - q_i/2}{\sqrt{(\sigma_{n_i}^2 + \sigma_{d_i}^2)}}$. Thus, (5.10) is obtained. ■

It is clear that both the implementation of the MMSE estimator and the evaluation of Fisher information involve integrals.

5.4.1.2 Sub-Optimal LMMSE Estimator

Though (5.8) is optimal, it requires the evaluation of two integrals. We would like to derive a more computationally efficient estimator. Combining (5.3) and (5.5), we have

$$y_i = \theta + g_i \quad (5.14)$$

where $g_i \triangleq n_i + w_i + d_i$. It can be shown that

$$E\{g_i\} = E\{n_i + w_i + d_i\} = 0 \quad (5.15)$$

and the variance is given as

$$\sigma_{g_i}^2 = \sigma_{n_i}^2 + \sigma_{d_i}^2 + q_i^2/12 \quad (5.16)$$

since noise n_i , w_i and q_i are independent of each other. g_i does not follow a Gaussian distribution. However, due to the linear relationship between y_i and θ as in (5.14), it is natural to use the LMMSE estimator for θ , which is provided as follows.

$$\hat{\theta}_{\text{LMMSE}} = \mu_\theta + \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \sum_{i=1}^N \frac{y_i - \mu_\theta}{\sigma_{g_i}^2} \quad (5.17)$$

where $\mu_\theta = 0$ is the mean of the prior PDF of θ . The corresponding estimation MSE is

$$E\{(\theta - \hat{\theta}_{\text{LMMSE}})^2\} = \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \quad (5.18)$$

(5.17) and (5.18) are derived based on the following results [79]:

$$\theta_{\text{LMMSE}} = E(\theta) + C_{\theta\mathbf{y}}C_{\mathbf{y}\mathbf{y}}^{-1}(\mathbf{y} - E(\mathbf{y}))$$

$$\text{MSE}(\hat{\theta}_{\text{LMMSE}}) = C_{\theta\theta} - C_{\theta\mathbf{y}}C_{\mathbf{y}\mathbf{y}}^{-1}C_{\mathbf{y}\theta}$$

where θ is a scalar random parameter to be estimated based on a set of measurements $\mathbf{y} = [y_1, \dots, y_N]^T$, $C_{\theta\theta} = E[(\theta - E(\theta))^2]$, $C_{\theta\mathbf{y}} = E[(\theta - E(\theta))(\mathbf{y} - E(\mathbf{y}))^T] = C_{\mathbf{y}\theta}^T$, and $C_{\mathbf{y}\mathbf{y}} = E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))^T]$.

Given that $E(\theta) = \mu_\theta$, $E(g_i) = 0$, θ is independent of g_i , and the relationship between θ and y_i in (5.14), for $i = 1, \dots, N$, we have

$$E(\mathbf{y}) = \mu_\theta \mathbf{1}$$

$$\begin{aligned} C_{\theta\mathbf{y}} &= E[(\theta - \mu_\theta)(\theta \mathbf{1} + \mathbf{g} - \mu_\theta \mathbf{1})^T] \\ &= E[(\theta - \mu_\theta)^2 \mathbf{1}^T + (\theta - \mu_\theta) \mathbf{g}^T] \\ &= \sigma_\theta^2 \mathbf{1}^T \end{aligned} \tag{5.19}$$

where $\mathbf{1}$ is a $N \times 1$ vector of all ones, and $\mathbf{g} = [g_1, \dots, g_N]^T$.

$$\begin{aligned} C_{\mathbf{y}\mathbf{y}} &= E[(\theta \mathbf{1} + \mathbf{g} - \mu_\theta \mathbf{1})(\theta \mathbf{1} + \mathbf{g} - \mu_\theta \mathbf{1})^T] \\ &= E[(\theta - \mu_\theta)^2 \mathbf{1} \mathbf{1}^T + \mathbf{g} \mathbf{g}^T + (\theta - \mu_\theta) \mathbf{1} \mathbf{g}^T + (\theta - \mu_\theta) \mathbf{g} \mathbf{1}^T] \\ &= \sigma_\theta^2 \mathbf{1} \mathbf{1}^T + \text{diag}(\sigma_{g_i}^2) \end{aligned} \tag{5.20}$$

where $\text{diag}(\sigma_{g_i}^2)$ represents a diagonal matrix with the i^{th} diagonal element given by $\sigma_{g_i}^2$.

Thus,

$$\begin{aligned} C_{\theta\mathbf{y}} C_{\mathbf{y}\mathbf{y}}^{-1} &= \sigma_\theta^2 \mathbf{1}^T (\sigma_\theta^2 \mathbf{1} \mathbf{1}^T + \text{diag}(\sigma_{g_i}^2))^{-1} \\ &\stackrel{(a)}{=} \sigma_\theta^2 \mathbf{1}^T \left\{ \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) - \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) \mathbf{1} \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) \right\} \\ &= \sigma_\theta^2 \left\{ \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) - \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) \right\} \\ &= \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) \end{aligned} \tag{5.21}$$

Therefore,

$$\begin{aligned} \hat{\theta}_{\text{LMMSE}} &= \mu_\theta + \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) (\mathbf{y} - \mu_\theta \mathbf{1}) \\ &= \mu_\theta + \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \sum_{i=1}^N \frac{y_i - \mu_\theta}{\sigma_{g_i}^2} \end{aligned}$$

The corresponding MSE is derived as

$$\begin{aligned}
\text{MSE}(\hat{\theta}_{\text{LMMSE}}) &= \sigma_\theta^2 - \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1} \mathbf{1}^T \text{diag} \left(\frac{1}{\sigma_{g_i}^2} \right) \mathbf{1} \sigma_\theta^2 \\
&= \sigma_\theta^2 \left(1 - \frac{\sum_{i=1}^N \frac{1}{\sigma_{g_i}^2}}{\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2}} \right) \\
&= \left(\frac{1}{\sigma_\theta^2} + \sum_{i=1}^N \frac{1}{\sigma_{g_i}^2} \right)^{-1}
\end{aligned}$$

As can be seen, the LMMSE and its MSE have closed-form solutions and are computationally efficient.

Proposition 5.3. *Given q_i such that (5.2) holds, the MSE of the LMMSE estimator is a monotonic increasing function of σ_{d_i} .*

Proof. Given q_i such that (5.2) holds, the MSE of the LMMSE estimator is given by (5.18). Since $\sigma_{g_i}^2 = \sigma_{n_i}^2 + \sigma_{d_i}^2 + q_i^2/12$, obviously, the MSE is a monotonic increasing function of σ_{d_i} . ■

Remark 1: Due to Proposition 5.3, we conjecture that the Fisher information J_i should be a monotonic decreasing function of σ_{d_i} , which will be shown numerically in Section 5.5. This is intuitively true, because a smaller σ_{d_i} means larger signal-to-noise ratio (SNR). Thus, once q_i is fixed, σ_{d_i} can be determined, *i.e.*, the smallest acceptable one. One should note that σ_{d_i} cannot be infinitely small. This can be interpreted from the CF domain of u_i . Since the bandwidth of d_i in the CF domain is inversely proportional to σ_{d_i} , the largest bandwidth acceptable is the one that completely covers the central lobe while does not cover the second side lobe to make sure that no aliasing error is introduced.

5.4.2 LPF-noise Design and Bit Allocation

We are interested in simultaneously designing the local quantizer parameter q_i and controlled noise d_i such that the performance is optimized. In fact, quantizer design is equivalent to the bit allocation problem, since a uniform quantizer is used. Note that sensors considered in this problem are not identical, in the sense that the variances of the observation noises are different from each other, *i.e.*, $\sigma_{n_i} \neq \sigma_{n_j}$, for $i \neq j$. If J given in (5.9) is used as the optimization metric, then the design problem can be formulated as

Optimization Problem:

$$\max_{\vec{B}, \vec{\sigma}_d} J(\vec{B}, \vec{\sigma}_d) \quad (5.22)$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^N b_i = R, \quad \text{and} \\ & \phi_{Z_i}(v) = 0, \quad |v| > \frac{\pi}{q_i}, \quad \text{for } i = 1, 2, \dots, N \end{aligned} \quad (5.23)$$

where $\vec{B} = (b_1, b_2, \dots, b_N)$, $\vec{\sigma}_d = (\sigma_{d_1}, \sigma_{d_2}, \dots, \sigma_{d_N})$, R is the total number of bits, and Z_i is the input variable of the quantizer.

Remark 2: (I) the problem can be solved without exhaustive search over the entire space of $\{\vec{B}, \vec{\sigma}_d\}$, due to Remark 1. The optimal solution for this problem can be obtained as follows: (i) for any possible \vec{B} , determine $\vec{\sigma}_d$ first, and then compute its corresponding Fisher information; (ii) the optimal solution is the combination which provides the maximum Fisher information in step (i). (II) When $N = 2$, there are only $R + 1$ different bit allocation solutions, and one can find the optimal solution by brute force. However, when N is large, the brute force method is not practical, and suboptimal solutions are more desirable. Algorithms such as the GBFOS algorithm [80], the convex relaxation [2], and the approximate dynamic programming method [81] can be used for this purpose.

5.5 Numerical Results

In this section, we first numerically show that J_i is a monotonically decreasing function of σ_{d_i} (Remark 1). Then, when $N = 2$, the optimal bit allocation scheme in the sense of maximizing the Fisher information by brute force is obtained. Besides, numerical results show that the optimal solution obtained by the proposed mechanism indeed yields the minimum MSE, compared to the method of equally distributing the bits and the method of allocating all the bits to the better sensor. We also show that the proposed sub-optimal LMMSE estimator can achieve comparable performance to the optimal one while alleviating the computational complexity.

5.5.1 Experiment 1

In this experiment, only 1 sensor is considered. We set $\sigma_n = 1$, and $q = 0.3$, such that (5.2) holds. The RV θ is Gaussian distributed with zero mean and variance $\sigma_\theta^2 = 4$. We can observe that, from Fig. 5.3, the Fisher information is a monotonically decreasing function of σ_d .

5.5.2 Experiment 2

There are a total of $N = 2$ sensors in the network, and R bits are available to be allocated between the two sensors. The two sensors are different from each other, with $\sigma_{n1} = 0.6$, and $\sigma_{n2} = 3$. Since Gaussian noise is considered in this chapter, and theoretically its bandwidth in the CF domain is not limited, we will truncate the bandwidth in the experiments in this section. That is, $\phi_x(v) \approx 0$, if $|v| \geq 4\sigma_{x_v}$, where $x \sim \mathcal{N}(0, \sigma_x^2)$ and $\sigma_{x_v} = 1/\sigma_x$ (which is because the CF of a Gaussian RV is still in the form of Gaussian, and the variance of the former is the inverse of that of the latter). To ensure that (5.2) holds, we have $\frac{\pi}{q_i} \geq 4(\frac{1}{\sigma_{n_i}})$,

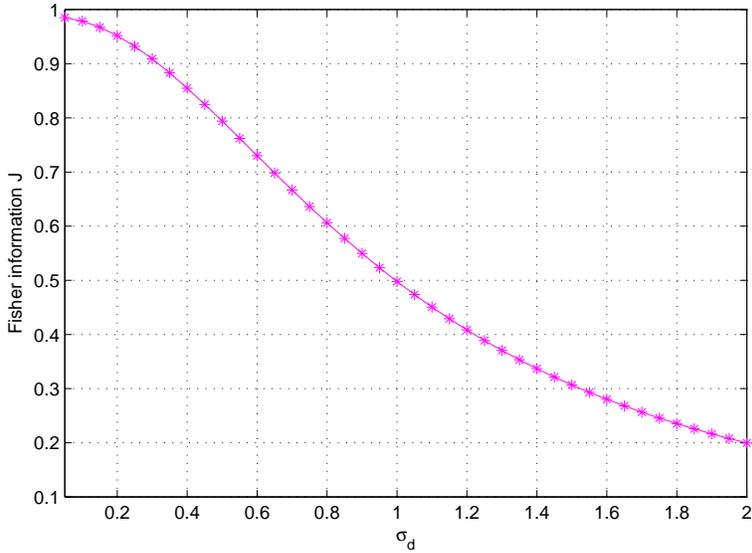


Figure 5.3: Fisher information as a function of σ_d

i.e., $q_i \leq \frac{\pi\sigma_{n_i}}{4}$. According to Remark 1, the smallest standard derivation σ_{d_i} should satisfy the condition $4(\frac{1}{\sigma_{d_i}}) = \frac{2\pi}{q_i} - 4(\frac{1}{\sigma_{n_i}})$, *i.e.*, $\sigma_{d_i} = \frac{1}{\frac{2\pi}{4q_i} - \frac{1}{\sigma_{n_i}}}$. Since a uniform quantizer is used at each local sensor, the relationship between the quantizer resolution q_i and the number of bits b_i for sensor i is given as $q_i = \frac{L_i}{2^{b_i}}$, where L_i is the observation data z_i 's range for sensor i , and $L_i = 8\sqrt{\sigma_\theta^2 + \sigma_{n_i}^2}$.

Table 5.1 shows the Fisher information comparison, when $R = 20$, for all the feasible bit allocation solutions. In the table, the combination (a, b) has the following meaning: a is the number of bits allocated to the first sensor while b is that allocated to the second one, and $a + b = R$. Note that non-feasible solutions (solutions violating (5.2)) are not listed in the table. Note also that, $(20, 0)$ and $(0, 20)$ are considered as feasible solutions, since 0 means one sensor is not active. We can observe that, the equal allocation $(10, 10)$ (which is usually used in sensor networks) does not yield the maximum Fisher information. Another interesting observation from Table 1 is that allocating all bits to the better sensor, *i.e.*, $(20, 0)$ is not the optimal solution, while the optimal one is $(12, 8)$, which implies that

Table 5.1: $R = 20$, Fisher Information Comparison

| Bit alloc. | J | Bit alloc. | J |
|------------|--------|------------|--------|
| (20, 0) | 3.0231 | (10, 10) | 3.1376 |
| (16, 4) | 3.1062 | (9, 11) | 3.1172 |
| (15, 5) | 3.1318 | (8, 12) | 3.1204 |
| (14, 6) | 3.1331 | (7, 13) | 3.0492 |
| (13, 7) | 3.1366 | (6, 14) | 2.7506 |
| (12, 8) | 3.1385 | (0, 20) | 0.3610 |
| (11, 9) | 3.1313 | – | – |

even if sensor 1 is better than sensor 2 in the sense of higher SNR, it is better to assign a few bits to sensor 2 to achieve the diversity gain. However, the Fisher information begins to decrease if more bits are assigned to sensor 2. Nevertheless, the difference between the Fisher information yielded by the optimal bit allocation solution (12, 8) and that yielded by the equal allocation (10, 10) is very small. This is because 10 is a large number of bits, which implies very high resolution of the quantizer.

In Table 5.2, R is reduced to 12. As in Table 5.1, the solution (12, 0) does not yield the optimal performance, neither does the solution (6, 6). And the optimal solution in the sense of maximizing the Fisher information is (8, 4). Obviously, when the total number of bits is decreased to a smaller number, the difference between the Fisher information yielded by the optimal bit allocation solution (8, 4) and that by the equal allocation (6, 6) is much larger than that in Table 5.1. We would like to justify the optimality of the proposed bit allocation scheme. Also, we would like to show that the performance of the sub-optimal LMMSE estimator is comparable to the optimal MMSE estimator. Therefore, in Table

5.2, we also provide the MSE comparison between different bit allocation solutions and between different estimators as well as the corresponding PCRLBs. Note that 1000 Monte Carlo runs are performed to compute the MSEs. Note also that MSE1 is the MSE of the optimal MMSE estimator, MSE2 is that of the sub-optimal LMMSE estimator and MSE3 is computed according to (5.18). Obviously, the optimal solution (8,4) yields the

Table 5.2: $R = 12$, Fisher information comparison

| Bit alloc. | J | PCRLB | MSE1 | MSE2 | MSE3 |
|------------|--------|--------|--------|--------|--------|
| (12, 0) | 3.0276 | 0.3303 | 0.3351 | 0.3353 | 0.3303 |
| (8, 4) | 3.0883 | 0.3238 | 0.3312 | 0.3310 | 0.3238 |
| (7, 5) | 3.0519 | 0.3277 | 0.3374 | 0.3375 | 0.3277 |
| (6, 6) | 2.7492 | 0.3637 | 0.3686 | 0.3685 | 0.3637 |
| (0, 12) | 0.3610 | 2.7698 | 2.7780 | 2.7756 | 2.7692 |

minimum MSE, which justifies the proposed bit allocation scheme. Another observation is that the MSE2 is comparable to MSE1, which means that the proposed LMMSE estimator can provide performance that is very close to the optimal estimator while saving a lot of computation efforts. By comparing MSE2 to MSE3, we can observe that the experimental MSEs are close to the analytic ones. Note that MSE3 is very close to the PCRLBs obtained by inverting (5.9), which further justifies that the LMMSE estimator is a very good alternative to the optimal MMSE estimator.

5.6 Summary

In this chapter, we have proposed a controlled noise aided Bayesian estimation scheme. The controlled noise acts like a low-pass filter in the CF domain. Assuming that the controlled noise is Gaussian, the problems of optimal controlled noise design and bit allocation, in the sense of maximizing the Fisher information at the FC, were solved jointly. A near-optimal linear MMSE estimator was also proposed in this chapter, which is computationally efficient. Numerical results corroborate our theoretical derivation. One interesting future work is to relax the Gaussian assumption on the controlled noise while designing the optimal realizable low-pass filter, so that the performance can be further improved.

Chapter 6

Closed-form Performance bound for Source Localization with Quantized Data in Wireless Sensor Networks

6.1 Motivation

With the significant advances in networking, wireless communications, micro-fabrication, and microprocessors, the wireless sensor network (WSN), once a futuristic technology, has become much more feasible. Due to their many applications in environmental monitoring, battlefield surveillance, and structural health management, WSNs have received considerable attention in recent years [82–84]. In the envisioned WSNs, there are a large number of inexpensive sensors which are densely deployed in a region of interest (ROI). This makes accurate intensity (energy) based object localization possible. Because signal intensity measurements are usually employed for object detection, it is very convenient and economical to utilize them to localize an object, without the need for more sophisticated and more expen-

sive sensor functionalities, such as the direction of arrival (DOA) [85] or time-delay of arrival (TDOA) [86] estimates. Since signal decays as a function of distance, signal intensity data contain the range information of the signal source. Energy-based localization methods have been proposed and developed in [87] for WSNs, where analog energy readings from sensors have been used to localize acoustic sources through the maximum-likelihood (ML) method. In [72], an ML based object localization algorithm using only quantized signal amplitudes was developed. Further, source localization based on quantized signals, which experience Rayleigh fading effect was investigated in [88]. Recently, an energy efficient iterative source localization scheme has been proposed in [89] where the algorithm begins with a coarse location estimate obtained from measurement data from a set of anchor sensors and then a few non-anchor sensors are activated in each iteration, sending their data to the fusion center to refine the location estimate. The non-anchor sensors are selected during each iteration based on some sensor selection criteria. Although sensor management is an important topic in WSNs ([90–94]), it is not the focus of this chapter.

The goal here is to determine bounds on the performance of location estimation. Since localization is an estimation problem, the performance measure in which we are interested is the mean-squared error (MSE) of the estimate. It is well known that the Cramér-Rao lower bound (CRLB) [19], a classical result, provides the lower bound on the MSE for any unbiased estimator. In [72], it has been shown that the ML localization algorithm can achieve its asymptotic performance bound, the CRLB, even with quantized data from a relatively small number of sensors. This implies that the theoretical CRLB is a reliable performance indicator. However, to compute the exact CRLB, the sensor location information must be available at the fusion center, which may not be practical when the sensor density is large. Besides, in some cases, one may be only interested in how the performance of the network is influenced by the sensor density of the system instead of the exact deployment of the sensors.

The impact of sensor density on the performance of an ML localization algorithm for a dense sensor network has been investigated in [95], where the fusion center receives analog data from the local sensors. A compact and closed-form result was derived therein. It was shown that the Fisher information increases linearly with the sensor density and with the signal-to-noise ratio (SNR). In this chapter, the impact of the sensor density on the performance of an ML localization algorithm for a dense sensor network is further investigated, where the fusion center receives quantized data from the local sensors. The main contributions of this chapter are: 1) we derive the closed-form CRLB for a general smooth, differentiable, monotonically decreasing and isotropic signal decay model; 2) we provide theoretical error analysis for the approximation based on LLN; 3) we use the closed-form CRLB as a metric to design a suboptimal quantizer, which is shown to achieve performance similar to the optimal one through numerical results.

6.2 Problem Formulation

As illustrated in Figure 6.1, we assume that a large number of sensors are deployed in the ROI, with a sensor density of λ sensors per unit area. The locations of sensors, denoted by (x_i, y_i) for $i = 1, \dots, N$, are independent realizations of a uniform distribution within the ROI, and they are assumed to be known to the fusion center.

We use the following model in which the signal emitted by the target attenuates isotropically:

$$a_i = g(d_i) \tag{6.1}$$

where $d_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2}$ is the distance between the target and the local sensor i , (x_t, y_t) are the coordinates of the target, and a_i is the received signal amplitude at sensor i . The function $g(\cdot)$ describes how signal power decays as the distance from the target increases.

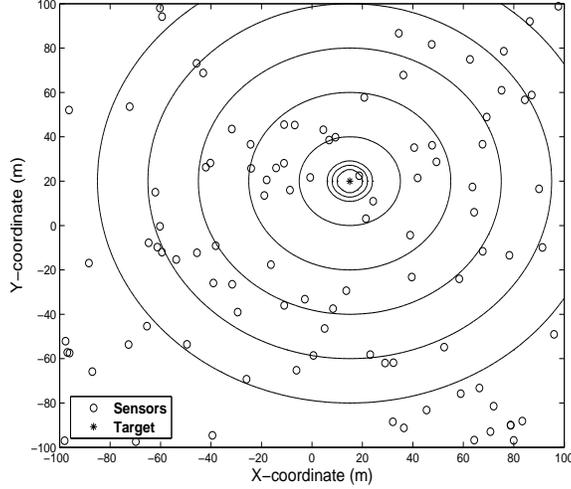


Figure 6.1: The signal intensity contours of a target located in a sensor field.

In this chapter, $g(\cdot)$ is considered as smooth, differentiable and monotonically decreasing.

The signal received at a specific sensor is corrupted by an additive noise. The noises at the local sensors are assumed to be identically and independently distributed (*i.i.d.*) and follow a Gaussian distribution with zero mean and variance σ^2 :

$$s_i = a_i + n_i \quad (6.2)$$

where s_i is the received signal, a_i is the signal amplitude, and

$$n_i \sim \mathcal{N}(0, \sigma^2)$$

The objective is to perform location estimation based on sensor measurements as we discuss next.

6.3 ML Location Estimator and Exact CRLB

If analog data are received at the fusion center, denoted as $\mathbf{S} = \{s_i : i = 1, \dots, N\}$, and let $\theta \triangleq [x_t \ y_t]'$ denote the target location, then the ML estimator based on analog data is given

as

$$\max_{\theta} \ln p(\mathbf{S}|\theta)$$

The corresponding CRLB is given in the following theorem.

Theorem 6.1. *Assuming that the location information of sensors is available at the fusion center, the CRLB on the covariance matrix of the estimation error of the unbiased estimator $\theta(\mathbf{S})$ based on analog sensor measurements is given by*

$$E \left\{ \left[\hat{\theta}(\mathbf{S}) - \theta \right] \left[\hat{\theta}(\mathbf{S}) - \theta \right]^T \right\} \geq \mathbf{J}_A^{-1} \quad (6.3)$$

in which \mathbf{J}_A is the Fisher information matrix (FIM)

$$\mathbf{J}_A = \sum_{i=1}^N \frac{(\partial g(d_i)/\partial d_i)^2}{\sigma^2 d_i^2} \mathbf{G}_i$$

where

$$\mathbf{G}_i \triangleq \begin{bmatrix} (x_i - x_t)^2 & (x_i - x_t)(y_i - y_t) \\ (x_i - x_t)(y_i - y_t) & (y_i - y_t)^2 \end{bmatrix} \quad (6.4)$$

Proof. The FIM can be obtained as follows:

$$\mathbf{J}_A = -E \left[\nabla_{\theta} \nabla_{\theta}^T \ln p(\mathbf{S}|\theta) \right]$$

Consider the element (1, 1) of \mathbf{J}_A :

$$\begin{aligned} \mathbf{J}_A(1, 1) &= E_{p(\mathbf{S}|\theta)} \left\{ -\frac{\partial^2}{\partial x_t^2} \ln p(\mathbf{S}|\theta) \right\} \\ &= \sum_i^N E_{p(s_i|\theta)} \left\{ -\frac{\partial^2}{\partial x_t^2} \ln p(s_i|\theta) \right\} \\ &= \sum_i^N E_{p(s_i|\theta)} \left\{ \frac{\partial}{\partial x_t} \ln p(s_i|\theta) \right\}^2 \end{aligned} \quad (6.5)$$

where

$$\begin{aligned}
\frac{\partial}{\partial x_t} \ln p(s_i|\theta) &= -\frac{\partial}{\partial x_t} \left\{ \frac{(s_i - a_i)^2}{2\sigma^2} \right\} \\
&= \left(\frac{s_i - a_i}{\sigma^2} \right) \frac{\partial a_i}{\partial x_t} \\
&= \left(\frac{s_i - a_i}{\sigma^2} \right) \frac{\partial g(d_i)}{\partial d_i} \frac{\partial d_i}{\partial x_t} \\
&= \left(\frac{s_i - a_i}{\sigma^2} \right) \frac{\partial g(d_i)}{\partial d_i} \frac{x_t - x_i}{d_i}
\end{aligned} \tag{6.6}$$

Then,

$$\begin{aligned}
&E_{p(s_i|\theta)} \left\{ \frac{\partial}{\partial x_t} \ln p(s_i|\theta) \right\}^2 \\
&= \frac{\sigma^2}{\sigma^4} \left(\frac{\partial g(d_i)}{\partial d_i} \right)^2 \frac{(x_t - x_i)^2}{d_i^2} \\
&= \frac{(\partial g(d_i)/\partial d_i)^2}{\sigma^2 d_i^2} (x_i - x_t)^2
\end{aligned} \tag{6.7}$$

where the second equation is due to the fact that $s_i|\theta \sim \mathcal{N}(a_i, \sigma^2)$. Therefore,

$$\mathbf{J}_A(1, 1) = \sum_i^N \frac{(\partial g(d_i)/\partial d_i)^2}{\sigma^2 d_i^2} (x_i - x_t)^2$$

Following a similar procedure, one can derive other elements of $\mathbf{J}_A(1, 1)$. ■

To save bandwidth in the network, we assume that each sensor sends quantized multibit (M -bit) data to the fusion center, which are denoted as $\mathbf{D} = \{D_i : i = 1, \dots, N\}$ where D_i can take any discrete value from 0 to $2^M - 1$. We define $L = 2^M$ and the thresholds for the i th sensor are given by $\vec{\eta}_i = [\eta_{i0}, \eta_{i1}, \dots, \eta_{iL}]$, where $\eta_{i0} = -\infty$ and $\eta_{iL} = \infty$. The quantization process for the i th sensor is such that

$$D_i = \begin{cases} 0 & \infty < s_i < \eta_{i1} \\ 1 & \eta_{i1} \leq s_i < \eta_{i2} \\ \vdots & \vdots \\ L-2 & \eta_{i(L-2)} \leq s_i < \eta_{i(L-1)} \\ L-1 & \eta_{i(L-1)} \leq s_i < \infty \end{cases}$$

Due to the Gaussian noise assumption, we can easily obtain

$$\begin{aligned} p_{il}(\vec{\eta}_i, \theta) &\triangleq p(D_i = l \mid \vec{\eta}_i, \theta) \\ &= Q\left(\frac{\eta_{il} - a_i}{\sigma}\right) - Q\left(\frac{\eta_{i(l+1)} - a_i}{\sigma}\right) \end{aligned} \quad (6.8)$$

where $Q(\cdot)$ is the complementary distribution function of the standard Gaussian distribution

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (6.9)$$

Based on the assumptions made in Section 6.2 and the notations above, it is straightforward to derive the likelihood function at the fusion center node which is given as

$$p(D \mid \theta) = \prod_{i=1}^N \prod_{l=0}^{L-1} p_{il}(\vec{\eta}_i, \theta)^{\delta_{D_i, l}} \quad (6.10)$$

where $\delta_{k,l}$ is the Kronecker delta function:

$$\delta_{k,l} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}$$

Therefore, the log-likelihood function of \mathbf{D} is as follows:

$$\ln p(\mathbf{D} \mid \theta) = \sum_{i=1}^N \sum_{l=0}^{L-1} \delta_{D_i, l} \ln[p_{il}(\vec{\eta}_i, \theta)] \quad (6.11)$$

Now, the ML estimation problem is the following optimization problem:

$$\max_{\theta} \ln p(\mathbf{D}|\theta)$$

The CRLB on the covariance matrix of the estimation error based on quantized sensors' observations is derived and provided in the following theorem:

Theorem 6.2. *Assuming the existence of an unbiased estimator $\hat{\theta}(\mathbf{D})$, and assuming that the location information of sensors is available at the fusion center, the CRLB is given by*

$$E \left\{ \left[\hat{\theta}(\mathbf{D}) - \theta \right] \left[\hat{\theta}(\mathbf{D}) - \theta \right]^T \right\} \geq \mathbf{J}^{-1} \quad (6.12)$$

in which \mathbf{J} is the Fisher information matrix (FIM)

$$\mathbf{J} = \sum_{i=1}^N \beta_i \mathbf{G}_i$$

where \mathbf{G}_i is defined in (6.4),

$$\beta_i = \frac{(\partial g(d_i)/\partial d_i)^2}{2\pi\sigma^2 d_i^2} \gamma_i \quad (6.13)$$

and

$$\gamma_i = \sum_l \frac{1}{p_{il}} \left[e^{-\frac{1}{2} \left(\frac{\eta_{il} - a_i}{\sigma} \right)^2} - e^{-\frac{1}{2} \left(\frac{\eta_{i(l+1)} - a_i}{\sigma} \right)^2} \right]^2 \quad (6.14)$$

Proof. The FIM can be obtained as follows:

$$\mathbf{J} = -E \left[\nabla_{\theta} \nabla_{\theta}^T \ln p(\mathbf{D}|\theta) \right]$$

First, we derive the (1,1) element of \mathbf{J} . Taking partial derivative twice of both sides of (6.11) with respect to x_t , we have

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{D}|\theta)}{\partial x_t^2} &= \sum_i \sum_l \frac{\delta_{D_i,l}}{p_{il}} \frac{\partial^2 p_{il}}{\partial x_t^2} \\ &- \sum_i \sum_l \frac{\delta_{D_i,l}}{p_{il}^2} \left[\frac{\partial p_{il}}{\partial x_t} \right]^2 \end{aligned} \quad (6.15)$$

Since

$$E[\delta(D_i - l)] = p_{il}(\vec{\eta}, \theta)$$

we have

$$E \left[\frac{\partial^2 \ln p(\mathbf{D}|\theta)}{\partial x_t^2} \right] = - \sum_i \sum_l \frac{1}{p_{il}} \left[\frac{\partial p_{il}}{\partial x_t} \right]^2 \quad (6.16)$$

where

$$\begin{aligned} \frac{\partial p_{il}}{\partial x_t} &= \frac{\partial}{\partial x_t} \left[Q \left(\frac{\eta_{il} - a_i}{\sigma} \right) - Q \left(\frac{\eta_{i(l+1)} - a_i}{\sigma} \right) \right] \\ &= \frac{e^{-\frac{1}{2} \left(\frac{\eta_{il} - a_i}{\sigma} \right)^2} - e^{-\frac{1}{2} \left(\frac{\eta_{i(l+1)} - a_i}{\sigma} \right)^2}}{\sqrt{2\pi}\sigma} \frac{\partial a_i}{\partial x_t} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial a_i}{\partial x_t} &= \frac{\partial g(d_i)}{\partial d_i} \frac{\partial d_i}{\partial x_t} \\ &= \frac{\partial g(d_i)/\partial d_i}{d_i} (x_t - x_i) \end{aligned}$$

Therefore, it is straightforward to derive that

$$\begin{aligned} \mathbf{J}_{11} &= -E \left[\frac{\partial^2 \ln p(\mathbf{D}|\theta)}{\partial x_t^2} \right] \\ &= \sum_i \beta_i (x_i - x_t)^2 \end{aligned} \quad (6.17)$$

where

$$\beta_i = \frac{(\partial g(d_i)/\partial d_i)^2}{2\pi\sigma^2 d_i^2} \gamma_i \quad (6.18)$$

and

$$\gamma_i = \sum_l \frac{1}{p_{il}} \left[e^{-\frac{1}{2} \left(\frac{\eta_{il} - a_i}{\sigma} \right)^2} - e^{-\frac{1}{2} \left(\frac{\eta_{i(l+1)} - a_i}{\sigma} \right)^2} \right]^2 \quad (6.19)$$

Following a similar procedure, other elements of \mathbf{J} can be derived. ■

6.4 Closed-Form Expressions for FIM and CRLB

In the previous section, the exact CRLB was derived and given in Theorem 6.2. In this section, we will derive closed-form expressions for FIM and CRLB, by assuming an isotropic signal decay model defined in (6.1) which is smooth, differentiable, as well as monotonically decreasing. As we can see, the FIM provided in Theorem 6.2 is quite complex, since it depends on the sensors' relative positions with respect to the target as well as the quantizer thresholds, which makes a compact solution very difficult to derive. In [95], a very simple expression for the FIM and CRLB for source localization based on analog data was obtained, by using the LLN. Here we again assume that the ROI, which is covered densely by uniformly distributed sensors, is very large. Based on this assumption, it is reasonable to assume that for a target that is located in the ROI, the received signal is non-negligible only within a circle surrounding the target, and hence the ROI can be approximately deemed to be without boundaries. The radius of the circle around the target is denoted as R . We assume that a received signal is negligible if its power $P_i = a_i^2 \leq \epsilon P_0$, where ϵ is a very small number. Then, R can be derived from (6.1), *i.e.*, $R = g^{-1}(\sqrt{\epsilon P_0})$.

First, we investigate the (1, 1) element of the FIM

$$J_{11} = \frac{1}{2\pi\sigma^2} \kappa_{11} \quad (6.20)$$

where

$$\kappa_{11} \triangleq \sum_{i=1}^N \frac{(\partial g(d_i)/\partial d_i)^2}{d_i^2} (x_i - x_t)^2 \gamma_i \quad (6.21)$$

Suppose the sensor density is high, and hence the number of sensors around the target that have non-negligible signal strength is large. In this case, we can apply the LLN [96] to approximate the summation in (6.21) by an integral. According to the LLN, as $N \rightarrow \infty$

$$\frac{1}{N}\kappa_{11} \rightarrow E \left[\frac{(\partial g(d_i)/\partial d_i)^2}{d_i^2} (x_i - x_t)^2 \gamma_i \right] \quad (6.22)$$

where the expectation is taken with respect to the distribution of the sensor position (x_i, y_i) .

Therefore, we have

$$\kappa_{11} \approx N \times E \left[\frac{(\partial g(d_i)/\partial d_i)^2}{d_i^2} (x_i - x_t)^2 \gamma_i \right]$$

We assume that the sensors are uniformly distributed within a circle around the target with a radius R , *i.e.*, the distribution is given as

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2} & (x - x_t)^2 + (y - y_t)^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

We denote the region inside this circle as \mathcal{R} and the region inside another circle with radius R and centered at $(0, 0)$ as \mathcal{R}' . Then,

$$\begin{aligned} \kappa_{11} &\approx \frac{N}{\pi R^2} \int \int_{\mathcal{R}} (x - x_t)^2 \left(\frac{\partial g(d)/\partial d}{d} \right)^2 \gamma' dx dy \\ &= \frac{N}{\pi R^2} \int \int_{\mathcal{R}'} x^2 \left(\frac{\partial g(d)/\partial d}{d} \right)^2 \gamma'' dx dy \\ &= \lambda \int \int_{\mathcal{R}'} x^2 \left(\frac{\partial g(d)/\partial d}{d} \right)^2 \gamma'' dx dy \end{aligned} \quad (6.23)$$

where λ is defined as the sensor density in the area, that is,

$$\lambda = \frac{N}{\pi R^2}$$

γ' and γ'' are similar to γ_i in (6.14), except that

in γ' ,

$$a_i = g(d) = g \left(\sqrt{(x - x_t)^2 + (y - y_t)^2} \right)$$

while in γ'' ,

$$a_i = g(d) = g(\sqrt{x^2 + y^2})$$

With the change of variables

$$x = r \cos \theta$$

$$y = r \sin \theta$$

integration in (6.23) can be converted from the cartesian coordinate system to the polar coordinate system. Then,

$$\begin{aligned} \kappa_{11} &= \lambda \int_0^R \int_0^{2\pi} \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \cos^2 \theta \tilde{\gamma}(r) dr d\theta \\ &= \lambda \pi \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \end{aligned} \quad (6.24)$$

where

$$\tilde{\gamma}(r) = \sum_{l=0}^{L-1} \frac{\left[e^{-\frac{1}{2} \left(\frac{\eta_{il} - a(r)}{\sigma} \right)^2} - e^{-\frac{1}{2} \left(\frac{\eta_{i(l+1)} - a(r)}{\sigma} \right)^2} \right]^2}{Q \left(\frac{\eta_{il} - a(r)}{\sigma} \right) - Q \left(\frac{\eta_{i(l+1)} - a(r)}{\sigma} \right)} \quad (6.25)$$

and $a(r) = g(r)$. Therefore,

$$J_{11} = \frac{\lambda}{2\sigma^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr$$

Following a similar procedure, we can also derive J_{22} , J_{12} and J_{21} in the FIM matrix as

$$J_{22} = J_{11} = \frac{\lambda}{2\sigma^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \quad (6.26)$$

$$J_{12} = J_{21} = 0 \quad (6.27)$$

Equation (6.27) can be derived as follows

$$\begin{aligned}
\kappa_{12} &\triangleq \sum_{i=1}^N \frac{(\partial g(d_i)/\partial d_i)^2}{d_i^2} (x_i - x_t)(y_i - y_t) \gamma_i \\
&\approx \frac{N}{\pi R^2} \int \int_{\mathcal{R}} (x - x_t)(y - y_t) \left(\frac{\partial g(d)/\partial d}{d} \right)^2 \gamma' dx dy \\
&= \lambda \int \int_{\mathcal{R}'} xy \left(\frac{\partial g(d)/\partial d}{d} \right)^2 \gamma'' dx dy \\
&= \lambda \int_0^R \int_0^{2\pi} \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \cos\theta \sin\theta \tilde{\gamma}(r) dr d\theta \\
&= \lambda \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \int_0^{2\pi} \cos\theta \sin\theta d\theta \\
&= 0
\end{aligned} \tag{6.28}$$

Therefore, we have the following theorem:

Theorem 6.3. *The Fisher information matrix (FIM) for the location estimation problem with quantized data considered in this chapter can be approximated as*

$$\mathbf{J} = \left[\frac{\lambda}{2\sigma^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \right] \mathbf{I}_2 \tag{6.29}$$

where \mathbf{I}_2 is a 2×2 identity matrix. The corresponding CRLB is the inverse of FIM

$$\mathbf{C} = \frac{2\sigma^2}{\lambda \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr} \mathbf{I}_2 \tag{6.30}$$

Thus, for the quantized data case, we have obtained a result that is very similar to that for the analog data case in [95], and the result is applicable to any smooth, differentiable and monotonically decreasing and isotropic signal decay model. Theorem 6.3 indicates that the Fisher information \mathbf{J} is a linearly increasing function of the sensor density λ given the SNR P_0/σ^2 . The relationship between the Fisher information and the SNR is not as obvious as the results obtained for analog data in [95], since the SNR is involved in the $Q(\cdot)$ and exponential functions in $\tilde{\gamma}(r)$ due to the quantization. Another important point that can be observed from Theorem 6.3 is that the performance of the sensor network of a given density can be assessed without requiring the exact location information of sensors.

6.5 Analysis of FIM Approximation Error

In the previous section, the closed-form expression for FIM was derived by making use of the LLN. In order to show that the LLN is an acceptable approximation in our work, we will theoretically analyze the error introduced by the use of LLN in this section.

For the target location estimation problem, the terms $\frac{(\partial g(d_i)/\partial d_i)^2}{d_i^2}(x_i - x_t)^2\gamma_i$ in (6.21) are *i.i.d.* random variables, since sensors are randomly deployed whose locations follow a uniform distribution in the ROI. Therefore,

$$\begin{aligned} \text{var}\{J_{11}\} &= \text{var}\left\{\frac{1}{2\pi\sigma^2}\kappa_{11}\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^2 N\sigma_{11}^2 \end{aligned} \quad (6.31)$$

where

$$\begin{aligned} \sigma_{11}^2 &= \text{var}\left\{\left(\frac{\partial g(d_i)/\partial d_i}{d_i}\right)^2 (x_i - x_t)^2\gamma_i\right\} \\ &= E\left[\left(\frac{\partial g(d_i)/\partial d_i}{d_i}\right)^2 (x_i - x_t)^2\gamma_i\right]^2 \\ &\quad - E^2\left[\left(\frac{\partial g(d_i)/\partial d_i}{d_i}\right)^2 (x_i - x_t)^2\gamma_i\right] \\ &\triangleq M_2 - M_1^2 \end{aligned}$$

and M_1, M_2 are the first and second moments of the random variable $\left(\frac{\partial g(d_i)/\partial d_i}{d_i}\right)^2 (x_i - x_t)^2\gamma_i$ respectively. They can be computed respectively as follows:

$$\begin{aligned} M_2 &= \int \int_R \frac{1}{\pi R^2} x^4 \left(\frac{\partial g(d)/\partial d}{d}\right)^4 (\gamma'')^2 dx dy \\ &= \frac{1}{\pi R^2} \int_R r^5 \left(\frac{\partial g(r)/\partial r}{r}\right)^4 \tilde{\gamma}^2(r) dr \int_0^{2\pi} \cos^4 \theta d\theta \\ &= \frac{3}{4R^2} \int_R r^5 \left(\frac{\partial g(r)/\partial r}{r}\right)^4 \tilde{\gamma}^2(r) dr \end{aligned} \quad (6.32)$$

$$\begin{aligned}
M_1 &= \int \int_R \frac{1}{\pi R^2} x^2 \left(\frac{\partial g(d)/\partial d}{d} \right)^2 (\gamma'') dx dy \\
&= \frac{1}{\pi R^2} \int_R r^3 \left(\frac{\partial g(r)/\partial r}{r} \right)^2 \tilde{\gamma}(r) dr \int_0^{2\pi} \cos^2 \theta d\theta \\
&= \frac{1}{R^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr
\end{aligned} \tag{6.33}$$

Similarly, for element J_{12} , we have

$$\begin{aligned}
\text{var}\{J_{12}\} &= \text{var} \left\{ \frac{1}{2\pi\sigma^2} \kappa_{12} \right\} \\
&= \left(\frac{1}{2\pi\sigma^2} \right)^2 N\sigma_{12}^2
\end{aligned} \tag{6.34}$$

where

$$\begin{aligned}
\sigma_{12}^2 &= \text{var} \left\{ \left(\frac{\partial g(d_i)/\partial d_i}{d_i} \right)^2 (x_i - x_t)(y_i - y_t)\gamma_i \right\} \\
&= E \left[\left(\frac{\partial g(d_i)/\partial d_i}{d_i} \right)^2 (x_i - x_t)(y_i - y_t)\gamma_i \right]^2 \\
&\quad - E^2 \left[\left(\frac{\partial g(d_i)/\partial d_i}{d_i} \right)^2 (x_i - x_t)(y_i - y_t)\gamma_i \right] \\
&= \frac{1}{\pi R^2} \int_R r^5 \left(\frac{\partial g(r)/\partial r}{r} \right)^4 \tilde{\gamma}^2(r) dr \\
&\quad \cdot \int_0^{2\pi} \cos^2 \theta \sin^2 \theta d\theta \\
&= \frac{1}{4R^2} \int_R r^5 \left(\frac{\partial g(r)/\partial r}{r} \right)^4 \tilde{\gamma}^2(r) dr
\end{aligned} \tag{6.35}$$

The equation above uses the result that $E \left[\left(\frac{\partial g(d_i)/\partial d_i}{d_i} \right)^2 (x_i - x_t)(y_i - y_t)\gamma_i \right] = 0$. Comparing (6.35) to (6.32), one can obtain that $\sigma_{12}^2 = \frac{M_2}{3}$.

One may be more interested in the relationship between the normalized variable J_{11}^r , which is defined as $J_{11}^r = \frac{J_{11}}{E\{J_{11}\}}$, and the sensor density or equivalently, the number of sensors in the ROI. Thus,

$$\begin{aligned}
\text{var}\{J_{11}^r\} &= \frac{\text{var}\{J_{11}\}}{E^2\{J_{11}\}} \\
&= \frac{1}{N} \frac{M_2 - M_1^2}{M_1^2}
\end{aligned} \tag{6.36}$$

J_{22}^r can be computed in the same manner, and it can be shown that $J_{22}^r = J_{11}^r$. We define the normalized diagonal variable as $J_{12}^r = \frac{J_{12}}{\sqrt{E\{J_{11}\}E\{J_{22}\}}}$, and obviously, $J_{12}^r = J_{21}^r$. Then,

$$\begin{aligned}
\text{var}\left\{\frac{J_{12}}{\sqrt{E\{J_{11}\}E\{J_{22}\}}}\right\} &= \frac{\text{var}\{J_{12}\}}{E\{J_{11}\}E\{J_{22}\}} \\
&= \frac{\text{var}\{J_{12}\}}{E^2\{J_{11}\}} \\
&= \frac{1}{N} \frac{M_2}{3M_1^2}
\end{aligned} \tag{6.37}$$

Remark: Equations (6.36) and (6.37) indicate that the variances of J_{11}^r and J_{12}^r decrease as the number of sensors increases, *i.e.*, the approximation by the use of LLN to obtain the closed-form FIM will be more accurate as expected, given the ROI and quantizer thresholds. Note that in this chapter we assume that all the sensors use identical quantizers.

6.6 Examples

In this section, we study two concrete examples of the signal decay model $g(d)$, both of which are smooth, differentiable, monotonically and isotropically decreasing.

6.6.1 Example 1: Gaussian-like Decay Model

In this subsection, we investigate the Gaussian-like decay model [73, 95]:

$$g(d) = \sqrt{P_0} e^{-\frac{\alpha d^2}{2}} \tag{6.38}$$

where P_0 is the signal power emitted by the target at distance zero, and α is a constant. P_0 and α are assumed to be determined through experiments, and are known to the fusion center.

6.6.1.1 Fisher Information Matrix

Corollary 6.4. *Assuming the existence of an unbiased estimator $\hat{\theta}(\mathbf{D})$, the CRLB for the Gaussian-like decay model (6.38) is given by*

$$E \left\{ \left[\hat{\theta}(\mathbf{D}) - \theta \right] \left[\hat{\theta}(\mathbf{D}) - \theta \right]^T \right\} \geq \mathbf{J}^{-1}$$

in which \mathbf{J} is the Fisher information matrix (FIM)

$$\mathbf{J} = \sum_{i=1}^N \beta_i \mathbf{G}_i$$

where G_i and γ_i are the same as (6.4) and (6.14) respectively, while

$$\beta_i = \frac{\alpha^2 P_0 e^{-\alpha d_i^2}}{2\pi\sigma^2} \gamma_i \quad (6.39)$$

Proof. The result of Corollary 6.4 can be obtained by directly applying the model $g(d) = \sqrt{P_0} e^{-\frac{\alpha d_i^2}{2}}$ to Theorem 6.2, *i.e.*, by applying

$$\begin{aligned} \left(\frac{\partial}{\partial d_i} \left[\sqrt{P_0} e^{-\alpha d_i^2/2} \right] \right)^2 &= \left(-\alpha d_i \sqrt{P_0} e^{-\alpha d_i^2/2} \right)^2 \\ &= \alpha^2 d_i^2 P_0 e^{-\alpha d_i^2} \end{aligned}$$

to (6.13) in Theorem 6.2. ■

It can be observed that the result in Corollary 6.4 is consistent with Theorem 1 in [73] where the derivation is directly conducted on the Gaussian-like signal decay model.

Corollary 6.5. *The FIM for the location estimation problem with quantized data for the Gaussian-like signal decay model can be approximated as*

$$\mathbf{J} = \left[\frac{\lambda P_0 \alpha^2}{2\sigma^2} \int_0^R e^{-\alpha r^2} r^3 \tilde{\gamma}(r) dr \right] \mathbf{I}_2 \quad (6.40)$$

where \mathbf{I}_2 is a 2×2 identity matrix, and $\tilde{\gamma}(r)$ is defined in (6.25)

Proof. With $\left[\frac{\partial g(d_i)/\partial d_i}{d_i} \right]^2 = \alpha^2 P_0 e^{-\alpha d_i^2}$,

$$\begin{aligned} J_{11} &= \frac{\lambda}{2\sigma^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \\ &= \frac{\lambda}{2\sigma^2} \int_0^R P_0 \alpha^2 e^{-\alpha r^2} r^3 \tilde{\gamma}(r) dr \\ &= \frac{\lambda P_0 \alpha^2}{2\sigma^2} \int_0^R e^{-\alpha r^2} r^3 \tilde{\gamma}(r) dr \end{aligned}$$

Then, Corollary 6.5 can be obtained using the result in Theorem 6.3. ■

Again, Corollary 6.5 is consistent with Theorem 2 in [73] where the derivation is directly conducted on the concrete signal decay model.

6.6.1.2 Error Analysis

For the Gaussian-like model, $\left[\frac{\partial g(d_i)/\partial d_i}{d_i} \right]^2 = \alpha^2 P_0 e^{-\alpha d_i^2}$, and M_2 and M_1 can be computed according to (6.32) and (6.33):

$$M_2^G = \frac{3}{4R^2} \int_R P_0^2 \alpha^4 e^{-2\alpha r^2} r^5 \tilde{\gamma}^2(r) dr$$

$$M_1^G = \frac{1}{R^2} \int_R P_0 \alpha^2 e^{-\alpha r^2} r^3 \tilde{\gamma}(r) dr$$

Then, $\text{var}\{J_{11}^G\}$ and $\text{var}\{J_{11}^{rG}\}$ can be determined respectively by using (6.36) and (6.37).

6.6.2 Example 2: Power Law Decay Model

In this subsection, we give another example where we use the popular signal decay model in wireless communication: $g(d) = \sqrt{\frac{P_0}{1+(d/d_0)^n}}$, which is called the power law decay model. Here, we set $d_0 = 1$ for simplicity without loss of generality.

6.6.2.1 Fisher Information Matrix

Corollary 6.6. *Assuming the existence of an unbiased estimator $\hat{\theta}(\mathbf{D})$, the CRLB for the power law decay model is given by*

$$E \left\{ \left[\hat{\theta}(\mathbf{D}) - \theta \right] \left[\hat{\theta}(\mathbf{D}) - \theta \right]^T \right\} \geq \mathbf{J}^{-1}$$

in which \mathbf{J} is the Fisher information matrix (FIM)

$$\mathbf{J} = \sum_{i=1}^N \beta_i \mathbf{G}_i$$

where \mathbf{G}_i and γ_i are the same as (6.4) and (6.14) respectively, while

$$\beta_i = \frac{n^2 P_0 d_i^{2n-4}}{8\pi\sigma^2 (1+d_i^n)^3} \gamma_i \quad (6.41)$$

Proof. Since $\frac{\partial g(d_i)}{\partial d_i} = \frac{-\sqrt{P_0} n d_i^{n-1}}{2(1+d_i^n)^{\frac{3}{2}}}$, applying it to (6.13) in Theorem 6.2 results in the corollary. ■

Corollary 6.7. *The FIM for the location estimation problem with quantized data for the power law signal decay model can be approximated as*

$$\mathbf{J} = \left[\frac{\lambda P_0 n^2}{8\sigma^2} \int_0^R \frac{r^{2n-1} \tilde{\gamma}(r)}{(1+r^n)^3} dr \right] \mathbf{I}_2 \quad (6.42)$$

where \mathbf{I}_2 is a 2×2 identity matrix.

Proof. The corollary can be deduced directly from Theorem 6.3. For the power law signal decay model, $\frac{\partial g(r)}{\partial r} = \frac{-\sqrt{P_0} n r^{n-1}}{2(1+r^n)^{\frac{3}{2}}}$. Applying it to Theorem 6.3 results in the corollary. ■

6.6.2.2 Error Analysis

Similar to Section 6.6.1.2,

$$M_2^P = \frac{3}{4R^2} \int_R \frac{P_0^2 n^4 r^{4n-3}}{16(1+r^n)^6} \tilde{\gamma}^2(r) dr$$

$$M_1^P = \frac{1}{R^2} \int_R \frac{P_0 n^2 r^{2n-1}}{4(1+r^n)^3} \tilde{\gamma}(r) dr$$

Then, $\text{var}\{J_{11}^P\}$ and $\text{var}\{J_{11}^{rP}\}$ can be determined by (6.36) and (6.37) respectively.

6.7 Simulation Results

In this section, we first compare the performance of the optimal quantizer that minimizes the exact CRLB to that of the uniform quantizer for the two examples we considered in Section 6.6, and then use the optimal quantizer that we find to perform the remaining simulations. We show that the ML estimator based on quantized data is efficient when the number of quantization bits is large. The closed-form FIMs provided by Corollary 6.5 and Corollary 6.7 will be compared to the exact FIMs given by Corollary 6.4 and Corollary 6.6 respectively in various scenarios. Finally, since the convergence of the exact FIM to the closed-form FIM is guaranteed, a sub-optimal quantizer design based on our closed-form FIM is proposed, which allows us to design the quantizer without requiring the location information of sensors while achieves performance similar to the optimal one.

6.7.1 Quantization Thresholds

In this subsection, we provide examples where the performance of the optimal quantizer is compared to that of the uniform quantizer. We assume that the target is located at $(0, 0)$, sensors are uniformly deployed in a square with side b , which is large enough so that the signal amplitude is negligible at the boundary of the square. The parameters are chosen

as $\sigma = 1$, $P_0 = 10000$, $\alpha = 1$, and $n = 4$. The $\text{SNR} = P_0/\sigma^2$. Thus, the M bit uniform quantizer employs thresholds which equally divide the amplitude region $[-4\sigma \sqrt{P_0} + 4\sigma]$ into 2^M partitions. In order to find the optimal quantizer thresholds, we assume that the fusion center knows the sensor locations. We first specify the lower bound and upper bound of the region in which the quantizer thresholds are located. Then, we randomly choose several starting points with dimension $2^i + 1$, where i is the number of quantization bits. Starting from each of those starting points, the function FMINCON in MATLAB helps us find the local minimum points. Then the minimum of those local minimum points yields the desired quantization thresholds. The corresponding CRLBs are presented in Figures 6.2 and 6.3.

Given the result in Figure 6.2 for the Gaussian-like decay model, where $\text{SNR} = 40 \text{ dB}$, we can observe that, when the number of bits is small, the optimal quantizer has much better performance in the sense that its CRLB is much smaller than that with the uniform quantizer. As the number of bits increases, the two quantizer design schemes lead to almost the same performance. Similar results can be observed from Figure 6.3 for the power law decay model. In the rest of this section, the optimal quantizer for each model is employed in the corresponding numerical examples. Note that the $\text{SNR} P_0/\sigma^2$ is defined to be the SNR at zero distance, and the signal decays rapidly as the distance from the target increases. Therefore, 40 dB is a relatively modest SNR value.

6.7.2 ML estimator

In this subsection, we show that the ML estimator is efficient, especially when the number of bits M is large, for both the Gaussian-like decay model and the power law decay model. We fix sensor density as $\lambda = 2$ in the $b \times b$ square region, and uniformly deploy the sensors in this region. In order to find the global maximum in ML estimation, a grid search is first used to find several coarse maximum points. Then, starting from those points, we use Matlab's

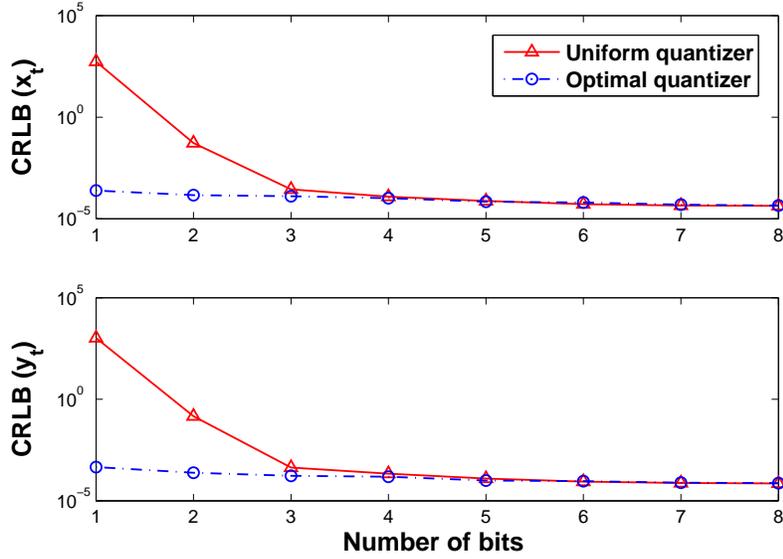


Figure 6.2: The CRLBs as a function of number of bits M (Gaussian-like Decay Model).

FMINCON to find the global maximum point.

For the Gaussian-like decay model, Figure 6.4 shows that the gap between the CRLB based the quantized data computed according to Theorem 6.2 and that based on the analog data computed according to Theorem 6.1 decreases rapidly as the number of bits M increases, and when M is large enough, the former converges to the latter. Furthermore, the root mean squared error (RMSE) of the MLE based on quantized data quickly converges to its CRLB when M increases. Similar results can be observed from Figure 6.5 for the power law decay model.

Hence, the CRLB based on the quantized data given by Theorem 6.2 is a tight bound for the ML estimator derived in the chapter, especially for a large M .

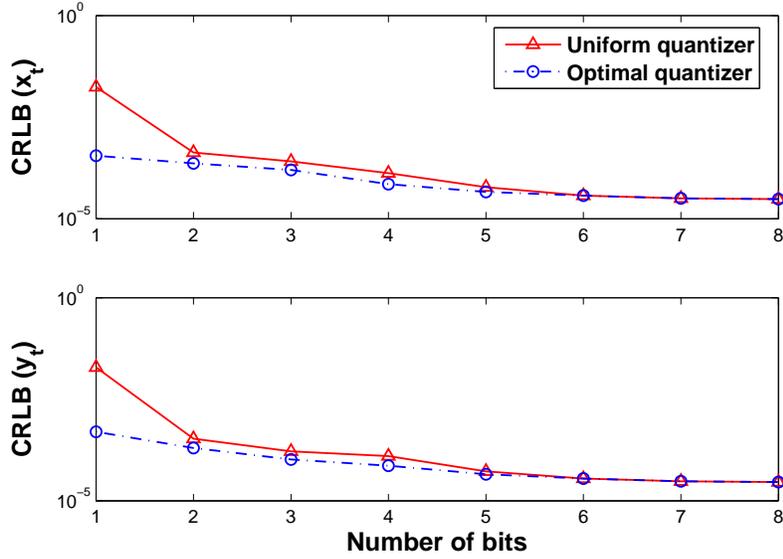


Figure 6.3: The CRLBs as a function of number of bits M (Power Law Decay Model)

6.7.3 Closed-Form CRLB

In this experiment, we compare the closed-form FIM and the exact FIM for the Gaussian-like decay model and the power law decay model respectively. We assume that the ROI is very large and the target is located at the center of the ROI without loss of generality. At the boundary of the ROI, which is a square with side b , the signal reduces to $1.8 \times 10^{-35}P_0$ for the Gaussian-like decay model and $10^{-7}P_0$ for the power law decay model respectively. A total of λb^2 sensors are randomly deployed in the ROI, whose positions follow *i.i.d.* uniform distribution within the ROI. The exact FIM is obtained via one Monte-Carlo run, meaning that only one realization of the sensor deployment is used.

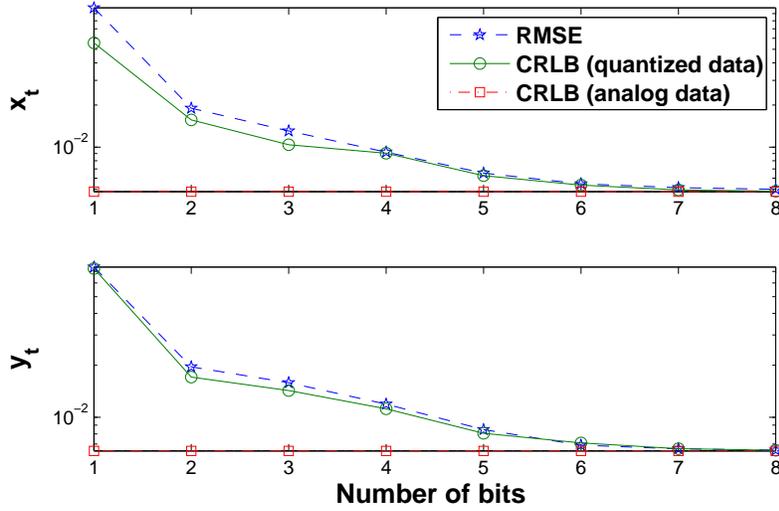


Figure 6.4: RMSE and CRLBs as a function of number of bits M (Gaussian-like Decay Model). $\text{SNR} = 40\text{dB}$.

6.7.3.1 Gaussian-like Decay Model

For the Gaussian-like decay model, it can be shown that on the average the number of sensors whose signal power is greater than ϵP_0 is

$$N = \lambda \pi R^2 = \frac{\lambda}{\alpha} \pi \ln(1/\epsilon) \quad (6.43)$$

From the above equation, it is clear that this number N is determined by $\frac{\lambda}{\alpha}$, for a fixed ϵ . In this experiment, we set $\alpha = 1$ without loss of generality. Furthermore, to compare the off-diagonal elements, we define a quantity analogous to the correlation coefficients for the covariance matrix [96]:

$$\rho = \frac{J_{12}}{\sqrt{J_{11}J_{22}}} \quad (6.44)$$

which is the off-diagonal element normalized by the square root of the product of the diagonal elements. This quantity captures the “correlation” between the two diagonal elements.

In Fig. 6.6, the elements of the FIM calculated by the exact and approximate methods

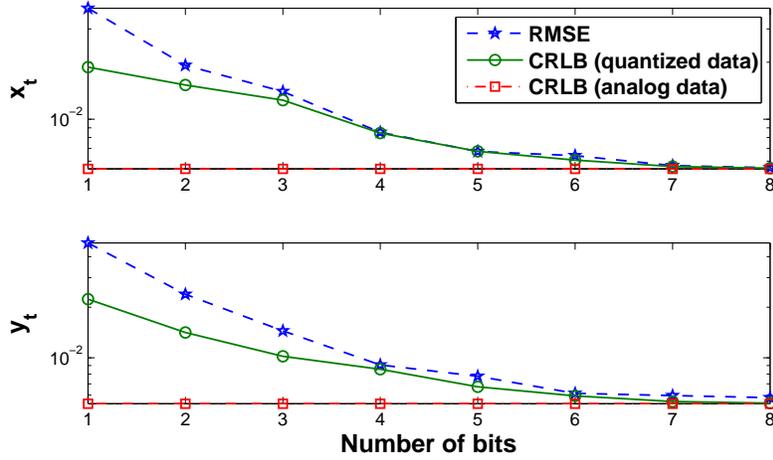


Figure 6.5: RMSE and CRLBs as a function of number of bits M (Power Law Decay Model). $\text{SNR} = 40\text{dB}$.

for a binary quantizer ($M = 1$) have been compared. As we can see, the approximate Fisher information increases linearly as sensor density increases, and the exact FIM oscillates around that obtained using the approximate expression. One should note that the jumpy curves in Fig. 6.6 are due to the fact that only one Monte-Carlo realization is employed for the exact FIM. It can be also observed that, when the sensor density increases, the amplitude of the oscillation around zero of the correlation defined in (6.44) decreases, which is consistent with the theoretical result that J_{12} is approximately zero when the sensor density is large. The convergence of the exact form to the closed-form is much more obvious in Fig. 6.7, where the trace of the CRLB matrix is provided as a function of λ . Note that the trace of the CRLB gives the lower bound on the summation of MSEs of estimating the target's two coordinates x_t and y_t . In Fig. 6.8, the normalized Fisher information J_{11}^r and J_{12}^r as a function of λ and their 1σ regions computed based on (6.36) and (6.37) in Section 6.5 are shown. One can observe that the 1σ region becomes smaller as the sensor density increases, *i.e.*, the variances of the normalized Fisher information J_{11}^r and J_{12}^r decrease as the sensor density

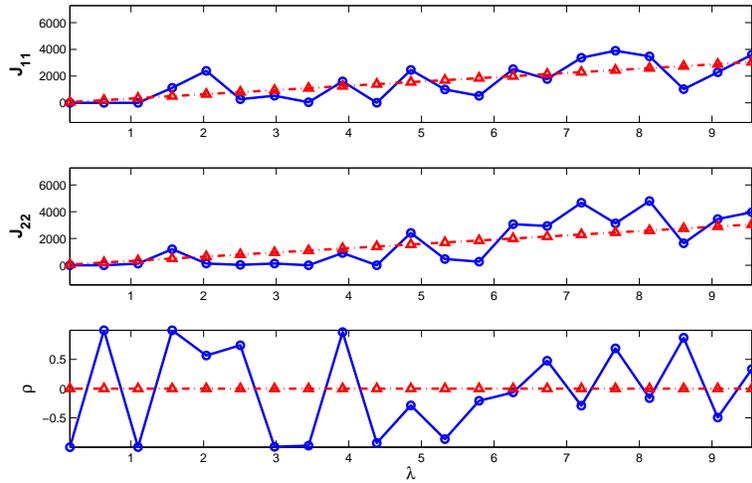


Figure 6.6: Elements of FIM as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Exact FIM; Dashed line + triangle: Approximate FIM. SNR = 40dB and $M = 1$.

increases, which indicates that the closed-form FIM by the use of LLN becomes more and more accurate as the sensor density increases.

6.7.3.2 Power Law Decay Model

As for the Gaussian-like decay model, the comparison between the exact FIM and the closed-form one for the power law decay model is provided in Fig. 6.9. Again, one can observe that the exact FIM oscillates around the closed-form one, due to the fact that only one Monte-Carlo run is performed. Furthermore, the correlation between the two diagonal elements in the FIM decreases as the sensor density increases. In Fig. 6.10, the convergence from the exact CRLB to the closed-form one is obvious. As for the Gaussian-like decay model, it can be observed that the estimation performance is better for a larger sensor density from Fig. 6.11. Note that, for the power law decay model, $M = 4$ bit quantizers are used instead of $M = 1$ bit quantizers which are used for the Gaussian-like decay model. The purpose is to

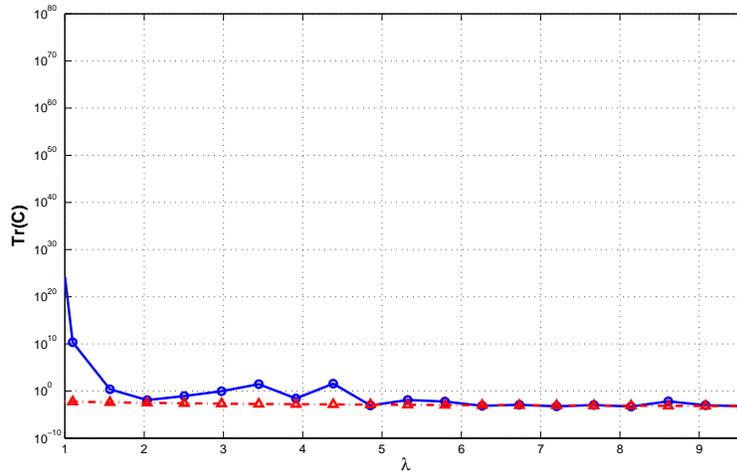


Figure 6.7: Trace of the CRLB matrix as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Exact CRLB; Dashed line + triangle: Approximate CRLB. SNR = 40dB, and $M = 1$.

show results with different parameters.

6.7.4 Suboptimal Quantizer Design Based on the Closed-Form CRLB

In the previous simulations, the optimal quantizer is designed based on the assumption that the fusion center knows each sensor's exact location. This assumption is strong and may not be practical. Moreover, it requires that the design of the optimal quantizer depend on the exact deployment of the sensors in the network. Therefore, it is desirable to design a quantizer which yields performance that is better than the uniform quantizer while relaxing the strong dependence on the information of the sensor location information. Since we have shown that the exact CRLB quickly converges to the closed-form CRLB in Section 6.7.3, we design a suboptimal quantizer based on the closed-form CRLB without acquiring the

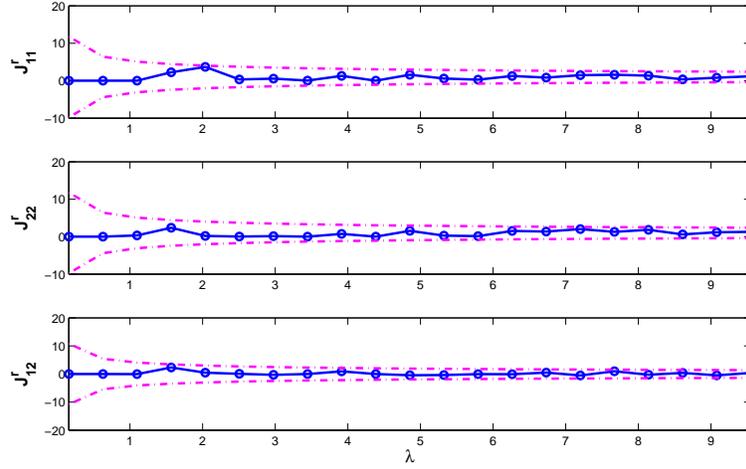


Figure 6.8: Relative FIM as a function of sensor density λ (Gaussian-like Decay Model). Solid line + circle: Relative FIM; Dashed line: 1σ region of the Relative FIM. SNR = 40dB, and $M = 1$.

location information for each sensor, hoping that it can achieve similar performance to the optimal one. Then, the thresholds of the suboptimal quantizer are given as

$$\hat{\eta} = \arg \max_{\eta} \mathbf{J} \quad (6.45)$$

where \mathbf{J} is given by (6.29). Note that \mathbf{J} is an identity matrix scaled by a factor, and therefore, it is sufficient to maximize the factor in (6.29), namely,

$$\hat{\eta} = \arg \max_{\eta} \left[\frac{\lambda}{2\sigma^2} \int_0^R \left(\frac{\partial g(r)/\partial r}{r} \right)^2 r^3 \tilde{\gamma}(r) dr \right] \quad (6.46)$$

In Figure 6.12 and Figure 6.13, we compare the performance for three different quantizers: 1) the optimal quantizer (we need to know the sensor positions); 2) the uniform quantizer; 3) the proposed suboptimal quantizer designed based on the closed-form FIM. The sensor density is chosen as $\lambda = 6$ for the Gaussian decay model in Figure 6.12 and $\lambda = 4$ for the power law decay model in Figure 6.13 respectively. In both models, we assume SNR = 40dB.

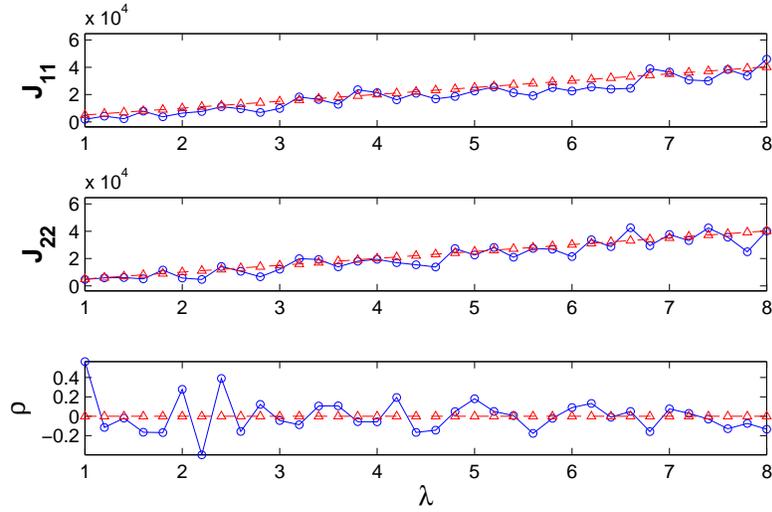


Figure 6.9: Element of FIM as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Exact FIM; Dashed line + triangle: Approximate FIM. SNR = 40dB and $M = 4$.

The advantage of the suboptimal quantizer is obvious when $M = 1$, which is a common case when bandwidth is stringent in sensor networks: it performs much better than the uniform quantizer. When the number of bits increases, both the uniform quantizer and the suboptimal quantizer achieve comparable performance to the optimal quantizer. Therefore, it can be concluded that the closed-form FIM is an acceptable metric for suboptimal quantizer design.

6.8 Summary

In this chapter, we have explored the source localization problem based on a ML localization algorithm, which utilizes the received quantized signal amplitudes at local sensors to estimate the coordinates of the target. Motivated by the goal of investigating the impact of sensor density on the ML algorithm, we have derived a closed-form and compact approximation to

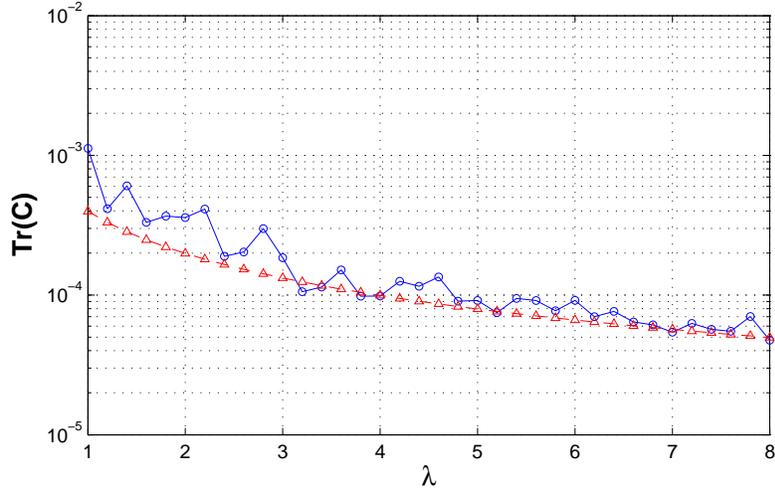


Figure 6.10: Trace of the CRLB matrix as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Exact CRLB; Dashed line + triangle: Approximate CRLB. SNR = 40dB and $M = 4$.

the CRLB based on the LLN for any smooth, differentiable and monotonically decreasing signal decay model. This closed-form result has helped us gain valuable insights into the relationship between the localization performance and the sensor density and SNR. Namely, the Fisher information is linear in the sensor density given the SNR. Numerical experiments have been conducted and the effectiveness of the approximation has been verified. A quantizer design method was also proposed, which determines the thresholds based on the main result of this chapter, namely the closed-form CRLB. The simulation results have shown us that the proposed quantizer achieves better performance than the uniform quantizer in general. Throughout this chapter, we assumed that the noises at the local sensors are *i.i.d.*. In practice, however, the sensor noises are often correlated, due to interference from common sources. Hence, our future research will take correlated noises into consideration. Further, the current work has only explored the problem when only one target exists in the ROI. One may be more interested in localizing multiple targets in the ROI. Hence, research work will

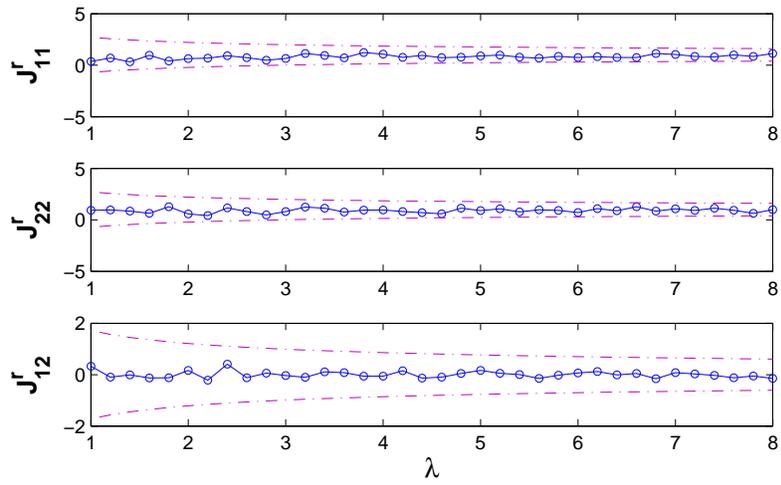


Figure 6.11: Relative FIM as a function of sensor density λ (Power Law Decay Model). Solid line + circle: Relative FIM; Dashed line: $1\text{-}\sigma$ region of the Relative FIM. SNR = 40dB and $M = 4$.

be extended to multi-target scenarios.

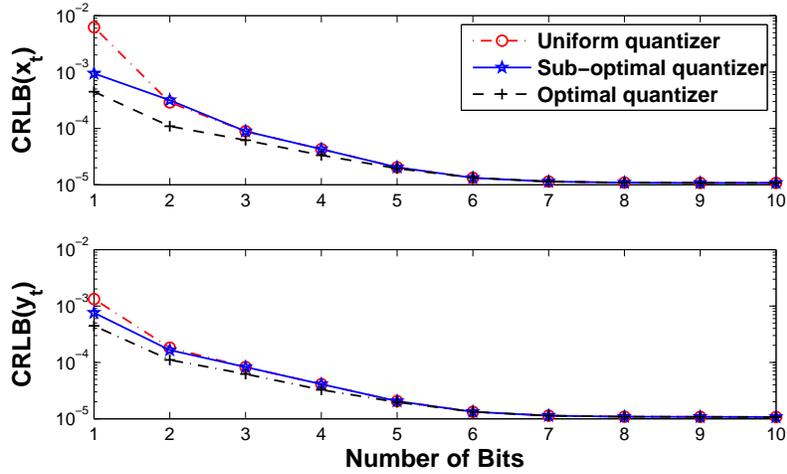


Figure 6.12: Performances comparison for different quantizers (Gaussian-like Decay Model).

SNR = 40dB and $\lambda = 6$

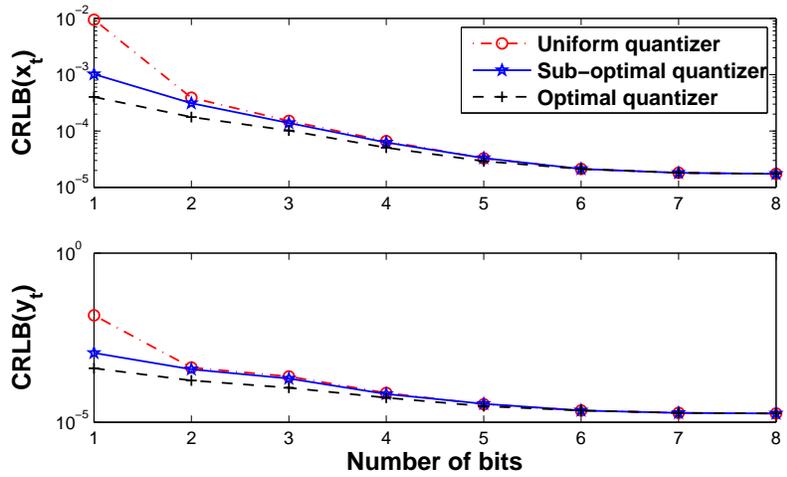


Figure 6.13: Performances comparison for different quantizers (Power Law Decay Model).

SNR = 40dB and $\lambda = 4$

Chapter 7

An Alternative Approach to Compute Conditional Posterior Cramér Rao Lower Bound for Nonlinear Filtering

7.1 Motivaton

The posterior (or Bayesian) Cramér-Rao lower bound (PCRLB, or BCRLB) is defined to be the inverse of the Fisher information matrix (FIM)¹ for a random vector [11] and provides a bound on the performance of estimators of the system state for filtering problems. In [12], Tichavsky et al. provided a recursive approach for calculating the sequential PCRLB for a general multi-dimensional discrete-time nonlinear filtering problem. The predictive and smoothing Cramér-Rao lower bounds for discrete-time nonlinear dynamic systems and their relations with filtering CRLB were discussed in [97]. However, the useful measurement information is averaged out making the unconditional PCRLB [12] [97] an off-line bound which is independent of the current state trajectory. Therefore, the unconditional PCRLB

¹FIM refers to the Bayesian information matrix which is used for random vectors.

does not reflect the nonlinear filtering performance for a particular system state realization very faithfully.

There are several modified versions of the PCRLB proposed in the literature, which are motivated by making the bound adaptive to the realization of the system state, so that it can be useful for online sensor management. In [98], a renewal strategy was used to restart the recursive unconditional PCRLB evaluation process, where the initial time is reset to a more recent past time, such that the prior knowledge of the initial system state is more useful and relevant to the sensor management problem. Therefore, the resulting PCRLB is conditioned on the measurements up to the reset time. Based on the PCRLB modified in this manner, a sensor deployment approach was developed to achieve better tracking accuracy with the efficient use of limited sensor resources. When a particle filter is used in the renewal strategy, the posterior probability density function (pdf) of the system state at the reset initial time is represented nonparametrically by a set of particles, from which it is difficult to derive the exact FIM. One may use Gaussian approximation as was done in [98], and then the FIM at the reset time can be taken as the inverse of the empirical covariance matrix estimated based on the particles. However, this approach may incur large errors and discrepancy, especially in a highly nonlinear and non-Gaussian system. Another modified version of the PCRLB, motivated by the problem of adaptive radar waveform design for target tracking, has been presented in [14]. The authors in [14] consider a linear Gaussian state dynamic model and a nonlinear measurement model, and propose to retain the unconditional recursive PCRLB derived in [12] with the exception of one term which corresponds to the contribution of the future measurements to the Fisher Information. The term with the future measurement contribution is modified in a heuristic way so that it includes the measurement history. Although the proposed method was shown to result in good performance for adaptive waveform design, the authors did not provide any theoretical

justification for this modification. Most recently, the notion of conditional PCRLB was introduced in [13], which was shown to be different from the modified PCRLBs proposed in [98] and [14]. The conditional PCRLB proposed in [13] provides a bound on the conditional mean squared error (MSE) of the system state estimate, based on the measurements up to the current time. Furthermore, the authors in [13] proposed a systematic recursive approach based on a certain approximation to evaluate the conditional PCRLB.

Our contributions in this chapter are as follows. We propose a new conditional PCRLB, which is based on the representation of the conditional PCRLB proposed in [13]. We call this bound the alternative conditional PCRLB (A-CPCRLB), since we discard the auxiliary FIM which is involved in the recursive update for the conditional PCRLB presented in [13]. Instead, an alternative approximate recursive update is proposed, which is direct, more compact and efficient than the one proposed in [13]. Furthermore, when the state dynamic model is linear and Gaussian, we show that this bound reduces to the modified PCRLB proposed in [14]. Hence, the proposed A-CPCRLB provides a generalization and theoretical justification for the one used in [14]. The analytical calculation of our proposed bound is not tractable except for very restricted cases such as linear Gaussian systems. For this reason, numerical computation methods such as the sequential Monte Carlo methods, i.e., particle filters [99] are used. We provide performance analysis in terms of computational complexities associated with the computation of the bound. A numerical example is provided to compare the original CPCRLB [13] with our proposed bound, namely the A-CPCRLB. For this particular example, we observe that the results are quite similar. Here, after presenting derivations of the proposed bounds, we provide a numerical example to compare the original CPCRLB [13] with our proposed bound, namely the A-CPCRLB. For this particular example, we observe that the results are quite similar.

7.2 Conditional Posterior Cramér-Rao Lower Bounds

Consider a n_x -dimensional state vector at time k , \mathbf{x}_k , whose discrete time dynamics is defined by

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k) \quad (7.1)$$

where $f_k : \mathfrak{R}^{n_x} \times \mathfrak{R}^{n_u} \rightarrow \mathfrak{R}^{n_x}$ and \mathbf{u}_k is the independent identically distributed (i.i.d.) process noise with dimension n_u . The measurement model is given by

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{v}_k), \quad (7.2)$$

where $h_k : \mathfrak{R}^{n_x} \times \mathfrak{R}^{n_v} \rightarrow \mathfrak{R}^{n_z}$, \mathbf{v}_k is the i.i.d. measurement noise, n_z and n_v are the dimensions of the measurement and measurement noise vectors, respectively. The process and the measurement noise distributions are denoted by $p_{\mathbf{u}_k}(\mathbf{u})$ and $p_{\mathbf{v}_k}(\mathbf{v})$, respectively. It is assumed that the estimator has complete information about the state dynamic model (7.1), the sensor measurement model (7.2) and the process and measurement noise distributions.

The conditional PCRLB sets a bound on the performance of estimating the state vector up to time $k + 1$, $\mathbf{x}_{0:k+1}$, when the new measurement \mathbf{z}_{k+1} becomes available given that the past measurements up to time k , $\mathbf{z}_{1:k}$, are all known, *i.e.*, the measurements up to time k are taken as realizations rather than random vectors. The sequence of conditional Fisher information $\{L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k})\}$ for estimating state vector $\{\mathbf{x}_{k+1}\}$ given the measurements up to time k can be computed as follows [13]:

$$L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}) = B_k^{22} - B_k^{21}[B_k^{11} + L_A(\mathbf{x}_k|\mathbf{z}_{1:k})]^{-1}B_k^{12} \quad (7.3)$$

where

$$B_k^{11} = E_{p_{\mathbf{u}_k}} \{-\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p(\mathbf{x}_{k+1}|\mathbf{x}_k)\} \quad (7.4)$$

$$B_k^{12} = E_{p_{\mathbf{u}_k}} \{-\Delta_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \ln p(\mathbf{x}_{k+1}|\mathbf{x}_k)\} = (B_k^{21})^T \quad (7.5)$$

$$B_k^{22} = E_{p_{k+1}^c} \{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} [\ln p(\mathbf{x}_{k+1}|\mathbf{x}_k) + \ln p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})] \} \quad (7.6)$$

and

$$L_A(\mathbf{x}_k|\mathbf{z}_{1:k}) = A_k^{22} - A_k^{21}(A_k^{11})^{-1}A_k^{12} \quad (7.7)$$

$$p_{k+1}^c \triangleq p(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1}|\mathbf{z}_{1:k}) \quad (7.8)$$

with

$$A_k^{11} = E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left[-\Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) \right]$$

$$\begin{aligned} A_k^{12} &= E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left[-\Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_k} \ln p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) \right] \\ &= (A_k^{21})^T \end{aligned}$$

$$A_k^{22} = E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left[-\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}) \right]$$

An approximate recursion to compute $L_A(\mathbf{x}_k|\mathbf{z}_{1:k})$ is also proposed in [13], which is

$$L_A(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx S_k^{22} - S_k^{21}[S_k^{11} + L_A(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})]^{-1}S_k^{12} \quad (7.9)$$

where

$$S_k^{11} = E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left[-\Delta_{\mathbf{x}_{k-1}}^{\mathbf{x}_{k-1}} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1}) \right] \quad (7.10)$$

$$S_k^{12} = E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left[-\Delta_{\mathbf{x}_{k-1}}^{\mathbf{x}_k} \ln p(\mathbf{x}_k|\mathbf{x}_{k-1}) \right] = (S_k^{21})^T \quad (7.11)$$

$$S_k^{22} = E_{p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})} \left\{ -\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} [\ln p(\mathbf{x}_k|\mathbf{x}_{k-1}) + \ln p(\mathbf{z}_k|\mathbf{x}_k)] \right\} \quad (7.12)$$

One should note that if (7.9) is used in (7.3), then (7.3) becomes an approximation, instead of equality.

7.3 Proposed A-CPCRLB

As shown in Section 7.2, an approximated recursive update of an auxiliary FIM is necessary to recursively compute the conditional Fisher information at each time step, which makes

the process complex. In this section, an alternative compact way is proposed, such that the process of computation is simplified and computation time can be saved at the same time.

Proposition 7.1. *The sequence of conditional Fisher information $L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k})$ for estimating state vectors \mathbf{x}_{k+1} given measurements up to time k can be computed as follows:*

$$L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}) \approx B_k^{22} - B_k^{21}[B_k^{11} + L(\mathbf{x}_k|\mathbf{z}_{1:k-1})]^{-1}B_k^{12} \quad (7.13)$$

where B_k^{11} , B_k^{12} , B_k^{21} and B_k^{22} are given by (7.4) through (7.6).

Proof. Since $p_k^c = p(\mathbf{x}_{0:k}, \mathbf{z}_k|\mathbf{z}_{1:k-1})$ according to (7.8), the conditional FIM given measurements up to time $k - 1$ can be decomposed as follows:

$$\begin{aligned} I(\mathbf{x}_{0:k}, \mathbf{z}_k|\mathbf{z}_{1:k-1}) &\triangleq \begin{bmatrix} O_k & P_k \\ P_k^T & Q_k \end{bmatrix} \\ &= \begin{bmatrix} E\{-\Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p_k^c\} & E\{-\Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_k} \ln p_k^c\} \\ E\{-\Delta_{\mathbf{x}_k}^{\mathbf{x}_{0:k-1}} \ln p_k^c\} & E\{-\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p_k^c\} \end{bmatrix} \end{aligned} \quad (7.14)$$

Thus, by applying matrix inversion formula [100], the inverse of the lower-right block of $I^{-1}(\mathbf{x}_{0:k}, \mathbf{z}_k|\mathbf{z}_{1:k-1})$, *i.e.*, the FIM for estimating \mathbf{x}_k given the measurements up to $k - 1$ is

$$L(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = Q_k - P_k^T O_k^{-1} P_k$$

Now, considering Fisher information given measurements up to time k , we have

$$\begin{aligned} &I(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1}|\mathbf{z}_{1:k}) \\ &= E \left[-1 \begin{bmatrix} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} & \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_k} & \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{k+1}} \\ \Delta_{\mathbf{x}_k}^{\mathbf{x}_{0:k-1}} & \Delta_{\mathbf{x}_k}^{\mathbf{x}_k} & \Delta_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \\ \Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{0:k-1}} & \Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_k} & \Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \end{bmatrix} \ln p_{k+1}^c \right] \end{aligned} \quad (7.15)$$

where p_{k+1}^c is defined in (7.8), which can be decomposed as follows:

$$\begin{aligned}
& p(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1} | \mathbf{z}_{1:k}) \\
&= p(\mathbf{x}_{k+1}, \mathbf{z}_{k+1} | \mathbf{x}_{0:k}, \mathbf{z}_k, \mathbf{z}_{1:k-1}) p(\mathbf{x}_{0:k} | \mathbf{z}_k, \mathbf{z}_{1:k-1}) \\
&= p(\mathbf{z}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{x}_k) \frac{p(\mathbf{x}_{0:k}, \mathbf{z}_k | \mathbf{z}_{1:k-1})}{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})}
\end{aligned} \tag{7.16}$$

Therefore,

$$\begin{aligned}
\ln p_{k+1}^c &= \ln p(\mathbf{z}_{k+1} | \mathbf{x}_{k+1}) + \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k) \\
&\quad + \ln p(\mathbf{x}_{0:k}, \mathbf{z}_k | \mathbf{z}_{1:k-1}) - \ln p(\mathbf{z}_k | \mathbf{z}_{1:k-1}) \\
&= \ln p(\mathbf{z}_{k+1} | \mathbf{x}_{k+1}) + \ln p(\mathbf{x}_{k+1} | \mathbf{x}_k) + \ln p_k^c \\
&\quad - \ln p(\mathbf{z}_k | \mathbf{z}_{1:k-1})
\end{aligned}$$

Hence,

$$\begin{aligned}
& I(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1} | \mathbf{z}_{1:k}) \\
&= \begin{bmatrix} -E_{p_{k+1}^c} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p_k^c & -E_{p_{k+1}^c} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_k} \ln p_k^c & 0 \\ -E_{p_{k+1}^c} \Delta_{\mathbf{x}_k}^{\mathbf{x}_{0:k-1}} \ln p_k^c & -E_{p_{k+1}^c} \Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p_k^c + B_k^{11} & B_k^{12} \\ 0 & B_k^{21} & B_k^{22} \end{bmatrix}
\end{aligned}$$

where B_k^{ij} , $i = 1, 2$, $j = 1, 2$ are defined in (7.4) through (7.6). Since the top-left submatrix is a function of \mathbf{z}_k , we can approximate it by its expectation with respect to $p(\mathbf{z}_k | \mathbf{z}_{1:k-1})$.

Then,

$$\begin{aligned}
& -E_{p_{k+1}^c} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p_k^c \\
&\approx -E_{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})} E_{p_{k+1}^c} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p_k^c \\
&= - \int p(\mathbf{z}_k | \mathbf{z}_{1:k-1}) p_{k+1}^c \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_{0:k-1}} \ln p_k^c d\mathbf{x}_{0:k+1} d\mathbf{z}_{k+1} d\mathbf{z}_k
\end{aligned} \tag{7.17}$$

$$\stackrel{(a)}{=} O_k \tag{7.18}$$

Note that (a) follows from plugging (7.16) in (7.17), and using the definition in (7.14).

Similarly,

$$\begin{aligned} -E_{p_{k+1}^c} \Delta_{\mathbf{x}_{0:k-1}}^{\mathbf{x}_k} \ln p_k^c &\approx P_k \\ -E_{p_{k+1}^c} \Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p_k^c &\approx Q_k \end{aligned}$$

Hence,

$$I(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1} | \mathbf{z}_{1:k}) \approx \begin{bmatrix} O_k & P_k & 0 \\ P_k^T & Q_k + B_k^{11} & B_k^{12} \\ 0 & B_k^{21} & B_k^{22} \end{bmatrix}$$

The conditional Fisher information $L(\mathbf{x}_{k+1} | \mathbf{z}_{1:k})$ is equal to the inverse of the lower right submatrix of $I^{-1}(\mathbf{x}_{0:k+1}, \mathbf{z}_{k+1} | \mathbf{z}_{1:k})$. Therefore, according to the matrix inversion formula,

$$\begin{aligned} L(\mathbf{x}_{k+1} | \mathbf{z}_{1:k}) &\approx B_k^{22} - [0 \quad B_k^{21}] \begin{bmatrix} O_k & P_k \\ P_k^T & Q_k + B_k^{11} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ B_k^{12} \end{bmatrix} \\ &= B_k^{22} - B_k^{21} (B_k^{11} + (Q_k - P_k^T O_k^{-1} P_k))^{-1} B_k^{12} \\ &= B_k^{22} - B_k^{21} (B_k^{11} + L(\mathbf{x}_k | \mathbf{z}_{1:k-1}))^{-1} B_k^{12} \end{aligned}$$

■

Based on Proposition 7.1, it is easy to show that the modified PCRLB in [14] is a special case of the A-CPCRLB, as stated in the following corollary.

Corollary 7.2. *For the particular case of linear state model with additive Gaussian noise, i.e., $\mathbf{x}_{k+1} = F_k \mathbf{x}_k + \mathbf{u}_k$, the conditional Fisher information $L(\mathbf{x}_{k+1} | \mathbf{z}_{1:k})$ is given by*

$$\begin{aligned} L(\mathbf{x}_{k+1} | \mathbf{z}_{1:k}) &\approx (Q_k + F_k (L(\mathbf{x}_k | \mathbf{z}_{1:k-1}))^{-1} F_k^T)^{-1} + \\ &E_{p_{k+1}^c} \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1} | \mathbf{x}_{k+1}) \right\} \end{aligned} \quad (7.19)$$

where F_k is the state transition matrix of the state process equation, and Q_k is the covariance matrix of the additive Gaussian noise u_k .

Proof. Since we consider a linear state model with additive Gaussian noise,

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k) = \frac{1}{(2\pi)^{\frac{n_x}{2}} |Q_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} [\mathbf{x}_{k+1} - F_k \mathbf{x}_k]^T Q_k^{-1} [\mathbf{x}_{k+1} - F_k \mathbf{x}_k] \right\}$$

Taking the logarithm of the probability density function (pdf) above, we can get

$$-\ln p(\mathbf{x}_{k+1}|\mathbf{x}_k) = c_0 + \frac{1}{2} [\mathbf{x}_{k+1} - F_k \mathbf{x}_k]^T Q_k^{-1} [\mathbf{x}_{k+1} - F_k \mathbf{x}_k]$$

where c_0 is a constant. Then, it is straightforward to get

$$B_k^{11} = F_k^T Q_k^{-1} F_k \quad (7.20)$$

$$B_k^{12} = -F_k^T Q_k^{-1} \quad (7.21)$$

$$B_k^{22} = Q_k^{-1} + E_{p_{k+1}^c} \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) \right\} \quad (7.22)$$

Therefore, using (7.13), we have

$$\begin{aligned} L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}) &= Q_k^{-1} + E_{p_{k+1}^c} \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) \right\} \\ &\quad - Q_k^{-1} F_k (F_k^T Q_k^{-1} F_k + L(\mathbf{x}_k|\mathbf{z}_{1:k-1}))^{-1} F_k^T Q_k^{-1} \\ &= (Q_k + F_k L^{-1}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) F_k^T)^{-1} \\ &\quad + E_{p_{k+1}^c} \left\{ -\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) \right\} \end{aligned} \quad (7.23)$$

where the last equation is due to the application of Woodbury matrix identity [101]:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(VA^{-1}U + C^{-1})^{-1}VA^{-1} \quad (7.24)$$

where A, U, C, V are matrices with proper dimensions. ■

One should note that the result in Corollary 7.2 is the same as the one used in [14]. Hence, the approximation in [14] is a special case of Proposition 7.1. Moreover, since the bound proposed in [14] is a heuristic one and not theoretically justified therein, Proposition 7.1 and Corollary 7.2 in this work provide it a theoretical justification.

Obviously, (7.13) is more compact than (7.3), since the conditional PCRLB is directly updated at each recursion in (7.13) without using the auxiliary FIM $L_A(\mathbf{x}_k|\mathbf{z}_{1:k})$. The computational efficiency of the A-CPCRLB will be analyzed in Section 7.4 in detail.

Another useful insight that can be deduced from Proposition 1 is that in a linear and Gaussian system, A-CPCRLB is identical to the offline PCRLB.

Corollary 7.3. *For the particular case of linear Gaussian dynamic model: $\mathbf{x}_{k+1} = F_k \mathbf{x}_k + \mathbf{u}_k$, $\mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{v}_k$, where \mathbf{u}_k and \mathbf{v}_k are Gaussian noises with covariance matrices Q_k and R_k respectively, the recursive conditional Fisher information (7.13) in Proposition 7.1 is the same as the recursive offline Fisher information J_{k+1} proposed in [12], i.e., $J_{k+1} = L(\mathbf{x}_{k+1}|\mathbf{z}_{1:k})$, given that $J_0 = L(\mathbf{x}_0)$.*

Proof. From Corollary 7.2, we already have (7.20), (7.21), and (7.22). Given the linear Gaussian observation model, we have

$$p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) = \frac{1}{(2\pi)^{\frac{n_z}{2}} |R_{k+1}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} [\mathbf{z}_{k+1} - H_{k+1} \mathbf{x}_{k+1}]^T R_{k+1}^{-1} [\mathbf{z}_{k+1} - H_{k+1} \mathbf{x}_{k+1}] \right\}$$

Then,

$$\begin{aligned} & E_{p_{\hat{\mathbf{x}}_{k+1}}^c} \left\{ -\Delta_{\hat{\mathbf{x}}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) \right\} \\ &= E_{p_{\hat{\mathbf{x}}_{k+1}}^c} \left\{ H_{k+1}^T R_{k+1}^{-1} H_{k+1} \right\} \\ &= H_{k+1}^T R_{k+1}^{-1} H_{k+1} \end{aligned}$$

Thus,

$$B_k^{22} = Q_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}$$

In [12], it has been shown that $J_{k+1} = D_k^{22} - D_k^{21}(J_k + D_k^{11})^{-1}D_k^{12}$, where $D_k^{11} = F_k^T Q_k^{-1} F_k$, $D_k^{12} = -F_k^T Q_k^{-1} = (D_k^{21})^T$, and $D_k^{22} = Q_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}$ when the system is linear and Gaussian. Hence, $J_{k+1} = L(\mathbf{x}_{k+1} | \mathbf{z}_{1:k})$, given the same initialization, *i.e.*, $J_0 = L(\mathbf{x}_0)$. ■

One should note that we use the same particle filter based method as that given in [13] to compute the A-CPCRLB. The details are not given here for the sake of brevity. Interested readers are referred to [13] for more information.

7.4 Computational Complexity

In this section, we analyze the computational complexity of the proposed bounds based on the total number of floating-point operations (*flops*).

The exact flops required for the derivative operations in (7.4)-(7.12) depend on the structures of the pdfs $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ and $p(\mathbf{z}_k | \mathbf{x}_k)$ and there is no universal count. Due to this dependence, we define new notations to represent the flops for these derivative operations: $\Delta_1^s \triangleq \text{fl}(\frac{\partial p(\mathbf{x}_{k+1} | \mathbf{x}_k)}{\partial x_{k+1_i}})$, $\Delta_2^s \triangleq \text{fl}(\frac{\partial^2 p(\mathbf{x}_{k+1} | \mathbf{x}_k)}{\partial x_{k+1_i} \partial x_{k+1_j}})$, $\Delta_3^s \triangleq \text{fl}(\frac{\partial p(\mathbf{x}_{k+1} | \mathbf{x}_k)}{\partial x_{k_i}})$, $\Delta_4^s \triangleq \text{fl}(\frac{\partial^2 p(\mathbf{x}_{k+1} | \mathbf{x}_k)}{\partial x_{k_i} \partial x_{k_j}})$, $\Delta_5^s \triangleq \text{fl}(\frac{\partial^2 p(\mathbf{x}_{k+1} | \mathbf{x}_k)}{\partial x_{k+1_i} \partial x_{k_j}})$, $\Delta_1^z \triangleq \text{fl}(\frac{\partial p(\mathbf{z}_k | \mathbf{x}_k)}{\partial \mathbf{x}_{k_i}})$, $\Delta_2^z \triangleq \text{fl}(\frac{\partial^2 p(\mathbf{z}_k | \mathbf{x}_k)}{\partial \mathbf{x}_{k_i} \partial \mathbf{x}_{k_j}})$, $1 \leq i, j \leq n_x$, where $\text{fl}(\cdot)$ represents the number of flops required for a given operation. When defining these notations, for simplicity, we have made the implicit assumption that the derivatives with respect to different elements of the state vector require the same number of flops. In the following calculations, we also assume that each particle has a non-identical weight, *i.e.*, there is no resampling.

We start with the calculation of flops required for the original CPCRLB in [13]. Note that the term B_k^{22} has two terms. Let us denote the first and the second terms in (7.6) as

$B_k^{22,a}$ and $B_k^{22,b}$, respectively. Then, the flops of the CPCRLB, *i.e.*, (7.3), can be represented as:

$$\begin{aligned} \text{fl}(L_{k+1}) = & \text{fl}(B_k^{22,a}) + \text{fl}(B_k^{22,b}) + \text{fl}(B_k^{12}) + \text{fl}(B_k^{11}) \\ & + \text{fl}(S_k^{11}) + \text{fl}(S_k^{12}) + \text{fl}(S_k^{22}) + \mathcal{O}(n_x^3). \end{aligned} \quad (7.25)$$

where $\mathcal{O}(n_x^3)$ represents the computational complexity associated with matrix inversions, multiplications and summations involved in (7.3) and (7.9). For matrix inversion, the exact flop count depends on the matrix and the specific technique used for inversion. Nevertheless, the flop count required for matrix inversion can be expressed as $\mathcal{O}(n_x^3)$ [102], which also subsumes the flops required for matrix multiplications and summations in (7.3) and (7.9). From the particle based computation in [13], we can calculate the flops required for the B terms and S terms as:

$$\text{fl}(B_k^{22,a}) = \frac{3}{2}Mn_x^2 + Mn_x\Delta_1^s + \frac{3}{2}Mn_x + 2M - 1 \quad (7.26)$$

$$\text{fl}(B_k^{12}) = \frac{3}{2}Mn_x^2 + Mn_x\Delta_3^s + \frac{3}{2}Mn_x + 2M - 1 \quad (7.27)$$

$$\text{fl}(B_k^{11}) = \frac{3}{2}Mn_x^2 + \frac{3}{2}Mn_x + 2M - 1 \quad (7.28)$$

$$\text{fl}(S_k^{11}) = \left(\frac{5}{2} + \frac{\Delta_4^s}{2}\right)Mn_x^2 + \left(\frac{5}{2} + \Delta_3^s + \frac{\Delta_4^s}{2}\right)Mn_x + 2M - 1 \quad (7.29)$$

$$\text{fl}(S_k^{12}) = (5 + \Delta_5^s)Mn_x^2 + \Delta_1^s Mn_x + M - 1 \quad (7.30)$$

$$\text{fl}(S_k^{22}) = \left(\frac{5}{2} + \frac{\Delta_2^z}{2}\right)Mn_x^2 + \left(\frac{5}{2} + \Delta_1^z + \frac{\Delta_2^z}{2}\right)Mn_x + 2M - 1 \quad (7.31)$$

By carefully investigating the computations required for the CPCRLB in [13], *i.e.*, expression in (7.3) and (7.9), and that of the A-CPCRLB, *i.e.*, expression in (7.13), we note that the extra computation of the CPCRLB at time k comes from the computations of S_k^{11} , S_k^{12} and S_k^{22} , as well as from a matrix inversion, multiplication and subtraction in equation (7.9).

Therefore, the A-CPCRLB saves the following number of flops, compared to the CPCRLB in [13]:

$$\begin{aligned}
fl_A^{save} &= fl(S_k^{22}) + fl(S_k^{11}) + fl(S_k^{12}) + \mathcal{O}(n_x^3) \\
&= (10 + \frac{\Delta_4^s}{2} + \Delta_5^s + \frac{\Delta_2^z}{2})Mn_x^2 + 5M + \mathcal{O}(n_x^3) \\
&\quad + (5 + \Delta_3^s + \frac{\Delta_4^s}{2} + \Delta_1^s + \Delta_1^z + \frac{\Delta_2^z}{2})Mn_x
\end{aligned} \tag{7.32}$$

From the above expression, it is clear that the savings are significant especially for large M and/or n_x .

7.5 Numerical Results

In this section, we consider the univariate nonstationary growth model (UNGM), which is a highly nonlinear and bimodal model, and perform a number of numerical experiments to compare the performance of the following bounds: 1) Offline PCRLB [12], 2) Conditional PCRLB proposed in [13], 3) A-CPCRLB, 4) D-CPCRLB proposed in [103], 5) D-CPCRLB with Gaussian approximation. The UNGM has been widely used in the nonlinear tracking literature as a benchmark problem [13, 62, 104]. The dynamic state space equations for a UNGM are given by

$$x_{k+1} = \alpha x_k + \beta \frac{x_k}{1 + x_k^2} + \gamma \cos(1.2k) + u_k \tag{7.33}$$

$$z_k = \kappa x_k^2 + v_k \tag{7.34}$$

where u_k and v_k are zero mean white Gaussian with variances σ_u^2 and σ_v^2 , respectively. The conditional mean-squared error (MSE) is calculated as follows. At time k , the posterior pdf is computed by the particle filter given the measurements up to time k . Then, 1000 independent realizations of z_{k+1} are generated according to (7.34), and the conditional MSE,

(*i.e.*, $\text{MSE}(\hat{x}_{k+1}|z_{1:k})$), is obtained based on the 1000 Monte Carlo runs. A single realization of z_{k+1} is randomly picked among the 1000 realizations above, and concatenated with the past measurement history to form $z_{1:k+1}$. The particles and weights corresponding to this particular z_{k+1} are stored and used for the iteration at time $(k + 1)$.

7.5.1 Highly Nonlinear Case

In this experiment, we set the parameters for UNGM as $\alpha = 1$, $\beta = 15$, $\gamma = 8$, $\sigma_u^2 = 4$, $\sigma_v^2 = 1$ and $\kappa = 1/20$ to make it highly nonlinear. Since it is difficult for conventional methods such as the Kalman Filter or the extended Kalman Filter to track the state when the system model is highly nonlinear, a particle filter is applied in the simulation, and its performance is illustrated in Figure 7.1(a).

In Figure 7.1(b), the conditional PCRLB in [13], A-CPCRLB derived in Section 7.3 and D-CPCRLB as well as the conditional MSE are plotted as functions of time. We can observe that all the three CPCRLBs follow the trends of the conditional MSE more faithfully than the (offline) PCRLB. It can also be observed that the original conditional PCRLB, A-CPCRLB and D-CPCRLB almost overlap with each other everywhere.

7.5.2 Weakly Nonlinear Case

In order to show that the proposed recursive update procedure is independent of the nonlinearity of the state process model, a weakly nonlinear example is provided in this section. Here, we set $\beta = 1$, resulting in a much smaller nonlinear component in the state process equation. The variance of the process noise is set as $\sigma_u^2 = 1$, smaller than the highly nonlinear case. We set the measurement noise variance $\sigma_v^2 = 0.01$, such that the signal-to-noise ratio (SNR) is high for the observation. Other parameters are kept the same as in Section 7.5.1. Similar to Figure 7.1(b), Figure 7.2 illustrates that there is almost no difference between the

original conditional PCRLB and the proposed bounds, and all the three CPCRLBs follow the trends of the conditional MSE more faithfully than the (offline) PCRLB. We can also observe that the convergence from the conditional MSE to the conditional PCRLB is better in Figure 7.2 than that in Figure 7.1(b), due to the weak nonlinearity in the former case.

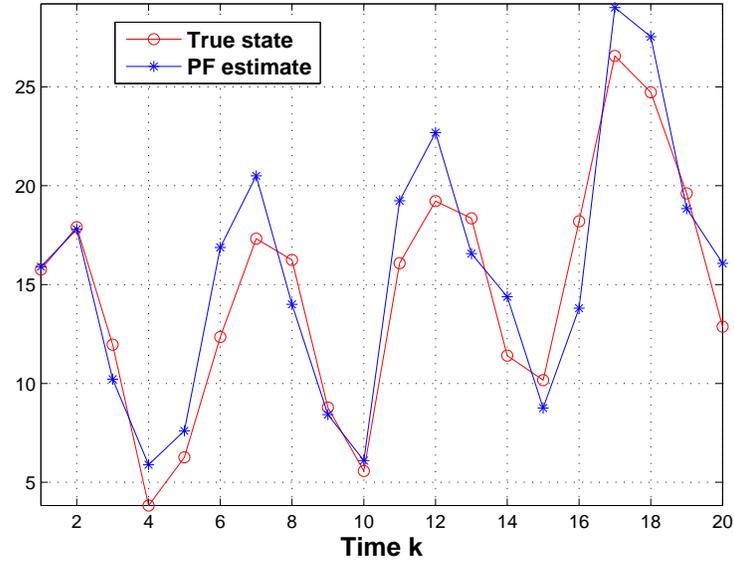
7.5.3 Discussion

In our particular examples shown in Figure 7.1(b) and Figure 7.2, it can be observed that the offline PCRLB is more optimistic than the conditional PCRLBs. It should be mentioned that this is not always the case, since for some specific realizations of the system state, the conditional PCRLB may result in values that are smaller than the unconditional PCRLB. This is due to the fact that the online (conditional) bounds depend on a specific realization of the system state and they provide bounds for that specific conditional MSE. Figure 7.1(b) and Figure 7.2 each depicts one particular realization, i.e., single Monte Carlo run.

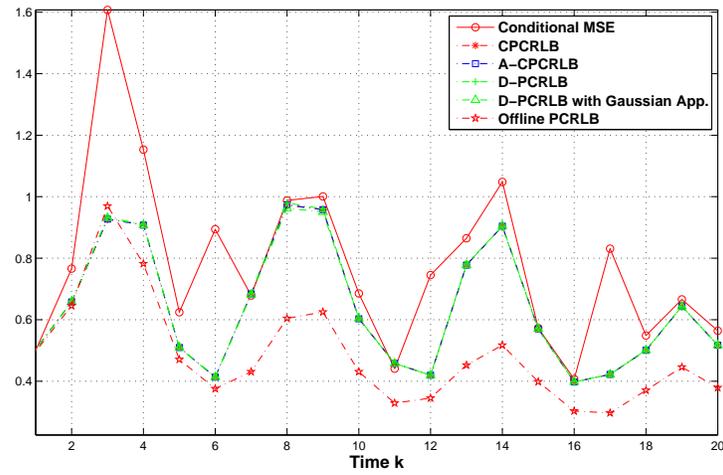
Simulation results show that the conditional bounds proposed in this chapter are almost the same as the original one proposed in [13]. Since the original CPCRLB in [13] and the A-CPCRLB both use approximations at each iteration, it is possible that the error due to approximations accumulates over time. It is difficult to perform an exact error comparison between the two recursive procedures. Nevertheless, the following intuitive analysis is helpful to interpret the simulation results. In [13], the approximation is only applied to the top left block of the auxiliary FIM in (7.7), while in A-PCRLB the approximation is applied to four blocks of the conditional FIM in (7.15). However, in [13], the approximated block is involved in three inversions to complete the update at each iteration, which makes the approximation propagate to all the elements of the conditional PCRLB. Therefore, both approximations result in almost the same order of errors, which explains why the gap between the two corresponding lower bounds is negligible as shown by numerical examples.

7.6 Summary

In this work, an alternative conditional PCRLB has been proposed, namely the A-CPCRLB for nonlinear sequential Bayesian estimation. It achieves almost the same performance as the conditional PCRLB proposed in [13] as demonstrated by numerical examples. The proposed A-CPCRLB is more compact and more efficient than the one in [13]. The computational complexity of both the A-CPCRLB and the original PCRLB is linear in the number of particles.



(a) Plot of true state x_k and its estimate \hat{x}_k by the particle filter



(b) Plot of offline PCRLB, CPCRLBs and conditional MSE

Figure 7.1: Highly Nonlinear Case

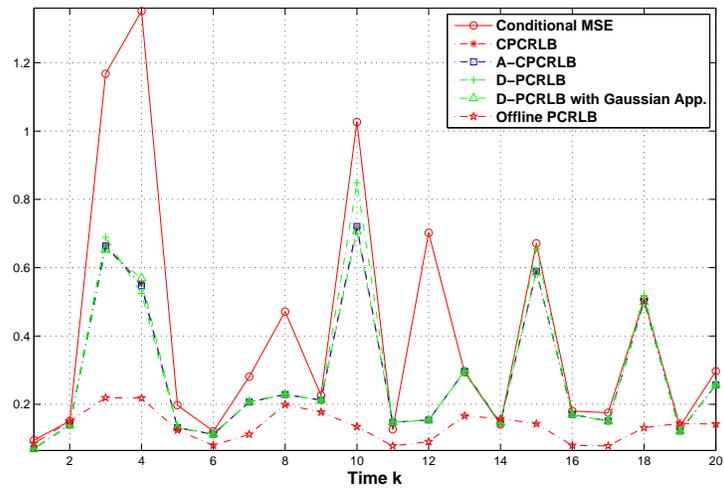


Figure 7.2: Weakly Nonlinear Case

Chapter 8

Conclusions and Future Directions

8.1 Conclusions

This dissertation investigated several distributed inference problems in resource-constrained sensor networks. Both static parameter estimation and dynamic state estimation have been considered. Due to the severe resource constraints, namely computation, energy and/or bandwidth, in sensor networks, we have proposed several methods that can be utilized either at the local sensors or at the fusion center or both to improve distributed inference performance.

We proposed a novel data selection and fusion algorithm for sequential Bayesian estimation which involves the participation of both sensors and the fusion center when the resources in the sensor networks are limited. To be more specific, this algorithm consists of two stages: at the first stage, each local sensor censors its observation, and sends it to the fusion center only if the observation is “informative” enough; at the second stage, the fusion center fuses both the received observations and the missing ones to infer the system state. We call this novel algorithm censoring and fusion with missing data (CFwMD), which essentially takes advantage of the fact that local censoring procedure can pick up more “informative” data

given limited energy and/or bandwidth constraints and the fusion rule at the fusion can utilize the missing information conveyed by the missing observations. It has been shown that the proposed algorithm attains performance that is very close to the benchmark scenario where all the sensors send their observations to the fusion center while saving much bandwidth. It has also been shown that, under the same bandwidth constraint, the proposed algorithm outperforms the one where observations are randomly selected and the one where missing observations are not fused.

In this dissertation, we developed an approach to solve the sensor management problem based on the theory of compressive sensing (CS). We employed a probabilistic transmission strategy and MACs, based on which, an equivalent standard CS representation was obtained. Therefore, the sensor management problem became the problem to determine the sensing matrix that is completely determined by the probabilities of transmission of sensors in the network which is obtained by solving a constrained optimization problem. Numerical results showed that the proposed methodology achieved comparable tracking performance to the case where all the sensors send their observations to the fusion center via parallel channels, but saves significant energy. Also, under the same energy constraint, the proposed methodology outperforms the random selection scheme significantly.

Furthermore, we considered the sensor management problem in terms of bit allocation for Bayesian parameter estimation. We proposed a methodology where quantized data from local sensors are not fused directly but independently corrupted with an additive Gaussian noise which performs as a low-pass filter in the characteristic function domain. The motivation of injecting independent noises is to recover the original raw analog measurements, inspired by Widrow's theory [10]. We derived the performance bound in terms of Fisher information matrix based on the new framework and proposed to allocate available bits to local sensors by optimizing the derived bound. We also showed that the new framework theoretically

justifies the additive noise modeling of quantization process, and therefore, facilitates the derivation of LMMSE.

The target localization problem with dense sensor deployment was also investigated in this dissertation. We considered a scenario where the number of sensors in the ROI is large, and quantized data are collected at the fusion center to locate the target. The relationship between the sensor density and the performance of the network in terms of Fisher information was analyzed. By using the LLN, we obtained a closed form solution of the Fisher information, which shows that the closed form Fisher information is a linearly increasing function of the sensor density, indicating that the more sensors the better. Furthermore, theoretical analysis of the error incurred by the approximation of LLN was carried out. Based on the closed form solution, we also designed a sub-optimal quantizer for local sensors, which was shown to attain similar performance as the optimal one.

To evaluate the online performance of target tracking, an alternative conditional PCRLB, namely, A-PCRLB was proposed in this dissertation. We showed that the proposed A-PCRLB is capable of utilizing the information of the measurements history, and therefore providing online performance evaluation for each realization of target tracking. Theoretical complexity analysis was provided to show that A-PCRLB is more computational efficient, compared to the original PCRLB proposed before in this field. Moreover, we have shown that the difference between A-PCRLB and the original PCRLB is negligible when plotted for any given target trajectory.

8.2 Suggestions for Future Work

Some promising directions for future work are listed in the following:

1. In Chapter 3, we intuitively provided a sensor censoring rule which depends on the

innovation. Though we have proved that it is equivalent to the KL-divergence criterion when the sensor measurement is a scalar, an overall optimal censoring rule in the sense of, say maximizing mutual information between the observation and the state, is desired. As a future work, one can investigate some information theory related metrics to derive the optimal per-sensor censoring rule for the WSN system.

2. The work in Chapter 3 was discussed under the assumption that the communication channels between the local sensors and the FC are perfect. However, this assumption is strong and ideal. In the future, one can study the effect of the communication channel statistics on the performance of the system, and develop a channel aware fusion rule at the FC.
3. In Chapter 3, we have shown that the proposed framework can only be generalized to a general nonlinear Gaussian system with feedback. For a non-feedback system, the framework cannot work since local sensors have no information about the state estimates of each other. In this case, a distributed scheme, such as a consensus algorithm, may be desired. But, how to explore the missing information conveyed by the missing data due to sensor censoring in the distributed network needs to be addressed, which is a promising future research direction.
4. Recently, Byzantine attacks on distributed detection in the WSN has become a hot topic [105] [106]. Simply speaking, Byzantine attack is the result of the existence of malicious sensors that send false information in the network. Discussions on estimation in the presence of Byzantine attack can be found in [107]. In our work thus far, sensors in the WSN are all assumed honest, *i.e.*, no malicious nodes exist in the WSN. However, the framework proposed in Chapter 3 is naturally attackable. For example, under the assumption that the communication channel is perfect, when an observation z_i from

sensor i is not received at the FC, the FC will automatically assume that z_i does not pass the censoring rule. However, if sensor i is malicious, even if z_i passes the censoring rule, it may not send it such that false information is gathered at the FC, and finally degrades the performance of the system. Therefore, as a future work, one can study the effect of the existence of malicious sensors on the inference performance, and explore robust solutions, where game theory can be a good tool.

5. In Chapter 4, we solved the sensor management problem using the theory of compressive sensing (CS). The projection matrix Φ was shown to be very sparse and provided good inference performance, but whether such a projection matrix can yield good recovery of the original signal is still in question. Therefore, one interesting future work is to analyze the restricted isometry property (RIP) of Φ to assess its recovery capability.
6. The result in Chapter 5 was derived by assuming that the injected noise is Gaussian. Even though such an assumption can make the derivation tractable, Gaussian noise incurs distortion when performing like a low-pass filter in the CF domain to filter out the signal due to its shape. Such distortion is definitely undesired when the filtered signal is used for inference. A promising research direction is to discard the Gaussian assumption and design an optimal low-pass filter which will not introduce any distortion or introduce a bounded distortion but is still a valid probability density function.

Bibliography

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38:393–422, 2002.
- [2] S. Joshi and S. Boyd. Sensor selection via convex optimization. *IEEE Trans. Signal Process.*, 57:451–462, Feb. 2009.
- [3] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42:251–260, February 2006.
- [4] W. Zhao, Y. Han, H. Wu, and L. Zhang. Weighted Distance Based Sensor Selection for Target Tracking in Wireless Sensor Networks . *IEEE Signal Process. Lett.*, 16(8):647–650, August 2009.
- [5] Y. Mo, R. Ambrosino, and B. Sinopoli. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*, 47:1330–1338, July 2011.
- [6] Eva A Riskin. Optimal bit allocation via the generalized BFOS algorithm. *IEEE Trans. Inf. Theory*, 37(2):400–402, March 1991.
- [7] E. Masazade, R. Niu, and P. K. Varshney. Dynamic bit allocation for object tracking in wireless sensor networks. *IEEE Trans. Signal Process.*, 60:5048–5063, October 2012.

- [8] D. Donoho. Compressed sensing. *IEEE Signal Processing Magazine*, pages 118–124, Jul 2007.
- [9] E. Candès and M. B. Wakin. An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine*, pages 21–30, Mar 2008.
- [10] B. Widrow and I. Kollár. *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*. Cambridge University Press, 2008.
- [11] H. L. Van Trees and K. L. Bell. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.
- [12] P. Tichavský, C. H. Muravchik, and A. Nehorai. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.*, 46(2):1386–1395, May 1998.
- [13] L. Zuo, R. Niu, and P. K. Varshney. Conditional posterior Cramér-Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.*, 59(1):1–14, January 2011.
- [14] M. Hurtado, T. Zhao, and A. Nehorai. Adaptive polarized waveform design for target tracking based on sequential Bayesian inference. *IEEE Trans. Signal Process.*, 56(3):1120–1133, March 2008.
- [15] G. Casella and R. L. Berger. *Statistical Inference*. Second Edition, 2002.
- [16] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley and Sons, 2001.

- [17] R. Niu and P. K. Varshney. Curvature nonlinearity measure and filter divergence detector for nonlinear tracking problems. In *11th International Conference on Information Fusion, 2008*, Cologne, Germany, July 2008.
- [18] M S Arulampalam, S Maskell, N Gordon, and T Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, Feb. 2002.
- [19] H. L. Van Trees. *Detection, Estimation and Modulation Theory*, volume 1. Wiley, New York, 1968.
- [20] P. Tichavsky, C.H. Muravchik, and A. Nehorai. Posterior Cramé-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.*, 46(5):1386–1396, May 1998.
- [21] Y. Bar-Shalom, P.K. Willett, and X. Tian. *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, Storrs, CT, 2011.
- [22] L. Zuo, R. Niu, and P. K. Varshney. A sensor selection approach for target tracking in sensor networks with quantized measurements. In *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2008)*, Las Vegas, NV, March 2008.
- [23] O. Ozdemir, R. Niu, and P. K. Varshney. Tracking in Wireless Sensor Networks Using Particle Filtering: Physical Layer Considerations. *IEEE Trans. Signal Processing*, 57(5):1987–1999, May 2009.
- [24] S. Grime and H. Durrant-Whyte. Data fusion in decentralized sensor networks. *IFAC Control Engineering Practice*, 2(5):849–863, 1994.

- [25] S.J. Julier and J.K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *Proceedings of the American Control Conference*, pages 2369–2373, 1997.
- [26] O. E. Drummond. On track and tracklet fusion filtering. In *SPIE Conf. on Signal and Data Processing of Small Targets, vol. 4728*, pages 176–195, August 2002.
- [27] H. Chen, T. Kirubarajan, and Y. Bar-Shalom. Performance limits of track-to-track fusion versus centralized estimation: theory and application. *IEEE Transactions on Aerospace and Electronic Systems*, 39(2):386–400, April 2003.
- [28] Y. M. Zhu, Z. S. You, J. Zhou, K.S. Zhang, and X. R. Li. The optimality for the distributed Kalman filtering fusion with feedback. *Automatica*, 37:1489–1493, September 2001.
- [29] F. Govaers and W. Koch. An exact solution to track-to-track-fusion at arbitrary communication rates. *IEEE Transactions on Aerospace and Electronic Systems*, 48(3):2718–2729, July 2012.
- [30] D. Castanon and D. Teneketzis. Distributed estimation algorithms for nonlinear systems. *IEEE Transactions on Automatic Control*, 30:418–425, May 1985.
- [31] C. Y. Chong, S. Mori, and K. C. Chang. Distributed multitarget multisensor tracking. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*. Artech House, 1990.
- [32] F. Bourgault and H.F. Durrant-Whyte. Communication in general decentralized filters and the coordinated search strategy. In *Proc. 2004 7th International Conference on Information Fusion*, pages 723–770, 2004.

- [33] L.L. Ong, T. Bailey, H. Durrant-Whyte, and B. Upcroft. Decentralised particle filtering for multiple target tracking in wireless sensor networks. In *Proc. 2008 11th International Conference on Information Fusion*, June 2008.
- [34] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proc. IEEE*, 98(11):1847-1864, Nov. 2010.
- [35] S. Kar and J.M.F. Moura. Gossip and distributed Kalman filtering: Weak consensus under weak detectability. *IEEE Trans. Signal Process.*, 59(4):1766–1784, April 2011.
- [36] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE*, 95(1):215–233, Jan. 2007.
- [37] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In *Proc. 46th IEEE Conference on Decision and Control*, December 2007.
- [38] R. Olfati-Saber. Kalman-consensus filter: Optimality, stability, and performance. In *Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, pages 7036–7042, Shanghai, China, December 2009.
- [39] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26:622–633, April 2008.
- [40] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *IEEE Trans. Signal Process.*, 57(7):2748–2761, July 2009.
- [41] H. Liu, H. So, F. Chan, and K. Lui. Distributed particle filtering for target tracking in sensor networks. *Progress in Electromagnetics Research C*, 11:171–182, 2009.

- [42] S. Farahmand, S. I. Roumeliotis, and G. B. Giannakis. Set-membership constrained particle filter: Distributed adaptation for sensor networks. *IEEE Trans. Signal Processing*, 59(9):4122–4138, September 2011.
- [43] O. Hlinka, O. Sluciak, F. Hlawatsch, P. M. Djuric, and M. Rupp. Likelihood consensus and its application to distributed particle filtering. *IEEE Trans. Signal Process.*, 60(8):4334–4349, August 2012.
- [44] A. Mohammadi and A. Asif. Distributed particle filter implementation with intermittent/irregular consensus convergence. *IEEE Transactions on Signal Processing*, 61(10):2572–2587, May 2013.
- [45] V. Savic, H. Wymeersch, and S. Zazo. Belief consensus algorithms for fast distributed target tracking in wireless sensor networks. ArXiv e-prints:1202.5261, Feb, 2012.
- [46] J. Williams, J. Fisher, and A. Willsky. Approximate dynamic programming for communication-constrained sensor network management. *IEEE Trans. Signal Process.*, 55:4300–4311, Aug. 2007.
- [47] H. Zhang, J. Moura, and B. Krogh. Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost. *IEEE Trans. Signal Process.*, 57:2383–2395, June 2011.
- [48] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin. On efficient sensor scheduling for linear dynamical systems. *Automatica*, 48:2482–2493, October 2012.
- [49] F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration. *IEEE Trans. Signal Process.*, 19(1):61–72, May 2002.

- [50] L. Zuo, R. Niu, and P. K. Varshney. Posterior CRLB based sensor selection for target tracking in sensor networks. In *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2007)*, Honolulu, HI, April 2007.
- [51] E. Masazade, M. Fardad, and P.K. Varshney. Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks. *IEEE Signal Processing Letters*, 19:845–848, June 2012.
- [52] C. Yu and P. K. Varshney. Bit allocation for discrete signal detection. *IEEE Trans. Communications*, 46:173–175, Feb. 1998.
- [53] C. Rago, P. Willett, and Y. Bar-Shalom. Censoring sensors: A low-communication-rate scheme for distributed detection. *IEEE Trans. Aerospace and Electronic Systems*, 32(1):554–568, 1996.
- [54] R. Jiang and B. Chen. Fusion of censored decisions in wireless sensor networks. *IEEE Trans. Wireless Comm.*, 4:2668–2673, November 2005.
- [55] S. Appadwedula, V. V. Veeravalli, and D. L. Jones. Decentralized detection with censoring sensors. *IEEE Trans. Signal Process.*, 56:1362–1373, April 2008.
- [56] R. S. Blum and B. M. Sadler. Energy efficient signal detection in sensor networks using ordered transmission. *IEEE Trans. Signal Process.*, 56:3229–3235, Jul. 2008.
- [57] X. Chen, R. Blum, and B. M. Sadler. A new scheme for energy-efficient estimation in a sensor network. In *43rd Annual Conference on Information Science and Systems*, Baltimore, MD, March 2009.
- [58] R. J. Little and D. B. Rubin. *Statistical analysis with missing data*. Wiley, New York, 2002.

- [59] E. J. Msechu and G. B. Giannakis. Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *IEEE Trans. Signal Process.*, 60:400–414, January 2012.
- [60] W. Koch. On exploiting ‘negative’ sensor evidence for target tracking and sensor data fusion. *Information Fusion, Special Issue on the Seventh International Conference on Information Fusion-Part II*, 8(10):28–39, January 2007.
- [61] Y. Ruan, P. Willett, A. Marrs, F. Palmieri, and S. Marano. Practical fusion of quantized measurements via particle filtering. *IEEE Trans. Aerospace and Electronic Systems*, 44(1):15–29, January 2008.
- [62] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, February 2002.
- [63] S. J. Roberts and W. D. Penny. Variational Bayes for generalized autoregressive models. *IEEE Trans. Signal Process.*, 50:2245–2257, September 2002.
- [64] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, and M. Metcalf. *Numerical Recipes in C: The Art of Scientific Computing*. Second Edition, Cambridge University Press, 1992.
- [65] U. A. Khan and J. M. F. Moura. Distributing the kalman filter for large-scale systems. *IEEE Trans. Signal Process.*, 56:4919–4935, October 2008.
- [66] A. Mohammadi and A. Asif. Distributed particle filtering for large scale dynamical systems. In *IEEE 13th International Multitopic Conference (INMIC), 2009*, December 2009.

- [67] A. Carmi and P. Gurfil. Sensor selection via compressed sensing. *Automatica*, pages 3304–3314, Nov 2013.
- [68] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, Feb 2006.
- [69] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Info. Theory*, 52(12):5406 – 5425, Dec. 2006.
- [70] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, Dec. 2007.
- [71] G. Yang, V.Y.F. Tan, C.K. Ho, S.H. Ting, and Y.L. Guan. Wireless compressive sensing for energy harvesting sensor nodes. *IEEE Trans. Signal Process.*, 61:4491 – 4505, Sep. 2013.
- [72] R. Niu and P. K. Varshney. Target location estimation in sensor networks with quantized data. *IEEE Trans. on Signal Process.*, 54(12):4519–4528, December 2006.
- [73] Y. Zheng, R. Niu, and P. K. Varshney. Closed-form performance for location estimation based on quantized data in sensor networks. In *Proc. 13th International Conference on Information Fusion*, pages 1–7, Edinburgh, Scotland, Jul. 2010.
- [74] O. Ozdemir, R. Niu, and P.K. Varshney. Channel aware target localization with quantized data in wireless sensor networks. *IEEE Trans. Signal Processing*, 57(3):1190–1202, March 2009.
- [75] S. Iyengar, R. Niu, and P. K. Varshney. Fusing dependent decisions for hypothesis testing with heterogeneous sensors. *IEEE Trans. Signal Processing*, 60(9):4888–4897, September 2012.

- [76] F. Gustafsson and R. Karlsson. Statistical results for system identification based on quantized observations. *Automatica*, 45:2794–2801, 2009.
- [77] Y. Kim and A. Ortega. Quantizer design for source localization in sensor networks. In *Proc. ICASSP'05 Conf.*, Philadelphia, Pennsylvania, USA, March 2005.
- [78] J. Chen and W. Su. Optimal Bit Allocation for Distributed Estimation in Wireless Sensor Networks. In *IEEE International Conference on Control and Automation*, Christchurch, New Zealand, Dec. 2009.
- [79] K. M. Steven. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [80] E. A. Riskin. Optimal bit allocation via the generalized BFOS algorithm. *IEEE Trans. Inf. Theory*, 37:400–402, Mar 1991.
- [81] E. Masazade, R. Niu, and P. K. Varshney. Dynamic bit allocation for object tracking in wireless sensor networks. *IEEE Trans. Signal Processing*, 60(10):5048–5063, October 2012.
- [82] S. Kumar, F. Zhao, and D. Shepherd Eds. Special issue on collaborative signal and information processing in microsensor networks. *IEEE Signal Processing Magazine*, 19(2), Mar. 2002.
- [83] H. Gharavi and S. Kumar Eds. Special issue on sensor networks and applications. *Proceedings of the IEEE*, 91(8), August 2003.
- [84] A. Sayeed, D. Estrin, G. Pottie, and K. Ramchandran Eds. Special Issue on Self-Organizing Distributed Collaborative Sensor Networks. *IEEE Journal on Selected Areas in Communications*, 23(4), April 2005.

- [85] A. L. Swindlehurst and T. Kailath. Passive direction-of-arrival and range estimation for near-field sources. In *Spectrum Estimation and Modeling, Fourth Annual ASSP Workshop on*, pages 123–128, Minneapolis, MN, USA, Aug 1988.
- [86] Y. Jiang and M. R. Azimi-Sadjadi. A Robust Source Localization Algorithm Applied to Acoustic Sensor Network. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007.
- [87] X. Sheng and Y. H. Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Trans. Signal Process.*, 53(1):44–53, Jan. 2005.
- [88] R. Niu and P. K. Varshney. Source Localization in Sensor Networks with Rayleigh Faded Signals. In *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume III, pages 1229–1232, Honolulu, Hawaii, April 2007.
- [89] E. Masazade, R. Niu, P. K. Varshney, and M. Keskinöz. Energy Aware Iterative Source Localization for Wireless Sensor Networks. *IEEE Trans. on Signal Process.*, 58:4824–4835, Sep 2010.
- [90] H. Wang, K. Yao, G. Pottie, and D. Estrin. Entropy-based sensor selection heuristic for target localization. In *3rd Int.Symp.Information Processing in Sensor Networks*, Berkeley, CA, April 2004.
- [91] C. M. Kreucher, K. D. Kastella, and A. O. Hero. Sensor management using an active sensing approach. *IEEE Trans. on Signal Process.*, 85:607–624, Mar 2005.

- [92] M. L. Hernandez, T. Kirubarajan, and Y. Bar-Shalom. Multisensor Resource Deployment Using Posterior Cramér-Rao Bounds. *IEEE Trans. Aerosp. Electron. Syst.*, 40(2):399–416, Apr 2004.
- [93] L. Zuo. *Conditional Posterior Cramer-Rao Lower Bound and Distributed Target Tracking in Sensor Networks*. PhD Thesis, 2010.
- [94] E. Masazade. *Resource Aware Distributed Detection and Estimation of Random Events in Wireless Sensor Networks*. PhD Thesis, 2010.
- [95] R. Niu and P. K. Varshney. Closed-Form Performance For Location Estimation Based on Fused Data in a Sensor Network. In *Proc. the 12th International Conference on Information Fusion*, Seattle, Washington, July 2009.
- [96] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, NY, 1984.
- [97] M. Šimandl, J. Královec, and P. Tichavský. Filtering, predictive, and smoothing cramer-rao bounds for discrete-time nonlinear dynamic systems. *Automatica.*, 11:1703–1716, April 2001.
- [98] M. L. Hernandez, T. Kirubarajan, and Y. Bar-Shalom. Multisensor resource deployment using posterior Cramér-Rao bounds. *IEEE Trans. Aerospace and Electronic Systems*, 40(2):399–416, April 2004.
- [99] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
- [100] R.A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 1985.

- [101] A. S. Householder. *The Theory of Matrices in Numerical Analysis*. Blaisdell Publishing Co., New York, 1964.
- [102] S. C. Chapra and R. P. Canale. *Numerical Methods for Engineers*. McGraw-Hill, 2002.
- [103] O. Ozdemir, R. Niu, and P. K. Varshney. Modified Bayesian Cramér-Rao lower bound for nonlinear tracking. In *Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2011)*, Prague, Czech Republic, May 2011.
- [104] J. H. Kotecha and P. M. Djurić. Gaussian particle filtering. *IEEE Trans. Signal Process.*, 51(10):2592–2601, October 2003.
- [105] S. Marano, V. Matta, and L. Tong. Distributed detection in the presence of byzantine attacks. *IEEE Trans. Signal Process.*, 57:16–29, January 2009.
- [106] A. Vempaty, L. Tong, and P. K. Varshney. Distributed inference with byzantine data: State-of-the-art review on data falsification attacks. *IEEE Signal Process. Mag.*, 30:65–75, September 2013.
- [107] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. K. Varshney. Localization in wireless sensor networks: Byzantines and mitigation techniques. *IEEE Trans. on Signal Process.*, 61:1495–1508, March 2013.

VITA

NAME OF AUTHOR: Yujiao Zheng

MAJOR: Electrical and Computer Engineering

EDUCATION:

M.S. Feb. 2011 Syracuse University, NY, USA

B.S. Jul. 2008 University of Science and Technology of China, Hefei, China

PUBLICATIONS:

1. Y. Zheng, R. Niu and P. Varshney, "Sequential Bayesian Estimation with Censored Data for Multi-Sensor Systems," submitted to *IEEE Trans. on Signal Process.* for the third round review, (AQ, Nov. 28, 2013).
2. Y. Zheng, O. Ozdemir, R. Niu, and P. K. Varshney, "New Conditional Posterior Cramér-Rao Lower Bounds for Nonlinear Sequential Bayesian Estimation," *IEEE Trans. on Signal Process.*, Oct. 2012.
3. Y. Zheng, R. Niu and P. Varshney, "Closed-Form Performance for Location Estimation Based on Quantized Data in Sensor Networks for General Decay Model," in preparation and to be submitted to *IEEE T-SP*.
4. Yujiao Zheng, Thakshila Wimalajeewa and Pramod K. Varshney, "Probabilistic sensor management for target tracking via compressive sensing," submitted to *the 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Nov. 2013.
5. Y. Zheng, R. Niu and P. Varshney, "Fusion of Quantized Data for Bayesian Estimation

Aided by controlled noise,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013

6. Y. Zheng, R. Niu and P. Varshney, “Sequential Bayesian Estimation with Censored Data,” *IEEE Statistical Signal Processing workshop*, Aug. 2012
7. Y. Zheng, R. Niu and P. Varshney, “Closed-Form Performance for Location Estimation Based on Quantized Data in Sensor Networks,” *13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.

AWARDS AND HONORS

- Master of Science Prize, Syracuse University, 2011
- Best Student Paper Award in the 13th International Conference on Information Fusion, Edinburgh, UK, July 2010
- Graduate Research Assistantship, Syracuse University, since Fall 2009
- Graduate Teaching Assistantship, Syracuse University, 2008-2009
- “Guanghua” Student Scholarship, University of Sci. & Tech. of China, 2007
- Outstanding Student Scholarship, University of Sci. & Tech. of China, 2005, 2006