

5-14-2010

State Estimation with Unconventional and Networked Measurements

Zhansheng Duan
University of New Orleans

Follow this and additional works at: <https://scholarworks.uno.edu/td>

Recommended Citation

Duan, Zhansheng, "State Estimation with Unconventional and Networked Measurements" (2010).
University of New Orleans Theses and Dissertations. 1133.
<https://scholarworks.uno.edu/td/1133>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

State Estimation with Unconventional and Networked Measurements

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Engineering and Applied Science

by

Zhansheng Duan

B.S. Xi'an Jiaotong University, 1999
Ph.D. Xi'an Jiaotong University, 2005

May 2010

Copyright 2010, Zhansheng Duan

Acknowledgments

This research was supported in part by NSFC grant 60602026, Project 863 through grant 2006AA01Z126, ARO through grant W911NF-04-1-0274, NASA/LEQSF through grant (2001-4)-01, ARO through grant W911NF-08-1-0409, ONR-DEPSCoR through grant N00014-09-1-1169, and NAVO through Contract # N62306-09-P-3S01.

I would like to express my sincerest gratitude to my major advisor, Dr. X. Rong Li, for his insightful and knowledgeable direction, continuous help and support, invaluable encouragement and discussion, the freedom allowed at doing research during the past four and a half years and even before I came to the United States, which made the completion of my current Ph.D. research work possible and will definitely benefit my career greatly.

I would also thank Dr. Tumulesh K. S. Solanky, Dr. Vesselin P. Jilkov, Dr. Huimin Chen, and Dr. Dongxiao Zhu for serving on my thesis committee, and for their constructive and valuable comments on the dissertation. Special thanks go to Dr. Li, Dr. Jilkov, Dr. Chen, Dr. Zhihai He from University of Missouri, Columbia, and their families, Dr. Ning Li, all international students stayed in section 113 of Louisiana Superdome during Hurricane Katrina, Fei Teng, Cao Liang, Zeyu Wu and Zhiquan He. Without them, I could never live through the hardship during and after Hurricane Katrina and continue my study in the United States up to now. Special thanks also go to Dr. Mahendra Mallick from Georgia Tech Research Institute (GTRI) for his continuous help.

In addition, I would like to acknowledge all my friends and members in the Information and Systems Laboratory (ISL), especially Dr. Jie Zhou, Dr. Meiqin Liu, Dr. Zhanlue Zhao,

Dr. Jifeng Ru, Dr. Ming Yang, Dr. Morten Stakkeland, Dr. Jian Lan, Dr. Xianghui Yuan, Anwer Bashi, Ryan Pitre, Trang Nguyen, Xiaomeng Bian, Yangsheng Chen, Lie Xiong, Yongqi Liang, Yu Liu, Jiande Wu, Gang Liu, Sowmya Bandarupalli, Ashwini Amara, Harika Rao Vamulapalli, Jaipal Reddy Katkuri, and Rastin Rastgoufard. With their collaboration and inspiring discussion, I enjoyed my research work in this pleasant environment. Furthermore, I want to thank all the members and staff of the Department of Electrical Engineering at the University of New Orleans for their service and support.

Last but not least at all, I want to thank my parents, my wife, Ping Yang, my brother and sister, for their unselfish, endless support, encouragement and love, and my daughter, Xiaowen Duan, for all the happiness she brought to me. Without them, I could not imagine that I can make this long journey. I dedicate this thesis to them.

Contents

Abstract	x
1 Introduction	1
1.1 Background	1
1.2 Research Objectives	7
1.3 Thesis Outline	8
2 Optimal Linear State Estimation with Noisy and Noise-free Measurements	10
2.1 Introduction and related research	10
2.2 Problem formulation	13
2.3 Noise-free measurements	14
2.3.1 Linear equality constraints	14
2.3.2 Nonlinear equality constraints	16
2.3.3 Autocorrelated measurement noise	16
2.3.4 Singular measurement noise	18
2.4 Batch LMMSE estimation	19
2.5 Sequential LMMSE estimation	21
2.5.1 Form 1	21
2.5.2 Form 2	23
2.6 Extension to nonlinear measurement	27
2.6.1 Nonlinear noisy measurement	28

2.6.2	Nonlinear noise-free measurement	29
2.6.3	Nonlinear noisy and noise-free measurements	29
2.7	Illustrative examples	31
2.7.1	Example 1	33
2.7.2	Example 2	35
2.7.3	Example 3	38
2.7.4	Example 4	41
2.8	Summary	43
3	State Estimation with Point and Set Measurements	46
3.1	Introduction and related research	46
3.2	Problem formulation	47
3.3	Set measurements	49
3.3.1	Linear inequality constraint	49
3.3.2	Nonlinear inequality constraint	50
3.3.3	Quantized measurement	51
3.4	Particle filter	52
3.5	MMSE filtering with point and set measurements	53
3.5.1	General form of MMSE filter	54
3.5.2	Some discussions on set constrained estimation	63
3.5.3	Relaxation of Gaussian assumption	68
3.6	Discretization of Gaussian-related distribution	70
3.7	Illustrative examples	72
3.7.1	Example 1—Mean and covariance of truncated Gaussian distribution	72
3.7.2	Example 2—Quantized estimation	76
3.7.3	Example 3—Inequality constrained estimation	79
3.7.4	Example 4—Noise-free measurements	83

3.8	Summary	86
4	Lossless Linear Transformation of Sensor Data for Distributed Estimation	
	Fusion	88
4.1	Introduction and related research	88
4.2	Problem formulation	92
4.3	Review of existing distributed fusion algorithms	93
4.3.1	Information matrix fusion	93
4.3.2	Simple convex combination	94
4.3.3	WLS fusion	95
4.4	Sensor measurement transformation	96
4.4.1	Transformation I	96
4.4.2	Transformation II	98
4.5	Distributed fusion with transformed data	101
4.6	Optimality of proposed algorithms	103
4.7	Reduction of computational complexity	106
4.8	Extension to reduce-rate communication	108
4.8.1	With invertible STM and no process noise	108
4.8.2	With not necessarily invertible STM and no process noise	110
4.9	Extension to singular measurement noise case	112
4.10	Summary	113
4.11	Appendix	114
4.11.1	Proof of Theorem 1	114
4.11.2	Proof of Theorem 2	116
4.11.3	Proof of Theorem 3	119
4.11.4	Proof of Theorem 4	120
4.11.5	Proof of Theorem 5	122

5	Optimal State Estimation in the Presence of Multiple Packet Dropouts	123
5.1	Introduction and related research	123
5.2	Problem formulation	125
5.3	Summary of two existing forms of LMMSE estimation [113, 114]	126
5.4	An alternative form of LMMSE estimation	130
5.5	MMSE estimation	135
5.6	Illustrative examples	139
5.6.1	Example 1— $\langle \gamma_k \rangle$ is a Bernoulli distributed white sequence	139
5.6.2	Example 2— $\langle \gamma_k \rangle$ is a Markov chain	140
5.7	Summary	144
6	Conclusions and Future Work	145
	Bibliography	164
	Vita	165

Abstract

This dissertation consists of two main parts. One is about state estimation with two types of unconventional measurements and the other is about two types of network-induced state estimation problems.

The two types of unconventional measurements considered are noise-free measurements and set measurements. State estimation with them has numerous real supports. For state estimation with noisy and noise-free measurements, two sequential forms of the batch linear minimum mean-squared error (LMMSE) estimator are obtained to reduce the computational complexity. Inspired by the estimation with quantized measurements developed by Curry [28], under a Gaussian assumption, the minimum mean-squared error (MMSE) state estimator with point measurements and set measurements of any shape is proposed by discretizing continuous set measurements. State estimation under constraints, which are special cases of the more general framework, has some interesting properties. It is found that under certain conditions, although constraints are indispensable in the evolution of the state, update by treating them as measurements is redundant in filtering.

The two types of network-induced estimation problems considered are optimal state estimation in the presence of multiple packet dropouts and optimal distributed estimation fusion with transformed data. An alternative form of LMMSE estimation in the presence of multiple packet dropouts, which can overcome the shortcomings of two existing ones, is proposed first. Then under a Gaussian assumption, the MMSE estimation is also obtained based on a hard decision by comparing the measurements at two consecutive time instants.

It is pointed out that if this comparison is legitimate, our simple MMSE solution largely nullifies existing work on this problem. By taking linear transformation of the raw measurements received by each sensor, two optimal distributed fusion algorithms are proposed. In terms of optimality, communication and computational requirements, three nice properties make them attractive.

Keywords: State Estimation, Noise-free Measurement, Set Measurement, Quantized Measurement, Constrained Estimation, Nonlinear Filtering, Distributed Fusion, Packet Dropout.

List of Figures

2.1	Vehicle trajectory in one run	32
2.2	Vehicle velocity in one run	32
2.3	RMS position error comparison for example 1. Note that \overline{POEC} overlaps with \overline{POEC} and $PO\overline{EC}$ overlaps with $POEC$	34
2.4	RMS velocity error comparison for example 1. Note that \overline{POEC} overlaps with \overline{POEC} and $PO\overline{EC}$ overlaps with $POEC$	35
2.5	RMS position error comparison for example 2. Note that NM-0 diverges and its error curve is far above the rest.	36
2.6	RMS position error comparison for example 2 without NM-0. Note that KF, NM-EC-Q and EC-NM-Q essentially overlap with each other.	37
2.7	RMS velocity error comparison for example 2. Note that KF, NM-EC-Q and EC-NM-Q essentially overlap with each other.	38
2.8	RMS position error comparison for example 3 with correct Q and $\hat{x}_{0 0} = \bar{x}_0$. Note that BNM, BNM-EC and EC-BNM essentially overlap with each other, and ENM, ENM-EC and EC-ENM essentially overlap with each other.	39
2.9	RMS velocity error comparison for example 3 with correct Q and $\hat{x}_{0 0} = \bar{x}_0$. Note that BNM, BNM-EC and EC-BNM essentially overlap with each other, and ENM, ENM-EC and EC-ENM essentially overlap with each other.	40

2.10	RMS position error comparison for example 3 with Q mismatch. Note that BNM-EC essentially overlaps with EC-BNM, and ENM-EC essentially overlaps with EC-ENM.	41
2.11	RMS velocity error comparison for example 3 with Q mismatch. Note that BNM-EC essentially overlaps with EC-BNM, and except for the small deviation at transient, ENM-EC essentially overlaps with EC-ENM.	42
2.12	RMS position error comparison for example 3 with mis-specified $\hat{x}_{0 0}$. Note that BNM-EC essentially overlaps with EC-BNM, and ENM-EC essentially overlaps with EC-ENM.	43
2.13	RMS velocity error comparison for example 3 with mis-specified $\hat{x}_{0 0}$. Note that BNM-EC essentially overlaps with EC-BNM, and except very little deviation at transient, ENM-EC essentially overlaps with EC-ENM.	44
2.14	RMS position error comparison for example 4	45
2.15	RMS velocity error comparison for example 4	45
3.1	Comparison of estimate to the mean of z in one run	74
3.2	Comparison of estimate to the covariance of z in one run	74
3.3	Comparison of estimate to the mean of z in another run	75
3.4	Comparison of estimate to the covariance of z in another run	75
3.5	RMS error comparison for quantized estimation	77
3.6	RMS position error comparison with correct Q and $\hat{x}_{0 0} = \bar{x}_0$. Note that KF overlaps with IEC.	82
3.7	RMS velocity error comparison with correct Q and $\hat{x}_{0 0} = \bar{x}_0$. Note that KF overlaps with IEC.	83
3.8	RMS position error comparison with Q mismatch. Note that KF and IEC-Q have no noticeable difference.	84

3.9	RMS velocity error comparison with Q mismatch. Note that KF and IEC-Q have no noticeable difference.	85
3.10	RMS position error comparison with mis-specified $\hat{x}_{0 0}$	85
3.11	RMS velocity error comparison with mis-specified $\hat{x}_{0 0}$	86
3.12	RMS position error comparison for example 4	86
3.13	RMS velocity error comparison for example 4	87
5.1	RMS error comparison with $p_k = 0.8$. Note that MMSE-a and MMSE-b overlap with each other.	140
5.2	RMS error comparison with $p_k = 0.5$. Note that MMSE-a and MMSE-b overlap with each other.	141
5.3	RMS error comparison with $p_k = 0.2$. Note that MMSE-a and MMSE-b overlap with each other.	142
5.4	RMS error comparison with $p_{11} = 1, p_{00} = 1$	143
5.5	RMS error comparison with $p_{11} = 0.5, p_{00} = 0.5$	143

List of Tables

2.1	LMMSE estimators used in example 1	34
2.2	LMMSE estimators used in example 2	36
2.3	State estimators used in example 3	39
3.1	Comparison of computational burden	77
3.2	Estimators used in inequality constrained estimation example	82
3.3	Estimators used in noise-free measurement example	84
5.1	Estimators used in Fig. 5.1 to Fig. 5.3	139
5.2	Estimators used in Fig. 5.4 and Fig. 5.5	142

Chapter 1

Introduction

1.1 Background

The process of inferring the value of a quantity of interest from indirect, inaccurate and uncertain observations is known as estimation [10]. More rigorously, estimation can be viewed as the process of selection of a point from a continuous space—the “best estimate.” The quantity of interest in an estimation problem can be a parameter—a time-invariant quantity (parameter estimation), or the state of a dynamic system (state estimation).

Dynamic systems are encountered almost everywhere in reality, e.g., in target tracking, chemical reaction, satellite guidance and navigation, power transmission and distribution, orbit determination, weather and financial forecasting, optimal control, fault diagnosis. We are interested to know the state of a dynamic system, but only indirect, inaccurate and uncertain observations are usually available to us. So state estimation is extremely important in practice.

In a typical state estimation problem, the estimator uses the knowledge about the evolution of the state (the system dynamics), the sensor (measurement system) and the probabilistic characterization of the various random factors (uncertainties) and the prior information. A celebrated solution to state estimation is the Kalman filter [62], which is optimal in the MMSE sense if the system satisfies a linear Gaussian assumption. In reality, the assumption

is usually not valid. Next, depending on the applications, some of the extensions to the classical state estimation will be briefly summarized. We describe these techniques following largely [76].

In practice, either the system dynamics or measurement system can be nonlinear and the driving noise may also be non-Gaussian. This leads to nonlinear filtering, which consists of point estimation and density estimation [76]. The goal of point estimation is to get some descriptive statistics of the random state, instead of its probability distribution. For example, most point estimation approaches focus on getting the first two moments of the state. Density estimation aims at estimating the probability distribution of the state.

To deal with the nonlinearity from system dynamics and/or measurement equations, in nonlinear point estimation the simplest way is approximation by linearization. For example, the most widely used Extended Kalman filter (EKF) successively linearizes around the latest estimate through a first-order Taylor series expansion (TSE). The EKF is popular mainly because of its simplicity, but its performance is not necessarily good. It is known that the performance of EKF relies heavily on the degree of nonlinearity of the system dynamics and/or measurement equations and the accuracy of the initial estimate [76]. From TSE, we know that the linearization is an adequate approximation of its nonlinear counterpart only when the latest estimate is sufficiently close to the quantity to be estimated, which can rarely be guaranteed in practice. The error may build up over time and result in filtering divergence. One compensation is to use iterated EKF (IEKF), in which the measurement equation is repeatedly linearized around the most recent updated state estimate several times. The IEKF usually outperforms the EKF somewhat, especially if the state update improves state prediction greatly, but an improvement is not guaranteed in general [76]. Experience indicates that after a couple of iterations the performance improvement diminishes. Another compensation is to replace the first-order TSE by some higher-order TSEs. A second-order EKF replaces the nonlinear functions involved in state and measurement predictions by a

second-order TSE and then applies the Kalman filter formulas for the gain calculation and update. Its implementation is not as simple as the EKF. TSEs with orders higher than two are rarely used in practice. Note that the EKF is derivative based, e.g., we need to calculate Jacobian and even Hessian matrices, which requires differentiability of the nonlinear function. In contrast, other nonlinear point estimation approaches introduced below are derivative free.

The idea of the EKF is to approximate a nonlinear function by its TSE approximation. Another popular idea proposed in recent years is to use interpolation. Unlike the TSE approximation which relies on the function value at only a single point, interpolation relies on the function values at multiple points and it is possible to be derivative free [76]. There are numerous interpolation formulas. Stirling's is among the best. Under the Gaussian assumption, the second-order Stirling interpolation based filter (DD2) [89] has fourth-order error, compared with the second-order error of the EKF and third-order error of the second-order EKF.

In contrast to the function approximation techniques [76], e.g., EKF, second-order EKF, first-order Stirling interpolation based filter (DD1) [89] and DD2, that approximate the involved nonlinear functions, moment matching techniques for nonlinear filtering approximate the first two moments involved in the LMMSE filter directly [76]. Different moment matching techniques differ from each other on the ways the first two moments involved in the LMMSE filter are approximated. For example, the unscented transformation (UT), which is the key to unscented filtering (UF) [60], approximates mean and covariance of a nonlinearly transformed random vector by deterministically sampled sigma points and their corresponding weights. The Gaussian filter [57] successively approximates each pdf by a moment-matching Gaussian distribution and then uses Gauss-Hermite quadratures to compute the involved first two moments.

Both TSE and Stirling's interpolation approximate the involved nonlinear functions in a deterministic way. They are simple but there is no optimality in general. Instead, stochastic

linearization [76] approximates the original nonlinear stochastic system locally by a stochastic model that is simple and linear in an optimal fashion so that linear filtering results can be applied. While the EKF is not accurate if the estimation error is not small, the filter based on stochastic linearization accounts for large errors within the expectations probabilistically and thus tends to have a more conservative gain and better performance for the case involving large estimation error. The main difficulty associated with stochastic linearization lies in the (analytical or numerical) evaluation of the expectations. Statistical linearization [69] approximates them by sample averages.

Density estimation is much harder than point estimation because the first two moments are only two descriptive statistics of a distribution. In theory, density estimation can be solved by Bayesian recursive filtering (BRF). The main difficulty for BRF comes from the evaluation of the expectations of the transition density and likelihood function. Particle filters [2] are sequential Monte-Carlo (SMC) simulation-based implementations of the BRF. It approximates the posterior distribution by a random probability mass function (pmf). The key to particle filtering is the choice of proposal distribution and resampling. As a breakthrough in the SMC method, resampling was introduced to counteract degeneracy, otherwise after several recursions weights of the random pmf will be concentrated at one particle leaving negligible weights to the others [76]. Although particle filtering is a close approximation to BRF, it achieves the significantly improved estimation performance at the cost of a heavily increased computational burden.

State estimation plays a key role in target tracking. In contrast to other state estimation, the challenges in target tracking are mainly from the measurement-origin uncertainty and target motion uncertainty.

By measurement-origin uncertainty in target tracking, it is meant that the origin of a measurement can not be determined without uncertainty. For example, in single-target tracking in clutter, the received measurements may be from the true target, but they may

also be from background clutter or countermeasures. In multiple-target tracking in clutter, a measurement may or may not be from a specific target of interest. It may also be from other targets or from background clutter or countermeasures. Most filters dealing with measurement-origin uncertainty in target tracking follow the traditional two-stage strategy: data association first and then estimate. That is, the origin of the measurements are determined first and then measurement-origin known filtering techniques can be applied. For instance, the nearest neighbor filter (NNF) [102,108] chooses the validated measurement that is closest to the gate center and ignores the rest, then it treats the selected measurement as if it were surely the true one and uses it to update the track. The NNF can be improved significantly [76] by accounting for, in the estimation step, the probability that the NN measurement is not really the true one if the validated measurement set is not empty, leading to the probabilistic nearest neighbor filter. The NNF only chooses the closest validated measurement to the gate center and discards the rest. This is hard decision. A major drawback is that the decision error (i.e., the closest validated measurement is not from the target) is not taken into account in estimation. The probabilistic data association filter (PDAF) [12] calculates the probability of each validated measurement to be from the true target and uses them as weights to sum up all validated measurements (soft or no decision) as a whole to update the track. The joint probabilistic data association filter (JPDAF) [48] is an extension of the PDAF to track maintenance of multiple targets by following the same soft decision idea. Three fundamental assumptions used by JPDAF [76] are that the established targets/tracks are known, a validated measurement can have only one source, and at most one measurement can originate from a target. Similar to the PDAF, the JPDAF calculates the probability of each measurement to be from the established targets. Another popular method for multiple-target tracking is multiple hypothesis tracking (MHT) [93], which differs from the JPDAF in two fundamental aspects [76]: the associations are measurement-oriented, rather than target-oriented in the JPDAF, and the hypotheses of data association are for the history,

not just for the current time as in the JPDAF.

Another major difficulty in maneuvering target tracking is due to the target motion uncertainty. In essence, maneuvering target tracking is a hybrid estimation problem in which we need to estimate both the continuous target state and discrete target motion mode [76]. Soft decision based multiple model (MM) approach, e.g., interacting multiple model (IMM) [16], is becoming the mainstream in maneuvering target tracking [76], as opposed to traditional hard decision based algorithms, e.g., variable dimension filter [5], and input estimation algorithm [20]. The conventional hard decision based approaches have two stages. They decide on a model first and then run a filter based on it as if it were the true one. There is only one model chosen at each time and only the filter based on it is run at any single time. The major drawback of the hard decision based approach is that the decision errors with respect to the model are not accounted for in estimation [76]. The MM approach assumes a set of models for the true mode, each having the possibility to be true at the time. A bank of elemental filters, each based on a unique model in the set, is run. The overall estimate is a weighted average of the results of elemental filters. In contrast to the hard decision based approaches, all possible decisions and their error probabilities are accounted for in the MM approach.

With the emerging of sensor networks, traditional state estimation problems are facing new challenges. For example, the measurements are available from multiple sensors. How to best utilize these multiple measurements is the key of estimation fusion [81]. There are two basic fusion architectures. One is centralized fusion and the other is distributed fusion, depending on whether the raw measurements are sent to the fusion center or not. In centralized fusion, all raw measurements are sent to the fusion center, while in distributed fusion, each sensor only sends in processed data. They have pros and cons in terms of performance, channel requirements, reliability, survivability, information sharing, etc. Although many practical issues do exist, theoretically, centralized fusion is nothing but an estimation

problem with distributed data. Distributed fusion is more challenging and has been a focal point of most fusion research.

Other challenges faced by network-based state estimation include: due to unreliable communication between local sensors and the processing center, packet transmission delay [4, 86, 112, 115, 129] and multiple packet dropouts [98–100, 113, 114] are usually inevitable; also, the constraints on communication bandwidth, power consumption [46, 70] and computational capability should be considered, which make the distributed estimation [81] and estimation problems with compressed [35, 128] or quantized data [33, 58] necessary; the sensors often work asynchronously [1, 55, 56, 74, 82, 88, 122] instead of synchronously, which is widely assumed in the literature; the sensors may also have different sampling rates and communicate rates.

In reality, the evolution of the state may be subject to constraints. For example, in ground target tracking [65], the road networks can be described by equality or inequality constraints. In the quaternion-based attitude estimation problem [15], the attitude vector must have a unit norm. In a compartmental model with zero net inflow [27], mass is conserved. In an undamped mechanical system, such as one with Hamiltonian dynamics, energy conservation law holds. Likewise, in circuit analysis, Kirchhoff’s current and voltage laws hold. In a chemical reaction, the species concentrations are nonnegative. All these make the constrained state estimation necessary.

1.2 Research Objectives

As briefly summarized above, depending on applications, there are different extensions to the classical state estimation. The main objective of this research is to deal with state estimation with two types of unconventional measurements and two types of network-induced problems. They are new extensions to the traditional state estimation. One type of unconventional

measurement considered is noise-free measurements and the other is set measurement. The first type of network-induced problem is distributed estimation fusion with transformed data and the second type is the one in the presence of multiple packet dropouts.

Conventionally, a measurement is assumed to be a point in the measurement space and driven by noise. But in reality, for some applications, the measurements may be noise-free or a subset of the measurement space. For example, state estimation problems under linear or nonlinear equality constraints, with correlated or singular measurement noise can all be formulated as those with noisy and noise-free measurements. State estimation problems under linear or nonlinear inequality constraints, with quantized measurements can all be formulated as those with point and set measurements. All these unconventional measurements surely need special treatment in order to obtain the optimal or near optimal solutions efficiently.

Considering the communication constraints, it is more beneficial for local sensors to send in processed data with less communication in distributed estimation fusion. In terms of estimation performance, it is better that the distributed estimation fusion with locally processed data can achieve as close performance to the centralized fusion as possible and is numerically more robust. The objective of research on this topic is to develop distributed fusion algorithms which can achieve these desirabilities.

Due to unreliable communication between local sensors and the processing center, packet dropouts may happen during transmission. The objective of research on this topic is to propose LMMSE-optimal and MMSE-optimal estimators for state estimation in the presence of packet dropouts.

1.3 Thesis Outline

This thesis consists of six chapters which are organized as follows:

Chapter 1 presents the background and objectives of this research work.

Chapter 2 deals with optimal state estimation with noisy and noise-free measurements. Given only the first two moments and without any assumption on the rank of the measurement matrix for noise-free measurement, it is pointed out that this estimation problem is in essence just one with singular measurement noise and is not really a big deal in theory, with the optimal solution provided by the batch LMMSE estimator. Two sequential forms of the batch LMMSE estimator are obtained to reduce the computational complexity.

Chapter 3 addresses state estimation problems with point and set measurements. Inspired by the estimation with quantized measurements developed by Curry [28], under a Gaussian assumption, the MMSE state estimator with point measurements and set measurements of any shape is proposed by discretizing continuous set measurements.

Chapter 4 discusses lossless linear transformation of sensor data for distributed estimation fusion. By taking linear transformation of the raw measurements received by each sensor, two optimal distributed fusion algorithms are proposed. Compared with existing fusion algorithms, they have three nice properties.

Chapter 5 is about optimal state estimation in the presence of multiple packet dropouts. An alternative form of LMMSE estimation in the presence of multiple packet dropouts is proposed first. Then under a Gaussian assumption, the MMSE estimation is also proposed.

Chapter 6 draws conclusions from this research work and lists some future work to continue and extend this research work.

Chapter 2

Optimal Linear State Estimation with Noisy and Noise-free Measurements

2.1 Introduction and related research

As will be shown below, numerous state estimation problems can be formulated as those with both noisy and noise-free measurements ¹ (e.g., state estimation under linear or nonlinear equality constraints, with correlated or singular measurement noise.)

For the case with noise-free measurements only, one simple heuristic [74,84] is to increase the zero diagonal elements of the measurement noise covariance matrix artificially to a small number, but optimality is lost and a stabilizing effect on the numerics of the Kalman filter occurs. Another well established method is the so-called reduced-order filter [74,84], in which a smaller dimensional state is used, as the name suggests. The key idea here is to express the original state as a linear combination of the current measurement and the reduced-order state from an invertible augmentation to the original measurement equation. In this way, the possible numerical problem resulted from zero covariance matrices of the measurement noises can be circumvented. But the computational complexity is not necessarily reduced since it is mainly determined by the dimension of the measurement instead of the state.

¹A noise-free measurement is also called a perfect measurement.

The state vectors of many dynamic systems evolve according to some linear or nonlinear equality constraints. For example, in ground target tracking [65, 123, 125], if we treat roads as curves without width, the road networks can then be described by equality constraints. In the quaternion-based attitude estimation problem, the attitude vector must have a unit norm [15]. In a compartmental model with zero net inflow [27], mass is conserved. In an undamped mechanical systems, such as one with Hamiltonian dynamics, energy conservation law holds. Likewise, in circuit analysis, Kirchhoff’s current and voltage laws hold. For state estimation with noise-free measurements due to equality constraints, numerous results and methods are available [51, 61, 66, 105, 116, 119, 124, 125, 131]. For example, the reparameterization method simply reparameterizes the system model so that the equality constraints are not required any more. It has several disadvantages. First, the physical meaning of the reparameterized system state may vary and be different at different time instants. Second, the interpretation of the reparameterized system state is less natural and more difficult [105]. Another popular method for equality constrained estimation is the projection method [51, 61, 66, 105, 124, 125], in which the estimate is projected onto the constraint subspace. Unfortunately, it has problems and limitations. Its main idea is to apply classical constrained optimization techniques to the constrained estimation problem. Some other methods, e.g., maximum probability method and mean square method, were also discussed in [105]. They are not free of the limitations, either. Also, all existing work processes the noisy measurement first and then the equality constraint. Is it the only choice or a good choice? If there are more than one choice, how should the end user choose among them? Unfortunately, such questions have not been answered in theory.

A white-noise observation model is widely used in target tracking. In practice, the measurement noise may be colored [18, 19, 49, 50, 52, 77, 83, 95, 121]. A “brute force” solution is to augment the state with measurement noise and thus the problem is reduced to a standard one, but with an increased dimension and “perfect state observations,” which yields

a singular error covariance—an undesirable feature [11, 74]. To avoid this, the measurement differencing approach was first proposed in [18].

Some but not all measurement components may be so accurate that it is occasionally reasonable to assume that they are perfect, i.e., they have no error [17, 19, 44, 45, 47, 54, 59, 83, 90, 91]. This is of practical importance [47] since it is often the case, e.g., in many industrial control systems [91]. For such a state estimation problem, the Kalman filter of a full system order, with a possible numerically ill-conditioned gain matrix, is inappropriate [91]. In such cases, reduced-order filters are preferable since there is no need to estimate those states which are known exactly [54]. Actually to circumvent numerical problems, as discussed below, *Moore-Penrose pseudoinverse* (*MP inverse* for short) can be simply used to replace the inverse if the optimality criterion is LMMSE. Since the computational complexity of the Kalman filter is mainly determined by the involved inverse, which has the same dimension as the measurement, it should be noted that the computational complexity can not necessarily be reduced by just reducing the order of the state.

As shown later, the state estimation problem with both noisy and noise-free measurements is in essence just one with singular measurement noise and is not really a big deal in theory. What matters is the computational complexity of the solution. So a main focus of this chapter is to find computationally efficient ways and analyze their applicability to different scenarios.

This chapter is organized as follows. Sec. 2.2 formulates the problem. Sec. 2.3 gives some cases with noise-free measurements so as to show our research work in this direction is useful. Sec. 2.4 presents the batch LMMSE estimator. Sec. 2.5 presents two equivalent forms of the sequential LMMSE estimator. Sec. 2.6 discusses extension to nonlinear measurements. Sec. 2.7 provides supporting numerical examples. Sec. 2.8 gives summary.

2.2 Problem formulation

Consider the following generic dynamic system

$$x_k = F_{k-1}x_{k-1} + G_{k-1}w_{k-1} \quad (2.1)$$

with zero-mean white noise w_k with $\text{cov}(w_k) = Q_k \geq 0$ and $x_k \in \mathbb{R}^n$, $E[x_0] = \bar{x}_0$, $\text{cov}(x_0) = P_0$.

Assuming that two types of measurements of the system state are available at the same time. The first type is noisy:

$$z_k^{(1)} = H_k^{(1)}x_k + v_k^{(1)} \quad (2.2)$$

with zero-mean white noise $v_k^{(1)}$ with $\text{cov}(v_k^{(1)}) = R_k^{(1)} > 0$ and $z_k^{(1)} \in \mathbb{R}^{m_1}$. $\langle w_k \rangle$, $\langle v_k^{(1)} \rangle$ and x_0 are uncorrelated with each other.

The assumption $R_k^{(1)} > 0$ is not restrictive in our framework, as explained in Sec. 2.3.4.

The second type of measurement is noise free:

$$z_k^{(2)} = H_k^{(2)}x_k \quad (2.3)$$

where $z_k^{(2)} \in \mathbb{R}^{m_2}$.

One may think that $z_k^{(2)}$ is always random since it is a measurement of the state. As shown later, this is not always true.

In this work, given only the first two moments, we try to obtain the LMMSE state estimation with both noisy and noise-free measurements. That is,

$$\hat{x}_{k|k}^{\text{LMMSE}} \triangleq E^* [x_k | z^k] = \arg \min_{\hat{x}_{k|k} = a_k + B_k Z_k} \text{MSE}(\hat{x}_{k|k})$$

where

$$\begin{aligned}
 z^k &= \{z_1, \dots, z_k\}, \quad Z_k = [z'_1, \dots, z'_k]' \\
 z_k &= [(z_k^{(1)})', (z_k^{(2)})']' \\
 \text{MSE}(\hat{x}_{k|k}) &= E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})']
 \end{aligned} \tag{2.4}$$

and a_k, B_k do not depend on Z_k .

2.3 Noise-free measurements

Before getting into details about how to obtain the optimal state estimation, let us discuss some cases with both types of measurements to show that our work is not only meaningful in theory but also useful for application.

2.3.1 Linear equality constraints

In this case, a linear equality constraint is placed on the *estimand* (quantity to be estimated):

$$Dx_k = d_k \tag{2.5}$$

where matrix D and vector d_k are both known.

Let

$$z_k^{(2)} = d_k, \quad H_k^{(2)} = D$$

It can be easily seen that the above linear equality constraint is exactly a noise-free measurement. So state estimation with linear equality constraints is a special case of the problem with both noisy and noise-free measurements.

This case has already been widely studied [51, 66, 105, 116, 119, 124, 131]. Several assump-

tions were made in the derivation of their results. For example, except for [131], they all made the Gaussian assumptions. Under the Gaussian assumption, the result in [119] was claimed to be optimal in the sense of *generalized maximum likelihood* (GML) in which the correlation between the pseudo measurement noise and the estimand was totally ignored; the result in [116] and the *maximum probability method* in [105] were claimed to be optimal in the sense of *maximum a posterior probability* (MAP); the *mean square method* in [105] was claimed to be optimal in the sense of MMSE. Necessary conditions need to be satisfied by a constrained linear system, and one way to construct a homogeneously constrained linear system based on the information of an unconstrained linear system was introduced in [66]. [51,66] formulated the problem with the Gaussian assumption just to agree with the assumptions in the standard Kalman filtering. [124] provided a geometric interpretation of the results in [105], so the Gaussian assumption was maintained. Since our optimality criterion is LMMSE, we only assume the first two moments to be known, which is the same as in [131]. Another common assumption is that D is of full row rank, otherwise, we can make it full row rank by simply removing linearly dependent rows. But to obtain the rank of D and find its linearly dependent rows may not be trivial. In addition, why is this assumption needed? What is the advantage of it? This will be analyzed later. On the contrary, we do not have any restrictions on the rank of $H_k^{(2)}$.

There was one common problem in the derivations of most existing results. The objective functions are inconsistent before and after using linear constraints, although some of the results are the same as a form of our sequential LMMSE filter. Specifically, the objective function before using the linear constraints is MSE (in the average sense), while afterwards it becomes fitting error in the least-squares sense, in which the estimate is treated as data. Then, in what sense is the final estimate optimal?

2.3.2 Nonlinear equality constraints

In this case, a nonlinear equality constraint is placed on the estimand:

$$c(x_k) = 0$$

where $c(\cdot)$ is some vector-valued nonlinear function.

Let

$$z_k^{(2)} = 0, \quad h_k^{(2)}(x_k) = c(x_k)$$

Then state estimation with nonlinear equality constraints is a special case of the problem with a nonlinear measurement model (see the section “extension to nonlinear measurements” for details.)

This case has also been widely studied [61, 105, 119, 124, 125]. Under the Gaussian assumption and linearization based on a first-order TSE, the result in [119] was claimed to be approximately GML-optimal. [105] extended their state estimation results for linear equality constraints to the case with nonlinear equality constraints. Under the Gaussian assumption, a second-order TSE was utilized in [124, 125] to get better estimation results. [61] even put constraints on the statistics of the distribution of the estimate and proposed a two-step projection method. The same problem and limitations mentioned in the above subsection exist in these results.

2.3.3 Autocorrelated measurement noise

Suppose that only noisy measurements are available for the dynamic system (2.1) as follows

$$z_k^{(2)} = H_k x_k + v_k$$

where the measurement noise $\langle v_k \rangle$ is autocorrelated instead of white:

$$v_{k+1} = B_k v_k + v_k^w$$

with zero-mean white $\langle v_k^w \rangle$, uncorrelated with $\langle w_k \rangle$ and x_0 , and $\text{cov}(v_k^w) = R_k^w > 0$,

A “brute force” solution to the optimal state estimation problem in this case is as follows.

Let

$$x_k^a = \begin{bmatrix} x'_k & v'_k \end{bmatrix}', \quad F_k^a = \begin{bmatrix} F_k & 0 \\ 0 & B_k \end{bmatrix}, \quad G_k^a = \begin{bmatrix} G_k & 0 \\ 0 & I \end{bmatrix}, \quad w_k^a = \begin{bmatrix} w'_k & (v_k^w)' \end{bmatrix}'$$

$$H_k^a = \begin{bmatrix} H_k & I \end{bmatrix}$$

Then the above dynamic system can be rewritten as

$$x_k^a = F_{k-1}^a x_{k-1}^a + G_{k-1}^a w_{k-1}^a$$

$$z_k^{(2)} = H_k^a x_k^a$$

As can be seen, the original noisy measurement $z_k^{(2)}$ is now noise free with respect to (w.r.t.) the augmented state x_k^a . As such, we only have noise-free measurements. As pointed out in, e.g., [11, 18, 74], due to the increased state dimension and zero measurement noise, numerical problems may arise if the inverse is still used as in the standard Kalman filtering, which is undesirable. That is also why the “difference measurement” method [11, 74] is popular for this case. But if the inverse is replaced by the MP inverse in the singular cases, the optimal state estimation can still be obtained based on this augmented noise-free form. Also, one bonus is that we will have the optimal estimate of the measurement noise at the same time.

2.3.4 Singular measurement noise

Suppose that the measurement available for the dynamic system (2.1) is:

$$\bar{z}_k = \bar{H}_k x_k + \bar{v}_k \quad (2.6)$$

where $\bar{R}_k = \text{cov}(\bar{v}_k)$ with $|\bar{R}_k| = 0$. We show that this case can be converted to our formulation with both noisy and noise-free measurements.

It follows that

$$\text{rank}(\bar{R}_k) \triangleq m_1 < m \triangleq \dim(\bar{v}_k)$$

It then follows from singular value decomposition (SVD) that there must exist a unitary matrix \bar{U}_k such that

$$\bar{U}_k \bar{R}_k \bar{U}_k' = \begin{bmatrix} R_k^{(1)} & 0_{m_1 \times (m-m_1)} \\ 0_{(m-m_1) \times m_1} & 0_{(m-m_1) \times (m-m_1)} \end{bmatrix}$$

where $R_k^{(1)} > 0$ is an $m_1 \times m_1$ diagonal matrix. That is also why we can assume $R_k^{(1)} > 0$ in our problem formulation without loss of generality.

Let

$$z_k = \bar{U}_k \bar{z}_k$$

Then, Eq. (2.6) can be rewritten as

$$z_k = H_k x_k + v_k$$

where

$$z_k = [(z_k^{(1)})', (z_k^{(2)})']', \quad H_k = [(H_k^{(1)})', (H_k^{(2)})']' = \bar{U}_k \bar{H}_k, \quad v_k = [(v_k^{(1)})', (v_k^{(2)})']' = \bar{U}_k \bar{v}_k$$

$$z_k^{(1)} \in \mathbb{R}^{m_1}, \quad z_k^{(2)} \in \mathbb{R}^{m_2}, \quad H_k^{(1)} \in \mathbb{R}^{m_1 \times n}, \quad H_k^{(2)} \in \mathbb{R}^{m_2 \times n}, \quad v_k^{(1)} \in \mathbb{R}^{m_1}, \quad v_k^{(2)} \in \mathbb{R}^{m_2}$$

$$m_2 = m - m_1, \quad \text{cov}(v_k^{(1)}) = R_k^{(1)}, \quad \text{cov}(v_k^{(1)}, v_k^{(2)}) = 0, \quad v_k^{(2)} = 0 \text{ a.s.}$$

Since \bar{U}_k is unitary and thus invertible, $z_k = \bar{U}_k \bar{z}_k$ is sufficient in the sense that the LMMSE based on z_k is equivalent to the LMMSE based on \bar{z}_k . That is, the original noisy measurements only Eq. (2.6) is equivalent to the following:

$$z_k^{(1)} = H_k^{(1)} x_k + v_k^{(1)}$$

$$z_k^{(2)} = H_k^{(2)} x_k$$

which is exactly in the form of our formulation.

2.4 Batch LMMSE estimation

With the augmented z_k in Eq. (2.4), the stacked measurement equation can be written as

$$z_k = \begin{bmatrix} H_k^{(1)} \\ H_k^{(2)} \end{bmatrix} x_k + \begin{bmatrix} v_k^{(1)} \\ 0_{m_2 \times 1} \end{bmatrix}$$

Define

$$H_k = [(H_k^{(1)})', (H_k^{(2)})']'$$

$$v_k = [(v_k^{(1)})', 0'_{m_2 \times 1}]'$$

Then the equation becomes

$$z_k = H_k x_k + v_k$$

where $z_k \in \mathbb{R}^{m_1+m_2}$, $\langle v_k \rangle$ is zero-mean white noise with $\text{cov}(v_k) = R_k = \text{diag}(R_k^{(1)}, 0_{m_2 \times m_2})$ and uncorrelated with $\langle w_k \rangle$ and x_0 .

Given $\hat{x}_{k-1|k-1} = E^*[x_{k-1}|z^{k-1}]$, $P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1})$ and z_k , it is well known (see, e.g., [73, 81]) that the **batch LMMSE estimator** of x_k is:

Prediction:

$$\begin{aligned}\hat{x}_{k|k-1} &= E^*[x_k|z^{k-1}] = F_{k-1}\hat{x}_{k-1|k-1} \\ P_{k|k-1} &= \text{MSE}(\hat{x}_{k|k-1}) = F_{k-1}P_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}\end{aligned}$$

Update:

$$\begin{aligned}\hat{x}_{k|k} &= E^*[x_k|z^{k-1}, z_k] = \hat{x}_{k|k-1} + P_{k|k-1}H'_k S_k^+(z_k - H_k \hat{x}_{k|k-1}) \\ P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = P_{k|k-1} - P_{k|k-1}H'_k S_k^+ H_k P_{k|k-1} \\ S_k &= H_k P_{k|k-1} H'_k + R_k\end{aligned}$$

In this dissertation, A^+ stands for the unique *MP inverse* of matrix A . It reduces to A^{-1} whenever A^{-1} exists.

Compared with the standard Kalman filtering, nothing is different except that the inverse, if not exist, is now replaced by the MP inverse.

Since $R_k \geq 0$ in general, the batch LMMSE estimator with both noisy and noise-free measurements needs to calculate the MP inverse S_k^+ , which is $(m_1 + m_2) \times (m_1 + m_2)$.

Even if x_0 , w_k and $v_k^{(1)}$ all have a Gaussian distribution with a nonsingular covariance matrix, v_k will be singular Gaussian and $P_{k|k}$ may not be positive definite in this case. Since a singular Gaussian has no density function, the posterior density $f(x_k|z^k) = \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$

of x_k is nonexistent and the MAP estimate of x_k does not exist if $P_{k|k}$ is not positive definite. That is also one of the reasons why we choose LMMSE as our optimality criterion.

2.5 Sequential LMMSE estimation

To reduce the computational complexity of the batch LMMSE estimator, we obtain two equivalent sequential forms of the LMMSE estimator. They process the noisy and noise-free measurements sequentially and thus have reduced computation.

2.5.1 Form 1

Theorem 1 (Sequential LMMSE, form 1). Given $\hat{x}_{k-1|k-1} = E^*[x_{k-1}|z^{k-1}]$, $P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1})$ and z_k , one form of the sequential LMMSE estimator of x_k is:

Prediction: Same as in the batch LMMSE estimator.

Update by the noisy measurement:

$$\hat{x}_{k|k}^{(1)} = \hat{x}_{k|k-1} + P_{k|k-1}(H_k^{(1)})'(S_k^{(1)})^{-1}(z_k^{(1)} - H_k^{(1)}\hat{x}_{k|k-1}) \quad (2.7)$$

$$P_{k|k}^{(1)} = P_{k|k-1} - P_{k|k-1}(H_k^{(1)})'(S_k^{(1)})^{-1}H_k^{(1)}P_{k|k-1}$$

$$S_k^{(1)} = H_k^{(1)}P_{k|k-1}(H_k^{(1)})' + R_k^{(1)} \quad (2.8)$$

Update by the noise-free measurement:

$$\hat{x}_{k|k} = \hat{x}_{k|k}^{(1)} + P_{k|k}^{(1)}(H_k^{(2)})'(S_k^{(2)})^+(z_k^{(2)} - H_k^{(2)}\hat{x}_{k|k}^{(1)}) \quad (2.9)$$

$$P_{k|k} = P_{k|k}^{(1)} - P_{k|k}^{(1)}(H_k^{(2)})'(S_k^{(2)})^+H_k^{(2)}P_{k|k}^{(1)} \quad (2.10)$$

$$S_k^{(2)} = H_k^{(2)}P_{k|k}^{(1)}(H_k^{(2)})' \quad (2.11)$$

Proof: Given $z_k^{(1)}$, the updated LMMSE estimate (2.7)-(2.8) follows directly from $\hat{x}_{k|k}^{(1)} =$

$E^*[x_k|z^{k-1}, z_k^{(1)}]$ and $P_{k|k}^{(1)} = \text{MSE}(\hat{x}_{k|k}^{(1)})$.

Since the LMMSE estimator $E^*[x_k|z^{k-1}, z_k]$ always has the *quasi-recursive form* [73]:

$$\hat{x}_{k|k} = E^*[x_k|z^{k-1}, z_k] = E^*[x_k|z^{k-1}, z_k^{(1)}, z_k^{(2)}] = \hat{x}_{k|k}^{(1)} + C_{1,2}C_{\tilde{z}_{2|1}^*}^+ \tilde{z}_{2|1}^*$$

where

$$\begin{aligned} \tilde{z}_{2|1}^* &= z_k^{(2)} - E^*[z_k^{(2)}|z^{k-1}, z_k^{(1)}] = z_k^{(2)} - H_k^{(2)}\hat{x}_{k|k}^{(1)} = H_k^{(2)}(x_k - \hat{x}_{k|k}^{(1)}) \\ C_{\tilde{z}_{2|1}^*} &= \text{cov}(\tilde{z}_{2|1}^*) = H_k^{(2)}P_{k|k}^{(1)}(H_k^{(2)})' \triangleq S_k^{(2)} \\ C_{1,2} &= \text{cov}(\tilde{x}_{k|k}^{(1)}, \tilde{z}_{2|1}^*) = P_{k|k}^{(1)}(H_k^{(2)})' \end{aligned}$$

(2.9) follows. Also,

$$P_{k|k} = \text{MSE}(\hat{x}_{k|k}) = P_{k|k}^{(1)} - C_{1,2}C_{\tilde{z}_{2|1}^*}^+ C_{1,2}' = P_{k|k}^{(1)} - P_{k|k}^{(1)}(H_k^{(2)})'(S_k^{(2)})^+ H_k^{(2)}P_{k|k}^{(1)}$$

□

Remark: This sequential LMMSE estimator needs to calculate an $m_1 \times m_1$ inverse $(S_k^{(1)})^{-1}$ and an $m_2 \times m_2$ MP inverse $(S_k^{(2)})^+$. This is less demanding than computing the $(m_1 + m_2) \times (m_1 + m_2)$ MP inverse S_k^+ in the batch LMMSE estimator.

Remark: If $P_{k|k}^{(1)} > 0$ and $H_k^{(2)}$ is of full row rank, then $(S_k^{(2)})^+$ can be replaced by $(S_k^{(2)})^{-1}$ and this sequential LMMSE estimator will be exactly the same as the one in [105] for a linear equality constrained state estimation problem. But there is still a significant difference. This sequential LMMSE estimator is free of the problem and limitations of the existing results discussed above.

Remark: If $H_k^{(2)}$ is of full row rank, a nice equivalent weighted average form of the

LMMSE estimator of x_k was presented in [131]:

$$\hat{x}_{k|k} = (I - J_k)\hat{x}_{k|k}^{(1)} + J_k(H_k^{(2)})^+ z_k^{(2)} \quad (2.12)$$

where

$$J_k = P_{k|k}^{(1)}(H_k^{(2)})'(S_k^{(2)})^+ H_k^{(2)}$$

This can also be easily proved from our sequential LMMSE estimator since $H_k^{(2)}(H_k^{(2)})^+ = I$. This weighted average form does provide a better understanding of the mathematical meaning behind $\hat{x}_{k|k}$, which is a weighted average of $\hat{x}_{k|k}^{(1)}$ and $(H_k^{(2)})^+ z_k^{(2)}$ (a particular solution to the linear equality constraint equation (2.5)). But computationally, this form is not preferred for two reasons. First, it requires two MP inverses: $(H_k^{(2)})^+ = (H_k^{(2)})'(H_k^{(2)}(H_k^{(2)})')^{-1}$ and $(S_k^{(2)})^+$. Our sequential LMMSE estimator only needs $(S_k^{(2)})^+$. Second, it is valid only when $H_k^{(2)}$ is of full row rank, but our sequential LMMSE estimator does not have this limitation.

2.5.2 Form 2

Theorem 2 (Sequential LMMSE, form 2). Given $\hat{x}_{k-1|k-1} = E^*[x_{k-1}|z^{k-1}]$, $P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1})$ and z_k , an alternative form of the sequential LMMSE estimator of x_k is:

Prediction: Same as in the batch LMMSE estimator.

Update by the noise-free measurement:

$$\begin{aligned} \hat{x}_{k|k}^{(2)} &= E^*[x_k|z^{k-1}, z_k^{(2)}] = \hat{x}_{k|k-1} + P_{k|k-1}(H_k^{(2)})'(S_k^{(2)})^+(z_k^{(2)} - H_k^{(2)}\hat{x}_{k|k-1}) \\ P_{k|k}^{(2)} &= \text{MSE}(\hat{x}_{k|k}^{(2)}) = P_{k|k-1} - P_{k|k-1}(H_k^{(2)})'(S_k^{(2)})^+ H_k^{(2)} P_{k|k-1} \\ S_k^{(2)} &= H_k^{(2)} P_{k|k-1} (H_k^{(2)})' \end{aligned}$$

Update by the noisy measurement:

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k}^{(2)} + P_{k|k}^{(2)}(H_k^{(1)})'(S_k^{(1)})^{-1}(z_k^{(1)} - H_k^{(1)}\hat{x}_{k|k}^{(2)}) \\ P_{k|k} &= P_{k|k}^{(2)} - P_{k|k}^{(2)}(H_k^{(1)})'(S_k^{(1)})^{-1}H_k^{(1)}P_{k|k}^{(2)} \\ S_k^{(1)} &= H_k^{(1)}P_{k|k}^{(2)}(H_k^{(1)})' + R_k^{(1)}\end{aligned}$$

Proof: Parallel to that of Theorem 1.

Remark: If the noise-free measurement is due to some linear equality constraint (2.5), this form is contrary to the common practice of processing the constraints later.

Remark: Form 1 and form 2 have the same computational complexity.

Remark: If $H_k^{(2)}$ is of full row rank, then a nice equivalent weighted average form of $\hat{x}_{k|k}^{(2)}$ is

$$\hat{x}_{k|k}^{(2)} = (I - L_k)\hat{x}_{k|k-1} + L_k(H_k^{(2)})^+ z_k^{(2)}$$

where

$$L_k = P_{k|k-1}(H_k^{(2)})'(S_k^{(2)})^+ H_k^{(2)}$$

which can be easily proved from form 2 using $H_k^{(2)}(H_k^{(2)})^+ = I$. The same remarks as for form 1 can be made here.

Remark: Since both forms of the sequential LMMSE estimator are equivalent to the batch LMMSE estimator, both forms have exactly the same performance. Then, the processing order for $z_k^{(1)}$ and $z_k^{(2)}$ does not matter in either performance or computation.

If the noise-free measurement is from a linear equality constraint (2.3) and the estimator is initialized by \bar{x}_0 and P_0 , the following theorem shows that the LMMSE update by the noise-free measurement in both of the sequential forms can be simply skipped without a performance loss. That is, both forms of the sequential LMMSE estimator reduce to one single form which only has the update by the noisy measurement.

Theorem 3. If the noise-free measurement is from a linear equality constraint (2.3) and $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$, then for sequential LMMSE estimator form 1, we have

$$\hat{x}_{k|k} = \hat{x}_{k|k}^{(1)}, \quad P_{k|k} = P_{k|k}^{(1)}$$

and for sequential LMMSE estimator form 2, we have

$$\hat{x}_{k|k}^{(2)} = \hat{x}_{k|k-1}, \quad P_{k|k}^{(2)} = P_{k|k-1}$$

Proof: In this case, $z_k^{(2)}$ is known deterministically. If $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$, it then follows from (2.3) that for form 1

$$z_k^{(2)} = E^*[z_k^{(2)} | z^{k-1}, z_k^{(1)}] = H_k^{(2)} \hat{x}_{k|k}^{(1)}$$

and for form 2

$$z_k^{(2)} = E^*[z_k^{(2)} | z^{k-1}] = H_k^{(2)} \hat{x}_{k|k-1}$$

The theorem can then be easily shown. □

Remark: This result for constrained state estimation agrees with Theorem 2 in [132] which dealt with constrained least-squares parameter estimation.

Remark: If the noise-free measurement is from a linear equality constraint, then \bar{x}_0 and P_0 can not be designated arbitrarily and should satisfy some necessary conditions, as pointed out in [66].

Remark: It should be noted that Theorem 3 is obtained under the assumption that the constrained LMMSE estimation can be obtained optimally. In practice, this optimality can not be guaranteed due to model mismatch (e.g., in target tracking) or nonlinearity in the system model, so the update by the noise-free measurement should still be kept to improve

estimation performance. Note also that if the noise-free measurement is random, Theorem 3 does not hold in general. Numerical examples to show these points are provided later.

If the noise-free measurement is from a linear equality constraint, then we need to double check whether the estimate really satisfies the constraint. The following theorem shows that all estimates really do so.

Theorem 4. For estimates from all forms, we have

$$H_k^{(2)} \hat{x}_{k|k-1} = z_k^{(2)}, \quad H_k^{(2)} \hat{x}_{k|k} = z_k^{(2)}, \quad H_k^{(2)} \hat{x}_{k|k}^{(1)} = z_k^{(2)}, \quad H_k^{(2)} \hat{x}_{k|k}^{(2)} = z_k^{(2)}$$

Proof: Since the batch and sequential LMMSE estimators are equivalent, only sequential form 1 is used to show that $H_k^{(2)} \hat{x}_{k|k} = z_k^{(2)}$ for simplicity.

It follows from Eq. (2.9) that

$$z_k^{(2)} - H_k^{(2)} \hat{x}_{k|k} = (I - S_k^{(2)}(S_k^{(2)})^+) H_k^{(2)} (x_k - \hat{x}_{k|k}^{(1)})$$

and thus, by the unbiasedness of $\hat{x}_{k|k}^{(1)}$,

$$E[z_k^{(2)} - H_k^{(2)} \hat{x}_{k|k}] = 0$$

and

$$\text{cov}(z_k^{(2)} - H_k^{(2)} \hat{x}_{k|k}) = (I - S_k^{(2)}(S_k^{(2)})^+) S_k^{(2)} (I - (S_k^{(2)})^+ S_k^{(2)}) = 0$$

So $z_k^{(2)} = H_k^{(2)} \hat{x}_{k|k}$ almost surely.

$H_k^{(2)} \hat{x}_{k|k}^{(2)} = z_k^{(2)}$ can be proved similarly. $H_k^{(2)} \hat{x}_{k|k}^{(1)} = z_k^{(2)}$ and $H_k^{(2)} \hat{x}_{k|k-1} = z_k^{(2)}$ then follow from Theorem 3. □

Remark: $H_k^{(2)} \hat{x}_{k|k} = z_k^{(2)}$ and $H_k^{(2)} \hat{x}_{k|k}^{(2)} = z_k^{(2)}$ actually hold for all noise-free measurement cases, e.g., when $z_k^{(2)}$ is random (if the noise-free measurement is not from a linear equality

constraint).

Since $\hat{x}_{k|k}$ is the LMMSE estimate of x_k given z^k and $P_{k|k}$ is the corresponding MSE matrix, they have the following properties:

- $P_{k|k} \leq P_{k|k}^{(1)} \leq P_{k|k-1}$ in general. But if the noise-free measurement is from a linear equality constraint and $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$, we have $P_{k|k} = P_{k|k}^{(1)}$.
- $P_{k|k} \leq P_{k|k}^{(2)} \leq P_{k|k-1}$ in general. But if the noise-free measurement is from a linear equality constraint and $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$, we have $P_{k|k}^{(2)} = P_{k|k-1}$.
- If x_0 , w_k and $v_k^{(1)}$ are jointly Gaussian distributed, $\hat{x}_{k|k}$ and $P_{k|k}$ are also optimal in the sense of MMSE.

2.6 Extension to nonlinear measurement

Although the batch LMMSE estimator is optimal, due to its heavier computational burden, it is not preferred. If the computational burden is concerned, the sequential forms are preferred.

Both sequential forms are equivalent to the batch LMMSE estimator and have the same computational complexity. Since form 1 is already simple enough, what is the specific reason or advantage to adopt and derive form 2? What is the benefit to have these two forms? How should the user choose between them? There should be no preference between these two forms if only linear measurements are involved. But if one or both of $z_k^{(1)}$ and $z_k^{(2)}$ are nonlinear, there is a preference. That is, with linear measurements only, the performance is independent of the processing order of $z_k^{(1)}$ and $z_k^{(2)}$. With nonlinear measurements, the order matters.

As demonstrated in [31, 41], for nonlinear filtering, sequential processing can not only reduce computational complexity but also improve performance (accuracy). This is because

when the less nonlinear measurement is processed first, prediction and update by the more nonlinear measurement will have a better reference point for function approximation based nonlinear filters [78] (e.g., EKF and DD2 [89]) and sigma points or quadrature points for moment approximation based nonlinear filters [78] (e.g., UF [60] and Gaussian Hermite filter (GHF) [57]). This will be used as guideline for sequential processing in nonlinear filtering.

2.6.1 Nonlinear noisy measurement

In this case, the noisy measurement is nonlinear

$$z_k^{(1)} = h_k^{(1)}(x_k, v_k^{(1)}) \quad (2.13)$$

while the noise-free measurement is still linear.

Since the noise-free measurement is linear, we should use it first for update. The update $\hat{x}_{k|k}^{(2)}$ and $P_{k|k}^{(2)}$ by using the noise-free measurement will be optimal. Then prediction and the update by using the noisy measurement should have better accuracy than if the noise-free measurement is used after the noisy one.

In each cycle, prediction and the update by noise-free measurement are exactly the same as in Theorem 2, and the update by noisy measurement can be done as follows:

$$\hat{x}_{k|k} = \hat{x}_{k|k}^{(2)} + C_{2,1} C_{\tilde{z}_{1|2}^*}^+ \tilde{z}_{1|2}^* \quad (2.14)$$

$$P_{k|k} = P_{k|k}^{(2)} - C_{2,1} C_{\tilde{z}_{1|2}^*}^+ C_{2,1}' \quad (2.15)$$

$$\tilde{z}_{1|2}^* = z_k^{(1)} - E^*[z_k^{(1)} | z^{k-1}, z_k^{(2)}] \quad (2.16)$$

$$C_{\tilde{z}_{1|2}^*} = \text{cov}(\tilde{z}_{1|2}^*) \quad (2.17)$$

$$C_{2,1} = \text{cov}(\tilde{x}_{k|k}^{(2)}, \tilde{z}_{1|2}^*) \quad (2.18)$$

2.6.2 Nonlinear noise-free measurement

In this case, the noisy measurement is still linear while the noise-free measurement is nonlinear:

$$z_k^{(2)} = h_k^{(2)}(x_k) \quad (2.19)$$

For the same reason as in the above subsection, we should update by the noisy measurement first, and then the noise-free measurement.

In each cycle, prediction and sequential update steps are the same as in the above subsection with $z_k^{(1)}$ and $z_k^{(2)}$ interchanged.

Remark: If the noise-free measurement is from a nonlinear equality constraint and the LMMSE estimator is initialized by $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$, it can be easily shown as in Theorem 3 that the update by the noise-free measurement can be skipped.

2.6.3 Nonlinear noisy and noise-free measurements

In this case, both the noisy measurement (2.13) and the noise-free measurement (2.19) are nonlinear.

In this case, we need to first measure the nonlinearity of $h_k^{(1)}(\cdot)$ and $h_k^{(2)}(\cdot)$. If $h_k^{(j)}(\cdot)$ is less nonlinear than $h_k^{(i)}(\cdot)$, $i, j = 1, 2$, $i \neq j$, we can use the following general equations to obtain the sequential LMMSE estimator.

In each cycle, the prediction step is exactly the same as in Theorems 1 or 2, and the

update by $z_k^{(j)}$ can be done as:

$$\begin{aligned}\hat{x}_{k|k}^{(j)} &= E^*[x_k|z^{k-1}, z_k^{(j)}] = \hat{x}_{k|k-1} + C_{\tilde{x}_{k|k-1}\tilde{z}_j^*} C_{\tilde{z}_j^*}^+ \tilde{z}_j^* \\ P_{k|k}^{(j)} &= P_{k|k-1} - C_{\tilde{x}_{k|k-1}\tilde{z}_j^*} C_{\tilde{z}_j^*}^+ C_{\tilde{x}_{k|k-1}\tilde{z}_j^*}' \\ \tilde{z}_j^* &= z_k^{(j)} - E^*[z_k^{(j)}|z^{k-1}] \\ C_{\tilde{z}_j^*} &= \text{cov}(\tilde{z}_j^*), \quad C_{\tilde{x}_{k|k-1}\tilde{z}_j^*} = \text{cov}(\tilde{x}_{k|k-1}, \tilde{z}_j^*)\end{aligned}$$

The update by $z_k^{(i)}$ can be done as (2.14) through (2.18) with $z_k^{(1)}$ replaced by $z_k^{(j)}$ and $z_k^{(2)}$ replaced by $z_k^{(i)}$.

Remark: Due to the nonlinearity of $z_k^{(i)}$ or $z_k^{(j)}$ or both, we do not have an elegant analytical form for $E^*[z_k^{(j)}|z^{k-1}]$, $E^*[z_k^{(i)}|z^{k-1}, z_k^{(j)}]$, $C_{\tilde{z}_j^*}$, $C_{\tilde{z}_i^*}$, $C_{\tilde{x}_{k|k-1}\tilde{z}_j^*}$ and $C_{j,i}$ in general, but they can be approximated by EKF, UF, DD2, GHF and even from the original definition of LMMSE estimation, as in [130].

Remark: If the above extension to nonlinear measurements is used to handle constrained estimation problem, then one natural theoretical question is whether the final estimate satisfies the given constraints. It should be noted that the optimality criterion used in this work is LMMSE conditioned on all measurement information up to the current time. Since the equality constraints are already included in the conditioning, $\hat{x}_{k|k}$ and $P_{k|k}$ should have achieved this goal approximately. If $\hat{x}_{k|k}$ satisfies the equality constraints (e.g., when $\hat{x}_{k|k}$ and $P_{k|k}$ are obtained precisely without any approximation), then it is what we are looking for, otherwise we have to make a choice between the LMMSE optimality criterion and the equality constraints. If the criterion is chosen, then $\hat{x}_{k|k}$ is what we are looking for, otherwise we need to project $\hat{x}_{k|k}$ to the closest point in the constraint subspace.

2.7 Illustrative examples

By Theorem 3, if the noise-free measurements are from equality constraints, then under certain conditions, the update by the equality constraints is redundant. We also pointed out that if there exists model mismatch or nonlinearity or if $z_k^{(2)}$ is random, two-step update will improve performance in general. In the following, we further verify these findings through numerical examples.

Consider the following dynamic system, which describes the motion of an on-road vehicle [66]:

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1}$$

where

$$x_k = [x_k \quad y_k \quad \dot{x}_k \quad \dot{y}_k]', \quad w_k \sim \mathcal{N}(0, Q_k), \quad x_0 \sim \mathcal{N}(\bar{x}_0, P_0), \quad P_0 = N_k P_0^u N_k$$

$$F_k = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G_k = \begin{bmatrix} 0 \\ 0 \\ T \sin \theta \\ T \cos \theta \end{bmatrix}, \quad Q_k = \begin{bmatrix} 30 & 10\sqrt{3} & 0 & 0 \\ 10\sqrt{3} & 10 & 0 & 0 \\ 0 & 0 & 10 & 10\sqrt{3}/3 \\ 0 & 0 & 10\sqrt{3}/3 & 10/3 \end{bmatrix}$$

$$N_k = I_4 - (H_k^{(2)})'(H_k^{(2)}(H_k^{(2)})')^{-1}H_k^{(2)}, \quad H_k^{(2)} = [0 \quad 0 \quad 1 \quad -\tan \theta]$$

$$\theta = \pi/3, \quad T = 2, \quad u_k = \begin{cases} 1, & \text{if } k \text{ is odd} \\ -1, & \text{if } k \text{ is even} \end{cases}$$

\bar{x}_0 and P_0^u will be given later. The state satisfies the following linear equality constraint [66]

$$H_k^{(2)}x_k = 0 \tag{2.20}$$

This is because the angle between the y axis and the road (treated as a straight line without

width) is θ .

Figs. 2.1 and 2.2 show the true vehicle trajectory and velocity in one run in the two-dimensional Cartesian coordinate plane. It can be seen that the given linear equality constraint is satisfied.

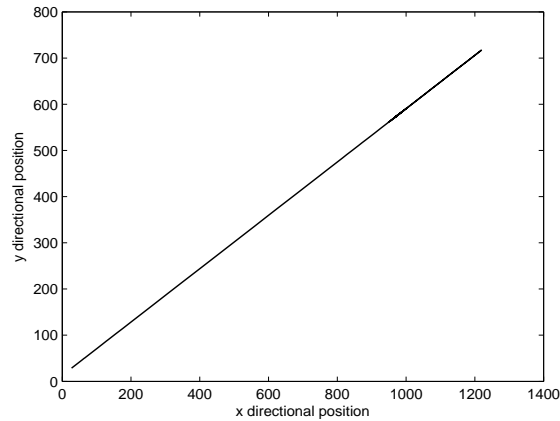


Figure 2.1: Vehicle trajectory in one run

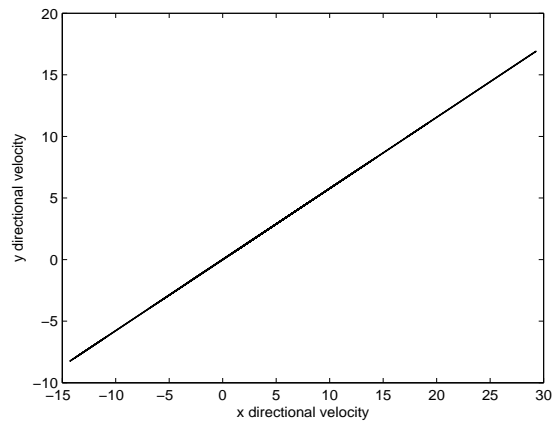


Figure 2.2: Vehicle velocity in one run

We want to estimate the state in the LMMSE sense based on the measurements and constraint. In the following, all estimators were initialized with $P_{0|0} = P_0$. A mismatched

Q_k :

$$Q_k^{\text{mis}} = \begin{bmatrix} 30 & 10\sqrt{3} & 0 & 0 \\ 10\sqrt{3} & 10 & 0 & 0 \\ 0 & 0 & 13.25 & 0.1443 \\ 0 & 0 & 0.1443 & 13.0833 \end{bmatrix}$$

is used in place of Q_k in some estimators, possibly along with a mis-specified initial estimate

$$\hat{x}_{0|0}^{\text{mis}} = \bar{x}_0 + [0\text{m} \quad 0\text{m} \quad 6\text{m/s} \quad 12\text{m/s}]'$$

Also, all results were averaged over 200 Monte Carlo runs.

2.7.1 Example 1

In this example, the measurement provided by one type of sensor is described by

$$\begin{aligned} z_k^{(1)} &= H_k^{(1)} x_k + v_k^{(1)} \\ z_k^{(3)} &= H_k^{(3)} x_k \end{aligned}$$

where

$$v_k^{(1)} \sim \mathcal{N}(0, R_k^{(1)}), \quad H_k^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad H_k^{(3)} = [2 \quad 3 \quad 0 \quad 0], \quad R_k^{(1)} = \text{diag}\{400\text{m}^2, 400\text{m}^2\}$$

That is, $z_k^{(3)}$ is a perfect measurement, which is random. It is also known that

$$\bar{x}_0 = [0\text{m} \quad 0\text{m} \quad 10\sqrt{3}\text{m/s} \quad 10\text{m/s}]', \quad P_0^u = \text{diag}(400\text{m}^2, 400\text{m}^2, 10(\text{m/s})^2, 10(\text{m/s})^2)$$

Figs. 2.3 and 2.4 show comparison results of LMMSE estimators in Table 2.1, all of

which use the true Q_k and are initialized with $\hat{x}_{0|0} = \bar{x}_0$.

Table 2.1: LMMSE estimators used in example 1

name	explanation
\overline{POEC}	with neither perfect observation nor equality constraint
\overline{POEC}	without perfect observation but with equality constraint
$POEC$	with perfect observation but without equality constraint
$POEC$	with both perfect observation and equality constraint

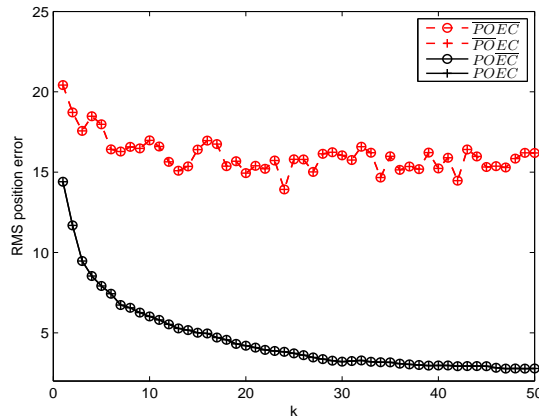


Figure 2.3: RMS position error comparison for example 1. Note that \overline{POEC} overlaps with $POEC$ and $POEC$ overlaps with $POEC$.

As is clear from the simulation results, on the one hand, the update by the equality constraint is really unnecessary—it does not improve performance. On the other hand, the update by the perfect observations is indispensable—it leads to a significant performance improvement. It should be noted that under our problem formulation, both the perfect observation $z_k^{(3)}$ and the linear equality constraint (2.20) are treated as noise-free measurements. But as demonstrated in this example, their effects are totally different.

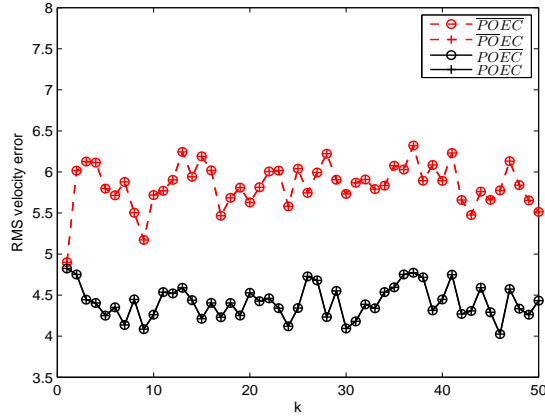


Figure 2.4: RMS velocity error comparison for example 1. Note that \overline{POEC} overlaps with $POEC$ and $POEC$ overlaps with $POEC$.

2.7.2 Example 2

In this example from [66], the measurement provided by another type of sensor is described by

$$z_k^{(1)} = H_k^{(1)} x_k + v_k^{(1)}$$

where

$$v_k^{(1)} \sim \mathcal{N}(0, R_k^{(1)}), \quad H_k^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad R_k^{(1)} = \text{diag}(400\text{m}^2, 400\text{m}^2, 10(\text{m/s})^2)$$

It is also known that

$$\bar{x}_0 = [0\text{m} \quad 0\text{m} \quad 10\sqrt{3}\text{m/s} \quad 10\text{m/s}]', \quad P_0^u = \text{diag}(400\text{m}^2, 400\text{m}^2, 10(\text{m/s})^2, 10(\text{m/s})^2)$$

Figs. 2.5, 2.6 and 2.7 show comparison results of LMMSE estimators in Table 2.2.

Checking against the condition stated in Theorem 3, it can be easily justified that the update by the linear equality constraint can be skipped in the LMMSE filter with correct Q

Table 2.2: LMMSE estimators used in example 2

name	explanation
KF	true Q_k , $\hat{x}_{0 0} = \bar{x}_0$, updated only by noisy measurement $z_k^{(1)}$
NM-Q	Q_k^{mis} , $\hat{x}_{0 0} = \bar{x}_0$, updated only by noisy measurement $z_k^{(1)}$
NM-EC-Q	Q_k^{mis} , $\hat{x}_{0 0} = \bar{x}_0$, updated by $z_k^{(1)}$ first and then by the equality constraint
EC-NM-Q	Q_k^{mis} , $\hat{x}_{0 0} = \bar{x}_0$, updated by the equality constraint first and then by $z_k^{(1)}$
NM-0	true Q_k , $\hat{x}_{0 0} = \hat{x}_{0 0}^{\text{mis}}$, updated only by $z_k^{(1)}$
NM-EC-0	true Q_k , $\hat{x}_{0 0} = \hat{x}_{0 0}^{\text{mis}}$, updated by $z_k^{(1)}$ first and then by the equality constraint
EC-NM-0	true Q_k , $\hat{x}_{0 0} = \hat{x}_{0 0}^{\text{mis}}$, updated by the equality constraint first and then by $z_k^{(1)}$

and $\hat{x}_{0|0}$. In other words, the KF above provides the optimal estimate for this constrained estimation problem. That is also why we did not consider the two processing orders described by Theorems 1 and 2 in this example of no model mismatch. Note that Fig. 2.6 is a zoomed-in version of Fig. 2.5 in that the NM-0 estimator is removed in order to see the difference among the other estimators more clearly.

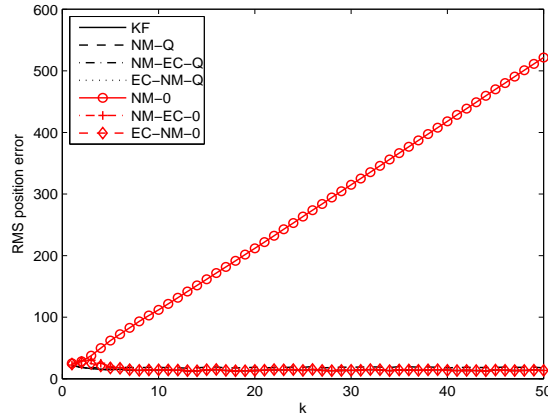


Figure 2.5: RMS position error comparison for example 2. Note that NM-0 diverges and its error curve is far above the rest.

It can be seen from the simulation results that the best performance is achieved by the KF, as analyzed above. If there is Q mismatch and the linear equality constraint is not

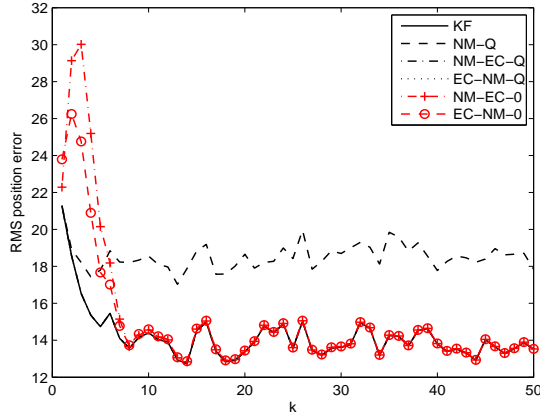


Figure 2.6: RMS position error comparison for example 2 without NM-0. Note that KF, NM-EC-Q and EC-NM-Q essentially overlap with each other.

taken into account in the update step, the NM-Q filter has much worse performance than the KF. But if there is Q mismatch and the linear equality constraint is fully accounted for in the update step, the performance of the NM-EC-Q and EC-NM-Q filters are almost the same as that of the KF. If $\hat{x}_{0|0}$ is mis-specified and the linear equality constraint is not taken into account in the update step, the NM-0 estimator has the worst performance and its position RMSE diverges. But if $\hat{x}_{0|0}$ is mis-specified and the linear equality constraint is fully taken into account in the update step, the NM-EC-0 and EC-NM-0 filters have almost the same performance as the KF during the steady state but are worse than the KF during the short transient. All these further verify our statements about the contribution of equality constraints in practical constrained estimation problems. We should take the linear equality constraint into account when there exists model mismatch. Also note that for performance there is almost no difference between the NM-EC-Q and EC-NM-Q filters, but there does exist difference between the NM-EC-0 and EC-NM-0 filters during transient. This means that the processing order of $z_k^{(1)}$ and the linear equality constraint does not matter at all when there exists Q mismatch, but it matters when $\hat{x}_{0|0}$ is mis-specified. And as expected, better performance is achieved when the linear equality constraint is processed before $z_k^{(1)}$.

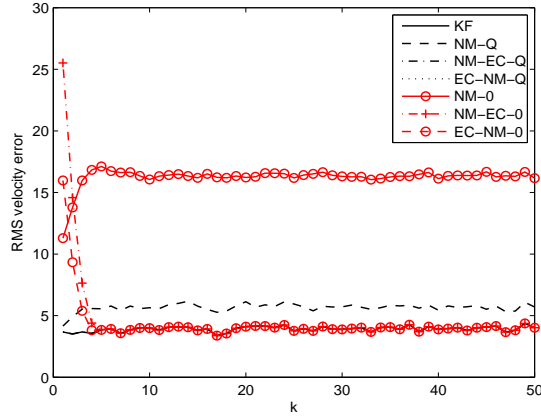


Figure 2.7: RMS velocity error comparison for example 2. Note that KF, NM-EC-Q and EC-NM-Q essentially overlap with each other.

2.7.3 Example 3

In this example, the range and bearing measurements provided by a radar located at the origin are described by

$$z_k^{(1)} = h_k^{(1)}(x_k) + v_k^{(1)}$$

where

$$h_k^{(1)}(x_k) = \begin{bmatrix} \sqrt{x_k^2 + y_k^2} \\ \tan^{-1}(y_k/x_k) \end{bmatrix}, v_k^{(1)} \sim \mathcal{N}(0, R_k^{(1)}), R_k^{(1)} = \text{diag}((80\text{m})^2, (10\text{mrad})^2)$$

It is also known that

$$\bar{x}_0 = [10^4\sqrt{3}\text{m} \quad 10^4\text{m} \quad 10^3\sqrt{3}\text{m/s} \quad 10^3\text{m/s}]'$$

$$P_0^u = \text{diag}((4000\text{m})^2, (4000\text{m})^2, (300\text{m/s})^2, (300\text{m/s})^2)$$

Figs. 2.8 through 2.13 show comparison results of state estimators in Table 2.3.

All estimators in Figs. 2.8 and 2.9 use the true Q_k and are initialized with $\hat{x}_{0|0} = \bar{x}_0$.

Table 2.3: State estimators used in example 3

name	explanation
BNM	updated only by noisy measurement $z_k^{(1)}$ (using LMMSE estimator in [130])
BNM-EC	updated by noisy measurement $z_k^{(1)}$ (using LMMSE estimator in [130]) first and then by the equality constraint
EC-BNM	updated by the equality constraint first and then by $z_k^{(1)}$ (using LMMSE estimator in [130])
ENM	updated only by $z_k^{(1)}$ (using EKF)
ENM-EC	updated by $z_k^{(1)}$ (using EKF) first and then by the equality constraint
EC-ENM	updated by the equality constraint first and then by $z_k^{(1)}$ (using EKF)

Estimators in Figs. 2.10 and 2.11 have Q mismatch but have no mis-specification on $\hat{x}_{0|0}$.

Estimators in Figs. 2.12 and 2.13 use correct Q but mis-specified $\hat{x}_{0|0}$.

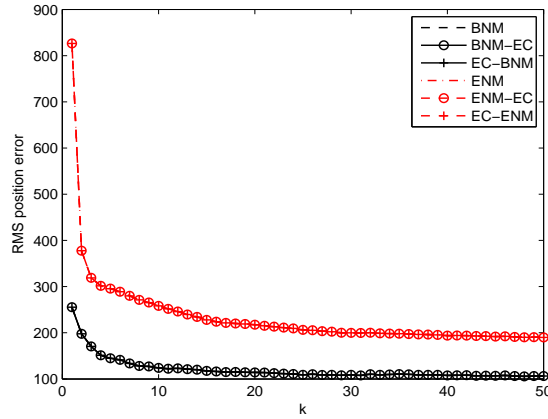


Figure 2.8: RMS position error comparison for example 3 with correct Q and $\hat{x}_{0|0} = \bar{x}_0$. Note that BNM, BNM-EC and EC-BNM essentially overlap with each other, and ENM, ENM-EC and EC-ENM essentially overlap with each other.

Due to the nonlinearity of the measurement $z_k^{(1)}$, state estimation in this example is a nonlinear filtering problem. In general, only suboptimal results can be obtained.

From the simulation results in Figs. 2.8 and 2.9, it can be seen that the BNM, BNM-EC and EC-BNM filters have almost the same performance. Also, the ENM, ENM-EC

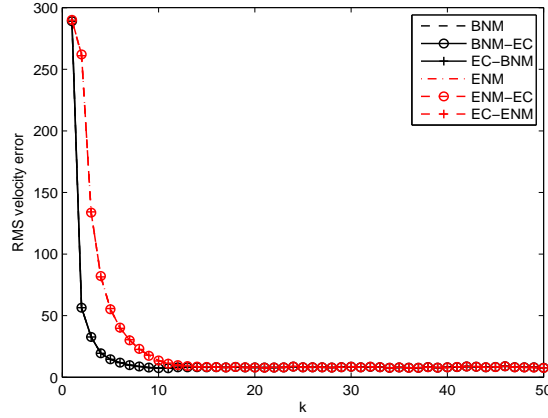


Figure 2.9: RMS velocity error comparison for example 3 with correct Q and $\hat{x}_{0|0} = \bar{x}_0$. Note that BNM, BNM-EC and EC-BNM essentially overlap with each other, and ENM, ENM-EC and EC-ENM essentially overlap with each other.

and EC-ENM filters have almost the same performance. This means that the update by linear equality constraint is still unnecessary even in this nonlinear filtering problem. The update does not change the filtering results at all, which also implies that the updated estimate by $z_k^{(1)}$ already satisfies the constraint. It can also be seen that a direct application of the LMMSE update for nonlinear measurement $z_k^{(1)}$ is completely feasible. For nonlinear measurement $z_k^{(1)}$, the LMMSE update beats the EKF update by far in terms of position RMS error and transient velocity RMS error. This is mainly due to the nice properties held by the approximate LMMSE update, as analyzed in detail in [130]. Surely, other approximation methods, e.g., UF, DD2, GHF, can also be applied. But since nonlinear filtering is not the focus of this work, they will not be compared here.

From the simulation results in Figs. 2.10 through 2.13, it can be seen that the LMMSE filter beats the EKF in every case. If there is Q mismatch or mis-specification on $\hat{x}_{0|0}$ and the linear equality constraint is not taken into account in the update step, the BNM filter or the ENM filter diverges in terms of position RMS error and has much worse performance than their counterparts which take the linear equality constraint into full account in the update step. This means that the update by linear equality constraint is really helpful when there is

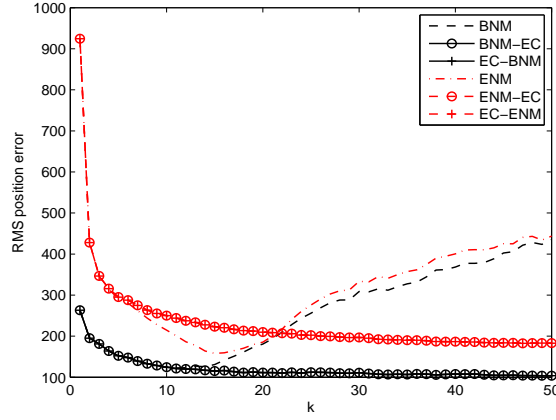


Figure 2.10: RMS position error comparison for example 3 with Q mismatch. Note that BNM-EC essentially overlaps with EC-BNM, and ENM-EC essentially overlaps with EC-ENM.

Q mismatch or mis-specification on $\hat{x}_{0|0}$ in nonlinear filtering. Also note that for performance there is almost no difference between the BNM-EC and EC-BNM filters, and between the ENM-EC and EC-ENM filters. This means that the processing order of $z_k^{(1)}$ and the linear equality constraint does not matter when there exists Q mismatch or mis-specification on $\hat{x}_{0|0}$ in this nonlinear filtering example.

2.7.4 Example 4

In this example from [105], the on-road vehicle is equipped to measure its ranges relative to two reference points $(x_r^{(1)}, y_r^{(1)})$ and $(x_r^{(2)}, y_r^{(2)})$ on the road as described by

$$z_k^{(1)} = h_k^{(1)}(x_k) + v_k^{(1)} \quad (2.21)$$

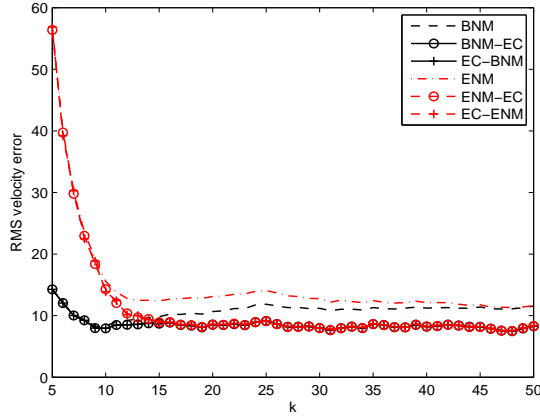


Figure 2.11: RMS velocity error comparison for example 3 with Q mismatch. Note that BNM-EC essentially overlaps with EC-BNM, and except for the small deviation at transient, ENM-EC essentially overlaps with EC-ENM.

where

$$h_k^{(1)}(x_k) = \begin{bmatrix} \sqrt{(x_k - x_r^{(1)})^2 + (y_k - y_r^{(1)})^2} \\ \sqrt{(x_k - x_r^{(2)})^2 + (y_k - y_r^{(2)})^2} \end{bmatrix}, \quad v_k^{(1)} \sim \mathcal{N}(0, R_k^{(1)})$$

$$R_k^{(1)} = \text{diag}((30\text{m})^2, (30\text{m})^2), \quad (x_r^{(1)}, y_r^{(1)}) = (0\text{m}, 0\text{m}), \quad (x_r^{(2)}, y_r^{(2)}) = (10^5\sqrt{3}\text{m}, 10^5\text{m})$$

It is also known that

$$\bar{x}_0 = [0\text{m} \quad 0\text{m} \quad 10\sqrt{3}\text{m/s} \quad 10\text{m/s}]', \quad P_0^u = \text{diag}(400\text{m}^2, 400\text{m}^2, 10(\text{m/s})^2, 10(\text{m/s})^2)$$

Figs. 2.14 and 2.15 show comparison results of the ENM-EC and EC-ENM filters, which use correct Q but mis-specified $\hat{x}_{0|0}$. The explanation for the names ENM-EC and EC-ENM are the same as in Table 2.3 except that $z_k^{(1)}$ is now defined by Eq. (2.21).

It can be clearly seen from the simulation results that the processing order of $z_k^{(1)}$ and the linear equality constraint matters a lot when there exists mis-specification on $\hat{x}_{0|0}$. As expected, better performance is achieved when the linear equality constraint is processed

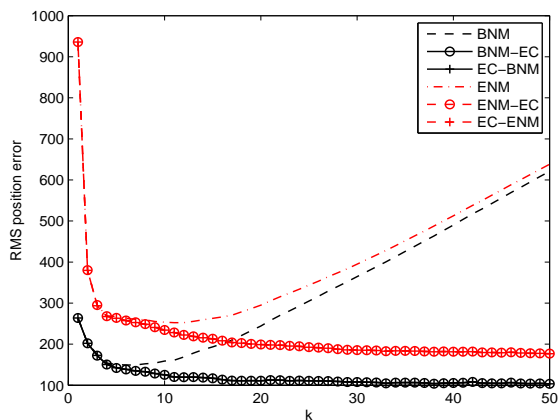


Figure 2.12: RMS position error comparison for example 3 with mis-specified $\hat{x}_{0|0}$. Note that BNM-EC essentially overlaps with EC-BNM, and ENM-EC essentially overlaps with EC-ENM.

before $z_k^{(1)}$.

2.8 Summary

This chapter targets the state estimation problem with both noisy and noise-free measurements. This more general framework has numerous real supports, e.g, state estimation under linear or nonlinear equality constraints, with correlated or singular measurement noise. Although state estimation with both noisy and noise-free measurements is not a big deal in theory, computationally efficient ways should be preferred. So two sequential forms of the LMMSE estimator which are equivalent to the batch LMMSE estimator have been proposed. How to extend the results to the nonlinear measurement case and how to choose between the two sequential forms have also been discussed. Numerical examples show that although both equality constraints and perfect measurements can be treated as noise-free measurements, they should be handled differently and with caution. It is also shown through numerical examples that the update by equality constraints is necessary and helpful when there exists model mismatch and the processing order of noisy measurements and equality constraints

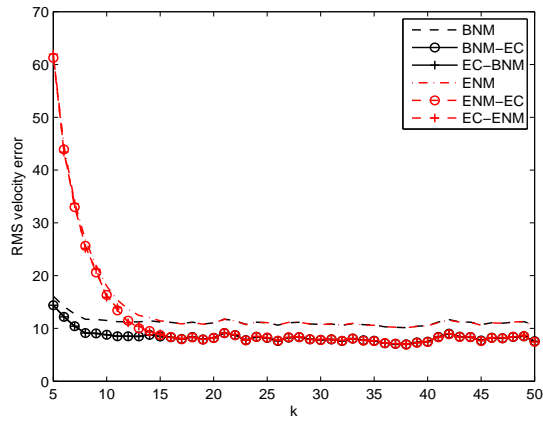


Figure 2.13: RMS velocity error comparison for example 3 with mis-specified $\hat{x}_{0|0}$. Note that BNM-EC essentially overlaps with EC-BNM, and except very little deviation at transient, ENM-EC essentially overlaps with EC-ENM.

matters for some filtering problems.

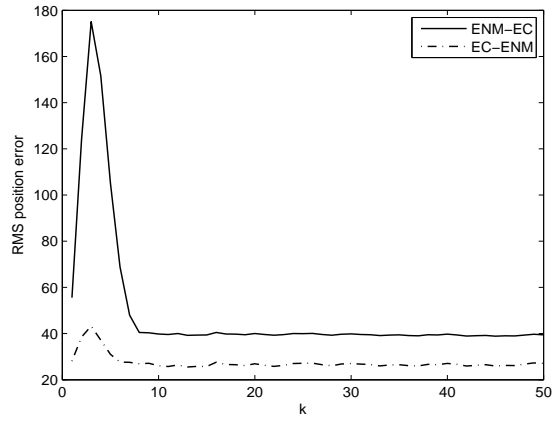


Figure 2.14: RMS position error comparison for example 4

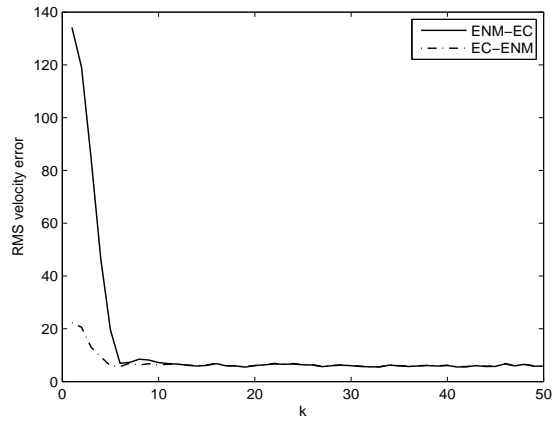


Figure 2.15: RMS velocity error comparison for example 4

Chapter 3

State Estimation with Point and Set Measurements

3.1 Introduction and related research

In this chapter, a *point measurement* is one that is a single point in the measurement space, while a *set measurement* is a subset of the measurement space.

As shown later, set measurements are available in numerous cases, for example, state estimation under linear, nonlinear inequality constraints, with quantized measurements. These cases of state estimation are much harder than those with point measurements.

Results about inequality constrained estimation are available [13,103,106,107]. The main idea is to extend the existing results in equality constrained estimation [105] to inequality constrained problem by utilizing classical optimization techniques (e.g., the active set method [103,106,107] and the interior point method [13]). Unfortunately, there are some problems associated with these methods which will be analyzed in detail later.

State estimation with quantized measurement is an old research topic [28,29], and it has attracted recent interests [33,58,63] with new developments in hybrid estimation, nonlinear filtering and numerical computation. For instance, given the quantization rules, [58] applied the first-order generalized pseudo-Bayesian (GPB1) idea in maneuvering target tracking

to improve the performance of Sheppard’s method; [63] applied particle filtering; and a numerically efficient implementation of the state estimator with quantized measurement in [28, 29] was proposed in [33]. Most of the work assumed that the quantization regions were hyper-rectangles and can not be easily extended to some other shaped regions, e.g., ellipsoidal regions. Also, the application of particle filtering to this type of problem is too costly and not affordable in practice as explained in detail later.

This work is an extension of our recent work in [33, 58] to a more general framework—state estimation with point and set measurements. In this chapter, inspired by the estimation with quantized measurements developed by Curry [28], under a Gaussian assumption, the MMSE state estimator with point measurements and set measurements of any shape is proposed by discretizing continuous set measurements. Possible ways to relax the Gaussian assumption and to discretize the involved Gaussian and truncated Gaussian distributions are also discussed. Supporting numerical examples are provided.

This chapter is organized as follows. Sec. 3.2 formulates the problem. Sec. 3.3 gives some cases with set measurements so as to show our research work in this direction is useful. Sec. 3.4 explains how huge the computational load by applying particle filtering to this type of problem can be. Sec. 3.5 presents the MMSE filter with point and set measurements. Sec. 3.6 discusses one possible way to discretize the involved Gaussian and truncated Gaussian distributions. Sec. 3.7 provides supporting numerical examples. Sec. 3.8 gives summary.

3.2 Problem formulation

Consider the following generic linear dynamic system

$$x_k = F_{k-1}x_{k-1} + G_{k-1}w_{k-1}$$

with zero-mean white Gaussian noise w_k with $\text{cov}(w_k) = Q_k \geq 0$ and $x_k \in \mathbb{R}^n$, $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$.

Assuming that two types of measurements of the state are available. The first is the conventional point measurement

$$z_k^{(1)} = H_k^{(1)} x_k + v_k^{(1)} \quad (3.1)$$

with zero-mean white Gaussian noise $v_k^{(1)}$ with $\text{cov}(v_k^{(1)}) = R_k^{(1)} > 0$ and $z_k^{(1)} \in \mathbb{R}^{m_1}$. $\langle w_k \rangle$, $\langle v_k^{(1)} \rangle$ and x_0 are independent of each other.

Remark: For simplicity, $R_k^{(1)} > 0$ is assumed although we can still handle it if $R_k^{(1)} \geq 0$.

The second type is the set measurement:

$$Y_k = \{z_k^{(2)} \in \mathcal{Z}_k\}, \mathcal{Z}_k \subseteq \mathbb{R}^{m_2}$$

where $z_k^{(2)}$ is a linear or nonlinear function of x_k which may or may not be driven by random noise, \mathcal{Z}_k is a subset of the measurement space \mathbb{R}^{m_2} .

Remark: In this work, the only requirement on \mathcal{Z}_k is that $\mathcal{Z}_k \subseteq \mathbb{R}^{m_2}$. Whether \mathcal{Z}_k is bounded or not, convex or concave, and how its geometric shape looks like do not really matter.

Remark: In the extreme case, \mathcal{Z}_k can be the empty set \emptyset , singleton or even $\Omega = \mathbb{R}^{m_2}$ (the measurement space). Both $\mathcal{Z}_k = \emptyset$ and $\mathcal{Z}_k = \Omega$ are trivial, since they do not provide any useful information normally and $z_k^{(2)}$ can be ignored. If \mathcal{Z}_k is a singleton, $z_k^{(2)} = \mathcal{Z}_k$ is a point measurement and is the same as $z_k^{(1)}$.

Remark: \mathcal{Z}_k can be either a *continuous* or *discrete* set. Only the continuous set case will be discussed in this chapter. For the discrete set case, we need to either reformulate the problem with a given probability mass for each set element, or obtain the probability mass online for each set element based on a filtering procedure.

In this work, we estimate the state as best as we can in the MMSE sense based on point measurement $z_k^{(1)}$ and set measurement Y_k , that is,

$$\hat{x}_{k|k}^{\text{MMSE}} = \arg \min_{\hat{x}_{k|k}} E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | \{z_t^{(1)}, Y_t\}_{t=1}^k]$$

where $\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}$.

3.3 Set measurements

Before getting into the details about how to obtain the best state estimation, let us discuss some cases with point and set measurements to show that our work is not only meaningful in theory but also useful for application.

3.3.1 Linear inequality constraint

In this case, a linear inequality constraint is placed on the estimand (quantity to be estimated):

$$Dx_k \leq d_k \text{ or } a_k \leq Dx_k \leq b_k \tag{3.2}$$

where matrix D and vectors d_k, a_k and b_k are all known. And the vector inequalities are interpreted componentwise.

Let

$$z_k^{(2)} = Dx_k$$

It can be easily seen that the inequality constraint defines a set measurement. So state estimation with linear inequality constraint is indeed a special case of the one with point and set measurements.

The one-sided linear inequality constraint of (3.2) has been widely studied [103, 106, 107]. The two-sided linear constraint case was also considered in [14, 94]. For example, [103, 106,

107] tried to extend the existing results of [105] for state estimation with linear equality constraints to the linear inequality constraint case by utilizing the active set method in optimization. There are problems with this extension. First, a search based numerical optimization technique is mandatory, which may be time consuming. Second, the objective functions are inconsistent before and after using the constraint. Specifically, the objective function before using the constraints is MSE, while afterwards it becomes fitting error in the least-squares sense, in which the estimate was treated as data. Thus, in what sense is the final estimate optimal? The two-sided linear inequality constraints were handled in [94] in an *ad hoc* way, while [14] proposed to treat them as the a priori information about the estimand for the special case of $D = I$, where the determination of the prior distribution of the estimand is the key.

3.3.2 Nonlinear inequality constraint

In this case, a nonlinear inequality constraint is placed on the estimand:

$$c(x_k) \leq 0 \text{ or } a_k \leq c(x_k) \leq b_k$$

where vectors a_k and b_k are both known and $c(\cdot)$ is some vector-valued nonlinear function.

Let

$$z_k^{(2)} = c(x_k)$$

We can easily see that state estimation with nonlinear inequality constraint is indeed a special case of the problem with point and set measurements.

Some relevant results were reported in [13,94]. Those of [13] are not free of the problems discussed above either and the nonlinear inequality constraints were handled in an *ad hoc* way in [94].

Remark: It should be noted that when the inequality constraints are treated as set mea-

surements, there will be no measurement noise (i.e., noise-free) in the measurement equation $z_k^{(2)} = Dx_k$ or $z_k^{(2)} = c(x_k)$. This may complicate the corresponding state estimation, to be explained later.

3.3.3 Quantized measurement

In this case, only a quantized measurement is available:

$$\begin{aligned} z_k^{(2)} &= H_k^{(2)} x_k + v_k^{(2)} \\ y_k &= Q(z_k^{(2)}) \end{aligned}$$

where $y_k \in \mathbb{R}^{n_y}$ is the quantized measurement with a quantization operator $Q(\cdot)$ and $z_k^{(2)}$ is the measurement before quantization.

The quantizer $Q(\cdot)$ is a nonlinear mapping and the quantization can be uniform (e.g., midrise or midtread [63,120]) or nonuniform (e.g., μ -law companding or A -law companding [96,97]). In general, with the help of a quantization rule $y_k = Q(z_k^{(2)})$, the only thing we can infer from a quantized measurement y_k about the measurement before quantization $z_k^{(2)}$ is that $z_k^{(2)} \in \mathcal{Z}_k$ for some $\mathcal{Z}_k \subset \mathbb{R}^{m_2}$ where the set of all quantization regions $\{\mathcal{Z}_k\}$ is a partition of \mathbb{R}^{m_2} . One simple example is the rectangular regions \mathcal{Z}_k defined by

$$\mathcal{Z}_k = \{z_k^{(2)} \in \mathbb{R}^{m_2} : a_k \leq z_k^{(2)} < b_k\}$$

where the vector inequalities are interpreted componentwise, i.e., $a_{k,i} \leq z_{k,i}^{(2)} < b_{k,i}$ for $\forall i = 1, \dots, m_2$.

Clearly, $z_k^{(2)} \in \mathcal{Z}_k$ inferred from the quantized measurement y_k is exactly a set measurement. That is, state estimation with quantized measurement is a special case of the problem with set measurement only.

3.4 Particle filter

As a close approximation to the recursive Bayesian filtering, particle filter significantly improves estimation accuracy at the cost of a heavily increased computational burden. It has many variants. In this work, the bootstrap particle filtering (BPF) is chosen as an example to analyze its feasibility for state estimation with point and set measurements.

Let $z_k = \{z_k^{(1)}, Y_k\}$. Then the BPF [76] for the given problem in one cycle can be summarized as follows:

- Initialization: $x_0^{(i)} \sim f_{x_0}(x)$, $i = 1, 2, \dots, N$.

- Proposal distribution:

$$\pi(x_k | x_{k-1}^{(i)}, z_k) = p(x_k | x_{k-1}^{(i)})$$

- Sampling:

$$x_k^{(i*)} = F_{k-1}x_{k-1}^{(i)} + G_{k-1}w_{k-1}^{(i)}$$

where $w_{k-1}^{(i)} \sim p_{w_{k-1}}(w)$.

- Weighting:

$$\begin{aligned} \tilde{\alpha}_k^{(i)} &= p(z_k | x_k^{(i*)}) = \mathcal{N}(z_k^{(1)}; H_k^{(1)}x_k^{(i*)}, R_k^{(1)})P\{z_k^{(2)} \in \mathcal{Z}_k | x_k = x_k^{(i*)}\} \\ \alpha_k^{(i)} &= \tilde{\alpha}_k^{(i)} / \sum_{j=1}^N \tilde{\alpha}_k^{(j)} \end{aligned}$$

- Resampling:

$$\{x_k^{(i)}, 1/N\}_{i=1}^N = \text{Resample}(\{x_k^{(i*)}, \alpha_k^{(i)}\}_{i=1}^N)$$

- Measurement update:

$$\hat{x}_{k|k} = \frac{1}{N} \sum_{j=1}^N x_k^{(j)}$$

$$P_{k|k} = \frac{1}{N} \sum_{j=1}^N (x_k^{(j)} - \hat{x}_{k|k})(x_k^{(j)} - \hat{x}_{k|k})'$$

Remark: In the above, $p(\cdot)$ denotes pdf and/or pmf, while $P\{\cdot\}$ denotes probability only. Also, N is the number of particles.

As can be seen from the above, by introducing the set measurement, one main difficulty is that at each time, for every particle, we need to calculate the following likelihood

$$P\{z_k^{(2)} \in \mathcal{Z}_k | x_k = x_k^{(i*)}\} = \int_{\mathcal{Z}_k} \mathcal{N}(z_k^{(2)}; H_k^{(2)} x_k^{(i*)}, R_k^{(2)}) dz_k^{(2)}$$

which is a multivariate integral. It is well known that for multivariate integration, if it can not be obtained analytically, numerical quadratures usually are time consuming unless some efficient ways can be utilized. For numerical quadrature at each time for every member of the thousands of particles, the computational burden is huge and can not be affordable for practical applications. So in the sequel, particle filtering is only used as a reference for estimation performance comparison.

3.5 MMSE filtering with point and set measurements

Due to the introduction of the set measurement, state estimation with point and set measurements is essentially a nonlinear filtering problem. Although nonlinearity destroys Gaussianity in general, to simplify derivation, we will still make Gaussian approximation to the updated state estimate at each time step. A similar idea can be found in the Gaussian filter developed for general nonlinear filtering problems in [57].

3.5.1 General form of MMSE filter

As was discussed above, state estimation with quantized measurement is a special type of the one with point and set measurements. Inspired by the ideas dealing with quantized measurements in [28, 29], we develop an approximate MMSE filtering algorithm with point and set measurements.

Define

$$z^k = \{z_1, z_2, \dots, z_k\}$$

It is well known that the MMSE filter is

$$\hat{x}_{k|k} = E[x_k | z^k] \quad (3.3)$$

Using the total expectation theorem, we have

$$\begin{aligned} \hat{x}_{k|k} &= E[x_k | z^k] = E[E[x_k | z_k^{(2)}, z^k] | z^k] = \int_{\mathcal{Z}_k} E[x_k | z_k^{(2)}, z^k] p(z_k^{(2)} | z^k) dz_k^{(2)} \\ &= \int_{\mathcal{Z}_k} E[x_k | z_k^{(2)}, z^{k-1}, z_k^{(1)}] p(z_k^{(2)} | z^k) dz_k^{(2)} \end{aligned} \quad (3.4)$$

with

$$\begin{aligned} P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z^k] = E[E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z_k^{(2)}, z^k] | z^k] \\ &= \int_{\mathcal{Z}_k} E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z_k^{(2)}, z^k] p(z_k^{(2)} | z^k) dz_k^{(2)} \\ &= \int_{\mathcal{Z}_k} E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z_k^{(2)}, z^{k-1}, z_k^{(1)}] p(z_k^{(2)} | z^k) dz_k^{(2)} \end{aligned} \quad (3.5)$$

That is, $\hat{x}_{k|k}$ and $P_{k|k}$ are the means of $E[x_k | z_k^{(2)}, z^{k-1}, z_k^{(1)}]$ and $E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z_k^{(2)}, z^{k-1}, z_k^{(1)}]$ (functions of $z_k^{(2)}$ given z^k) w.r.t. the distribution $p(z_k^{(2)} | z^k)$.

From the definition of z^k , it follows that

$$p(z_k^{(2)}|z^k) = p(z_k^{(2)}|z^{k-1}, z_k^{(1)}, Y_k) = \frac{1}{c_z} p(z_k^{(2)}|z^{k-1}, z_k^{(1)}) I_{\mathcal{Z}_k}(z_k^{(2)})$$

where

$$I_{\mathcal{Z}_k}(z_k^{(2)}) = \begin{cases} 1 & \text{if } z_k^{(2)} \in \mathcal{Z}_k \\ 0 & \text{otherwise} \end{cases}, \quad c_z = \int_{\mathcal{Z}_k} p(z_k^{(2)}|z^{k-1}, z_k^{(1)}) dz_k^{(2)}$$

That is, $p(z_k^{(2)}|z^k)$ is nothing but a truncated version of $p(z_k^{(2)}|z^{k-1}, z_k^{(1)})$.

In general, it is hard to solve (3.4) and (3.5) analytically except in some special cases due to several difficulties. First, it is hard to obtain the truncated distribution $p(z_k^{(2)}|z^k)$ exactly. Second, it is also hard to find an analytical form for the nonlinear filtering problem $E[x_k|z_k^{(2)}, z^{k-1}, z_k^{(1)}]$ and $E[\tilde{x}_{k|k}\tilde{x}'_{k|k}|z_k^{(2)}, z^{k-1}, z_k^{(1)}]$ if $z_k^{(2)}$ is a nonlinear function of x_k . Third, even if we have an analytical form of $p(z_k^{(2)}|z^k)$, $E[x_k|z_k^{(2)}, z^{k-1}, z_k^{(1)}]$ and $E[\tilde{x}_{k|k}\tilde{x}'_{k|k}|z_k^{(2)}, z^{k-1}, z_k^{(1)}]$, to evaluate the involved integrals in a closed form is not easy and only numerical quadrature can be done in practice.

Given

$$\hat{x}_{k-1|k-1} = E[x_{k-1}|z^{k-1}], \quad P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1}|z^{k-1})$$

assume that

$$p(x_{k-1}|z^{k-1}) = \mathcal{N}(x_{k-1}; \hat{x}_{k-1|k-1}, P_{k-1|k-1}) \tag{3.6}$$

Then the state at time k can be estimated as follows

$$\hat{x}_{k|k-1} = E[x_k | z^{k-1}] = F_{k-1} \hat{x}_{k-1|k-1} \quad (3.7)$$

$$P_{k|k-1} = \text{MSE}(\hat{x}_{k|k-1} | z^{k-1}) = F_{k-1} P_{k-1|k-1} F_{k-1}' + G_{k-1} Q_{k-1} G_{k-1}' \quad (3.8)$$

$$\hat{x}_{k|k}^{(1)} = E[x_k | z^{k-1}, z_k^{(1)}] = \hat{x}_{k|k-1} + K_k^{(1)} (z_k^{(1)} - H_k^{(1)} \hat{x}_{k|k-1}) \quad (3.9)$$

$$P_{k|k}^{(1)} = \text{MSE}(\hat{x}_{k|k}^{(1)} | z^{k-1}, z_k^{(1)}) = P_{k|k-1} - K_k^{(1)} S_k^{(1)} (K_k^{(1)})' \quad (3.10)$$

$$K_k^{(1)} = P_{k|k-1} (H_k^{(1)})' (S_k^{(1)})^{-1}$$

$$S_k^{(1)} = H_k^{(1)} P_{k|k-1} (H_k^{(1)})' + R_k^{(1)}$$

This is nothing but an application of the Kalman filter.

Depending on whether $z_k^{(2)}$ is a linear or nonlinear function of x_k , we discuss next how to solve (3.4) and (3.5) approximately in an efficient way.

Linear case

When $z_k^{(2)}$ takes the linear form

$$z_k^{(2)} = H_k^{(2)} x_k + v_k^{(2)} \quad (3.11)$$

where $v_k^{(2)}$ is zero-mean white Gaussian noise with $\text{cov}(v_k^{(2)}) = R_k^{(2)} \geq 0$ and uncorrelated with $\langle w_k \rangle$, $\langle v_k^{(1)} \rangle$ and x_0 , it follows from the Kalman filter that

$$\hat{x}_{k|k}^* = E[x_k | z^{k-1}, z_k^{(1)}, z_k^{(2)}] = \hat{x}_{k|k}^{(1)} + K_k (z_k^{(2)} - \hat{z}_{k|k}^{(2)})$$

$$P_{k|k}^* = \text{MSE}(\hat{x}_{k|k}^* | z^{k-1}, z_k^{(1)}, z_k^{(2)}) = P_{k|k}^{(1)} - K_k S_k^{(2)} K_k'$$

where

$$\begin{aligned}\hat{z}_{k|k}^{(2)} &= E[z_k^{(2)} | z^{k-1}, z_k^{(1)}] = H_k^{(2)} \hat{x}_{k|k}^{(1)} \\ K_k &= P_{k|k}^{(1)} (H_k^{(2)})' (S_k^{(2)})^{-1} \\ S_k^{(2)} &= \text{MSE}(\hat{z}_{k|k}^{(2)} | z^{k-1}, z_k^{(1)}) = H_k^{(2)} P_{k|k}^{(1)} (H_k^{(2)})' + R_k^{(2)}\end{aligned}$$

Note that $\hat{x}_{k|k}^{(1)}$, K_k and $H_k^{(2)}$ have nothing to do with $z_k^{(2)}$. Substituting $\hat{x}_{k|k}^*$ into Eq. (3.4) yields

$$\hat{x}_{k|k} = \hat{x}_{k|k}^{(1)} + K_k (E[z_k^{(2)} | z^k] - \hat{z}_{k|k}^{(2)}) \quad (3.12)$$

Since

$$\tilde{x}_{k|k} = x_k - \hat{x}_{k|k} = x_k - \hat{x}_{k|k}^* + \hat{x}_{k|k}^* - \hat{x}_{k|k} = x_k - \hat{x}_{k|k}^* + K_k (z_k^{(2)} - E[z_k^{(2)} | z^k])$$

we have

$$E[\tilde{x}_{k|k} \tilde{x}_{k|k}' | z_k^{(2)}, z^{k-1}, z_k^{(1)}] = P_{k|k}^* + K_k (z_k^{(2)} - E[z_k^{(2)} | z^k]) (z_k^{(2)} - E[z_k^{(2)} | z^k])' K_k'$$

Note that $P_{k|k}^*$ and K_k have nothing to do with $z_k^{(2)}$. Substituting this into Eq. (3.5) yields

$$P_{k|k} = P_{k|k}^* + K_k \text{cov}(z_k^{(2)} | z^k) K_k' \quad (3.13)$$

These are exactly the formulas for state estimation with quantized measurements in [28, 29, 33], except that now we have both point measurement $z_k^{(1)}$ and set measurement Y_k .

Remark: The reason why $z_k^{(1)}$ is processed first and $z_k^{(2)}$ is processed later is that the update by $z_k^{(1)}$ is optimal and it is better to put the approximation step as close to the output of final result as possible.

It follows from the Gaussian assumption (3.6) that $p(z_k^{(2)} | z^k)$ is a truncated Gaussian

distribution

$$p(z_k^{(2)}|z^k) = \frac{1}{c_z} \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)}) I_{\mathcal{Z}_k}(z_k^{(2)})$$

where

$$c_z = \int_{\mathcal{Z}_k} \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)}) dz_k^{(2)}$$

Clearly, the key to this MMSE filter is to compute the mean and covariance matrix of the truncated Gaussian distribution $p(z_k^{(2)}|z^k)$. For the scalar case, i.e., $m_2 = 1$, they can be obtained analytically [33] in terms of the error function $\Phi(\cdot)$ and the density function of the standard Gaussian distribution. For the multi-dimensional case, i.e., $m_2 > 1$, their computation usually relies on multi-dimensional integration over the truncation region without a closed form and only numerical quadrature can be done in practice.

There are several difficulties associated with the numerical quadratures for $E[z_k^{(2)}|z^k]$ and $\text{cov}(z_k^{(2)}|z^k)$. For example, the Gaussian distribution $\mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)})$ may be singular¹ in some cases, e.g., due to $|R_k^{(2)}| = 0$. As we know, the density function for a singular Gaussian distribution is not defined, let alone the quadratures based on it. Even if $\mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)})$ is nonsingular, the following two aspects may still complicate the quadratures in general. First, due to the high dimensionality, the curse of dimensionality may occur. Second, the numerical quadrature may be over a truncation region not in a regular shape. Recall that in the problem formulation, the truncation region is only required to be a subset of the m_2 -dimensional measurement space, so it can be of any shape depending on the problem.

Now let us use two examples to illustrate how irregular the shapes of the truncation region can be for the case $m_2 = 3$. For example, the domain of integration can be unbounded in a three-dimensional space like

$$-\infty < z_{k,1}^{(2)} \leq b_1, \quad a_2 \leq z_{k,2}^{(2)} < b_2, \quad a_3 < z_{k,3}^{(2)} < +\infty$$

¹The Gaussian distribution $\mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)})$ with $|S_k^{(2)}| = 0$ is called a singular Gaussian.

Another example is that the domain of integration may be an ellipsoid [126] like

$$\sum_{i=1}^3 ((z_{k,i}^{(2)} - a_i)/b_i)^2 = 1$$

where $z_{k,i}^{(2)}$ is the i -th component of the vector $z_k^{(2)}$, a_i and b_i are constants.

If the quadratures used to evaluate $E[z_k^{(2)}|z^k]$ and $\text{cov}(z_k^{(2)}|z^k)$ are well defined, we may consider to use a mature numerical quadrature rule like the Newton–Cotes rule, which includes the rectangle rule, trapezoidal rule and Simpson’s rule as special cases, and the Clenshaw–Curtis rule, but they are mainly for integrations over a one-dimensional bounded region and can not be easily extended to other cases. Efficient numerical quadrature rules like the Gaussian quadrature can indeed handle integration over unbounded domain, but they are still mainly for the one-dimensional case. There do exist exceptions, for instance, the Gaussian-Hermite quadrature can be applied in the high dimensional case because the domain of integration is from $-\infty$ to $+\infty$, which is kept the same after changing variables through decoupling to the original ones. The other Gaussian quadratures are not so lucky because after decoupling of the original variables, the domain of each new variable may be strongly coupled, which will complicate the quadrature greatly and the existing Gaussian quadrature rules can not be applied. To handle multi-dimensional integration, Monte Carlo and sparse grid methods are usually suggested. The sparse grid method is based on a one-dimensional quadrature rule, but performs a more sophisticated combination of univariate results. It is mainly for integration with a hyper-rectangular domain. Monte Carlo integration may yield better accuracy for the same number of function evaluations than repeated integrations using one-dimensional methods, but the computational complexity is usually too high to be affordable online.

Next, a new way to obtain $E[z_k^{(2)}|z^k]$ and $\text{cov}(z_k^{(2)}|z^k)$ without resorting to any conventional quadrature rules is discussed, which is more general in the sense that it can be applied

to set measurements of any shape, of any dimension, regardless if the involved truncated Gaussian distribution is singular or not.

Suppose that there exists a discrete random vector $z_{k,td}^{(2)}$ which is an approximation to the conditional distribution $p(z_k^{(2)} | z^{k-1}, z_k^{(1)}, Y_k)$:

$$P\{z_{k,td}^{(2)} = z_k^{(2,i)}\} = \mu_k^{(i)}, \quad i = 1, 2, \dots, M$$

where

$$\mu_k^{(i)} \geq 0, \quad \sum_{i=1}^M \mu_k^{(i)} = 1$$

Then it follows that

$$\hat{z}_{k|k} = E[z_k^{(2)} | z^{k-1}, z_k^{(1)}, Y_k] \approx E[z_{k,td}^{(2)}] = \sum_{i=1}^M \mu_k^{(i)} z_k^{(2,i)}$$

$$\begin{aligned} \text{cov}(z_k^{(2)} | z^{k-1}, z_k^{(1)}, Y_k) &\approx \text{cov}(z_{k,td}^{(2)}) = \sum_{i=1}^M \mu_k^{(i)} (z_k^{(2,i)} - \hat{z}_{k|k})(z_k^{(2,i)} - \hat{z}_{k|k})' \\ &= \sum_{i=1}^M \mu_k^{(i)} z_k^{(2,i)} (z_k^{(2,i)})' - \hat{z}_{k|k} \hat{z}_{k|k}' \end{aligned}$$

How to discretize a truncated Gaussian distribution will be discussed in detail later.

Nonlinear case

If $z_k^{(2)}$ takes a nonlinear form

$$z_k^{(2)} = h_k(x_k, v_k^{(2)})$$

even under the Gaussian assumption (3.6), we do not have $p(z_k^{(2)} | z^{k-1}, z_k^{(1)}) = \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k}^{(2)}, S_k^{(2)})$ and it is usually hard to obtain the exact distribution $p(z_k^{(2)} | z^{k-1}, z_k^{(1)})$, let alone $p(z_k^{(2)} | z^k)$, which is a truncated version of $p(z_k^{(2)} | z^{k-1}, z_k^{(1)})$. Thus the MMSE filter with a nonlinear $z_k^{(2)}$ is much harder to obtain than the one with a linear $z_k^{(2)}$.

Under the Gaussian assumption (3.6), we have $p(x_k|z^{k-1}, z_k^{(1)}) = \mathcal{N}(x_k; \hat{x}_{k|k}^{(1)}, P_{k|k}^{(1)})$. Suppose that there exists close discrete approximations to $p(x_k|z^{k-1}, z_k^{(1)})$ and $p(v_k^{(2)}) = \mathcal{N}(v_k^{(2)}; 0, R_k^{(2)})$

$$\begin{aligned} P\{x_{k,d} = x_k^{(i)}\} &= \alpha_k^{(i)}, \quad i = 1, 2, \dots, M_x \\ P\{v_{k,d}^{(2)} = v_k^{(2,j)}\} &= \beta_k^{(j)}, \quad j = 1, 2, \dots, M_v \end{aligned}$$

where

$$\alpha_k^{(i)} \geq 0, \quad \sum_{i=1}^{M_x} \alpha_k^{(i)} = 1, \quad \beta_k^{(j)} \geq 0, \quad \sum_{j=1}^{M_v} \beta_k^{(j)} = 1$$

Then a close discrete approximation to $p(z_k^{(2)}|z^{k-1}, z_k^{(1)})$ is

$$P\{z_{k,d}^{(2)} = h_k(x_k^{(i)}, v_k^{(2,j)})\} = \gamma_k^{(i,j)} = \alpha_k^{(i)} \beta_k^{(j)}$$

where $i = 1, 2, \dots, M_x, j = 1, 2, \dots, M_v$.

Selecting only those $h_k(x_k^{(i)}, v_k^{(2,j)}) \in \mathcal{Z}_k$ and normalizing their corresponding probability masses yield a close discrete approximation to the truncated distribution $p(z_k^{(2)}|z^k)$ which has the pmf

$$P\{z_{k,td}^{(2)} = z_k^{(2,l)}\} = \zeta_k^{(l)}, \quad l = 1, 2, \dots, L$$

with

$$\zeta_k^{(l)} \geq 0, \quad \sum_{l=1}^L \zeta_k^{(l)} = 1$$

Then we can approximate $\hat{x}_{k|k}$ and $P_{k|k}$ by

$$\hat{x}_{k|k} \approx \sum_{l=1}^L \hat{x}_{k|k}^{(l*)} \zeta_k^{(l)} \tag{3.14}$$

$$\begin{aligned}
P_{k|k} &\approx \sum_{l=1}^L E[\tilde{x}_{k|k} \tilde{x}'_{k|k} | z_{k,td}^{(2)} = z_k^{(2,l)}, z^{k-1}, z_k^{(1)}] \zeta_k^{(l)} \\
&= \sum_{l=1}^L \zeta_k^{(l)} E[\tilde{x}_{k|k}^{(l*)} (\tilde{x}_{k|k}^{(l*)})' | z_{k,td}^{(2)} = z_k^{(2,l)}, z^{k-1}, z_k^{(1)}] + \sum_{l=1}^L \zeta_k^{(l)} (\hat{x}_{k|k}^{(l*)} - \hat{x}_{k|k}) (\hat{x}_{k|k}^{(l*)} - \hat{x}_{k|k})' \\
&= \sum_{l=1}^L [P_{k|k}^{(l*)} + (\hat{x}_{k|k}^{(l*)} - \hat{x}_{k|k}) (\hat{x}_{k|k}^{(l*)} - \hat{x}_{k|k})'] \zeta_k^{(l)} \\
&= \sum_{l=1}^L \zeta_k^{(l)} P_{k|k}^{(l*)} + \sum_{l=1}^L \zeta_k^{(l)} \hat{x}_{k|k}^{(l*)} (\hat{x}_{k|k}^{(l*)})' - \hat{x}_{k|k} \hat{x}'_{k|k}
\end{aligned} \tag{3.15}$$

where

$$\begin{aligned}
\tilde{x}_{k|k} &= x_k - \hat{x}_{k|k}, \quad \tilde{x}_{k|k}^{(l*)} = x_k - \hat{x}_{k|k}^{(l*)} \\
\hat{x}_{k|k}^{(l*)} &= E[x_k | z_{k,td}^{(2)} = z_k^{(2,l)}, z^{k-1}, z_k^{(1)}] \\
P_{k|k}^{(l*)} &= \text{MSE}(\hat{x}_{k|k}^{(l*)})
\end{aligned}$$

Note that $\hat{x}_{k|k}^{(l*)} = E[x_k | z_k^{d,(2)} = z_k^{(2,l)}, z^{k-1}, z_k^{(1)}]$ is a standard nonlinear filtering problem. Existing nonlinear filtering methods for point estimation can be applied here. As an illustration, only the general form of state estimator that is optimal in the sense of LMMSE is listed here:

$$\begin{aligned}
\hat{x}_{k|k}^{(l*)} &= E[x_k | z_{k,td}^{(2)} = z_k^{(2,l)}, z^{k-1}, z_k^{(1)}] = \hat{x}_{k|k}^{(1)} + K_k^{(2,l)} \tilde{z}_{k|k}^{(2,l)} \\
P_{k|k}^{(l*)} &= P_{k|k}^{(1)} - K_k^{(2,l)} S_k^{(2,l)} (K_k^{(2,l)})'
\end{aligned}$$

where

$$\begin{aligned}
K_k^{(2,l)} &= \text{cov}(\tilde{x}_{k|k}^{(1)}, \tilde{z}_{k|k}^{(2,l)}) (S_k^{(2,l)})^+, \quad S_k^{(2,l)} = \text{cov}(\tilde{z}_{k|k}^{(2,l)}) \\
\tilde{z}_{k|k}^{(2,l)} &= z_k^{(2,l)} - E^*[z_k^{(2)} | z^{k-1}, z_k^{(1)}]
\end{aligned}$$

Remark: Due to the nonlinearity of $z_k^{(2)}$, we do not have an elegant analytical form for $E^*[z_k^{(2)} | z^{k-1}, z_k^{(1)}]$, $\text{cov}(\tilde{x}_{k|k}^{(1)}, \tilde{z}_{k|k}^{(2,l)})$ and $S_k^{(2,l)}$ in general, but they can be approximated by the

extended Kalman filter, unscented filtering [60], DD2 [89], Gaussian Hermite filter [57] or even from the original definition of LMMSE estimation, as in [130].

Remark: If $\text{cov}(\tilde{x}_{k|k}^{(1)}, \tilde{z}_{k|k}^{(2,l)})$ and $S_k^{(2,l)}$ do not depend on $z_k^{(2,l)}$, then (3.14) and (3.15) are equivalent to (3.12) and (3.13), which is more preferable due to its efficiency in implementation.

3.5.2 Some discussions on set constrained estimation

As was done in [35] to handle equality constraints, the set constraint

$$c_k(x_k) \in \mathcal{Z}_k \tag{3.16}$$

can be treated as another piece of set measurement information. In this way, state estimation with set constraints can be solved simply by direct application of the general form of the approximate MMSE filter developed in the above subsection. Although set constraints can be treated as set measurements, they do have their own characteristics. For example, \mathcal{Z}_k is nonrandom and the validity of set constraints is known beforehand. Next, we provide another way to solve state estimation with set constraints by the use of this special characteristic.

Define

$$z_k^{(2)} = c_k(x_k)$$

From the definition of z^k , it follows that

$$p(x_k|z^k) = p(x_k|z^{k-1}, z_k^{(1)}, Y_k) = \frac{1}{c_x} p(x_k|z^{k-1}, z_k^{(1)}) I_{\mathcal{X}_k}(x_k)$$

where

$$I_{\mathcal{X}_k}(x_k) = \begin{cases} 1 & \text{if } x_k \in \mathcal{X}_k \\ 0 & \text{otherwise} \end{cases}, \quad \mathcal{X}_k = \{x_k : c_k(x_k) \in \mathcal{Z}_k\}, \quad c_x = \int_{\mathcal{X}_k} p(x_k | z^{k-1}, z_k^{(1)}) dx_k$$

That is, $p(x_k | z^k)$ is nothing but a truncated version of $p(x_k | z^{k-1}, z_k^{(1)})$. Furthermore, since

$$P_{k|k} = \text{MSE}(\hat{x}_{k|k} | z^k) = \text{cov}(x_k | z^k) \quad (3.17)$$

it can be easily seen that $\hat{x}_{k|k}$ and $P_{k|k}$ are nothing but the mean and covariance matrix of the truncated distribution $p(x_k | z^k)$.

Remark: This was also pointed out in the so called *pdf truncation approach* of [104] for linear inequality constrained state estimation, where the mean and covariance matrix of the truncated Gaussian distribution is obtained based on Jordan canonical decomposition and Gram-Schmidt orthogonalization procedure. But unfortunately, after consideration of the first constraint, all subsequent computation relies on an incorrect assumption that the state satisfies untruncated Gaussian distribution, which is actually truncated, after the application of all past constraints.

Remark: When \mathcal{Z}_k is random, the above analysis still holds. This is the case when $z_k^{(2)} = c_k(x_k)$ is noise free.

In general, it is very hard to obtain (3.3) and (3.17) analytically except in some special cases. But as is analyzed next, for constrained estimation problem, if the estimation before update by the set constraint can be done optimally, then the update by set constraints can be skipped and all estimates satisfy the constraints automatically.

Suppose that we can obtain the following estimates optimally

$$\begin{aligned}\hat{x}_{k|k-1} &= E[x_k|z^{k-1}], \quad P_{k|k-1} = \text{MSE}(\hat{x}_{k|k-1}|z^{k-1}) \\ \hat{x}_{k|k}^{(1)} &= E[x_k|z^{k-1}, z_k^{(1)}], \quad P_{k|k}^{(1)} = \text{MSE}(\hat{x}_{k|k}^{(1)}|z^{k-1}, z_k^{(1)})\end{aligned}$$

Since it is known beforehand that the ground truth satisfies the set constraints (3.16), it follows that

$$P\{x_k \in \mathcal{X}_k\} = 1$$

That is, x_k outside \mathcal{X}_k will never occur. Also, the fact that the ground truth really satisfies the set constraints (3.16) is independent of the measurements: whether we observe it. No matter what kind of measurement is observed, the reality that the ground truth satisfies the set constraints (3.16) can not be changed. This implies that

$$P\{x_k \in \mathcal{X}_k|z^{k-1}\} = 1, \quad P\{x_k \in \mathcal{X}_k|z^{k-1}, z_k^{(1)}\} = 1$$

Now it can be easily seen that

$$\begin{aligned}\hat{x}_{k|k}^{(1)} &= E[x_k|z^{k-1}, z_k^{(1)}] = \int x_k p(x_k|z^{k-1}, z_k^{(1)}) dx_k = \int_{\mathcal{X}_k} x_k p(x_k|z^{k-1}, z_k^{(1)}) dx_k \\ &= \int x_k p(x_k|z^{k-1}, z_k^{(1)}, x_k \in \mathcal{X}_k) dx_k = \int x_k p(x_k|z^{k-1}, z_k^{(1)}, Y_k) dx_k = E[x_k|z^k] = \hat{x}_{k|k}\end{aligned}$$

That is, set measurement information from the set constraint is imposed implicitly. From orthogonal projection, we know that $\hat{x}_{k|k}^{(1)}$ satisfies the set constraint (3.16) for sure.

Similarly, it can be shown that for one-step ahead prediction $\hat{x}_{k|k-1}$, the set constraint (3.16) is also imposed implicitly and $\hat{x}_{k|k-1}$ satisfies it automatically.

Remark: Although we treat the set constraint as set measurement information in this work, there is significant difference between the set constraint and traditional measurement.

As stated above, the set constraint is known beforehand but the traditional measurement is never known before the sensing process takes place. Also the set constraint stays the same from run to run of the dynamic system, while the traditional measurement is random and thus changing from run to run.

Now that the update by set constraints can be skipped in the MMSE filter, so why do we still need to study the constrained estimation problem? Why do we still bother with set constraints? Note that in the above analysis, a strong assumption is that we can obtain the estimates before update by set constraints optimally. This is not always true in reality. For example, in practice, we may have model mismatch, nonlinearity from dynamic system or measurement equation, etc. In these cases, it is hard to obtain the unconstrained estimates optimally. If we still treat set constraints as pieces of set measurement information and take them into account in the practical estimation procedure, the estimation performance can be improved due to the use of more pieces of information.

Clearly, the key to set constrained state estimation is to compute the mean and covariance matrix of the truncated distribution $p(x_k|z^k)$. There are several difficulties associated with this computation. First, it is hard to obtain this truncated distribution exactly in some cases even if we have the exact form of $p(x_k|z^{k-1}, z_k^{(1)})$. Second, even though the analytical form of $p(x_k|z^k)$ is available, to evaluate the integrals for $E[x_k|z^k]$ and $\text{cov}(x_k|z^k)$ in a closed form is not easy and only numerical quadrature can be usually done in reality. Then all those difficulties discussed above concerning the numerical quadratures for $E[z_k^{(2)}|z^k]$ and $\text{cov}(z_k^{(2)}|z^k)$ may exist again. To develop a more general approach, a similar way by the use of a discretization method for $E[x_k|z^k]$ and $\text{cov}(x_k|z^k)$ is discussed next.

Under the Gaussian assumption in (3.6), $p(x_k|z^k)$ will be truncated Gaussian. Suppose that for the truncated Gaussian distribution $p(x_k|z^k)$ there exists a close discrete approxi-

mation which has the pmf

$$P\{x_{k,td} = x_k^{(j)}\} = \eta_k^{(j)}, \quad j = 1, 2, \dots, N$$

where

$$\eta_k^{(j)} \geq 0, \quad \sum_{j=1}^N \eta_k^{(j)} = 1$$

Then it follows that

$$\begin{aligned} \hat{x}_{k|k} &= E[x_k|z^k] \approx \sum_{j=1}^N \eta_k^{(j)} x_k^{(j)} \\ \text{cov}(x_k|z^k) &\approx \sum_{j=1}^N \eta_k^{(j)} (x_k^{(j)} - \hat{x}_{k|k})(x_k^{(j)} - \hat{x}_{k|k})' = \sum_{j=1}^N \eta_k^{(j)} x_k^{(j)} (x_k^{(j)})' - \hat{x}_{k|k} \hat{x}_{k|k}' \end{aligned}$$

Remark: For linear set constrained state estimation, we now have two choices: (a) to directly apply the general form of the approximate MMSE filter and (b) to compute the mean and covariance matrix of the truncated version of $p(x_k|z^{k-1}, z_k^{(1)})$ directly. As is clear from above, when applying the general form of the approximate MMSE filter to the linear set constrained case, the key is to compute the mean and covariance matrix of $p(z_k^{(2)}|z^k)$. So in practice, if m_2 is less than n , the mean and covariance of $p(z_k^{(2)}|z^k)$ will be relatively easier to compute than that of $p(x_k|z^k)$, as shown in the illustrative example below. In this case, the general form of the approximate MMSE filter should be chosen. Otherwise, we should compute the mean and covariance matrix of the truncated version of $p(x_k|z^{k-1}, z_k^{(1)})$ directly.

Remark: For nonlinear set constraints, if the nonlinear filters $\hat{x}_{k|k}^{(l*)}$, $l = 1, 2, \dots, L$ can not be obtained efficiently, it is preferable to compute the mean and covariance matrix of the truncated version of $p(x_k|z^{k-1}, z_k^{(1)})$ directly.

Remark: If the framework of state estimation with both point and set measurements is used for state estimation with set constraints, then one natural theoretical question is whether the final estimate satisfies the given constraints. It should be noted that the optimality criterion used in this work is MMSE conditioned on all measurement information up

to the current time. Since the given set constraints are already included in the conditioning, $\hat{x}_{k|k}$ and $P_{k|k}$ should have achieved this goal approximately. If $\hat{x}_{k|k}$ satisfies the given set constraints (e.g., due to projection when $\hat{x}_{k|k}$ and $P_{k|k}$ are obtained precisely without any approximation), then it is what we are looking for. Unfortunately due to approximation and computational error, this may not be the case. In this case, we have to make a choice between the MMSE optimality criterion and the inequality constraints. If the criterion is chosen, then $\hat{x}_{k|k}$ is what we are looking for, otherwise we need to project $\hat{x}_{k|k}$ to the closest point in the constraint subspace.

3.5.3 Relaxation of Gaussian assumption

Clearly, the key to the above derivation of the approximate MMSE filter is the Gaussian assumption (3.6). In reality due to the involvement of set measurements, this assumption is not true in general. An idea is to use a Gaussian mixture to approximate $p(x_{k-1}|z^{k-1})$ after each cycle of filtering as

$$p(x_{k-1}|z^{k-1}) = \sum_{i=1}^m \beta_{k-1}^{(i)} \mathcal{N}(x_{k-1}; \hat{x}_{k-1|k-1}^{(i)}, P_{k-1|k-1}^{(i)})$$

This is supported by the fact that any distribution can be approximated as closely as desired by a Gaussian mixture distribution [57].

For the next filtering cycle, the approximate MMSE filter is applied to each component to obtain the corresponding updated estimate $\mathcal{N}(x_k; \hat{x}_{k|k}^{(i)}, P_{k|k}^{(i)})$ autonomously. The weight $\beta_k^{(i)}$ of each component in the Gaussian mixture can be updated as follows when $z_k^{(2)}$ takes

the linear form (3.11)

$$\begin{aligned}
\tilde{\beta}_k^{(i)} &= \beta_{k-1}^{(i)} p(z_k | \hat{x}_{k|k-1}^{(i)}) \\
&= \beta_{k-1}^{(i)} \mathcal{N}(z_k^{(1)}; \hat{z}_{k|k-1}^{(1,i)}, S_k^{(1,i)}) P\{z_k^{(2)} \in \mathcal{Z}_k | z_k^{(2)} \sim \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k-1}^{(2,i)}, S_k^{(2,i)})\} \\
\beta_k^{(i)} &= \tilde{\beta}_k^{(i)} / \sum_{j=1}^m \tilde{\beta}_k^{(j)}, \quad i = 1, 2, \dots, m
\end{aligned}$$

where

$$\begin{aligned}
\hat{x}_{k|k-1}^{(i)} &= F_{k-1} \hat{x}_{k-1|k-1}^{(i)} \\
P_{k|k-1}^{(i)} &= F_{k-1} P_{k-1|k-1}^{(i)} F_{k-1}' + G_{k-1} Q_{k-1} G_{k-1}' \\
\hat{z}_{k|k-1}^{(j,i)} &= H_k^{(j)} \hat{x}_{k|k-1}^{(i)} \\
S_k^{(j,i)} &= H_k^{(j)} P_{k|k-1}^{(i)} (H_k^{(j)})' + R_k^{(j)}, \quad j = 1, 2
\end{aligned}$$

Remark: If a set measurement is available, the computation of the likelihood $P\{z_k^{(2)} \in \mathcal{Z}_k | z_k^{(2)} \sim \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k-1}^{(2,i)}, S_k^{(2,i)})\}$ in the above weight update is really a headache. It is suggested to use our proposed discretization method to get this probability as

$$P\{z_k^{(2)} \in \mathcal{Z}_k | z_k^{(2)} \sim \mathcal{N}(z_k^{(2)}; \hat{z}_{k|k-1}^{(2,i)}, S_k^{(2,i)})\} = \sum_{l=1}^L \mu_k^{(i,l)} I_{\mathcal{Z}_k}(y_k^{(i,l)})$$

where

$$I_{\mathcal{Z}_k}(y_k^{(i,l)}) = \begin{cases} 1 & \text{if } y_k^{(i,l)} \in \mathcal{Z}_k \\ 0 & \text{otherwise} \end{cases}$$

and $\{\mu_k^{(i,1)}, \mu_k^{(i,2)}, \dots, \mu_k^{(i,L)}\}$ is the set of probability masses of a discrete random vector $y_k^{(i)}$ over the space $\{y_k^{(i,1)}, y_k^{(i,2)}, \dots, y_k^{(i,L)}\}$, which is an approximation of the Gaussian distribution $\mathcal{N}(z_k^{(2)}; \hat{z}_{k|k-1}^{(2,i)}, S_k^{(2,i)})$.

Remark: The other forms of weight update in [57], derived under different optimality

criteria, can also be applied here similarly.

3.6 Discretization of Gaussian-related distribution

In summary, we met two types of discretization problems above. One is for a Gaussian distribution, and the other is for a truncated Gaussian distribution. Next we discuss how to find a close discretization method for each type.

From [80], it is known that a continuous random variable can be approximated by a discrete random variable as closely as desired, as described by the following lemma.

Lemma 1 [80]. Given a distance metric tolerance ϵ and a scalar continuous random variable x with range \mathcal{X} and cdf $F_x(t)$, the scalar discrete random variable y characterized by range \mathcal{Y}^* (with minimum cardinality $L = \lceil 1/2\epsilon \rceil =$ smallest integer not smaller than $1/2\epsilon$) and pmf μ^* , which is closest to x in the sense of minimum distribution mismatch, is given by

$$\begin{aligned}\mathcal{Y}^* &= \{y^{(1)}, y^{(2)}, \dots, y^{(L)}\} \\ \mu^* &= \{\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(L)}\} \\ y^{(i)} &= \arg\min_{t \in \mathcal{X}} [F_x(t) - (i - 1/2)/L], \quad i = 1, 2, \dots, L \\ \mu^{(i)} &= P\{y = y^{(i)} | y \in \mathcal{Y}^*\} = 1/L\end{aligned}$$

Unfortunately, we do not have such a nice result for the vector case. Now let us discuss how to discretize a general singular or nonsingular multivariate Gaussian distribution.

Given an m -dimensional Gaussian distribution

$$p(z) = \mathcal{N}(z; \mu_z, R_z)$$

it follows from singular value decomposition that there must exist a unitary matrix U such that

$$UR_zU' = \begin{bmatrix} R_z^{(\#)} & 0_{r \times (m-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (m-r)} \end{bmatrix}$$

where $R_z^{(\#)}$ is an $r \times r$ diagonal matrix with $R_z^{(\#)} > 0$ and $r = \text{rank}(R_z)$.

Let

$$a = [((R_z^{(\#)})^{1/2}b)', 0_{1 \times (m-r)}]'$$
 (3.18)

where $b \in \mathbb{R}^r$ and $b \sim \mathcal{N}(0_{r \times 1}, I_{r \times r})$. It can then be easily verified that

$$U'a + \mu_z \sim \mathcal{N}(\mu_z, R_z)$$
 (3.19)

The discretization of a Gaussian distribution can then be summarized as follows.

Step 1: Following Lemma 1, obtain the optimal discrete approximation to the standard Gaussian distribution and get L mass points in the one-dimensional space, each of which has a probability mass $1/L$. Construct r copies of these L mass points w.r.t. r component of b and then obtain the L^r mass points in an r -dimensional space simply through combining points in the r copies, which are a close approximation to the distribution of b , each having probability mass $1/L^r$.

Step 2: Transform all L^r mass points through Eqs. (3.18) and (3.19) to get a close discrete approximation of the Gaussian distribution $p(z) = \mathcal{N}(z; \mu_z, R_z)$.

Remark: As L increases, the approximation accuracy becomes better, to be shown next in the illustrative example part.

Remark: Since our set measurement formulation includes the linear inequality constraint as a special case, it follows that $S_k^{(2)} > 0$ may not always be true. In this case, the untruncated distribution $p(z_k^{(2)} | z^{k-1}, z_k^{(1)})$ may be a singular Gaussian distribution. That is exactly why we have the block forms above.

Remark: In essence, this discretization method is a *quasi Monte Carlo* method and it samples deterministically rather than randomly as in the Monte Carlo method.

Remark: The optimality criterion above for distribution approximation is the minimum distribution mismatch. The other criteria in [80], e.g., minimum distance and moment-matching, can also be applied.

In order to discretize a truncated Gaussian distribution, we can discretize the untruncated Gaussian distribution first, and then reject all those mass points located outside the truncation region \mathcal{Z} and renormalize the probability masses of all those mass points within the truncation region to sum up to the unity. Note, however, that this *rejection method* is not efficient, although it works in general. An efficient way was proposed in [67].

3.7 Illustrative examples

3.7.1 Example 1—Mean and covariance of truncated Gaussian distribution

As can be seen above, the truncated Gaussian distribution plays a key role in state estimation with point and set measurements. In the following, through a scalar truncated Gaussian distribution, we compare the accuracy, efficiency and stability of the Monte Carlo approach and our discretization approach in obtaining its mean and covariance.

For any scalar truncated Gaussian distribution

$$p(z) = \frac{1}{c_z} \mathcal{N}(z; \mu_z, R_z) I_{\mathcal{Z}}(z)$$

where

$$\mathcal{Z} = \{z : l_b \leq z \leq u_b\}$$

from [68], it is known that its mean and covariance can be calculated analytically by

$$E[z] = \mu_z + \frac{Z\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right) - Z\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right)}{\Phi\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right) - \Phi\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right)} \sqrt{R_z} \quad (3.20)$$

and

$$\text{cov}(z) = \left[1 + \frac{\frac{l_b - \mu_z}{\sqrt{R_z}} Z\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right) - \frac{u_b - \mu_z}{\sqrt{R_z}} Z\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right)}{\Phi\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right) - \Phi\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right)} - \left(\frac{Z\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right) - Z\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right)}{\Phi\left(\frac{u_b - \mu_z}{\sqrt{R_z}}\right) - \Phi\left(\frac{l_b - \mu_z}{\sqrt{R_z}}\right)} \right)^2 \right] R_z \quad (3.21)$$

where

$$Z(t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\}, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left\{-\frac{1}{2}s^2\right\} ds$$

The Monte Carlo approach used to obtain the mean and covariance of the above scalar truncated Gaussian distribution are the same as the discretization approach except that we need to randomly sample from the untruncated Gaussian distribution first as

$$z_u^{(i)} \sim \mathcal{N}(\mu_z, R_z), \quad i = 1, 2, \dots, L$$

where L is the number of samples. Then select only those $z_u^{(i)} \in \mathcal{Z}$ and denote them as $z^{(j)}$, $j = 1, 2, \dots, L_s$, where L_s is the number of selected samples from $\{z_u^{(i)}\}_{i=1}^L$, and thus

$$E[z] = \bar{z} \approx \frac{1}{L_s} \sum_{j=1}^{L_s} z^{(j)}$$

$$\text{cov}(z) \approx \frac{1}{L_s} \sum_{j=1}^{L_s} (z^{(j)} - \bar{z})^2 = \frac{1}{L_s} \sum_{j=1}^{L_s} (z^{(j)})^2 - \bar{z}^2$$

For the case $\mu_z = 0$, $R_z = 1$, $l_b = -1$, $u_b = 2$ and $L = 10, 20, 30, \dots, 600$, the comparison results of the estimates of the mean and covariance of z obtained from the analytical approach (using (3.20) and (3.21)), discretization approach and Monte Carlo approach (MC) are as

follows

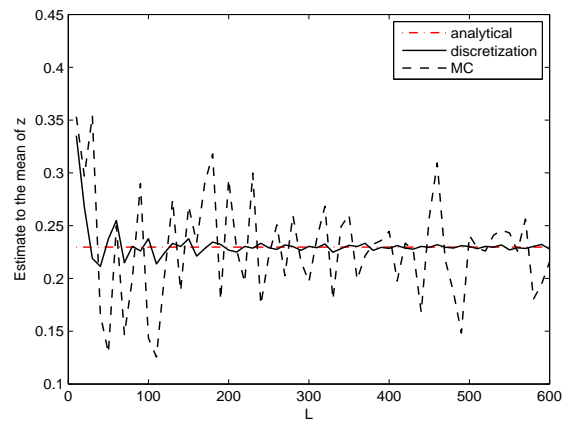


Figure 3.1: Comparison of estimate to the mean of z in one run

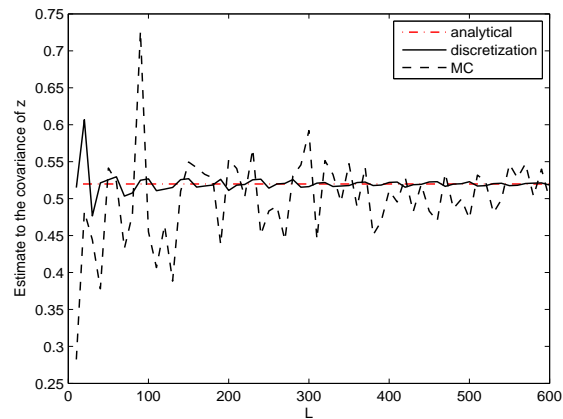


Figure 3.2: Comparison of estimate to the covariance of z in one run

From (3.20) and (3.21), it is known that the analytical solutions to the mean and covariance of z are

$$E[z] = 0.2296, \text{ cov}(z) = 0.5198$$

As can be seen clearly from the above plots, both the estimates from the discretization approach and the Monte Carlo approach fluctuate around these analytical solutions. By taking all L 's into account, the overall estimation accuracy of the discretization approach is

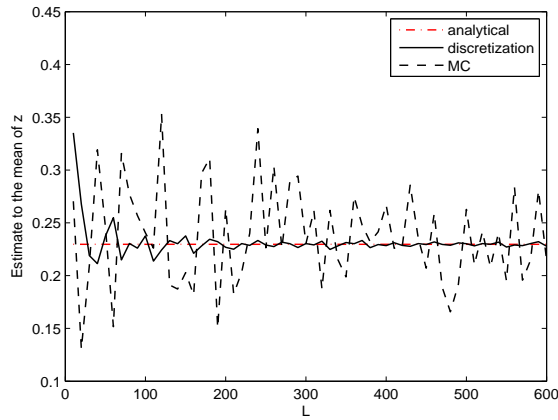


Figure 3.3: Comparison of estimate to the mean of z in another run

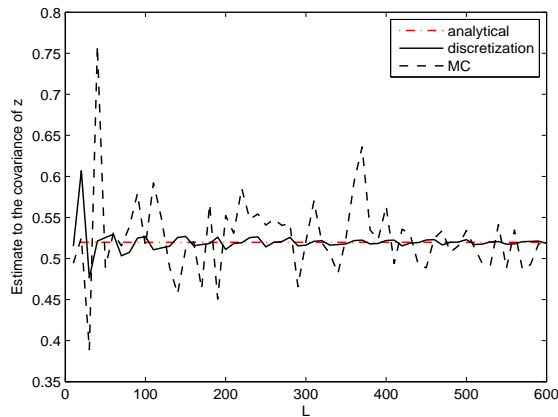


Figure 3.4: Comparison of estimate to the covariance of z in another run

much better than that of the Monte Carlo approach. Also, as L increases, the estimation accuracy for both the mean and covariance of z are getting better. It can also be seen that as L increases, the estimates from the discretization approach converge to the analytical solutions in a much faster rate than the Monte Carlo approach. When $L = 200$, the estimates from the discretization approach are already very close to the analytical solutions, while even when $L = 600$, there are still significant difference between the estimates from the Monte Carlo approach and the analytical solutions. From run to run, the estimates from the discretization approach demonstrate no change at all, while the estimates from the Monte Carlo approach

change substantially. This is because the discretization approach uses deterministic sampling while the Monte Carlo approach uses random sampling. In summary, in terms of estimation accuracy, convergence rate and stability, the discretization approach beats the Monte Carlo approach. That is also the reason why we chose to use the discretization approach above.

3.7.2 Example 2—Quantized estimation

As summarized above, the quantized measurement is a special case of set measurements. In the following, we illustrate the feasibility, accuracy and efficiency of applying our proposed approximate MMSE filter to state estimation with quantized measurements through a numerical example.

Consider the following scalar dynamic system [63] with quantized measurements only

$$\begin{aligned}x_k &= Fx_{k-1} + w_{k-1} \\z_k &= x_k + v_k \\y_k &= \Delta \text{round}(z_k/\Delta)\end{aligned}$$

where

$$\begin{aligned}F &= 0.95, w_k \sim \mathcal{N}(0, Q), Q = 0.1^2, \Delta = 5 \\v_k &\sim \mathcal{N}(0, R), R = 0.58^2, x_0 \sim \mathcal{N}(0, P_0), P_0 = 20\end{aligned}$$

and the function $\text{round}(\cdot)$ rounds the input argument to the nearest integer.

Fig. 3.5 shows comparison results of our proposed approximate MMSE filter (AMMSE, using (3.12), (3.13) and our discretization approach), the ideal Kalman filter with the measurements before quantization (iKF), bootstrap particle filter (BPF) and the posterior Cramer-Rao lower bound (PCRLB). The results are averaged over 100 Monte Carlo

runs. The cardinality of the discrete random variable which approximates the untruncated Gaussian distribution is 100 and the number of particles is 3,000.

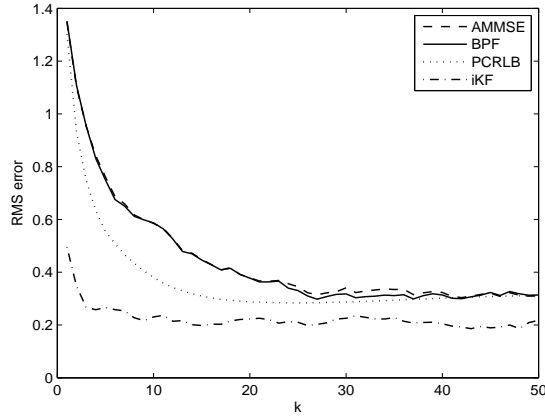


Figure 3.5: RMS error comparison for quantized estimation

It can be seen from the simulation result that our proposed approximate MMSE filter has very close performance to that of the particle filter but with much less computation, as shown in Table 3.1. There is a gap between the performance of our proposed filter and the PCRLB. We do not know whether the PCRLB for this problem is tight or not. The big difference between the PCRLB and the performance of the ideal Kalman filter is due to the quantization, which actually reveals the information loss caused by quantization.

Table 3.1: Comparison of computational burden

AMMSE	BPF
1	82.9822

The PCRLB $1/J_k$ for this problem can be recursively calculated as follows [32, 118]

$$J_{k+1} = D_k^{22} - D^{21} D^{12} / (J_k + D^{11})$$

where

$$J_0 = E[-\Delta_{x_0}^{x_0} \log p(x_0)] = 1/P_0$$

$$D^{11} = E[-\Delta_{x_k}^{x_k} \log p(x_{k+1}|x_k)] = F^2/Q$$

$$D^{12} = E[-\Delta_{x_k}^{x_{k+1}} \log p(x_{k+1}|x_k)] = -F/Q$$

$$D^{21} = E[-\Delta_{x_{k+1}}^{x_k} \log p(x_{k+1}|x_k)] = D^{12}$$

$$\begin{aligned} D_k^{22} &= E[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(x_{k+1}|x_k)] + E[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1}|x_{k+1})] \\ &= 1/Q + E[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1}|x_{k+1})] \end{aligned}$$

That is

$$\begin{aligned} J_{k+1} &= 1/Q + E[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1}|x_{k+1})] - F^2/(Q^2(J_k + F^2/Q)) \\ &= (Q + F^2/J_k)^{-1} + E[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1}|x_{k+1})] \end{aligned}$$

Clearly, to calculate PCRLB for this problem, the key is then to calculate $E[-\Delta_{x_k}^{x_k} \log p(y_k|x_k)]$.

Since $y_k \in \{\dots, -2\Delta, -\Delta, 0\Delta, \Delta, 2\Delta, \dots\}$ we have

$$\begin{aligned} p(y_k|x_k) &= P(y_k = j\Delta|x_k) = P\{j\Delta - \frac{\Delta}{2} \leq x_k + v_k < j\Delta + \frac{\Delta}{2}|x_k\} \\ &= P\{(j - \frac{1}{2})\Delta - x_k \leq v_k < (j + \frac{1}{2})\Delta - x_k\} = \int_{(j-\frac{1}{2})\Delta-x_k}^{(j+\frac{1}{2})\Delta-x_k} \frac{1}{\sqrt{2\pi R}} \exp\{-\frac{t^2}{2R}\} dt \\ &= \Phi(\frac{(j + \frac{1}{2})\Delta - x_k}{\sqrt{R}}) - \Phi(\frac{(j - \frac{1}{2})\Delta - x_k}{\sqrt{R}}) \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial p(y_k|x_k)}{\partial x_k} &= -\frac{1}{\sqrt{2\pi R}} (e^{-\frac{((j+\frac{1}{2})\Delta-x_k)^2}{2R}} - e^{-\frac{((j-\frac{1}{2})\Delta-x_k)^2}{2R}}) \\ \frac{\partial^2 p(y_k|x_k)}{\partial x_k^2} &= -\frac{1}{\sqrt{2\pi R^{\frac{3}{2}}}} ((j + \frac{1}{2})\Delta - x_k) e^{-\frac{((j+\frac{1}{2})\Delta-x_k)^2}{2R}} - ((j - \frac{1}{2})\Delta - x_k) e^{-\frac{((j-\frac{1}{2})\Delta-x_k)^2}{2R}} \end{aligned}$$

$$\frac{\partial^2 \log p(y_k|x_k)}{\partial x_k^2} = \frac{p(y_k|x_k) \frac{\partial^2 p(y_k|x_k)}{\partial x_k^2} - (\frac{\partial p(y_k|x_k)}{\partial x_k})^2}{p^2(y_k|x_k)}$$

From the Monte Carlo simulation, $E[-\Delta_{x_k}^{x_k} \log p(y_k|x_k)] = E[-\frac{\partial^2 \log p(y_k|x_k)}{\partial x_k^2}]$ can be calculated approximately by

$$E[-\Delta_{x_k}^{x_k} \log p(y_k|x_k)] \approx \frac{1}{M} \sum_{i=1}^M -\frac{\partial^2 \log p(y_k(i)|x_k(i))}{\partial x_k^2(i)}$$

where $y_k(i)$ and $x_k(i)$ are the realizations of y_k and x_k on the i -th Monte-Carlo run, respectively, and M is the number of runs.

3.7.3 Example 3—Inequality constrained estimation

As stated above, under certain conditions, the update by the set constraints is redundant. We also pointed out that if there exists model mismatch or nonlinearity or if $z_k^{(2)}$ is random, two-step update will improve performance in general. In the following, we verify these findings through numerical examples.

Consider the following dynamic system², which describes the motion of an on-road vehicle [66]:

$$x_k = F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \tag{3.22}$$

²This is the same dynamic system used in the illustrative examples part of chapter 2. For convenience, its description is repeated below.

where

$$x_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]', \quad w_k \sim \mathcal{N}(0, Q_k), \quad x_0 \sim \mathcal{N}(\bar{x}_0, P_0), \quad \bar{x}_0 = [0 \ 0 \ 10\sqrt{3} \ 10]'$$

$$F_k = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G_k = \begin{bmatrix} 0 \\ 0 \\ T \sin \theta \\ T \cos \theta \end{bmatrix}, \quad Q_k = \begin{bmatrix} 30 & 10\sqrt{3} & 0 & 0 \\ 10\sqrt{3} & 10 & 0 & 0 \\ 0 & 0 & 10 & 10\sqrt{3}/3 \\ 0 & 0 & 10\sqrt{3}/3 & 10/3 \end{bmatrix}$$

$$P_0 = N_k \text{diag}\{400, 400, 10, 10\} N_k, \quad N_k = I_4 - (H_k^{(2)})' (H_k^{(2)} (H_k^{(2)})')^{-1} H_k^{(2)}$$

$$H_k^{(2)} = [0 \ 0 \ 1 \ -\tan \theta], \quad \theta = \pi/3, \quad T = 2, \quad u_k = \begin{cases} 1, & \text{if } k \text{ is odd} \\ -1, & \text{if } k \text{ is even} \end{cases}$$

The state satisfies the following linear equality constraint [66]

$$H_k^{(2)} x_k = 0 \tag{3.23}$$

This is because the angle between the y axis and the road (treated as a straight line without width) is θ . For the filter, it only knows the following linear inequality constraint

$$-1 \leq H_k^{(2)} x_k \leq 1 \tag{3.24}$$

Figs. 2.1 and 2.2 show the true vehicle trajectory and velocity in one run in the two-dimensional Cartesian coordinate plane. It can be seen that the implicit linear equality constraint (3.23) is satisfied, and thus the linear inequality constraint (3.24) is also satisfied.

The measurement is described by

$$z_k^{(1)} = H_k^{(1)} x_k + v_k^{(1)} \tag{3.25}$$

where

$$v_k^{(1)} \sim \mathcal{N}(0, R_k^{(1)}), H_k^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, R_k^{(1)} = \begin{bmatrix} 400 & 0 & 0 \\ 0 & 400 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

We want to estimate the state in the MMSE sense based on the measurement $z_k^{(1)}$ and linear inequality constraint (3.24). In the following, all estimators were initialized with $P_{0|0} = P_0$. A mismatched Q_k :

$$Q_k^{\text{mis}} = \begin{bmatrix} 30 & 10\sqrt{3} & 0 & 0 \\ 10\sqrt{3} & 10 & 0 & 0 \\ 0 & 0 & 13.25 & 0.1443 \\ 0 & 0 & 0.1443 & 13.0833 \end{bmatrix}$$

is used in place of Q_k in some estimators, possibly along with a mis-specified initial estimate

$$\hat{x}_{0|0}^{\text{mis}} = \bar{x}_0 + [0 \ 0 \ 6 \ 12]'$$

Also, all results were averaged over 200 Monte Carlo runs.

Figs. 3.6 through 3.11 show comparison results of estimators in Table 3.2.

Since the noisy point measurement $z_k^{(1)}$ is linear in x_k , under the given knowledge and with correct Q and $\hat{x}_{0|0}$, the KF updated only by noisy measurement is optimal. Checking against the condition stated above, it can be easily justified that the update by the linear inequality constraint can be skipped. In other words, the KF updated only by noisy measurement provides the optimal estimate for this constrained estimation problem. It can be seen from Figs. 3.6 through 3.11 that the best performance is achieved by the KF, as just analyzed.

From Figs. 3.8 and 3.9, it can be seen that if there is Q mismatch and linear inequality constraint is not taken into account in the update step, the PM-Q filter has much worse

Table 3.2: Estimators used in inequality constrained estimation example

name	explanation
KF	true Q_k , $\hat{x}_{0 0} = \bar{x}_0$, updated only by point measurement $z_k^{(1)}$
IEC	true Q_k , $\hat{x}_{0 0} = \bar{x}_0$, updated by point measurement $z_k^{(1)}$ first and then by the inequality constraint (3.24) using (3.12), (3.13)
PM-Q	Q_k^{mis} , $\hat{x}_{0 0} = \bar{x}_0$, updated only by point measurement $z_k^{(1)}$
IEC-Q	Q_k^{mis} , $\hat{x}_{0 0} = \bar{x}_0$, updated by point measurement $z_k^{(1)}$ first and then by the inequality constraint (3.24) using (3.12), (3.13)
PM-0	true Q_k , $\hat{x}_{0 0} = \hat{x}_{0 0}^{\text{mis}}$, updated only by point measurement $z_k^{(1)}$
IEC-0	true Q_k , $\hat{x}_{0 0} = \hat{x}_{0 0}^{\text{mis}}$, updated by point measurement $z_k^{(1)}$ first and then by the inequality constraint (3.24) using (3.12), (3.13)

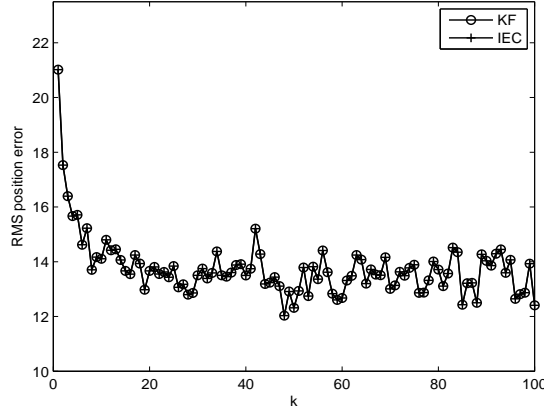


Figure 3.6: RMS position error comparison with correct Q and $\hat{x}_{0|0} = \bar{x}_0$. Note that KF overlaps with IEC.

performance than the KF. But if there is Q mismatch and the linear inequality constraint is fully accounted for in the update step, the performance of the IEC-Q filter is almost the same as that of the KF. From Figs. 3.10 and 3.11, it can be seen that if $\hat{x}_{0|0}$ is mis-specified and the linear inequality constraint is not taken into account in the update step, the PM-0 estimator has the worst performance and its position RMSE diverges. But if $\hat{x}_{0|0}$ is mis-specified and the linear inequality constraint is fully taken into account in the update step, except during the initial short transient, IEC-0 and KF have comparable position RMSE and

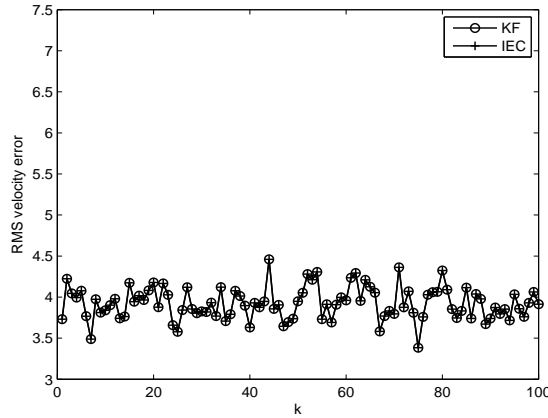


Figure 3.7: RMS velocity error comparison with correct Q and $\hat{x}_{0|0} = \bar{x}_0$. Note that KF overlaps with IEC.

almost identical velocity RMSE. All these verify our statements about the contribution of set constraints in practical constrained estimation problems. We should take the set constraint into account when there exists model mismatch.

3.7.4 Example 4—Noise-free measurements

As discussed above, set constraints can be treated as noise-free set measurements. Also, under certain conditions, update by set constraints is redundant. Through the following numerical example, we want to point out that this argument is not true for all types of noise-free set measurements. That is, not all types of noise-free set measurements are redundant in the update step of state estimation. They should be handled with care.

In this example, the dynamic system considered is the same as the one described by (3.22). The measurement provided by another sensor consists of two types. The first type is the noisy point measurement described by (3.25), and the other type is noise-free set measurement which is described by

$$z_k^{(3)} = H_k^{(3)} x_k \in \mathcal{Z}_k \quad (3.26)$$

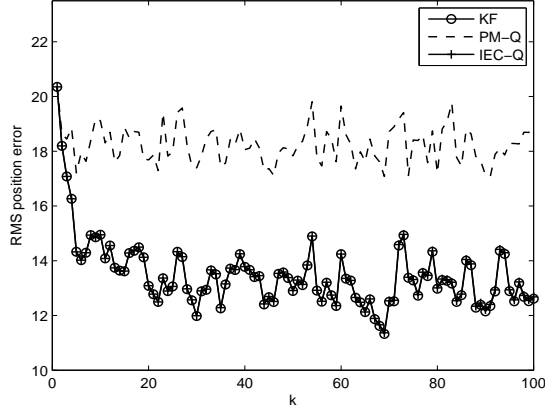


Figure 3.8: RMS position error comparison with Q mismatch. Note that KF and IEC-Q have no noticeable difference.

where

$$\mathcal{Z}_k = \{z_k^{(3)} : H_k^{(3)} x_k - \Delta_z \leq z_k^{(3)} \leq H_k^{(3)} x_k + \Delta_z\}, \quad H_k^{(3)} = [2 \quad 3 \quad 0 \quad 0], \quad \Delta_z = 6$$

For the filter, it only knows the interval \mathcal{Z}_k in the real line but not how \mathcal{Z}_k was generated.

We want to estimate the state in MMSE sense based on the noisy point measurement $z_k^{(1)}$ described by (3.25), linear inequality constraint (3.24) and the noise-free set measurement \mathcal{Z}_k described by (3.26). In the following, all estimators were initialized with $\hat{x}_{0|0} = \bar{x}_0$, $P_{0|0} = P_0$ and the true Q_k . Also, all results were averaged over 200 Monte Carlo runs.

Figs. 3.12 and 3.13 show comparison results of estimators in Table 3.3.

Table 3.3: Estimators used in noise-free measurement example

name	explanation
KF	updated only by noisy point measurement $z_k^{(1)}$
NFSM	updated by noisy point measurement $z_k^{(1)}$ first and then by noise-free set measurement \mathcal{Z}_k using (3.12) and (3.13)

As was already illustrated in the above inequality constrained example, the update by the linear inequality constraint (3.24) is redundant. This means that even if the KF and NFSM

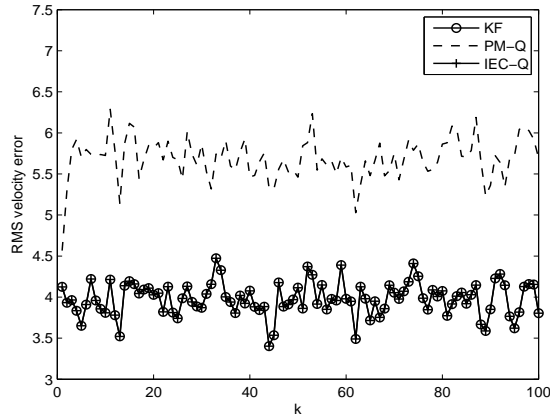


Figure 3.9: RMS velocity error comparison with Q mismatch. Note that KF and IEC- Q have no noticeable difference.

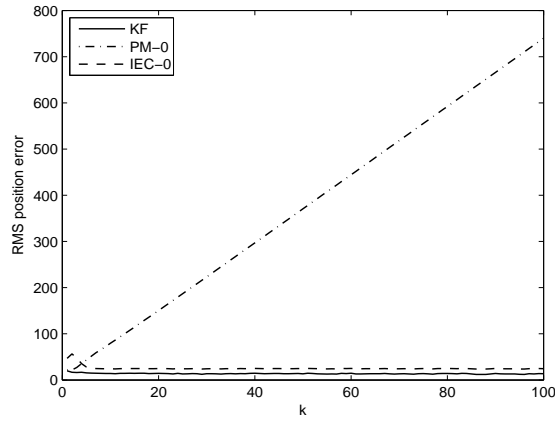


Figure 3.10: RMS position error comparison with mis-specified $\hat{x}_{0|0}$

filter were also updated by the linear inequality constraint (3.24), their performance should stay the same. But as is clear from the simulation results, the update by the noise-free set measurement \mathcal{Z}_k is indispensable—it leads to a significant performance improvement. That is, under our problem formulation, although both \mathcal{Z}_k and the linear inequality constraint (3.24) are treated as noise-free set measurements, their effects are totally different, as demonstrated in this example.

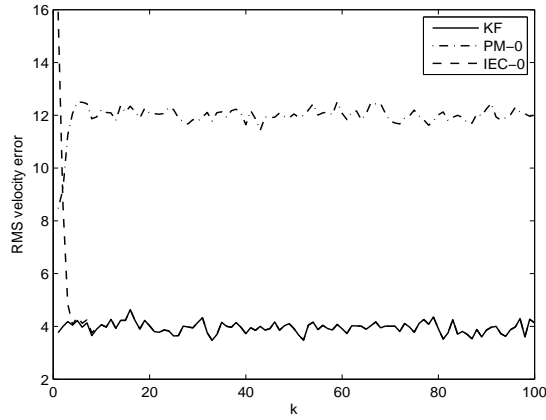


Figure 3.11: RMS velocity error comparison with mis-specified $\hat{x}_{0|0}$

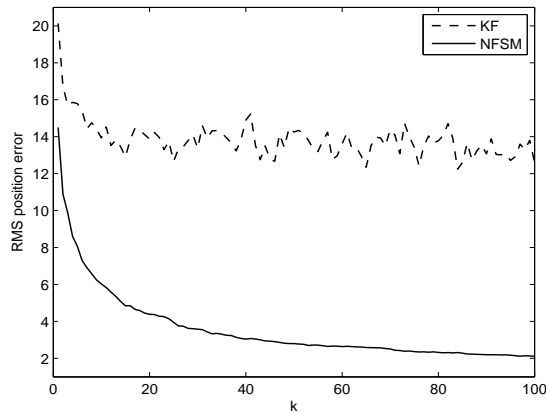


Figure 3.12: RMS position error comparison for example 4

3.8 Summary

This chapter targets the state estimation problem with point and set measurements. The main difficulty and intractability for solving this type of problem comes from the uncertainty associated with set measurements. The key idea proposed in this work is to approximate a continuous set measurement by discrete point measurements under a Gaussian assumption. Possible ways to relax the Gaussian assumption and to discretize the involved Gaussian and truncated Gaussian distributions are discussed. Numerical examples show that in terms of

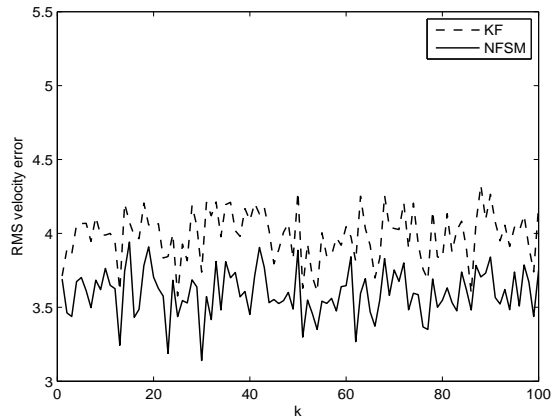


Figure 3.13: RMS velocity error comparison for example 4

accuracy, efficiency and stability, the proposed discretization method should be preferred when compared with the Monte Carlo approach. It is also shown that the estimator with quantized measurements with the proposed discretization method has very close performance to that of the particle filter but with a far less computational load. For state estimation subject to inequality constraints, under certain conditions, the update by them is redundant. Also, although set constraints are redundant for the update in some filters, this is not true for all types of noise-free set measurements.

Chapter 4

Lossless Linear Transformation of Sensor Data for Distributed Estimation Fusion

4.1 Introduction and related research

Estimation fusion [81], or data fusion for estimation, is the problem of how to best utilize useful information contained in multiple sets of data for the purpose of estimating a quantity—a parameter or process. It has widespread applications in military and civilian fields, e.g., target tracking and localization, air traffic control, guidance and navigation, fault diagnosis, surveillance and monitoring. Estimation fusion is used for potential improved estimation accuracy [43], enhanced reliability and survivability, extended coverage and observability, etc.

There are many difficulties facing estimation fusion. For example, in wireless sensor networks, the computational resources, communication bandwidth and power consumption are usually limited. Also communication delay is inevitable. One well-known phenomenon about communication delay is the so-called out-of-sequence measurement [4, 7–9, 127, 129]. In reality, multiple sensors often work asynchronously [1, 55, 56, 74, 82, 88, 122] instead of synchronously, which is widely assumed in the literature. Multiple sensors may also have different sampling rates and communication rates. Some other difficulties include multiple

targets and multipaths, which may need the introduction of measurement or track association, etc.

Two basic fusion architectures are well known [81]: *centralized* and *decentralized/distributed* (also referred to as *measurement fusion* and *track fusion* in target tracking, respectively), depending on whether the raw measurements are sent to the fusion center or not. In centralized fusion, all raw measurements are sent to the fusion center, while in distributed fusion, each sensor only sends in processed data. They have pros and cons in terms of performance, channel requirements, reliability, survivability, information sharing, etc. Although many practical issues do exist, theoretically, centralized fusion is nothing but estimation with distributed data. Distributed fusion is more challenging and has been a focal point of most fusion research. With regard to communication, there is an unresolved dispute about which of these two basic architectures is a better choice. For example, some argue that standard distributed fusion (where local estimates are transmitted) should be preferred since sending raw measurements is usually more demanding. This argument seems reasonable. Unfortunately, to obtain the cross-correlation of the estimation errors across sensors, extra communication of filtering gain, measurement matrix, etc., are also needed. Then it is doubtful that distributed fusion can still beat centralized fusion communicationwise. But as will be shown in this chapter, there does exist a distributed fusion algorithm which can beat the centralized one in terms of communication while having the same performance.

Distributed fusion has been researched for several decades and numerous results are available. Two general approaches were used [75]. One is the equivalence to centralized fusion [23, 25, 32, 35, 53, 111]. This approach is more popular but it is much harder to follow. The other is the estimation approach that converts an estimation fusion problem to a standard estimation problem by treating whatever is available to the fuser as data in a standard estimation problem [6, 21, 22, 24, 32, 64, 81]. For example, [6] and [21] proposed a two-sensor track-to-track fusion algorithm based on the maximum likelihood (ML) method for the

Gaussian case. For more than two sensors, [24,64] proposed a track-to-track fusion algorithm based on the ML (for the Gaussian case) and weighted least-squares (WLS) methods. [25,53] proposed a decentralized structure to reconstruct the optimal centralized estimate when the measurement noises across sensors are uncorrelated, and [111] for the correlated case. [81] proposed unified fusion rules in the sense of best linear unbiased estimate (BLUE) and WLS for all fusion architectures with arbitrary correlation, which includes distributed fusion as a special class. [22] proposed a distributed fusion algorithm which is optimal in the sense of maximum a posteriori (MAP). [117] proposed a distributed fusion algorithm which is optimal in the sense of MMSE for the Gaussian case or linear MMSE by the additional use of local predicted estimates when compared with [22].

Most existing distributed fusion algorithms, e.g., the ones based on the information form of the Kalman filter and the optimal WLS estimator, presume the existence of the inverses of error covariance matrices. Theoretically speaking, the error covariance matrices are only at least positive semi-definite, not necessarily invertible. For example, some measurement components may be accurate enough to be reasonably assumed perfect, i.e., they have no error [17,44,45,47,54,59,90,91]. This is of practical importance [47] since it is often the case, e.g., in many industrial control systems [91]. In this case, the measurement noise and the error covariance matrix will be singular. This means that we can not always use the existing distributed fusion algorithms directly. One simplest solution is to replace the matrix inverse when it does not exist by some generalized inverse (e.g., the *MP inverse*). But in this way, the original distributed fusion algorithm is not necessarily optimal any longer.

Considering communication constraints (e.g., on communication bandwidth or power consumption or both) and affordable computational resources at the fusion center (e.g., in sensor networks), it is more beneficial for local sensors to send in compressed data. There has been some discussion [33,101,110,111,128,133] in the literature about compression in the hope of reducing the communication requirements. They mainly differ in the following

aspects. For example, where are the compression rules constructed, the fusion center as a whole or at each sensor separately? Which of the distributed fusion approaches mentioned above is used? Does construction of the compression rule at one sensor need information from other sensors? Some of the existing methods may be questionable: besides the transmission of compressed data, information necessary for the construction of compression rules also needs to be transmitted over the communication network. This may run counter to the motivation for compression if all these transmissions are taken into account although the results are indeed optimal in some sense.

In this chapter, by taking two linear transformations of the raw measurement at each sensor, two optimal distributed fusion algorithms are proposed. Compared with the existing fusion algorithms, they have three nice properties which make them attractive. Pros and cons of these two new algorithms are analyzed. Sequential processing of the transformed data is discussed, which is intended to reduce the computational complexity of the proposed algorithms. Compression along time in the case of reduced-rate communication for some simple cases and an extension to the singular measurement noise case are also discussed.

This chapter is organized as follows. Sec. 4.2 formulates the multi-sensor distributed fusion problem. Sec. 4.3 briefly reviews the background of this research work. Sec. 4.4 describes two new linear transformations of the raw measurement at each sensor. Sec. 4.5 presents distributed fusion algorithms with the transformed data. Sec. 4.6 analyzes their optimality. Sec. 4.7 provides one way to reduce their computational complexity. Sec. 4.8 discusses compression along time in the case of reduced-rate communication for some simple cases. Sec. 4.9 discusses extension to the singular measurement noise case. Sec. 4.10 gives summary. Sec. 4.11 provides proofs to all theorems.

4.2 Problem formulation

Consider the following generic dynamic system

$$x_k = F_{k-1}x_{k-1} + G_{k-1}w_{k-1} \quad (4.1)$$

with zero-mean white noise w_k with $\text{cov}(w_k) = Q_k \geq 0$ and $x_k \in \mathbb{R}^n$, $E[x_0] = \bar{x}_0$, $\text{cov}(x_0) = P_0$.

Assume that altogether N_s sensors are used to observe the state simultaneously

$$z_k^{(i)} = H_k^{(i)}x_k + v_k^{(i)}, \quad i = 1, 2, \dots, N_s \quad (4.2)$$

with zero-mean white noise $v_k^{(i)}$ with $\text{cov}(v_k^{(i)}) = R_k^{(i)} > 0$ and $z_k^{(i)} \in \mathbb{R}^{m_i}$. $\langle w_k \rangle$, $\langle v_k^{(i)} \rangle$, and x_0 are uncorrelated with each other. It is also assumed that the measurement noises across sensors are uncorrelated.

Remark: The assumption $R_k^{(i)} > 0$ is not restrictive in our framework, as explained in Sec. 4.9.

In distributed fusion, the fusion center tries to get the best estimate of the state with the processed data received from each sensor.

In this chapter, by distributed fusion, it is meant that data-processed observations, not necessarily the local estimates, are available at the fusion center. Systems with only local estimates available at the fusion center, referred to as the standard distributed fusion in [81], are not the focus of this chapter.

4.3 Review of existing distributed fusion algorithms

Before introducing our proposed approaches, we first give a brief review of three existing popular distributed fusion algorithms in terms of optimality, communication requirements, and dependence on the inverses of error covariance matrices to explain the necessity of our work.

4.3.1 Information matrix fusion

Denote the previous fused updated estimate as $\hat{x}_{k-1|k-1}^d$, $P_{k-1|k-1}^d$ and the current local predicted and updated estimates as $\hat{x}_{k|k-1}^{(i)}$, $P_{k|k-1}^{(i)}$, $\hat{x}_{k|k}^{(i)}$, $P_{k|k}^{(i)}$, $i = 1, 2, \dots, N_s$. [25] first proposed the information matrix distributed fusion algorithm, which is equivalent to the centralized one.

There are two different but equivalent forms. The one without feedback is

$$\begin{aligned} (P_{k|k}^d)^{-1} &= (P_{k|k-1}^d)^{-1} + \sum_{i=1}^{N_s} ((P_{k|k}^{(i)})^{-1} - (P_{k|k-1}^{(i)})^{-1}) \\ (P_{k|k}^d)^{-1} \hat{x}_{k|k}^d &= (P_{k|k-1}^d)^{-1} \hat{x}_{k|k-1}^d + \sum_{i=1}^{N_s} ((P_{k|k}^{(i)})^{-1} \hat{x}_{k|k}^{(i)} - (P_{k|k-1}^{(i)})^{-1} \hat{x}_{k|k-1}^{(i)}) \end{aligned}$$

and the other one with feedback is

$$\begin{aligned} (P_{k|k}^d)^{-1} &= (P_{k|k-1}^d)^{-1} + \sum_{i=1}^{N_s} ((P_{k|k}^{(i)})^{-1} - (P_{k|k-1}^d)^{-1}) \\ (P_{k|k}^d)^{-1} \hat{x}_{k|k}^d &= (P_{k|k-1}^d)^{-1} \hat{x}_{k|k-1}^d + \sum_{i=1}^{N_s} ((P_{k|k}^{(i)})^{-1} \hat{x}_{k|k}^{(i)} - (P_{k|k-1}^d)^{-1} \hat{x}_{k|k-1}^d) \end{aligned}$$

As we know, this algorithm came from an ingenious equivalent transformation of the optimal centralized fusion algorithm, in which all the raw measurements are replaced by their corresponding local predicted and updated estimates. It can be seen that, to make this algorithm work and to guarantee its optimality, all the involved in-

verses $(P_{k|k}^d)^{-1}$, $(P_{k|k-1}^d)^{-1}$, $(P_{k|k}^{(i)})^{-1}$, $(P_{k|k-1}^{(i)})^{-1}$ (needed only when there is no feedback), $i = 1, 2, \dots, N_s$ should exist. Sensor i needs to transmit $\hat{x}_{k|k}^{(i)} \in \mathbb{R}^n$ and $P_{k|k}^{(i)} \in \mathbb{R}^{n \times n}$ to the fusion center. When there is no feedback, it also needs to transmit $\hat{x}_{k|k-1}^{(i)} \in \mathbb{R}^n$ and $P_{k|k-1}^{(i)} \in \mathbb{R}^{n \times n}$. The communication is one-way in this case and the total communication requirements from each sensor to the fusion center are $2(n + n^2)$. When there is feedback, the fusion center also needs to feedback $\hat{x}_{k|k-1}^d \in \mathbb{R}^n$ and $P_{k|k-1}^d \in \mathbb{R}^{n \times n}$ to each sensor. The communication will be two-way then, but the total communication requirements between each sensor and the fusion center are still $2(n + n^2)$.

4.3.2 Simple convex combination

The key idea of the simple convex combination method [10, 26, 87] is to ignore the cross-correlation of the local estimation errors across sensors, so it is usually worse compared with other distributed fusion methods, let alone with centralized fusion. Under this idea, the fused estimate can be simplified to

$$\begin{aligned}\hat{x}_{k|k}^d &= P_{k|k}^d \cdot \sum_{i=1}^{N_s} (P_{k|k}^{(i)})^{-1} \hat{x}_{k|k}^{(i)} \\ P_{k|k}^d &= \left(\sum_{i=1}^{N_s} (P_{k|k}^{(i)})^{-1} \right)^{-1}\end{aligned}$$

Clearly in this method, sensor i only needs to transmit $\hat{x}_{k|k}^{(i)}$ and $P_{k|k}^{(i)}$ to the fusion center and the total communication requirements from each sensor to the fusion center are just $n + n^2$. This method also requires the existence of $(P_{k|k}^{(i)})^{-1}$.

4.3.3 WLS fusion

For any two given sensors i and j , denote the cross-covariance between their state estimation errors at time k as $P_{k|k}^{(i,j)}$. From [3, 10], it is known that $P_{k|k}^{(i,j)}$ can be obtained by

$$P_{k|k}^{(i,j)} = (I - K_k^{(i)} H_k^{(i)}) P_{k|k-1}^{(i,j)} (I - K_k^{(j)} H_k^{(j)})' = P_{k|k}^{(i)} (P_{k|k-1}^{(i)})^{-1} P_{k|k-1}^{(i,j)} (P_{k|k-1}^{(j)})^{-1} P_{k|k}^{(j)}$$

$$P_{k|k-1}^{(i,j)} = F_{k-1} P_{k-1|k-1}^{(i,j)} F_{k-1}' + G_{k-1} Q_{k-1} G_{k-1}'$$

When there is feedback from the fusion center to local estimators, the calculation of $P_{k|k}^{(i,j)}$ can be simplified to

$$P_{k|k}^{(i,j)} = (I - K_k^{(i)} H_k^{(i)}) P_{k|k-1}^d (I - K_k^{(j)} H_k^{(j)})' = P_{k|k}^{(i)} (P_{k|k-1}^d)^{-1} P_{k|k}^{(j)}$$

The WLS fusion [24, 64] can then be described as

$$\hat{x}_{k|k}^d = P_{k|k}^d \mathbf{I}'_{N_s} (\mathbf{P}_{k|k}^{N_s})^{-1} \hat{\mathbf{x}}_{k|k}^{N_s}$$

$$P_{k|k}^d = (\mathbf{I}'_{N_s} (\mathbf{P}_{k|k}^{N_s})^{-1} \mathbf{I}_{N_s})^{-1}$$

where

$$\mathbf{I}_{N_s} = \underbrace{\begin{bmatrix} I_{n \times n} & I_{n \times n} & \cdots & I_{n \times n} \end{bmatrix}}_{N_s \ I_{n \times n}}', \quad \hat{\mathbf{x}}_{k|k}^{N_s} = \begin{bmatrix} (\hat{x}_{k|k}^{(1)})' & (\hat{x}_{k|k}^{(2)})' & \cdots & (\hat{x}_{k|k}^{(N_s)})' \end{bmatrix}'$$

$$\mathbf{P}_{k|k}^{N_s} = \begin{bmatrix} P_{k|k}^{(1)} & P_{k|k}^{(1,2)} & \cdots & P_{k|k}^{(1,N_s)} \\ P_{k|k}^{(2,1)} & P_{k|k}^{(2)} & \cdots & P_{k|k}^{(2,N_s)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{k|k}^{(N_s,1)} & P_{k|k}^{(N_s,2)} & \cdots & P_{k|k}^{(N_s)} \end{bmatrix}$$

Compared with centralized fusion, the WLS fusion algorithm is suboptimal. For any

sensor i , it needs to transmit $\hat{x}_{k|k}^{(i)}$ and $P_{k|k}^{(i)}$ to the fusion center. If $P_{k|k-1}^{(i)}$ is invertible, then it also needs to transmit $P_{k|k-1}^{(i)}$; otherwise it also needs to transmit $K_k^{(i)} \in \mathbb{R}^{n \times m_i}$ and $H_k^{(i)} \in \mathbb{R}^{m_i \times n}$. With the feedback, the fusion center also needs to feedback $\hat{x}_{k|k-1}^d$ and $P_{k|k-1}^d$. If $P_{k|k-1}^d$ is not invertible, then sensor i also needs to transmit $K_k^{(i)}$ and $H_k^{(i)}$ to the fusion center. In sum, the total communication requirements for all cases are larger than $n + n^2$. Also, clearly this algorithm requires the existence of $(\mathbf{P}_{k|k}^{N_s})^{-1}$.

In the following, we will discuss two new optimal distributed fusion algorithms which are equivalent to the optimal centralized one and are still applicable even if the error covariance matrices may not be invertible. Meanwhile, it is expected that the communication between local sensors and the fusion center can be reduced to some extent.

4.4 Sensor measurement transformation

4.4.1 Transformation I

Let

$$T_k^{(i)} = (H_k^{(i)})'(R_k^{(i)})^{-1} \quad (4.3)$$

$$\bar{z}_k^{(i)} = T_k^{(i)} z_k^{(i)} \in \mathbb{R}^n \quad (4.4)$$

Then from Eq. (4.2), it follows that

$$\bar{z}_k^{(i)} = T_k^{(i)} H_k^{(i)} x_k + T_k^{(i)} v_k^{(i)}$$

Furthermore, let

$$\begin{aligned}\bar{H}_k^{(i)} &= T_k^{(i)} H_k^{(i)} = (H_k^{(i)})' (R_k^{(i)})^{-1} H_k^{(i)} \in \mathbb{R}^{n \times n} \\ \bar{v}_k^{(i)} &= T_k^{(i)} v_k^{(i)}\end{aligned}\tag{4.5}$$

Then the above equation can be rewritten as

$$\bar{z}_k^{(i)} = \bar{H}_k^{(i)} x_k + \bar{v}_k^{(i)}, \quad i = 1, 2, \dots, N_s\tag{4.6}$$

with zero-mean white noise $\langle \bar{v}_k^{(i)} \rangle$ with

$$\bar{R}_k^{(i)} = \text{cov}(\bar{v}_k^{(i)}) = T_k^{(i)} R_k^{(i)} (T_k^{(i)})' = \bar{H}_k^{(i)}\tag{4.7}$$

uncorrelated with $\langle w_k \rangle$ and x_0 , uncorrelated across sensors.

Remark: For the multi-sensor measurement system (4.2), if $H_k^{(i)}$ and $R_k^{(i)}$, $i = 1, 2, \dots, N_s$, are time-varying, then in centralized fusion, sensor i may also need to transmit $R_k^{(i)}$ to the fusion center besides $z_k^{(i)}$ and $H_k^{(i)}$. But for our new transformed measurement equation (4.6), although it has a similar form of the original measurement equation (4.2), there is no need to transmit $\bar{R}_k^{(i)}$ to the fusion center at all, even when it is time-varying, due to its equivalence to $\bar{H}_k^{(i)}$. This will certainly help reduce communication from each sensor to the fusion center and is a nice property of our new algorithm.

In distributed fusion, each sensor sends the processed data $\bar{z}_k^{(i)}$ and the corresponding measurement matrix $\bar{H}_k^{(i)}$ to the fusion center. In all, the communication requirements from each sensor to the fusion center are $n + n^2$ at any time instant. This is equal to or less than what is required by most existing distributed fusion algorithms.

Remark: Clearly the new transformed data $\bar{z}_k^{(i)} = T_k^{(i)} z_k^{(i)}$ is actually the so-called information state in the information form of the Kalman filter. That is, originally we are handling

the estimation fusion problem in the measurement space with dimension m_i , $i = 1, 2, \dots, N_s$.

The linear transformation (4.4) converts it into the estimate space.

Remark: For any sensor i , if $(P_{k|k-1}^{(i)})^{-1}$ and $(P_{k|k}^{(i)})^{-1}$ exist, then it follows from the information form of the Kalman filter that

$$(P_{k|k}^{(i)})^{-1} \hat{x}_{k|k}^{(i)} - (P_{k|k-1}^{(i)})^{-1} \hat{x}_{k|k-1}^{(i)} = (H_k^{(i)})' (R_k^{(i)})^{-1} z_k^{(i)}$$

This is exactly where our idea for the transformation (4.4) comes from.

4.4.2 Transformation II

In distributed fusion with compressed data, a mapping $g_k^{(i)}(\cdot)$, $i = 1, 2, \dots, N_s$, is applied to compress local raw measurement first:

$$\bar{z}_k^{(i)} = g_k^{(i)}(z_k^{(i)})$$

where $\dim(\bar{z}_k^{(i)}) \leq \dim(z_k^{(i)})$. Then the compressed data $\bar{z}_k^{(i)}$ is sent to the fusion center for estimation.

In this work, only a linear compression rule is considered:

$$\bar{z}_k^{(i)} = C_k^{(i)} z_k^{(i)}$$

Also, it is required that the compression rule at each sensor be constructed locally based on local information only and there is no feedback from the fusion center to sensors.

Given

$$\text{rank}(H_k^{(i)}) = r_i, \quad r_i \leq \min(m_i, n)$$

from full rank decomposition, we have

$$H_k^{(i)} = M_k^{(i)} N_k^{(i)} \quad (4.8)$$

where $M_k^{(i)} \in \mathbb{R}^{m_i \times r_i}$ is of full column rank and $N_k^{(i)} \in \mathbb{R}^{r_i \times n}$ is of full row rank.

Substituting $H_k^{(i)}$ of Eq. (4.8) into Eq. (4.3), we have

$$T_k^{(i)} = (N_k^{(i)})' (M_k^{(i)})' (R_k^{(i)})^{-1}$$

Since $N_k^{(i)}$ is of full row rank, if both sides are premultiplied by $((N_k^{(i)})')^+$, we have an intermediate transformation matrix

$$\tilde{T}_k^{(i)} = (M_k^{(i)})' (R_k^{(i)})^{-1}$$

Finally, let

$$C_k^{(i)} = ((M_k^{(i)})' (R_k^{(i)})^{-1} M_k^{(i)})^{-1/2} \tilde{T}_k^{(i)} \quad (4.9)$$

Then from Eq. (4.2), it follows that

$$\bar{z}_k^{(i)} = C_k^{(i)} z_k^{(i)} = C_k^{(i)} H_k^{(i)} x_k + C_k^{(i)} v_k^{(i)} = \bar{H}_k^{(i)} x_k + \bar{v}_k^{(i)} \quad (4.10)$$

where

$$\begin{aligned} \bar{H}_k^{(i)} &= C_k^{(i)} H_k^{(i)} \\ \bar{v}_k^{(i)} &= C_k^{(i)} v_k^{(i)} \end{aligned} \quad (4.11)$$

with zero-mean white noise $\langle \bar{v}_k^{(i)} \rangle$ with

$$\bar{R}_k^{(i)} = \text{cov}(\bar{v}_k^{(i)}) = C_k^{(i)} R_k^{(i)} (C_k^{(i)})' = I_{r_i \times r_i} \quad (4.12)$$

which is uncorrelated with $\langle w_k \rangle$ and x_0 , uncorrelated across sensors and $\bar{z}_k^{(i)} \in \mathbb{R}^{r_i}$, and

$$\bar{H}_k^{(i)} = ((M_k^{(i)})'(R_k^{(i)})^{-1}M_k^{(i)})^{1/2}N_k^{(i)} \in \mathbb{R}^{r_i \times n} \quad (4.13)$$

is of full row rank.

Remark: Since $M_k^{(i)}$ is of full column rank and $R_k^{(i)} > 0$, $(M_k^{(i)})'(R_k^{(i)})^{-1}M_k^{(i)}$ is always invertible.

Remark: Similar to the above analysis for transformation I, even though the multi-sensor measurement system (4.2) is time-varying, there is still no need to transmit $\bar{R}_k^{(i)}$ to the fusion center at all since $\bar{R}_k^{(i)} = I_{r_i \times r_i}$. This will certainly help reduce communication from each sensor to the fusion center and is a nice property of our new algorithm.

In distributed fusion, each sensor sends the processed data $\bar{z}_k^{(i)}$ and the corresponding measurement matrix $\bar{H}_k^{(i)}$ to the fusion center. In all, the communication requirements from each sensor to the fusion center are $r_i + r_i \times n$ at any time instant. This beats centralized and most existing distributed fusion algorithms since $r_i \leq \min(m_i, n)$, which also means that the local raw measurements are compressed before transmission.

Remark: The introduction of full rank decomposition may be computationally demanding at each sensor. For efficient ways to calculate full rank decomposition, see [92].

4.5 Distributed fusion with transformed data

Let

$$\begin{aligned}
 z_k^d &= [(\bar{z}_k^{(1)})', (\bar{z}_k^{(2)})', \dots, (\bar{z}_k^{(N_s)})']' \\
 H_k^d &= [(\bar{H}_k^{(1)})', (\bar{H}_k^{(2)})', \dots, (\bar{H}_k^{(N_s)})']' \\
 v_k^d &= [(\bar{v}_k^{(1)})', (\bar{v}_k^{(2)})', \dots, (\bar{v}_k^{(N_s)})']'
 \end{aligned} \tag{4.14}$$

where for transformation I

$$\begin{aligned}
 E[v_k^d] &= 0_{N_s n \times 1} \\
 R_k^d &= \text{cov}(v_k^d) = \text{diag}(\bar{H}_k^{(1)}, \bar{H}_k^{(2)}, \dots, \bar{H}_k^{(N_s)})
 \end{aligned} \tag{4.15}$$

and for transformation II

$$\begin{aligned}
 E[v_k^d] &= 0_{r \times 1}, \quad r = \sum_{i=1}^{N_s} r_i \\
 R_k^d &= \text{diag}(I_{r_1 \times r_1}, I_{r_2 \times r_2}, \dots, I_{r_{N_s} \times r_{N_s}}) = I_{r \times r}
 \end{aligned} \tag{4.16}$$

Then the stacked measurement equation at the fusion center with respect to (w.r.t.) the transformed data from all N_s sensors can be written as

$$z_k^d = H_k^d x_k + v_k^d$$

Given the previous fused estimate $\hat{x}_{k-1|k-1}^d$, $P_{k-1|k-1}^d$, in the sense of LMMSE [73, 74, 81], the **LMMSE Distributed Fusion with Transformation I** can be computed recursively

as follows:

$$\hat{x}_{k|k-1}^d = F_{k-1} \hat{x}_{k-1|k-1}^d \quad (4.17)$$

$$P_{k|k-1}^d = F_{k-1} P_{k-1|k-1}^d F_{k-1}' + G_{k-1} Q_{k-1} G_{k-1}' \quad (4.18)$$

$$\hat{x}_{k|k}^d = \hat{x}_{k|k-1}^d + K_k^d (z_k^d - H_k^d \hat{x}_{k|k-1}^d) \quad (4.19)$$

$$K_k^d = P_{k|k-1}^d (H_k^d)' (S_k^d)^+$$

$$P_{k|k}^d = P_{k|k-1}^d - P_{k|k-1}^d (H_k^d)' (S_k^d)^+ H_k^d P_{k|k-1}^d \quad (4.20)$$

$$S_k^d = H_k^d P_{k|k-1}^d (H_k^d)' + R_k^d$$

and the **LMMSE Distributed Fusion with Transformation II** is the same as that with transformation I except that

$$\begin{aligned} K_k^d &= P_{k|k-1}^d (H_k^d)' (S_k^d)^{-1} \\ P_{k|k}^d &= P_{k|k-1}^d - P_{k|k-1}^d (H_k^d)' (S_k^d)^{-1} H_k^d P_{k|k-1}^d \end{aligned} \quad (4.21)$$

Remark: It can be seen that the only difference between distributed fusion with transformations I and II is that $(S_k^d)^+$ is used in the former and $(S_k^d)^{-1}$ in the latter. For transformation I, we only know that $\bar{H}_k^{(i)} = (H_k^{(i)})' (R_k^{(i)})^{-1} H_k^{(i)} \geq 0$ and $R_k^d = \text{diag}(\bar{H}_k^{(1)}, \bar{H}_k^{(2)}, \dots, \bar{H}_k^{(M)}) \geq 0$ in general, and that is why we use the MP inverse. While for transformation II, we know that $R_k^d = I_{r \times r} > 0$ for sure, and so we use the inverse.

Remark: For transformation I, in some special cases, we do have $\bar{H}_k^{(i)} > 0$, $i = 1, 2, \dots, N_s$, e.g., when all $H_k^{(i)}$'s are of full column rank. In these cases, $R_k^d > 0$ and $(S_k^d)^+$ reduces to $(S_k^d)^{-1}$.

4.6 Optimality of proposed algorithms

Although the above distributed fusion algorithms are already optimal with the transformed data, we still want to know whether they are equivalent to centralized fusion. For convenience of comparison, centralized fusion is described first.

Let

$$\begin{aligned}
 z_k^c &= [(z_k^{(1)})', (z_k^{(2)})', \dots, (z_k^{(N_s)})']' \\
 H_k^c &= [(H_k^{(1)})', (H_k^{(2)})', \dots, (H_k^{(N_s)})']' \\
 v_k^c &= [(v_k^{(1)})', (v_k^{(2)})', \dots, (v_k^{(N_s)})']'
 \end{aligned} \tag{4.22}$$

where

$$\begin{aligned}
 E[v_k^c] &= 0_{l \times 1}, \quad l = \sum_{i=1}^{N_s} m_i \\
 R_k^c &= \text{cov}(v_k^c) = \text{diag}(R_k^{(1)}, R_k^{(2)}, \dots, R_k^{(N_s)})
 \end{aligned} \tag{4.23}$$

Then the stacked measurement equation at the fusion center w.r.t. the raw measurements from all N_s sensors can be written as

$$z_k^c = H_k^c x_k + v_k^c$$

Given the previous centralized fused estimate $\hat{x}_{k-1|k-1}^c$, $P_{k-1|k-1}^c$, the **LMMSE Central-**

ized **Fusion** can be computed recursively as follows:

$$\hat{x}_{k|k-1}^c = F_{k-1}\hat{x}_{k-1|k-1}^c \quad (4.24)$$

$$P_{k|k-1}^c = F_{k-1}P_{k-1|k-1}^c F_{k-1}' + G_{k-1}Q_{k-1}G_{k-1}' \quad (4.25)$$

$$\hat{x}_{k|k}^c = \hat{x}_{k|k-1}^c + K_k^c(z_k^c - H_k^c\hat{x}_{k|k-1}^c) \quad (4.26)$$

$$K_k^c = P_{k|k-1}^c(H_k^c)'(S_k^c)^{-1}$$

$$P_{k|k}^c = P_{k|k-1}^c - P_{k|k-1}^c(H_k^c)'(S_k^c)^{-1}H_k^cP_{k|k-1}^c \quad (4.27)$$

$$S_k^c = H_k^cP_{k|k-1}^c(H_k^c)' + R_k^c \quad (4.28)$$

Compared with the LMMSE centralized fusion, it is found that the following theorems hold for the above proposed distributed fusion algorithms.

Theorem 1 If $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, then for the LMMSE distributed fusion with transformation I, we have

$$(H_k^d)'(S_k^d)^+H_k^d = (H_k^c)'(S_k^c)^{-1}H_k^c$$

Proof: See the Appendix.

Theorem 2. If $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, then for the LMMSE distributed fusion with transformation II, we have

$$(H_k^d)'(S_k^d)^{-1}H_k^d = (H_k^c)'(S_k^c)^{-1}H_k^c$$

Proof: See the Appendix.

Theorem 3 If $\hat{x}_{k-1|k-1}^d = \hat{x}_{k-1|k-1}^c$ and $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, then the above LMMSE distributed fusion algorithms are globally optimal in the sense that they are equivalent to the centralized one, that is,

$$\hat{x}_{k|k}^d = \hat{x}_{k|k}^c, \quad P_{k|k}^d = P_{k|k}^c$$

Proof: See the Appendix.

Remark: It should be noted that the compression rule for transformation II is not unique for two reasons. First, the full rank decomposition of $H_k^{(i)}$ is not unique and they all lead to the same result (equivalent to centralized fusion). Second, any transformation in the form of

$$C_k^{(i)} = (A_k^{(i)}(M_k^{(i)})'(R_k^{(i)})^{-1}M_k^{(i)}(A_k^{(i)})')^{-1/2}A_k^{(i)}(M_k^{(i)})'(R_k^{(i)})^{-1}$$

is optimal if $A_k^{(i)} \in \mathbb{R}^{r_i \times r_i}$ is an invertible matrix.

In summary, our proposed distributed fusion algorithms have three nice properties:

- They are globally optimal—they are equivalent to centralized fusion.
- In terms of communication requirements from each sensor to the fusion center, fusion with transformation I just requires $n+n^2$, which is the same or better than most existing distributed fusion algorithms; fusion with transformation II just requires $r_i + r_i \times n$, which beats centralized and most existing distributed fusion algorithms.
- They do not need the inverses of error covariance matrices, which are assumed in most existing distributed fusion algorithms but can not be guaranteed to exist, so they can be applied in more cases.

Now let us make a comparison between the fusion algorithms with transformations I and II. Transformation II beats transformation I in communication because $r_i + r_i \times n = r_i(n+1) \leq n(n+1) = n+n^2$. It is hard to say which is better in computation. On the one hand, transformation I is relatively easier, as can be seen from (4.3) since $H_k^{(i)}$ itself is directly involved, while transformation II involves the full rank decomposition of $H_k^{(i)}$ and the inverse square root matrix $((M_k^{(i)})'(R_k^{(i)})^{-1}M_k^{(i)})^{-1/2}$, as can be seen from (4.9). On the other hand, the computation at the fusion center is slightly harder for transformation I due to the involvement of the $N_s n \times N_s n$ MP inverse $(S_k^d)^+$, while transformation II uses the $r \times r$ inverse $(S_k^d)^{-1}$. In sum, transformation I has better local computation but transformation II has better computation at the fusion center.

Remark: After the results of this work had been worked out, we found that a similar idea given in the appendix of [111] to reduce the dimensionality of the raw measurements also uses full rank decomposition of the measurement matrix at each sensor. By comparison, it can be seen that the distributed fusion algorithm of [111] is based on the information form of the Kalman filter and its optimality (equivalence to the centralized fusion) was proved also based on this form, which requires the existence of the inverses of error covariance matrices and this may not be satisfied in more general cases. Also, one goal of this work is to eliminate the transmission of the covariance matrix of the compressed measurement noise to save communication, as was made clear in the above, which is not the case for the algorithm of [111]. Therefore, our distributed fusion with transformation II is more applicable.

4.7 Reduction of computational complexity

The above three nice properties make the proposed algorithms attractive. But for implementation, the computational complexity seems a little bit large due to the involvement of the $N_s n \times N_s n$ MP inverse $(S_k^d)^+$ and $r \times r$ inverse $(S_k^d)^{-1}$ in the two algorithms, respectively.

In the following, we show that in the proposed algorithms, transformed data from different sensors can be sequentially processed without loss of fusion performance. In this way, the computational complexity can be reduced to a certain degree.

Theorem 4 (Sequential LMMSE Distributed Fusion) For the proposed algorithms, the fused estimate can be equivalently computed in a sequential way as follows.

Prediction:

$$\begin{aligned}\hat{x}_{k|k}^{(0)} &= \hat{x}_{k|k-1}^d = F_{k-1} \hat{x}_{k-1|k-1}^d \\ P_{k|k}^{(0)} &= P_{k|k-1}^d = F_{k-1} P_{k-1|k-1}^d F_{k-1}' + G_{k-1} Q_{k-1} G_{k-1}'\end{aligned}$$

Update by sensor i , $i = 1, 2, \dots, N_s$:

$$\begin{aligned}\hat{x}_{k|k}^{(i)} &= \hat{x}_{k|k}^{(i-1)} + K_k^{(i)}(\bar{z}_k^{(i)} - \bar{H}_k^{(i)}\hat{x}_{k|k}^{(i-1)}) \\ P_{k|k}^{(i)} &= P_{k|k}^{(i-1)} - K_k^{(i)}\bar{S}_k^{(i)}(K_k^{(i)})' \\ K_k^{(i)} &= P_{k|k}^{(i-1)}(\bar{H}_k^{(i)})'(\bar{S}_k^{(i)})^+ \\ \bar{S}_k^{(i)} &= \bar{H}_k^{(i)}P_{k|k}^{(i-1)}(\bar{H}_k^{(i)})' + \bar{H}_k^{(i)}\end{aligned}$$

Finally,

$$\hat{x}_{k|k}^d = \hat{x}_{k|k}^{(N_s)}, \quad P_{k|k}^d = P_{k|k}^{(N_s)}$$

Proof: See the Appendix.

Due to sequential processing, the original $N_s n \times N_s n$ MP inverse $(S_k^d)^+$ in transformation I is now replaced by N_s $n \times n$ MP inverses $(\bar{S}_k^{(i)})^+$, $i = 1, 2, \dots, N_s$, and the original $r \times r$ inverse $(S_k^d)^{-1}$ in transformation II is now replaced by N_s $r_i \times r_i$ inverses $(\bar{S}_k^{(i)})^{-1}$, $i = 1, 2, \dots, N_s$. The computational complexity is indeed reduced to some extent.

Remark: The key to our proof of the feasibility of sequential processing is the uncorrelatedness of the transformed measurement noises across sensors. There is a major difference between this proof and the conventional proof, which is based on the information form of the Kalman filter that requires the error covariance matrices and measurement noise covariance matrices to be invertible. Our proof is based on the recursibility of the LMMSE estimation proposed in [73] which is more general in theory.

Remark: In general, performance of the sequential LMMSE estimation depends on the order in which the data are processed and it may differ from that of the batch LMMSE estimation. As is clear from the above, however, our sequential LMMSE distributed fusion is equivalent to the batch LMMSE distributed fusion and the data processing order does not matter at all.

Remark: It is assumed above that all N_s sensors work synchronously. But in practice,

most sensors may work asynchronously [1, 71, 82, 88, 122] due to different sampling rates, different initial times, etc. Following a similar sequential processing idea, the optimal asynchronous distributed fusion algorithm which uses the transformed data in Eq. (4.6) or (4.10) can also be easily obtained. The key to this optimal asynchronous distributed fusion algorithm is to describe the state transition from one arbitrary time instant to another. For more details, the interested reader is referred to [55, 56, 72].

4.8 Extension to reduce-rate communication

In the above, it is assumed that the local sensors have full rate communication with the fusion center. However, in some applications, due to communication constraints, it is more meaningful for the sensors to send in processed data in a reduced rate. In the following, it is assumed that the sensors send their compressed data to the fusion center at every N time instants and then the fusion center fuses correspondingly.

Depending on whether a dynamic system is driven by process noise or not [30] and its state transition matrix (STM) is invertible or not, we can divide dynamic systems into four classes. Due to the increased complexity, only two simpler classes are discussed here. The other two classes are left for future work.

4.8.1 With invertible STM and no process noise

In this case, for $j = 1, 2, \dots, N - 1$,

$$x_{k+j} = \prod_{l=j}^{N-1} F_{k+l}^{-1} x_{k+N}$$

$$z_{k+j}^{(i)} = H_{k+j}^{(i)} \prod_{l=j}^{N-1} F_{k+l}^{-1} x_{k+N} + v_{k+j}^{(i)}$$

Then it follows that

$$z_{k+N}^i = H_{k+N}^i x_{k+N} + v_{k+N}^i$$

where

$$\begin{aligned} z_{k+N}^i &= [(z_{k+N}^{(i)})', (z_{k+N-1}^{(i)})', \dots, (z_{k+1}^{(i)})']', \quad H_{k+N}^i = [(H_{k+N}^{(i)})', \dots, (H_{k+1}^{(i)} \prod_{l=1}^{N-1} F_{k+l}^{-1})']' \\ v_{k+N}^i &= [(v_{k+N}^{(i)})', (v_{k+N-1}^{(i)})', \dots, (v_{k+1}^{(i)})']', \quad R_{k+N}^i = \text{cov}(v_{k+N}^i) = \text{diag}(R_{k+N}^{(i)}, \dots, R_{k+1}^{(i)}) > 0 \end{aligned}$$

For transformation I, at sensor i , its local raw measurements from $k+1$ up to $k+N$ can be transformed losslessly to

$$\bar{z}_{k+N}^i = \bar{H}_{k+N}^i x_{k+N} + \bar{v}_{k+N}^i$$

where

$$\begin{aligned} \bar{z}_{k+N}^i &= T_{k+N}^i z_{k+N}^i, \quad \bar{H}_{k+N}^i = T_{k+N}^i H_{k+N}^i, \quad T_{k+N}^i = (H_{k+N}^i)' (R_{k+N}^i)^{-1} \\ \bar{R}_{k+N}^{(i)} &= \text{cov}(\bar{v}_{k+N}^i) = (H_{k+N}^i)' (R_{k+N}^i)^{-1} H_{k+N}^i \end{aligned}$$

For transformation II, given the full rank decomposition

$$H_{k+N}^i = \check{M}_{k+N}^i \check{N}_{k+N}^i$$

at sensor i , its local raw measurements from $k+1$ up to $k+N$ can be compressed losslessly to

$$\bar{z}_{k+N}^i = \bar{H}_{k+N}^i x_{k+N} + \bar{v}_{k+N}^i$$

where

$$\begin{aligned}\bar{z}_{k+N}^i &= C_{k+N}^i z_{k+N}^i, \quad \bar{H}_{k+N}^i = C_{k+N}^i H_{k+N}^i \\ C_{k+N}^i &= ((\check{M}_{k+N}^i)'(R_{k+N}^i)^{-1}\check{M}_{k+N}^i)^{-1/2}(\check{M}_{k+N}^i)'(R_{k+N}^i)^{-1} \\ \bar{R}_{k+N}^{(i)} &= \text{cov}(\bar{v}_{k+N}^i) = I_{d_i \times d_i}, \quad d_i = \text{rank}(H_{k+N}^i)\end{aligned}$$

Note that in this case, the prediction at the fusion center is N steps ahead:

$$\begin{aligned}\hat{x}_{k+N|k}^d &= \prod_{l=1}^N F_{k+N-l} \hat{x}_{k|k}^d \\ P_{k+N|k}^d &= \prod_{l=1}^N F_{k+N-l} P_{k|k}^d (\prod_{l=1}^N F_{k+N-l})'\end{aligned}$$

Then it can be updated by the locally transformed data \bar{z}_{k+N}^i , $i = 1, 2, \dots, N_s$, to obtain $\hat{x}_{k+N|k+N}^d$ and $P_{k+N|k+N}^d$.

4.8.2 With not necessarily invertible STM and no process noise

In this case, for $j = 1, 2, \dots, N$,

$$\begin{aligned}x_{k+j} &= \prod_{l=1}^j F_{k+j-l} x_k \\ z_{k+j}^{(i)} &= H_{k+j}^{(i)} \prod_{l=1}^j F_{k+j-l} x_k + v_{k+j}^{(i)}\end{aligned}$$

Then it follows that

$$z_k^i = H_k^i x_k + v_k^i$$

where

$$\begin{aligned}z_k^i &= [(z_{k+1}^{(i)})', (z_{k+2}^{(i)})', \dots, (z_{k+N}^{(i)})']', \quad H_k^i = [(H_{k+1}^{(i)} F_k)', \dots, (H_{k+N}^{(i)} \prod_{l=1}^N F_{k+N-l})']' \\ v_k^i &= [(v_{k+1}^{(i)})', (v_{k+2}^{(i)})', \dots, (v_{k+N}^{(i)})']', \quad R_k^i = \text{cov}(v_k^i) = \text{diag}(R_{k+1}^{(i)}, \dots, R_{k+N}^{(i)}) > 0\end{aligned}$$

For transformation I, at sensor i , its local raw measurements from $k + 1$ up to $k + N$ can be transformed losslessly to

$$\bar{z}_k^i = \bar{H}_k^i x_k + \bar{v}_k^i$$

where

$$\bar{z}_k^i = T_k^i z_k^i, \bar{H}_k^i = T_k^i H_k^i, T_k^i = (H_k^i)'(R_k^i)^{-1}, \bar{R}_k^{(i)} = \text{cov}(\bar{v}_k^{(i)}) = (H_k^i)'(R_k^i)^{-1}H_k^i$$

For transformation II, given the full rank decomposition

$$H_k^i = \check{M}_k^i \check{N}_k^i$$

at sensor i , its local raw measurements from $k + 1$ up to $k + N$ can be compressed losslessly to

$$\bar{z}_k^i = \bar{H}_k^i x_k + \bar{v}_k^i$$

where

$$\begin{aligned} \bar{z}_k^i &= C_k^i z_k^i, \bar{H}_k^i = C_k^i H_k^i, C_k^i = ((\check{M}_k^i)'(R_k^i)^{-1}\check{M}_k^i)^{-1/2}(\check{M}_k^i)'(R_k^i)^{-1} \\ \bar{R}_k^{(i)} &= \text{cov}(\bar{v}_k^{(i)}) = I_{d_i \times d_i}, d_i = \text{rank}(H_k^i) \end{aligned}$$

In this case, $\hat{x}_{k|k}^d$ and $P_{k|k}^d$ are first updated by the locally transformed data \bar{z}_k^i , $i = 1, 2, \dots, N_s$, to obtain the smoothed estimate $\hat{x}_{k|k+N}^d$ and $P_{k|k+N}^d$, and then:

$$\begin{aligned} \hat{x}_{k+N|k+N}^d &= \prod_{l=1}^N F_{k+N-l} \hat{x}_{k|k+N}^d \\ P_{k+N|k+N}^d &= \prod_{l=1}^N F_{k+N-l} P_{k|k+N}^d \left(\prod_{l=1}^N F_{k+N-l} \right)' \end{aligned}$$

Remark: For the cases with process noise, the problem becomes much harder due to the

correlation between common process noise if we still try to align the measurements obtained at different time instants w.r.t. the state vector at one fixed time instant as done above.

4.9 Extension to singular measurement noise case

In the above, it is assumed that $R_k^{(i)} > 0$, $i = 1, 2, \dots, N_s$, which may limit the application of the proposed algorithms. We now extend it to the singular case, i.e., $|R_k^{(i)}| = 0$.

If $|R_k^{(i)}| = 0$ and suppose $\text{rank}(R_k^{(i)}) = a_i < m_i$, then it follows from the singular value decomposition that there must exist a unitary matrix $U_k^{(i)}$ such that

$$U_k^{(i)} R_k^{(i)} (U_k^{(i)})' = \begin{bmatrix} \bar{R}_{k,1}^{(i)} & 0_{a_i \times b_i} \\ 0_{b_i \times a_i} & 0_{b_i \times b_i} \end{bmatrix}$$

where $b_i = m_i - a_i$ and $\bar{R}_{k,1}^{(i)} > 0$ is an $a_i \times a_i$ diagonal matrix.

Let

$$\bar{z}_k^{(i)} = U_k^{(i)} z_k^{(i)}$$

Then from Eq. (4.2), it follows that

$$\bar{z}_k^{(i)} = U_k^{(i)} H_k^{(i)} x_k + U_k^{(i)} v_k^{(i)} = \bar{H}_k^{(i)} x_k + \bar{v}_k^{(i)}$$

where

$$\bar{z}_k^{(i)} = [(\bar{z}_{k,1}^{(i)})', (\bar{z}_{k,2}^{(i)})']', \quad \bar{H}_k^{(i)} = [(\bar{H}_{k,1}^{(i)})', (\bar{H}_{k,2}^{(i)})']' = U_k^{(i)} H_k^{(i)}, \quad \bar{v}_k^{(i)} = [(\bar{v}_{k,1}^{(i)})', (\bar{v}_{k,2}^{(i)})']' = U_k^{(i)} v_k^{(i)}$$

$$\bar{z}_{k,1}^{(i)} \in \mathbb{R}^{a_i}, \quad \bar{z}_{k,2}^{(i)} \in \mathbb{R}^{b_i}, \quad \bar{H}_{k,1}^{(i)} \in \mathbb{R}^{a_i \times n}, \quad \bar{H}_{k,2}^{(i)} \in \mathbb{R}^{b_i \times n}, \quad \bar{v}_{k,1}^{(i)} \in \mathbb{R}^{a_i}, \quad \bar{v}_{k,2}^{(i)} \in \mathbb{R}^{b_i}$$

$$\text{cov}(\bar{v}_{k,1}^{(i)}) = \bar{R}_{k,1}^{(i)}, \quad \text{cov}(\bar{v}_{k,1}^{(i)}, \bar{v}_{k,2}^{(i)}) = 0, \quad \bar{v}_{k,2}^{(i)} = 0 \text{ a.s.}$$

Since $U_k^{(i)}$ is unitary and thus invertible, $\bar{z}_k^{(i)} = U_k^{(i)} z_k^{(i)}$ is sufficient in that the LMMSE

estimation based on $z_k^{(i)}$ is equivalent to the one based on $\bar{z}_k^{(i)}$. That is, the original measurement equation (4.2) is equivalent to the following:

$$\begin{aligned}\bar{z}_{k,1}^{(i)} &= \bar{H}_{k,1}^{(i)}x_k + \bar{v}_{k,1}^{(i)} \\ \bar{z}_{k,2}^{(i)} &= \bar{H}_{k,2}^{(i)}x_k\end{aligned}$$

Here, the noisy measurement $\bar{z}_{k,1}^{(i)} = \bar{H}_{k,1}^{(i)}x_k + \bar{v}_{k,1}^{(i)}$ can be losslessly transformed to an n -dimensional new measurement in transformation I or $\text{rank}(\bar{H}_{k,1}^{(i)})$ -dimensional new measurement in transformation II. The noise-free measurement $\bar{z}_{k,2}^{(i)} = \bar{H}_{k,2}^{(i)}x_k$, by the following theorem, can also be compressed losslessly. One disadvantage of this extension is that we need to use MP inverse in general, instead of just matrix inverse, especially to handle the noise-free part.

Theorem 5. The noise-free measurement $\bar{z}_{k,2}^{(i)} = \bar{H}_{k,2}^{(i)}x_k$ can be compressed losslessly to a $\text{rank}(\bar{H}_{k,2}^{(i)})$ -dimensional new measurement by simply selecting a maximal set of linearly independent row vectors of $\bar{H}_{k,2}^{(i)}$.

Proof: See the Appendix.

4.10 Summary

By taking linear transformation of the raw measurement at each sensor, two optimal distributed fusion algorithms have been proposed. Three nice properties make them attractive. First, they are globally optimal in that they are equivalent to centralized fusion. Second, in terms of communication requirements from each sensor to the fusion center, distributed fusion with transformation I is the same or better than most existing distributed fusion algorithms, and distributed fusion with transformation II is better than centralized and most existing distributed fusion algorithms. Third, they do not need the inverses of error covariance matrices, which are assumed in most existing distributed fusion algorithms but can not

be guaranteed to exist, so they can be applied in more general cases. Pros and cons of the two new algorithms have been analyzed. Sequential processing of the transformed data to reduce the computational complexity of the proposed algorithms has been discussed. Extensions to reduced-rate communication for some simpler cases and to the singular measurement noise case have also been discussed.

4.11 Appendix

4.11.1 Proof of Theorem 1

Since $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, it follows from Eqs. (4.18) and (4.25) that $P_{k|k-1}^d = P_{k|k-1}^c$.

Let

$$\check{H}_k = \text{diag}(H_k^{(1)}, H_k^{(2)}, \dots, H_k^{(N_s)})$$

Then from Eqs. (4.15) and (4.5), we have

$$R_k^d = \text{diag}(\bar{H}_k^{(1)}, \bar{H}_k^{(2)}, \dots, \bar{H}_k^{(N_s)}) = \check{H}_k'(R_k^c)^{-1}\check{H}_k$$

Furthermore, it follows from Eqs. (4.14) and (4.5) that

$$H_k^d = [\bar{H}_k^{(1)}, \bar{H}_k^{(2)}, \dots, \bar{H}_k^{(M)}]' = \check{H}_k'(R_k^c)^{-1}H_k^c$$

Then

$$\begin{aligned} & (H_k^d)'(S_k^d)^+H_k^d \\ &= (H_k^c)'(R_k^c)^{-1}\check{H}_k(\check{H}_k'(R_k^c)^{-1}H_k^cP_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}\check{H}_k + \check{H}_k'(R_k^c)^{-1}\check{H}_k)^+ \check{H}_k'(R_k^c)^{-1}H_k^c \\ &= (H_k^c)'(R_k^c)^{-1}\check{H}_k(\check{H}_k'(R_k^c)^{-1}H_k^cP_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}\check{H}_k + \check{H}_k'(R_k^c)^{-1}R_k^c(R_k^c)^{-1}\check{H}_k)^+ \check{H}_k'(R_k^c)^{-1}H_k^c \\ &= (H_k^c)'(R_k^c)^{-1}\check{H}_k(\check{H}_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}\check{H}_k)^+ \check{H}_k'(R_k^c)^{-1}H_k^c \end{aligned}$$

It can be easily seen that

$$(H_k^c)'(R_k^c)^{-1}\check{H}_k = (H_k^c)'(S_k^c)^{-1}S_k^c(R_k^c)^{-1}\check{H}_k$$

Also, from matrix inversion lemma [74], we have

$$(S_k^c)^{-1} = (R_k^c)^{-1} - (R_k^c)^{-1}H_k^cU_k^{-1}P_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}$$

where

$$U_k = P_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}H_k^c + I$$

Thus

$$(H_k^c)'(S_k^c)^{-1} = (H_k^c)'(R_k^c)^{-1} - (H_k^c)'(R_k^c)^{-1}H_k^cU_k^{-1}P_{k|k-1}^d(H_k^c)'(R_k^c)^{-1} = A_k(H_k^c)'(R_k^c)^{-1}$$

and

$$(H_k^c)'(R_k^c)^{-1}\check{H}_k = A_k(H_k^c)'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}\check{H}_k$$

where

$$A_k = I - (H_k^c)'(R_k^c)^{-1}H_k^cU_k^{-1}P_{k|k-1}^d$$

Note that

$$(H_k^c)' = [(H_k^{(1)})', (H_k^{(2)})', \dots, (H_k^{(N_s)})'] = \mathbf{I}_{N_s}\check{H}_k'$$

where

$$\mathbf{I}_{N_s} = [I_{n \times n}, I_{n \times n}, \dots, I_{n \times n}]$$

Thus

$$(H_k^c)'(R_k^c)^{-1}\check{H}_k = A_k\mathbf{I}_{N_s}\check{H}_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}\check{H}_k$$

Taking transpose on both sides, we have

$$\check{H}'_k(R_k^c)^{-1}H_k^c = \check{H}'_k(R_k^c)^{-1}S_k^c(R_k^c)^{-1}\check{H}_k\mathbf{I}'_{N_s}B_k$$

where

$$V_k = (H_k^c)'(R_k^c)^{-1}H_k^c P_{k|k-1}^d + I, \quad B_k = I - P_{k|k-1}^d V_k^{-1}(H_k^c)'(R_k^c)^{-1}H_k^c$$

Then

$$\begin{aligned} & (H_k^d)'(S_k^d)^+ H_k^d \\ &= A_k \mathbf{I}_{N_s} \check{H}'_k(R_k^c)^{-1} S_k^c(R_k^c)^{-1} \check{H}_k (\check{H}_k^T(R_k^c)^{-1} S_k^c(R_k^c)^{-1} \check{H}_k)^+ \check{H}'_k(R_k^c)^{-1} S_k^c(R_k^c)^{-1} \check{H}_k \mathbf{I}'_{N_s} B_k \\ &= A_k \mathbf{I}_{N_s} \check{H}'_k(R_k^c)^{-1} S_k^c(R_k^c)^{-1} \check{H}_k \mathbf{I}'_{N_s} B_k \\ &= A_k (H_k^c)'(R_k^c)^{-1} S_k^c(R_k^c)^{-1} H_k^c B_k \\ &= (H_k^c)'(S_k^c)^{-1} S_k^c(S_k^c)^{-1} H_k^c \\ &= (H_k^c)'(S_k^c)^{-1} H_k^c \end{aligned}$$

This completes the proof. □

4.11.2 Proof of Theorem 2

Since $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, from Eqs. (4.18) and (4.25), it follows that $P_{k|k-1}^d = P_{k|k-1}^c$.

Let

$$T_k = \text{diag}(C_k^{(1)}, C_k^{(2)}, \dots, C_k^{(N_s)})$$

Then it follows from Eqs. (4.16), (4.12), (4.23), (4.14) and (4.11) that

$$R_k^d = I_{r \times r} = T_k R_k^c T_k', \quad H_k^d = T_k H_k^c$$

Thus

$$(H_k^d)'(S_k^d)^{-1}H_k^d = (H_k^c)'T_k'(T_kH_k^cP_{k|k-1}^d(H_k^c)'T_k' + T_kR_k^cT_k')^{-1}T_kH_k^c = (H_k^c)'T_k'(T_kS_k^cT_k')^{-1}T_kH_k^c$$

Furthermore, let

$$M_k = \text{diag}(M_k^{(1)}, M_k^{(2)}, \dots, M_k^{(N_s)})$$

$$L_k = M_k'(R_k^c)^{-1}M_k$$

Then

$$T_k = L_k^{-1/2}M_k'(R_k^c)^{-1}$$

$$\begin{aligned} & (H_k^d)'(S_k^d)^{-1}H_k^d \\ &= (H_k^c)'(R_k^c)^{-1}M_kL_k^{-1/2}(L_k^{-1/2}M_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_kL_k^{-1/2})^{-1}L_k^{-1/2}M_k'(R_k^c)^{-1}H_k^c \\ &= (H_k^c)'(R_k^c)^{-1}M_k(M_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k)^{-1}M_k'(R_k^c)^{-1}H_k^c \end{aligned}$$

It can be easily seen that

$$(H_k^c)'(R_k^c)^{-1}M_k = (H_k^c)'(S_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k$$

Also, from Eq. (4.28) and matrix inversion lemma, we have

$$(S_k^c)^{-1} = (R_k^c)^{-1} - (R_k^c)^{-1}H_k^cU_k^{-1}P_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}$$

where

$$U_k = P_{k|k-1}^d(H_k^c)'(R_k^c)^{-1}H_k^c + I$$

Thus

$$(H_k^c)'(S_k^c)^{-1} = A_k(H_k^c)'(R_k^c)^{-1}$$

and

$$(H_k^c)'(R_k^c)^{-1}M_k = A_k(H_k^c)'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k$$

where

$$A_k = I - (H_k^c)'(R_k^c)^{-1}H_k^cU_k^{-1}P_{k|k-1}^d$$

Note that

$$(H_k^c)' = \check{N}_k M_k'$$

where

$$\check{N}_k = [(N_k^{(1)})', (N_k^{(2)})', \dots, (N_k^{(N_s)})']'$$

Thus

$$(H_k^c)'(R_k^c)^{-1}M_k = A_k \check{N}_k M_k' (R_k^c)^{-1} S_k^c (R_k^c)^{-1} M_k$$

Taking transpose on both sides, we have

$$M_k'(R_k^c)^{-1}H_k^c = M_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k\check{N}_k' B_k$$

where

$$V_k = (H_k^c)'(R_k^c)^{-1}H_k^cP_{k|k-1}^d + I, \quad B_k = I - P_{k|k-1}^d V_k^{-1}(H_k^c)'(R_k^c)^{-1}H_k^c$$

Then

$$\begin{aligned}
& (H_k^d)'(S_k^d)^{-1}H_k^d \\
&= A_k\check{N}_kM_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k(M_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k)^{-1}M_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k\check{N}_k'B_k \\
&= A_k\check{N}_kM_k'(R_k^c)^{-1}S_k^c(R_k^c)^{-1}M_k\check{N}_k'B_k \\
&= (H_k^c)'(S_k^c)^{-1}S_k^c(S_k^c)^{-1}H_k^c \\
&= (H_k^c)'(S_k^c)^{-1}H_k^c
\end{aligned}$$

This completes the proof. □

4.11.3 Proof of Theorem 3

Since $\hat{x}_{k-1|k-1}^d = \hat{x}_{k-1|k-1}^c$, from Eqs. (4.17) and (4.24), we have $\hat{x}_{k|k-1}^d = \hat{x}_{k|k-1}^c$. Also, since $P_{k-1|k-1}^d = P_{k-1|k-1}^c$, from Eqs. (4.18) and (4.25), it follows that $P_{k|k-1}^d = P_{k|k-1}^c$ and Theorems 1 and 2 hold. Thus from Eqs. (4.20) and (4.27), we have $P_{k|k}^d = P_{k|k}^c$ for distributed fusion with transformation I; and from Eqs. (4.21) and (4.27), we have $P_{k|k}^d = P_{k|k}^c$ for distributed fusion with transformation II. Finally, from the almost sure uniqueness of the LMMSE estimators that two LMMSE estimators of the same quantity to be estimated using the same set of data are almost surely identical if and only if their MSE matrices are equal [79], we also have $\hat{x}_{k|k}^d = \hat{x}_{k|k}^c$. This completes the proof. □

4.11.4 Proof of Theorem 4

In the sense of LMMSE, it is easy to get the predictor of x_k at the fusion center as

$$\begin{aligned}\hat{x}_{k|k}^{(0)} &= \hat{x}_{k|k-1}^d = E^*[x_k|z_1^d, z_2^d, \dots, z_{k-1}^d] = F_{k-1}\hat{x}_{k-1|k-1}^d \\ P_{k|k}^{(0)} &= P_{k|k-1}^d = \text{MSE}(\hat{x}_{k|k-1}^d) = F_{k-1}P_{k-1|k-1}^d F_{k-1}' + G_{k-1}Q_{k-1}G_{k-1}'\end{aligned}$$

Given $\bar{z}_k^{(1)}$, the updated LMMSE estimator of x_k is

$$\begin{aligned}\hat{x}_{k|k}^{(1)} &= E^*[x_k|z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^{(1)}] = \hat{x}_{k|k-1}^d + P_{k|k-1}^d(\bar{H}_k^{(1)})'(\bar{S}_k^{(1)})^+(\bar{z}_k^{(1)} - \bar{H}_k^{(1)}\hat{x}_{k|k-1}^d) \\ &= \hat{x}_{k|k}^{(0)} + K_k^{(1)}(\bar{z}_k^{(1)} - \bar{H}_k^{(1)}\hat{x}_{k|k}^{(0)}) \\ P_{k|k}^{(1)} &= \text{MSE}(\hat{x}_{k|k}^{(1)}) = P_{k|k-1}^d - P_{k|k-1}^d(\bar{H}_k^{(1)})'(\bar{S}_k^{(1)})^+\bar{H}_k^{(1)}P_{k|k-1}^d = P_{k|k}^{(0)} - K_k^{(1)}\bar{S}_k^{(1)}(K_k^{(1)})'\end{aligned}$$

Let

$$\bar{z}_k^i = \{\bar{z}_k^{(1)}, \bar{z}_k^{(2)}, \dots, \bar{z}_k^{(i-1)}, \bar{z}_k^{(i)}\} = \{\bar{z}_k^{i-1}, \bar{z}_k^{(i)}\}$$

for $i = 2, 3, \dots, N_s$. Since the LMMSE estimator $E^*[x_k|z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^i]$ always has the *quasi-recursive form* [73], we have

$$\begin{aligned}\hat{x}_{k|k}^{(i)} &= E^*[x_k|z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^i] = E^*[x_k|z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^{i-1}, \bar{z}_k^{(i)}] \\ &= \hat{x}_{k|k}^{(i-1)} + C_{i-1,i}C_{\bar{z}_k^{i-1}}^+ \bar{z}_{i|i-1}^*\end{aligned}$$

where

$$\begin{aligned}\tilde{z}_{i|i-1}^* &= \bar{z}_k^{(i)} - E^*[\bar{z}_k^{(i)} | z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^{i-1}] = \bar{z}_k^{(i)} - E^*[\bar{H}_k^{(i)} x_k | z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^{i-1}] \\ &= \bar{z}_k^{(i)} - \bar{H}_k^{(i)} \hat{x}_{k|k}^{(i-1)} = \bar{H}_k^{(i)} (x_k - \hat{x}_{k|k}^{(i-1)}) + \bar{v}_k^{(i)}\end{aligned}$$

$$C_{\tilde{z}_{i|i-1}^*} = \text{cov}(\tilde{z}_{i|i-1}^*) = \bar{S}_k^{(i)}$$

$$C_{i-1,i} = \text{cov}(\tilde{x}_{k|k}^{(i-1)}, \tilde{z}_{i|i-1}^*) = P_{k|k}^{(i-1)} (\bar{H}_k^{(i)})'$$

that is,

$$\hat{x}_{k|k}^{(i)} = E^*[x_k | z_1^d, z_2^d, \dots, z_{k-1}^d, \bar{z}_k^i] = \hat{x}_{k|k}^{(i-1)} + K_k^{(i)} (\bar{z}_k^{(i)} - \bar{H}_k^{(i)} \hat{x}_{k|k}^{(i-1)})$$

Also,

$$P_{k|k}^{(i)} = \text{MSE}(\hat{x}_{k|k}^{(i)}) = \text{MSE}(\hat{x}_{k|k}^{(i-1)}) - C_{i-1,i} C_{\tilde{z}_{i|i-1}^*}^+ C'_{i-1,i} = P_{k|k}^{(i-1)} - K_k^{(i)} \bar{S}_k^{(i)} (K_k^{(i)})'$$

This is actually also the recursive LMMSE estimator with transformed data \bar{z}_k^{i-1} and $\bar{z}_k^{(i)}$ since $\hat{x}_{k|k}^{(i)}$ depends on \bar{z}_k^{i-1} only through $\hat{x}_{k|k}^{(i-1)}$.

Repeating the same procedure until the transformed data from sensor N_s is also used, we have

$$\begin{aligned}\hat{x}_{k|k}^d &= \hat{x}_{k|k}^{(N_s)} = \hat{x}_{k|k}^{(N_s-1)} + K_k^{(N_s)} (\bar{z}_k^{(N_s)} - \bar{H}_k^{(N_s)} \hat{x}_{k|k}^{(N_s-1)}) \\ P_{k|k}^d &= P_{k|k}^{(N_s)} = P_{k|k}^{(N_s-1)} - K_k^{(N_s)} \bar{S}_k^{(N_s)} (K_k^{(N_s)})'\end{aligned}$$

This completes the proof. □

4.11.5 Proof of Theorem 5

Premultiplying a series of elementary row transformation matrices to $\bar{H}_{k,2}^{(i)}$ that the final transformed measurement matrix is only composed of a maximal set of linearly independent row vectors of $\bar{H}_{k,2}^{(i)}$ and zero row vectors. Since elementary row transformation matrices are invertible, the LMMSE estimation based on $\bar{z}_{k,2}^{(i)}$ must be equivalent to the LMMSE estimation based on the final transformed measurement. Furthermore, since except a maximal set of linearly independent row vectors of $\bar{H}_{k,2}^{(i)}$, all the remaining rows of the final transformed measurement matrix are zero row vectors and the final transformed measurement is noise-free, premultiplying a series of those aforementioned elementary row transformation matrices to $\bar{H}_{k,2}^{(i)}$ is equivalent to simply selecting a maximal set of linearly independent row vectors of $\bar{H}_{k,2}^{(i)}$. From linear algebra, we know that the rank of $\bar{H}_{k,2}^{(i)}$ is also its row rank, which is defined to be the maximal number of linearly independent rows of $\bar{H}_{k,2}^{(i)}$. So the final transformed measurement is $\text{rank}(\bar{H}_{k,2}^{(i)})$ -dimensional.

□

Chapter 5

Optimal State Estimation in the Presence of Multiple Packet Dropouts

5.1 Introduction and related research

With the emerging of sensor networks, traditional estimation problems are facing new challenges. For example, due to unreliable communication between local sensors and the processing center, packet transmission delay [4, 86, 112, 115, 129] and multiple packet dropouts [98–100, 113, 114] are usually inevitable.

In this chapter, we deal with state estimation in the presence of multiple packet dropouts. As in [98–100, 113, 114], by multiple packet dropouts, it is meant that the received data is either the current raw measurement or the last received data in a probabilistic manner. If the time stamp for the corresponding raw measurement is available for each received packet, then this estimation problem reduces to one with intermittent measurements [109] or missing data [74] by comparing the time stamp of the received packet with the current time. Correspondingly, existing methods can be applied directly. What may make the problem harder is that the time-stamp information may not be available in some cases (see the formulation in [98–100, 113, 114]) and what is known to the estimator is only the data arrival probability.

In [98–100], in order to have a similar form as the system used by the Kalman filter, the original system with multiple packet dropouts is converted to a stochastic parameter system by augmentation. Then by defining the stochastic \mathcal{H}_2 -norm of the system with stochastic parameter, the optimal \mathcal{H}_2 filters were designed through the linear matrix inequality approach. In [114], by the innovations approach, the LMMSE estimation in the presence of multiple packet dropouts was obtained completely from the same stochastic parameter system as in [98–100]. In [113], for a time-invariant system, by designating the optimal estimator to be of a specific linear form and from the unbiasedness and minimum error covariance properties of the optimal estimator, the associated coefficient matrices were obtained. This sequential LMMSE estimator is based partially on the same stochastic parameter system as in [98–100]. The characterization of its measurement residual part is not clear. The relying on a higher dimensional stochastic parameter system increases the computational load of the estimators of [114] and [113]. Also, both of their optimal initializations depend on information about the past two pieces of received data before the first physically received data. This dependency seems to be too much for real implementation.

To overcome the shortcomings of the existing LMMSE estimation methods with multiple packet dropouts, an alternative form is derived first in this work. Then under a Gaussian assumption, the MMSE estimation is also derived by hard decision, which without the Gaussian assumption is LMMSE w.r.t. the raw measurement sequence without the unrarrived packets. This MMSE estimation method has two nice properties. First, unlike the proposed LMMSE estimation method, its performance does not depend on the value of the initial data at the processing center. Second, it does not depend on the probability of the data arrival events. Numerical examples are provided to compare performance of the proposed estimators.

This chapter is organized as follows. Sec. 5.2 formulates the problem. Sec. 5.3 summarizes two existing forms of LMMSE estimation in the presence of multiple packet dropouts. Sec. 5.4 derives an alternative form of the LMMSE estimation. Sec. 5.5 discusses the MMSE

estimation and its further relaxation. Sec. 5.6 provides numerical examples to compare performance of the proposed estimators. Sec. 5.7 gives summary.

5.2 Problem formulation

Consider the following generic dynamic system:

$$x_k = F_{k-1}x_{k-1} + G_{k-1}w_{k-1} \quad (5.1)$$

where $x_k \in \mathcal{R}^n$, $\langle w_k \rangle$ is zero-mean white noise with covariance $Q_k \geq 0$, $E[x_0] = \bar{x}_0$, $\text{cov}(x_0) = P_0$, and x_0 is uncorrelated with $\langle w_k \rangle$.

The raw measurement observed by a generic local sensor is given by

$$z_k = H_k x_k + v_k, \quad k = 1, 2, \dots \quad (5.2)$$

where $z_k \in \mathcal{R}^m$, $\langle v_k \rangle$ is zero-mean white noise with covariance $R_k \geq 0$, and $\langle v_k \rangle$ is uncorrelated with x_0 and $\langle w_k \rangle$.

Instead of estimating the state locally by using $\langle z_k \rangle$ directly, they are transmitted to an estimator through an unreliable network where packet dropouts are possible. It is assumed that the data received by the estimator can be modeled by:

$$y_k = \gamma_k z_k + (1 - \gamma_k)y_{k-1} \quad (5.3)$$

where γ_k has Bernoulli distribution with $P\{\gamma_k = 1\} = p_k$ and $P\{\gamma_k = 0\} = 1 - p_k = q_k$, γ_k is uncorrelated with all the other random variables, and $\langle \gamma_k \rangle$ is a white sequence.

Remark: The data arrival probability p_k is a measure of the reliability and transmission quality of the network. In [98–100,113,114], it was assumed that p_k is known to the estimator.

But as shown later, there is no need to make this assumption for some estimators.

Remark: The assumption that γ_k is uncorrelated with all the other random variables seems reasonable for many situations. But as shown later, whether this assumption is valid does not matter for some estimators.

Remark: It is assumed that y_{k-1} is still available when y_k is received. This is also the case in [98–100, 113, 114].

In this work, given only the first two moments, we first try to obtain the LMMSE state estimation in the presence of multiple packet dropouts. That is,

$$\hat{x}_{k|k}^{\text{LMMSE}} \triangleq E^*[x_k|y^k] = \arg \min_{\hat{x}_{k|k}=a_k+B_k Y_k} \text{MSE}(\hat{x}_{k|k})$$

where

$$y^k = \{y_1, \dots, y_k\}, Y_k = [y'_1, \dots, y'_k]', \text{MSE}(\hat{x}_{k|k}) = E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})']$$

and a_k, B_k do not depend on Y_k .

Also we try to obtain the MMSE state estimation in the presence of multiple packet dropouts

$$\hat{x}_{k|k}^{\text{MMSE}} = E[x_k|y^k]$$

for some special cases.

5.3 Summary of two existing forms of LMMSE estimation [113, 114]

First, in order to have a similar form as the system used by the Kalman filter, the original system (5.1), (5.2) and (5.3) is converted to the following stochastic parameter system by

augmentation:

$$X_{k+1} = \tilde{F}_k X_k + \tilde{G}_k W_k \quad (5.4)$$

$$y_k = \tilde{H}_k X_k + \gamma_k v_k \quad (5.5)$$

where

$$X_k = [x'_k \quad y'_{k-1}]' \in \mathcal{R}^{n+m}, \quad W_k = [w'_k \quad v'_k]', \quad \tilde{F}_k = \begin{bmatrix} F_k & 0 \\ \gamma_k H_k & (1 - \gamma_k) I_m \end{bmatrix}$$

$$\tilde{G}_k = \begin{bmatrix} G_k & 0 \\ 0 & \gamma_k I_m \end{bmatrix}, \quad \tilde{H}_k = [\gamma_k H_k \quad (1 - \gamma_k) I_m]$$

Under the given assumption and from (5.4), it follows that

$$c_{k+1} = \bar{F}_k c_k \bar{F}'_k + p_k q_k \begin{bmatrix} 0 & 0 \\ H_k & -I_m \end{bmatrix} c_k \begin{bmatrix} 0 & 0 \\ H_k & -I_m \end{bmatrix}' + U_k \quad (5.6)$$

where

$$c_k = E[X_k X'_k], \quad \bar{F}_k = E[\tilde{F}_k] = \begin{bmatrix} F_k & 0 \\ p_k H_k & q_k I_m \end{bmatrix}, \quad U_k = \begin{bmatrix} G_k Q_k G'_k & 0 \\ 0 & p_k R_k \end{bmatrix}$$

and it is assumed that

$$c_0 = \begin{bmatrix} P_0 + \bar{x}_0 \bar{x}'_0 & 0 \\ 0 & 0 \end{bmatrix}$$

In [114], for the augmented stochastic parameter system (5.4) and (5.5), the sequential LMMSE estimate of X_{k+1} was obtained as follows through the innovations approach.

Prediction:

$$\hat{X}_{k+1|k} = \bar{F}_k \hat{X}_{k|k}$$

$$\text{MSE}(\hat{X}_{k+1|k}) = p_k q_k \begin{bmatrix} 0 & 0 \\ H_k & -I_m \end{bmatrix} c_k \begin{bmatrix} 0 & 0 \\ H_k & -I_m \end{bmatrix}' + \bar{F}_k \text{MSE}(\hat{X}_{k|k}) \bar{F}_k' + U_k$$

Update:

$$\hat{X}_{k+1|k+1} = \hat{X}_{k+1|k} + K_{k+1} \varepsilon_{k+1}$$

$$\varepsilon_{k+1} = y_{k+1} - \bar{H}_{k+1} \hat{X}_{k+1|k}$$

$$K_{k+1} = \text{MSE}(\hat{X}_{k+1|k}) \bar{H}_{k+1}' S_{k+1}^{-1}$$

$$\bar{H}_{k+1} = E[\tilde{H}_{k+1}] = [p_{k+1} H_{k+1} \quad q_{k+1} I_m]$$

$$S_{k+1} = p_{k+1} q_{k+1} [H_{k+1} \quad -I_m] c_{k+1} [H_{k+1} \quad -I_m]' + \bar{H}_{k+1} \text{MSE}(\hat{X}_{k+1|k}) \bar{H}_{k+1}' + p_{k+1} R_{k+1}$$

$$\text{MSE}(\hat{X}_{k+1|k+1}) = \text{MSE}(\hat{X}_{k+1|k}) - K_{k+1} S_{k+1}^{-1} K_{k+1}'$$

with the initial conditions

$$\hat{X}_{0|-1} = [\bar{x}'_0 \quad 0]', \quad \text{MSE}(\hat{X}_{0|-1}) = \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix}$$

Finally,

$$\hat{x}_{k+1|k} = [I_n \quad 0] \hat{X}_{k+1|k}, \quad \hat{x}_{k+1|k+1} = [I_n \quad 0] \hat{X}_{k+1|k+1}$$

$$P_{k+1|k} = [I_n \quad 0] \text{MSE}(\hat{X}_{k+1|k}) [I_n \quad 0]', \quad P_{k+1|k+1} = [I_n \quad 0] \text{MSE}(\hat{X}_{k+1|k+1}) [I_n \quad 0]'$$

In [113], it was further assumed that

$$F_k = F, G_k = G, H_k = H, Q_k = Q, R_k = R, p_k = p, q_k = q = 1 - p$$

By designating the optimal of x_{k+1} to be of the form

$$\hat{x}_{k+1|k+1} = \Phi_k \hat{x}_{k|k} + K_k^1 y_{k+1} + K_k^2 y_k \quad (5.7)$$

and from the unbiasedness and minimum error covariance properties of $\hat{x}_{k+1|k+1}$, it was found that

$$\Phi_k = (I_n - pK_k^1 H)F$$

$$K_k^1 = P_{k+1|k} H' \Lambda_k^{-1}$$

$$K_k^2 = -qK_k^1$$

$$P_{k+1|k} = FP_{k|k}F' + GQG'$$

$$\begin{aligned} \Lambda_k &= p(HP_{k+1|k}H' + R) + qHGG'G'H' + q \begin{bmatrix} pH - HF & qI_m \end{bmatrix} c_k \begin{bmatrix} pH - HF & qI_m \end{bmatrix}' \\ &\quad + pq^2 \begin{bmatrix} H & -I_m \end{bmatrix} c_k \begin{bmatrix} H & -I_m \end{bmatrix}' + (1 - p^2)R \end{aligned}$$

$$P_{k+1|k+1} = P_{k+1|k} - pK_k^1 \Lambda_k (K_k^1)'$$

The one-step ahead prediction is

$$\hat{x}_{k+1|k} = F\hat{x}_{k|k}$$

with the initial conditions

$$\hat{x}_{0|0} = \bar{x}_0, P_{0|0} = P_0$$

Remark: As can be clearly seen from the above, the LMMSE estimator of [114] is purely and the LMMSE estimator of [113] is partially based on the augmented stochastic parameter

form (5.4) and (5.5) of the original system. Compared with the original system described by (5.1) – (5.3), the dimension of the state vector is increased from n to $n + m$, so the computational complexity is increased also.

Remark: The final form (5.7) of the LMMSE estimator in [113] is not similar to that of the Kalman filter, so it is not easy to analyze the measurement residual of this estimator. For example, how big the measurement residual is and what about its statistical characteristics are not clear. As we know, information about the measurement residual is useful for some hypothesis testing problem, e.g., data association in target tracking.

Remark: To initialize the LMMSE estimator in [113], (besides $\hat{x}_{0|0} = \bar{x}_0$ and $P_{0|0} = P_0$) y_0 as well as $E[x_0 y'_{-1}]$ and $E[y_{-1} y'_{-1}]$, both assumed zero therein, is needed. To initialize the LMMSE estimator in [114], (besides $\hat{x}_{0|-1} = \bar{x}_0$ and $P_{0|-1} = P_0$) $E[x_0 y'_{-1}]$, $E[y_{-1} y'_{-1}]$, $\hat{y}_{-1|-1}$, $\text{MSE}(\hat{y}_{-1|-1})$, $E[(x_0 - \hat{x}_{0|-1})(y_{-1} - \hat{y}_{-1|-1})']$, all assumed zero therein, and y_0 are needed.

5.4 An alternative form of LMMSE estimation

From the appendix of [130], it is easy to verify the following two lemmas for LMMSE estimation.

Lemma 1 For a scalar-valued γ , if γ is uncorrelated with x and z , then

$$E^*[\gamma x|z] = E[\gamma]E^*[x|z]$$

Lemma 2 For the LMMSE estimate \hat{x} of x , we have

$$E[\hat{x}\hat{x}'] = E[xx'] - \text{MSE}(\hat{x})$$

With these two lemmas, besides the two forms of LMMSE estimation obtained in [113] and [114], an alternative form is given in Theorem 1.

Theorem 1 (LMMSE estimation). Given p_k , $\hat{x}_{k-1|k-1} = E^*[x_{k-1}|y^{k-1}]$, $P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1})$, an alternative form of the LMMSE estimate of x_k for system (5.1)–(5.3) is:

Prediction:

$$\begin{aligned}\hat{x}_{k|k-1} &= E^*[x_k|y^{k-1}] = F_{k-1}\hat{x}_{k-1|k-1} \\ P_{k|k-1} &= \text{MSE}(\hat{x}_{k|k-1}) = F_{k-1}P_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}\end{aligned}$$

Update:

$$\begin{aligned}\hat{x}_{k|k} &= E^*[x_k|y^k] = \hat{x}_{k|k-1} + C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} C_{\tilde{y}_{k|k-1}}^+ \tilde{y}_{k|k-1} \\ P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = P_{k|k-1} - C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} C_{\tilde{y}_{k|k-1}}^+ C'_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} \\ \tilde{y}_{k|k-1} &= y_k - p_k H_k \hat{x}_{k|k-1} - q_k y_{k-1} \\ C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} &= p_k P_{k|k-1} H'_k \\ C_{\tilde{y}_{k|k-1}} &= p_k^2 H_k P_{k|k-1} H'_k + p_k R_k + p_k q_k (H_k C_k H'_k - H_k E_k - E'_k H'_k + D_{k-1}) \\ C_k &= F_{k-1} C_{k-1} F'_{k-1} + G_{k-1} Q_{k-1} G'_{k-1} \\ D_{k-1} &= p_{k-1} (H_{k-1} C_{k-1} H'_{k-1} + R_{k-1}) + q_{k-1} D_{k-2} \\ E_k &= p_{k-1} F_{k-1} C_{k-1} H'_{k-1} + q_{k-1} F_{k-1} E_{k-1}\end{aligned}$$

with the initial conditions

$$\begin{aligned}\hat{x}_{0|0} &= \bar{x}_0, \quad P_{0|0} = P_0, \quad C_0 = P_0 + \bar{x}_0 \bar{x}'_0 \\ D_0 &= E[y_0 y'_0], \quad E_1 = F_0 E[x_0 y'_0] + G_0 E[w_0 y'_0]\end{aligned}$$

Proof:

Given $\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$, it follows easily from the property of LMMSE estimation

that

$$\begin{aligned}\hat{x}_{k|k-1} &= E^*[x_k|y^{k-1}] = E^*[F_{k-1}x_{k-1} + G_{k-1}w_{k-1}|y^{k-1}] = F_{k-1}\hat{x}_{k-1|k-1} \\ \tilde{x}_{k|k-1} &= x_k - \hat{x}_{k|k-1} = F_{k-1}\tilde{x}_{k-1|k-1} + G_{k-1}w_{k-1} \\ P_{k|k-1} &= \text{MSE}(\hat{x}_{k|k-1}) = F_{k-1}P_{k-1|k-1}F_{k-1}' + G_{k-1}Q_{k-1}G_{k-1}'\end{aligned}$$

The LMMSE estimator $E^*[x_k|y^k]$ always has the following *quasi-recursive* form [73]

$$\begin{aligned}\hat{x}_{k|k} &= E^*[x_k|y^k] = E^*[x_k|y^{k-1}, y_k] = \hat{x}_{k|k-1} + C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} C_{\tilde{y}_{k|k-1}}^+ \tilde{y}_{k|k-1} \\ P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = P_{k|k-1} - C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} C_{\tilde{y}_{k|k-1}}^+ C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}}'\end{aligned}$$

From Lemma 1 above, we have

$$\begin{aligned}\tilde{y}_{k|k-1} &= y_k - E^*[y_k|y^{k-1}] = y_k - E^*[\gamma_k z_k + (1 - \gamma_k)y_{k-1}|y^{k-1}] \\ &= y_k - E[\gamma_k]E^*[H_k x_k + v_k|y^{k-1}] - E[1 - \gamma_k]E^*[y_{k-1}|y^{k-1}] \\ &= y_k - p_k H_k \hat{x}_{k|k-1} - q_k y_{k-1}\end{aligned}$$

Furthermore, $\tilde{y}_{k|k-1}$ can be rewritten as

$$\begin{aligned}\tilde{y}_{k|k-1} &= y_k - p_k H_k \hat{x}_{k|k-1} - q_k y_{k-1} \\ &= \gamma_k (H_k x_k + v_k) + (1 - \gamma_k)y_{k-1} - p_k H_k \hat{x}_{k|k-1} - (1 - p_k)y_{k-1} \\ &= \gamma_k H_k x_k - p_k H_k \hat{x}_{k|k-1} + \gamma_k v_k + (p_k - \gamma_k)y_{k-1} \\ &= \gamma_k H_k x_k - \gamma_k H_k \hat{x}_{k|k-1} + \gamma_k H_k \hat{x}_{k|k-1} - p_k H_k \hat{x}_{k|k-1} + \gamma_k v_k + (p_k - \gamma_k)y_{k-1} \\ &= \gamma_k H_k \tilde{x}_{k|k-1} + (\gamma_k - p_k)H_k \hat{x}_{k|k-1} + \gamma_k v_k + (p_k - \gamma_k)y_{k-1}\end{aligned}$$

By the principle of orthogonality, $\tilde{x}_{k|k-1}$ is orthogonal to $\hat{x}_{k|k-1}$ and y_{k-1} . Thus

$$C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}} = \text{cov}(\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}) = p_k P_{k|k-1} H'_k$$

$$\begin{aligned} C_{\tilde{y}_{k|k-1}} &= \text{cov}(\tilde{y}_{k|k-1}) = E[\tilde{y}_{k|k-1} \tilde{y}'_{k|k-1}] \\ &= E[\gamma_k^2 (H_k P_{k|k-1} H'_k + R_k) + E[(\gamma_k - p_k)^2] (H_k E[\hat{x}_{k|k-1} \hat{x}'_{k|k-1}] H'_k - H_k E[\hat{x}_{k|k-1} y'_{k-1}]) \\ &\quad - E[y_{k-1} \hat{x}'_{k|k-1}] H'_k + E[y_{k-1} y'_{k-1}]) \\ &= p_k (H_k P_{k|k-1} H'_k + R_k) + (p_k - p_k^2) (H_k E[\hat{x}_{k|k-1} \hat{x}'_{k|k-1}] H'_k - H_k E[(x_k - \tilde{x}_{k|k-1}) y'_{k-1}] \\ &\quad - E[y_{k-1} (x_k - \tilde{x}_{k|k-1})'] H'_k + E[y_{k-1} y'_{k-1}]) \\ &= p_k (H_k P_{k|k-1} H'_k + R_k) + (p_k - p_k^2) (H_k (C_k - P_{k|k-1}) H'_k - H_k E_k - E'_k H'_k + D_{k-1}) \\ &= p_k^2 H_k P_{k|k-1} H'_k + p_k R_k + p_k q_k (H_k C_k H'_k - H_k E_k - E'_k H'_k + D_{k-1}) \end{aligned}$$

where we have used

$$E[\hat{x}_{k|k-1} \hat{x}'_{k|k-1}] = E[x_k x'_k] - P_{k|k-1} = C_k - P_{k|k-1}$$

$$\begin{aligned} C_k &\triangleq E[x_k x'_k] = E[(F_{k-1} x_{k-1} + G_{k-1} w_{k-1})(F_{k-1} x_{k-1} + G_{k-1} w_{k-1})'] \\ &= F_{k-1} C_{k-1} F'_{k-1} + G_{k-1} Q_{k-1} G'_{k-1} \end{aligned}$$

$$\begin{aligned}
D_k &\triangleq E[y_k y_k'] \\
&= E[(\gamma_k z_k + (1 - \gamma_k) y_{k-1})(\gamma_k z_k + (1 - \gamma_k) y_{k-1})'] \\
&= E[\gamma_k^2 z_k z_k'] + E[\gamma_k(1 - \gamma_k) z_k y_{k-1}'] + E[\gamma_k(1 - \gamma_k) y_{k-1} z_k'] + E[(1 - \gamma_k)^2 y_{k-1} y_{k-1}'] \\
&= p_k E[z_k z_k'] + q_k D_{k-1} \\
&= p_k E[(H_k x_k + v_k)(H_k x_k + v_k)'] + q_k D_{k-1} \\
&= p_k (H_k C_k H_k' + R_k) + q_k D_{k-1}
\end{aligned}$$

$$\begin{aligned}
E_k &\triangleq E[x_k y_{k-1}'] \\
&= E[(F_{k-1} x_{k-1} + G_{k-1} w_{k-1})(\gamma_{k-1} z_{k-1} + (1 - \gamma_{k-1}) y_{k-2})'] \\
&= E[(F_{k-1} x_{k-1} + G_{k-1} w_{k-1})(\gamma_{k-1} (H_{k-1} x_{k-1} + v_{k-1}) + (1 - \gamma_{k-1}) y_{k-2})'] \\
&= p_{k-1} F_{k-1} C_{k-1} H_{k-1}' + (1 - p_{k-1}) F_{k-1} E_{k-1}
\end{aligned}$$

with the initial conditions

$$\begin{aligned}
C_0 &= E[x_0 x_0'] = \text{cov}(x_0) + E[x_0] E'[x_0] = P_0 + \bar{x}_0 \bar{x}_0' \\
D_0 &= E[y_0 y_0'], \quad E_1 = E[x_1 y_0'] = E[(F_0 x_0 + G_0 w_0) y_0'] = F_0 E[x_0 y_0'] + G_0 E[w_0 y_0']
\end{aligned}$$

□

Remark: It can be easily seen that when $p_k = 1$, this LMMSE estimator reduces to the Kalman filter.

Remark: It can be clearly seen that to calculate $\hat{x}_{k|k}$, we do need both y_k and y_{k-1} . That is why we assume that y_{k-1} is still available when y_k is received in the problem formulation part. This is similar to what is done in difference measurement method [11, 74] for estimation under autocorrelated measurement noise.

Remark: Since two forms of the LMMSE estimation in the presence of multiple packet

dropouts are already available in [113] and [114], why derive still another one? There are three sources of motivation. The first is from computational consideration, the second is from the angle of measurement residual characterization and the third is from the initialization perspective. Compared with the two existing forms, a system of a higher dimension like the one in (5.4) and (5.5) and matrix operation with a higher dimension like the one in (5.6) are never used, so the computational burden is reduced. The LMMSE estimator of Theorem 1 is in a similar form as the Kalman filter and its measurement residual is well characterized by $\tilde{y}_{k|k-1}$, $C_{\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}}$ and $C_{\tilde{y}_{k|k-1}}$. To initialize the LMMSE estimator of Theorem 1, besides $\hat{x}_{0|0} = \bar{x}_0$ and $P_{0|0} = P_0$, we need $E[y_0 y_0']$, $E[x_0 y_0']$, $E[w_0 y_0']$ and y_0 itself, which is clearly easier to obtain than what is required by the two existing forms.

As we know, LMMSE achieves the smallest MSE within the linear class w.r.t. y^k . Given only the first two moments of other random quantities except γ_k , can we do better than the LMMSE estimator of Theorem 1 in terms of MSE? The answer is yes, as shown in the next section.

5.5 MMSE estimation

For an estimation problem involving intermediate decision, hard decision may be worse than soft decision. That is also one of the reasons why soft decision based algorithms, e.g., interacting multiple model (IMM) algorithm [16], are popular in maneuvering target tracking, as opposed to hard decision based algorithms, e.g., variable dimension filter [5] and input estimation (IE) algorithm [20]. But as will be shown in the following, the best state estimation performance in the presence of multiple packet dropouts is achieved by hard decision. Note that in this case hard decision is equivalent to soft decision and hard decision can be done without any decision error.

Consider system (5.1)–(5.3). If we further assume that x_0 , $\langle w_k \rangle$ and $\langle v_k \rangle$ are Gaussian

distributed, then the MMSE estimation can be simply done as in Theorem 2.

Theorem 2 (MMSE estimation)¹. If x_0 , $\langle w_k \rangle$ and $\langle v_k \rangle$ are Gaussian distributed and given $\hat{x}_{k-1|k-1} = E[x_{k-1}|y^{k-1}]$, $P_{k-1|k-1} = \text{MSE}(\hat{x}_{k-1|k-1})$, then the MMSE estimate of x_k for system (5.1) – (5.3) is:

Prediction:

$$\begin{aligned}\hat{x}_{k|k-1} &= E[x_k|y^{k-1}] = F_{k-1}\hat{x}_{k-1|k-1} \\ P_{k|k-1} &= \text{MSE}(\hat{x}_{k|k-1}) = F_{k-1}P_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}\end{aligned}$$

Update:

If $y_k = y_{k-1}$, then

$$\begin{aligned}\hat{x}_{k|k} &= E[x_k|y^k] = \hat{x}_{k|k-1} \\ P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = P_{k|k-1}\end{aligned}$$

Otherwise (i.e., if $y_k \neq y_{k-1}$)

$$\begin{aligned}\hat{x}_{k|k} &= E[x_k|y^k] = \hat{x}_{k|k-1} + P_{k|k-1}H'_k(H_kP_{k|k-1}H'_k + R_k)^+(y_k - H_k\hat{x}_{k|k-1}) \\ P_{k|k} &= \text{MSE}(\hat{x}_{k|k}) = P_{k|k-1} - P_{k|k-1}H'_k(H_kP_{k|k-1}H'_k + R_k)^+H_kP_{k|k-1}\end{aligned}$$

Proof:

The prediction follows easily from the property of MMSE estimation.

From the total expectation theorem, the MMSE estimate of x_k can be written as:

$$E[x_k|y^k] = E[x_k|\gamma_k = 1, y^k]P\{\gamma_k = 1|y^k\} + E[x_k|\gamma_k = 0, y^k]P\{\gamma_k = 0|y^k\}$$

¹On Apr. 11, 2010, i.e., three days before the defense of this dissertation, it was found that the same idea to obtain MMSE estimation was published in [85]. We worked out this result in September 2009 and presented it publicly in the ISL group seminar of UNO on November 6, 2010.

From the property of a continuous distribution, it follows that

$$P\{z_k = y_{k-1}\} = 0$$

With this, we can easily see that if $y_k = y_{k-1}$, then

$$P\{\gamma_k = 0|y^k\} = 1, \quad P\{\gamma_k = 1|y^k\} = 0$$

and in this case the updated MMSE estimate of x_k simplifies to

$$E[x_k|y^k] = E[x_k|\gamma_k = 0, y^k] = E[x_k|y^{k-1}]$$

which is nothing but the one-step ahead prediction of x_k .

If $y_k \neq y_{k-1}$, then

$$y_k = z_k, \quad P\{\gamma_k = 1|y^k\} = 1, \quad P\{\gamma_k = 0|y^k\} = 0$$

and in this case the updated MMSE estimate of x_k simplifies to

$$E[x_k|y^k] = E[x_k|\gamma_k = 1, y^k] = E[x_k|y^{k-1}, z_k = y_k]$$

And since the LMMSE estimation turns out to be the MMSE estimation under the Gaussian assumption, it follows easily that

$$E[x_k|y^k] = E[x_k|y^{k-1}, z_k = y_k] = \hat{x}_{k|k-1} + P_{k|k-1}H'_k(H_kP_{k|k-1}H'_k + R_k)^+(y_k - H_k\hat{x}_{k|k-1})$$

if $y_k \neq y_{k-1}$.

□

Remark: From Theorem 2, the test of the equivalence between y_k and y_{k-1} clearly has nothing to do with the distribution of x_0 , $\langle w_k \rangle$ and $\langle v_k \rangle$, so even if only the first two moments of x_0 , $\langle w_k \rangle$ and $\langle v_k \rangle$ are available, we can still use Theorem 2 to obtain an estimate of x_k . The simple hard decision between $y_k = y_{k-1}$ and $y_k \neq y_{k-1}$ is a nonlinear operation and is not allowed in LMMSE estimation. It can help reduce the uncertainty associated with y_k , so the MMSE estimator should have better performance than the LMMSE estimator of Theorem 1. This is further verified by the illustrative examples in the next section. In this case, the estimator is still optimal in the LMMSE sense, but not w.r.t. y^k any more. Instead, it is optimal w.r.t. the raw measurement sequence $\langle z_k \rangle$ without the unrarrived packets.

Remark: It is clear from Theorem 2 that the test of the equivalence between y_k and y_{k-1} has nothing to do with the distribution of γ_k . That is, we do not need to know the exact value of p_k if $\langle \gamma_k \rangle$ is a Bernoulli distributed white sequence. Also $\langle \gamma_k \rangle$ can be any binary random sequence, e.g., a Markov chain.

Remark: It can be easily seen that performance of the LMMSE estimation algorithms in [113, 114] and the LMMSE estimation algorithm of Theorem 1 all depend on the value of y_0 . But for the MMSE estimator of Theorem 2, its performance does not depend on the value of y_0 because y_0 is used only for comparison when y_1 is received.

Remark: As is clear from Theorem 2, state estimation in the presence of multiple packet dropouts as formulated by system (5.1)–(5.3) is easy since the MMSE estimate can be obtained based on the simple hard decision by comparing y_k and y_{k-1} . For problems for which this comparison is legitimate, our simple solution by Theorem 2 largely nullifies the existing work on this problem.

Remark: Taken into account of the finite word length effect, performance loss will occur in the implementation of Theorem 2 when a digital quantity is utilized in place of an analog one. This is due to the many-to-one mapping from analog to digital. However, performance loss should be small for modern digital equipment based implementation.

5.6 Illustrative examples

5.6.1 Example 1— $\langle \gamma_k \rangle$ is a Bernoulli distributed white sequence

In this example, we verify the effect of y_0 and data arrival probability p_k on the performance of our proposed LMMSE and MMSE estimators through numerical examples.

Consider the system (5.1)–(5.3), where

$$F_k = 0.95, G_k = 1, w_k \sim \mathcal{N}(0, Q_k), Q_k = 1, k = 0, 1, 2, \dots, 50$$

$$x_0 \sim \mathcal{N}(\bar{x}_0, P_0), \bar{x}_0 = 0, P_0 = 20$$

$$v_k \sim \mathcal{N}(0, R_k), R_k = 9, k = 1, 2, \dots, 50$$

It is also known that

$$y_0 \sim \mathcal{N}(\bar{y}_0, \sigma_y^2), \bar{y}_0 = 0$$

and y_0 is uncorrelated with x_0 and w_0 .

All results in the following are averaged over 1000 Monte Carlo runs.

Figs. 5.1, 5.2 and 5.3 show comparison results of estimators listed in Table 5.1 for $p_k = 0.8, 0.5$ and 0.2 , respectively.

Table 5.1: Estimators used in Fig. 5.1 to Fig. 5.3

name	explanation
KF	Kalman filter without packet dropouts
MMSE-a	MMSE estimator with $\sigma_y^2 = 100$
MMSE-b	MMSE estimator with $\sigma_y^2 = 900$
LMMSE-a	LMMSE estimator with $\sigma_y^2 = 100$
LMMSE-b	LMMSE estimator with $\sigma_y^2 = 900$

As is clear from the simulation results, the KF achieves the best performance since there is no loss in the raw measurement information. The difference between the LMMSE-a

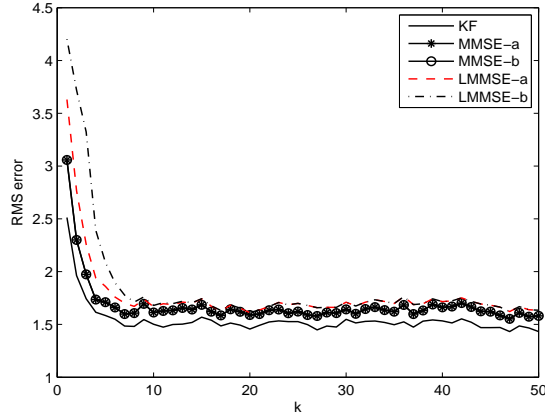


Figure 5.1: RMS error comparison with $p_k = 0.8$. Note that MMSE-a and MMSE-b overlap with each other.

filter and LMMSE-b filter verifies that the LMMSE estimator of Theorem 1 depends on the distribution and value of y_0 . The overlap of MMSE-a and MMSE-b verifies that the MMSE estimator does not depend on the value of y_0 and it is only used for comparison. Also, it can be seen that the MMSE estimator outperforms the LMMSE estimator, as expected. The gap between the MMSE estimator and the KF discloses the effect of packet dropouts over network transmission. As p_k decreases, the gap increases. Also, as p_k decreases, the performance of LMMSE and MMSE estimators becomes worse and their convergence rates become slower. All these are because as p_k decreases, raw measurement packets will arrive at the estimator less frequently. Correspondingly, the estimators will rely on the prediction more and more.

5.6.2 Example 2— $\langle \gamma_k \rangle$ is a Markov chain

In this example, the case in which $\langle \gamma_k \rangle$ is a Markov chain is considered. We verify the effect of the parameters of the Markov chain on the performance of our proposed MMSE estimator through numerical examples.

Consider still the system (5.1)–(5.3) with the same parameters as in example 1 except

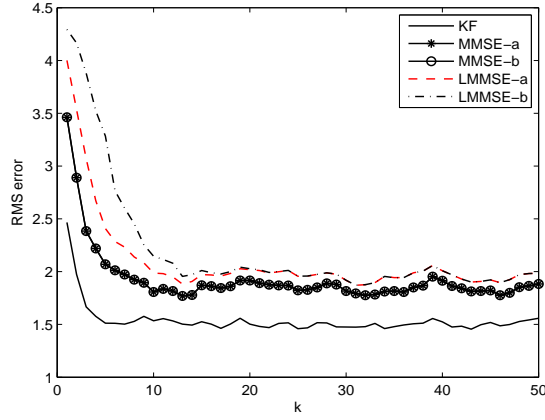


Figure 5.2: RMS error comparison with $p_k = 0.5$. Note that MMSE-a and MMSE-b overlap with each other.

that

$$\sigma_y^2 = 100$$

and $\langle \gamma_k \rangle$ is a homogeneous Markov chain described by initial probabilities

$$P\{\gamma_1 = 1\} = p_1, \quad P\{\gamma_1 = 0\} = q_1 = 1 - p_1$$

and transition probabilities

$$P\{\gamma_{k+1} = 1 | \gamma_k = 1\} = p_{11}, \quad P\{\gamma_{k+1} = 0 | \gamma_k = 1\} = p_{10} = 1 - p_{11}$$

$$P\{\gamma_{k+1} = 0 | \gamma_k = 0\} = p_{00}, \quad P\{\gamma_{k+1} = 1 | \gamma_k = 0\} = p_{01} = 1 - p_{00}$$

where γ_k is uncorrelated with all the other random variables.

All results in the following are averaged over 1000 Monte Carlo runs.

Figs. 5.4 and 5.5 show comparison results of estimators listed in Table 5.2 for two extreme cases where $(p_{11}, p_{00}) = (1, 1)$ and $(0.5, 0.5)$, respectively.

For the case of $p_{11} = 1, p_{00} = 1$, it means that if $\gamma_1 = 1$, then $\gamma_k = 1, \forall k \geq 2$; otherwise if $\gamma_1 = 0$, then $\gamma_k = 0, \forall k \geq 2$. That is, if z_1 is received at $k = 1$, then all z_k 's, $k = 2, 3, \dots$,

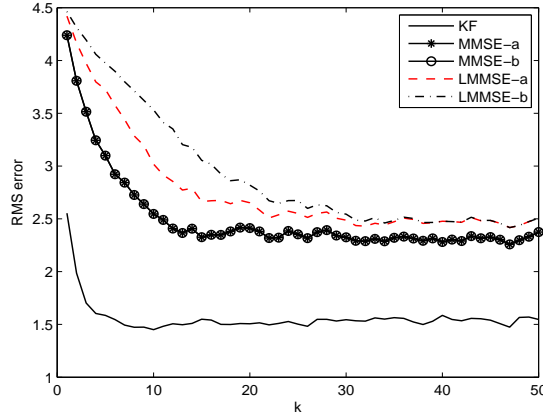


Figure 5.3: RMS error comparison with $p_k = 0.2$. Note that MMSE-a and MMSE-b overlap with each other.

Table 5.2: Estimators used in Fig. 5.4 and Fig. 5.5

name	explanation
KF	Kalman filter without packet dropouts
MMSE-a	MMSE estimator with initial probability $p_1 = 0.8$
MMSE-b	MMSE estimator with initial probability $p_1 = 0.5$
MMSE-c	MMSE estimator with initial probability $p_1 = 0.2$

will also be received; otherwise if z_1 is dropped at $k = 1$, then all z_k 's, $k = 2, 3, \dots$, will also be dropped. Thus p_1 will control the performance of the MMSE filters and these filters with different p_1 values will have different performance.

For the case of $p_{11} = 0.5$, $p_{00} = 0.5$, it means that regardless of p_1 , for $k = 2, 3, \dots$, we always have

$$P\{\gamma_k = 1\} = 0.5, P\{\gamma_k = 0\} = 0.5, \forall k \geq 2$$

This is equivalent to having $\langle \gamma_k \rangle$, $k = 2, 3, \dots$, as a Bernoulli distributed white sequence with $p_k = 0.5$ ($\gamma_k = 1$ and $\gamma_k = 0$ have an equal chance of occurrence). That is, except p_1 , probabilistically there is no difference among the three MMSE filters listed in Table 5.2 in terms of p_k , $k = 2, 3, \dots$. Correspondingly, on the average, except during the initial short transient, the three MMSE filters should have very close performance.

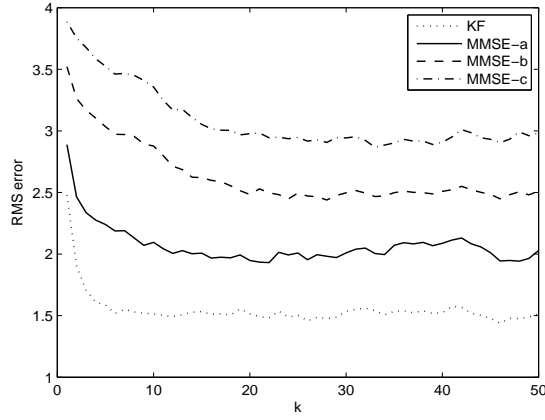


Figure 5.4: RMS error comparison with $p_{11} = 1$, $p_{00} = 1$

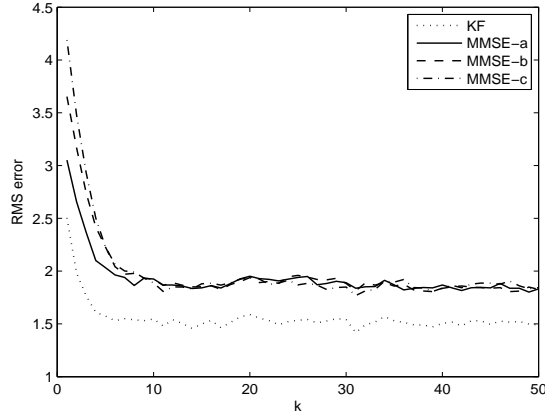


Figure 5.5: RMS error comparison with $p_{11} = 0.5$, $p_{00} = 0.5$

As is clear from the simulation results, the KF achieves the best performance among all filters since there is no loss in the raw measurement information. For the case of $p_{11} = 1$, $p_{00} = 1$, MMSE-a has the best performance among all three MMSE filters and MMSE-c has the worst. For the case of $p_{11} = 0.5$, $p_{00} = 0.5$, except for the noticeable difference during the initial short transient, all three MMSE filters do have very close performance. The simulation results are completely in accordance with the above theoretical analysis. The reason is that the hard decision involved in our proposed MMSE filter has no decision error and it can capture each realization of the binary sequence $\langle \gamma_k \rangle$ exactly.

5.7 Summary

In network based applications, communication between local sensors and the processing center is usually not perfectly reliable, so packet dropouts may happen. Two existing forms of the LMMSE estimation in the presence of multiple packet dropouts are based on a stochastic parameter system constructed by augmenting the original state and measurement. Concerning the computational load, measurement residual characterization, and requirements on initialization, these two forms are not preferred. We have first derived an alternative form of the LMMSE estimator. Then under the Gaussian assumption, we have also derived the MMSE estimator. The MMSE estimator has two nice properties. First, unlike the LMMSE estimator, its performance does not depend on the value of the initial data at the processing center. Second, it does not depend on the distribution of the data arrival events. These have been verified by numerical examples. The MMSE estimator is obtained based on a hard decision by comparing the measurement values at two consecutive time instants. If this comparison is legitimate, our simple optimal solution largely nullifies the existing work on this problem.

Chapter 6

Conclusions and Future Work

This dissertation addresses state estimation with two types of unconventional measurements and two types of network-induced estimation problems. The two types of unconventional measurements are noise-free measurements and set measurements. The two types of network-induced estimation problems are optimal state estimation in the presence of multiple packet dropouts and optimal distributed estimation fusion with transformed data.

State estimation with noisy and noise-free measurements has been formulated in a general setting. It has numerous real supports, e.g., state estimation under linear or nonlinear equality constraints, with correlated or singular measurement noise. Two sequential forms equivalent to the batch LMMSE estimator have been obtained to reduce the computational complexity. The contribution of update by treating equality constraints as noise-free measurements has been analyzed especially. This work has been documented in [35, 37].

State estimation with point and set measurements has also been formulated in a general setting. It has numerous real supports, e.g., state estimation under linear or nonlinear inequality constraints, with quantized measurements. Inspired by the estimation with quantized measurements developed by Curry [28], under a Gaussian assumption, the MMSE state estimator with point measurements and set measurements of any shape has been proposed by discretizing continuous set measurements. The contribution of update by treating set constraints as noise-free measurements has been analyzed especially. This work has been

documented in [32, 33, 42, 58].

State estimation in the presence of multiple packet dropouts originated from unreliable communication between local sensors and the processing center. To overcome the shortcomings of two existing LMMSE estimators, an alternative form of LMMSE estimation has been proposed first. Then under a Gaussian assumption, the MMSE estimation with multiple packet dropouts has also been obtained based on a hard decision by comparing the measurements at two consecutive time instants. It is also pointed out that if this comparison is legitimate, our simple MMSE solution largely nullifies the existing work on this problem. This work has been documented in [39, 40].

By taking linear transformation of the raw measurements received by each sensor, two optimal distributed fusion algorithms have been proposed. They have three nice properties. First, they are globally optimal in that they are equivalent to centralized fusion. Second, in terms of communication requirements from each sensor to the fusion center, distributed fusion with transformation I is the same or better than most existing distributed fusion algorithms, and distributed fusion with transformation II is better than centralized and most existing distributed fusion algorithms. Third, they do not need the inverses of error covariance matrices, which are assumed in most existing distributed fusion algorithms but can not be guaranteed to exist, so they can be applied in more general cases. This work has been documented in [34, 36, 38].

In our work on distributed fusion with transformed data, the compressed dimension for each sensor is the rank of the measurement matrix. Whether this is the minimal compressible dimension for the problem is a topic for future work. Also, for the extension to reduced-rate communication, we have only considered the case without process noise. Extension to the case with process noise is another topic for future work.

I am also interested in constrained parameter estimation, e.g., constrained least-squares estimation. We have successively solved isoform proportion estimation problem in next

generation gene sequence analysis where it is required that the isoform proportions sum up to one and they are nonnegative. One key to solve this problem is the constrained least-squares estimation.

Bibliography

- [1] A. T. Alouani, J. E. Gray, and D. H. McCabe. Theory of distributed estimation using multiple asynchronous sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 41(2):717–722, April 2005.
- [2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, February 2002.
- [3] Y. Bar-Shalom. On the track-to-track correlation problem. *IEEE Transactions on Automatic Control*, 26(2):571–572, Apr 1981.
- [4] Y. Bar-Shalom. Update with out-of-sequence measurements in tracking: exact solution. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):769–777, July 2002.
- [5] Y. Bar-Shalom and K. Birmiwal. Variable dimension filter for maneuvering target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 18(5):621–629, September 1982.
- [6] Y. Bar-Shalom and L. Campo. The effect of the common process noise on the two-sensor fused-track covariance. *IEEE Transactions on Aerospace and Electronic Systems*, 22(6):803–805, November 1986.

- [7] Y. Bar-Shalom and H. M. Chen. IMM estimator with out-of-sequence measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 41(1):90–98, January 2005.
- [8] Y. Bar-Shalom and H. M. Chen. Removal of out-of-sequence measurements from tracks. *IEEE Transactions on Aerospace and Electronic Systems*, 45(2):612–619, April 2009.
- [9] Y. Bar-Shalom, H. M. Chen, and M. Mallick. One-step solution for the multistep out-of-sequence-measurement problem in tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 40(1):27–37, January 2004.
- [10] Y. Bar-Shalom and X. R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, 1995.
- [11] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley Interscience, 2001.
- [12] Y. Bar-Shalom and E. Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11:451–460, 1975.
- [13] B. M. Bell, J. V. Burke, and G. Pillonetto. An inequality constrained nonlinear Kalman-Bucy smoother by interior point likelihood maximization. *Automatica*, 45(1):25–33, January 2009.
- [14] A. Benavoli, L. Chisci, A. Farina, L. Ortenzi, and G. Zappa. Hard-constrained versus soft-constrained parameter estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 42(4):1224–1239, October 2006.
- [15] D. S. Bernstein and D. C. Hyland. Compartmental modelling and second-moment analysis of state space systems. *SIAM Journal on Matrix Analysis and Applications*, 14(3):880–901, 1993.

- [16] H. A. P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, August 1988.
- [17] K. Brammer. Lower order optimal linear filtering of nonstationary random sequences. *IEEE Transactions on Automatic Control*, 13(2):198–199, April 1968.
- [18] A. E. Bryson and L. J. Henrikson. Estimation using sampled data containing sequentially correlated noise. *Journal of Spacecraft*, 5(6):662–665, June 1968.
- [19] A. E. Bryson and D. E. Johansen. Linear filtering for time-varying systems using measurements containing colored noise. *IEEE Transactions on Automatic Control*, 10(1):4–10, January 1965.
- [20] Y. T. Chan, A. G. C. Hu, and J. B. Plant. A Kalman filter based tracking scheme with input estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 15(2):237–244, March 1979.
- [21] K. C. Chang, R. K. Saha, and Y. Bar-Shalom. On optimal track-to-track fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1271–1276, October 1997.
- [22] K. C. Chang, Z. Tian, and S. Mori. Performance evaluation for MAP state estimate fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 40(2):706–714, April 2004.
- [23] K. C. Chang, Z. Tian, and R. K. Saha. Performance evaluation of track fusion with information matrix filter. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):455–466, April 2002.

- [24] H. M. 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.
- [25] C. Y. Chong. Hierarchical estimation. In *Proceedings of the MIT/ONR Workshop on C3*, Monterey, CA, 1979.
- [26] C. Y. Chong, S. Mori, W. H. Barker, and K. C. Chang. Architectures and algorithms for track association and fusion. *IEEE Aerospace and Electronic Systems Magazine*, 15(1):5–13, January 2000.
- [27] J. L. Crassidis and F. L. Markley. Unscented filtering for spacecraft attitude estimation. *AIAA Journal of Guidance, Control, and Dynamics*, 26(4):536–542, 2003.
- [28] R. E. Curry. *Estimation and Control with Quantized Measurements*. MIT Press, February 1970.
- [29] R. E. Curry, W. E. V. Velde, and J. E. Potter. Nonlinear estimation with quantized measurements - PCM, predictive quantization, and data compression. *IEEE Transactions on Information Theory*, 16(2):152–161, March 1970.
- [30] O. E. Drummond. Feedback in track fusion without process noise. In *Proceedings of the 1995 SPIE Conference on Signal and Data Processing of Small Targets*, volume 2561, pages 369–383, Orlando, FL, Sep 1995.
- [31] Z. S. Duan, C. Z. Han, and X. R. Li. Sequential nonlinear tracking filter with range-rate measurements in spherical coordinates. In *Proceedings of the 7th International Conference on Information Fusion*, volume 1, pages 599–605, Stockholm, Sweden, June 2004.

- [32] Z. S. Duan, V. P. Jilkov, and X. R. Li. Posterior Cramer-Rao bounds for state estimation with quantized measurement. In *Proceedings of the 40th IEEE Southeastern Symposium on System Theory*, pages 376–380, New Orleans, LA, USA, March 2008.
- [33] Z. S. Duan, V. P. Jilkov, and X. R. Li. State estimation with quantized measurements: Approximate MMSE approach. In *Proceedings of the 11th International Conference on Information Fusion*, pages 1067–1072, Cologne, Germany, July 2008.
- [34] Z. S. Duan and X. R. Li. Optimal distributed estimation fusion with transformed data. In *Proceedings of the 11th International Conference on Information Fusion*, pages 1291–1297, Cologne, Germany, 2008.
- [35] Z. S. Duan and X. R. Li. Best linear unbiased state estimation with noisy and noise-free measurements. In *Proceedings of the 12th International Conference on Information Fusion*, pages 2193–2200, Seattle, Washington, July 2009.
- [36] Z. S. Duan and X. R. Li. Optimal distributed estimation fusion with compressed data. In *Proceedings of the 12th International Conference on Information Fusion*, pages 563–570, Seattle, Washington, July 6-9 2009.
- [37] Z. S. Duan and X. R. Li. Optimal linear state estimation with noisy and noise-free measurements. *Submitted to IEEE Transactions on Aerospace and Electronic Systems*, September 2009.
- [38] Z. S. Duan and X. R. Li. Lossless linear transformation of sensor data for distributed estimation fusion. *Submitted to IEEE Transactions on Signal Processing*, revised, January 2010.
- [39] Z. S. Duan and X. R. Li. On optimal state estimation with multiple packet dropouts. *Submitted to IEEE Transactions on Signal Processing*, March 2010.

- [40] Z. S. Duan and X. R. Li. Optimal state estimation with multiple packet dropouts. *Submitted to the 13th International Conference on Information Fusion*, Edinburgh, UK, 2010.
- [41] Z. S. Duan, X. R. Li, C. Z. Han, and H. Y. Zhu. Sequential Unscented Kalman filter for radar target tracking with range rate measurements. In *Proceedings of the 8th International Conference on Information Fusion*, volume 1, pages 130–137, Philadelphia, PA, USA, July 2005.
- [42] Z. S. Duan, X. R. Li, and V. P. Jilkov. State estimation with point and set measurements. *Submitted to the 13th International Conference on Information Fusion*, Edinburgh, UK, 2010.
- [43] Z. S. Duan, X. R. Li, and M. Yang. Why are more sensors better in estimation fusion? In *Proceedings of 2007 International Colloquium on Information Fusion*, pages 173–179, Xi’an, China, August 22-25 2007.
- [44] F. W. Fairman. Hybrid estimators for discrete-time stochastic systems. *IEEE Transactions on Systems, Man and Cybernetics*, 8(12):849–854, December 1978.
- [45] F. W. Fairman and L. Luk. On reducing the order of Kalman filters for discrete-time stochastic systems having singular measurement noise. *IEEE Transactions on Automatic Control*, 30(11):1150–1152, November 1985.
- [46] J. Fang and H. B. Li. Power constrained distributed estimation with correlated sensor data. *IEEE Transactions on Signal Processing*, 57(8):3292–3297, August 2009.
- [47] E. Fogel and Y. F. Huang. Reduced-order optimal state estimator for linear systems with partially noise corrupted measurement. *IEEE Transactions on Automatic Control*, 25(5):994–996, October 1980.

- [48] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal Oceanic Engineering*, 8:173–184, July 1983.
- [49] R. Gazit. Digital tracking filters with high order correlated measurement noise. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):171–177, January 1997.
- [50] M. L. Graham, F. O’Brien, K. F. Gong, and S. E. Hammel. Adaptive filtering in underwater tracking with correlated measurement noise. In *Proceedings of 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 285–288, Minneapolis, MN, April 1993.
- [51] N. Gupta. Kalman filtering in the presence of state space equality constraints. In *Proceedings of the 26th Chinese Control Conference*, pages 107–113, Zhangjiajie, Hunan, China, July 2007.
- [52] J. A. Guu and C. H. Wei. Maneuvering target tracking using IMM method at high measurement frequency. *IEEE Transactions on Aerospace and Electronic Systems*, 27(3):514–519, May 1991.
- [53] H. R. Hashemipour, S. Roy, and A. J. Laub. Decentralized structures for parallel Kalman filtering. *IEEE Transactions on Automatic Control*, 33(1):88–94, January 1988.
- [54] C. S. Hsieh and F. C. Chen. Optimal minimal-order least-squares estimators via the general two-stage Kalman filter. *IEEE Transactions on Automatic Control*, 46(11):1772–1776, November 2001.
- [55] Y. Y. Hu, Z. S. Duan, and C. Z. Han. Optimal batch asynchronous fusion algorithm. In *Proceedings of 2005 IEEE International Conference on Vehicular Electronics and Safety*, pages 237–240, Xi’an, China, Oct 14-16 2005.

- [56] Y. Y. Hu, Z. S. Duan, and D. H. Zhou. Estimation fusion with general asynchronous multi-rate sensors. *IEEE Transactions on Aerospace and Electronic Systems*, accepted, September 2009.
- [57] K. Ito and K. Q. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, May 2000.
- [58] V. P. Jilkov, Z. S. Duan, and X. R. Li. State estimation with quantized measurements: Multiple model approach. In *Proceedings of 2007 International Colloquium on Information Fusion*, pages 1–7, Xi’an, China, August 2007.
- [59] Y. T. Ju and V. B. Haas. A duality principle for state estimation with partially noise corrupted measurements. In *Proceedings of the 20th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, volume 20, pages 623–624, San Diego, CA, December 1981.
- [60] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, March 2000.
- [61] S. J. Julier and J. J. LaViola. On Kalman filtering with nonlinear equality constraints. *IEEE Transactions on Signal Processing*, 55(6):2774–2784, June 2007.
- [62] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- [63] R. Karlsson and F. Gustafsson. Particle filtering for quantized sensor information. In *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, September 2005.

- [64] K. H. Kim. Development of track to track fusion algorithms. In *Proceedings of the 1994 American Control Conference*, pages 1037–1041, Baltimore, MD, June 1994.
- [65] T. Kirubarajan, Y. Bar-Shalom, K. R. Pattipati, and I. Kadar. Ground target tracking with variable structure IMM estimator. *IEEE Transactions on Aerospace and Electronic Systems*, 36(1):26–46, January 2000.
- [66] S. Ko and R. R. Bitmead. State estimation for linear systems with state equality constraints. *Automatica*, 43(8):1363–1368, August 2007.
- [67] J. H. Kotecha and P. M. Djuric. Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 1757–1760, Phoenix, AZ, USA, March 1999.
- [68] S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous Multivariate Distributions Volume 1: Models and Applications*. Wiley, 2000.
- [69] T. Lefebvre, H. Bruyninckx, and J. De Schutter. Comment on "a new method for nonlinear transformation of means and covariances in filters and estimators". *IEEE Transactions on Automatic Control*, 47(8):1406–1409, August 2002.
- [70] J. L. Li and G. AlRegib. Distributed estimation in energy-constrained wireless sensor networks. *IEEE Transactions on Signal Processing*, 57(10):3746–3758, October 2009.
- [71] T. Li, A. Ekpenyong, and Y. F. Huang. Source localization and tracking using distributed asynchronous sensors. *IEEE Transactions on Signal Processing*, 54(10):3991–4003, October 2006.

- [72] X. R. Li. Optimal linear estimation fusion - part VII: dynamic systems. In *Proceedings of the 6th International Conference on Information Fusion*, pages 455–462, Cairns, Australia, July 2003.
- [73] X. R. Li. Recursibility and optimal linear estimation and filtering. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 1761–1766, Atlantis, Paradise Island, Bahamas, December 14-17 2004.
- [74] X. R. Li. *Applied Estimation and Filtering*. Course Notes, University of New Orleans, February 2006.
- [75] X. R. Li. *Introduction to Information Fusion*. Course Notes, University of New Orleans, February 2008.
- [76] X. R. Li. *Target Tracking*. Course Notes, University of New Orleans, August 2008.
- [77] X. R. Li, C. Z. Han, and J. Wang. Discrete-time linear filtering in arbitrary noise. In *Proceedings of the 39th IEEE Conference on Decision and Control*, volume 2, pages 1212–1217, Sydney, Australia, December 2000.
- [78] X. R. Li and V. P. Jilkov. A survey of maneuvering target tracking - approximation techniques for nonlinear filtering. In *Proceedings of 2004 SPIE Conference on Signal and Data Processing of Small Targets*, volume 5428, pages 537–550, San Diego, CA, USA, April 2004.
- [79] X. R. Li and K. S. Zhang. Optimal linear estimation fusion - part IV: Optimality and efficiency of distributed fusion. In *Proceedings of the 4th International Conference on Information Fusion*, pages WeB1–19–WeB1–26, Montreal, QC, Canada, August 2001.

- [80] X. R. Li, Z. L. Zhao, and X. B. Li. General model-set design methods for multiple-model approach. *IEEE Transactions on Automatic Control*, 50(9):1260–1276, September 2005.
- [81] X. R. Li, Y. M. Zhu, J. Wang, and C. Z. Han. Optimal linear estimation fusion - part I: Unified fusion rules. *IEEE Transactions on Information Theory*, 49(9):2192–2208, September 2003.
- [82] X. D. Lin, Y. Bar-Shalom, and T. Kirubarajan. Multisensor multitarget bias estimation for general asynchronous sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 41(3):899–921, April 2005.
- [83] R. K. Mehra and A. E. Bryson. Linear smoothing using measurements containing correlated noise with an application to inertial navigation. *IEEE Transactions on Automatic Control*, 13(5):496–503, October 1968.
- [84] J. M. Mendel. *Lessons in Digital Estimation Theory*. Prentice Hall, Englewood Cliffs, NJ, 1987.
- [85] M. Moayedi, Y. K. Foo, and Y. C. Soh. Adaptive Kalman filtering in networked systems with random sensor delays, multiple packet dropouts and missing measurements. *IEEE Transactions on Signal Processing*, 58(3):1577–1588, March 2010.
- [86] M. Moayedi, Y. C. Soh, and Y. K. Foo. Optimal Kalman filtering with random sensor delays, packet dropouts and missing measurements. In *Proceedings of 2009 American Control Conference*, pages 3405–3410, St. Louis, MO, June 2009.
- [87] S. Mori, W. H. Barker, C.Y. Chong, and K. C. Chang. Track association and track fusion with nondeterministic target dynamics. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):659–668, April 2002.

- [88] D. Musicki. Multi-target tracking using multiple passive bearings-only asynchronous sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 44(3):1151–1160, July 2008.
- [89] M. Norgaard, N. K. Poulsen, and O. Ravn. New developments in state estimation for nonlinear systems. *Automatica*, 36(11):1627–1638, November 2000.
- [90] J. O’Reilly. On linear least-squares estimators for discrete-time stochastic systems. *IEEE Transactions on Systems, Man and Cybernetics*, 10(5):276–279, May 1980.
- [91] J. O’Reilly. Comments on two recent papers on reduced-order optimal state estimation for linear systems with partially noise corrupted measurement. *IEEE Transactions on Automatic Control*, 27(1):280–282, February 1982.
- [92] R. Piziak and P. L. Odell. Full rank factorization of matrices. *Mathematics Magazine*, 72(3):193–201, June 1999.
- [93] D. B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24:843–854, December 1979.
- [94] P. W. Richards. Constrained Kalman filtering using pseudo-measurements. In *Proceedings of 1995 IEE Colloquium on Algorithms for Target Tracking*, pages 75–79, London, UK, May 1995.
- [95] S. R. Rogers. Alpha-beta filter with correlated measurement noise. *IEEE Transactions on Aerospace and Electronic Systems*, 23(4):592–594, July 1987.
- [96] Y. Ruan and P. Willett. A quantization architecture for track fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 41(2):671–681, April 2005.

- [97] Y. Ruan, P. Willett, and A. MARRS. Fusion of quantized measurements via particle filtering. In *Proceedings of 2003 IEEE Aerospace Conference*, volume 4, pages 1967–1978, Big Sky, MT, March 2003.
- [98] M. Sahebsara. Optimal filtering in multiple channel networked control systems with multiple packet dropout. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 3367–3371, Cancun, Mexico, December 2008.
- [99] M. Sahebsara, T. Chen, and S. L. Shah. Optimal H₂ filtering in networked control systems with multiple packet dropout. *IEEE Transactions on Automatic Control*, 52(8):1508–1513, August 2007.
- [100] M. Sahebsara, T. Chen, and S. L. Shah. Optimal H₂ filtering with random sensor delay, multiple packet dropout and uncertain observations. *International Journal of Control*, 80(2):292–301, February 2007.
- [101] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo. Distributed estimation using reduced-dimensionality sensor observations. *IEEE Transactions on Signal Processing*, 55(8):4284–4299, August 2007.
- [102] R. G. Sea. An efficient suboptimal decision procedure for associating sensor data with stored tracks in real-time surveillance systems. In *Proceedings of 1971 IEEE Conference on Decision and Control*, pages 33–37, Miami Beach, Florida, December 1971.
- [103] D. Simon. A game theory approach to constrained minimax state estimation. *IEEE Transactions on Signal Processing*, 54(2):405–412, February 2006.
- [104] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.

- [105] D. Simon and T. L. Chia. Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems*, 38(1):128–136, January 2002.
- [106] D. Simon and D. L. Simon. Aircraft turbofan engine health estimation using constrained Kalman filtering. *ASME Journal of Engineering for Gas Turbines and Power*, 127(2):323–328, April 2005.
- [107] D. Simon and D. L. Simon. Kalman filtering with inequality constraints for turbofan engine health estimation. *IEE Proceedings Control Theory and Applications*, 153(3):371–378, May 2006.
- [108] R. A. Singer and R. G. Sea. New results in optimizing surveillance system tracking and data correlation performance in dense multitarget environments. *IEEE Transactions on Automatic Control*, 18(6):571–582, December 1973.
- [109] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poola, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, September 2004.
- [110] E. B. Song, Y. M. Zhu, and J. Zhou. Sensors’ optimal dimensionality compression matrix in estimation fusion. *Automatica*, 41(12):2131–2139, December 2005.
- [111] E. B. Song, Y. M. Zhu, J. Zhou, and Z. S. You. Optimal Kalman filtering fusion with cross-correlated sensor noises. *Automatica*, 43(8):1450–1456, August 2007.
- [112] S. L. Sun. Linear minimum variance estimators for systems with bounded random measurement delays and packet dropouts. *Signal Processing*, 89(7):1457–1466, July 2009.

- [113] S. L. Sun, L. H. Xie, and W. D. Xiao. Optimal full-order and reduced-order estimators for discrete-time systems with multiple packet dropouts. *IEEE Transactions on Signal Processing*, 56(8):4031–4038, August 2008.
- [114] S. L. Sun, L. H. Xie, W. D. Xiao, and Y. C. Soh. Optimal linear estimation for systems with multiple packet dropouts. *Automatica*, 44(5):1333–1342, May 2008.
- [115] S. L. Sun, L. H. Xie, W. D. Xiao, and N. Xiao. Optimal filtering for systems with multiple packet dropouts. *IEEE Transactions on Circuits and Systems—II: Express Briefs*, 55(7):695–699, July 2008.
- [116] B. O. S. Teixeira, J. Chandrasekar, L. A. B. Torres, L. A. Aguirre, and D. S. Bernstein. State estimation for equality-constrained linear systems. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 6220–6225, New Orleans, LA, USA, December 2007.
- [117] X. Tian and Y. Bar-Shalom. Exact algorithms for four track-to-track fusion configurations: All you wanted to know but were afraid to ask. In *Proceedings of the 12th International Conference on Information Fusion*, pages 537–544, Seattle, Washington, July 6-9 2009.
- [118] P. Tichavsky, C. H. Muravchik, and A. Nehorai. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46(5):1386–1396, May 1998.
- [119] L. S. Wang, Y. T. Chiang, and F. R. Chang. Filtering methods for nonlinear systems with constraints. *IEE Proceedings Control Theory and Applications*, 149(6):525–531, November 2002.
- [120] B. Widrow, I. Kollar, and M. C. Liu. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2):353–361, April 1996.

- [121] W. R. Wu and D. C. Chang. Maneuvering target tracking with colored noise. *IEEE Transactions on Aerospace and Electronic Systems*, 32(4):1311–1320, October 1996.
- [122] L. P. Yan, B. S. Liu, and D. H. Zhou. Asynchronous multirate multisensor information fusion algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 43(3):1135–1146, July 2007.
- [123] C. Yang, M. Bakich, and E. Blasch. Nonlinear constrained tracking of targets on road. In *Proceedings of 2005 International Conference on Information Fusion*, pages 235–242, Philadelphia, PA, July 25-29 2005.
- [124] C. Yang and E. Blasch. Fusion of tracks with road constraints. *Journal of Advances in Information Fusion*, 3(1):14–32, June 2008.
- [125] C. Yang and E. Blasch. Kalman filtering with nonlinear state constraints. *IEEE Transactions on Aerospace and Electronic Systems*, 45(1):70–84, January 2009.
- [126] F. W. Yang and Y. M. Li. Set-membership filtering with state constraints. *IEEE Transactions on Aerospace and Electronic Systems*, 45(4):1619–1629, October 2009.
- [127] K. S. Zhang, X. R. Li, and H. M. Chen. Multi-sensor multi-target tracking with out-of-sequence measurements. In *Proceedings of the 6th International Conference on Information Fusion*, volume 1, pages 672–679, Cairns, Queensland, Australia, July 2003.
- [128] K. S. Zhang, X. R. Li, P. Zhang, and H. F. Li. Optimal linear estimation fusion - part VI: Sensor data compression. In *Proceedings of the 6th International Conference on Information Fusion*, pages 221–228, Cairns, Queensland, Australia, July 2003.
- [129] K. S. Zhang, X. R. Li, and Y. M. Zhu. Optimal update with out-of-sequence measurements. *IEEE Transactions on Signal Processing*, 53(6):1992–2004, June 2005.

- [130] Z. L. Zhao, X. R. Li, and V. P. Jilkov. Best linear unbiased filtering with nonlinear measurements for target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 40(4):1324–1336, October 2004.
- [131] J. Zhou and Y. M. Zhu. The linear minimum mean-square error estimation with constraints and its applications. In *Proceedings of 2006 International Conference on Computational Intelligence and Security*, pages 1801–1804, Guangzhou, China, November 2006.
- [132] Y. M. Zhu and X. R. Li. Recursive least squares with linear constraints. *Communications in Information and Systems*, 7(3):287–312, 2007.
- [133] Y. M. Zhu, E. B. Song, J. Zhou, and Z. S. You. Optimal dimensionality reduction of sensor data in multisensor estimation fusion. *IEEE Transactions on Signal Processing*, 53(5):1631–1639, May 2005.

Vita

Zhansheng Duan received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, China, in 1999 and 2005, respectively, both in Electrical Engineering.

From Feb. 2001 to Aug. 2005, he was a research assistant in the department of automation, Xi'an Jiaotong University. His research was mainly on adaptive estimation, radar target tracking and multi-sensor fusion. Since Aug. 2005, he has been a research assistant in the Department of Electrical Engineering, University of New Orleans. Since Jan. 2010, he has also been an assistant professor/research in the Department of Computer Science, University of New Orleans. His current research interest includes estimation and detection theory, target tracking, information fusion, nonlinear filtering, performance evaluation, gene regulatory network inference, next generation sequence analysis. He has authored or coauthored one book: *Multisource Information Fusion* (Tsinghua University Publishing House, 2006), and 46 journal and conference proceedings papers.