5-22-2006

Performance Appraisal of Estimation Algorithms and Application of Estimation Algorithms to Target Tracking

Zhanlue Zhao
University of New Orleans

PERFORMANCE APPRAISAL OF ESTIMATION ALGORITHMS AND APPLICATION OF ESTIMATION ALGORITHMS TO TARGET TRACKING

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

Zhanlue Zhao

B.S. Xi'an Jiaotong University, China, 1997
M.S. Xi'an Jiaotong University, China, 2000

May, 2006

To My Parents

Acknowledgements

I would like to express my deepest gratitude to my major advisor, Dr. X. Rong Li, for his invaluable inspiration, continuous support, and constructive suggestions, which made it possible for me to complete the Ph.D. degree. The positive and open research atmosphere created by Dr. Li is the best nutrition for me to gain considerable knowledge and insights into problems and techniques in estimation and decision theory, statistical signal processing and information fusion. The research attitude and enthusiasm exemplified by Dr. Li will give me everlasting inspiration and encouragement to my future work and research.

I am also indebted to Dr. Huimin Chen. His help, inspiration and encouragement made my research much more interesting and enjoyable. The collaborations with Dr. Chen often start with heated discussions and end with insightful ideas. Special thanks go to Dr. Vesselin Jilkov for the long-term collaboration and his continuous support. I would also thank Dr. Vesselin Jilkov, Dr. Huimin Chen, Dr. Tumulesh K.S. Solanky, Dr. Zhide Fang and Dr. Stephen C. Lipp for serving on my thesis committee, and for their constructive and valuable comments on the dissertation. I am also very grateful to all my friends and members in the Information and Systems Laboratory, especially Jifeng Ru, Ming Yang, Zhansheng Duan, Keshu Zhang, Peng Zhang, Anwer Bash, Ryan Pitre, Trang Nguyen and Victor Alvarado. With their support and friendship, I had the great pleasure of carrying out research in such a pleasant working environment. And I have to thank many friends who love me and whom I love. They make so much difference in my study life in New Orleans. Furthermore, I would

like to thank all the members and staff of the Department of Electrical Engineering at the University of New Orleans for their support.

Last but not least, I want to thank my parents for their love and encouragement. Their love gives me so much meaning in my life. I dedicate this dissertation to them.

Contents

Abstract

This dissertation consists of two parts. The first part deals with the performance appraisal of estimation algorithms. The second part focuses on the application of estimation algorithms to target tracking.

Performance appraisal is crucial for understanding, developing and comparing various estimation algorithms. In particular, with the evolvement of estimation theory and the increase of problem complexity, performance appraisal is getting more and more challenging for engineers to make comprehensive conclusions. However, the existing theoretical results are inadequate for practical reference. The first part of this dissertation is dedicated to performance measures which include ***local performance measures***, ***global performance measures*** and ***model distortion measure***.

The second part focuses on application of the recursive best linear unbiased estimation (BLUE) or lineae minimum mean square error (LMMSE) estimation to nonlinear measurement problem in target tracking. Kalman filter has been the dominant basis for dynamic state filtering for several decades. Beyond Kalman filter, a more fundamental basis for the recursive best linear unbiased filtering has been thoroughly investigated in a series of papers by Dr. X. Rong Li. Based on the so-called quasi-recursive best linear unbiased filtering technique, the constraints of the Kalman filter ***Linear-Gaussian*** assumptions can be re-

laxed such that a general linear filtering technique for nonlinear systems can be achieved. An approximate optimal BLUE filter is implemented for nonlinear measurements in target tracking which outperforms the existing method significantly in terms of accuracy, credibility and robustness.

Chapter 1

Introduction

This dissertation consists of two parts. The first part deals with the performance appraisal of estimation algorithms. The goal is to explore and initiate systematic, comprehensive and scientifically defensible performance appraisal approaches. The second part is an application of estimation algorithms to target tracking. The goal is to promote the application of the recursive best linear unbiased estimation or linear minimum mean square error(LMMSE) estimation technique. We will apply the recursive best linear unbiased filtering method to handle nonlinear measurements in target tracking.

1.1 Appraisal of Estimation Performance

The ability to meaningfully assess the performance of an estimation algorithm is crucial for development and comparison of practical systems. With evolution and proliferation of estimation methods, such as least squares, the method of moments, maximum likelihood method, Bayes and empirical Bayes procedures, uniformly minimum variance unbiased estimation, risk-reduction approaches, robust technique and resampling technique, various algorithms have been developed. In addition, as estimation problems and the related systems of interest are getting more and more complex, huge difficulties are encountered for reconciling various measures of performance, effectiveness and robustness.

For example, study on performance metrics for target discrimination and tracking is a

research focal point sponsored by several agencies of the U.S. Department of Defense. Also, as pointed out in [58], the emergence of fusion strategies such as adaptive fusion and fusion management requires that Measures of Performance (MoPs), Effectiveness (MoEs), and Robustness (MoRs) to be examined with greater seriousness. Huge difficulties are encountered. Further, the optimization of an algorithm with respect to one particular criterion will frequently result in performance deterioration as measured by another criterion. Because of this, one might be led to define "global" metrics which measure the overall effectiveness of an algorithm. One could, for example, try to measure composite algorithm competence using some ad hoc weighted average of local metrics. In practice, however, such composite metrics are very difficult to interpret. U.S. Navy investigators on the (Joint Operational Tactical System) JOTS program concluded after extensive research on practical performance evaluation techniques that [59]:

"Examination of the host of MoEs as presented earlier does not always provide clear cut objective distinction of good versus poor performance. During the course of the evaluation, attempts were made to establish a single MoE as a composite [i.e. weighted sum] of correlator performance .... several problems were encountered with this MoE, however.... Clearly a better method of combining MoEs is required .... Such an approach should codify the philosophy of differing severity of the various types of errors, and nonlinearly increasing cost associated with those errors as they increase and confusion compounds."

Therefore, a systematic, comprehensive and scientifically defensible performance appraisal is clearly needed. Towards this goal, we identify important performance aspects and explore feasible performance appraisal measures. The first part of the dissertation will be dedicated to theoretical investigations of this topic.

The analysis and comparison of estimation performance have two main aspects.

One is to investigate the properties of estimators and to measure the information extracted from the corrupted data or observations, such as precision, unbiasedness, efficiency,

consistency, credibility and so on [21, 3, 57, 53, 13]. These investigations will provide a profile of estimation performance which is mainly determined by the formulation of the problem. For example, the performance bound depends on the likelihood function or the posterior distribution. The unbiasedness could be either a requirement of problem formulation or a desired property. These topics have been thoroughly addressed in the subject of statistics.

The other focuses on performance requirements for application purposes or meaningful interpretation of estimation results. It emphasizes the methods to measure the goodness of an estimation result relative to the application requirements. In contrast to the former, existing work dealing with the latter is quite limited, although substantial work has been done on performance metrics for evaluation of target tracking and data fusion algorithms [15, 58, 16, 36, 35].

We will focus on the following four topics which are organized as follows.

In Chapter 2, the characteristics of parameter inference are presented. The inherent gap between estimation problem formulation and application requirements is clarified. A brief discussion on estimation criteria versus estimation measures is presented. A general formulation of performance measures is introduced.

In Chapter 3, a group of local performance measures is introduced or proposed. The pros and cons of most popularly used measures are investigated. New performance measures are introduced, such as relative error measures, frequency count-type measures and pairwise comparison measures.

Chapter 4 focuses on global performance measures. In this chapter, the interaction between estimation criteria and estimator is examined. The performance spectrum is proposed to reveal such interrelationship. The desired estimation error probability density function (PDF) is introduced to provide a more comprehensive characterization of the estimation performance requirements. The performance measures relative to the desired PDF is proposed.

Chapter 5 is dedicated to a model distortion measure. In parameter estimation and state

filtering, model approximation is quite common in engineering research and development. It heavily affects estimation performance. And it is of great interest to have a model distortion measure to indicate the divergence or difference between the original model and the approximated one. The significance of a model distortion measure is illustrated by its application to two important problems—model set design and model probability behavior—associated with the multiple model (MM) algorithm.

The related work on local and global performance measures has been published in [38, 34, 33, 35, 36, 37, 64]. The main results on model distortion measure have been appeared in [65, 63].

1.2 Application of Estimation Algorithm to Target Tracking

The second part of the dissertation focuses on the application of the recursive best linear unbiased estimation (BLUE) to target tracking with nonlinear measurements. The Kalman filter has been popularly used for state estimation for several decades. Beyond the Kalman filter, a more fundamental basis is the recursive best linear unbiased filtering which has been thoroughly investigated in a series of papers by Dr. X. Rong Li [69, 45, 29, 46, 32, 30, 31, 24]. Based on the so-called quasi-recursive best unbiased linear filtering technique, the Kalman filter's **Linear-Gaussian** assumptions can be relaxed such that a general linear filtering scheme for nonlinear systems could be obtained.

In target tracking, target dynamics is usually modelled in the Cartesian coordinates, while the measurements are directly available in the original sensor coordinates which are highly nonlinear with respect to the target state. For a long while, this problem has been handled by various algorithms based on the Kalman filter but with serious flaws [2, 20, 47, 48, 49, 50, 55, 56]. An approximate BLUE filter is implemented to target tracking with nonlinear measurements which outperforms the existing method significantly in terms of

accuracy, credibility and robustness without increase in computation. This part is mainly based on our published papers [67, 66, 68].

In Chapter 6, we present the BLUE filtering technique. Its significance and contribution are addressed.

In Chapter 7, we apply the BLUE filtering technique to target tracking with nonlinear measurements. The fundamental flaws of existing measurement conversion methods are investigated. An approximate implementation of the BLUE filter is presented. The simulation studies are illustrated. We conclude that the BLUE filter is more fundamental than the Kalman filter. It outperforms the existing methods significantly.

Chapter 8 concludes the dissertation and provides directions to future work.

1.3 Contributions of the Dissertation

The contributions of this dissertation are summarized as below.

First, a thorough examination of the pros and cons of the most often used local performance measures is conducted. The theoretical and practical justifications and interpretations of these measures are clarified. In particular, the misinterpretation of the most popularly used root mean square error is elucidated. This part is mainly based on the papers [38, 34, 33, 35, 36].

Second, we initiate a systematic study for comprehensive and global estimation performance appraisal. A new interpretation of the optimality of the Bayesian estimator is presented which has a similarity to the lever principle. The interrelationship between the estimator and estimation criterion is revealed which can help to produce new optimal estimators with predictable estimation performance. Accordingly, an estimation performance spectrum is proposed for estimation performance comparison. Also, the desired estimation error probability density function is proposed such that various application concerns and requirements can be taken into account. The concentration and dispersion levels of an esti-

mation error distribution relative to the desired one can be measured as well by the proposed concentration and dispersion measures.

Third, a model distortion measure for statistical inference is proposed to indicate the model difference between the original model and the approximated one. The distortion measure turns out to be the Kullback-Leibler divergence. This measure is highly related to the behavior of the multiple model algorithm. It can be applied to the model set design problem for multiple model estimation. Based on the behavior of model probability of the multiple model estimation, the model probability update structure is extended to the multiple hypotheses comparison such that an online structure is proposed to carry out the multiple estimators' performance comparison and fusion. The illustrative example is demonstrated as well.

Fourth, we apply the recursive linear minimum mean square error estimation to target tracking with measurements in polar and spherical coordinates. An approximate LMMSE filter is achieved which outperforms the existing methods significantly in terms of the accuracy, robustness and credibility.

Part I

Performance Appraisal of Estimation Algorithms

Chapter 2

On Estimation Performance Measures

2.1 Parameter Estimation

Estimation theory is concerned with inferring the value of a quantity of interest from indirect, inaccurate and uncertain observations. It consists of four components: parameter space, observation space, probabilistic relation between the parameter space and the observation space, estimation criteria [57]. The probabilistic relation between parameter space and observation space determines the formulation of the problem. The parameter is estimated from observations based on the probabilistic relation, most often, through optimization procedure with an objective function defined by estimation criterion.

We demonstrate the inference structure by the following diagram.

The complete information on the uncertainty of parameter or state is carried by the posterior probability density function or likelihood function. For Bayesian estimation, the prior information of parameter is updated by using observation through Bayesian formula to obtain the posterior PDF. The estimator is a special point within the domain of the parameter or state which replaces its posterior PDF with estimation errors. This special point is determined through optimization procedure. In other words, from a posterior PDF to an estimator through optimization procedure, we obtain one possible value at the cost of losing the complete information and producing estimation errors. For classical estimation, the complete information about the parameter or state is represented by the likelihood

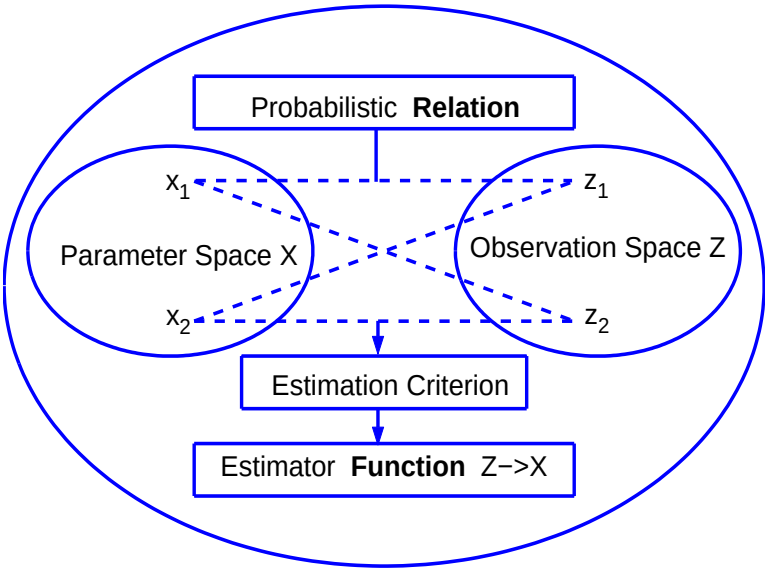


Fig. 2.1: Estimation Diagram

function. Usually, the maximum likelihood principle is employed to determine the estimator. Again, the likelihood function is replaced by a point with associated estimation errors.

For estimation, besides the estimator itself it is also necessary to measure or specify estimation error in a meaningful and comprehensive manner.

The estimation criterion serves to locate the optimal estimator. It needs to meet two objectives. First, it should measure the practical requirements adequately. Second, it should lend itself most easily to manipulation and computation from a technical and mathematical viewpoint. Most often, an estimation criterion is a compromise of these two objectives and mainly determined by the second one. It is even more instructive to retrospect the history of the minimum mean square error (MMSE) as an estimation criterion proposed by Gauss as follows [17]:

... The integral $\int_{-\infty}^{\infty} x^2\phi(x)dx$*, i.e., the average value of* x^2*, seems very suitable for defining and measuring, in a general way, the uncertainty of a system of observations.*

If one objects that this convention is arbitrary and does not appear necessary, we readily

agree. The question which concerns us here has something vague about it from its very nature, and can not be made really precise except by some principle which is arbitrary to a certain degree.

It is clear to begin with that the loss should not be proportional to the error committed, for under this hypothesis, since a positive error would be considered as a loss, a negative error would be considered as a gain; the magnitude of loss ought, on the contrary, to be evaluated by a function of the error whose value is always positive. Among the infinite number of functions satisfying this condition, it seems natural to choose the simplest, which is, without doubt, the square of the error, and is the way proposed above.

Clearly, according to the arguments given by Gauss, the selection of estimation criterion is arbitrary to a certain degree and can not reflect most of the application concerns. In order to evaluate the estimation errors or performance for application purposes, we need use performance measures. We discuss estimation criterion, estimation performance measures and performance analysis in details next.

2.2 Criteria vs. Measures

Performance optimization, performance evaluation, and performance analysis are closely related but different. Simply put, performance optimization is based on a theoretical criterion (objective function); performance evaluation is usually done in terms of some performance measures; and performance analysis aims at developing a performance model.

Practical measures for performance evaluation of estimation algorithms are closely related with theoretical criteria for performance optimization. They both should reflect the performance and fit well with the intended applications. However, they are not to be confused with each other. A performance measure is a ruler/metric that quantifies or sizes up the performance of the algorithm, while a criterion defines the optimality of the solution to the problem in some sense. Loosely speaking, most estimation criteria are theoretical

performance measures that serve as the objective functions for the estimation algorithms to optimize. As a result, a criterion should be tractable in that it lends itself to easy mathematical manipulations, rather than merely a meaningful performance measure. In other words, a criterion formulates the estimation problem as a tractable optimization problem and serves as the basis for algorithm development. While mathematical tractability is a highly desirable property for an optimality criterion—an intractable criterion is hardly useful—it is not an important consideration for a performance measure. Compared to estimation criterion, estimation performance measure only concerns the fitness of an estimator to application requirements. It can be very complex and comprehensive.

There is also a significant difference between performance *analysis* and performance *evaluation*. Performance analysis focuses on relationships between the performance and its key factors and hence relies heavily on analytic tools. This is not the case for performance evaluation, which simply evaluates or appraises the performance. Our focal point is performance evaluation, not performance analysis. Tractability of a performance measure is an important consideration in performance analysis, but not in performance evaluation, since a highly complex performance measure would most likely render performance analysis intractable, but it can only make performance evaluation costlier.

2.3 Formulation of Performance Measures

The following **notational convention** will be maintained throughout this dissertation. We refer to any quantity to be estimated as an ***estimatee***. It can be a time-invariant (or slowly varying) parameter, a (deterministic or random) process or signal, in particular, the state of a (deterministic or random) system. We will use the term ***estimator*** to mean both a parameter estimator and a filter. The measures are presented in a form suitable for parameter estimators directly, but are applicable to filters at any given time in a straightforward way. The estimatee, its estimate, and estimation error are denoted by x, $\hat{x}$, and $\tilde{x}$ or $\mathbf{x}$, $\hat{\mathbf{x}}$ and

$\tilde{\mathbf{x}}$, respectively. They are always assumed to be n-dimensional unless otherwise is stated explicitly. Subscript i stands for quantities pertaining to the ith run of a Monte-Carlo simulation. It is always assumed that $\tilde{x}_i$ and $\tilde{x}_j$ are independent for $i \neq j$. All default vectors are column vectors. The Euclidean norm of a vector a is denoted as

$$\|a\| = (a'a)^{1/2}$$

where a' stands for the transpose of the column vector a.

An estimator is a function of observations which provides a technical guess of a parameter or state. Since it is data-dependent, its performance is random and should be evaluated in a statistical sense. In other words, we need to know which estimator is more accurate or closer to the true value on average than the other. In order to meaningfully quantify such closeness for comparison purpose, the performance appraisal has to answer the following two questions:

1. How to quantify the distance between an estimatee and its estimator given each piece of data?

2. How to measure the performance difference statistically in terms of either marginal distribution or joint distribution when comparing multiple estimators?

Next, a general formulation of performance measure which incorporates the above two questions is presented.

First, we give the general performance measure for a single estimator. Define the performance measure of a classical estimator as below,

$$\begin{aligned} M = E_{\mathbf{z}}\left[m(\mathbf{x}, \hat{\mathbf{x}})\right] &= \int m(\mathbf{x}, \hat{\mathbf{x}}) f(\mathbf{z}; \mathbf{x}) d\mathbf{z} \\ &=< m(\mathbf{x}, \hat{\mathbf{x}}), f(\mathbf{z}; \mathbf{x}) > \end{aligned}$$

where parameter $\mathbf{x}$ is constant, $f(\mathbf{z}; \mathbf{x})$ is the likelihood function of $\mathbf{x}$, $m(\mathbf{x}, \hat{\mathbf{x}})$ is the measure function of the difference between estimatee $\mathbf{x}$ and estimator $\hat{\mathbf{x}}$, and usually is a quantification of accuracy, concentration probability or costs.

Similarly, a performance measure of a Bayesian estimator can be defined as

$$\begin{aligned} M &= E_{\mathbf{x},\mathbf{z}}[m(\mathbf{x},\hat{\mathbf{x}})] \\ &= \int\int m(\mathbf{x},\hat{\mathbf{x}}) f(\mathbf{x},\mathbf{z}) d\mathbf{x} d\mathbf{z} \\ &= \int \hat{M}_1(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} \\ &= \int \hat{M}_2(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \end{aligned}$$

where $\hat{M}_1 = \int\int m(\mathbf{x},\hat{\mathbf{x}}) f(\mathbf{x}|\mathbf{z}) d\mathbf{x} f(\mathbf{z}) d\mathbf{z}$ and $\hat{M}_1 = \int\int m(\mathbf{x},\hat{\mathbf{x}}) f(\mathbf{z}|\mathbf{x}) d\mathbf{z} f(\mathbf{x}) d\mathbf{x}$, $f(\mathbf{x},\mathbf{z})$ is the joint PDF of $\mathbf{x}$ and $\mathbf{z}$, again, $m(\mathbf{x},\hat{\mathbf{x}})$ is the deviation quantification function. Clearly, the Bayesian estimator's performance measure M is determined by $m(\mathbf{x},\hat{\mathbf{x}})$ and $f(\mathbf{x},\mathbf{z})$. $\hat{M}_1(\mathbf{z})$ is the performance measure based on the posterior PDF $f(\mathbf{x}|\mathbf{z})$, while $\mathbf{z}$ is given; $\hat{M}_2(\mathbf{x})$ is the performance measure based on the likelihood function $f(\mathbf{z}|\mathbf{x})$ conditioned on $\mathbf{x}$.

Second, when comparing multiple estimators, the difference between estimators relative to the true estimatee can be measured directly. In other words, how often or how much one estimator is better than the other relative to the true parameter $\mathbf{x}$ will be counted. In this case, the joint distribution information of estimators and estimatee is taken into consideration [17]. Here, we only stress pairwise comparisons.

The performance measure can be defined for direct pairwise comparison of two classical estimators relative to the parameter itself as

$$\begin{aligned} M(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x}) &= E\left[m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x})\right] \\ &= \int m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x}) f(\mathbf{z}; \mathbf{x}) d\mathbf{z} \end{aligned}$$

and for direct pairwise comparison of two Bayesian estimators as

$$\begin{aligned} M(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x}) &= E_{\mathbf{x},\mathbf{z}}[m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x})] \\ &= \int\int m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x}) f(\mathbf{x},\mathbf{z}) d\mathbf{x} d\mathbf{z} \end{aligned}$$

where $m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x})$ is the difference between $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j$ relative to the true estimatee $\mathbf{x}$, $i, j \in \{1, 2\}$ and $i \neq j$. For example, for pairwise comparison, a meaningful measure could be Pitman Closeness

$$m(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x}) = \begin{cases} 1 & if\ \hat{\mathbf{x}}_i \overset{\mathbf{x}}{\succ} \hat{\mathbf{x}}_j \\ 0.5 & if\ \hat{\mathbf{x}}_i \overset{\mathbf{x}}{=} \hat{\mathbf{x}}_j \\ 0 & if\ \hat{\mathbf{x}}_i \overset{\mathbf{x}}{\prec} \hat{\mathbf{x}}_j \end{cases}$$

where $\hat{\mathbf{x}}_i \overset{\mathbf{x}}{\succ} \hat{\mathbf{x}}_j$ and $\hat{\mathbf{x}}_i \overset{\mathbf{x}}{\prec} \hat{\mathbf{x}}_j$ denotes that $\hat{\mathbf{x}}_i$ prefers to $\hat{\mathbf{x}}_j$ and $\hat{\mathbf{x}}_j$ prefers to $\hat{\mathbf{x}}_i$ relative to $\mathbf{x}$, respectively. If $M(\hat{\mathbf{x}}_\mathbf{i} : \hat{\mathbf{x}}_\mathbf{j}; \mathbf{x})$ is larger than 0.5, we say $\hat{\mathbf{x}}_i$ is better than $\hat{\mathbf{x}}_j$. A detailed discussion will be given in next chapter.

Usually, the analytical form of performance measures can not be obtained and computer simulations are resorted to. The use of computer simulations for assessing estimation performance, although quite valuable for gaining insight and motivating conjectures, is hardly conclusive. Nevertheless, when the performance of some particular scenarios is of interest, simulation studies can provide a more direct comparison. Therefore, on the one hand, computer simulations are scenario-dependent; on the other, they also give a direct evaluation under a specific scenario.

When Monte Carlo simulation is employed, all the performance measures simply become

$$M = E[m(*)] = \frac{1}{N} \sum_{i=1}^{N} m^l(*)$$

where $m^i(*)$ is the measure of the difference between $\mathbf{x}$ and its estimator $\hat{\mathbf{x}}$ or the difference between $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j$ relative to the true estimatee $\mathbf{x}$, $i, j \in \{1, 2\}$ and $i \neq j$ on the lth run of N Monte Carlo runs.

Local performance measures, global performance measures and model distortion measure all follow this general formulation. They differ from each other in the specific definitions of $m^i(*)$ so as to emphasize different aspects of estimation performance concerns.

Chapter 3

Local Performance Measures

This chapter deals with performance measures that are local in the sense that they rely on relatively local characteristics (e.g., moments) of the distribution of the estimation error. Interpretation of existing performance measures will be examined and new performance measures will be proposed. They are classified into the following four categories: absolute error measures (without a reference), relative error measures (with a reference), frequency counts (of some events), and pairwise comparison measures. The first three categories are devised to measure the performance of a single estimator. The last category is for comparison purpose where the joint distribution of the two estimators relative to the parameter is involved.

This Chapter is organized as follows. In Sec. 3.1, we first discuss several measures of *absolute error* of estimation, including the root-mean-square error (RMSE) and what we called average Euclidean error (AEE). Their pros and cons are presented. It is advocated that the widely-used RMSE should be replaced by AEE in many applications. Also, two new measures of absolute error—geometric average error and harmonic average error—are introduced. Secs. 3.2 and 3.3 are dedicated entirely to new measures. In Sec. 3.2 we present several measures of *relative error* in that they are normalized estimation errors with respect to three references, respectively: magnitude of the estimatee, error of prior mean, and measurement error. We propose in Sec. 3.3 three new metrics—success rate, failure rate, and

concentration probability—of an estimation algorithm, which measure how concentrated the estimates are around the estimatee. They stem from frequency counts of estimates within some neighborhood of the estimatee, especially within what we called success region and feasible region. They are complementary to the error measures and are useful for some applications. Measures for pairwise comparison—Pitman closeness measure, loss measure and gain measure—are provided in Sec. 3.4. They compare two estimators relative to the estimatee case by case directly. The performance difference of these two estimators is concluded by counting on the difference in each case. Illustrative examples are presented in Sec. 3.5. Sec. 3.6 is concluding remarks.

3.1 Absolute Error Measures[0]

3.1.1 RMS Error and Average Euclidean Error

RMSE. By far the most popular measure of *estimation error* is the *root-mean-square* (*RMS*) *error*, or ***RMSE*** for short, defined by

$$\mathrm{RMSE}(\hat{x}) = \left(\frac{1}{M} \sum\nolimits_{i=1}^{M} \|\tilde{x}_i\|^2 \right)^{1/2} \tag{3.1}$$

where $\|\tilde{x}_i\| = (\tilde{x}_i' \tilde{x}_i)^{1/2}$.

The main nice feature of RMSE is that it is the finite-sample approximation of the *standard error* $\sqrt{E[\tilde{x}'\tilde{x}]}$, which is closely related with standard deviation. For scalar *unbiased* estimators, it is actually the most natural finite-sample approximation of standard deviation of estimation error. Since standard deviation is an important parameter for probabilistic analysis, RMSE is useful for probabilistic analysis in the scalar case. For example, if $\tilde{x}$ is scalar, zero-mean, and Gaussian distributed with variance σ^2, then $P\{|\tilde{x}| < \mathrm{RMSE}\} \approx P\{|\tilde{x}| < \sigma\} = 0.683$.

[0] © [2006] IEEE. Reprinted, with permission, from X. R. Li, Z. L. Zhao, "Evaluation of Estimation Algorithms Part I: Local Performance Measures". *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, Jul. 2006.

AEE. An emphasis of this chapter is a proposal to replace in many situations the use of RMSE with the following ***average Euclidean error*** (***AEE***):

$$\mathrm{AEE}(\hat{x}) = \frac{1}{M}\sum_{i=1}^{M} \|\tilde{x}_i\| \tag{3.2}$$

The term "Euclidean error" stems from the concept of Euclidean distance or Euclidean norm. AEE (in the scalar case) is sometimes called *mean absolute error* (MAE), but this is not recommended because the term "absolute error" is ambiguous for a vector. Also, it may have a misleading implication that it is an absolute error, as opposed to a relative error.

Denote by $e = \|\tilde{x}\|$ the Euclidean norm of the estimation error, which is a random variable, and let $\bar{e}$ and σ_e^2 be the mean and variance of e. Then

$$E[\mathrm{AEE}(\hat{x})] = \bar{e}, \quad \mathrm{var}[\mathrm{AEE}(\hat{x})] = \sigma_e^2/M \tag{3.3}$$

where $\sigma_e^2 = \mathrm{var}(e) = E(e^2) - \bar{e}^2 = \mathrm{tr}(\Sigma) + \mu'\mu - \bar{e}^2$, and μ and Σ are the mean and covariance of estimation error $\tilde{x}$.

The mean Euclidean error $\bar{e}$ is clearly a good indicator for the estimation error and, albeit unknown, it can be estimated. Since AEE and $\bar{e}$ are respectively the sample average and the true mean of $\|\tilde{x}\|$, AEE is the most natural approximation/estimator of $\bar{e}$.

RMSE vs. AEE. RMSE and AEE focus on large errors in that their values are dominated by the large error terms. For example, if all 100 terms are around 1 except that one term is 400, then AEE would be about 5, but RMSE would be approximately 400, that is, determined by one term while all the other 99 terms are essentially ignored! We say that such measures are ***pessimistic*** since they pay so much attention to how *bad* the performance is. Clearly, RMSE is much more pessimistic than AEE. This example also reveals that RMSE could be very unreasonable since it is highly distorted—it penalizes too severely on large errors (i.e., bad behavior) and virtually ignores small errors (i.e., good behavior). This is highly undesirable or even unacceptable for many applications, particularly as a performance measure. This is part of the reason why a special treatment is needed while using RMSE in

the case involving a few huge errors due to, say, filtering divergence: What is the RMSE of a filter if it diverges on one or few of many runs?

Geometrically speaking, AEE($\hat{x}$) is the arithmetic average of the Euclidean distance between x and $\hat{x}$, while RMSE($\hat{x}$) $= \frac{1}{\sqrt{M}}(\sum_{i=1}^{M} \|\tilde{x}_i\|^2)^{1/2}$ is one $\sqrt{M}$th of the Euclidean distance between the (Mn)-dimensional vectors $[x_1', x_2', \ldots, x_M']'$ and $[\hat{x}_1', \hat{x}_2', \ldots, \hat{x}_M']'$, which hardly has a physical interpretation, where x_i and $\hat{x}_i$ are the estimatee and estimator on run i. As such, AEE is truly the average distance between the estimate and the estimatee in our physical (i.e., Euclidean) space. By contrast, RMSE does not have a simple *physical* interpretation—it is not actually an average distance in our Euclidean space, although it is often so interpreted incorrectly. Clearly, an erroneous interpretation invites confusion and mistakes.

Consider the case of a scalar measurement z with mean $\bar{z}$. Mean deviation $E[|z - \bar{z}|]$ is no doubt the most natural abstraction of the practical concept of measurement error since measurement errors in most situations are actually average absolute errors. This should be clear from the calibration process of a measurement instrument. On the other hand, RMSE reflects the standard error (or standard deviation $\sqrt{E[(z - \bar{z})^2]}$ in the unbiased case). Due to the popularity of standard deviation, measurement errors are more often expressed in terms of RMSE or standard deviation than mean deviation. This is mainly because an analysis based on the former is more tractable mathematically than on the latter. A problem in practice is that many practitioners do not understand or are not even aware of the difference—the former is also interpreted as the average error in magnitude.

The popularity of RMSE arises mostly from the fact that it is in general *the* best finite-sample approximation of the standard error, which is the most popular optimality criterion in terms of error. The widespread use of standard error (in fact, MSE) as a theoretical criterion is well justified by its mathematical tractability. Such a justification is invalid as far as practical metrics for performance evaluation are concerned since mathematical

tractability is no longer a concern here. Instead, a clear and correct interpretation of the metrics is essential.

It follows from the Kolmogorov strong law of large numbers, RMSE^2 and AEE tend (almost surely) to mean-square error $E(e^2)$ and *mean Euclidean error* $\bar{e}$, respectively, as $M \to \infty$. They are consistent estimators of $E(e^2)$ and $\bar{e}$, respectively.

In the scalar Gaussian case, the mean deviation $\bar{e} = E[|\tilde{x}|]$ and the standard deviation σ are related by $E[|\tilde{x}|] = \sigma\sqrt{2/\pi} \approx 0.8\sigma$. Given an AEE in such a case, it can be converted to RMSE and thus a probabilistic analysis can still be carried out since such an analysis almost always assumes a Gaussian error. Note that for the vector case, it is usually not straightforward to carry out a probabilistic analysis based on RMSE without additional assumptions, even in the Gaussian case.

It is well known [23] that the conditional mean and median (i.e., the mean and median of the posterior distribution) minimizes, respectively, MSE and mean Euclidean error and hence RMSE and AEE. Consequently, the use of RMSE implicitly favors the conditional mean (i.e., MMSE) based estimators rather than any other estimator. This is usually a weakness for performance evaluation, but could be a strength for some other purposes, especially for verifying that an MMSE-based algorithm is properly designed and/or implemented. By contrast, AEE favors estimators that minimize mean Euclidean error, which is, however, rarely used as an optimality criterion due to the difficulty involved.

On the basis of the above arguments, *we recommend that RMSE be replaced by AEE when one is mainly concerned with the magnitude of the estimation errors*, since the latter has a more direct and natural interpretation. When a probabilistic analysis is needed, the RMSE is usually more convenient.

3.1.2 Harmonic Average Error

As explained before, RMSE and AEE are dominated by their large individual terms and hence focus on bad behavior. In this sense they are pessimistic measures. However, we may want to pay more attention to good behavior in some situations, that is, to use *optimistic* measures.[1] For this purpose, we propose the following ***harmonic average error*** (***HAE***):

$$\mathrm{HAE}(\hat{x}) = \left(\frac{1}{M} \sum\nolimits_{i=1}^{M} \|\tilde{x}_i\|^{-1} \right)^{-1} = \frac{M}{1/\|\tilde{x}_1\| + \cdots + 1/\|\tilde{x}_M\|} \tag{3.4}$$

based on the well-known harmonic average. Clearly, HAE is dominated by the small error terms and hence is an ***optimistic*** measure. It focuses on how *good* (rather than how *bad*, as with a pessimistic measure) the performance of an estimator is.

Harmonic average has the following minmax property. Consider the problem of approximating (estimating) the true error $e = \|\tilde{x}\|$ that is bound to be within the interval $[a, b]$. Denote by $\hat{e}$ an approximate of e. Then the harmonic average $H(a, b)$ of a and b yields the minimum of the greatest possible relative error $|e - \hat{e}|/e$, that is, $H(a, b) = \arg\min_{a \le \hat{e} \le b} \max_{a \le e \le b}(|e - \hat{e}|/e)$, while the arithmetic average $A(a, b)$ of a and b yields the minimum of the greatest possible absolute error, that is, $A(a, b) = \arg\min_{a \le \hat{e} \le b} \max_{a \le e \le b} |e - \hat{e}|$ [9].

3.1.3 Geometric Average Error

Neither RMSE and AEE nor HAE is good as a *balanced* measure for performance evaluation and, in particular, as a metric for "*average*" error in magnitude. As far as the *average* error in magnitude is concerned, for many problems it is reasonable to expect that any large error (bad behavior) should be possibly balanced by a sufficiently small error (good behavior) and vice versa. From this standpoint, the use of the arithmetic average as an average error in

[1] From the Bayesian estimation viewpoint, the MAP and maximum likelihood (ML) estimation minimizes an optimistic measure—the expected value of a small (infinitesimal in theory) "golf hole" of estimation error [23]. This is in contrast to MMSE and least-squares estimation that minimizes a pessimistic measure, that is, the quadratic error.

magnitude is flawed. For example, assume AEE is equal to a. Then, an error that is larger than $2a$ cannot be balanced by any sufficiently small error. The situation is reversed for HAE.

In view of the above, we propose the following ***geometric average error*** (***GAE***): $\mathrm{GAE}(\hat{x}) = \left(\prod_{i=1}^{M} \|\tilde{x}_i\|\right)^{1/M}$ based on the well-known geometric average, which is better computed recursively

$$\mathrm{GAE}_M(\hat{x}) = \mathrm{GAE}_{M-1}(\hat{x}) \left[\|\tilde{x}_M\|/\mathrm{GAE}_{M-1}(\hat{x})\right]^{1/M}$$

or through its logarithm for numerical reasons

$$\log[\mathrm{GAE}(\hat{x})] = \frac{1}{M} \sum\nolimits_{i=1}^{M} \log \|\tilde{x}_i\| \tag{3.5}$$

$\mathrm{GAE}_M(\hat{x})$ is defined as zero if one or more $\|\tilde{x}_i\|$ is zero. Clearly, it does not suffer from the flaw mentioned above—any large error can be balanced by a sufficiently small error and vice versa.

Denote by $e = \|\tilde{x}\|$ the random variable that is the Euclidean norm of the estimation error. It follows from Hölder's inequality that [18]

$$\bar{e} = E[\mathrm{AEE}(\hat{x})] = E[\mathrm{GAE}_1(\hat{x})] \geq E[\mathrm{GAE}_M(\hat{x})] \geq E[\mathrm{GAE}_L(\hat{x})] \geq \mathrm{GM}(e), \quad L > M > 1$$

with all inequalities strict if $\bar{e} < \infty$ and $\mathrm{GM}(e) > 0$, where $\mathrm{GM}(e) = \exp[E(\ln \|\tilde{x}\|)]$ is the geometric mean of the Euclidean norm $\|\tilde{x}\|$ of the random error $\tilde{x}$ in theory. As shown in [10], if $e = \|\tilde{x}\|$ has zero probability outside the interval $[a, b]$ then $\sigma_e^2/b \leq \bar{e} - \mathrm{GM}(e) \leq \sigma_e^2/a$, where $\sigma_e^2 = \mathrm{var}(e)$. Also, by the Kolmogorov strong law of large numbers, $\mathrm{GAE}(\hat{x})$ tends (almost surely) to $\mathrm{GM}(e)$ as $M \to \infty$. In other words, $\mathrm{GAE}(\hat{x})$ has an expected value that converges monotonically to $\mathrm{GM}(e)$ from above as the number M of Monte-Carlo runs increases, and thus it is a consistent, albeit biased, estimator of the geometric mean. In addition, the application of the central limit theorem to $\log \|\tilde{x}_1\|, \ldots, \log \|\tilde{x}_M\|$ implies that $\log[\mathrm{GAE}(\hat{x})]$ is asymptotically Gaussian (normal) with $\mathcal{N}(E(\log \|\tilde{x}\|), \mathrm{var}(\log \|\tilde{x}\|)/M)$ as $M \to \infty$ and hence $\mathrm{GAE}(\hat{x})$ is asymptotically log-normal.

3.2 Relative Error Measures[0]

A relative error is one that is relative to some reference. There are many choices for the reference. Relative error often reveals better the inherent error characteristics of an estimator than the absolute error. For example, it is usually reasonable to expect that relative error of an estimator is less variant than the absolute error as the magnitude of the estimatee varies. Given two estimators and their performance for two different problems, *respectively*, it would be misleading to use any absolute error measure for their performance comparison, but relative error measures can be used. For such reasons, we recommend evaluating the performance of estimators in terms of a relative error in most cases, although the literature is full of performance evaluation in terms of absolute error.

Clearly, estimation error $\tilde{x}$ depends on the magnitude of the estimatee x and the accuracy of the data (measurements) z (as well as the prior distribution of x for a Bayesian estimator) as the input to the estimator $\hat{x}(z)$. As a result, probably the most natural relative error is $\tilde{x}/\|x\|$ or $\|\tilde{x}\|/\|x\|$. Another good choice of the reference is the measurement error; that is, we express the estimation error $\tilde{x}$ in terms of the measurement error. We deal with measures for such relative errors in this section.

3.2.1 Estimation error relative to estimatee

All the measures discussed in the previous section were given in terms of the absolute error $\tilde{x}$. This is often not desirable. For instance, an absolute error of $\tilde{x} = 1$ is only 1% for an estimatee of $x = 100$ but 50% for an estimatee of $x = 2$.

The relative error $\tilde{x}/\|x\|$ versions of the measures discussed in the previous section are simply given by their corresponding formulas with $\tilde{x}_i$ replaced by $\tilde{x}_i/\|x_i\|$. The absolute error norm $\|\tilde{x}_i\|$ used in these measures are replaced by the relative error norm $\|\tilde{x}_i\|/\|x_i\|$.

For example, the *RMS relative error* (RMSRE) and *average relative error* (ARE) are

$$\text{RMSRE}(\hat{x}) = \left(\frac{1}{M} \sum_{i=1}^{M} \|\tilde{x}_i\|^2 / \|x_i\|^2 \right)^{1/2} \tag{3.6}$$

$$\text{ARE}(\hat{x}) = \frac{1}{M} \sum_{i=1}^{M} \|\tilde{x}_i\| / \|x_i\| \tag{3.7}$$

where x_i is the estimatee on run i (i.e., the ith realization of the estimatee x). Such measures are very simple but limited. Relative measures presented below are more appealing.

Improvement of posterior over prior. We now introduce a metric of relative error for Bayesian (and recursive) estimation. We call it ***Bayesian estimation error quotient*** (***BEEQ***). It quantifies the improvement, in terms of error, of a Bayesian estimator $\hat{x}$ over the prior mean $\bar{x}$ or of the updated estimate $\hat{x}$ of a recursive estimator over the predicted estimate $\bar{x}$. We define it as

$$r^*(\hat{x}) = \frac{\text{AEE}(\hat{x})}{\text{AEE}(\bar{x})} = \frac{\sum_{i=1}^{M} \|x_i - \hat{x}_i\|}{\sum_{i=1}^{M} \|x_i - \bar{x}\|} \tag{3.8}$$

where x_i and $\hat{x}_i$ are the ith realizations of x and $\hat{x}$, respectively, and $\bar{x}$ is the prior mean (or prediction) of x. We do not recommend the use of RMSE here, such as $r^*(\hat{x}) = \frac{\text{RMSE}^*(\hat{x})}{\text{RMSE}^*(\bar{x})} = \left(\sum_{i=1}^{M} \|x_i - \hat{x}_i\|^2 / \sum_{i=1}^{M} \|x_i - \bar{x}\|^2 \right)^{1/2}$, because of the shortcomings of the RMSE (e.g., undue dominance of large terms and lack of a natural interpretation) and the fact that merits of RMSE are largely irrelevant here.

One may be tempted to define BEEQ as $\frac{1}{M} \sum_{i=1}^{M} r_i$, that is, the *arithmetic average* of individual error quotients $r_i = \|x_i - \hat{x}_i\| / \|x_i - \bar{x}\|$. This definition is not appropriate. As for any arithmetic average of a positive quantity, particularly a ratio, the average will be dominated by its large terms (i.e., by the cases in which errors are amplified). In other words, good (small) terms should be counted but would be essentially ignored in this definition. What is much worse and in fact fatal is that, as a result of this drawback, this measure is significantly greater than 1 even for many optimal estimators, which is misleading and unacceptable since a good estimator should have a BEEQ significantly smaller than 1. Instead,

the *geometric average* is much more appropriate here and thus BEEQ is better defined as $r(\hat{x}) = \left(\prod_{i=1}^{M} r_i\right)^{1/M}$, which is better computed through its logarithm for numerical reasons:

$$\log[r(\hat{x})] = \frac{1}{M}\sum_{i=1}^{M} \log r_i, \quad r_i = \frac{\|x_i - \hat{x}_i\|}{\|x_i - \bar{x}\|} \tag{3.9}$$

Error amplification and error reduction are thus balanced.

BEEQ quantifies the contribution of the data to Bayesian (and recursive) estimation. As the error $\|x - \bar{x}\|$ increases, an approximately constant BEEQ indicates that data has an insignificant contribution to estimation, because the measurement error is in effect much larger than $\|x - \bar{x}\|$; the degree of drop in BEEQ reflects that of increase in the contribution of the data; BEEQ will not increase unless an increase in $\|x - \bar{x}\|$ will lead to an even larger increase in the measurement error.

It follows from the properties of arithmetic and geometric averages (see, e.g., [9]) that BEEQ is always bounded by the smallest and largest individual error quotients: $r_{\min} \leq r(\hat{x}) \leq r_{\max}$, $r_{\min} \leq r^*(\hat{x}) \leq r_{\max}$, where $r_{\min} = \min\{r_1, \ldots, r_M\}$ and $r_{\max} = \max\{r_1, \ldots, r_M\}$.

3.2.2 Estimation Error Relative to Measurement Error

Clearly, the estimation accuracy depends on the accuracy of the data (measurements) as the input to the estimator. In fact, a primary benefit of estimation is that estimates are more accurate than measurements (after they are converted to the same space). As a measure of this accuracy improvement, we introduce a metric, named ***estimate-measurement error ratio*** (***EMER***).

Assume that a (vector-valued) measurement z is related to the estimatee x by $z = g(x, v)$, where v is the measurement error. Let $h(x) = E[g(x, v)|x] = E[z|x]$. For the additive zero-mean noise case with $z = a(x) + v$, we have $h(x) = a(x)$. Then, EMER is defined by

$$\rho^*(\hat{x}) = \frac{\mathrm{AEE}^*(h(\hat{x}))}{\mathrm{AEE}^*(z)} = \frac{\sum_{i=1}^{M} \|h(x_i) - h(\hat{x}_i)\|}{\sum_{i=1}^{M} \|h(x_i) - z_i\|} \tag{3.10}$$

where AEE* represents AEE in the measurement space:

$$\begin{aligned} \text{AEE}^*(h(\hat{x})) &= \frac{1}{M}\sum_{i=1}^{M}\|h(x_i) - h(\hat{x}_i)\| \\ \text{AEE}^*(z) &= \frac{1}{M}\sum_{i=1}^{M}\|h(x_i) - z_i\| \end{aligned}$$

That is, EMER is defined as—after converting x and $\hat{x}$ to the measurement space—the ratio of the average distance between x and $\hat{x}$ over the average distance between x and z. Note that the use of AEE* is preferable to that of RMSE* since RMSE is too much dominated by the large error terms, although measurement errors are often given in terms of standard deviation.

Similar to BEEQ, the *geometric average* of individual estimate-measurement error ratios is more appealing and thus EMER $\rho(\hat{x})$ is better defined by:

$$\log[\rho(\hat{x})] = \frac{1}{M}\sum_{i=1}^{M}\log\rho_i, \quad \rho_i = \frac{\|h(x_i) - h(\hat{x}_i)\|}{\|h(x_i) - z_i\|} \tag{3.11}$$

Also similar to BEEQ, EMER is always bounded by the smallest and largest individual error ratios: $\rho_{\min} \le \rho(\hat{x}) \le \rho_{\max}$, $\rho_{\min} \le \rho^*(\hat{x}) \le \rho_{\max}$, where $\rho_{\min} = \min\{\rho_1, \ldots, \rho_M\}$ and $\rho_{\max} = \max\{\rho_1, \ldots, \rho_M\}$.

Clearly, we expect that EMER < 1 for a good estimator and the smaller the EMER the better the estimator.

The ultimate goal of estimation is usually to approximate the estimatee as closely as possible. The closeness is usually better measured in the estimatee space directly, rather than in the measurement space, as the above EMER does. In some situations, such as positioning or localization applications, the mapping $h(\cdot)$ defined above is invertible; that is, the mapping $h^{-1}(\cdot)$ from the measurement space to the estimatee space is known. Then, a better definition of EMER is

$$\log[\rho_x(\hat{x})] = \frac{1}{M}\sum_{i=1}^{M}\log\rho_i^x, \quad \rho_i^x = \frac{\|x_i - \hat{x}_i\|}{\|x_i - h^{-1}(z_i)\|} \tag{3.12}$$

or alternatively

$$\rho_x^*(\hat{x}) = \frac{\mathrm{AEE}(\hat{x})}{\mathrm{AEE}(h^{-1}(z))} = \frac{\sum_{i=1}^{M} \|x_i - \hat{x}_i\|}{\sum_{i=1}^{M} \|x_i - h^{-1}(z_i)\|} \tag{3.13}$$

where $\mathrm{AEE}(h^{-1}(z)) = \mathrm{AEE}(\hat{x})|_{\hat{x}=h^{-1}(z)}$. It quantifies the amount of improvement an estimator has on the estimate provided by the measurement directly.

3.3 Frequency Counts of Concentration[0]

Simply put, each metric discussed so far provides a ruler to measure how large the estimation error is: those of Sec. 3.2 are normalized with respect to a variety of references, while those of Sec. 3.1 are not normalized. In contrast with these error measures, the performance measures introduced in this section stem from counting the occurrence of some events.

3.3.1 Success Region and Success Rate

Recall that the least-squares (LS) estimation and MMSE estimation differ from the maximum likelihood (ML) estimation and MAP estimation in their underlying ideas. The former seeks an estimator that has the smallest error, while the latter uses the "most frequently occurred" value of the estimatee as the estimator. As such, ML and MAP estimators may have larger *average* errors but a higher probability of staying close to the true estimatee. This has important implications. This fundamental difference should be kept in mind while deciding on which estimation method to employ for a particular application.

Take the problem of estimation for interception/weapon control as an example. Here what is important is that the estimates should be within a close neighborhood of the estimatee (the kill zone of the weapon) such that the target can be destroyed, rather than the average distance between them. Consider two estimators. The first delivers most estimates within the kill zone but has some really bad misses. The second misses the kill zone quite often but has few really bad misses and thus has a smaller average error (squared). Clearly, it is the first estimator, not the second one, that should be chosen for this application. This example

clearly demonstrates another important point: Use of RMSE alone for a comparison between MMSE (or LS) and MAP (or ML) estimators could be quite unfair, which is unfortunately quite common in practice. In fact, it should be clear that the use of RMSE to evaluate estimators for such applications as weapon control and interception is not so appropriate.

With such applications in mind, we introduce a concept of a success of an estimate and a corresponding region named success region. An estimator has a ***success*** given a particular set of data if its estimate falls inside the *success region* R_s. Clearly, a good example of a success region is the kill zone of a weapon, say, a missile—a region in which a target will be destroyed by the missile. If an estimate is in the success region, a missile detonated at the location of the estimate will destroy the target, hence a success.

The actual success region is clearly application dependent. We introduce the following general definition of a success region for the cases when the actual success region is not known for the application at hand. The ***success region*** R_s of an estimator $\hat{x}$ is a function of the estimatee x and data z defined by

$$R_s = \{\hat{x}_i : \rho_i(\hat{x}_i) \leq \delta_s\} \tag{3.14}$$

where $\rho_i = \|h(x_i) - h(\hat{x}_i)\|/\|h(x_i) - z_i\|$ is the individual estimate-measurement error ratio (or better use $\rho_i^x = \|x_i - \hat{x}_i\|/\|x_i - h^{-1}(z_i)\|$ if $h(\cdot)$ is invertible) for some small δ_s (say, $\delta_s = 0.1$). Loosely speaking, for a given x a success region is a ball $\|\tilde{x}\| \leq \delta_s\|h(x) - z\|$ of a radius $\delta_s\|h(x) - z\|$ centered at x, meaning that an estimate with an error at most δ_s times the measurement error is treated as a success (e.g., a hit). For a random x, the center of the ball is also random.

This definition of success region certainly has drawbacks. For instance, it depends on data z, while for some applications, e.g., the above interception problem, the success region should not depend on the data. However, to expect a high success rate of any estimator based on a set of poor data is unrealistic. This definition is recommended only when the actual success region is not known.

The ***success rate*** of an estimator is defined as the frequency (percentage) that its estimates $\hat{x}_i$ are *inside* the corresponding success region. It is a finite-sample approximation of the ***success probability*** $P\{\hat{x} \in R_s\}$.

3.3.2 Feasible Region and Failure Rate

Similarly, it is also beneficial to introduce the concept of failure for an estimator. We say an estimator has a ***failure*** if its estimate falls outside a region called feasible region. A *feasible region* R_f is one (in the estimatee space) that an estimate falling outside it would result in a serious consequence—the estimator would usually not be able to recover by itself (due to, e.g., divergence). For example, in radar tracking, the so-called radar gate is a region (in the measurement space) in which a target can be tracked in the track mode (rather than the search mode) of the radar. It can be treated as the feasible region because the target will be lost if its state estimate falls outside it.

Clearly, the actual feasible region is also application dependent. We introduce the following general definition of a feasible region for the cases when the actual feasible region is not known for the application at hand. The ***feasible region*** R_f of an estimator $\hat{x}$ is a function of the estimatee x and data z defined by

$$R_f = \{\hat{x}_i : \rho_i(\hat{x}_i) \leq \delta_f\} \tag{3.15}$$

for some large δ_f (say, $\delta_f = 10$). Loosely speaking, a feasible region is a ball $\|\tilde{x}\| \leq \delta_f \|h(x) - z\|$ of a radius $\delta_f \|h(x) - z\|$ centered at x, meaning that an estimate with an equivalent error in the measurement space greater than δ_f times the measurement error is treated as a failure. Note that the complement (outside) of the feasible region is deemed the failure region.

The ***failure rate*** of an estimator is defined as the frequency (percentage) that its estimates $\hat{x}_i$ are *outside* the corresponding feasible region. It is a finite-sample approximation of the ***failure probability*** $P\{\hat{x} \notin R_f\}$.

3.3.3 Concentration Region and Probability

The success (or failure) rate is with respect to a specified success (or feasible) region. It is sometimes more convenient to examine how large the "success" region of an estimator is given a required success rate. Consider the region

$$R_\delta = \{\hat{x}_i : \rho_i(\hat{x}_i) \leq \delta\} \tag{3.16}$$

If $\delta > 0$ is such that, say, 80% of the estimates fall inside R_δ, then we can refer to R_δ as the (normalized[2]) 80% ***concentration region*** of the estimator and δ as the 80% ***radius***. As such, an estimator with a smaller radius δ for a given required success rate (say, 80%) delivers estimates closer to the estimatee. Given an arbitrary region R_δ, $P\{\hat{x} \in R_\delta\}$ is the ***concentration probability*** of the estimator on R_δ. Its finite-sample approximation is the ***concentration percentile***. Note that although somewhat similar to the concept of a confidence interval, the concentration region relies on the (individual) estimate-measurement error ratio.

3.4 Measures for Pairwise Comparison

When two estimators are employed simultaneously for competition, e.g., in a game, their preference will depend on the pairwise comparison in all cases. In other words, their performance difference involves the joint information of these two estimators relative to the truth. If using measures involves only the marginal information of each estimate, the conclusions could be misleading. For example when two absolute estimation error sequence are given as $\tilde{\mathbf{x}}_1 = [1\ 2\ 3]$ and $\tilde{\mathbf{x}}_2 = [2\ 3\ 1]$, no matter which of the performance measures in Secs. 3.1, 3.2 and 3.3 is used, the difference can not be discriminated. The need for performance measures for pairwise comparison arises in such situations.

[2]Normalized in the sense that they are relative to the measurement error since a ratio is used.

3.4.1 Pitman Closeness Measure

The ***Pitman closeness measure*** (PCM) is based on the probabilities of the relative closeness of competing estimators to the parameter [17]. We explain it next. Let $m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$ denote the measure of the difference between $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ relative to the parameter $\mathbf{x}$, and

$$m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) = \begin{cases} 1 & if\ \hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2 \\ 0.5 & if\ \hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2 \\ 0 & if\ \hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2 \end{cases},$$

where $\hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2$ means that $\hat{\mathbf{x}}_1$ is preferred to $\hat{\mathbf{x}}_2$ relative to $\mathbf{x}$, for example, $||\hat{\mathbf{x}}_1 - \mathbf{x}|| < ||\hat{\mathbf{x}}_2 - \mathbf{x}||$. The Pitman closeness measure of $\hat{\mathbf{x}}_1$ superior to $\hat{\mathbf{x}}_2$ relative to $\mathbf{x}$**,** is defined as follows

$$\begin{aligned} \text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) &= E\left[m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})\right] \\ &= Pr(\hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2) + 0.5 * Pr(\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2) \end{aligned}$$

where $E[*]$ stands for the expectation of $*$ and $Pr(*)$ is the probability of $*$. Clearly

$$\text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) = 1 - \text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$$

When $\text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) > 0.5$ or equivalently $\text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) > \text{PCM}(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$, we say $\hat{\mathbf{x}}_i$ is better than $\hat{\mathbf{x}}_j$. Pitman closeness measure is the relative frequency in which the estimator $\hat{\mathbf{x}}_1$ will be superior to its competitor $\hat{\mathbf{x}}_2$.

PCM is generally applicable and robust against the choice of error measure function. More specifically, the Pitman closeness always exists; it does not change when different measures, say the L^p norms, of the estimation error are used. It uses the joint information of two competing estimators, and more information can be extracted. Some peculiar phenomena about the Pitman closeness have been observed by Rao [17]. He successfully argued for the Pitman closeness as an intrinsic measure of acceptability and presented many diverse univariate examples in which shrinking the MSE of an unbiased estimator to an MMSE estimator did not yield a better estimator in the sense of Pitman closeness [17]. Therefore, the Pitman closeness reveals some deep insights about the estimation performance which can

not be revealed by MSE. Also, it can discriminate two estimators that are indistinguishable by measures based on individual information.

However, because PCM is based on the joint information of pairwise estimators, this invokes some deficiencies [17, 51, 53]. For example, they may take into account a subset of the sample space that occurs only with probability slightly greater than 0.5; the "winner" can then be arbitrarily bad when the observation does not belong to this subset. In other words, they ignore many cases that are still relevant. Due to this peculiarity, they are not transitive. In other words, if $\hat{\mathbf{x}}_1$ is better than $\hat{\mathbf{x}}_2$ and $\hat{\mathbf{x}}_2$ is better than $\hat{\mathbf{x}}_3$, according to PCM, we can not conclude $\hat{\mathbf{x}}_1$ is better than $\hat{\mathbf{x}}_3$. Consequently, when more than two estimators are compared, it might be that none is dominant in the sense of PCM, even though under some special conditions, the transitiveness does hold. The details can be found in [17, 51].

3.4.2 Loss Measure and Gain Measure

Based on the total probability theorem, a more specified measure with different costs under conditions of "winning", "losing" and "even", respectively, can be defined as below

$$\begin{aligned} M(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) &= \Pr(\hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2) E(m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) | \hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2) \\ &+ \Pr(\hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2) E(m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) | \hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2) \\ &+ \Pr(\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2) E(m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) | \hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2), \end{aligned}$$

where $m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$ is the difference measure between $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j$ relative to $\mathbf{x}$, $i, j \in \{1, 2\}$ and $i \neq j$, $E(m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) | \hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2)$ is the expectation of $m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$ conditioned on $\hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2$. Then , given

$$m(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) = \begin{cases} l(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) & if\ \hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2 \\ 0 & if\ \hat{\mathbf{x}}_1 \succeq \hat{\mathbf{x}}_2 \end{cases}$$

we have the ***loss measure*** (LM) under condition that $\hat{\mathbf{x}}_1$ is inferior to $\hat{\mathbf{x}}_2$ relative to $\mathbf{x}$ as below:

$$LM(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) = \Pr(\hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2) E(l(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) | \hat{\mathbf{x}}_1 \prec \hat{\mathbf{x}}_2),$$

where $l(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ is the penalty for $\hat{\mathbf{x}}_1$ to be inferior to $\hat{\mathbf{x}}_2$. The smaller $LM(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$ is, the better the estimator $\hat{\mathbf{x}}_1$ is.

Similarly, the ***gain measure*** (GM) that $\hat{\mathbf{x}}_1$ is superior to $\hat{\mathbf{x}}_2$ relative to $\mathbf{x}$ can be defined as

$$GM(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x}) = \Pr(\hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2) E(g(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) | \hat{\mathbf{x}}_1 \succ \hat{\mathbf{x}}_2),$$

where $g(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ is the reward for $\hat{\mathbf{x}}_1$ to be superior to $\hat{\mathbf{x}}_2$. The larger $G(\hat{\mathbf{x}}_1 : \hat{\mathbf{x}}_2; \mathbf{x})$ is, the better the estimator $\hat{\mathbf{x}}_1$ is. Obviously, when $g(\mathbf{x}, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \equiv 1$, this measure reduces to the PCM.

The estimators compete against each other case by case directly. The joint information of the estimators and estimatee is involved, more aspects of the difference between estimators is revealed.

The loss measure and gain measure can greatly alleviate the peculiarity of the PCM that "the winner can then be arbitrarily bad when the observation does not belong to this subset". Because they still depends on the joint information of estimators, generally speaking, they are still not transitive.

3.5 Illustrative Examples

Example 1. In this example, we focus on demonstrating that error measures and frequency counts are complementary.

Given a single linear noisy measurement z of an estimatee x:

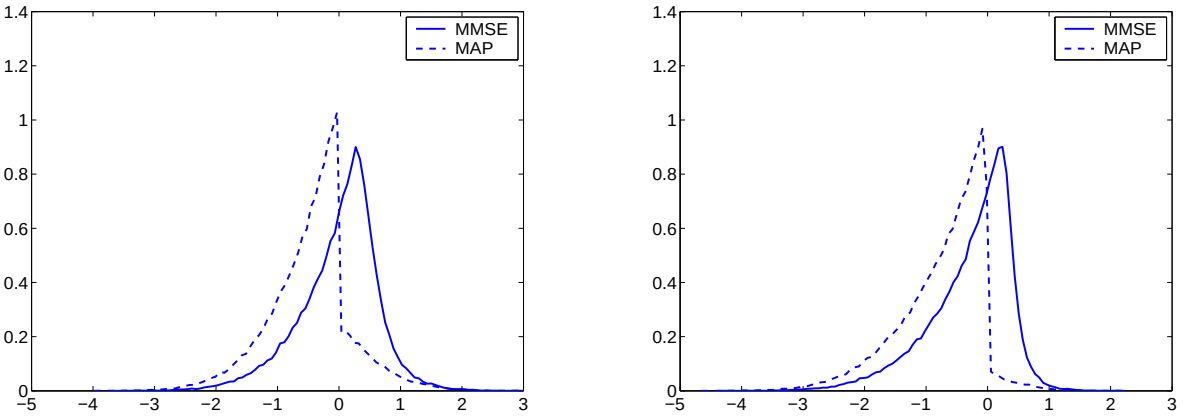
$$z = x + v \qquad \text{with} \qquad v \sim \mathcal{N}(0, 1)$$

The maximum likelihood (ML) estimator is $\hat{x}^{\mathrm{ML}} = z$. Assume that x is an exponentially distributed random variable with prior pdf $f(x) = \lambda e^{-\lambda x}\mathbf{1}(x)$, where $\mathbf{1}(x)$ is the unit step function. It can be shown [23] that two classes of MAP and MMSE estimators, respectively, with different values of λ, are given by

$$\begin{aligned} \hat{x}^{\mathrm{MAP}}(\lambda) &= \max(z - \lambda, 0) \\ \hat{x}^{\mathrm{MMSE}}(\lambda) &= (\sqrt{2\pi}[1 - \Phi(\lambda - z)])^{-1} e^{-(z-\lambda)^2/2} + z - \lambda \end{aligned}$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function.

A simulation with $100,000$ Monte Carlo runs was conducted in which x was generated as a random variable with the true density $f(x) = e^{-x}\mathbf{1}(x)$ (i.e., $\lambda = 1$). In the following plots, the x-axis is the λ values assumed in the MAP and MMSE estimators, while the true λ is always equal to 1.



(a) $\lambda = 1$ (b) $\lambda = 1.8$

Fig. 3.1: Empirical probability densities of estimation errors of MAP and MMSE estimators.

Fig. 3.1 shows the empirical error pdfs of the MAP and MMSE estimators, each for the two cases of $\lambda = 1$ and 1.8, respectively. Note that the estimation errors of the MAP estimators are

$$\tilde{x}^{\mathrm{MAP}} = x - \hat{x}^{\mathrm{MAP}} = x - \max(x + v - \lambda, 0) = \begin{cases} \lambda - v & z > \lambda \\ x & z \le \lambda \end{cases}$$

They are highly non-Gaussian. Their density functions are made by cutting and pasting a Gaussian pdf $\mathcal{N}(\tilde{x};\lambda,1)/c$ and a mixture of an exponential pdf and a Gaussian pdf. Those of the MMSE estimators, albeit non-Gaussian, are not far from Gaussian because they are the sum of a Gaussian variable and something complex:

$$\begin{aligned}\tilde{x}^{\mathrm{MMSE}}(\lambda) &= x(1)-\hat{x}^{\mathrm{MMSE}}(\lambda) = x(1)-(z(1)-\lambda)-(\sqrt{2\pi}[1-\Phi(\lambda-z)])^{-1}e^{-(z-\lambda)^2/2} \\ &= \lambda - v - (\sqrt{2\pi}[1-\Phi(\lambda-z)])^{-1}e^{-(z-\lambda)^2/2} \\ &\approx \lambda - v - [0.661|\lambda-z| + 0.3999\sqrt{(\lambda-z)^2+5.51}]\end{aligned}$$

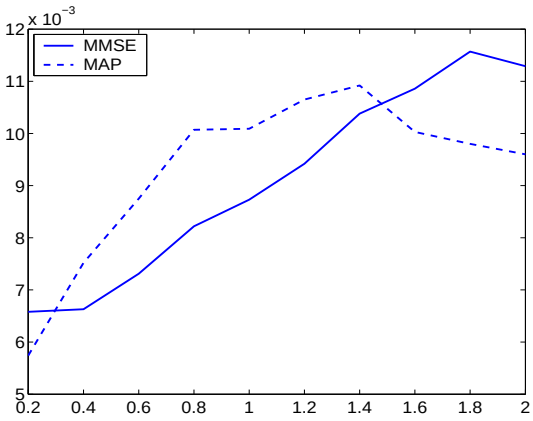
where use has been made of the following approximation (see, e.g., [21])

$$\Phi(x) \approx 1 - \frac{e^{-x^2/2}/\sqrt{2\pi}}{0.661x + 0.3999\sqrt{x^2+5.51}}, \quad x \geq 0$$

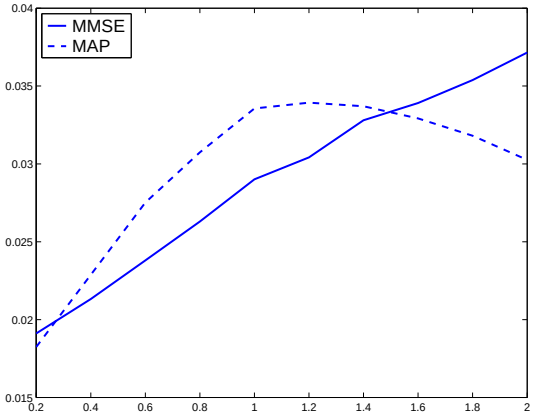
Fig. 3.2 shows the success rates for the success region defined by (3.14) with several δ_s^2 values, and the failure rates for the feasible region (3.15) with $\delta_f = 10$. It can be seen that $\hat{x}^{\mathrm{MAP}}(\lambda=1)$ has about highest success rate among MAP estimators. However, its success rate is lower than those of the MMSE estimators with large λ values. This is understandable from the error distribution given in Fig. 3.1 because these MMSE estimators have a greater probability (larger area) for a small error. However, in the limit as $\delta_s \to 0$, the $\hat{x}^{\mathrm{MAP}}(\lambda=1)$ would have a greater probability of an extremely small error than that of the MMSE estimators, as the theory predicts.[3] In fact, the error pdfs of the MMSE estimators do not peak at zero, but as the assumed λ increases, the peak moves towards zero and thus the success rate is higher. This example demonstrates that caution should be exercised when using success rate if the estimation errors are not Gaussian distributed. However, we believe it is generally safe in the Gaussian case.

Fig. 3.2(d) shows that the MAP estimators have worse failure rates than the MMSE estimators. This is understandable: The MAP estimators in theory should have higher

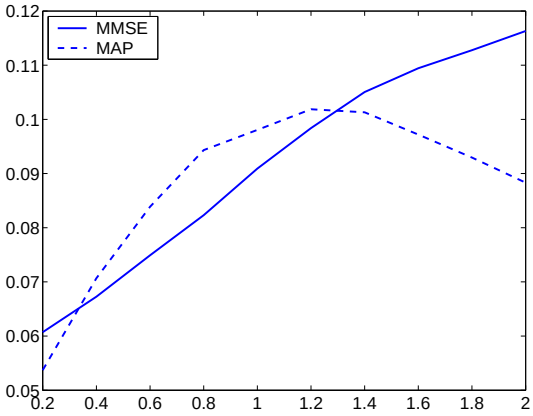
[3]The peak of the empirical density of $\tilde{x}^{\mathrm{MAP}}(\lambda=1)$ would be higher if more bins were used in generating and plotting the density.



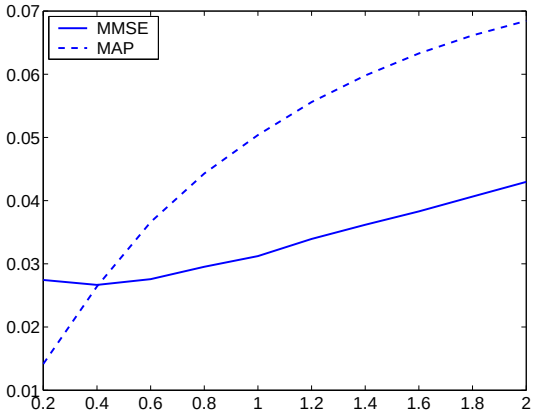
(a) Success rate with $\delta_s^2 = 0.0001$



(b) Success rate with $\delta_s^2 = 0.001$



(c) Success rate with $\delta_s^2 = 0.01$



(d) Failure rate with $\delta_f = 10$

Fig. 3.2: Success rate and failure rate of MAP and MMSE estimators vs. assumed δ values.

success rate (i.e., more small error), but larger average errors, and thus also more very bad errors, which implies a higher failure rate. It is also interesting to note that both failure rates increase with λ.

Example 2. In this example, we demonstrate that RMSE, AEE, PCM and LM are complementary. Given the noisy measurements $\mathbf{z}^n = (\mathbf{z}_1, ..., \mathbf{z}_n)$ of an estimatee $\mathbf{x}$, where

$$\mathbf{z}_i = \mathbf{x} + v_i, \quad v \sim \mathcal{N}(0, \sigma_v^2) \quad i = 1, ..., n$$

Assume that $\mathbf{x}$ is a ternary random variable with point masses $P\{\mathbf{x} = -1\} = P\{\mathbf{x} = 0\} = P\{\mathbf{x} = 1\} = 1/3$ and uncorrelated with v.

It can be shown [23] that two classes of MAP and MMSE estimators, respectively, with different values of σ_v^2, are given by

$$\hat{\mathbf{x}}^{MAP}(\sigma_v^2) = \begin{cases} -1 & \hat{\mathbf{z}} \leq -0.5 \\ 0 & -0.5 < \hat{\mathbf{z}} \leq -0.5 \\ 1 & \hat{\mathbf{z}} > 0.5 \end{cases}$$

$$\hat{\mathbf{x}}^{MMSE}(\sigma_v^2) = c(z)[\exp(-\frac{(1-\hat{\mathbf{z}})^2}{2\sigma_{\hat{\mathbf{z}}}^2}) - \exp(-\frac{(-1-\hat{\mathbf{z}})^2}{2\sigma_{\hat{\mathbf{z}}}^2})]$$

where $\hat{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$, $\sigma_{\hat{\mathbf{z}}}^2 = \frac{\sigma_v^2}{n}$, $c(z) = \exp(-\frac{(1-\hat{\mathbf{z}})^2}{2\sigma_{\hat{\mathbf{z}}}^2}) + \exp(-\frac{(-1-\hat{\mathbf{z}})^2}{2\sigma_{\hat{\mathbf{z}}}^2}) + \exp(-\frac{\hat{\mathbf{z}}^2}{2\sigma_{\hat{\mathbf{z}}}^2})$.

A simulation with 100,000 Monte Carlo runs was conducted in which $\mathbf{x}$ was generated as a random variable. In this example, the loss measure is defined in terms of the absolute estimation error.

Table 3.1. shows the RMSE, AEE, PCM and LM of the MAP and MMSE estimators at $\sigma_v = 2, \sigma_v = 6$ and $\sigma_v = 10$, respectively. The RMSE of the MMSE estimator is always smaller than that of the MAP estimator since MMSE estimator minimizes MSE, but it does not minimize the mean error and hence AEE.

For $\sigma_v = 2$, the AEE of the MAP estimator is smaller than that of the MMSE estimator. As we know, the MMSE estimator places "undue emphasis on large deviations which may occur with small probability"— this happens here. From PCM it is clear that the former is better than the latter in 74.54 percent of cases. The LM indicates that although large errors

Table 3.1: RMSE, AEE, PCM and LM of MAP and MMSE estimators at $\sigma_v = 2, \sigma_v = 6$ and $\sigma_v = 10$.

σ_v	$\sigma_v = 2$		$\sigma_v = 6$		$\sigma_v = 10$	
	$\hat{\mathbf{x}}^{MMSE}$	$\hat{\mathbf{x}}^{MAP}$	$\hat{\mathbf{x}}^{MMSE}$	$\hat{\mathbf{x}}^{MAP}$	$\hat{\mathbf{x}}^{MMSE}$	$\hat{\mathbf{x}}^{MAP}$
RMSE	0.4776	0.5506	0.7506	0.9798	0.7910	1.1040
AEE	0.3635	0.2914	0.6459	0.6736	0.6786	0.7955
PCM	25.46%	74.54%	47.30%	52.70%	54.47%	45.53%
LM	0.1345	0.0663	0.1702	0.2914	0.1543	0.4120

happen in the cases that MAP estimator is inferior to MMSE estimator, its average loss is still smaller than that caused by MMSE estimator. Therefore, in this case, we conclude that the MAP estimator is better than MMSE estimator.

For $\sigma_v = 6$, although the MMSE estimator is better than the MAP estimator in terms of RMSE, AEE and LM, the latter is more often superior to the former. The large absolute errors of the MAP estimator result in that its average error is larger than that of the MMSE estimator.

For $\sigma_v = 10$, the MMSE estimator is better than the MAP estimator in terms of all above measures. And we can conclude that in this case the MMSE estimator is preferred.

Figs. 3.3 and 3.4 show the concentration probabilities (CP) of the MMSE estimator and MAP estimator versus the absolute estimation error range c at $\sigma_v = 2$ and 6, respectively. Here the CP is simply the CDF of the absolute estimation error. Because the MAP estimator can only have an absolute estimation error c of $0, 1$ or 2, the CP of the MAP estimator is a step function vs. c. However, the c of the MMSE estimator is continuous, and its CP is also continuous. Obviously, neither of them can dominate the other. For example, while $\sigma_v = 2$ and $c < 0.503$, the MAP estimator has the higher concentration probability. Whereas for $0.503 < c < 1$, the MMSE estimator's CP is higher. From Figs. 3.3 and 3.4, we can easily

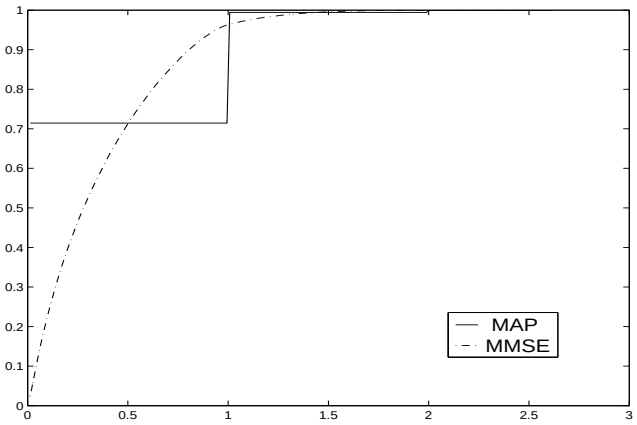


Fig. 3.3: CP vs. c ($\sigma_v = 2$)

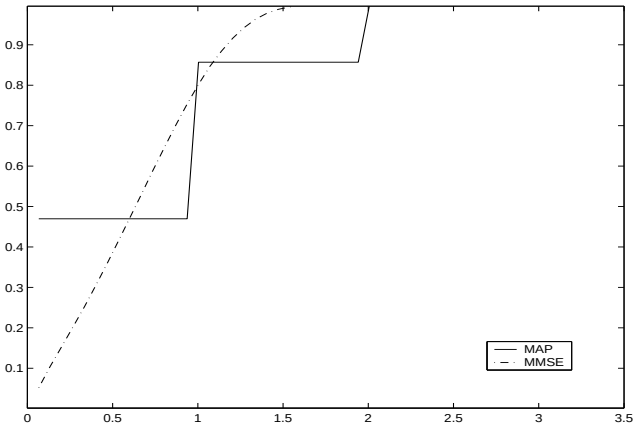


Fig. 3.4: CP vs. c ($\sigma_v = 6$)

tell which estimator is more concentrated for a given c.

In Figs. 3.5 and 3.6, the RMSE and AEE vs. the noise deviation σ_v are shown. Clearly, the RMSE of the MMSE estimator is always smaller than that of the MAP estimator. However, for $\sigma_v < 5.3$, the AEE of the MAP estimator is smaller than that of the MMSE estimator.

As is clear from Fig. 3.7, for the Pittam Closeness measure of the MAP estimator relative to. MMSE estimator vs. σ_v, for $\sigma_v < 7.3$, the MAP estimator is better than the MMSE estimator. In Fig. 3.8 ($LM_{MAP} - LM_{MMSE}$), the MAP estimator outperforms the MMSE

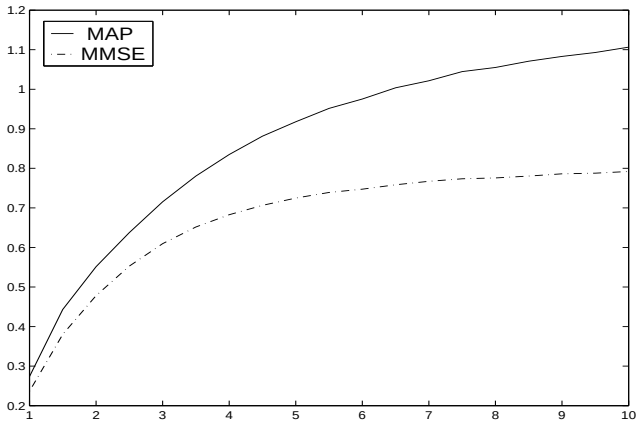


Fig. 3.5: RMSE vs. σ_v

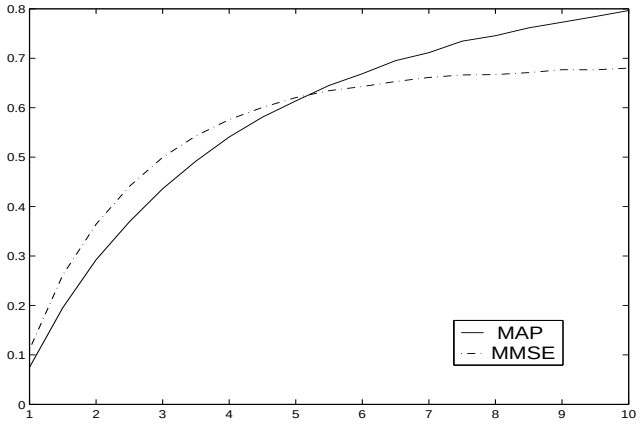


Fig. 3.6: AEE vs. σ_v

estimator for $\sigma_v < 3.7$.

Except the RMSE measure, all the others give a quite reasonable conclusion from different aspects and tradeoffs. This makes a lot of sense because the MAP estimator as a hard decision here is better than the MMSE estimator when the noise is small, while the noise become bigger the hard decision will take a much greater risk, and consequently, the MMSE estimator becomes better.

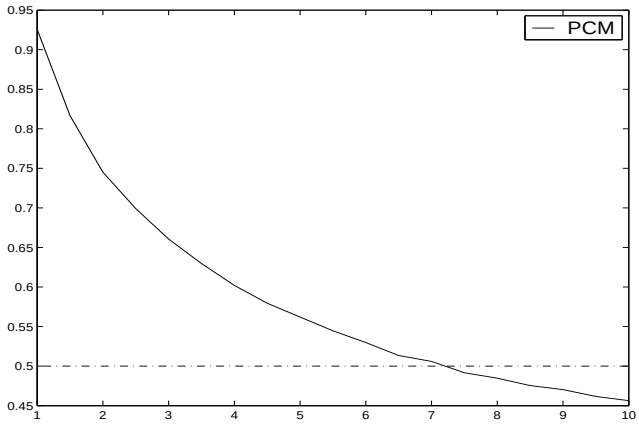


Fig. 3.7: PCM of $\hat{\mathbf{x}}^{MAP}$ relative to $\hat{\mathbf{x}}^{MMSE}$ vs. σ_v

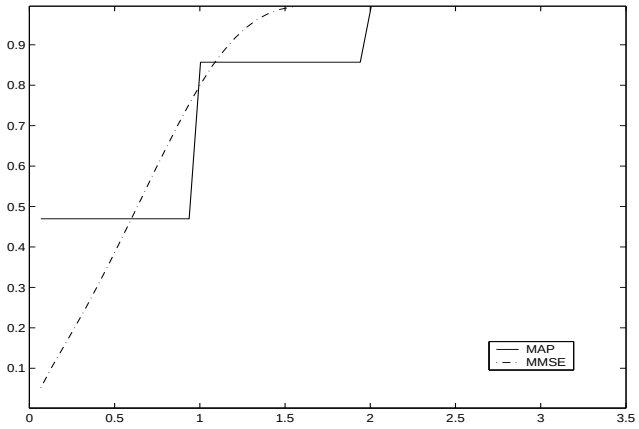


Fig. 3.8: $(LM_{MAP} - LM_{MMSE})$ vs. σ_v)

3.6 Concluding Remarks

A variety of new practical metrics for measuring performance of estimation algorithms has been proposed and justified. They are presented in four groups: (a) *absolute* error measures, including average Euclidean error (AEE), geometric average error (GAE), harmonic average error (HAE), median and mode of error norm, as well as root-mean-square error (RMSE); (b) *relative* error measures, including Bayesian estimation error quotient (BEEQ), estimate-measurement error ratio (EMER), and relative RMSE, AEE; (c) *frequency* counts, including success rate, failure rate, and concentration probability; (d) *measures for pairwise compar-*

ison, including Pitman closeness, loss measure and gain measure. These metrics are useful for measuring or comparing different aspects of the performance of estimation algorithms.

It is an illusion that performance evaluation can be done completely fairly and impartially. This is partly because simple metrics cannot capture a complete picture of the performance of an estimation algorithm and those that are more complete (e.g., empirical error distribution functions) are more complex and subject to subjective interpretations. More specifically, use of any metric in performance evaluation implicitly favors the estimator that tries to optimize this same metric. Nevertheless, all is not lost. What one should do is to choose the metrics that are more relevant to the case under consideration. That is also the value of having a wide spectrum of metrics available, along with a good understanding of them. While it is certainly superior in theory to have a unified metric, to be discussed in next chapter, specific metrics can be significantly more useful for particular applications.

Chapter 4

Global Performance Measures

Local performance measures are good at indicating local statistical performance concerns but not global ones. For example, they can not reflect the performance preference shift in a spectrum of metrics. A comprehensive view about estimation performance is clearly desired. In particular, the insights into the global performance need to be explored. This chapter is dedicated to two types of global performance measures.

The first one pinpoints the interaction between estimation criteria and optimal estimation such that what is called ***performance spectrum*** is proposed accordingly. The location of estimator within the domain of estimatee is determined by the estimation criterion. When estimation criterion changes, the location my also shift. Insight into such interaction will help elucidate performance prediction for producing new estimators and change of preference in performance comparison.

The second one identifies and specifies the connotation of comprehensive performance requirements such that what is called ***the desired estimation error PDF*** is proposed. The desired error PDF can provide a more comprehensive characterization of the performance requirements. Various performance concerns can be taken into account as well with ease. The performance measures associated with the desired error PDF are also proposed.

This Chapter is organized as follows. In Sec. 4.1, we investigate the relationship and interaction among estimator, estimation criterion and performance measures. Accordingly,

the performance spectrum is proposed, along with the discussion of its properties. In Sec. 4.2, we introduce the desired estimation error PDF to characterize the application requirements. New performance measures are proposed to indicate the dispersion or convergence level of estimation error PDF relative to the desired error PDF. Examples for both measures are illustrated in Sec. 4.3. Conclusions are given in Sec. 4.4.

4.1 Estimation Performance Spectrum

The performance spectrum is designed to display the performance preference shift when estimation error is measured in different norms. We first investigate the procedure of Bayesian point estimation so as to show the interaction between estimators and estimation criteria.

4.1.1 Bayesian Point Estimation

Given a joint probability density function of $f(x, z)$ and observation z, Bayesian estimation is to infer the random variable x of interest from z based on $f(x, z)$ by optimizing a cost or gain function. For example, assuming $g(\cdot)$ is a cost function, the corresponding estimator $\hat{x}$ is defined by

$$\hat{x} = \arg\min_{\hat{x}} E\left[g(\tilde{x})|z\right]$$

where $\tilde{x} = x - \hat{x}$, $E\left[g(\tilde{x})|z\right]$ is the conditional expectation of cost $g(\tilde{x})$. $g(\tilde{x})$ needs to satisfy some admissibility conditions which include $g(\tilde{x}) = g(-\tilde{x})$ and $g(\tilde{x})$ is a convex function. Note the complete information of x is carried by its posterior PDF $f(x|z)$ and the Bayesian estimator is the location in the support set of x which follows the estimation criterion.

After we obtain the estimator $\hat{x}$, this special point serve as a representative of the posterior PDF of x. In other words, from a posterior PDF to an estimator through the optimization procedure, we obtain one possible value at the price of losing the complete information and producing the estimation errors. For estimation purpose, it is also necessary to quantify the estimation error or performance.

Towards a comprehensive view of the estimation performance, we further investigate the procedure of producing an estimator so as to gain a physical interpretation on the relationship between an estimator and an estimation criterion.

4.1.2 A Physical Interpretation

First, we consider the scalar case. Assume $g(e)$, $e \geq 0$, is a nondecreasing function and first order differentiable, i.e., $g'(e) \geq 0$. The estimator is defined by

$$\hat{x} = \arg\min_{\hat{x}} E\left[g(||x - \hat{x}||)|z\right]$$

Note that

$$\begin{aligned} E\left[g(||x - \hat{x}||)|z\right] &= \int_{-\infty}^{\infty} g(||x - \hat{x}||) f(x|z) dx \\ &= \int_{-\infty}^{\hat{x}} g(\hat{x} - x) f(x|z) dx + \int_{\hat{x}}^{\infty} g(x - \hat{x}) f(x|z) dx \end{aligned} \tag{4.1}$$

Following Leibniz's rule of differentiation

$$\frac{d}{dy}\int_{a(y)}^{b(y)} g(x,y)dx = \int_{a(y)}^{b(y)} \frac{\partial}{\partial y} g(x,y)dx + \frac{db(y)}{dy} g(b(y),y) - \frac{da(y)}{dy} g(a(y),y) \tag{4.2}$$

and taking derivative of $E\left[g(||x - \hat{x}||)|z\right]$ with respect to $\hat{x}$ yield:

$$\frac{dE\left[g(||x - \hat{x}||)|z\right]}{d\hat{x}} = \int_{-\infty}^{\hat{x}} g'(\hat{x} - x) f(x|z) dx - \int_{\hat{x}}^{\infty} g'(x - \hat{x}) f(x|z) dx \tag{4.3}$$

A necessary condition for $\hat{x}$ to be the optimal estimator is

$$\frac{dE\left[g(||x - \hat{x}||)|z\right]}{d\hat{x}} = 0$$

i.e.,

$$\int_{-\infty}^{\hat{x}} g'(\hat{x} - x) f(x|z) dx = \int_{\hat{x}}^{\infty} g'(x - \hat{x}) f(x|z) dx \tag{4.4}$$

Let $t = \hat{x} - x$, we have

$$\int_{0}^{\infty} g'(t) f(\hat{x} - t|z) dt = \int_{0}^{\infty} g'(t) f(\hat{x} + t|z) dt \tag{4.5}$$

(4.5) says that $g'(t)$ acts upon $f(\hat{x}-t|z)$ and $f(\hat{x}+t|z)$, $t \geq 0$, respectively, and $f(x|z)$ is balanced at $x = \hat{x}$, i.e.,

$$\langle g', f_{\hat{x}-t} \rangle = \langle g', f_{\hat{x}+t} \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in the Hilbert space. This is similar to the lever principle: *An equilibrium is established when the moment of the force acting in a clockwise direction is equal to the moment of the force acting in a counterclockwise direction.* In physics the product of a force by its effort arm is called a moment of the force. The lever principle is simply

$$F_1 \cdot L_1 = F_2 \cdot L_2$$

where F_1 and F_2 are two forces, L_1 and L_2 are their effort arms. The estimator minimizing the mean $g(e)$ error is a balance pivot for the "force" of $g'(t)$ acting upon two sides of a "lever" of $f(x|z)$. This gives a "physical" interpretation about the relationship between estimator and estimation criterion.

Moreover, if $g(e)$ is a second order differentiable concave function, i.e., $g''(e) > 0$, we have

$$\frac{d^2 E\left[g(||x-\hat{x}||)|z\right]}{d\hat{x}^2} = \int_{-\infty}^{\hat{x}} g''(\hat{x}-x) f(x|z) dx + \int_{\hat{x}}^{\infty} g''(x-\hat{x}) f(x|z) dx > 0 \tag{4.6}$$

at the balance point. Thus, the balance pivot minimizes the cost function and is the estimator.

For the vector case and the cost function

$$g(\mathbf{x}-\hat{\mathbf{x}}) = \sum\nolimits_{i=1}^{n} g(||x_i - \hat{x}_i||) \tag{4.7}$$

where x_i is the ith element of $\mathbf{x}$ and $\hat{x}_i$ is the estimator of x_i, because

$$\min_{\hat{\mathbf{x}}} E\left[\sum_{i=1}^{n} g(||x_i - \hat{x}_i||)|\mathbf{z}\right] \geq \sum_{i=1}^{n} \min_{x_i} E\left[g(||x_i - \hat{x}_i||)|\mathbf{z}\right] \tag{4.8}$$

where the equality holds for the optimal estimator $\hat{\mathbf{x}}^* = (\hat{x}_1^*, ..., \hat{x}_n^*)$ with $\hat{x}_i^*$ being the marginal defined by $\hat{x}_i^* = \arg\min_{\hat{x}} E\left[g(||x_i - \hat{x}_i||)|\mathbf{z}\right]$, $\hat{\mathbf{x}}^*$. Therefore, the vector optimal estimator

consists of the marginal optimal estimators of each dimension and it is still an equilibrium along each dimension.

Moreover, from (4.5), we have

$$\int_0^\infty g'(t)(f(\hat{x}-t|z)-f(\hat{x}+t|z))dt=0$$

Property 1. If there exists a point $\hat{x}$ that satisfies

$$f(\hat{x}-t|z)-f(\hat{x}+t|z)=0 \tag{4.9}$$

for all $t \geq 0$, this symmetrical point $\hat{x}$ of the posterior PDF will always be the optimal estimator regardless which Bayes admissible criterion is used. In other words, if the posterior density satisfies (4.9), the estimates will be optimal for a large class of cost functions. Clearly, if the posterior density function is Gaussian, it satisfies the above condition. In this case, all admissible criteria lead to the same estimator, which is the conditional mean. However, the performance of the estimation still needs to be measured by the estimation measures from different aspects of practical requirements such as concentration probability.

Property 2. For an $\hat{x}$, if $f(\hat{x}-t|z)-f(\hat{x}+t|z)<0$ (or >0) for all $t \geq 0$, because $g'(e) \geq 0$, $e \geq 0$, $\int_0^\infty g'(t)(f(\hat{x}-t|z)-f(\hat{x}+t|z))dt<0$ (or >0), this point $\hat{x}$ can not be an optimal estimator regardless which admissible criterion is used. Therefore, a necessary condition that a point $\hat{x}$ can be an optimal estimator is $f(\hat{x}-t|z)-f(\hat{x}+t|z) \geq 0$ for some t and $f(\hat{x}-t|z)-f(\hat{x}+t|z) \leq 0$ for others. We say these points are admissible. The set of all these points is the admissible region. Only an admissible point can be a balance pivot.

4.1.3 Optimal Estimator Locations and p-power Error Costs

In this section, we investigate the interaction between the estimator locations and p-power error costs. Define the p-power error cost function as $g(\tilde{x})=||\tilde{x}||^p$, $p \geq 1$.

Theorem 1. *For a posterior PDF, if $\hat{x}_1$ and $\hat{x}_2$ are two optimal estimators corresponding to the p-power error cost functions with $p=p_1$ and p_2, respectively, any point $\hat{x}_*$ between $\hat{x}_1$*

and $\hat{x}_2$ is an optimal estimator that minimizes the average p-power error with a suitable p_* between p_1 and p_2. In particular, any point between the posterior mean $\bar{x}$ and median $x_{1/2}$ is an optimal estimator w.r.t. a p-power error cost with $1 < p < 2$.

Proof: For any point $\hat{x} = \hat{x}_*$ between $\hat{x}_1$ and $\hat{x}_2$, we only need to show that there exists an appropriate p, which is between p_1 and p_2, and $\hat{x} = \hat{x}_*$ minimizes

$$E\left[|\hat{x} - x|^p | z\right] = \int_{-\infty}^{\hat{x}} (\hat{x} - x)^p f(x|z) dx + \int_{\hat{x}}^{\infty} (x - \hat{x})^p f(x|z) dx \tag{4.10}$$

Let

$$\begin{aligned} c(p, \hat{x}) &\triangleq \frac{dE\left[|\hat{x} - x|^p | z\right]}{d\hat{x}} \\ &= \int_{-\infty}^{\hat{x}} p(\hat{x} - x)^{p-1} f(x|z) dx - \int_{\hat{x}}^{\infty} p(x - \hat{x})^{p-1} f(x|z) dx \end{aligned} \tag{4.11}$$

Taking derivative of $c(p, \hat{x})$ with respect to $\hat{x}$ yields:

$$\frac{dc(p, \hat{x})}{d\hat{x}} = \int_{-\infty}^{\hat{x}} p(p-1)(\hat{x} - x)^{p-2} f(x|z) dx + \int_{\hat{x}}^{\infty} p(p-1)(x - \hat{x})^{p-2} f(x|z) dx \tag{4.12}$$

Note that for $p \geq 1$ each term in (4.12) is nonnegative; if $\frac{dc(p,\hat{x})}{d\hat{x}}$ is equal to zero, both terms have to be zero. Under the condition of $p > 1$, $x - \hat{x}$ would be equal to zero except over a zero-measure set, i.e. $x = \hat{x}$ almost surely. Otherwise,

$$\frac{dc(p, \hat{x})}{d\hat{x}} > 0 \tag{4.13}$$

Therefore, for any given $p > 1$, $E\left[|\hat{x} - x|^p | z\right]$ is a strictly convex function and $c(p, \hat{x})$ is strictly increasing in $\hat{x}$.

Let $\hat{x} = \hat{x}_*$ be between $\hat{x}_1$ and $\hat{x}_2$. Without loss of generality assuming $\hat{x}_1 < \hat{x}_* < \hat{x}_2$, we have

$$c(p_1, \hat{x}_*) > c(p_1, \hat{x}_1) = 0 \tag{4.14}$$

and

$$c(p_2, \hat{x}_*) < c(p_2, \hat{x}_2) = 0 \tag{4.15}$$

Since the integrant of $c(p,\hat{x})$ is differentiable with respect to p, $c(p,\hat{x})$ is continuous in p. Therefore, (4.14) and (4.15) guarantee that there exists $p=p_*$ to make $c(p_*,\hat{x}_*)=0$ where p_* is between p_1 and p_2. Further, $c(p,\hat{x})$ is monotonic in $\hat{x}$, so $\hat{x}=\hat{x}_*$ is the unique minimum point of $E\left[|\hat{x}-x|^{p_*}|z\right]$. The uniqueness also indicates the balance point is unique in this case. Similarly, for $\hat{x}_1>\hat{x}_2$, we have the same conclusion.

Obviously, when $c(1,\hat{x})=0$, $\int_{-\infty}^{\hat{x}} f(x|z)dx=\int_{\hat{x}}^{\infty} f(x|z)dx$ and $\hat{x}$ is the median; when $c(2,\hat{x})=0$, $\hat{x}=\int_{-\infty}^{\infty} xf(x|z)dx$ and $\hat{x}$ is the posterior mean. Then, according to the above conclusion, any point between the posterior mean and median will be an optimal estimator corresponding to an appropriate p-power error cost function with $1<p<2$. This completes the proof.

Theorem 1 gives a very general and intuitive characterization of the interaction between the locations of the estimators and corresponding p-power error costs. Next, we see the reverse of the theorem also holds for a large class of posterior PDFs.

Theorem 2. Assume for a given posterior PDF $f(x|z)$ and any value $\hat{x}$ of x, there exists at most one $T_{\hat{x}}>0$ such that

$$\begin{cases} f(\hat{x}-t|z)\geq f(\hat{x}+t|z) \text{ when } t>T_{\hat{x}} \\ f(\hat{x}-t|z)\leq f(\hat{x}+t|z) \text{ when } t<T_{\hat{x}} \end{cases} \tag{4.16}$$

or

$$\begin{cases} f(\hat{x}-t|z)\leq f(\hat{x}+t|z) \text{ when } t>T_{\hat{x}} \\ f(\hat{x}-t|z)\geq f(\hat{x}+t|z) \text{ when } t<T_{\hat{x}} \end{cases} \tag{4.17}$$

and assume $c(p_1,\hat{x}_1)=c(p_2,\hat{x}_2)=0$ for some $p_i\geq 1$, i=1,2, and $\hat{x}_1,\hat{x}_2$. Given p_*, $1\leq p_1<p_*<p_2$, then there exists $\hat{x}_*$ between $\hat{x}_1$ and $\hat{x}_2$ such that $c(p_*,\hat{x}_*)=0$.

(4.16) and (4.17) characterize a large class of unimodal PDFs of a moderate asymmetry, such as the chi-square distributions, exponential and log-normal distributions.

Proof: Without loss of generality, assume for any $\hat{x}$ that the condition (4.16) holds. Then

for $p > p_1$,

$$\begin{aligned} c(p,\hat{x}_1) =& p\int_0^{\infty} t^{p-1}(f(\hat{x}_1 - t|z) - f(\hat{x}_1 + t|z))dt \\ =& p\int_0^{T_{\hat{x}}} t^{p-p_1}t^{p_1-1}(f(\hat{x}_1 - t|z) - f(\hat{x}_1 + t|z))dt \\ & - p\int_{T_{\hat{x}}}^{\infty} t^{p-p_1}t^{p_1-1}(f(\hat{x}_1 + t|z) - f(\hat{x}_1 - t|z))dt \\ <& p\int_0^{T_{\hat{x}}} T_{\hat{x}}^{p-p_1}t^{p_1-1}(f(\hat{x}_1 - t|z) - f(\hat{x}_1 + t|z))dt \\ & - p\int_{T_{\hat{x}}}^{\infty} T_{\hat{x}}^{p-p_1}t^{p_1-1}(f(\hat{x}_1 + t|z) - f(\hat{x}_1 - t|z))dt \\ =& pT_{\hat{x}}^{p-p_1}c(p_1,\hat{x}_1) = 0 \end{aligned} \tag{4.18}$$

Similarly, for $p < p_2$

$$\begin{aligned} c(p,\hat{x}_2) &= \int_0^{\infty} t^{p-1}(f(\hat{x}_2 - t|z) - f(\hat{x}_2 + t|z))dt \\ &> T_{\hat{x}}^{p-p_2}c(p_2,\hat{x}_2) = 0 \end{aligned} \tag{4.19}$$

According to (4.13), $c(p,\hat{x})$ is continuous and monotonically increasing in $\hat{x}$. Therefore, for any p_*, $1 \le p_1 < p_* < p_2$, (4.18) and (4.19) guarantee that there exists a point $\hat{x}_*$ between $\hat{x}_1$ and $\hat{x}_2$ that is the optimal estimator w.r.t. the p_*-power error cost.

Similarly, for condition (4.17), $\hat{x}_1 > \hat{x}_2$, we have the same conclusion. This completes the proof.

For the vector case with a cost function (4.7), although Theorem 1 and Theorem 2 do not hold, each dimension of the estimator does because the optimal estimator consists of the marginal optimal estimators of each dimension. Along each dimension, each element of the vector-valued estimator is still an equilibrium.

4.1.4 Estimation Performance Spectrum

The previous analysis on characteristics of estimation and its performance evaluation can be summarized as follows.

First, the estimation criterion is a compromise between application requirements and mathematic tractability. Most often, the estimation criterion is determined by mathematical tractability such as the MSE. As we have pointed out, the estimator is just a balance point in the support set. On the one hand, many estimation criteria may lead to the same point; on the other, an estimation criterion easy to handle can greatly facilitate the estimation procedure. This justifies the feasibility to pick a simple cost function as estimation criterion.

Second, the optimal estimator is a balance point only for the given cost function. By no means could an optimal estimator with respect to one particular estimation criterion be always the "best" estimator for all purposes. In practical problems, it is still necessary to know its behavior in terms of other performance measures.

Third, Many estimation requirements could be equivalent to some suitable L_p norm costs in the sense that the estimation requirement and one particular L_p norm cost may share the same optimal estimator.

Accordingly, we propose the ***estimation performance spectrum*** as below,

$$S_p = \left(E\left(||\tilde{x}||^p\right)\right)^{1/p} \approx \left(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left|\left|\tilde{x}^i\right|\right|^p\right)^{1/p} \qquad p > 0$$

where $\tilde{x}$ is the scalar or vector estimation error, superscript i stands for quantities pertaining to the ith run of a Monte-Carlo simulation.

The performance spectrum can provide the following insights and information for performance comparison of several estimators. First, $S_{p=1}$ and $S_{p=2}$ are the AEE and root MSE, and these two points are of particular interest in most cases. For a posterior PDF with the properties defined in Theorem 2, the magnitudes of S_p of different estimators at $p = 1$ and $p = 2$ can also indicate the relative position relationship to the median and mean. The smaller $S_{p=1}$ is, the closer the corresponding estimator is to the median. The smaller $S_{p=2}$ is, the closer the corresponding estimator is to the mean. Second, the preference of one estimator along the p-power axis can also be available. Third, because S_p depends more on the large values of $||\tilde{x}||$ as p increases, the degree of concentration can be revealed by check-

ing the slope $\frac{dS_p}{dp}$. As p increases, if the slope of an estimator is sharper than the other, its estimation error PDF has a heavier tail. Fourth, according to Theorem 1, for the scalar case, every point between two estimators is another estimator which is a compromise between the two estimation criteria. For the vector case, a convex combination of two estimators may not be optimal w.r.t. any p-power error criterion. However, it is still a compromise between the given estimation criteria along each dimension. Also, we have a rough idea about the performance of the combined estimator.

Therefore we can have a comprehensive view on the performance shift along with the p-power error costs of these estimators. And moreover, it is possible to construct new estimators from the given ones. An illustrative example is presented later in Sec. 4.3.

4.2 Desired Error PDF and Performance Measures

In this part, we argue that the estimation requirements can be more comprehensively characterized by an error PDF. And the performance relative to the desired estimation error PDF is proposed as well.

4.2.1 The Desired Error PDF

In practice, the relationship between the estimation final goal and the estimation error distribution is, most often, not straightforward. For example, in speech signal processing, the mean opinion score is used to evaluate the performance, and the error criterion which can best fit people's listening comfort is still unclear. Based on MSE or AEE, the conclusion could be far from satisfactions. For another example, given a real system with an unknown parameter, an efficient way to quantify the cost or loss of the system caused by the estimation error of the parameter is usually not straightforward.

We simply apply the philosophy coined centuries ago by Descartes:"... when it is not in our power to know what is true, we ought to do what is more probable." As estimation

error is inevitable, the practical performance requirements and concerns can be much more easily characterized by an estimation error PDF rather than by MSE, AEE or concentration probability. Using error PDF can integrate various application concerns and requirements, in particular, when some prior knowledge need to be taken into considerations. For example, the estimation error PDF of an unknown parameter that a stochastic system can tolerate gives a clear picture on the performance requirements. Therefore, we propose the desired error PDF to characterize the application requirements and concerns.

The shape of a desired PDF to characterize application requirements and concerns is determined by each application. Generally speaking, the more concentrated the desired PDF is, the more desired it is. The ideal error PDF is a Delta function $\delta(\tilde{x})$ which does not tolerate any error. However, except for the asymptotical case, none of estimators can satisfy this requirement. Another candidate is the Gaussian distribution with zero mean and a specified covariance because it is not unreasonable to assume many effects can affect the parameter of interest and a Gaussian distribution of the error due to these effects is natural according to the central limit theorem. Its covariance defines the tolerance of the system, and the zero mean specifies the requirement of unbiasedness. Of course, some other specific distributions can also be alternative candidates.

If one objects the proposed desired PDF to be too arbitrary, "we readily agree this convention is arbitrary and does not appear necessary to some degree. The question which concerns us here has something vague about it from its very nature." We only use a simple form of PDF to characterize the desired properties. Hopefully, it can capture these major desired properties of requirements for estimation errors.

We stress that even though we only use a Gaussian distribution as the desired PDF, we compare the PDF of estimation error with the whole desired distribution rather than only the first and second moments. The shape of the whole distribution is taken into account.

The concentration or dispersion measure w.r.t. the desired error PDF will be defined to

indicate estimation performance relative to the desired error PDF.

4.2.2 Relative Concentration and Dispersion Measures

In order to measure the concentration or dispersion level of an estimation error PDF relative to the desired one, we propose the following measures. For simplicity, we use x and y but not $\tilde{x}$ and $\tilde{y}$ for estimation errors in this section.

Given $x \sim f(x) = N(x; m, P)$, where

$$N(x; m, P) = |2\pi P|^{-\frac{1}{2}} \exp(-(x-m)'P^{-1}(x-m)/2)$$

from the density function we observe that the equal-height contours are defined by the relation [57]

$$t^2 = (x-m)'P^{-1}(x-m) \tag{4.20}$$

which is the description for an ellipsoid. The ellipsoid expands monotonically as t increase, and the cumulative probability $Pr(t > T)$ of being outside the ellipsoid $T^2 = (x-m)'P^{-1}(x-m)$ is only a function of T

$$Pr(t > T) = \frac{k}{2^{k/2}\Gamma(k/2+1)} \int_T^\infty x^{k-1} e^{-x^2/2} dx \tag{4.21}$$

where $k = dim(x)$.

Similarly, the concentration probability associated with the ellipsoid (4.20) is $Pr(t < T) = 1 - Pr(t > T)$. We let $Pr(t > T)$ and $Pr(t < T)$ be the concentration level and dispersion level of each point on the ellipsoid (4.20), respectively.

Fig. 4.1 shows the probability density at $k = 2$, and the concentration level and dispersion level for $k = 1$ is shown in Fig. 4.2.

Once the concentration or dispersion level of each point is quantified, the overall relative concentration and dispersion measures can be defined in the following two different manners.

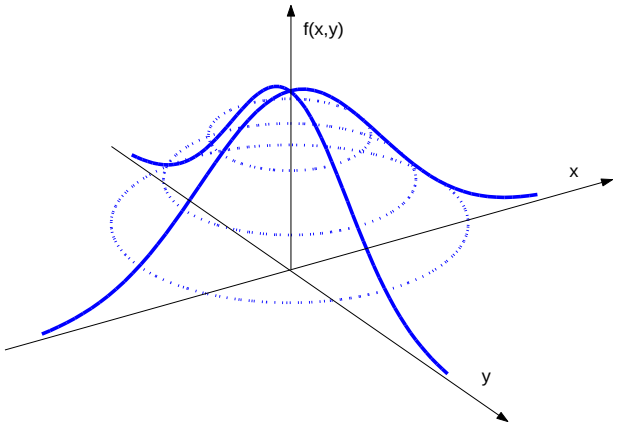


Fig. 4.1: 2D Gaussian Density

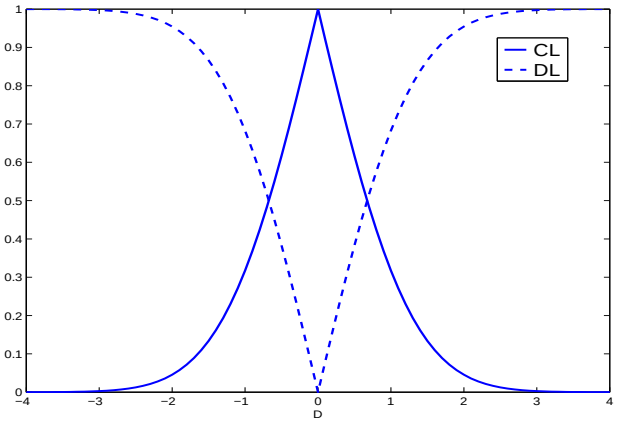


Fig. 4.2: 1D Concentration and Dispersion Levels

Concentration and Dispersion Measures of Combinatorially independent $\mathbf{Y}_i$. When Y_i, $i = 1, ..., N$, are combinatorially independent to reach their own estimation results, their averaged concentration or dispersion measure can be their arithmetic mean of each $Y_i^{'}$s concentration or dispersion level according to the Addition Principle in combinatorics. The Addition Principle in combinatorics can be simply explained as that the total is produced by the addition of each relevant quantities, which is the defining reason to use addition operation.

We define the concentration measure for a combinatorially independent error population

Y_i, $i = 1, ..., N$, relative to $x \sim N(x; m, P)$ as below

$$\text{CMI}(y : x) = E_y(\Pr(t > T(y))) \approx \frac{1}{N} \sum_{i=1}^{N} \Pr(t > T(Y_i))$$

or the dispersion measure as

$$\text{DMI}(y : x) = E_y(\Pr(t < T(y))) \approx \frac{1}{N} \sum_{i=1}^{N} \Pr(t < T(Y_i))$$

where $T^2(Y_i) = (Y_i - m)' P^{-1} (Y_i - m)$, $\Pr(t > T(Y_i))$ is relative to $f(x)$.

Concentration and Dispersion Measures of Combinatorially Dependent $\mathbf{Y}_i$. When Y_i, $i = 1, ..., N$, are combinatorially dependent to affect the final result, their average concentration or dispersion measure can be their geometric mean of all $Y_i^{'}$s concentration or dispersion levels according to the Multiplication Principle in combinatorics. The Multiplication Principle says: assuming that a procedure consists of k series steps, and each step j has n_j methods to perform regardless of the choices made on the previous steps. The total number of methods to perform the entire procedure is $n_1 n_2 \cdots n_k$.

We define the concentration measure for a combinatorially dependent population Y_i, $i = 1, ..., N$, relative to $x \sim N(x; m, P)$ as below

$$\text{CMD}(y : x) = e^{E_y(\ln(\Pr(t > T(y))))} \approx e^{\frac{1}{N} \sum_{i=1}^{N} \ln(\Pr(t > T(Y_i)))}$$

or the dispersion measure as

$$\text{DMD}(y : x) = e^{E_y(\ln(\Pr(t < T(y))))} \approx e^{\frac{1}{N} \sum_{i=1}^{N} \ln(\Pr(t < T(Y_i)))}$$

$\text{CMI}(x : x)$, $\text{DMI}(x : x)$, $\text{CMD}(x : x)$ and $\text{DMD}(x : x)$ have the following property.

Lemma 1: $CMI(x : x) = DMI(x : x) = \frac{1}{2}$; $CMD(x : x) = DMD(x : x) = e^{-1}$. □

Therefore, when $\text{CMI}(y : x) > \frac{1}{2}$, we say $f(y)$ is more concentrated than the desired $f(x)$; conversely, when $\text{CMI}(y : x) < \frac{1}{2}$, $f(y)$ is less concentrated than the desired $f(x)$ in terms of CMI. Similarly, when $\text{CMD}(y : x) > 0.368$, we say the sequence of Y_i is more concentrated

than the desired; conversely, when $\text{CMD}(y:x) < 0.368$, the estimation error sequence Y_i is less concentrated.

Comments: Given $m = 0$ and $k = 2n$, since $Pr(t > T) = e^{-T^2/2} \sum_{i=1}^{n} \frac{2^{-i+1}}{\Gamma(i)} T^{2(i-1)}$,

$$\text{CMD}(y:x) = e^{(-\frac{trace(CP^{-1})}{2} + E_y(\ln \sum_{i=1}^{n} \frac{2^{-i+1}}{\Gamma(i)} T^{2(i-1)}))}$$

where C is the MSE corresponding to the estimation error with PDF $f(y)$. The term $L = \frac{-trace(CP^{-1})}{2}$ plays an important role on the geometric mean of the dispersion level of each Y_i relative to $f(x)$. When $k = 2$, $\text{CMD}(y:x) = e^L$. L quantifies the concentration measure completely. $L < -1$ indicates that $f(y)$ is less concentrated than $f(x)$; otherwise, $f(y)$ is more concentrated than $f(x)$. In particular, when $P = C$, $L = -1$ and we say an unbiased estimator with MSE=C has the same dispersion level as a Gaussian distribution $N(x; 0, C)$.

In view of the above, first, the MSE of an estimator carries important information associated with a Gaussian distribution with zero mean and the covariance of the MSE to indicate how concentrated the error PDF is. Second, the above measures provide more information than MSE itself about the concentration or dispersion information of $f(y)$ relative to $f(x)$. This will be illustrated in the following section.

4.3 Illustrative Examples

Example 1. For Estimation Performance Spectrum

We use the first example in Sec. 3.5 again (refer to page 38) to illustrate the application of performance spectrum. Given $\hat{x}^{\text{MAP}}$ and $\hat{x}^{\text{MMSE}}$, there exists an approximate relationship among mean, median and mode for unimodal probability density function of moderate asymmetry [54]:

$$\text{mean} - \text{mode} \approx 3(\text{mean} - \text{median})$$

Thus

$$\text{median} \approx \frac{2}{3}\text{mean} + \frac{1}{3}\text{mode} \tag{4.22}$$

We fuse $\hat{x}^{\mathrm{MAP}}$ and $\hat{x}^{\mathrm{MMSE}}$ according to (4.22) and we can expect the fused estimator is near the median and is superior to the other two in terms of $S_{p=1}$. We compare their estimation performance spectra.

A simulation with $100,000$ Monte Carlo runs was conducted in which x was generated as a random variable with the density $f(x) = e^{-x}\mathsf{1}(x)$.

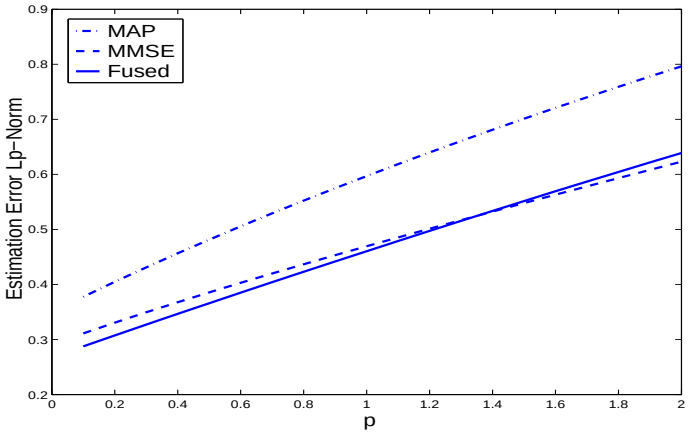


Fig. 4.3: Estimation Performance Spectrum $p > 0.1$

Fig. 4.3 shows the estimation performance spectrum S_p of $\hat{x}^{\mathrm{MAP}}$, $\hat{x}^{\mathrm{MMSE}}$ and the fused one for $0.1 < p < 2$. $\hat{x}^{\mathrm{MMSE}}$ dominates $\hat{x}^{\mathrm{MAP}}$. This indicates that $\hat{x}^{\mathrm{MMSE}}$ is more concentrated than $\hat{x}^{\mathrm{MAP}}$. The slope of $\hat{x}^{\mathrm{MAP}}$ is sharpest, and this indicates its estimation error is more dispersed than the other two. The slope of the fused estimator is also sharper than that of $\hat{x}^{\mathrm{MMSE}}$, this indicates the fused one's estimation error is distributed with a heavier tail. Although, theoretically speaking, $\hat{x}^{\mathrm{MAP}}$ should have larger concentration probability within a small region, its superiority as p becomes small is quite negligible in this situation. Its large dispersion results from the location of the MAP estimator is on the boundary of the admissible region and even the support set. In this case, most often we say $\hat{x}^{\mathrm{MAP}}$ is less desirable.

As we expected, the fused estimator has smallest S_p at $p = 1$, which indicates it is close

to the median. Moreover, because the S_p of the fused one is the smallest as $p < 1.2$, the estimation error of the fused one is also more concentrated but with a heavier distribution tail than $\hat{x}^{\mathrm{MMSE}}$, which is verified by Fig. 4.4.

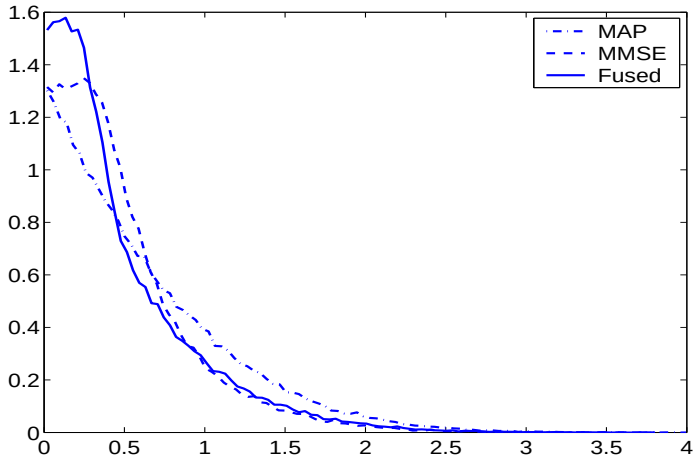


Fig. 4.4: Empirical PDF of $||\tilde{x}||$

Example 2. For the Desired PDF

As pointed out in [35], the estimation performance is most completely described by the estimation error PDF. In the following example, we directly show how the above measures reveal the difference among the following error PDFs. In Fig. 4.5, we illustrate the following five error PDFs, all with the zero mean and unity covariance.

$$
\begin{aligned}
f_1(\tilde{x}) =& N(\tilde{x}; 0, 1) \\
f_2(\tilde{x}) =& U(\tilde{x}; -\sqrt{12}, \sqrt{12}) \\
f_3(\tilde{x}) =& \alpha N(\tilde{x}; -5, 0.2^2) + \beta N(\tilde{x}; 0, 0.2^2) + \alpha N(\tilde{x}; 5, 0.2^2) \\
f_4(\tilde{x}) =& 0.5N(\tilde{x}; -0.9, 0.19) + 0.5N(\tilde{x}; 0.9, 0.19) \\
f_5(\tilde{x}) =& 0.5N(\tilde{x}; 0, 0.1) + 0.5N(\tilde{x}; 0, 1.9)
\end{aligned}
$$

where $\alpha = 0.0192, \beta = 1 - 2\alpha$, and $p_*(\tilde{x}) = N(\tilde{x}; 0, 0.5)$ is the desired PDF. According to MSE, there is no difference among the above five PDFs . However, as is clear from Fig. 4.5,

they are quite different.

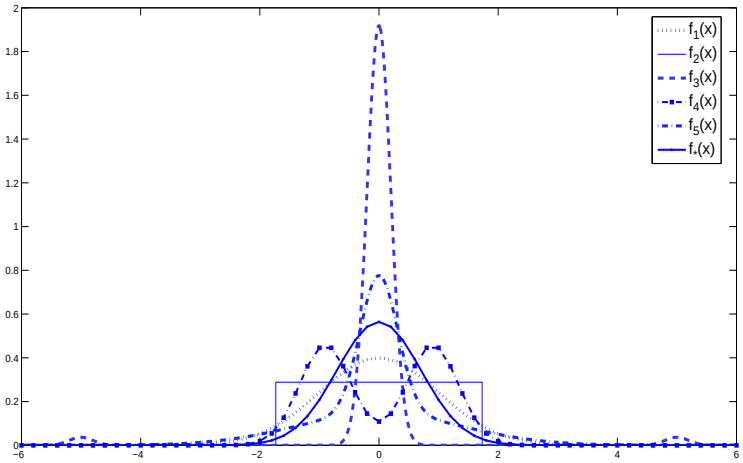


Fig. 4.5: Five PDFs with the same Mean and Covariance

Given the desired error PDF $f_*(\tilde{x}) = N(\tilde{x}; 0, P)$, $P = 0.5 : 0.1 : 0.8$, we examine CMIs and CMDs of the above five PDFs in Figs. 4.6 and 4.7 by generating 3000 samples from each PDF.

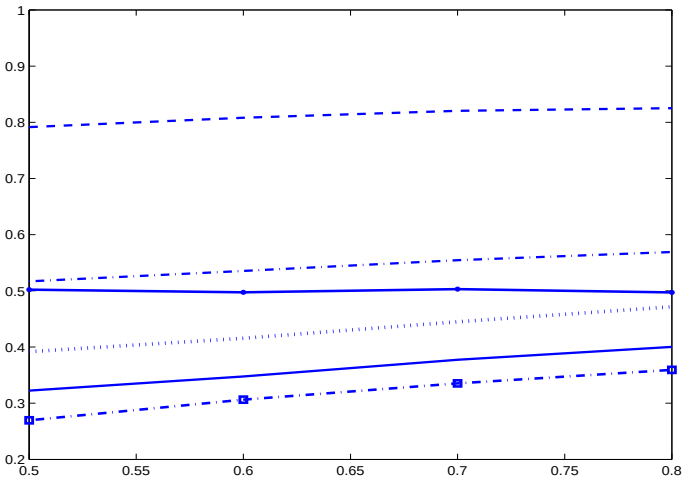


Fig. 4.6: CMI

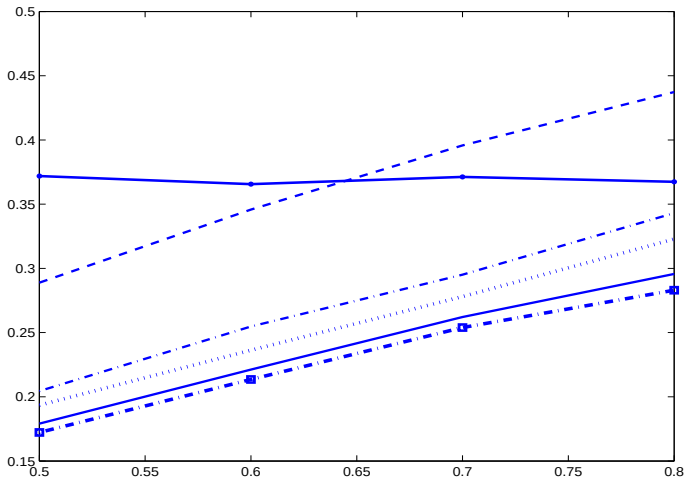


Fig. 4.7: CMD

According to CMI, $f_3(\tilde{x})$ has the highest concentration level; the next place is $f_5(\tilde{x})$; and they are more concentrated than the desired one. $f_4(\tilde{x})$ is the least concentrated one, followed by $f_2(\tilde{x})$ and $f_1(\tilde{x})$, and they are all less concentrated than the desired one. It is also clear that CMI of the desired error PDF w.r.t. itself is around 0.5. Similar to CMI, the preference level of these five PDFs are in the order of $f_3(\tilde{x}) \succ f_5(\tilde{x}) \succ f_1(\tilde{x}) \succ f_2(\tilde{x}) \succ f_4(\tilde{x})$. However, because less concentrated samples have a larger effect on CMD than on CMI, except for $f_3(\tilde{x})$, all others are less concentrated than the desired one as $P < 0.65$. Until the tolerance level $P > 0.65$, $f_3(\tilde{x})$ starts to become more concentrated. It is also clear that CMD of the desired error PDF is around 0.37. Both CMI and CMD will increase as the tolerance level P goes up.

4.4 Conclusions

The performance spectrum and the desired estimation error PDF with associated performance measures are proposed in this chapter. Both of them are designed to provide a global view on estimation performance.

The performance spectrum is based on the discussion on the interaction between

estimators and estimation criteria. A comprehensive view on the relationship among estimator, estimation criterion, and performance measures is achieved. It can be summarized as below:

- The estimation criterion is a compromise between the estimation requirements and mathematical tractability. There exists a gap between the estimation criterion and the application requirements. The estimation performance should be examined by the application requirements.

- The optimal estimator has a vivid *physical* interpretation: it is an equilibrium to balance the effects generated by the cost function. Each balance point could correspond to many different effects. From this viewpoint, using a commensurate estimation criterion easy to handle may significantly simplify the optimization procedure.

- We propose the estimation performance spectrum to compare estimators. Compared to the existing isolated and independent performance evaluation, we can gain a comprehensive view about the behavior of these estimators.

- Theorem 1 and Theorem 2 provide theoretical insights to fuse estimators with different estimation criteria into a new one with a predictable compromised performance.

The desired error PDF gives a more comprehensive characterization of application requirements. The proposed relative concentration measure and dispersion measure can reveal the concentration and dispersion level relative to the desired PDF. They provide the more detailed and comprehensive discriminations among error PDFs. The Lemma 1 also gives a reference that can indicate an estimation error PDF of an estimator with MSE C is more or less concentrated or dispersed than a Gaussian PDF with zero mean and covariance C. For example, when CMI $> \frac{1}{2}$, we say the examined error PDF will be more concentrated than the Gaussian PDF with zero mean and the specific covariance in the sense of the CMI.

Chapter 5

Model Distortion Measure and Multiple Model Algorithm

In parameter estimation and state filtering, model approximation is quite common in engineering research and development. It heavily affects estimation performance and it is of great interest to have a model distortion measure to indicate the divergence or difference between the original model and the approximate one. This Chapter is dedicated to the model distortion measure.

For model approximation, various strategies have been employed [52]. Local linearization of the model around the "best" point or simplification conditioned on past data is often the first choice. Another choice is to approximate the model by its finite-order expansion, which usually is optimal in the sense of the expansion criteria, of the measurement model or probabilistic model. For example, Taylor series expansion is quite often used to approximate a nonlinear function, which is optimal in the sense of the least-squares among all polynomial expansions at the working point; Gram-Charlie expansions of probability densities have been employed for state estimation. In contrast to the above approaches, approximating the posterior density is another choice. This type of methods includes point-mass approximation, gaussian density mixture, and spline approximation. These methods have good accuracy but high complexity.

For the above model approximation approaches, the deviation of the approximate model

from the original one seems little understood. Moreover, these approximations distort the original relation between the parameter of interest and the observation and cause performance deterioration. For example, the optimality of Taylor series expansion does not guarantee a good approximation quality. The divergence and credibility of the extended Kalman filter are serious problems in practice.

Once the original model has been approximated, the inference will completely depend on the approximate model. The deviation of approximate model from the original one is often not taken into account. As a result, it gives little idea about the consequences of the approximation. In order to analyze performance deterioration and develop new algorithms, it is quite crucial to have a measure to indicate the quality of these approximations.

Towards this goal, we analyze the setup and modeling of parameter inference and clarify its inherent vagueness. Accordingly, we draw the connection between model distortion and the difference between two probability density functions. We work out a distortion measure, and it turns out that the Kullback-Leibler (K-L) divergence, which has many nice properties, can serve this purpose.

We apply the K-L divergence to multiple model estimation. Two important problems associated with multiple model estimation are discussed. One is on the model set design; the other is on the model probability behavior. We demonstrate that the K-L divergence and K-L information are of great significance for estimation performance indication.

The rest of the Chapter is organized as follows. In Sec. 5.1, we discuss some characteristics of the parameter estimation. In Sec. 5.2, we work out a distortion measure: the K-L divergence. Its properties are also presented. In Sec. 5.3, we introduce the Multiple Model (MM) approach to parameter estimation. Then, we address the model set design problem and the model probability behavior analysis which are highly related to the proposed distortion measure. A model set design example and a multiple algorithm fusion example are illustrated in Sec. 5.4. Conclusions are presented in Sec. 5.5.

5.1 Probabilistic Model of Parameter Inference

The parameter estimation consists of four components: parameter space, observation space, probabilistic relation between the parameter space and the observation space and estimation criteria [57]. The probabilistic relation between the parameter space and the observation space governs the effect of the parameter on the observation. The parameter is estimated from the observation based on the probabilistic relation, most often, by optimization procedure according to the estimation criterion.

The probabilistic **relation** between the parameter space and the observation space can be characterized by an observation model describing a physical phenomenon, or by a probabilistic model, such as the joint distribution of the parameter and the observation, or likelihood of the parameter. For example, the contaminated observation z of parameter x by noise v can be modelled as $z = x + v$ or by their joint PDF $f(x, z)$. The parameter can be inferred based only on its probabilistic model. Therefore, the probabilistic model is complete and sufficient in the sense of carrying all the information needed for inference.

The goal of inference is defined by the estimation criterion. The inference is to find a **function** from the observation space to the parameter space which is optimal according to the estimation criterion. The relationship between estimation criterion and model approximation is not clear. In other words, how to approximate the model to favor a specific estimation criterion has not been revealed theoretically. Also the credibility of the estimator could also be different [34].

One way to measure the model distortion is to appraise the model in the sense that it can facilitate inference with good enough results in most situations and for most estimation criteria. Accordingly, the quality of the approximate model depends only on its ability to characterize the underlying truth and to provide a credible probabilistic relation for inference. Therefore, we will measure the model distortion independent of the the selection of estimation

criteria.

Since the probabilistic relation between the parameter space and observation space is completely characterized by their joint probability density function, the model difference is equivalent to the difference between two PDFs. We will quantify the model difference by comparing two PDFs.

5.2 Model Distortion Measure

5.2.1 Defining the Measure

Model approximation distorts the relation between the parameter of interest and its observation. How to quantify this distortion is the key to having a meaningful measure. Practically speaking, given the data and several models (multiple hypotheses), hypothesis testing is a way to find the fittest model to the data with minimum decision error or cost [19]. We follow this procedure to quantify the difference between two probabilistic models.

The joint PDF $f(x, z)$ of parameter x and observation z can be written as $f(\mathbf{x})$, where $\mathbf{x} = (x, z)$. Given two PDFs $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$, we measure their difference as follows.

First, we use likelihood ratio $\lambda(\mathbf{x}) = \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})}$ to characterize the difference of the parameter and observation pair (x, z) in both models because it has the most separation power in the Neyman-Pearson sense. Second, the average of $\lambda(\mathbf{x})$ in respect of either $f_0(\mathbf{x})$ or $f_1(\mathbf{x})$ gives their average power to separate them, respectively. It is reasonable to expect that $\lambda(\mathbf{x}) = \alpha$ and $\lambda(\mathbf{x}) = 1/\alpha$ have the same power to quantify their difference due to the symmetrical position of $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$ in this definition. So the geometric mean of the likelihood ratio $\bar{\lambda}_i$, $i = 0,\ 1$,

$$\bar{\lambda}_i = \exp(\int f_i(\mathbf{x}) \ln \lambda(\mathbf{x}) d\mathbf{x})$$

is suitable to measure the average difference. The geometrical mean is balanced around 1 as

argued in [36]. Third, we use the ratio of these averages to define the total difference D:

$$\begin{aligned} D(f_0(\mathbf{x}), f_1(\mathbf{x})) =& \bar{\lambda}_0/\bar{\lambda}_1 \\ =& \exp(\int (f_0(\mathbf{x}) - f_1(\mathbf{x})) \ln \lambda(\mathbf{x}) d\mathbf{x}) \end{aligned}$$

Note, according to the maximum entropy theory, $\int f_0(\mathbf{x}) \ln \lambda(\mathbf{x}) d\mathbf{x} = \int f_0(\mathbf{x}) \ln f_0(\mathbf{x}) d\mathbf{x} - \int f_1(\mathbf{x}) \ln f_0(\mathbf{x}) d\mathbf{x} > 0$, $\bar{\lambda}_0 > 1$. Likewise, $\bar{\lambda}_1 < 1$. The larger D is, the larger the distortion would be. Let $J = \ln D$, we have

$$\begin{aligned} J(f_0(\mathbf{x}), f_1(\mathbf{x})) =& \int (f_0(\mathbf{x}) - f_1(\mathbf{x})) \ln \lambda(\mathbf{x}) d\mathbf{x} \\ =& \int f_0(\mathbf{x}) \ln \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} d\mathbf{x} + \int f_1(\mathbf{x}) \ln \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} d\mathbf{x} \\ =& I(0:1) + I(1:0) \end{aligned}$$

where $I(i:j) = \int f_i(\mathbf{x}) \ln \frac{f_i(\mathbf{x})}{f_j(\mathbf{x})} d\mathbf{x}$, $i, j = 0, 1$. Similarly, for $\lambda(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}$, following the same procedure, we have $J(f_1(\mathbf{x}), f_0(\mathbf{x})) = J(f_0(\mathbf{x}), f_1(\mathbf{x}))$. Later, we simply use $J(0, 1)$ to measure the difference between $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$. Again, the larger J is, the larger the distortion would be.

5.2.2 Comments

J is the Kullback-Leibler (K-L) divergence and I is the K-L information. The K-L information is always non-negative and is zero if and only if $f_0(\mathbf{x}) = f_1(\mathbf{x})$. A stronger property known as Pinsker's inequality is [14]

$$||f_0(\mathbf{x}) - f_1(\mathbf{x})|| \leq \sqrt{2I(1:0)} \tag{5.1}$$

where $||f_0(\mathbf{x}) - f_1(\mathbf{x})|| = \int |f_0(\mathbf{x}) - f_1(\mathbf{x})| d\mathbf{x}$ is the L^1 distance between $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$, serving as the lower bound of K-L information.

From the viewpoint of **model selection**, given the truth $\mathbf{x} \sim f_\theta(\mathbf{x})$, for model selection from candidates $f_{\theta_j}(\mathbf{x})$ using the minimum K-L information is equivalent to using the

maximum likelihood principle and the maximum entropy theorem, i.e.

$$\begin{aligned}\min_{\theta_j} I(\theta : \theta_j) &= \min_{\theta_j} \lim_{n\to\infty} \sum\nolimits_{i=1}^{n} \left(\ln f_\theta(\mathbf{x}_i) - \ln f_{\theta_j}(\mathbf{x}_i)\right) \\ &\Rightarrow \max_{\theta_j} \lim_{n\to\infty} \sum\nolimits_{i=1}^{n} \ln f_{\theta_j}(\mathbf{x}_i) \\ &\Rightarrow \max_{\theta_j} \lim_{n\to\infty} \prod\nolimits_{i-1}^{n} f_{\theta_j}(\mathbf{x}_i)\end{aligned}$$

From the viewpoint of **making decision**, given an observation, H_i, $i = 0, 1$, is the hypothesis that $\mathbf{x}$ is from the population with PDF $f_i(\mathbf{x})$, and then by using the likelihood ratio decision rule it follows that the definition of the Kullback-Leibler information $I(1:0)$ is the mean information per observation from population $f_1(\mathbf{x})$ for discrimination in favor of H_1 against H_0 [19].

Moreover, from the viewpoint of the **large deviation theory**, defining the false alarm as the probability $P_F = Pr(\text{decide } H_1|H_0 \text{ is true})$, by the weak law of large numbers, P_F will converge to zero for a fixed detection probability as the sample size Nincreases. The large deviation theory says it converges exponentially fast and the convergence rate is determined by the K-L information $I(1:0)$ [8, 12]. That is $P_F = \lim_{N\to\infty} \exp(-NI(1:0))$. The smaller the convergence rate is, the more probable the false alarm occurs, and the more difficult it is to discriminate them.

The K-L information is often called *distance*, but it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality. The K-L divergence is the sum of $I(1:0)$ and $I(0:1)$. It is symmetric, but still not a distance for not satisfying the triangle inequality.

Suppose $f_0(\mathbf{x})$ is the true model and $f_1(\mathbf{x})$ an approximate one. For model approximation, both the true probabilistic model and the approximate one are available. As to tell the approximate model from the truth by the data, $I(0:1)$ measures the deviation; the distortion of the probabilistic measure caused by using the approximate model for inference is measured by $I(1:0)$. For statistical inference, the K-L divergence is more comprehensive than the

K-L information. It can quantify the deviation from the truth and can indicate the effects of model approximation on estimation.

We stress that the K-L divergence will be infinite once two PDFs have different support sets. This implies that the approximate model will lead to estimators which can never reach some true parameter values or may produce some values the truth can not be. In this case, there may exist a better way to do approximation. For some applications, this may not matter; for others, it may. Practitioners need to pay more attention to this case to have a reasonable approximation appraisal.

The significance of the model distortion measure will be exemplified by its application to the multiple model (MM) algorithm. We will apply the model distortion measure to address two important problems—model set design and model behavior analysis—associated with multiple model algorithms next. Not only can the Kullback-Leibler information indicate the model distortion but it is also highly related to the model probability behavior.

5.3 Multiple Model Algorithm

Multiple-model (MM) estimation is a powerful approach to adaptive estimation. It is particularly good for problems involving structural as well as parametric changes, such as hybrid estimation and problems involving model uncertainty [1]. Compared to the conventional estimation, which relies entirely on a single "best" model for inference, a multiple model algorithm also takes the model uncertainty as an integral part of the inference. It uses a group of models to do inference simultaneously and produces the overall output according to each model's fitness to observations. The fundamental idea is to utilize multiple models to cover the true model and combine the estimator of each model by the total expectation theorem through Bayesian procedure [22, 42, 44, 43, 27, 26]. It is cost-effective, robust, and of a parallel structure.

Given a probabilistic model $f(z, x, W)$, where z is the observation, x is the parame-

ter of interest and W is a random nuisance continuous-valued parameter, the MMSE-MM algorithm for estimating x from z can be formulated as

$$\begin{aligned}
\hat{x} \equiv & E_{x|z(W)}\left[x|z\right] = E_{x|z(W)}[E_{x|W,z}\left[x|W,z\right]|z] && (5.2)\\
\approx & E_{x|z(\omega)}[E_{x|\omega,z}\left[x|\omega,z\right]|z] && (5.3)\\
= & \sum_{i=1}^{M} E_{x|z(\omega_i)}\left[x|\omega_i,z\right] Pr(\omega_i|z) && (5.4)\\
\text{mse}(\hat{x}|z) \equiv & E_{x|z(W)}\left[((x-\hat{x})'(x-\hat{x})|z\right] && (5.5)\\
= & E_{x|z(W)}\left[\text{mse}(\hat{x}|W,z)|z\right] && (5.6)\\
\approx & E_{x|z(\omega)}\left[\text{mse}(\hat{x}|\omega,z)|z\right] && (5.7)\\
= & \sum\nolimits_{i=1}^{M}\left[\text{mse}\left(\hat{x}|\omega_i,z\right) + (\hat{x}_i-\hat{x})'(\hat{x}_i-\hat{x})\right] Pr(\omega_i|z) && (5.8)\\
Pr(\omega_i|z) = & Pr_0(\omega_i) f(z;\omega_i)/f_C(z) && (5.9)
\end{aligned}$$

where $E_{x|z(W)}\left[x|z\right]$ stands for the expectation of x conditioned on z, ω denotes a discrete random variable with a finite sample set $\{\omega_i\}$, $i = 1, ..., M$, to approximate W, $Pr_0(\omega_i)$, $f(z;\omega_i)$ and $Pr(\omega_i|z)$ are the initial probability, likelihood and posterior probability of the ith model, respectively, $\text{mse}(\hat{x})$ is the mean square error of $\hat{x}$, $\hat{x}_i = E\left[x|\omega_i,z\right]$, and $f_C(z) = \sum_{j=1}^{M} f(z;\omega_j)Pr_0(\omega_j)$ is the common normalization term. From (5.2) to (5.3) and (5.6) to (5.7), the original nuisance parameter W is approximated by the model set $\{\omega_i\}$ with the initial probabilities $\{Pr_0(\omega_i)\}$.

The approximation is mainly due to the complexity of the original model.

As the multiple observation z^N is concerned, the probabilistic model is denoted by $f(z^N, x, W)$, and the model probability can be calculated either in a batch or in a recursive way as below

$$\begin{aligned}
Pr(\omega_i|z^N) = & \frac{Pr_0(\omega_i) f(z^N;\omega_i)}{\sum_{j=1}^{M} Pr_0(\omega_j) f(z^N;\omega_j)} && (5.10)\\
= & \frac{Pr_0(\omega_i|z^{N-1}) f(z_N;\omega_i z^{N-1})}{\sum_{j=1}^{M} Pr_0(\omega_j|z^{N-1}) f(z_N;\omega_j,z^{N-1})} && (5.11)
\end{aligned}$$

We will investigate the behavior characteristics of (5.9), (5.10) and (5.11) next.

5.3.1 Model Set Design

Performance of the MM estimator depends heavily on the approximation of $\mathbf{W}$ from (5.6) to (5.7) or the design of $\{\omega_i\}$ in (5.7). A meaningful definition of the coverage of each model ω_i in the parameter space is the key to model set design. Our results on quantization of the PDF $f(\mathbf{W})$ w.r.t. different criteria are given in [22, 42, 44, 43, 27, 26] where the nuisance parameter is quantized in its own parameter space without considerations of the model distortion. We, here, take further considerations of the deviation of the approximate model from the true one. This is essential to model set design. For an extreme example, if $f(\mathbf{x}, \mathbf{z}, \mathbf{W}) = f(\mathbf{x}, \mathbf{z}, \omega)$ for all possible $\mathbf{W}$ and a corresponding ω, the quantization of $f(\mathbf{W})$ will not affect the inference of $\mathbf{x}$ from $\mathbf{z}$.

Once we have this distortion measure, we can follow some quantization methods in [42] such that the model distortion is also taken into consideration and the performance should not deteriorate too much. Especially, when $f(\mathbf{W})$ has a uniform distribution, we can design the model set as

$$\omega^N = \arg\min_{\omega^n} \max_{w \in s_{\omega_i}} J(w, \omega_i) < D$$

where ω^n is any model set $\{\omega_i, i = 1, ..., n\}$, s_{ω_i} is the parameter region covered by ω_i, D is a predetermined value of the maximum allowable distortion level. We pick the model set ω^N with N equal to the minimum allowable n.

In the next section, we apply the K-L divergence as the model distortion measure to design a proper model set for an illustrative example.

5.3.2 Model Probability Behavior

An important issue in the multiple model estimation is to understand its behavior, such as the estimation consistency when the true model is in the model set, and the consistent discrimination between models [5, 6, 4, 7]. We will focus on investigating the behavior of the model probability such that we can perceive its evolution when the truth is inside or

outside the model set. It turns out that model probability behavior heavily depends on the model *distance* between the model in the model set and the true one, which is defined by the Kullback-Leibler (K-L) information.

First, examining (5.9), we take *log* on both sides and have

$$\ln Pr(\omega_i|z) = \ln Pr_0(\omega_i) + \ln \frac{f(z;\omega_i)}{f(z;w)} - \ln \frac{f_C(z)}{f(z;w)} \tag{5.12}$$

where w is one realization of W and $f(z;w)$ is the true observation generation model at w. Take geometric average on both sides of (5.9) w.r.t. $f(z;w)$. Then

$$\begin{aligned} e^{E_{z|w}[\ln Pr(\omega_i|z)]} =& e^{\int \ln Pr(\omega_i|z) f(z;w) dz} = Pr_0(\omega_i) e^{I(w:C)-I(w:\omega_i)} \\ =& Pr_0(\omega_i) e^{I(w:C)-I(w:\omega_i)} \end{aligned} \tag{5.13}$$

where $I(w:\omega_i)$ is the Kullback-Leibler (K-L) information between $f(z;w)$ and $f(z;\omega_i)$.

Because geometric mean is no more than the arithmetical mean, the total of geometric mean of the model probability is no more than 1. After renormalization, we have

$$\begin{aligned} \bar{Pr}^G(\omega_i|z) =& Pr_0(\omega_i) \frac{e^{-I(w:\omega_i)}}{\sum_{j=1}^M Pr_0(\omega_j) e^{-I(w:\omega_j)}} \\ =& Pr_0(\omega_i) \left(\sum_{j=1}^M Pr_0(\omega_j) e^{I(w:\omega_i)-I(w:\omega_j)} \right)^{-1} \end{aligned}$$

Although $\bar{Pr}^G(\omega_i|z)$ is not a probability, it is the normalized geometric average of the model probability w.r.t. the true model. The summation of all $\bar{Pr}^G(\omega_i|z)$ is still one. Compared to the prior model probability, on average, the model probability update will change at the rate

$$r(\omega_i) = \frac{\bar{Pr}^G(\omega_i|z)}{Pr_0(\omega_i)} = \left(\sum_{j=1}^M Pr_0(\omega_j) e^{I(w:\omega_i)-I(w:\omega_j)} \right)^{-1} \tag{5.14}$$

Therefore, besides the prior model probability, on average, the change rate of model probability only depends on the K-L information between the true model and the models in model set. Assuming $I(w:\omega_i)$ is the minimum among $I(w:\omega_j)$, $j = 1, ..., M$, we say model $f(z;\omega_i)$

is closest to the true observation generation model $f(z;w)$ in terms of the K-L information. Then

$$r(\omega_i) = \left(\sum_{j=1}^{M} Pr_0(\omega_j) e^{I(w:\omega_i)-I(w:\omega_j)}\right)^{-1} \geq \left(\sum_{j=1}^{M} Pr_0(\omega_j)\right)^{-1} = 1 \tag{5.15}$$

The model probability of the model closest to the true one will increase on average. Similarly, the model probability of the farthest model will decrease on average.

Moreover,

$$\frac{\bar{Pr}^G(\omega_i|z)}{\bar{Pr}^G(\omega_j|z)} = \frac{Pr_0(\omega_i)}{Pr_0(\omega_j)} \exp(I(w:\omega_j) - I(w:\omega_i)) \tag{5.16}$$

The ratio between the two average model probabilities $\bar{Pr}^G(\omega_i|z)$ and $\bar{Pr}^G(\omega_j|z)$ will increase or decrease exponentially fast, which only depends on their K-L information difference. Both (5.15) and (5.16) also indicate that the model probability of the closest model will increase at the highest rate on average.

Second, examining (5.11) for the recursive form, similarly, we have

$$\begin{aligned}
\ln Pr(\omega_i|z^N) =& \ln Pr(\omega_i|z^{N-1}) + \ln \frac{f(z_N|\omega_i, z^{N-1})}{f(z_N, |w, z^{N-1})} - \ln \frac{f_C(z_N|z^{N-1})}{f(z_N, w, z^{N-1})} \\
e^{\left(E\left[\ln Pr(\omega_i|z^N)\right]\right)} =& Pr(\omega_i|z^{N-1}) e^{I(w:C;z^{N-1})-I(w:\omega_i;z^{N-1})} \qquad (5.17) \\
\bar{Pr}^G(\omega_i|z^N) =& Pr(\omega_i|z^{N-1}) \frac{e^{-I(w:\omega_i;z^{N-1})}}{\sum_{j=1}^{M} Pr(\omega_j|z^{N-1}) e^{-I(w:\omega_j;z^{N-1})}} \\
=& Pr(\omega_i|z^{N-1}) \times \left(\sum_{j=1}^{M} Pr(\omega_j|z^{N-1}) e^{I(w:\omega_i;z^{N-1})-I(w:\omega_j;z^{N-1})}\right)^{-1} \qquad (5.18)
\end{aligned}$$

where $I(w:\omega_i;z^{N-1})$ is the K-L information between $f(z_N|w,z^{N-1})$ and $f(z_N|\omega_i,z^{N-1})$, and $f_C(z_N|z^{N-1}) = \sum_{j=1}^{M} Pr_0(\omega_j|z^{N-1}) f(z_N;\omega_j,z^{N-1})$. The average change rate of the model probability update is

$$r(\omega_i) = \left(\sum_{j=1}^{M} Pr_0(\omega_j) e^{I(w:\omega_i;z^{N-1})-I(w:\omega_j;z^{N-1})}\right)^{-1} \tag{5.19}$$

In the recursive case, the average change rate $r(\omega_i)$ is random—it depends on the previous observations z^{N-1}. Given z^{N-1}, the model probability of the closest model to the true one

will increase on average, and the model probability of the farthest model will decrease on average.

Again,

$$\frac{\bar{Pr}^G(\omega_i|z^N)}{\bar{Pr}^G(\omega_j|z^N)} = \frac{Pr_0(\omega_i|z^{N-1})}{Pr_0(\omega_j|z^{N-1})} e^{I(w:\omega_i;z^{N-1})-I(w:\omega_j;z^{N-1})} \tag{5.20}$$

the ratio between two average model probabilities $\bar{Pr}^G(\omega_i|z^N)$ and $\bar{Pr}^G(\omega_j|z^N)$ will increase or decrease exponentially fast. Given z^{N-1}, the rate depends only on their K-L information difference.

Third, we examine (5.10) as N goes to infinity to reveal the convergence of the model probability .

A very general and rigorous results about the asymptotic behavior of posterior probability when the true model is not inside the model set has been presented in [7], but the results rely on heavy mathematical machinery and the insight is not straightforward. We present the asymptotic behavior of the model probability in a very succinct way. It is easy to show

$$\begin{aligned}
\lim_{N\to\infty} Pr(\omega_i|z^N) &= \lim_{N\to\infty} \frac{Pr_0(\omega_i) f(z^N;\omega_i)}{\sum_{j=1}^M Pr_0(\omega_j) f(z^N;\omega_j)} \\
&= \lim_{N\to\infty} \frac{Pr_0(\omega_i) e^{N\times\frac{1}{N}(\ln f(z^N;\omega_i)-\ln f(z^N;\omega))}}{\sum_{j=1}^M Pr_0(\omega_j) e^{N\times\frac{1}{N}(\ln f(z^N;\omega_j)-\ln f(z^N;\omega))}} \\
&= \lim_{N\to\infty} \frac{Pr_0(\omega_i) e^{-NI(w:\omega_i)}}{\sum_{j=1}^M Pr_0(\omega_j) e^{-NI(w:\omega_j)}} \\
&= \lim_{N\to\infty} \frac{Pr_0(\omega_i)}{\sum_{j=1}^M P_0(\omega_j) \exp^{N(I(w:\omega_i)-I(w:\omega_j))}}
\end{aligned} \tag{5.21}$$

Assume $i = \arg\min_j(I(w:\omega_j))$, and $I(w:\omega_i) \neq I(w:\omega_j)$, $j = 1, ..., M$ and $j \neq i$, that is ω_i is the unique model closest to the true w. then

$$\lim_{N\to\infty} P_0(\omega_j) e^{N(I(w:\omega_i)-I(w:\omega_j))} = 0 \tag{5.22}$$

and

$$\lim_{N\to\infty} Pr(\omega_i|z^N) = 1, \quad \lim_{N\to\infty} Pr(\omega_j|z^N) = 0, \quad \forall j \neq i \tag{5.23}$$

This indicates the posterior model probability of the model closest to the true one will converge to 1 as the sample size goes to infinity.

Moreover,

$$\begin{aligned}\lim_{N\to\infty}\frac{Pr(\omega_i|z^N)}{Pr(\omega_j|z^N)} &= \lim_{N\to\infty}\frac{Pr_0(\omega_i)}{Pr_0(\omega_j)}e^{N(I(w:\omega_j)-I(w:\omega_i))} \\ &= \lim_{N\to\infty}\frac{Pr(\omega_i|z^{N-1})}{Pr(\omega_j|z^{N-1})}e^{I(w:\omega_j)-I(w:\omega_i)} \end{aligned} \tag{5.24}$$

The model probability ratio between any two models will increase or decrease exponentially fast at the rate of their K-L information difference as sample size goes to infinity.

5.4 Illustrative Examples

Two examples are designed to illustrate the model set design and the fusion of multiple algorithms, respectively.

Example 1. Model Set Design

Consider a single linear noisy measurement z of a parameter x:

$$z = wx + v$$

$$z = \omega x + v$$

where $x \sim N(x; m, 1)$, $v \sim N(v; 0, 1)$, $w \sim U(w; 0, 1)$ and they are independent of each other. Then the probabilistic model of this estimation problem is its joint probability density function

$$f(z, x, w) = N(z; wx, 1)N(x; m, 1)U(w; 0, 1)$$

Following the procedure in [3, pp. 123-128], its best linear unbiased estimator (BLUE) is

$$\begin{aligned}\hat{x}^{BLUE} &= \bar{x} + C_{xz}C_z^{-1}(z-\bar{z}) \\ &= m + 0.5(1.333 + 0.083m^2)^{-1}(z - 0.5m) \qquad (5.25)\\ P &= P_0 - C_{xz}C_z^{-1}C'_{xz} \\ &= 1 - 0.25(1.333 + 0.083m^2)^{-1} \qquad (5.26)\end{aligned}$$

where P is the mean-sqrare error of $\hat{x}^{BLUE}$.

First, we use $z = 0.5x + v$ to approximate the original system, then the estimator is simply

$$\hat{x}^{UM} = m + 0.4(z - 0.5m),\ P^{UM} = 0.8 \qquad (5.27)$$

where $\hat{x}^{UM}$ stands for the the BLUE based on the uni-model of ω= 0.5.

We compare only the average normalized estimation error square (ANEES) and root-mean-square error (RMSE) of these two estimators. The ANEES is defined by [34]

$$\text{ANEES} = \frac{1}{Nn}\sum_{i=1}^{N}(x_i - \hat{x}_i)'P_i^{-1}(x_i - \hat{x}_i) \qquad (5.28)$$

where $(x_i - \hat{x}_i)$ and P_i are the estimation error and error covariance in the ith run, n is the dimension of x, and N is the total number of runs. If the estimation error and the estimated covariance match each other, i.e., the estimator is credible, the ANEES is will be close to 1.

As is clear from RMSE and ANEES of both $\hat{x}^{BLUE}$ and $\hat{x}^{UM}$ in Fig. 5.4 (a) and (b), respectively and, not surprisingly, the approximation by replacing $w \sim U(w; 0, 1)$ with ω= 0.5 is too crude, the performance deteriorates dramatically when $|m| > 2$.

We measure the K-L divergence between two PDFs of $f(z, x, w_1)$ and $f(z, x, w_2)$. Assuming one is the truth and the other is the model used by estimator, the K-L divergence is

$$\begin{aligned}J(1,2) &= \iint (f(z,x,w_1) - f(z,x,w_2)) \ln\frac{f(z,x,w_1)}{f(z,x,w_2)}dxdz \\ &= 2\left(m^2+1\right)(w_1 - w_2)^2 \qquad (5.29)\end{aligned}$$

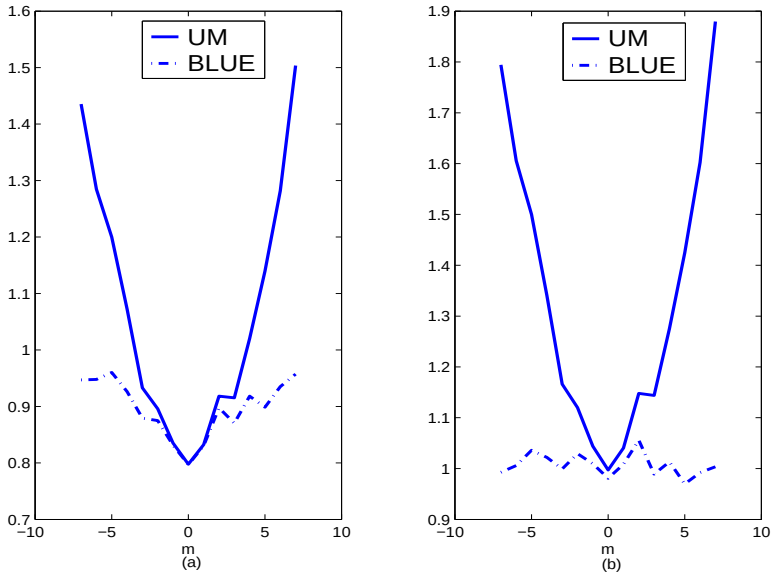


Fig. 5.1: $\hat{x}^{BLUE}$ vs. $\hat{x}^{MM}$ (a) RMSE (b) ANEES

where $f(z, x, w = w_i) = \frac{1}{2\pi} \exp\left(-\frac{(z-xw_i)^2+(x-m)^2}{2}\right)$, $i = 1, 2$.

First, $J(1, 2)$ is symmetrical around $m = 0$ w.r.t. m for all possible w_1 and w_2. Previously, we argued that the K-L divergence can indicate the performance deterioration due to the model approximation, and the smaller the K-L divergence is, the better the estimation performance is. Accordingly, the performance of the approximate model of each true w w.r.t. m should have a symmetrical pattern, so does the average. That is, the uni-model estimation performance should be symmetrical. By checking Fig. 5.4, the symmetrical patten of RMSE and ANEES agrees with this assumption. So the K-L divergence can quite effectively indicate performance changes.

Second, $\hat{x}^{UM}$ and $\hat{x}^{BLUE}$ have similar RMSE and ANEES when $-2 \leq m \leq 2$. By the minimax principle, $\min\limits_{0\leq\omega\leq 1} \max\limits_{|m|\leq 2, 0\leq w\leq 1} J(1, 2) = 2\left(m^2 + 1\right)(w - \omega)^2 = 2.5$ with $\omega = 0.5$, $m = \pm 2$, $w = 1$ or 0. We can assume that $J(1, 2) \leq 2.5$ indicates $f(z, x, w = w_1)$ and $f(z, x, w = w_2)$ are quite close for this problem such that the estimation performance of the approximate model will not deteriorate too much.

Third, we also use multiple models to handle this problem. Let the minimum divergence between the true w and the nearest model ω in the model set be less than 2.5 in the worst case. That is $\max_{|m|\leq 7} 2\left(m^2+1\right)(w-\omega)^2 \leq 2.5$. We have $100(w-\omega)^2 \leq 2.5$ and $|w-\omega| \leq \sqrt{\frac{1}{40}}$. The minimum model number over $w \in [0,1]$ will be $ceil\left[\frac{1}{2|w-\omega|}\right] = 4$. We pick $\omega = [0,\ 0.33,\ 0.66,\ 0.99]$ as our model set.

For this estimation problem, the multiple model algorithm simply becomes a typical Bayesian inference procedure as is summarized in Table 5.1.

Table 5.1: MM Approach to This Example

1. Model prior information (for i=1,2,...,M):
   Model probability: $\mu^{(i)} \triangleq P\{\omega^{(i)}\} = 1/M$
   State: $x_0^{(i)} = x_0$
   Covariance: $P_0^{(i)} = P_0$
2. Model-conditioned Estimation (for i=1,2,..., M):
   Measurement residual: $\tilde{z}^{(i)} = z - \omega^{(i)} x_0^{(i)}$
   Residual covariance: $S^{(i)} = \omega^{(i)} P_0^{(i)} (\omega^{(i)})' + \sigma_v^2$
   Estimation gain: $K^{(i)} = P_0^{(i)} (\omega^{(i)})' \left(S^{(i)}\right)^{-1}$
   Update estimator: $\hat{x}^{(i)} = x_0^{(i)} + K^{(i)} \tilde{z}^{(i)}$
   Update covariance: $P^{(i)} = P_0^{(i)} - K^{(i)} (S^{(i)})^{-1} (K^{(i)})'$
3. Model probability update (for i=1,2,...,M):
   Model likelihood: $L^{(i)} = f(\tilde{z}^{(i)} | \omega^{(i)}) = \frac{\exp[-(\tilde{z}^{(i)})'(S^{(i)})^{-1}\tilde{z}^{(i)}/2]}{|2\pi S^{(i)}|^{1/2}}$
   Model probability: $\mu^{(i)} = \frac{\mu^{(i)} L^{(i)}}{\sum_j^M \mu^{(j)} L^{(j)}}$
4. Estimate fusion:
   Overall estimate: $\hat{x} = \sum_i^M \mu^{(i)} \hat{x}^{(i)}$
   Overall covariance: $P = \sum_i^M [P^{(i)} + (\hat{x} - \hat{x}^{(i)})(\hat{x} - \hat{x}^{(i)})'] \mu^{(i)}$

As is clear from RMSE and ANEES of both $\hat{x}^{BLUE}$ and $\hat{x}^{MM}$ in Fig. 5.4 (a) and (b), the MM estimator has better accuracy and is very credible. The MM algorithm with the designed model set has better performance than$\hat{x}^{BLUE}$ does. Obviously, using the mm method to approximate the original problem improves the performance considerably compared with the estimator using only one model.

The above example demonstrates the validity of the K-L divergence as a measure of performance deterioration is significant. It can predict the performance deterioration with quite good accuracy, and may serve an an effective model distortion measure for model set design for MM approach.

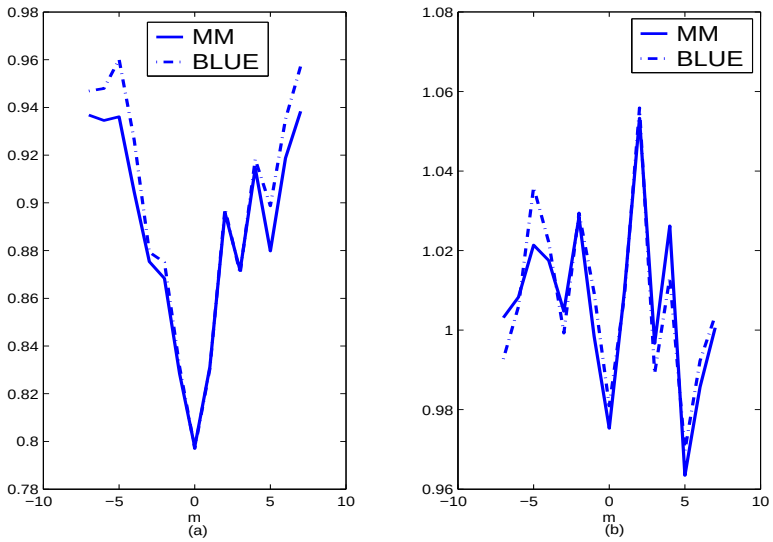


Fig. 5.2: $\hat{x}^{BLUE}$ vs. $\hat{x}^{MM}$ (a) MSE (b) ANEES

Example 2. Fusion of Multiple Algorithms

In an MM algorithm, each model can also be viewed as a hypothesis to characterize the observations and each model contributes to the overall output. The overall output is the summation of the estimator inferred from each model and weighted by its model probability which measures the fitness of the observation to each model or hypothesis. The behavior of the model probability indicates that on average the model probability update increases the weight of the estimator of the fittest model. Moreover, the model probability of the fittest model finally approaches one as the observation size goes to infinity. This remark applies to both the parametric case and filtering case, regardless of whether the true model is inside or outside the model set.

In practice, we constantly need to compare multiple algorithms and make hard decisions to pick the better one for real applications. We can also view each algorithm as a hypothesis to characterize the observations associated with a state or parameter of interest. Similarly, we can apply the model probability update structure to evaluate the quality of each algorithm or hypothesis, and accordingly combine these algorithms by their weighted summation, where the weight is calculated in a similar manner as for the model probability in a multiple model algorithm. Particularly, in nonlinear estimation problems, the performance of an algorithm depends on the working points. Therefore, the preference of an algorithm to another may change at different working points. This exhibits difficulties to conclude which algorithm is better in real applications. In such a case, combining the multiple algorithms by a weighted summation may produce a robust result at all working points, because on average the weight update can automatically increase the share of the better hypothesis.

Next, we present an illustrative example.

Consider target localization using a radar or active sonar with range r and bearing b measurements:

$$r = \sqrt{\mathsf{x}^2 + \mathsf{y}^2} + v_r, \quad b = \tan^{-1}\frac{\mathsf{y}}{\mathsf{x}} + v_b$$

where (x, y) is the true location of a stationary target in the Cartesian coordinates, $v = [v_r, v_b]' \sim \mathcal{N}(0, R)$ is the zero-mean Gaussian measurement noise with covariance $R = \text{diag}(\sigma_r^2, \sigma_b^2)$.

Suppose that estimates of the target location in the Cartesian coordinates are needed. The known prior information of the target location $x = [\mathsf{x}, \mathsf{y}]'$ is: $x \sim \mathcal{N}(\bar{x}, C_x)$, that is, x is Gaussian distributed with mean $\bar{x}$ and covariance C_x. We consider three estimators: $\hat{x}_{\text{UB}}$, $\hat{x}_{\text{BLUE}}$—based on the unbiased measurement conversion method of [50], and the recursive best linear unbiased estimation (BLUE) method of [68], respectively—and the fused one. The fused estimator is the weighted summation of $\hat{x}_{\text{UB}}$ and $\hat{x}_{\text{BLUE}}$, i.e.,

$$\hat{x}_{\text{Fused}} = Pr(\hat{x}_{\text{BLUE}}|z)\hat{x}_{\text{BLUE}} + Pr(\hat{x}_{\text{UB}}|z)\hat{x}_{\text{UB}}$$

The weight is calculated by the model probability update formulation. Assuming two estimators have the same prior probability, and their posterior probabilities are

$$Pr(\hat{x}_{\text{BLUE}}|z) = \frac{Pr_0(\hat{x}_{\text{BLUE}})f(z;\hat{x}_{\text{BLUE}})}{Pr_0(\hat{x}_{\text{BLUE}})f(z;\hat{x}_{\text{BLUE}}) + Pr_0(\hat{x}_{\text{UB}})f(z;\hat{x}_{\text{UB}})} \tag{5.30}$$

$$= \frac{f(z;\hat{x}_{\text{BLUE}})}{f(z;\hat{x}_{\text{BLUE}}) + f(z;\hat{x}_{\text{UB}})} \tag{5.31}$$

The second equality is because of the equal prior assumption. Similarly, $Pr(\hat{x}_{\text{UB}}|z) = 1 - Pr(\hat{x}_{\text{BLUE}}|z)$.

A simulation with 500 Monte Carlo runs was conducted in which the target location was generated as a random variable $x \sim \mathcal{N}(\bar{x}, C_x)$, $\bar{x} = [-50000\ 20000]'$, $C_x = diag([500^2\ 500^2])$. Unless stated otherwise, all plots were obtained from $\hat{x}_{\text{UB}}$, $\hat{x}_{\text{BLUE}}$ and the fused one.

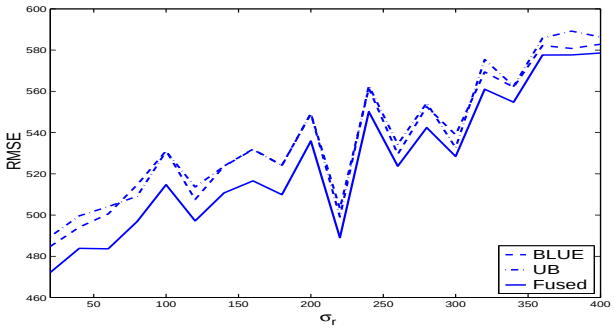


Fig. 5.3: RMSE vs.σ_r

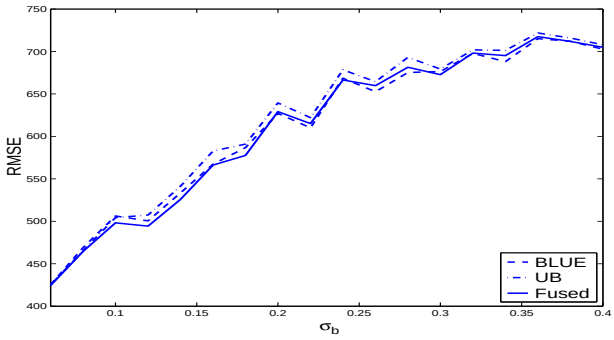


Fig. 5.4: RMSE vs.σ_b

Fig. 5.3 shows the root mean square error (RMSE) of $\hat{x}_{\mathrm{UB}}$, $\hat{x}_{\mathrm{BLUE}}$ and the fused one vs. σ_r, $20 < \sigma_r < 400$, at fixed $\sigma_b = 0.1$. As is clear from the figure, neither $\hat{x}_{\mathrm{UB}}$ nor $\hat{x}_{\mathrm{BLUE}}$ can dominate the other. This is not a surprise for such a nonlinear estimation problem because the performance preference of one algorithm to another depends on the working points. The fused one has a smaller estimation error than the other two. This indicates that the weight calculated by (5.30) can help pick the better estimator more often on average such that the fused estimator has better performance.

Fig. 5.4 shows the RMSE of $\hat{x}_{\mathrm{UB}}$, $\hat{x}_{\mathrm{BLUE}}$ and the fused one vs. σ_b, $0.06 < \sigma_b < 0.4$, at fixed $\sigma_r = 40$. The performance of $\hat{x}_{\mathrm{BLUE}}$ dominates that of $\hat{x}_{\mathrm{UB}}$. When $\sigma_b < 0.2$, the fused one outperform the other two. When $\sigma_b > 0.2$, the fused one is a little bit worse than $\hat{x}_{\mathrm{BLUE}}$ but is better than $\hat{x}_{\mathrm{UB}}$. This is because although on average more weight is assigned to a better hypothesis, this hypothesis may not lead to a better estimator. However, on average, we still prefer the result from a better hypothesis.

Therefore, the structure of model probability update also provides a feasible way to fuse multiple estimation algorithms.

5.5 Conclusions

In this Chapter, we discussed the model distortion measure for model approximations. We have proposed to use the K-L divergence to measure the model distortion. The properties associated with K-L divergence have been explored. We applied the K-L divergence to design the model sets for the MM approach. Through an illustrative example, we demonstrate that the K-L divergence can predict the performance deterioration with quite good accuracy and can provide new insight into model approximation. Moreover, it is quite general, conceptually simple, and has high potential to provide a guideline to the development of highly adaptive algorithms.

Moreover, we investigate the model probability behavior of MM algorithm. Although

there are many varieties of MM algorithms, the insight into the behavior of the posterior model probability is quite general. It is primarily determined by the model *distance* between the model in the model set and the true one, which is quantified by the K-L information.

Regardless of whether the true model is inside or outside the model set, in a multiple model algorithm, the model closest to the true one tends to have more weight to the overall output. In other words, the performance of multiple model algorithm depends more on the model closest to the true one than the others. The ratio between two average model probabilities will increase or decrease exponentially fast as the observation size increases, and the changing rate is determined by the difference between two K-L information of each model relative to the true model.

We also draw a connection between multiple model algorithms and the comparison of multiple estimation algorithms from the viewpoint of multiple hypotheses. By using the model probability update structure, we can also have a weighted sum of multiple algorithms that assigns more weight to the result from a better hypothesis. This provides a feasible way to combine multiple algorithms and produce more robust estimators.

Part II

Application of Best Linear Unbiased Filtering Method to Target Tracking

Chapter 6

Best Linear Unbiased Filtering Method[4]

6.1 Introduction

The Kalman filter has been popularly used for dynamic state filtering for several decades. Beyond the Kalman filter, a more fundamental basis of the recursive best linear unbiased filtering has been thoroughly investigated in a series of papers by Dr. X. Rong Li [69, 45, 29, 46, 32, 30, 31, 24, 25]. Many aspects of the best linear unbiased estimation have been examined thereby to reveal different application concerns. They cover unified optimal linear estimation fusion for central and distributed sensor fusion [45, 29, 46], optimality and efficiency of distributed fusion [32], relationships among weighted least square methods and BLUE method with prior, partial prior and no prior [30], and recursibility of linear estimation [25]. What is called quasi-recursive best unbiased linear filtering technique in [30] is of particular interest. Based on such formulation, the Kalman filter's **Linear-Gaussian** assumptions can be relaxed such that a general linear filtering technique for nonlinear systems could be obtained. Hence, we abandon the Kalman filter and go to a more fundamental basis of optimal recursive linear estimation.

[4] © [2004] IEEE. Reprinted, with permission, from Z. L. Zhao, X. R. Li, V. P. Jilkov, "Best Linear Unbiased Filtering with Nonlinear Measurements for Target Tracking". *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, pp. 1324-1336, Oct. 2004

The rest of this chapter presents the recursive best unbiased linear estimation filter.

6.2 Recursive BLUE Filter

Given the observation $\mathbf{z} = h(\mathbf{x}, v)$, where v is the observation noise and $h(\mathbf{x}, v)$ can be a linear or nonlinear function of state $\mathbf{x}$ and v, the well-known BLUE or linear minimum mean square error (LMMSE) estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ given $\mathbf{z}$ (see [3, pp. 123-128]) is

$$\hat{\mathbf{x}} = E^*[\mathbf{x}|\mathbf{z}] = \bar{\mathbf{x}}_0 + C_{\mathbf{xz}} C_{\mathbf{z}}^{-1} (\mathbf{z} - \bar{\mathbf{z}}_0) \tag{6.1}$$

$$P = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{'}] = P_0 - C_{\mathbf{xz}} C_{\mathbf{z}}^{-1} C_{\mathbf{zx}} \tag{6.2}$$

where $\bar{\mathbf{x}}_0$ and P_0 are the prior mean and covariance of $\mathbf{x}$; $\bar{\mathbf{z}}_0$ and $C_{\mathbf{z}}$ are the mean and covariance of $\mathbf{z}$. $C_{\mathbf{xz}}$ is the covariance of $\mathbf{x}$ and $\mathbf{z}$. The BLUE or LMMSE estimator of $\mathbf{x}$ given observation $\mathbf{z}$ is denoted by $E^*[\mathbf{x}|\mathbf{z}]$ because it has many properties of the conditional mean $E[\mathbf{x}|\mathbf{z}]$, although it is not really conditional mean. For example, it is a linear operator: $E^*[A\mathbf{x}+B\mathbf{y}+c|\mathbf{z}] = AE^*[\mathbf{x}|\mathbf{z}] + BE^*[\mathbf{y}|\mathbf{z}] + c$. Also, when $\mathbf{x}$ and $\mathbf{z}$ are jointly gaussian, $E^*[\mathbf{x}|\mathbf{z}] = E[\mathbf{x}|\mathbf{z}]$. Obviously, the estimator is in a linear function of the observation, and it does not matter whether the observation is linear or nonlinear.

Let $\mathbf{z}_k$ and $\mathbf{x}_k$ be the measurement and state at time k, and let $\mathbf{z}^{k-1}$ stand for the set of all past measurements before time k. The best linear unbiased estimator of $\mathbf{x}_k$ given $\mathbf{z}^k$ is $E^*\left[\mathbf{x}_k|\mathbf{z}^k\right]$. The recursive BLUE requires that the current state can only be estimated based on the previous estimator and the current measurement. It is known that for all $\mathbf{z}_k$ and $\mathbf{x}_k$ without any special assumptions the BLUE $E^*\left[\mathbf{x}_k|\mathbf{z}^k\right]$ always has the quasi-recursive form (see. e.g. [30, 25]).

$$\hat{\mathbf{x}}_k = E^*\left[\mathbf{x}_k|\mathbf{z}^k\right] = \bar{\mathbf{x}}_k + K_k \tilde{\mathbf{z}}_k \tag{6.3}$$

$$P_k = \bar{P}_k - K_k S_k K_k^{'} \tag{6.4}$$

where

$$\bar{\mathbf{x}}_k = E^* \left[\mathbf{x}_k | \mathbf{z}^{k-1}\right]$$

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \bar{\mathbf{x}}_k$$

$$\bar{\mathbf{z}}_k = E^* \left[\mathbf{z}_k | \mathbf{z}^{k-1}\right]$$

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \bar{\mathbf{z}}_k$$

$$\bar{P}_k = \mathrm{cov}(\tilde{\mathbf{x}}_k)$$

$$S_k = \mathrm{cov}(\tilde{\mathbf{z}}_k)$$

$$K_k = \mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k) S_k^{-1}$$

$$P_k = \mathrm{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_k).$$

This form is quasi-recursive, not necessarily truly recursive, because given the state estimator $\hat{\mathbf{x}}_{k-1}$ of $\mathbf{x}_{k-1}$ and its error covariance P_{k-1} at time $k-1$, the calculations of the predicted state $\bar{\mathbf{x}}_k$ and its error covariance $\bar{P}_k$, the predicted measurement $\bar{\mathbf{z}}_k$ and its error covariance S_k, and cross-covariance $\mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$ may depend on $\mathbf{z}^{k-1}$ and some second-order moment information prior to time $k-1$ not only through $\hat{\mathbf{x}}_{k-1}$ and P_{k-1}; however, if $\bar{\mathbf{x}}_k$, $\bar{\mathbf{z}}_k$, S_k, $\bar{P}_k$ and $\mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$ are available precisely through $\hat{\mathbf{x}}_{k-1}$ and P_{k-1} only, the above form will be truly recursive and optimal in the sense of BLUE or LMMSE. Even if this BLUE filter can not be truly recursive, it provides a solid framework to do approximated filtering.

For a linear system with the standard Kalman filter assumptions, the BLUE filter exactly becomes the Kalman filter.

From the above analysis, as long as these required terms $\bar{\mathbf{x}}_k$, $\bar{\mathbf{z}}_k$, $\bar{P}_k$, S_k and $\mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$ are available based on $\hat{\mathbf{x}}_{k-1}$and P_{k-1} as well as $\mathbf{z}_k$, the recursive BLUE is perfectly valid for nonlinear observation and thus can be applied to filtering with nonlinear measurements. Therefore, we abandon the Kalman filter framework by adopting the BLUE filter directly. All the fundamental flaws that were pointed our in the Introduction are circumvented.

Consider the stochastic system with linear dynamics and polar or spherical measurements

$$\mathbf{x}_k = F_{k-1}\mathbf{x}_{k-1} + \Gamma_{k-1}w_{k-1} \tag{6.5}$$

$$\mathbf{z}_k = h(\mathbf{x}_k, v_k) \tag{6.6}$$

where $\{w_k\}$ and $\{v_k\}$ are, in the sequel, uncorrelated white process noise and measurement noise sequences, both independent of the initial state $\mathbf{x}_0$, F_{k-1} is the transition matrix, and $h(\mathbf{x}_k, v_k)$ is the measurement function. Clearly, the BLUE or LMMSE state prediction can be done exactly the same as in the Kalman filter:

$$\bar{\mathbf{x}}_k = E^*\left[\mathbf{x}_k|\mathbf{z}^{k-1}\right] = F_{k-1}\hat{\mathbf{x}}_{k-1} + \Gamma_{k-1}\bar{w}_{k-1} \tag{6.7}$$

$$\bar{P}_k = \mathrm{cov}(\tilde{\mathbf{x}}_k) = F_{k-1}P_{k-1}F'_{k-1} + \Gamma_{k-1}Q_{k-1}\Gamma'_{k-1}. \tag{6.8}$$

What remains is to calculate

$$\bar{\mathbf{z}}_k = E^*\left[\mathbf{z}_k|\mathbf{z}^{k-1}\right] \tag{6.9}$$

$$S_k = \mathrm{cov}(\tilde{\mathbf{z}}_k) \tag{6.10}$$

$$K_k = \mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)S_k^{-1}. \tag{6.11}$$

Once these terms are available, the state update follows as (6.3) and (6.4). BLUE filter provides a more fundamental basis to nonlinear filtering techniques than the Kalman filter. We apply the best linear unbiased filtering method to target tracking problem with measurements in polar and spherical coordinates next.

Chapter 7

Application to Target Tracking with Nonlinear Measurements[4]

7.1 Introduction

In tracking applications, target dynamics are usually modelled in Cartesian coordinates, while the measurements are directly available in the original sensor coordinates, most often in spherical coordinates. For tracking in Cartesian coordinates using polar or spherical measurements, the measurement conversion method is widely used [28]. The basic idea is to transform the nonlinear measurement model into a pseudo-linear form in the Cartesian coordinates, estimate the bias and covariance of the converted measurement noise, and then use the Kalman filter, which clearly has better accuracy and consistency than the EKF.

A number of improved techniques for measurement conversion have been proposed [2, 20, 47, 48, 49, 50, 55, 56]. They differ in the ways of obtaining the bias and covariance of the noise of the converted measurement. Measurement models used in target tracking, including measurement conversion methods, are surveyed in [28]. The pros and cons of various measurement-conversion techniques have been revealed there with succinct explanation. In particular, it was pointed out that these conversion methods have fundamental flaws, as elaborated in Section 7.2. The idea of developing a theoretically sounder filter for tracking with linear dynamics and nonlinear measurements was briefly discussed in [28].

In this chapter, such a filter is developed, which stems from the recognition that the Kalman filter is nothing else but a recursive BLUE for a linear system; furthermore, although the Kalman filter cannot optimally handle any nonlinear measurement directly, we may still be able to develop the recursive BLUE filter for nonlinear measurements. This chapter presents such a recursive BLUE filter for tracking with linear dynamics and nonlinear measurements.

The rest of this chapter is organized as follows. In Sec. 7.2, we summarize the previous measurement-conversion method and explain its flaws. In Secs. 7.3 and 7.4, tracking in polar coordinates and spherical coordinates is given. The approximate implementations are discussed in section 7.5. The simulation comparison of this new method with two state-of-the-art measurement conversion techniques is presented in Sec. 7.6. The conclusions are given in Sec. 7.7.

7.2 Measurement Conversion Approach

To be simple and clear, the main idea and mathematical derivation are given for the polar measurements, and then the results are extended to the spherical measurements.

The measured range r_m and bearing θ_m are defined with respect to the true range r and bearing θ

$$r_m = r + \tilde{r} \qquad \theta_m = \theta + \tilde{\theta}.$$

The noise $\tilde{r}$ and $\tilde{\theta}$ are assumed to be independent with zero mean and standard deviations σ_r and σ_θ, respectively. The measurement conversion method converts the polar measurements into the Cartesian coordinates by

$$x_m = r_m \cos\theta_m = (r + \tilde{r}) \cos(\theta + \tilde{\theta})$$

$$y_m = r_m \sin\theta_m = (r + \tilde{r}) \sin(\theta + \tilde{\theta})$$

and then transform the above nonlinear form into a pseudo-linear form with respect to (x, y),

$$x_c = x + \tilde{x}_c \qquad y_c = y + \tilde{y}_c \tag{7.1}$$

where (x_c, y_c) is the converted debiased measurement [2, 20] or converted unbiased measurement [49, 50], $(x, y) = (r\cos\theta, r\sin\theta)$ is the true position of the target and $(\tilde{x}_c, \tilde{y}_c)$ is the noise of the converted measurement . Observe that (7.1) just divides the converted measurement into the true position (x, y) and the measurement noise $(\tilde{x}_c, \tilde{y}_c)$, however, $(\tilde{x}_c, \tilde{y}_c)$ is actually dependent on (x, y), namely, components of the state.

Once the linear form is obtained, the measurement conversion method applies the Kalman filter to do tracking. All measurement conversion techniques differ from each other in the ways that the bias and covariance of $(\tilde{x}_c, \tilde{y}_c)$ are calculated. Specific comparisons can be found in [48]. Basically, one way is to compute the mean and covariance conditioned on the measurement as $E[[\tilde{x}_c,\ \tilde{y}_c]'|r_m, \theta_m]$ and $\mathrm{cov}[[\tilde{x}_c,\ \tilde{y}_c]'|r_m, \theta_m]$, referred to as fixed-measurement [48] or unbiased method [49, 50]; another way is to compute them firstly conditioned on the true state and then conditioned on measurement as $E[E[[\tilde{x}_c,\ \tilde{y}_c]'|r, \theta]|r_m, \theta_m]$ and $E[\mathrm{cov}([\tilde{x}_c,\ \tilde{y}_c]'|r, \theta)|r_m, \theta_m]$, referred to as fixed-truth [48] or debiased method [2, 20]. As argued in [28], however, we will call them measurement-conditioned (MC) approach and nested conditioning (NC) approach, respectively.

From the above analysis, all previously proposed measurement conversion techniques have the following fundamental flaws. First, $(\tilde{x}_c, \tilde{y}_c)$ is state dependent; second, its covariance is estimated conditioned on the current measurement or state/measurement; third, the measurement noise sequence $\{(\tilde{x}_c, \tilde{y}_c)_k\}$ is not white anymore. However, in the assumptions of Kalman filter, the measurement noise is independent of the state, its covariance is unconditional, and it is white. As pointed out in [28], these fundamental flaws were ignored or overlooked before. As a result, they are by no means optimal.

From another point of view, there are two steps inside the measurement conversion method. Firstly, the state or the measurement is assumed to be nonrandom, then the mean

and covariance of the converted measurement error can be calculated. In the second step, the state is viewed as random, then the kalman filter can be applied. As we know, the Kalman filter framework requires the state should be only random. This inconsistency leads to some undesirable consequences. For example, in the first step because either the measurement or the state is nonrandom, and then the uncertainty of the state is not taken into account, the value of the computed covariance of the converted measurement error is smaller than the true covariance, where the state is random. Consequently, the state estimations of the MC and NC approaches will perform optimistically. Furthermore, when the measurement noise is state-dependent, the measurement prediction covariance formula of the Kalman filter should be modified to take into account the cross-correlation between the state and the measurement noise. All these techniques fail to have such modifications.

In Sections 7.3 and 7.4, the specific form of $\bar{\mathbf{z}}_k$, S_k and $\text{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$ for the polar and spherical measurements will be derived, respectively.

7.3 Tracking in Polar Coordinates

For any time instant k, the measured range and bearing in Polar coordinates are defined with respect to the true range r and bearing θ as

$$r_m = r + \tilde{r}, \quad \theta_m = \theta + \tilde{\theta}.$$

The noises $\tilde{r}$ and $\tilde{\theta}$ are assumed to be independent of each other and of r, θ, and be Gaussian distributed with zero means and standard deviations σ_r and σ_θ, respectively. Converting the

polar measurements into the Cartesian coordinates yields

$$\begin{aligned} h_x(\mathbf{x}, v) =& x_m = (r + \tilde{r})\cos(\theta + \tilde{\theta}) \\ =& x\cos\tilde{\theta} - y\sin\tilde{\theta} + \frac{x}{\sqrt{x^2+y^2}}\tilde{r}\cos\tilde{\theta} - \frac{y}{\sqrt{x^2+y^2}}\tilde{r}\sin\tilde{\theta} \\ h_y(\mathbf{x}, v) =& y_m = (r + \tilde{r})\sin(\theta + \tilde{\theta}) \\ =& y\cos\tilde{\theta} + x\sin\tilde{\theta} + \frac{y}{\sqrt{x^2+y^2}}\tilde{r}\cos\tilde{\theta} + \frac{x}{\sqrt{x^2+y^2}}\tilde{r}\sin\tilde{\theta} \end{aligned} \tag{7.2}$$

Then the tracking system has the nonlinear measurements of formula (6.6):

$$\mathbf{z}_k = h(\mathbf{x}_k, v_k) \tag{7.3}$$

where $\mathbf{z}_k = [x_m,\ y_m]'_k$, $\mathbf{x}_k = [x,\ \dot{x},\ y,\ \dot{y}]'_k$ and $v_k = [\tilde{r},\ \tilde{\theta}]'_k$.

Before we derive $\bar{\mathbf{z}}_k = E^*[\mathbf{z}_k|\mathbf{z}^{k-1}]$, S_k and $\mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$, *for brevity, we drop the time index k and the conditioning on* $\mathbf{z}^{k-1}$, i.e. let $\bar{\mathbf{z}}$, S and $\mathrm{cov}(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$ stand for $\bar{\mathbf{z}}_k$, S_k and $\mathrm{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$, respectively. Now denote $\bar{\mathbf{x}}_k$ and $\bar{P}_k$ by

$$\begin{aligned} \bar{\mathbf{x}} =& [\bar{x}\ \bar{\dot{x}}\ \bar{y}\ \bar{\dot{y}}]' \\ \bar{P} =& \begin{bmatrix} \mathrm{cov}(\tilde{x}) & \mathrm{cov}(\tilde{x}, \tilde{\dot{x}}) & \mathrm{cov}(\tilde{x}, \tilde{y}) & \mathrm{cov}(\tilde{x}, \tilde{\dot{y}}) \\ \mathrm{cov}(\tilde{\dot{x}}, \tilde{x}) & \mathrm{cov}(\tilde{\dot{x}}) & \mathrm{cov}(\tilde{\dot{x}}, \tilde{y}) & \mathrm{cov}(\tilde{\dot{x}}, \tilde{\dot{y}}) \\ \mathrm{cov}(\tilde{y}, \tilde{x}) & \mathrm{cov}(\tilde{y}, \tilde{\dot{x}}) & \mathrm{cov}(\tilde{y}) & \mathrm{cov}(\tilde{y}, \tilde{\dot{y}}) \\ \mathrm{cov}(\tilde{\dot{y}}, \tilde{x}) & \mathrm{cov}(\tilde{\dot{y}}, \tilde{\dot{x}}) & \mathrm{cov}(\tilde{\dot{y}}, \tilde{y}) & \mathrm{cov}(\tilde{\dot{y}}) \end{bmatrix}. \end{aligned}$$

Let $\lambda_1 = E[\cos\tilde{\theta}] = e^{-\sigma_\theta^2/2}$, $\lambda_2 = E[\cos^2\tilde{\theta}] = \frac{1}{2}(1 + e^{-2\sigma_\theta^2})$, $\lambda_3 = E[\sin^2\tilde{\theta}] = \frac{1}{2}(1 - e^{-2\sigma_\theta^2})$ and $E[\sin\tilde{\theta}] = E[\sin\tilde{\theta}\cos\tilde{\theta}] = 0$. Since (r, θ) and $(\tilde{r}, \tilde{\theta})$ are independent, (x, y) and $(\tilde{r}, \tilde{\theta})$ are also independent. Because the noise sequence $[\tilde{r}, \tilde{\theta}]'_k$ is white, $E[[\tilde{r}, \tilde{\theta}]'_k|\mathbf{z}^{k-1}] = E[[\tilde{r}, \tilde{\theta}]'_k] = 0$.

Thus, by Lemma 1 in Appendix A, the predicted measurement is

$$\bar{\mathbf{z}} = E^* \begin{bmatrix} x_m \\ y_m \end{bmatrix} = \lambda_1 \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \lambda_1 \bar{\mathbf{x}} \tag{7.4}$$

where $\bar{\mathbf{x}} \triangleq [\bar{x},\ \bar{y}]' \triangleq E^*\left[\mathbf{x}_k|\mathbf{z}^{k-1}\right]$.

By Lemma 2 in Appendix A, the cross-covariance between state prediction error $\tilde{\mathbf{x}}$ and

measurement prediction error $\tilde{\mathbf{z}}$ is

$$\begin{aligned}\mathrm{cov}(\tilde{\mathbf{x}},\tilde{\mathbf{z}}) &= E\left[\mathbf{x}\mathbf{z}^{'}\right] - E[\bar{\mathbf{x}}\bar{\mathbf{z}}^{'}] \\ &= \lambda_1\left(E\begin{bmatrix} x^2 & xy \\ x\dot{x} & \dot{x}y \\ xy & y^2 \\ x\dot{y} & \dot{y}y \end{bmatrix} - E\begin{bmatrix} \bar{x}^2 & \bar{x}\bar{y} \\ \bar{x}\bar{\dot{x}} & \dot{x}\bar{y} \\ \bar{x}\bar{y} & \bar{y}^2 \\ \bar{x}\bar{\dot{y}} & \bar{\dot{y}}\bar{y} \end{bmatrix}\right) \\ &= \lambda_1\left[\bar{P}(:,1),\ \bar{P}(:,3)\right] \end{aligned} \tag{7.5}$$

where, $\bar{P}(:,i)$ $(i=1,3)$ is the ith column vector of $\bar{P}$.

The predicted measurement error covariance is

$$S = \mathrm{cov}\left(\tilde{\mathbf{z}}\right) = \begin{bmatrix} \mathrm{cov}(\tilde{x}_m) & \mathrm{cov}(\tilde{x}_m,\tilde{y}_m) \\ \mathrm{cov}(\tilde{y}_m,\tilde{x}_m) & \mathrm{cov}(\tilde{y}_m) \end{bmatrix}. \tag{7.6}$$

Since

$$\begin{aligned} E[x_m^2] &= E[(x\cos\tilde{\theta} - y\sin\tilde{\theta} + \frac{x\tilde{r}\cos\tilde{\theta}}{\sqrt{x^2+y^2}} - \frac{y\tilde{r}\sin\tilde{\theta}}{\sqrt{x^2+y^2}})^2] \\ &= \lambda_2 E\left[x^2\right] + \lambda_3 E\left[y^2\right] + \frac{1}{2}\sigma_r^2 + \frac{1}{2}\sigma_r^2 e^{-2\sigma_\theta^2} E\left[\frac{x^2-y^2}{x^2+y^2}\right] \end{aligned}$$

we have

$$\begin{aligned} S_{11} &= \mathrm{cov}(\tilde{x}_m) = E[x_m^2] - E[\bar{x}_m^2] \\ &= \lambda_2\mathrm{cov}(\tilde{x}) + \lambda_3\mathrm{cov}(\tilde{y}) + \frac{1}{2}\sigma_r^2 + \lambda_3 E[\bar{y}^2] + (\lambda_2 - \lambda_1^2)E[\bar{x}^2] + \frac{1}{2}\sigma_r^2 e^{-2\sigma_\theta^2} E\left[\frac{x^2-y^2}{x^2+y^2}\right]. \end{aligned}$$

Similarly

$$\begin{aligned} S_{22} &= \mathrm{cov}(\tilde{y}_m) \\ &= \lambda_2\mathrm{cov}(\tilde{y}) + \lambda_3\mathrm{cov}(\tilde{x}) + \frac{1}{2}\sigma_r^2 + \lambda_3 E[\bar{x}^2] + (\lambda_2 - \lambda_1^2)E[\bar{y}^2] + \frac{1}{2}\sigma_r^2 e^{-2\sigma_\theta^2} E\left[\frac{y^2-x^2}{x^2+y^2}\right] \\ S_{12} &= \mathrm{cov}(\tilde{x}_m,\tilde{y}_m) \\ &= e^{-2\sigma_\theta^2}\mathrm{cov}\left(\tilde{x},\tilde{y}\right) + (e^{-2\sigma_\theta^2} - \lambda_1^2)E[\bar{x}\bar{y}] + \sigma_r^2 e^{-2\sigma_\theta^2} E\left[\frac{xy}{x^2+y^2}\right] \end{aligned}$$

where S_{ij} is the ith row and jth column element of S. All the $\mathrm{cov}(*)$ terms are directly available from $\bar{P}$. However, unlike $\bar{\mathbf{z}}$ and $\mathrm{cov}(\tilde{\mathbf{x}},\tilde{\mathbf{z}})$, S_{ij} can not be calculated precisely

because the $E[*]$ terms have no closed form. Fortunately, S can be approximated quite accurately, which will be discussed in detail in Section 5.

7.4 Tracking in Spherical Coordinates

In spherical coordinates, the measured range, bearing and elevation are defined with respect to the true range r, bearing θ and elevation ϕ as

$$r_m = r + \tilde{r}, \;\; \theta_m = \theta + \tilde{\theta}, \;\; \phi_m = \phi + \tilde{\phi}.$$

The noises $\tilde{r}$, $\tilde{\theta}$ and $\tilde{\phi}$ are assumed to be independent of each other and of r, θ, ϕ, and be Gaussian distributed with zero means and standard deviations σ_r, σ_θ and σ_ϕ, respectively. Converting the spherical measurements into the Cartesian coordinates yields

$$\begin{aligned}
h_x(\mathbf{x}, v) &= x_m = r_m \cos\theta_m \cos\phi_m \\
&= (r + \tilde{r}) \cos(\theta + \tilde{\theta}) \cos(\phi + \tilde{\phi}) \\
h_y(\mathbf{x}, v) &= y_m = r_m \sin\theta_m \cos\phi_m \\
&= (r + \tilde{r}) \sin(\theta + \tilde{\theta}) \cos(\phi + \tilde{\phi}) \\
h_z(\mathbf{x}, v) &= z_m = r_m \sin\phi_m = (r + \tilde{r}) \sin(\phi + \tilde{\phi}).
\end{aligned}$$

Then the tracking system has the nonlinear measurements of formula (6.6):

$$\mathbf{z}_k = h(\mathbf{x}_k, v_k) \tag{7.7}$$

where $\mathbf{z}_k = [x_m,\ y_m,\ z_m]'_k$, $\mathbf{x}_k = [x,\ \dot{x},\ y,\ \dot{y},\ z,\ \dot{z}]'_k$ and $v_k = [\tilde{r},\ \tilde{\theta},\ \tilde{\phi}]'_k$.

Similar to the polar coordinate system case, the predicted state and its error covariance are denoted by $\bar{\mathbf{x}}$ and $\bar{P}$.

$$\bar{\mathbf{x}}=[\bar{x}\ \bar{\dot{x}}\ \bar{y}\ \bar{\dot{y}}\ \bar{z}\ \bar{\dot{z}}]'$$

$$\bar{P}=\begin{bmatrix} \mathrm{cov}(\tilde{x}) & \mathrm{cov}(\tilde{x},\tilde{\dot{x}}) & \mathrm{cov}(\tilde{x},\tilde{y}) & \mathrm{cov}(\tilde{x},\tilde{\dot{y}}) & \mathrm{cov}(\tilde{x},\tilde{z}) & \mathrm{cov}(\tilde{x},\tilde{\dot{z}}) \\ \mathrm{cov}(\tilde{\dot{x}},\tilde{x}) & \mathrm{cov}(\tilde{\dot{x}}) & \mathrm{cov}(\tilde{\dot{x}},\tilde{y}) & \mathrm{cov}(\tilde{\dot{x}},\tilde{\dot{y}}) & \mathrm{cov}(\tilde{\dot{x}},\tilde{z}) & \mathrm{cov}(\tilde{\dot{x}},\tilde{\dot{x}}) \\ \mathrm{cov}(\tilde{y},\tilde{x}) & \mathrm{cov}(\tilde{y},\tilde{\dot{x}}) & \mathrm{cov}(\tilde{y}) & \mathrm{cov}(\tilde{y},\tilde{\dot{y}}) & \mathrm{cov}(\tilde{y},\tilde{z}) & \mathrm{cov}(\tilde{y},\tilde{\dot{z}}) \\ \mathrm{cov}(\tilde{\dot{y}},\tilde{x}) & \mathrm{cov}(\tilde{\dot{y}},\tilde{\dot{x}}) & \mathrm{cov}(\tilde{\dot{y}},\tilde{y}) & \mathrm{cov}(\tilde{\dot{y}}) & \mathrm{cov}(\tilde{\dot{y}},\tilde{z}) & \mathrm{cov}(\tilde{\dot{y}},\tilde{\dot{z}}) \\ \mathrm{cov}(\tilde{z},\tilde{x}) & \mathrm{cov}(\tilde{z},\tilde{\dot{x}}) & \mathrm{cov}(\tilde{z},\tilde{y}) & \mathrm{cov}(\tilde{z},\tilde{\dot{y}}) & \mathrm{cov}(\tilde{z}) & \mathrm{cov}(\tilde{z},\tilde{\dot{z}}) \\ \mathrm{cov}(\tilde{\dot{z}},\tilde{x}) & \mathrm{cov}(\tilde{\dot{z}},\tilde{\dot{x}}) & \mathrm{cov}(\tilde{\dot{z}},\tilde{y}) & \mathrm{cov}(\tilde{\dot{z}},\tilde{\dot{y}}) & \mathrm{cov}(\tilde{\dot{z}},\tilde{z}) & \mathrm{cov}(\tilde{\dot{z}}) \end{bmatrix}$$

Let λ_i $(i = 1, 2, 3)$ be the same as before, and $\mu_1 = E[\cos\tilde{\phi}] = e^{-\sigma_\phi^2/2}$, $E[\sin\tilde{\phi}] = 0$, $\mu_2 = E[\cos^2\tilde{\phi}] = \frac{1}{2}(1 + e^{-2\sigma_\phi^2})$, $\mu_3 = E[\sin^2\tilde{\phi}] = \frac{1}{2}(1 - e^{-2\sigma_\phi^2})$, $E[\sin\tilde{\phi}\cos\tilde{\phi}] = 0$. Again, $E[[\tilde{r}, \tilde{\theta}, \tilde{\phi}]'_k|\mathbf{z}^{k-1}] = E[[\tilde{r}, \tilde{\theta}, \tilde{\phi}]'_k] = 0$ because the noise sequence $[\tilde{r}, \tilde{\theta}, \tilde{\phi}]'_k$ is white.

Thus the predicted measurement is

$$\bar{\mathbf{z}} = E^* \begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix} = \begin{bmatrix} \lambda_1\mu_1\bar{x} \\ \lambda_1\mu_1\bar{y} \\ \mu_1\bar{z} \end{bmatrix} \tag{7.8}$$

where $\bar{\mathbf{x}} \triangleq [\bar{x},\ \bar{y},\ \bar{z}]' \triangleq E^*\left[\mathbf{x}_k|\mathbf{z}^{k-1}\right]$.

The cross-covariance between the state prediction error $\tilde{\mathbf{x}}$ and measurement prediction error $\tilde{\mathbf{z}}$ is

$$\begin{aligned} \mathrm{cov}(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) &= E\left[\mathbf{x}\mathbf{z}'\right] - E[\bar{\mathbf{x}}\bar{\mathbf{z}}'] \\ &= \begin{bmatrix} \mu_1\lambda_1\mathrm{cov}(\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{x},\tilde{y}) & \mu_1\mathrm{cov}(\tilde{x},\tilde{z}) \\ \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{x}},\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{x}},\tilde{y}) & \mu_1\mathrm{cov}(\tilde{\dot{x}},\tilde{z}) \\ \mu_1\lambda_1\mathrm{cov}(\tilde{y},\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{y}) & \mu_1\mathrm{cov}(\tilde{y},\tilde{z}) \\ \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{y}},\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{y}},\tilde{y}) & \mu_1\mathrm{cov}(\tilde{\dot{y}},\tilde{z}) \\ \mu_1\lambda_1\mathrm{cov}(\tilde{z},\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{z},\tilde{y}) & \mu_1\mathrm{cov}(\tilde{z}) \\ \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{z}},\tilde{x}) & \mu_1\lambda_1\mathrm{cov}(\tilde{\dot{z}},\tilde{y}) & \mu_1\mathrm{cov}(\tilde{\dot{z}},\tilde{z}) \end{bmatrix} \\ &= \mu_1\left[\lambda_1\bar{P}(:,1),\ \lambda_1\bar{P}(:,3),\ \bar{P}(:,5)\right] \end{aligned} \tag{7.9}$$

where $\bar{P}(:, i)$ $(i = 1, 3, 5)$ is the ith column vector of $\bar{P}$.

The predicted-measurement error covariance is

$$\begin{aligned} S &= \mathrm{cov}\left(\tilde{\mathbf{z}}\right) = E[\mathbf{z}\mathbf{z}'] - E[\bar{\mathbf{z}}\bar{\mathbf{z}}'] \\ &= \begin{bmatrix} \mathrm{cov}(\tilde{x}_m) & \mathrm{cov}(\tilde{x}_m,\tilde{y}_m) & \mathrm{cov}(\tilde{x}_m,\tilde{z}_m) \\ \mathrm{cov}(\tilde{y}_m,\tilde{x}_m) & \mathrm{cov}(\tilde{y}_m) & \mathrm{cov}(\tilde{y}_m,\tilde{z}_m) \\ \mathrm{cov}(\tilde{z}_m,\tilde{x}_m) & \mathrm{cov}(\tilde{z}_m,\tilde{y}_m) & \mathrm{cov}(\tilde{z}_m) \end{bmatrix}. \end{aligned} \tag{7.10}$$

Let $r = \sqrt{x^2+y^2+z^2}$ and $r_1 = \sqrt{x^2+y^2}$. Since

$$\begin{aligned}
E[x_m^2] =& E[(x\cos\tilde{\theta}\cos\tilde{\phi} - y\sin\tilde{\theta}\cos\tilde{\phi} + \frac{x}{r}\tilde{r}\cos\tilde{\theta}\cos\tilde{\phi} \\
& - \frac{y}{r}\tilde{r}\sin\tilde{\theta}\cos\tilde{\phi} - \frac{xz}{r_1}\cos\tilde{\theta}\sin\tilde{\phi} + \frac{yz}{r_1}\sin\tilde{\theta}\sin\tilde{\phi} \\
& - \frac{x}{r_1}\frac{z}{r}\tilde{r}\cos\tilde{\theta}\sin\tilde{\phi} + \frac{y}{r_1}\frac{z}{r}\tilde{r}\sin\tilde{\theta}\sin\tilde{\phi})^2] \\
=& \lambda_2\mu_2 E\left[\tilde{x}^2\right] + \lambda_3\mu_2 E\left[\tilde{y}^2\right] + \lambda_2\mu_2\sigma_r^2 E\left[\frac{x^2}{r^2}\right] \\
& + \lambda_3\mu_2\sigma_r^2 E\left[\frac{y^2}{r^2}\right] + \lambda_2\mu_3\sigma_r^2 E\left[\frac{x^2z^2}{r_1^2r^2}\right] \\
& + \lambda_2\mu_3 E\left[\frac{x^2z^2}{r_1^2}\right] + \lambda_3\mu_3 E\left[\frac{y^2z^2}{r_1^2}\right] \\
& + \lambda_3\mu_3\sigma_r^2 E\left[\frac{y^2z^2}{r_1^2r^2}\right],
\end{aligned}$$

we have

$$\begin{aligned}
S_{11} =& \mathrm{cov}(\tilde{x}_m) = E[x_m^2] - E[\bar{x}_m^2] \\
=& \lambda_2\mu_2\mathrm{cov}(\tilde{x}) + \lambda_3\mu_2\mathrm{cov}(\tilde{y}) + \lambda_2\mu_2\sigma_r^2 E\left[\frac{x^2}{r^2}\right] \\
& + \lambda_3\mu_2\sigma_r^2 E\left[\frac{y^2}{r^2}\right] + \lambda_2\mu_3 E\left[\frac{x^2z^2}{r_1^2}\right] + \lambda_3\mu_3 E\left[\frac{y^2z^2}{r_1^2}\right] \\
& + \lambda_2\mu_3\sigma_r^2 E\left[\frac{x^2z^2}{r_1^2r^2}\right] + \lambda_3\mu_3\sigma_r^2 E\left[\frac{y^2z^2}{r_1^2r^2}\right] \\
& + (\lambda_2\mu_2 - \lambda_1^2\mu_1^2)E[\bar{x}^2] + \lambda_3\mu_2 E[\bar{y}^2]
\end{aligned}$$

and similarly

$$
\begin{aligned}
S_{22} =& \operatorname{cov}(\tilde{y}_m^2) \\
=& \lambda_2\mu_2\operatorname{cov}(\tilde{y}) + \lambda_3\mu_2\operatorname{cov}(\tilde{x}) + \lambda_2\mu_2\sigma_r^2 E\left[\frac{y^2}{r^2}\right] + \lambda_3\mu_2\sigma_r^2 E\left[\frac{x^2}{r^2}\right] \\
&+ \lambda_3\mu_3 E\left[\frac{x^2z^2}{r_1^2}\right] + \lambda_2\mu_3 E\left[\frac{y^2z^2}{r_1^2}\right] + \lambda_3\mu_3\sigma_r^2 E\left[\frac{x^2z^2}{r_1^2r^2}\right] + \lambda_2\mu_3\sigma_r^2 E\left[\frac{y^2z^2}{r_1^2r^2}\right] \\
&+ (\lambda_2\mu_2 - \lambda_1^2\mu_1^2)E[\bar{y}^2] + \lambda_3\mu_2 E[\bar{x}^2] \\
S_{33} =& \operatorname{cov}(\tilde{z}_m) \\
=& \mu_2\operatorname{cov}(\tilde{z}) + \mu_3(\operatorname{cov}(\tilde{x}) + \operatorname{cov}(\tilde{y})) + \mu_2\sigma_r^2 E\left[\frac{z^2}{r^2}\right] \\
&+ \mu_3\sigma_r^2 E\left[\frac{r_1^2}{r^2}\right] + (\mu_2 - \mu_1^2)E[\bar{z}^2] + \mu_3 E[\bar{x}^2 + \bar{y}^2] \\
S_{12} =& \operatorname{cov}(\tilde{x}_m, \tilde{y}_m) \\
=& \mu_2(\lambda_2 - \lambda_3)\operatorname{cov}(\tilde{x}, \tilde{y}) + \sigma_r^2\mu_2(\lambda_2 - \lambda_3)E\left[\frac{xy}{r^2}\right] + (\lambda_2 - \lambda_3)\mu_3 E\left[\frac{xyz^2}{r_1^2}\right] \\
&+ \sigma_r^2(\lambda_2 - \lambda_3)\mu_3 E\left[\frac{xyz^2}{r_1^2r^2}\right] + (\mu_2(\lambda_2 - \lambda_3) - \lambda_1^2\mu_1^2)E[\bar{x}\bar{y}] \\
S_{13} =& \operatorname{cov}(\tilde{x}_m, \tilde{z}_m) \\
=& \lambda_1(\mu_2 - \mu_3)(\operatorname{cov}(\tilde{x}, \tilde{z}) + \sigma_r^2 E\left[\frac{xz}{r^2}\right]) + ((\mu_2 - \mu_3)\lambda_1 - \lambda_1\mu_1^2)E[\bar{x}\bar{z}] \\
S_{23} =& \operatorname{cov}(\tilde{y}_m, \tilde{z}_m) \\
=& \lambda_1(\mu_2 - \mu_3)(\operatorname{cov}(\tilde{y}, \tilde{z}) + \sigma_r^2 E\left[\frac{yz}{r^2}\right]) + ((\mu_2 - \mu_3)\lambda_1 - \lambda_1\mu_1^2)E[\bar{y}\bar{z}]
\end{aligned}
$$

where S_{ij} is the ith row and jth column element of S. All the $\operatorname{cov}(*)$ terms are available from $\bar{P}$ and the $E[*]$ terms have to be approximated. The details are discussed next.

7.5 Approximation of Measurement Residual Covariance S

In Sections 7.3 and 7.4, if the state estimator $\hat{\mathbf{x}}_{k-1}$ and its error covariance P_{k-1} are given, formulas for the predicted state $\bar{\mathbf{x}}_k$ and its error covariance $\bar{P}_k$, the predicted measurement $\bar{\mathbf{z}}_k$ and cross-covariance $\operatorname{cov}(\tilde{\mathbf{x}}_k, \tilde{\mathbf{z}}_k)$ are precise. Only the measurement residual covariance

S needs approximation.

First, the terms $E[\bar{x}^2]$, $E[\bar{y}^2]$, $E[\bar{z}^2]$, $E[\bar{x}\bar{y}]$, $E[\bar{x}\bar{z}]$ and $E[\bar{y}\bar{z}]$ in S in both the polar and spherical cases are replaced by the quantities inside their expectations, respectively. For example, $E[\bar{x}^2]$ and $E[\bar{y}^2]$ are replaced by $\bar{x}^2$ and $\bar{y}^2$, and then S_{11} in the polar case becomes

$$\begin{aligned} S_{11}=&\mathrm{cov}(\tilde{x}_m) = E[x_m^2] - E[\bar{x}_m^2] \\ \approx&\lambda_2 \mathrm{cov}(\tilde{x}) + \lambda_3 \mathrm{cov}(\tilde{y}) + \frac{1}{2}\sigma_r^2 + \lambda_3 \bar{y}^2 \\ &+(\lambda_2 - \lambda_1^2)\bar{x}^2 + \frac{1}{2}\sigma_r^2 e^{-2\sigma_\theta^2} E\left[\frac{x^2-y^2}{x^2+y^2}\right]. \end{aligned}$$

We justify this approximation as follows.

Actually, the difference between S_{11} and the above approximation is $(\lambda_2 - \lambda_1^2)(E[\bar{x}^2] - \bar{x}^2) + \lambda_3(E[\bar{y}^2] - \bar{y}^2)$. As we know, $(\bar{x}^2 - E[\bar{x}^2])$ and $(E[\bar{y}^2] - \bar{y}^2)$ are usually much smaller than $\bar{x}^2$ and $\bar{y}^2$ in a tracking problem. Furthermore, by the Taylor series expansion equations

$$\begin{aligned} \lambda_2 =& \frac{1}{2}(1 + e^{-2\sigma_\theta^2}) = 1 - \sigma_\theta^2 + O(\sigma_\theta^4) \\ \lambda_2 - \lambda_1^2 =& \frac{1}{2}(1 + e^{-2\sigma_\theta^2}) - e^{-\sigma_\theta^2} = \frac{\sigma_\theta^4}{2} - O(\sigma_\theta^6) \\ \lambda_3 =& \frac{1}{2}(1 - e^{-2\sigma_\theta^2}) = \sigma_\theta^2 - \sigma_\theta^4 + O(\sigma_\theta^6) \end{aligned}$$

we see that λ_2 is on the order of $(1 - \sigma_\theta^2)$; $(\lambda_2 - \lambda_1^2)$ and λ_3 are on the order of σ_θ^4 and σ_θ^2, respectively. In practice for most radar systems, the bearing noise standard deviation σ_θ is below 0.4 rad [2], so the contributions of $(E[\bar{x}^2] - \bar{x}^2)$ and $(E[\bar{y}^2] - \bar{y}^2)$ to S_{11} are greatly decreased by their weights, $(\lambda_2 - \lambda_1^2)$ and λ_3, compared with the term $\lambda_2 \mathrm{cov}(x)$. For example, even for large error $\sigma_\theta \leq 0.2$ rad, $\frac{(\lambda_2 - \lambda_1^2)}{\lambda_2} \leq 0.07\%$ and $\frac{\lambda_3}{\lambda_2} \leq 4\%$. Therefore, this approximation difference is usually negligible.

All substitutions of $E[\bar{x}^2]$, $E[\bar{y}^2]$, $E[\bar{z}^2]$, $E[\bar{x}\bar{y}]$, $E[\bar{x}\bar{z}]$ and $E[\bar{y}\bar{z}]$ by the quantities inside their expectations can be justified in a similar way for both polar and spherical cases.

Second, it seems that all the other $E[*]$ terms involved in S in both the polar and spherical cases have either no closed form or a complex expression. We approximate them

through Taylor series expansion. The Taylor series expansion of a function $f(x,y)$ around (x_0,y_0) is

$$\begin{aligned} f(x,y) \approx & f(x_0,y_0) + (x-x_0)\frac{\partial f(x_0,y_0)}{\partial x} + (y-y_0)\frac{\partial f(x_0,y_0)}{\partial y} + (x-x_0)^2\frac{\partial^2 f(x_0,y_0)}{2\partial x^2} \\ & + (x-x_0)(y-y_0)\frac{\partial^2 f(x_0,y_0)}{\partial x \partial y} + (y-y_0)^2\frac{\partial^2 f(x_0,y_0)}{2\partial y^2}. \end{aligned}$$

Taking expectation yields

$$\begin{aligned} E[f(x,y)] \approx & f(x_0,y_0) + \mathrm{cov}(\tilde{x})\frac{\partial^2 f(x_0,y_0)}{2\partial x^2} \\ & + \mathrm{cov}(\tilde{x},\tilde{y})\frac{\partial^2 f(x_0,y_0)}{\partial x \partial y} + \mathrm{cov}(\tilde{y})\frac{\partial^2 f(x_0,y_0)}{2\partial y^2}. \end{aligned}$$

Except for the terms $E\left[\frac{x^2z^2}{d_1^2}\right]$, $E\left[\frac{y^2z^2}{d_1^2}\right]$ and $E\left[\frac{xyz^2}{d_1^2}\right]$ in S of the spherical case, the other $E[*]$ terms are smaller than 1. For example, in S_{11} of the polar case,

$$\begin{aligned} E\left[\frac{y^2-x^2}{x^2+y^2}\right] \approx & \frac{\bar{y}^2-\bar{x}^2}{\bar{x}^2+\bar{y}^2} + \frac{2\bar{y}^2(\bar{y}^2-3\bar{x}^2)}{(\bar{x}^2+\bar{y}^2)^3}\mathrm{cov}(\tilde{x}) \\ & + \frac{4\bar{x}\bar{y}(\bar{x}^2-\bar{y}^2)}{(\bar{x}^2+\bar{y}^2)^3}\mathrm{cov}(\tilde{x},\tilde{y}) - \frac{2\bar{x}^2(\bar{x}^2-3\bar{y}^2)}{(\bar{x}^2+\bar{y}^2)^3}\mathrm{cov}(\tilde{y}). \end{aligned}$$

Since the denominator has higher order than the numerator in the differential terms and $(\bar{x}^2+\bar{y}^2)$ is usually much larger than $\mathrm{cov}(\tilde{x})$, $\mathrm{cov}(\tilde{y})$ and $\mathrm{cov}(\tilde{x},\tilde{y})$, the second order terms are therefore negligible. For simplicity, $E\left[\frac{y^2-x^2}{x^2+y^2}\right]$ is approximated only by $\frac{\bar{y}^2-\bar{x}^2}{\bar{x}^2+\bar{y}^2}$. A similar justification can be applied to all $E[*]$ terms that are smaller than 1. That is $E[f(\mathbf{x})]$ is simply replaced by $f(\bar{\mathbf{x}})$, the value of $f(\mathbf{x})$ at $\mathbf{x}=\bar{\mathbf{x}}$. This assertion has been verified term by term carefully. Notice that all the $\mathrm{cov}(*)$ terms above are directly available from the state prediction error covariance $\bar{P}$.

The terms $E\left[\frac{x^2z^2}{r_1^2}\right]$, $E\left[\frac{y^2z^2}{r_1^2}\right]$ and $E\left[\frac{xyz^2}{r_1^2}\right]$ in S of spherical case could be larger than 1. The expectation of the second-order Taylor series expansion of $E\left[\frac{x^2z^2}{r_1^2}\right]$ is

$$\begin{aligned} E\left[\frac{x^2z^2}{r_1^2}\right] \approx & \frac{\bar{x}^2\bar{z}^2}{\bar{r}_1^2} + \frac{\bar{x}^2}{\bar{r}_1^2}\mathrm{cov}(\tilde{z}) + \frac{2\bar{x}\bar{z}\bar{y}^2}{\bar{r}_1^4}\mathrm{cov}(\tilde{x},\tilde{z}) - \frac{2\bar{y}\bar{z}\bar{x}^2}{\bar{r}_1^4}\mathrm{cov}(\tilde{y},\tilde{z}) \\ & + 2\bar{x}\bar{y}\bar{z}^2\frac{\bar{x}^2-\bar{y}^2}{\bar{r}_1^6}\mathrm{cov}(\tilde{x},\tilde{y}) - \bar{x}^2\bar{z}^2\frac{\bar{x}^2-3\bar{y}^2}{\bar{r}_1^6}\mathrm{cov}(\tilde{y}) - \bar{y}^2\bar{z}^2\frac{3\bar{x}^2-\bar{y}^2}{\bar{r}_1^6}\mathrm{cov}(\tilde{x}) \end{aligned}$$

where $\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}$ and $\bar{r}_1 = \sqrt{\bar{x}^2 + \bar{y}^2}$. Because not only each second-order term is much smaller than $\frac{\bar{x}^2\bar{z}^2}{\bar{r}_1^2}$ in the above formula, but also the weight $\lambda_2\mu_3$ of $E\left[\frac{x^2z^2}{r_1^2}\right]$ in S_{11} is very small, we again approximate $E\left[\frac{x^2z^2}{r_1^2}\right]$ simply by $\frac{\bar{x}^2\bar{z}^2}{\bar{r}_1^2}$. Similarly, $E\left[\frac{y^2z^2}{r_1^2}\right]$ and $E\left[\frac{xyz^2}{r_1^2}\right]$ are approximated by $\frac{\bar{y}^2\bar{z}^2}{\bar{r}_1^2}$ and $\frac{\bar{x}\bar{y}\bar{z}^2}{\bar{r}_1^2}$, respectively.

Based on the above analysis, it can be expected that the approximations of the components of S are quite accurate for most target tracking applications. This will be verified by the simulation results in the next section. We can see the true estimation errors are almost always perfectly consistent with the computed covariance by the filter in all simulations.

7.6 Simulation and Comparison

Before we do the simulations, for the convenience of implementation, we summarize the BLUE filter for the spherical coordinates in Table 7.1.

Considering that it has been shown in [2, 20] that the measurement conversion method outperforms the EKF clearly at long range for root-mean-square azimuth of 1.5° (26 milliradians) or more in terms of the estimation accuracy and filter credibility [34], in the following we only compare two state-of-the-art conversion techniques with our proposed method. We choose the measurement-conditioned (MC) approach and one version of the nested conditioning (NC) approach [28], which are among the top choices in the class of measurement conversion methods. They were referred to as "fixed-measurement" and "additive fixed truth II" approaches in [48] and unbiased and debiased methods in [20, 49, 50], respectively.

Here the formulas of "fixed-measurement" and "additive fixed truth II" in [48] are adopted for MC and NC, respectively. The comparison results of the same scenario as that of [48] are presented as follows.

Consider a scenario of a three-dimensional Cartesian x-y-z space with a single sensor located at the origin. The target sampling period is 1 sec. The coordinates (x, y, z) of the target object at time zero are determined by random draws from three independent, Gaussian

Table 7.1: One Iteration of the BLUE Filter for a system with linear dynamics and converted spherical measurements specified by (6.5), (6.6) and (7.7)

1. Prediction:

$$\bar{\mathbf{x}}_k = [\bar{x},\ \bar{\dot{x}},\ \bar{y},\ \bar{\dot{y}},\ \bar{z},\ \bar{\dot{z}}]' = F_{k-1}\hat{\mathbf{x}}_{k-1} + \Gamma_{k-1}\bar{w}_{k-1}$$

$$\bar{P} = F_{k-1}P_{k-1}F'_{k-1} + \Gamma_{k-1}Q_{k-1}\Gamma'_{k-1}$$

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}, \quad \bar{r}_1 = \sqrt{\bar{x}^2 + \bar{y}^2}$$

$$\alpha = \left(\frac{\mu_2\sigma_r^2}{\bar{r}^2} + \frac{\mu_3\bar{z}^2}{\bar{r}_1^2} + \frac{\mu_3\sigma_r^2\bar{z}^2}{\bar{r}_1^2\bar{r}^2}\right)$$

$$S(1,1) \approx \lambda_2\mu_2\bar{P}(1,1) + \lambda_3\mu_2\bar{P}(3,3) + \alpha(\lambda_2\bar{x}^2 + \lambda_3\bar{y}^2)$$

$$S(2,2) \approx \lambda_2\mu_2\bar{P}(3,3) + \lambda_3\mu_2\bar{P}(1,1) + \alpha(\lambda_3\bar{x}^2 + \lambda_2\bar{y}^2)$$

$$S(3,3) \approx \mu_2\bar{P}(5,5) + \mu_3(\bar{P}(1,1) + \bar{P}(3,3)) + \mu_2\sigma_r^2\frac{\bar{z}^2}{\bar{r}^2} + \mu_3\sigma_r^2\frac{\bar{r}_1^2}{\bar{r}^2}$$

$$S(1,2) = S(2,1) \approx (\lambda_2 - \lambda_3)(\mu_2\bar{P}(1,3) + \alpha\bar{x}\bar{y}$$

$$S(1,3) = S(3,1) \approx \lambda_1(\mu_2 - \mu_3)(\bar{P}(1,5) + \sigma_r^2\frac{\bar{x}\bar{z}}{\bar{r}^2})$$

$$S(2,3) = S(3,2) \approx \lambda_1(\mu_2 - \mu_3)(\bar{P}(3,5) + \sigma_r^2\frac{\bar{y}\bar{z}}{\bar{r}^2})$$

$$K_k = \mu_1\left[\lambda_1\bar{P}(:,1),\ \lambda_1\bar{P}(:,3),\ \bar{P}(:,5)\right]S^{-1}$$

2. Update:

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + K_k(\mathbf{z}_k - \mu_1\left[\lambda_1\bar{x},\ \lambda_1\bar{y},\ \bar{z}\right]')$$

$$P_k = \bar{P}_k - K_kSK_k^{'}$$

Constants:

$$\lambda_1 = e^{-\sigma_\theta^2/2},\ \lambda_2 = \tfrac{1}{2}(1 + e^{-2\sigma_\theta^2}),\ \lambda_3 = \tfrac{1}{2}(1 - e^{-2\sigma_\theta^2})$$
$$\mu_1 = e^{-\sigma_\phi^2/2},\ \mu_2 = \tfrac{1}{2}(1 + e^{-2\sigma_\phi^2}),\ \mu_3 = \tfrac{1}{2}(1 - e^{-2\sigma_\phi^2})$$

distributions with means -50 km, 200 km and 0 km, respectively, and a common standard deviation of 5 km. The target moves at a nearly constant high velocity, whose components $\dot{x}$, $\dot{y}$ and $\dot{z}$, obtained by a random draw from a Gaussian distribution with means 1000m/s, 0m/s and 0m/s, respectively, and a common standard deviation of 0.1 km/s. The sensor's independent measurement errors have standard deviations $\sigma_r = 4$ m and $\sigma_\theta = \sigma_\phi = 10$ millirad.

Following [48], all the filters are initialized with an effectively infinite initial state error

covariance, and a highly inaccurate initial state estimate. The tracking period begins at 100 sec after time zero and continues for 100 secs.

We compare only the average normalized estimation error square (ANEES) and position root-mean-square error (RMSE) of these three filters by increasing the process noise first, and then increasing the measurement noise. The velocity RMSEs will not be presented, because the velocity RMSEs of the three filters have little difference in all cases. Moreover, due to the initial condition used here, it is not easy to tell which one is better in the transient part. So we only present the results which start from 150 sec.

The ANEES is defined by [34]

$$\text{ANEES} = \frac{1}{Nn}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\mathbf{x}}_i)' P_i^{-1}(\mathbf{x}_i - \hat{\mathbf{x}}_i) \tag{7.11}$$

where $(\mathbf{x}_i - \hat{\mathbf{x}}_i)$ and P_i are the state estimation error and error covariance in ith run , n is the state dimension, and N is the total number of runs. If the estimation error and the estimated covariance match each other, the ANEES will be close to 1, we say the filter is credible.

The performance of the three filters in different cases can be summarized as follows.

Case 1. *No process noise, measurement errors* $\sigma_r = 4m$ *and* $\sigma_\theta = \sigma_\phi = 10$ *millirad, 500 runs*

Fig. 7.1 compares the ANEES' of the MC and NC techniques with that of the BLUE filter. The BLUE filter is much more credible than the other two. The ANEES of the BLUE filter is very close to the ideal unity, which indicates the estimator has almost perfect credibility. Fig. 7.2 compares the position RMSEs of the three techniques. The BLUE filter is more accurate than the other two after the initial transient. Notice that the MC method is more credible than the NC method, but the NC method is more accurate than the MC method in this situation.

Case 2. *Process noise* $Q = 10^2 I$*, measurement errors* $\sigma_r = 4m$ *and* $\sigma_\theta = \sigma_\phi = 10$ *millirad, 500 runs*

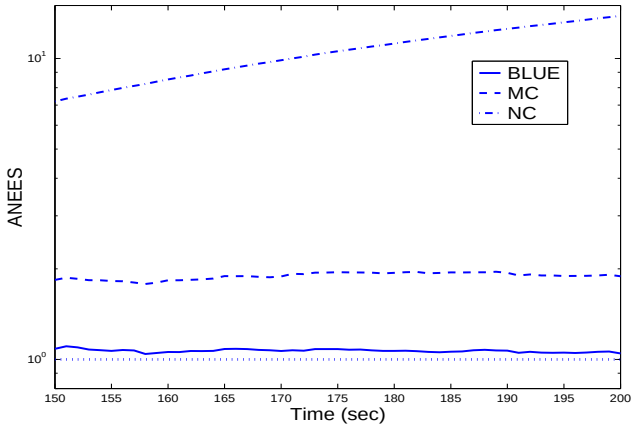


Fig. 7.1: ANEES (Case 1)

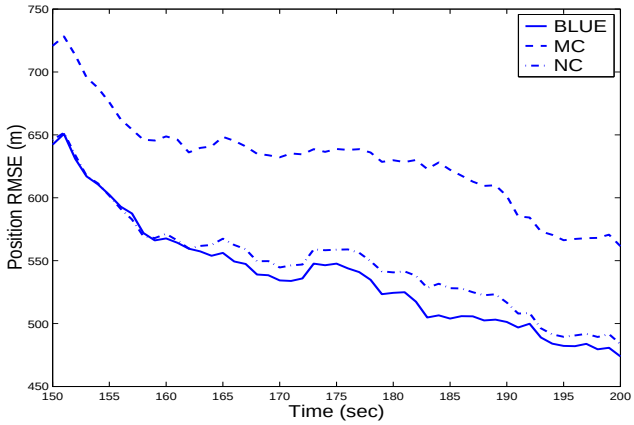


Fig. 7.2: RMSE of position after transient (Case 1)

In this case, only the process noise is increased. As is clear from Figs. 7.3 and 7.3, the ANEES of the BLUE filter is very close to ideal unity and the BLUE filter is also more accurate than the other two. Moreover, with the increase of the process noise, the credibilities of the MC and NC methods become better. The increase of the position RMSE can be explained as follows. The motion of the target is flying away from the origin.With the constant bearing and elevation noise, the measurement noise becomes larger and larger. When the filter enters the steady state, the position RMSE will become larger and larger also.

Case 3. *No process noise, Measurement errors* $\sigma_r = 40m$ *and* $\sigma_\theta = \sigma_\phi = 100$ *millirad,*

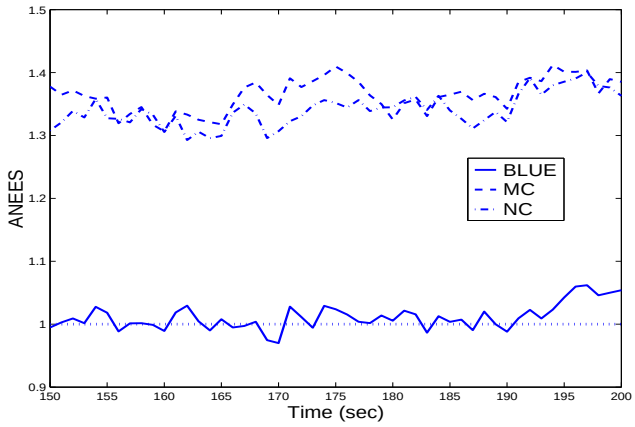


Fig. 7.3: ANEES (Case 2)

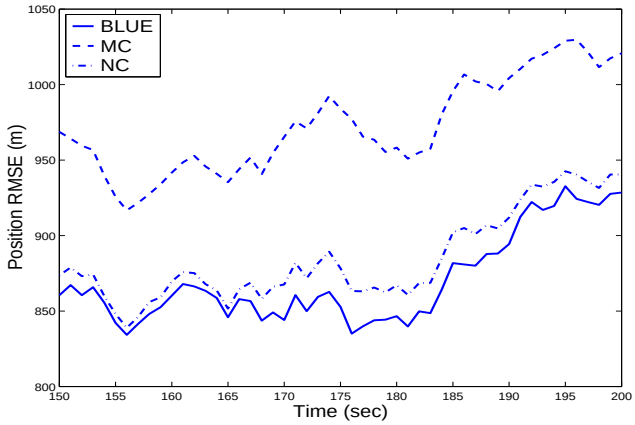


Fig. 7.4: RMSE of position after transient (Case 2)

500 runs

In this case, only the measurement noise is increased. The simulation results are plotted in Figs. 7.5 and 7.5. The BLUE filter is much more accurate than the other two. Also, its ANEES is still very close to unity, even though the position RMSE is rather large.

Case 4. *process noise* $Q = 10^2 I$, *Measurement errors* $\sigma_r = 40m$ *and* $\sigma_\theta = \sigma_\phi = 100$ *millirad, 500 runs*

In this case, both the process noise and measurement noise are increased. The simulation results in Figs. 7.7 and 7.7 show that the BLUE filter is still very credible and much more accurate than the other two.

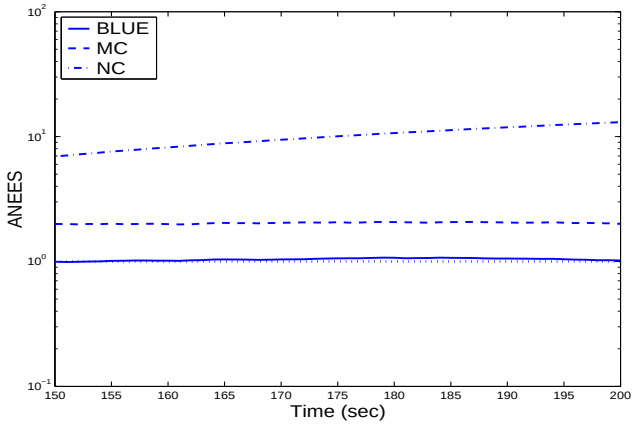


Fig. 7.5: ANEES (Case 3)

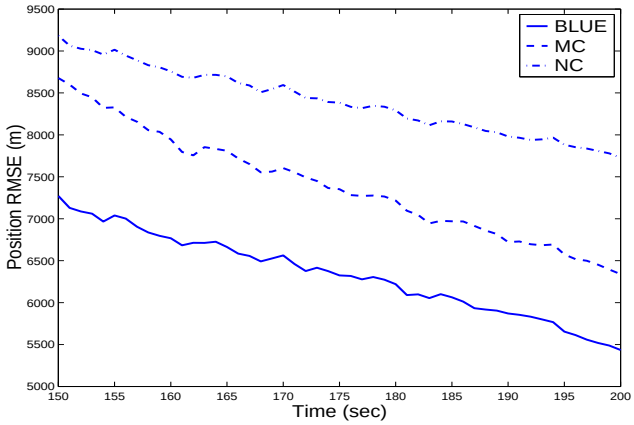


Fig. 7.6: RMSE of position after transient (Case 3)

In the MC and NC methods, because the computation of the covariance of the converted measurement error assumes that either the measurement or the state/measurement is nonrandom, and then the uncertainty of the state is not taken into account, the computed covariance is smaller than the true covariance, where the state is random. Consequently, the state estimations of the MC and NC methods perform optimistically, which explains why their ANEES' are above 1.

In the simulation conducted, the standard implementation of the MC and NC conversion techniques sometimes exhibited numerical problems; we had to use the numerically more robust Joseph form. On the contrary, we did not experience any numerical problem with

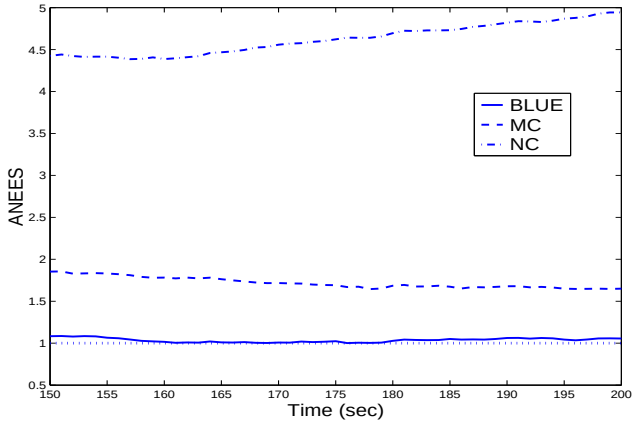


Fig. 7.7: ANEES (Case 2)

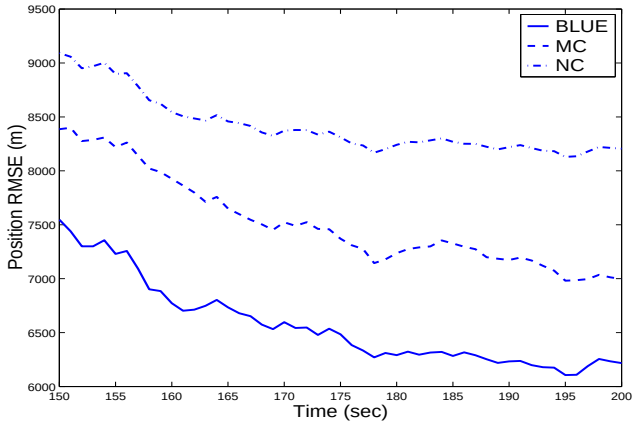


Fig. 7.8: RMSE of position after transient (Case 2)

the approximate implementation of our proposed BLUE filter.

We also compared the BLUE filter with the EKF as [2, 20] did. When the initial condition is poor or the process or measurement noise is intensive, the EKF is subject to divergence. Even though the EKF works well in some scenarios, its performance is still inferior to that of the measurement conversion method. The specific conclusion can be found in [2, 20].

Furthermore, we have run these three methods under the same conditions using Matlab 6.1v on a P4 1.7GHz CPU with 256M RAM and compared their computational costs for one iteration of each method. Table 7.2 contains those costs. Their average times of each run are quite close. So the good performance of BLUE filter is achieved without increasing the

computational complexity.

Table 7.2: Computational Costs for One Iteration of Each Method

Filters	BLUE	MC	NC
CPU time (ms)	104	102	101

7.7 Conclusions

An optimal recursive filter in the LMMSE sense has been presented for linear dynamics with nonlinear measurements. For the recursive BLUE is perfectly valid for nonlinear measurements, the proposed filter operates entirely in the Cartesian coordinates but is free of the fundamental flaws of the measurement-conversion method.

The simulation-based comparison of the proposed BLUE filter with two state-of-the-art measurement-conversion techniques demonstrates the following. In terms of estimation errors and filter credibility, our proposed BLUE filter significantly outperforms these two measurement-conversion techniques in all simulations; in particular, our proposed filter is almost always perfectly credible in the sense that the true estimation errors are consistent with the filter's self-assessment. Because of abandoning the Kalman filter framework and using optimal BLUE filter in the implementation, the good performance is achieved without increasing the computational complexity and algorithm complexity.

7.8 Appendix

Lemma 2: *For scalar-valued x, y and z, if y is independent of x and z, then*

$$E^*[xy|z] = E[y]E^*[x|z]$$

Proof: Since y is independent of x and z,

$$\begin{aligned} E[xy] &= E[x]E[y] \\ cov(xy,z) &= E[xyz] - E[xy]E[z] \\ &= E[y]E[xz] - E[y]E[x]E[z] \\ &= E[y](E[xz] - E[x]E[z]) \\ &= E[y]cov(x,z) \end{aligned}$$

we have

$$\begin{aligned} E^*[xy|z] &= E[xy] + cov(xy,z)cov^-1(z)(z - E[z]) \\ &= E[y](E[x] + cov(x,z)cov^-1(z)(z - E[z])) \\ &= E[y]E^*[x|z] \end{aligned}$$

□

Lemma 3: *For BLUE estimation error covariance,*

$$\begin{aligned} cov(\tilde{\mathbf{x}}) &= E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})'] = E[\mathbf{x}\mathbf{x}'] - E[\hat{\mathbf{x}}\hat{\mathbf{x}}'] \\ cov(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) &= E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{y} - \hat{\mathbf{y}})'] = E[\mathbf{x}\mathbf{y}'] - E[\hat{\mathbf{x}}\hat{\mathbf{y}}'] \end{aligned}$$

where $\hat{\mathbf{x}} = \mathbf{E}^*[\mathbf{x}|\mathbf{z}]$, $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$, $\hat{\mathbf{y}} = \mathbf{E}^*[\mathbf{y}|\mathbf{z}]$, $\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}}$.
Proof:

$$\begin{aligned} cov(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) &\triangleq E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{y} - \hat{\mathbf{y}})'] \\ &= E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{y}'] - E[(\mathbf{x} - \hat{\mathbf{x}})\hat{\mathbf{y}}'] \\ &= E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{y}'] \\ &= E[\mathbf{x}\mathbf{y}'] - E[\hat{\mathbf{x}}\mathbf{y}'] \\ &= E[\mathbf{x}\mathbf{y}'] - E[\hat{\mathbf{x}}(\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}})'] \\ &= E[\mathbf{x}\mathbf{y}'] - E[\hat{\mathbf{x}}\hat{\mathbf{y}}'] \end{aligned}$$

The second line and the last line come from orthogonality principle: $(\mathbf{x} - \hat{\mathbf{x}}) \perp \hat{\mathbf{y}}$ and $(\mathbf{y} - \hat{\mathbf{y}}) \perp \hat{\mathbf{x}}$. Similarly

$$cov(\tilde{\mathbf{x}}) = E[\mathbf{x}\mathbf{x}'] - E[\hat{\mathbf{x}}\hat{\mathbf{x}}']$$

□

Chapter 8

Summary and Future Work

This dissertation discusses the performance measures for estimation performance appraisal and the application of an estimation algorithm to target tracking with nonlinear measurement.

In the first part, this dissertation covers three types of measures which include local performance measures, global performance measures and model distortion measure. Local measures and global measures are designed to leverage the interpretation on and insights into various aspects of estimation performance, such as accuracy, concentration probability and uncertainty reduction. They address these aspects from local and global viewpoints, respectively. Many useful local performance measures has been published in [38, 34, 33, 35, 36, 37]. [64] presents the global performance measure. For parameter estimation, the model approximation is quite common and it inevitably affects estimation performance. Model distortion measure is devised to indicate the difference between the approximated model and the original one. [65] and [63] exemplify applications of the model distortion measure to MM algorithm.

The ongoing work along this line focuses on tests for and measures of estimator's credibility. Most estimators and filters provide assessments of their own estimation error, often in the form of mean-square error. Are these self-assessments trustable? What is the degree to which they are trustable? The credibility tests and credibility measures which have been

addressed in the submitted papers [39, 40, 41] are designed to answer these questions.

I am also aiming at the topics on model comparison and model selection. The essential idea is still to find efficient performance measures to size up the stochastic dependence relation between model and data. The published paper on this topic includes [62]. Future work will emphasize the exploration of these existing measures and potential novel measures.

Another goal of performance evaluation is on the following questions: how to quantify the accuracy of the accuracy measures, how to measure the credibility measures' credibility. The statistics seems lack of thorough theoretical conclusion on such questions. We will explore the insights into such questions and gain meaningful interpretations. Such questions could be rooted in the profound discoveries of Kurt Godel and Alan M. Turing in the 1930's. "Godel showed that no finite set of axioms and methods of reasoning could encompass all the mathematical properties of the positive integers. Turing later couched Godel's ingenious and complicated proof in a more accessible form. He showed that Godel's incompleteness theorem is equivalent to the assertion that there can be no general method for systematically deciding whether a computer program will ever halt, that is, whether it will ever cause the computer to stop running (see [11])." After all, the performance evaluation will finally reach the limit of the foundation of probability theory and statistical inference theory. The intuition tells us that the ultimate conclusions we can make will depend on the discoveries of equivalence of the incompleteness theorem to its counterpart in probability theory and statistical inference. The ultimate goal is towards such an understanding and insight.

For estimation algorithm application, we apply the best linear unbiased estimation technique to nonlinear measurements in target tracking. One main contribution of this work is to relax the Kalman filter's assumptions and go to a more fundamental basis. We emphasize that the best linear unbiased estimation is valid for nonlinear as well as linear measurements. The proposed method outperforms the measurement conversion methods significantly. Some detailed applications to polar and spherical coordinate measurements of the recursive best

linear unbiased filtering technique can be found in [67, 66, 68].

The ongoing and future work along this line is on the comparison among these most often used filtering techniques, such as EKF, unscented filter, our proposed BLUE filter and particle filter. Our interest is to cast these methods into a general theoretical framework such that we can elucidate the pros and cons of these methods theoretically. Related work is presented in [61, 60]. A natural extension of such work is to have a more general viewpoint on these methods from parametric, nonparametric and semiparametric setting. Novel techniques are desired.

I am also interested in the research topic of sensor fusion. An efficient multi-sensor fusion procedure could efficiently extract more useful information, which is hidden in distorted multi-sensor data. Information from multiple sensors can reduce overall uncertainty and thus increase the accuracy of the measurement, which finally improves the detection performance. Sensor fusion has great potential for applications in system identification, pattern recognition, fault detection, image processing, and target classification and tracking.

Bibliography

[1] Y. Bar-Shalom and X. R. Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Boston, MA, 1993. (Reprinted by YBS Publishing, 1998).

[2] Y. Bar-Shalom and X. R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, Storrs, CT, 1995.

[3] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*. Wiley, New York, 2001.

[4] Y. Baram. A Sufficient Condition for Consistent Discrimination Between Stationary Gaussian Models. *IEEE Trans. Automatic Control*, AC-23(5):958–960, Oct. 1978.

[5] Y. Baram and N. R. Sandell, Jr. An Information Theoretic Approach to Dynamical Systems Modeling and Identification. *IEEE Trans. Automatic Control*, AC-23(1):61–66, Feb. 1978.

[6] Y. Baram and N. R. Sandell Jr. Consistent Estimation on Finite Parameter Sets with Application to Linear Systems Identification. *IEEE Trans. Automatic Control*, AC-23(3):451–454, June 1978.

[7] R. H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.*, 37:51–58, 1966.

[8] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation*. Wiley, New York, 1990.

[9] P. S. Bullen. *Handbook of Means and Their Inequalities*. Kluwer Academic, Dordrescht, The Netherlands, 2003.

[10] D. I. Cartwright and M. J. Field. An Efficient Approximate Solution to the Kiefer-Weiss Problem. *Proc. Amer. Math. Soc.*, 71:36–38, 1978.

[11] G. J. Chaitin. Randomness in Arithmetic. *Scientific American*, 259:80–85, July 1988.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1959.

[13] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey, 1946.

[14] I. Csiszar. Information theoretic methods in probability and statistics. *IEEE Information Theory Society Newsletter*, 48:21–30, 1998.

[15] A. I. El-Fallah, R. P. Mahler, T. Zajic, E. Sorensen, M. G. Alford, and R. K. Mehra. Scientific Performance Evaluation for Sensor Menagement. In *Signal Processing, Sensor Fusion, and Target Recognition IX*, volume SPIE 4052, pages 183–194, 2000.

[16] B. E. Fridling and O. E. Drummond. Performance Evaluation Methods for Multiple Target Tracking Algorithms. In *Proc. SPIE on Signal and Data Processing for Small Targets,* vol. 1481, 1991.

[17] J. P. Keating, R. L. Mason, and P. K. Sen. *Pitman's Measure of Closeness: A Comparison of Statistical Estimators*. SIAM, Philadelhia, 1993.

[18] S. Kotz and N. L. Johnson. *Encyclopedia of Statistical Sciences*, volume 3. Wiley, New York, 1983.

[19] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[20] D. Lerro and Y. Bar-Shalom. Tracking with Debiased Consistent Converted Measurements vs. EKF. *IEEE Trans. Aerospace and Electronic Systems*, AES-29(3):1015–1022, July 1993.

[21] X. R. Li. *Probability, Random Signals, and Statistics*. CRC Press, Boca Raton, FL, 1999.

[22] X. R. Li. Model-Set Design for Multiple-Model Estimation—Part I. In *Proc. 2002 International Conf. on Information Fusion*, pages 26–33, Annapolis, MD, USA, July 2002.

[23] X. R. Li. Applied Estimation and Filtering, May 2003. Course Notes, Univ. of New Orleans.

[24] X. R. Li. Optimal Linear Estimation Fusion–Part VII: Dynamic Systems. In *Proc. 2003 International Conf. on Information Fusion*, pages 455–462, Cairns, Austrailia, July 2003.

[25] X. R. Li. Recursibility and Optimal Recursive Linear Estimation and Filtering. In *Proc. 43th IEEE Conf. on Decision and Control*, pages 1761–1766, Dec. 2004.

[26] X. R. Li and C. He. Model-Set Choice for Multiple-Model Estimation. In *Proc. IFAC 14th World Congress*, pages 169–174, Beijing, China, July 1999. Paper no. 3a-154.

[27] X. R. Li and C. He. Model-Set Design, Choice, and Comparison for Multiple-Model Estimation. In *Proc. 1999 SPIE Conf. on Signal and Data Processing of Small Targets,* vol. 3809, pages 501–513, Denver, CO, USA, July 1999.

[28] X. R. Li and V. P. Jilkov. A Survey of Maneuvering Target Tracking—Part III: Measurement Models. In *Proc. 2001 SPIE Conf. on Signal and Data Processing of Small Targets,* vol. 4473, pages 423–446, San Diego, CA, USA, July-Aug. 2001.

[29] X. R. Li and J. Wang. Unified Optimal Linear Estimation Fusion—Part II: Discussions and Examples. In *Proc. 2000 International Conf. on Information Fusion*, pages MoC2.18–MoC2.25, Paris, France, July 2000.

[30] X. R. Li and K. S. Zhang. Optimal Linear Estimation Fusion—Part IV: Optimality and Efficiency of Distributed Fusion. In *Proc. 2001 International Conf. on Information Fusion*, pages WeB1.19–WeB1.26, Montreal, QC, Canada, Aug. 2001.

[31] X. R. Li, K.-S. Zhang, J. Zhao, and Y.-M. Zhu. Optimal Linear Estimation Fusion—Part V: Relationships. In *Proc. 2002 International Conf. on Information Fusion*, pages 497–504, Annapolis, MD, USA, July 2002.

[32] X. R. Li and P. Zhang. Optimal Linear Estimation Fusion—Part III: Cross-Correlation of Local Estimation Errors. In *Proc. 2001 International Conf. on Information Fusion*, pages WeB1.11–WeB1.18, Montreal, QC, Canada, Aug. 2001.

[33] X. R. Li, Z. Zhao, and V. P. Jilkov. Practical Measures and Test for Credibility of an Estimator. In *Proc. Workshop on Estimation, Tracking, and Fusion — A Tribute to Yaakov Bar-Shalom*, pages 481–495, Monterey, CA, May 2001.

[34] X. R. Li, Z. Zhao, and V. P. Jilkov. Estimator's Credibility and Its Measures. In *Proc. IFAC 15th World Congress*, Barcelona, Spain, July 2002. Paper no. 2779.

[35] X. R. Li and Z. L. Zhao. Measures of Performance for Evaluation of Estimators and Filters. In *Proc. 2001 SPIE Conf. on Signal and Data Processing of Small Targets,* vol. 4473, pages 530–541, San Diego, CA, USA, July-August 2001.

[36] X. R. Li and Z. L. Zhao. Practical Measures for Performance Evaluation of Estimators and Filters. In *Proc. Workshop on Estimation, Tracking, and Fusion — A Tribute to Yaakov Bar-Shalom*, pages 467–480, Monterey, CA, USA, May 2001.

[37] X .R. Li and Z. L. Zhao. Relative Error Measures for Evaluation of Estimation Performance. In *Proc. 2005 International Conf. on Information Fusion*, Philadelphia, PA, July 2005.

[38] X .R. Li and Z. L. Zhao. Evaluation of Estimation Algorithms Part I: Local Performance Measures. *IEEE Trans. Aerospace and Electronic Systems*, Jul. 2006.

[39] X .R. Li and Z. L. Zhao. Measuring Estimator's Credibility: Noncredibility Index. Submitted to 2006 International Conference on Information Fusion, July 2006.

[40] X .R. Li and Z. L. Zhao. Testing Estimator's Credibility-Part I: Tests for MSE. Submitted to 2006 International Conference on Information Fusion, July 2006.

[41] X .R. Li and Z. L. Zhao. Testing Estimator's Credibility-Part II: Other Tests. Submitted to 2006 International Conference on Information Fusion, July 2006.

[42] X .R. Li, Z. L. Zhao, and X. Bai Li. General Model-Set Design Methods for Multiple-Model Approach. *IEEE Trans. Automatic Control*, AC-50:1260–1276, Sep. 2005.

[43] X. R. Li, Z. L. Zhao, P. Zhang, and C. He. Model-Set Design, Choice, and Comparison for Multiple-Model Approach to Hybrid Estimation. In *Proc. Workshop on Signal Processing, Communications, Chaos and Systems*, pages 59–92, Newport, RI, USA, June 2002.

[44] X. R. Li, Z. L. Zhao, P. Zhang, and C. He. Model-Set Design for Multiple-Model Estimation—Part II: Examples. In *Proc. 2002 International Conf. on Information Fusion*, pages 1347–1354, Annapolis, MD, USA, July 2002.

[45] X. R. Li, Y. M. Zhu, and C. Z. Han. Unified Optimal Linear Estimation Fusion—Part I: Unified Models and Fusion Rules. In *Proc. 2000 International Conf. on Information Fusion*, pages MoC2.10–MoC2.17, Paris, France, July 2000.

[46] X. R. Li, Y. M. Zhu, J. Wang, and C. Z. Han. Optimal Linear Estimation Fusion—Part I: Unified Fusion Rules. *IEEE. Trans. Information Theory*, IT-49(9):2192–2208, Sept. 2003.

[47] M. Miller and O. Drummond. Coordinate Transformation Bias in Target Tracking. In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets 1999*, pages 409–424, 1999. SPIE Vol. 3809.

[48] M. D. Miller and O. E. Drummond. Comparison of Methodologies for Mitigating Coordinate Transformation Bias in Target Tracking. In *Proc. 2000 SPIE Conf. on Signal and Data Processing of Small Targets,* vol. 4048, pages 414–427, Orlando, Florida, USA, April 2000.

[49] L. Mo, X. Song, Y. Zhou, and Z. Sun. An Alternative Unbiased Consistent Converted Measurements for Target Tracking. In *Proceedings of SPIE: Acquisition, Tracking, and Pointing XI, Vol. 3086*, pages 308–310, 1997.

[50] L. Mo, X. Song, Y. Zhou, Z. Sun, and Y. Bar-Shalom. Unbiased Converted Measurements for Tracking. *IEEE Trans. Aerospace and Electronic Systems*, AES-34(3):1023–1026, 1998.

[51] C. P. Robert, J. T. G. Hwang, and W. E. Strawderman. Is Pitman closeness a reasonable criterion? *Journ. Amer. Statist. Assoc.*, 88:57–6, 1993.

[52] K. Rudolf. *Recursive Nonlinear Estimation: A Geometric Approach.* Springer, New York, 1996.

[53] L. J. Savage. *The Foundations of Statistics.* Wiley, New York, 1954.

[54] A. Stuart and K. Ord. *Kendall's Advanced Theory of Statistics—Vol. 1: Distribution Theory.* Edward Arnold, London, 1994.

[55] P. Suchomski. Explicit Expressins for Debiased Statistics of 3D Converted Measurements. *IEEE Trans. Aerospace and Electronic Systems*, AES-35(1):368–370, 1999.

[56] M. Toda and R. Patel. Performance Bounds for Continuous-Time Filters in the Presence of Modeling Errors. *IEEE Trans. Aerospace and Electronic Systems*, AES-14(6):912–920, Nov. 1978.

[57] H. Van Trees. *Detection, Estimation and Modulation Theory.* Wiley, New York, 1968.

[58] T. Zajic, J. L. Hoffman, and R. P. Mahler. Scientific Performance Metrics for Data Fusion: New Results. In *Signal Processing, Sensor Fusion, and Target Recognition IX*, volume SPIE 4052, pages 172–182, 2000.

[59] T. Zajic and R. P. Mahler. Practical Information-based Data Fusion Evaluation. volume SPIE 3720, pages 92–103, 1999.

[60] Z. L. Zhao, H. M. Chen, G. Chen, C. Kwan, and X .R. Li. Comparison of Several Ballistic Target Tracking Filters. to appear in 2006 American Control Conf., June 2006.

[61] Z. L. Zhao, H. M. Chen, G. Chen, C. Kwan, and X .R. Li. IMM-LMMSE Filtering Algorithm for Ballistic Target Tracking with Unknown Ballistic Coefficient. In *2006 SPIE Conf. on Signal and Data Processing of Small Targets*, Orlando, Florida, USA, April 2006.

[62] Z. L. Zhao, H. M. Chen, and X .R. Li. Semi-parametric Model Selection With Applications to Regression. In *IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, June 2005.

[63] Z. L. Zhao and X .R. Li. Probabilistic Model Distortion Measure and Its Application to Multiple Model Approach. In *Proc. 38th Asilomar Conference on Signals, Systems and Computers*, pages 2146–2150, Pacific Grove, CA, Nov. 2004.

[64] Z. L. Zhao and X .R. Li. Interaction between the Estimator and Estimation Criteria and Its application. In *Proc. 2005 International Conf. on Information Fusion*, Philadelphia, PA, July 2005.

[65] Z. L. Zhao and X .R. Li. The Behavior of the Model Probability in the Multiple Model Approach. In *Proc. 2005 International Conf. on Information Fusion*, Philadelphia, PA, July 2005.

[66] Z. L. Zhao, X .R. Li, and V. P. Jilkov. Best linear unbiased filtering with spherical measurements for target tracking. In *Proc. 2003 SPIE Conf. on Signal and Data Processing of Small Targets,* vol. 5204, San Diego, CA, Aug. 2003.

[67] Z. L. Zhao, X .R. Li, V. P. Jilkov, and Y. M. Zhu. Optimal linear unbiased filtering with polar measurements for target tracking. In *Proc. 2002 International Conf. on Information Fusion*, pages 1527–1534, Annaplis, MD, USA, July 2002.

[68] Z. L. Zhao, X .R. Li, and V.P.Jilkov. Best Linear Unbiased Filtering with Nonlinear Measurements for Target Tracking. *IEEE Trans. Aerospace and Electronic Systems*, AES-40:1324–1336, Oct. 2004.

[69] Y. M. Zhu and X. R. Li. Best Linear Unbiased Estimation Fusion. In *Proc. 1999 International Conf. on Information Fusion*, pages 1054–1061, Sunnyvale, CA, USA, July 1999.

Vita

Zhanlue Zhao received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, P. R. China, in 1997 and 2000, respectivelyand Ph.D. degree in electrical engineering from University of New Orleans, U.S.A. He joined Sentient Corporation, ID, U.S.A. in 2006 where he is a research engineer.

He has authored and coauthored 25 journal and conference proceedings papers. His current research interests include signal and data processing, information fusion, detection and estimation theory, information theory and RKHS approach with applications to fault diagnosis/prognosis and target tracking.