IMPROVED STATE ESTIMATION FOR JUMP MARKOV LINEAR SYSTEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

UMUT ORGUNER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

DECEMBER 2006

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. İsmet Erkmen
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Mübeccel Demirekler
Supervisor

Examining Committee Members

Prof. Dr. Kemal Leblebicioğlu	(METU, EE)	_____
Prof. Dr. Mübeccel Demirekler	(METU, EE)	_____
Prof. Dr. Aydan M. Erkmen	(METU, EE)	_____
Prof. Dr. Ömer Morgül	(Bilkent University, EE)	_____
Prof. Dr. Hitay Özbay	(Bilkent University, EE)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Umut Orguner

Signature :

ABSTRACT

IMPROVED STATE ESTIMATION FOR JUMP MARKOV LINEAR SYSTEMS

Orguner, Umut

Ph.D., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Mübeccel Demirekler

December 2006, 183 pages

This thesis presents a comprehensive example framework on how current multiple model state estimation algorithms for jump Markov linear systems can be improved. The possible improvements are categorized as:

- Design of multiple model state estimation algorithms using new criteria.
- Improvements obtained using existing multiple model state estimation algorithms.

In the first category, risk-sensitive estimation is proposed for jump Markov linear systems. Two types of cost functions namely, the instantaneous and cumulative cost functions related with risk-sensitive estimation are examined and for each one, the corresponding multiple model estate estimation algorithm is derived. For the cumulative cost function, the derivation involves the reference probability method where one defines and uses a new probability measure under which the involved processes has independence properties. The performance of the proposed risk-sensitive filters are illustrated and compared with conventional algorithms using simulations.

The thesis addresses the second category of improvements by proposing

- Two new online transition probability estimation schemes for jump Markov linear systems.
- A mixed multiple model state estimation scheme which combines desirable properties of two different multiple model state estimation methods.

The two online transition probability estimators proposed use the recursive Kullback-Leibler (RKL) procedure and the maximum likelihood (ML) criteria to derive the corresponding identification schemes. When used in state estimation, these methods result in an average error decrease in the root mean square (RMS) state estimation errors, which is proved using simulation studies.

The mixed multiple model estimation procedure which utilizes the analysis of the single Gaussian approximation of Gaussian mixtures in Bayesian filtering, combines IMM (Interacting Multiple Model) filter and GPB2 (2nd Order Generalized Pseudo Bayesian) filter efficiently. The resulting algorithm reaches the performance of GPB2 with less Kalman filters.

Keywords: Multiple model, state estimation, jump Markov linear system, transition probability, Markov chain, interacting multiple model, IMM, risk sensitive

ÖZ

MARKOV ATLAMALI DOĞRUSAL SİSTEMLER İÇİN GELİŞTİRİLMİŞ DURUM KESTİRİMİ

Orguner, Umut
Doktora, Elektrik Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Mübeccel Demirekler

Aralık 2006, 183 sayfa

Bu tez çalışması Markov atlamalı doğrusal sistemlerde çoklu modelli durum kestiriminin geliştirilmesi için kapsamlı bir örnek çerçeve sunmaktadır. Olası geliştirimler iki kategoride incelenmektedir:

- Yeni kriterler kullanarak çoklu modelli durum kestirimi.
- Varolan çoklu modelli durum kestirim algoritmaları kullanılarak yapılabilecek geliştirmeler.

Birinci kategoride Markov atlamalı doğrusal sistemler için risk duyarlı kestirim önerilmektedir. Risk duyarlı kestirim için literatürde bulunan anlık ve birikimli maliyet fonksiyonları incelenmiş ve herbirine takabül eden risk duyarlı kestirim yöntemleri türetilmiştir. Birikimli maliyet fonksiyonu için yapılan türetimde ilgili süreçlerin altında bağımsız olduğu yeni bir olasılık ölçüsünün tanımlanıp kullanıldığı referans olasılık yöntemi kullanılmıştır. Önerilen risk duyarlı algoritmaların başarımları benzetim çalışmaları ile gösterilip geleneksel yöntemlerle karşılaştırılmıştır.

Tez çalışması ikinci kategoride

- Markov atlamalı doğrusal sistemler için iki çevrimiçi geçiş olasılığı kestirim algoritması
- Bir karma çoklu modelli durum kestirim algoritması

önermektedir.

Çevrimiçi geçiş olasılığı kestirim algoritmaları, tanılama yöntemlerini türetebilmek için sırasıyla özyineli Kullback-Leibler yöntemini ve en büyük olabilirlik kriterini kullanmaktadır. Durum kestiriminde kullanıldıkları zaman, bu algoritmaların durum hatalarının etkin değerlerini düşürdüğü benzetim çalışmaları ile gösterilmiştir.

Gaussian karışımlarının tek Gaussian ile yaklaşıklanmasının etkilerini inceleyen bir analizin sonuçlarını kullanan karma çoklu modelli durum kestirim algoritması IMM (etkileşimli çoklu model) süzgeci ve GPB2 (ikinci derece genelleştirilmiş yalancı Bayesian) süzgeçlerini verimli bir şekilde birleştirmektedir. Sonuçta elde edilen algoritma GPB2'nin başarımına daha az Kalman süzgeci kullanarak ulaşmaktadır.

Anahtar Kelimeler: Çoklu model, durum kestirimi, Markov atlamalı doğrusal sistem, geçiş olasılığı, Markov zinciri, etkileşimli çoklu model, IMM, risk duyarlı

To those who contributed to this thesis study,

ACKNOWLEDGMENTS

It is a pleasure for me to express my sincere gratitude to my thesis supervisor Prof. Dr. Mübeccel Demirekler for her belief, patience, encouragement and guidance throughout the study. I greatly appreciate her share in every step taken in the development of the thesis.

I would also like to thank the members of my thesis committee, Prof. Dr. Kemal Leblebicioğlu and Prof. Dr. Ömer Morgül and the visitors of my thesis progress presentations, Prof. Dr. Hitay Özbay and Asst. Prof. Dr. Afşar Saranlı for their support, encouragement and suggestions which improved the quality of the thesis.

I should also mention the part of the faculty of my department, especially of the faculty members whose courses I took during my M.S. and Ph.D. studies, in shaping the path of this research. I am also grateful for the ideal studying conditions that my department and university has provided me with during my 7 year research and teaching assistantship. Especially, the extensive experience I have gained about digital signal processing will contribute a lot to my academic life.

Moreover, I owe much to my office, department and project mates Emre Özkan, Evren İmre and İ. Yücel Özbek for their sharing my (otherwise alone) days and nights in the department.

I will also never forget the unending support and love my family has provided me all my life.

PREFACE

This thesis is the result of the author's 4-year Ph.D. study under the supervision of Prof. Dr. Mübeccel Demirekler. Originally, each chapter of it was separate research on the general framework of multiple model estimation. While writing the final copy, I tried hard to integrate the material as much as possible into a single and whole study. Most of the material presented has already been submitted to academic journals for publication. The following is a list of these publications:

- U. Orguner and M. Demirekler, "An online sequential algorithm for the estimation of transition probabilities for jump Markov linear systems," *Automatica*, vol. 42, no. 10, pp. 1735–1744, Oct. 2006.

- U. Orguner and M. Demirekler, "Risk-sensitive filtering for jump Markov linear systems," Submitted to *Automatica*.

- U. Orguner and M. Demirekler, "Analysis of the effects of the single Gaussian approximation of Gaussian mixtures in Bayesian filtering with applications to mixed multiple-model estimation algorithms," Submitted to *International Journal of Control*.

- U. Orguner and M. Demirekler, "Maximum likelihood estimation of transition probabilities of jump Markov linear systems," To be submitted to *IEEE Transactions on Signal Processing*.

At this stage, I must admit that, maybe for the future graduate students, the material presented here did not actually pop up in my mind instantly. Chronologically (and maybe ironically[1]), it was the research in Chapter 6 on mixed multiple model estimation algorithm that was partially completed as the result of my long desire to analyze the approximations involved in well-known multiple model estimation algorithms. The results of Chapter 6 were actually a beginning study on the deviations of the multiple model filtering

[1] This is because Chapter 6 is the last chapter before the conclusions.

methods from optimality. The mixed IMM-GPB2 algorithm came into picture as a by-product of this beginning which could not go any further due to the high amount of nonlinearity existing in the filtering methods.

It was after this discouragement that I realized the work of Jilkov & Li [1] on the online estimation of transition probabilities associated with jump Markov linear systems (JMLSs). The subject was new and there happened to exist many methods for a similar problem in hidden Markov models (HMMs) which are not touched by Jilkov & Li. Using this motivation, I attempted to apply the recursive Kullback-Leibler (RKL) algorithm [2] to JMLSs which is presented in Chapter 4. The outcome was more impressive than I could foresee in that the whole method was to be re-derived for JMLSs due to the fact that the Markov chain is buried under the measurement process deeper in JMLSs than in HMMs.

While studying the HMM literature for possible transition probability estimation methods, I was also carrying out research on state estimation which has always sounded more academically fruitful. At those times, I realized the risk-sensitive state estimation for HMMs. Since the state process of HMMs is discrete-valued, it was not directly possible to apply the ideas to JMLSs which have both continuous and discrete-valued states. What was more interesting was actually the reference probability method used for deriving the risk-sensitive estimator for HMMs. This has motivated me to turn towards some measure theoretical probability and reference probability method which had always intrigued me under a different title "Stochastic Differential Equations". After spending many months on these and risk-sensitive estimator theory, I could derive the risk-sensitive filters for JMLSs which are presented in Chapters 2 and 3. I actually set off first to derive the work in Chapter 2 using the reference probability method but it did not take much for me to notice that our old favorite classical probability theory was, in fact, enough. The risk-sensitive estimator given in Chapter 3 was quite more challenging (and indeed much more fun!) to derive. That derivation triggered a big-bang for me to understand (at last) that this was only the tip of the iceberg that

I am wandering on. Using the techniques used for Chapter 3, it was quite straightforward to obtain the results of Chapter 5 after (a couple of months of) studying maximum likelihood estimation theory using the reference probability method.

As easily observed from the mentioned story above, the process of forming this Ph.D. thesis was full of inspiration and perspiration.[2] For the perspiration, I tried to mention the ones I owe much in the dedication and acknowledgements parts but now, I would like to thank V. Krishnamurthy, V. Jilkov, X.R. Li, S. Dey, J.B. Moore, R.J. Elliott, L. Aggoun, I.B. Collins, J.J. Ford (none of whom I have ever met) and others that I could not add here[3] for their research which contributed to the inspiration part of this thesis.

Umut Orguner
Ankara, Turkey
August 25th, 2006

[2] At the time being, unfortunately and surprisingly, I can not say which one was more in it.

[3] The names in the references part are at the beginning of this list of others.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Even with his great knowledge and imagination, it is not known whether Norbert Wiener could dream of the current expansion of the subject he worked on when he passed away in 1964 after only four years Kalman published his seminal paper [3]. What is known about him is that it was the anti-aircraft fire control problems that led him to apply statistical ideas to linear filtering and prediction problem which are presented in his famous monograph [4]. This work was a revolution which still echoes in the engineering community and it inspired many great researchers including the "founder" of the information theory, Claude E. Shannon, who is quoted in Kailath's linear filtering survey [5] as

> Credit should also be given to Professor N. Wiener, whose elegant solution of the problems of filtering and prediction of stationary ensembles has considerably influenced the writer's thinking in this field.

As the title of Wiener's monograph expresses, Wiener worked on stationary signals whose characteristics are summarized in terms of covariance functions. The generalization of his study to non-stationary signals came after about ten years with the work [3] of R.E. Kalman whose ideas dominated the modern control theory ever after. During the race of the space programs between the two countries of the world in the sixties, Kalman's ideas proved quite useful. According to Green and Limebeer [6], there were two main reasons for this:

1. For the space vehicles, it is possible to develop mathematical models of their behaviour thanks to the essentially ballistic character of their

dynamics. In addition, descriptions of external disturbances based on white noise are often appropriate in aerospace applications.

2. Many of the control problems from the space program are concerned with resource management. Performance criteria of this type are easily embedded in the so-called linear quadratic Gaussian (LQG) framework .

The benefit gained by the usage of Kalman's work in space technology led the researchers to try the same algorithms in conventional industrial applications, which resulted in a disappointment. In industrial processes, the mathematical models used by the estimators are never exact and the true model of the system might actually be changing. Moreover, the external disturbances affecting the system and measurements hardly conform to statistical descriptions like whiteness or stationarity. These constraints have made the communities reassess the status of estimation (and control) theory and surged an enormous number of publications under the title of robust estimation theory in the last three decades. In addition to the combination of the estimation algorithms with model adaptation rules, $\mathcal{H}_\infty$, guaranteed-cost, set-valued, multiple-model, and risk-sensitive estimation algorithms appeared as only some of the tools that have been developed during this period. In fact, being exhaustive about the the existing methods is impossible and beyond the scope. In this thesis, the concentration will be on the multiple model estimation in which the model uncertainty of the system under consideration is covered using a finite number of models.

The history of multiple model estimation dates back to 1970 when the work of Ackerson and Fu [7] named "On state estimation in switching environments" first appeared. Specifically, they considered discrete-time linear systems whose disturbances (noise terms) are assumed to come from one of several Gaussian distributions with different means and variances. The changes in the noise distributions, which represented different environments the system is running under, were according to the state of a Markov chain. This model was actually a specific case of the so-called jump Markov linear systems (JMLSs)

which are linear systems whose parameters evolve according to a finite-state Markov chain. In the analysis of [7], Ackerson and Fu realized the exponentially growing memory and computation requirements of the optimal minimum mean square error (MMSE) estimator for this type of systems. The suboptimal algorithm they proposed to solve this problem involved the approximation of Gaussian mixtures by a single Gaussian density and was actually a specific case of a broader type of algorithms which are called by Tugnait [8] as the generalized pseudo Bayesian (GPB) algorithms.

The exponential growing of the requirements of the optimal MMSE algorithm for JMLSs is caused simply by the exponentially increasing number of possible state histories (hypotheses) of the underlying Markov chain. Limiting the growing number of hypotheses is the general characteristics of the proposed solutions. Akashi and Kumamoto proposed the random selection of a predetermined number of hypotheses and discarding the remaining ones in [9]. More wisely, the selection of the most likely predetermined number of hypotheses is proposed in [10] based on the posterior probability of the hypotheses. In 1982, Tugnait published an extensive survey and comparison (using equal computation resources) paper on the existing multiple model estimation algorithms which concluded that GPB algorithms are, in general, to be preferred to other algorithms. However, it was also emphasized that the performances of the algorithms are scenario dependent and no single algorithm is best for all situations.

After approximately six years of silence in the community, Blom and Bar-Shalom [11] suggested an algorithm which has a better place than the GPB algorithms on the performance vs. computation curve in 1988. The algorithm, which was called as the interacting multiple model (IMM) algorithm, in the words of Johnston and Krishnamurthy [12], revitalized the field of multiple model estimation and attracted much attention. This algorithm has been used in many real world applications the most important of which is target tracking [13, 14]. After the IMM algorithm became popular, the interest has been shifted to different cost functions and filter structures for multiple model

estimation. In this regard, the linear MMSE [15, 16], optimal control based [17] and maximum a posteriori (MAP) estimation based [18, 12] approaches has followed the IMM filter. The stochastic sampling based algorithms, which have become popular after the introduction of the particle filters into the area of state estimation [19], were also applied to the case of JMLSs in [20, 21, 22].

This thesis is about (further) improvement of the multiple model estimation algorithms. The wide coverage of methods applied in the field as described above makes the improvement over the existing algorithms a challenging work. The existing multiple model algorithms, in author's opinion, can be improved in two different and possibly intersecting ways:

- The first way is to generate completely new algorithms which use different criteria and means to obtain optimal estimates.

- The second way is to use the algorithms in the literature with optimized parameters and optimized schemes. In this way, the best possible performance can be obtained from the existing algorithms.

The first approach, although being more fruitful and academically interesting, is much more difficult to achieve due to the wide spectrum of existing algorithms. In this thesis, both of the ways described above are used to achieve an improvement over the existing multiple model estimation methods in the literature. With this content, the overall thesis can be divided into two parts which cover the first and the second way of improvements respectively.[1]

As mentioned above, the first way to achieve an improvement over the existing multiple model estimation algorithms, which is to find optimal methods with novel criteria is much more challenging than the second way. The key for answering this challenge is to look at the development of the estimation theory through the years. It is interesting that the theory of state estimation has been developed as a response to its control theory counterpart most of

[1] Originally, the thesis was formatted to include these part units but the graduate school insisted on removing them for format standardization. In the current format, Chapters 2 and 3 form Part-I and Chapters 4, 5 and 6 form Part-II.

the time. This aspect of estimation theory is evident from the earlier linear quadratic (Gaussian) control days of the 60's to the more recent $\mathcal{H}_\infty$ control period of the 90's. It was with this idea in our mind that the risk sensitive control and estimation has attracted our attention. Risk sensitive control theory is related with the control problems in which one minimizes the expected exponential of a quadratic cost criterion [23, 24]. Risk sensitive estimation, which appeared after its control counterpart deals with estimation problems where the expected exponential of quadratic estimation error is minimized. There exist two different criteria for risk-sensitive estimation in the literature. The first one considers the instantaneous estimation error [25]. The risk-sensitive filter corresponding to this cost function, which is derived for linear Gauss-Markov systems appears to be the same as the Kalman filter [25]. The second type of cost function is a cumulative one which considers sum of all the estimation errors from the initial time to the current time. The risk sensitive filter for this cost function is also derived for the linear Gauss-Markov systems [26]. The resulting filter is still linear but it is different from the Kalman filter. In the first part of the thesis which is composed of Chapters 2 and 3, risk sensitive filters are derived for now JMLSs using both of the cost functions.

For the derivation of the first algorithm, which minimizes the instantaneous cost function and is given in Chapter 2, we use classical probabilistic methods. The resulting filter uses the IMM filter statistics and it differs from the IMM filter only at its output calculation step. The second filter, which minimizes the more complicated cumulative cost function and is presented in Chapter 3, is derived using the reference probability methods. This time, like the case in the conditional-mean estimators (or equivalently in MMSE estimators), optimal risk-sensitive multiple model filter turns out to be impossible to implement with exponentially growing memory and computation requirements. This problem is solved using IMM-type approximations in the resulting filter. The reference probability method used in the derivation provides a perfectly transparent framework for this purpose.

As for the methods of the second approach, we present two new online estimators for the transition probabilities associated with JMLSs and a mixed multiple model estimation algorithm in the second part of the thesis which is formed by Chapters 4, 5 and 6. Almost all multiple model estimation algorithms in the literature use constant and heuristically selected transition probabilities for the estimation task. In this framework, being able to estimate the transition probabilities in an online fashion would make the algorithms closer to optimality. To this end, we present two methods namely the recursive Kullback-Leibler (RKL) algorithm and maximum likelihood (ML) estimation algorithms in Chapters 4 and 5 respectively.

The RKL approach is an application of a technique previously applied to hidden Markov models (HMMs) to JMLSs. The resulting transition probability estimator minimizes a Kullback-Leibler divergence [27] based cost function using stochastic approximation [28] type recursions. On the oher hand, the ML transition probability estimation problem is solved by making use of the famous expectation-maximization (EM) procedure. The erratically (even more than exponentially) growing memory and computation requirements of the resulting exact EM algorithm makes us to approximate it by an N^3 component IMM filter where N is the number of models in the JMLS. The parameter estimates are then found using the mode-weights of this IMM filter.

Multiple model estimation algorithms in the literature generally use multiple Kalman filters for each measurement and they are amenable to parallel implementation. Making use of these properties, in Chapter 6 of the thesis, we propose a mixed (IMM-GPB2) multiple model estimation algorithm which can reach the performance of GPB2 algorithm with an average number of Kalman filters near to IMM algorithm. In order to obtain this result, the difference between the state estimates of IMM and GPB2 algorithms is examined analytically and a formula is found to quantify this difference.

Although some of the material presented in the thesis requires no special background or the related background is presented along with the results, the reference probability method which is used extensively in Chapters 3 and

5 needs special attention. Due to this, a brief background on the subject is presented in Appendix A which includes some theoretical aspects of the method as well as the derivation of the Kalman and risk-sensitive filters for linear Gauss-Markov systems using the reference probability method. Throughout the thesis, since the problems involved are changed from chapter to chapter, each chapter has its own problem definition even though some overlapping exists in the general framework. In addition to this, when same example is used in more than one chapter, the example statement is re-expressed for the sake of completeness. Moreover, there are some minor but inevitable repetitions in the issues and the references in the introduction sections of each chapter to avoid jumping to conclusions too early and for the sake of a clear presentation and easy reading.

1.1 Contributions

The major contributions of this thesis can be summarized as follows:

- A new multiple model estimation algorithm minimizing the expected exponential of cumulative estimation error is presented. The algorithm is shown to be superior to the IMM filter for an unknown parameter scenario.

- A new output calculation scheme for the IMM filter is found. The new output calculation mechanism makes the IMM filter minimize the expected exponential of instantaneous estimation error and the resulting estimates are slightly more robust to parameter uncertainties.

- Two new online transition probability estimation algorithms are proposed. With the online transition probability estimation, the existing multiple model state estimation algorithms work better than the ones using constant heuristic transition probabilities.

- An analysis of the effects of single Gaussian approximation is made. Using the results of this analysis, a mixed (IMM-GPB2) multiple model

estimation algorithm is proposed. The proposed algorithm can reach the performance of GPB2 with less Kalman filters and hence less computation.

CHAPTER 2

RISK-SENSITIVE MULTIPLE-MODEL STATE ESTIMATION: INSTANTANEOUS CASE

2.1 Introduction

Kalman filter is the most well-known state estimation tool in the literature for linear Gauss-Markov systems. One disadvantage of it is its sensitivity to modeling errors. The literature is abundant with approaches to overcome this drawback. Multiple-model filtering is a solution to this disadvantage when the uncertainty in the modeling can be covered by a finite number of models. The well-known interacting multiple model (IMM) filter [11, 13] and the generalized pseudo Bayesian (GPB) methods [7, 8] are the most famous of the multiple-model filtering algorithms. IMM filter, being actually an approximation of the second order GPB method (called as GPB2), has a computational load very near to a GPB1 filter and therefore, it is an efficient and popular estimation tool especially in target tracking community [14]. Other multiple-model filtering methods range from linear minimum mean-square error [15, 16], optimal control based [17], MAP estimation based [12, 18] approaches which can be classified as the classical approaches to stochastic sampling oriented methods [20, 21, 22] which have become popular after the introduction of the particle filters into the area of state estimation [19].

Although multiple-model filtering is itself a solution to the uncertainty in-

herent in the modeling process, generally, it is quite unlikely that the models covering the modes of a system under investigation are perfectly known. Two types of solutions is possible for this problem. The first solution is to add more models into the multiple-model filtering algorithm to cover more uncertainty. Nevertheless, adding more models may also lower the performance of the algorithm [29] and usage of the so-called variable structure multiple-model filtering algorithms might be necessary [29]. The second way to solve the problem is to apply the results of the research which can be named as the robust multiple model filtering. In [30, 31], the $\mathcal{H}_\infty$ methodologies have been applied to jump Markov linear systems (JMLSs) which are basically the linear systems whose parameters evolve according to a finite state Markov chain. Also, [32] considered the case of uncertain model parameters. However, in all these approaches, the underlying mode-sequence of the JMLS is assumed to be known. Later, [33] presented results related with the mode-independent case. The so called guaranteed-cost approaches, where the state estimators are designed such that the covariance of the estimation error is guaranteed to be within a certain bound for all admissible uncertainties, have also been applied to the case of JMLSs [34, 35]. However, these works also assume that the underlying mode-sequences are known. Quite recently, [36] dealt with the case of uncertain JMLSs and obtained solutions using the linear matrix inequality (LMI) approaches.

Risk-sensitive estimation, which appeared in the literature mainly after its control counterpart, is the general name given to the area of (robust [37]) estimation where the exponential of the (instantaneous or cumulative) quadratic estimation error is minimized. As mentioned in Chapter 1, there exist two different criteria for risk-sensitive estimation in the literature. The first one considers the instantaneous estimation error [25] and the second one uses a cumulative cost function which considers the sum of all the estimation errors from the initial time to current time [26]. In this chapter, the instantaneous cost function will be considered and the results for the cumulative cost function, which are more involved, will be given in Chapter 3. This chapter is the

least demanding part of the thesis in that classical probabilistic techniques are used for the derivation.

The risk-sensitive filter corresponding to the instantaneous cost function, which is derived for linear Gauss-Markov systems, appears to be the same as the Kalman filter [25]. This property basically results from the fact that the expected values of all even moments of the estimation error are minimized simultaneously by the Kalman filter state estimates when the conditional state distributions are Gaussian like the case in linear Gauss-Markov systems. Since the exponential of the quadratic estimation error is a weighted sum of (infinite number of) even moments of the quadratic estimation error (by Taylor series representation), expected exponential of the quadratic estimation error is minimized by the Kalman filter estimates for linear Gauss-Markov systems. In JMLSs, the case is different because, at each time step k, the conditional state densities turn out to be Gaussian mixtures.

This chapter is organized as follows. In Sec. 2.2, problem definition is made. An approximate solution for the risk sensitive multiple model estimation problem, which is called as IRS-IMM algorithm, will be derived in Sec. 2.3. The performance of the IRS-IMM algorithm is examined on a simulation scenario in Sec. 2.4. The chapter is finalized with conclusions in Sec. 2.5.

2.2 Problem Definition

The following jump Markov linear system model is considered

$$x_{k+1} = A(r_{k+1})x_k + B(r_{k+1})w_{k+1}, \tag{2.1}$$

$$y_k = C(r_k)x_k + D(r_k)v_k \tag{2.2}$$

where

- $\{x_k \in \mathbb{R}^n\}$ is the continuous-valued base-state sequence with initial distribution

$$x_0 \sim \mathcal{N}(x_0; \bar{x}_0, \Sigma_0), \tag{2.3}$$

where the notation $\mathcal{N}(x; \bar{x}, \Sigma)$ stands for a Gaussian probability density function for dummy variable x which has a mean $\bar{x}$ and covariance Σ. We assume $\Sigma_0 > 0$.

- $\{r_k\}$ is the unknown discrete-valued modal-state sequence,
- $\{y_k \in \mathbb{R}^m\}$ is the noisy observation sequence,
- $\{w_k \in \mathbb{R}^n\}$ is a white process noise sequence with distribution,

$$w_k \sim \mathcal{N}(w_k; 0, I_n), \tag{2.4}$$

where I_n denotes the identity matrix of size $n \times n$,

- $\{v_k \in \mathbb{R}^m\}$ is a white measurement noise sequence independent from the process noise w_k with distribution

$$v_k \sim \mathcal{N}(v_k; 0, I_m). \tag{2.5}$$

The discrete-valued modal-state $r_k \in \{1, 2, \ldots, N\}$ is assumed to be a first-order finite-state homogenous Markov chain with transition probability matrix $\Pi = [\pi_{ij}]$. The basic variables w_k, v_k, x_0 and the modal-state sequence r_k are assumed to be mutually independent for all k. The time-varying matrices $A(r_k)$, $B(r_k)$, $C(r_k)$, and $D(r_k)$ are assumed to be known for each value of r_k.

Our aim is to find a recursive (instantaneous) risk-sensitive estimate $\hat{x}_{k|k}^{IRS}$ defined as

$$\hat{x}_{k|k}^{IRS} \triangleq \arg\min_{\xi} E\left[\exp\left\{ \frac{\theta}{2}(x_k - \xi)^T Q_k (x_k - \xi) \right\} \middle| \mathcal{Y}_k \right] \tag{2.6}$$

where θ is a scalar generally called as the risk-sensitive parameter, $Q_k > 0$ is a known weighting matrix and $\mathcal{Y}_k \triangleq \sigma\{y_1, y_2, \ldots, y_k\}$ denotes the σ-algebra generated by the random variables $\{y_1, y_2, \ldots, y_k\}$.

2.3 Risk-Sensitive IMM Estimation

Interacting multiple model (IMM) filter is an approximate and efficient solution to the multiple-model minimum mean-square estimation problem [11, 13]. At

each time step k, IMM filter approximates the information state $p(x_k|\mathcal{Y}_k)$ as a Gaussian mixture defined as

$$p(x_k|\mathcal{Y}_k) = \sum_{j=1}^{N} \mu_k^j p(x_k|r_k = j, \mathcal{Y}_k) = \sum_{j=1}^{N} \mu_k^j \mathcal{N}(x_k; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \tag{2.7}$$

where the state estimates

$$\hat{x}_{k|k}^j \triangleq E[x_k|r_k = j, \mathcal{Y}_k] \tag{2.8}$$

and the covariances

$$\Sigma_{k|k}^j \triangleq E[(x_k - \hat{x}_{k|k}^j)(x_k - \hat{x}_{k|k}^j)^T|r_k = j, \mathcal{Y}_k] \tag{2.9}$$

are the approximate mode-conditioned state estimates and covariances respectively calculated by the IMM filter. The quantities μ_k^j denote the mode probabilities defined as

$$\mu_k^j \triangleq P(r_k = j|\mathcal{Y}_k). \tag{2.10}$$

In this respect, IMM filter is an efficient finite dimensional solution to calculate the conditional density $p(x_k|\mathcal{Y}_k)$ which is also required in the calculation of the expectation on the right hand side of Eqn. 2.6. For this reason, in the solution of the instantaneous risk-sensitive estimation problem, we are going to assume that the IMM calculated information state is available. In other words, for obtaining the the risk-sensitive state estimate $\hat{x}_{k|k}^{RS}$ defined by Eqn. 2.6, the IMM calculated statistics $\{\mu_k^j, \hat{x}_{k|k}^j, \Sigma_{k|k}^j\}_{j=1}^N$ will be necessary.

2.3.1 Derivation

Expectation involved in Eqn. 2.6 is by definition

$$E\left[\exp\left\{\frac{\theta}{2}(x_k - \xi)^T Q_k (x_k - \xi)\right\}\middle|\mathcal{Y}_k\right] = \int \exp\left\{\frac{\theta}{2}(x_k - \xi)^T Q_k (x_k - \xi)\right\} p(x_k|\mathcal{Y}_k) dx_k. \tag{2.11}$$

Using the approximated density $p(x_k|\mathcal{Y}_k)$ calculated by the IMM filter, we obtain

$$E\left[\exp\left\{\frac{\theta}{2}(x_k - \xi)^T Q_k (x_k - \xi)\right\}\middle|\mathcal{Y}_k\right]$$

$$= \sum_{j=1}^{N} \mu_k^j \int \exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\} \mathcal{N}(x_k; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) dx_k$$
$$= \sum_{j=1}^{N} \mu_k^j \mathcal{I}_j(\xi) \tag{2.12}$$

where

$$\mathcal{I}_j(\xi) \triangleq \int \exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\} \mathcal{N}(x_k; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) dx_k. \tag{2.13}$$

In order to calculate the integrals $\{\mathcal{I}_j(\xi)\}_{j=1}^N$, we expand the Gaussian term as

$$\begin{aligned}\mathcal{I}_j(\xi) &= \frac{1}{\sqrt{|2\pi\Sigma_{k|k}^j|}} \int \exp\left\{\frac{\theta}{2}(\xi-x_k)^T Q_k (\xi-x_k)\right\} \\ &\qquad \times \exp\left\{-\frac{1}{2}(x_k-\hat{x}_{k|k}^j)^T (\Sigma_{k|k}^j)^{-1}(x_k-\hat{x}_{k|k}^j)\right\} dx_k \end{aligned} \tag{2.14}$$

where the notation $|.|$ denotes the matrix determinant. If we define

$$M_k^j \triangleq \left[(\Sigma_{k|k}^j)^{-1} - \theta Q_k\right]^{-1} > 0, \tag{2.15}$$

$$S_k^j \triangleq \left[\frac{1}{\theta} Q_k^{-1} - \Sigma_{k|k}^j\right]^{-1} > 0, \tag{2.16}$$

we can take the integral in Eqn. 2.14 using the result of App. B.3 as

$$\mathcal{I}_j(\xi) = \sqrt{\frac{|M_k^j|}{|\Sigma_{k|k}^j|}} \exp\left\{\frac{1}{2}(\xi-\hat{x}_{k|k}^j)^T S_k^j (\xi-\hat{x}_{k|k}^j)\right\}. \tag{2.17}$$

Substituting this result into Eqn. 2.12, we obtain

$$\begin{aligned} E\left[\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\} \middle| \mathcal{Y}_k\right] \\ = \sum_{j=1}^{N} \mu_k^j \sqrt{\frac{|M_k^j|}{|\Sigma_{k|k}^j|}} \exp\left\{\frac{1}{2}(\xi-\hat{x}_{k|k}^j)^T S_k^j (\xi-\hat{x}_{k|k}^j)\right\}. \end{aligned} \tag{2.18}$$

Consequently, the instantaneous risk-sensitive state estimate is given as

$$\hat{x}_{k|k}^{IRS} \triangleq \arg\min_{\xi} \sum_{j=1}^{N} \mu_k^j \sqrt{\frac{|M_k^j|}{|\Sigma_{k|k}^j|}} \exp\left\{\frac{1}{2}(\xi-\hat{x}_{k|k}^j)^T S_k^j (\xi-\hat{x}_{k|k}^j)\right\}. \tag{2.19}$$

The mixture of exponentials on the right hand side of Eqn. 2.19 is impossible to minimize analytically. The numerical algorithms can be applied for the

minimization. This type of numerical solution which needs multiple iterations for each time step k is quite costly for many applications, especially for target tracking where real-time algorithms are required. Moreover, even if a numerical algorithm is selected for making the minimization, the cost function evaluation procedure might cause an overflow in the computer (or processor) due to the blowing characteristics of the exponential functions. These issues necessitate an approximation of the cost function to make an analytical minimization. We, at this point, choose to approximate the exponentials on the right hand side of Eqn. 2.19 by their first order Taylor series expansion i.e., $\exp(x) \approx 1 + x$. Then, we get

$$\hat{x}^{IRS}_{k|k} \approx \arg\min_{\xi} \sum_{j=1}^{N} \mu^j_k \sqrt{\frac{|M^j_k|}{|\Sigma^j_{k|k}|}} \frac{1}{2} (\xi - \hat{x}^j_{k|k})^T S^j_k (\xi - \hat{x}^j_{k|k}). \tag{2.20}$$

The cost function on the right hand side of Eqn. 2.20 is quadratic in the minimization variable ξ. Therefore, after taking the gradient with respect to ξ, equating to zero and solving for ξ, we obtain the following unique solution

$$\hat{x}^{IRS}_{k|k} = \left[\sum_{j=1}^{N} \mu^j_k \sqrt{\frac{|M^j_k|}{|\Sigma^j_{k|k}|}} S^j_k \right]^{-1} \sum_{j=1}^{N} \mu^j_k \sqrt{\frac{|M^j_k|}{|\Sigma^j_{k|k}|}} S^j_k \hat{x}^j_{k|k}. \tag{2.21}$$

Since $M^j_k = \Sigma^j_{k|k} S^j_k \frac{1}{\theta} Q^{-1}_k$, we can simplify the solution as

$$\hat{x}^{IRS}_{k|k} = \left[\sum_{j=1}^{N} \mu^j_k \sqrt{|S^j_k|} S^j_k \right]^{-1} \sum_{j=1}^{N} \mu^j_k \sqrt{|S^j_k|} S^j_k \hat{x}^j_{k|k} \tag{2.22}$$

where

$$S^j_k \;=\; \left[\frac{1}{\theta} Q^{-1}_k - \Sigma^j_{k|k} \right]^{-1}. \tag{2.23}$$

As a result, using the statistics of IMM algorithm, the risk-sensitive filtering modifies the output estimate calculation of the IMM to include the covariance weights depending on the IMM calculated mode-conditioned covariances. We will call the overall algorithm which is composed of IMM filter (information state or statistics) recursions and the risk-sensitive output calculation as the IRS-IMM algorithm (or filter) where the abbreviation "IRS" stands for "instantaneous risk-sensitive".

2.3.2 Properties of IRS-IMM Filter

In this section, we are going to emphasize some properties of the filter derived in the previous section. For this purpose, we define a hypothetical (quite non-rigorous) operator $\mathcal{T}$ which replaces every exponential in its operand expression with its first order Taylor series approximation, i.e.,

$$\mathcal{T}(\exp(x)) = 1 + x. \tag{2.24}$$

Using this operator, we can define MMSE estimate as

$$\hat{x}_{k|k}^{MMSE} \triangleq \arg\min_{\xi} E\left[\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\middle|\mathcal{Y}_k\right] \tag{2.25}$$

$$= \arg\min_{\xi} E\left[\mathcal{T}\left(\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\}\right)\middle|\mathcal{Y}_k\right]. \tag{2.26}$$

In [25], by showing the equivalence of the MMSE estimate and the risk-sensitive estimates for linear Gauss-Markov systems, it is shown that the equality

$$\arg\min_{\xi} E\left[\mathcal{T}\left(\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\}\right)\middle|\mathcal{Y}_k\right]$$
$$= \arg\min_{\xi} E\left[\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\}\middle|\mathcal{Y}_k\right] \tag{2.27}$$

is satisfied for linear Gauss-Markov systems. In other words, the existence of the operator $\mathcal{T}$ inside the expectation does not change the minimizing point of the cost function. What is proved above by the derivation of the risk-sensitive output estimate calculation formula is that this is not the case for jump Markov linear systems. Instead, our approximate output calculation method finds the risk-sensitive estimate using the following formula.

$$\arg\min_{\xi} E\left[\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\}\middle|\mathcal{Y}_k\right]$$
$$\approx \arg\min_{\xi} \mathcal{T}\left(E\left[\exp\left\{\frac{\theta}{2}(x_k-\xi)^T Q_k (x_k-\xi)\right\}\middle|\mathcal{Y}_k\right]\right). \tag{2.28}$$

A general property for risk-sensitive filters is that their estimates converge to MMSE estimates when the risk sensitive parameter θ goes to zero. Using the definition of the matrices S_k^j, we see that

$$S_k^j \to \theta Q_k \quad \text{for} \quad j = 1, \ldots, N \quad \text{when} \quad \theta \to 0. \tag{2.29}$$

Substituting this result into final formula for $\hat{x}^{RS}_{k|k}$ given in Eqn. 2.22, we obtain

$$\hat{x}^{IRS}_{k|k} \rightarrow \left[\sum_{j=1}^{N} \mu^j_k \sqrt{|\theta Q_k|}\theta Q_k\right]^{-1} \sum_{j=1}^{N} \mu^j_k \sqrt{|\theta Q_k|}\theta Q_k \hat{x}^j_{k|k} \tag{2.30}$$

$$= \sum_{j=1}^{N} \mu^j_k \hat{x}^j_{k|k} \triangleq \hat{x}^{IMM}_{k|k} \tag{2.31}$$

when $\theta \rightarrow 0$. Therefore, the risk-sensitive output estimate goes to the IMM estimate when the risk-sensitive parameter θ goes to zero in spite of the Taylor series approximation made in the derivation.

Note that the derivation of the risk-sensitive output estimate calculation formula requires the positive definiteness of M^j_k and S^j_k, which are equivalent conditions since $M^j_k = \Sigma^j_{k|k} S^j_k \frac{1}{\theta} Q^{-1}_k$ and the matrices $\Sigma^j_{k|k}$, Q_k are positive definite. The definition of S^j_k in Eqn. 2.16 shows that the matrix S^j_k will be positive definite if and only if the matrix θQ_k is sufficiently "small", which is satisfied if the risk-sensitive parameter θ is sufficiently small. As a result, like other risk-sensitive filters in the literature, the result of the output estimate calculation formula (state estimate of IRS-IMM) converges to the MMSE output estimate (state estimate of IMM) if the risk-sensitive parameter θ is selected sufficiently small.

2.4 Simulation Results

In this section, the performance of the IRS-IMM algorithm will be observed and compared to that of the IMM algorithm. For this purpose, we consider a simplified example of a moving target whose acceleration evolves according to a finite-state Markov chain. This example is a slightly modified version of the one given in [1]. Only the target dynamics in one-dimension, which is given as

$$\underbrace{\begin{bmatrix} p_k \\ v_k \end{bmatrix}}_{x_k} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{k-1} \\ v_{k-1} \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} [a_k + w_k], \tag{2.32}$$

will be considered. Here, p_k, v_k and a_k denote the target position, velocity and acceleration respectively. The initial state x_0 is normally distributed with

mean $\bar{x}_0$ and covariance Σ_0 which are given as

$$\bar{x}_0 = \begin{bmatrix} 80000 \\ 400 \end{bmatrix}, \qquad \Sigma_0 = \begin{bmatrix} 10000 & 1000 \\ 1000 & 10000 \end{bmatrix}. \tag{2.33}$$

The acceleration process a_k is a finite-state Markov chain with states in the set $\{0, 20, -20\}$. The transition probability matrix for the finite-state Markov chain is

$$\Pi = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} \tag{2.34}$$

which corresponds to a highly maneuvering target. The white process noise $w_k \sim \mathcal{N}(w_k; 0, 2^2)$ represents small acceleration changes. It is assumed that only the positions are measured, i.e.,

$$y_k = p_k + \nu_k \tag{2.35}$$

where the terms $\nu_k \sim \mathcal{N}(w_k; 0, 100^2)$ stands for the normally distributed white measurement noise. The sampling period T is taken to be 10 sec's.

The IMM and IRS-IMM algorithms are run on the artificially generated measurements of the system defined above for 1000 Monte-Carlo runs with $\theta = 8 \times 10^{-5}$ and Q_k selected as

$$Q_k = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \quad \text{for all} \quad k. \tag{2.36}$$

The algorithms are assumed not to know the true probability transition matrix given in Eqn. 2.34 and use the diagonally dominant probability transition matrix Π_{IMM} given as

$$\Pi_{IMM} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}. \tag{2.37}$$

This type of matrices is generally used in practice in the cases where the true transition matrix is not known [13]. These are actually almost all cases in real life problems.

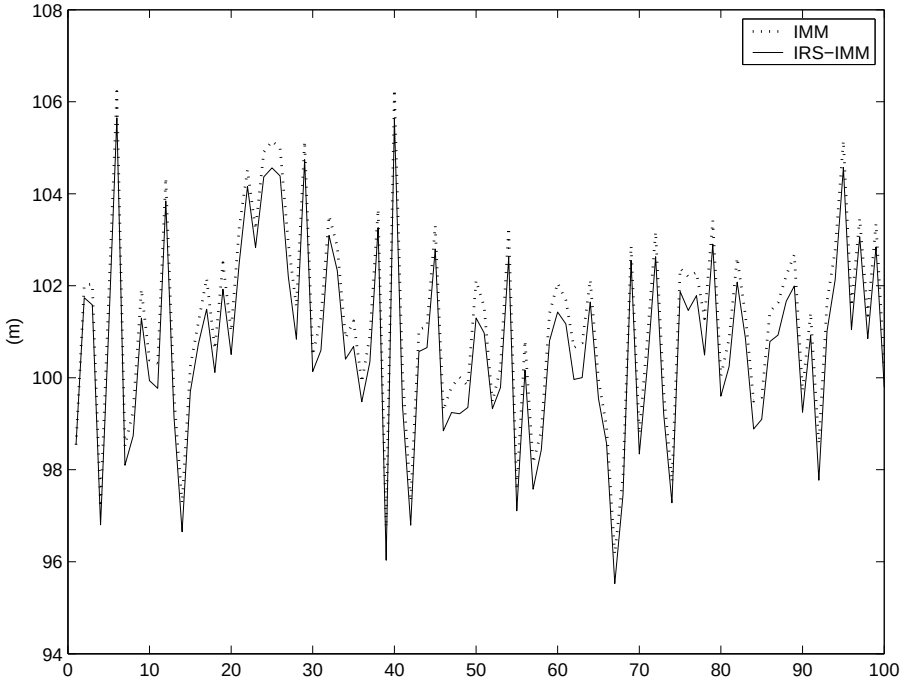

Figure 2.1: RMS position errors of the IMM and IRS-IMM algorithms

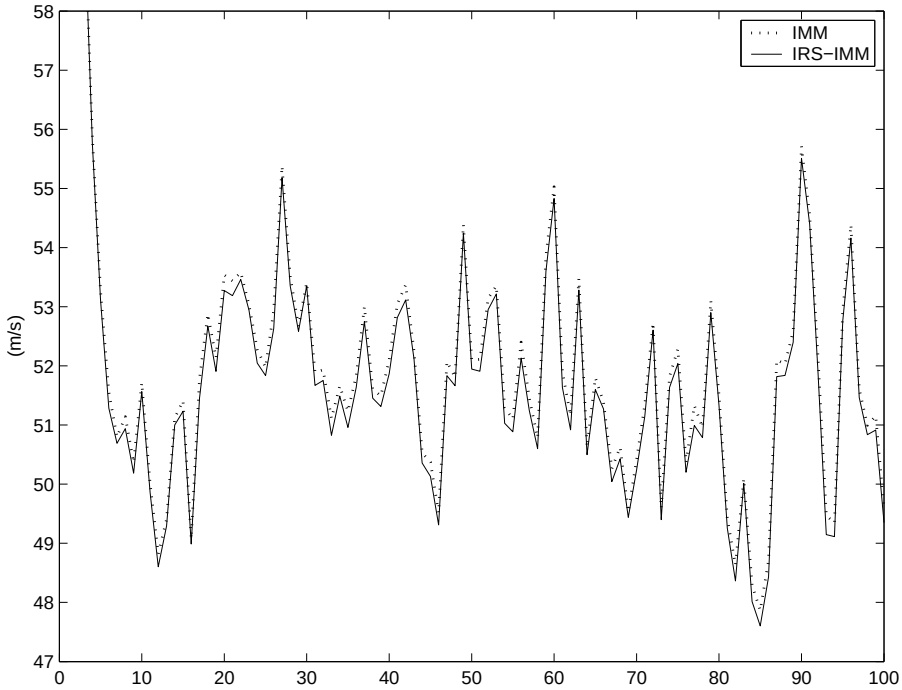

Figure 2.2: RMS velocity errors of the IMM and IRS-IMM algorithms

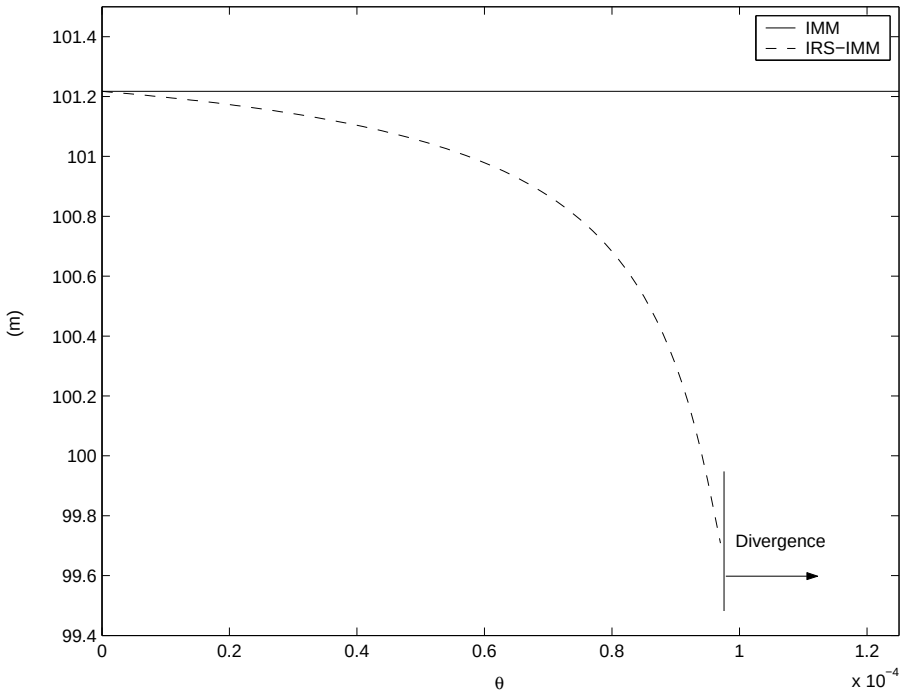

Figure 2.3: Average RMS position errors (per sample) of the IMM and IRS-IMM algorithms for different θ values.

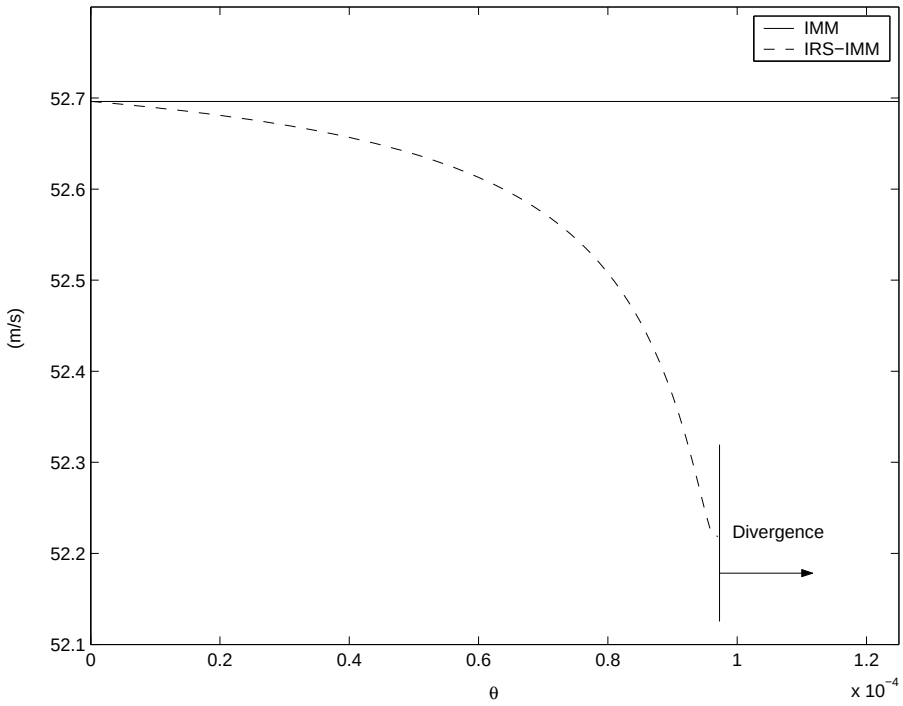

Figure 2.4: Average RMS velocity errors (per sample) of the IMM and IRS-IMM algorithms for different θ values.

In Fig. 2.1 and Fig. 2.2, the RMS position and velocity errors of the algorithms are presented. As easily realized from the figures, the RMS errors of the IRS-IMM algorithm are slightly lower than the standard IMM in this unknown parameter scenario. These results show that the risk-sensitive output calculation mechanism can yield better results in uncertain environments where modeling cannot be done accurately.

The average RMS position and velocity errors per measurement sample of the algorithms for different θ values are plotted in Fig. 2.3 and Fig. 2.4 respectively. The figures show that the error reduction (compared to IMM) obtained by new output calculation algorithm increases (with increasing θ) with growing rate until the algorithm diverges at around $\theta = 1 \times 10^{-4}$. The figures also make it clear that much better performance than those shown in Figures 2.1 and 2.2 can be obtained from the algorithm if better θ selection mechanisms are used.

2.5 Conclusion

In this chapter, a multiple model risk-sensitive estimation algorithm which minimizes the expected exponential of the instantaneous quadratic estimation error has been derived. The algorithm uses the IMM filter's statistics and differs from the IMM filter only in the output estimate calculation step. The estimate calculation requires not only mode probabilities and the mode-conditioned state estimates (like the case in IMM) but also the mode conditioned covariances of the individual Kalman filters. The estimate of the algorithm has been shown to reduce to the estimate of the standard IMM algorithm when the risk-sensitive parameter θ goes to zero. The algorithm has achieved better performance than the IMM filter in a scenario with model uncertainties.

CHAPTER 3

RISK-SENSITIVE MULTIPLE-MODEL STATE ESTIMATION: CUMULATIVE CASE

3.1 Introduction

Due to the (over-)simplicity and limitedness of instantaneous cost function examined in Chapter 2, the instances of the cumulative cost function dominates the risk-sensitive filtering and control literature. Although, due to the cumulative characteristics of the cost function that makes the analysis and synthesis more involved, there are many methods to derive cumulative risk-sensitive filters based on the cumulative cost function. The first cumulative risk-sensitive filter for linear Gauss-Markov systems is derived using dynamic programming [26]. The reference probability methods[1] [38, 39] in which a new probability measure is defined and exploited have been used in [40, 41] to derive filters and smoothers for nonlinear systems.[2] Furthermore, using the derivation procedures based on game theory and $\mathcal{H}_\infty$ filtering theory with which the risk sensitive estimation has been shown to be related [26, 40] is also theoretically possible. Risk-sensitive filters can be obtained even by applying Kalman filter in an indefinite metric (Krein) space [42].

In this chapter, we consider the cumulative risk-sensitive estimation prob-

[1] A comprehensive tutorial about the reference probability method is given in App. A.

[2] The risk-sensitive filter for linear Gauss-Markov systems is also derived using the reference probability method in App. A.

lem for jump Markov linear systems using the reference probability methods. Like the case in the conditional-mean estimators (or equivalently in minimum mean square error (MMSE) estimators), optimal risk-sensitive multiple model filter turns out to be impossible to implement with exponentially growing memory requirements. This problem is solved here using IMM-type approximations which are identified by making a derivation of the IMM filter using the reference probability method. The reference probability method used in the derivation provides a perfectly transparent framework for this identification by making the process of obtaining recursive expectations possible. Notice that the reference probability method has a crucial importance in this chapter. Therefore, if the reader is not familiar with this method, he/she should consult App. A which includes the background theory along with the derivation of Kalman and risk-sensitive filters for linear Gauss-Markov systems using the reference probability method.

The organization of the chapter is as follows. In Sec. 3.2, we make a problem definition and define the required probability measures for the reference probability method. Sec. 3.3 makes a brief derivation of the IMM filter using the reference probability method and identifies the approximations made in the IMM filter, which are stated in normalized probability domains in the literature, in the reference probability domain. The main results of the chapter are given in Sec. 3.4 where the derivation of the cumulative risk-sensitive filtering algorithm (which is called as CRS-IMM algorithm) is completed. In Sec. 3.5, the properties of the resulting filter is stated along with some implementation issues. The performance of the CRS-IMM algorithm is illustrated on a simulated unknown parameter scenario in Sec. 3.6. The chapter is finalized with conclusions in Sec. 3.7.

3.2 Problem Definition

The following jump Markov linear system model is considered

$$x_{k+1} = A(r_{k+1})x_k + B(r_{k+1})w_{k+1}, \tag{3.1}$$

$$y_k \;=\; C(r_k)x_k + D(r_k)v_k \tag{3.2}$$

where

- $\{x_k \in \mathbb{R}^n\}$ is the continuous-valued base-state sequence with initial distribution

$$x_0 \sim \mathcal{N}(x_0; \bar{x}_0, \Sigma_0), \tag{3.3}$$

  where the notation $\mathcal{N}(x; \bar{x}, \Sigma)$ stands for a Gaussian probability density function for dummy variable x which has a mean $\bar{x}$ and covariance Σ. We assume $\Sigma_0 > 0$.

- $\{r_k\}$ is the unknown discrete-valued modal-state sequence,

- $\{y_k \in \mathbb{R}^m\}$ is the noisy observation sequence,

- $\{w_k \in \mathbb{R}^n\}$ is a white process noise sequence with distribution,

$$w_k \sim \mathcal{N}(w_k; 0, I_n), \tag{3.4}$$

  where I_n denotes the identity matrix of size $n \times n$.

- $\{v_k \in \mathbb{R}^m\}$ is a white measurement noise sequence independent from the process noise w_k with distribution

$$v_k \sim \mathcal{N}(v_k; 0, I_m). \tag{3.5}$$

The discrete-valued modal-state $r_k \in \{e_1, e_2, \ldots, e_N\}$ is assumed to be a first-order finite-state homogenous Markov chain with transition probability matrix $\Pi = [\pi_{ij}]$. Here the variable $e_j \in \mathbb{R}^N$ denotes the canonical unit vector with unity at the jth position and zeros elsewhere. The basic variables w_k, v_k, x_0 and the modal-state sequence r_k are assumed to be mutually independent for all k. The time-varying matrices $A(r_k)$, $B(r_k)$, $C(r_k)$, and $D(r_k)$ are assumed to be known for each value of r_k. Moreover, the matrices $B(r_k)$ and $D(r_k)$ are assumed to be invertible. This is a requirement of the derivation using the reference probability method.[3]

[3] We will elaborate on bypassing this restriction using limiting arguments in Sec. 3.5.

Our aim is to calculate the recursive cumulative risk-sensitive estimate called as $\hat{x}_{k|k}^{CRS}$ given as

$$\hat{x}_{k|k}^{CRS} = \arg\min_{\zeta \in \mathbb{R}^n} E[\exp\{\theta \Psi_{0,k}(\zeta)\} | \mathcal{Y}_k] \tag{3.6}$$

where $\mathcal{Y}_k$ denotes the complete filtration generated by $\{y_0, y_1, \ldots, y_k\}$ and

$$\Psi_{0,k}(\zeta) \triangleq \hat{\Psi}_{0,k-1} + \frac{1}{2}(x_k - \zeta)^T Q_k (x_k - \zeta), \tag{3.7}$$

$$\hat{\Psi}_{m,n} \triangleq \sum_{l=m}^{n} \frac{1}{2}(x_l - \hat{x}_{l|l}^{CRS})^T Q_l (x_l - \hat{x}_{l|l}^{CRS}). \tag{3.8}$$

The matrices Q_k are known, positive definite weighting matrices and the real number $\theta > 0$ is called as the risk-sensitive parameter like the case in Chapter 2. Note that the structure of the cost function implies a filtering (and not smoothing) framework in that the estimates $\hat{x}_{l|l}^{CRS}$ for $l = 0, \ldots, k-1$ have been already calculated and will not change at time k.

3.2.1 Change of Measure

In the derivations given in the subsequent sections, it is initially assumed that we are in an ideal probability space $(\Omega, \mathcal{F}, \overline{P})$ where, under the probability measure $\overline{P}$

- $\{x_k\}$, $k \in \mathbb{N}$ is a sequence of independent, identically distributed (i.i.d.) random variables which are Gaussian distributed with zero mean and covariance I_n. Call their density function as $\phi(x) = \mathcal{N}(x; 0, I_n)$.

- $\{y_k\}$, $k \in \mathbb{N}$ is a sequence of i.i.d. random variables which are Gaussian distributed with zero mean and covariance I_m. Call their density function as $\psi(x) = \mathcal{N}(x; 0, I_m)$.

- $\{r_k\}$, $k \in \mathbb{N}$ is a first-order finite-state homogenous Markov chain with transition probability matrix $\Pi = [\pi_{ij}]$ and initial distribution

$$\pi_0 = [\ \pi_0^1 \quad \pi_0^2 \quad \cdots \quad \pi_0^N\ ]. \tag{3.9}$$

We define the sequence of random variables $\{\overline{\lambda}_l\}$ and $\{\overline{\Lambda}_k\}$, $k, l \in \mathbb{N}$ as

$$\begin{aligned}
\overline{\lambda}_l &= \begin{cases} \frac{\phi(\sqrt{\Sigma_0}^{-1}(x_l - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_l)} \frac{\psi(D^{-1}(r_l)(y_l - C(r_l)x_l))}{|D(r_l)|\psi(y_l)}, & l = 0 \\ \frac{\phi(B^{-1}(r_l)(x_l - A(r_l)x_{l-1}))}{|B(r_l)|\phi(x_l)} \frac{\psi(D^{-1}(r_l)(y_l - C(r_l)x_l))}{|D(r_l)|\psi(y_l)}, & l > 0 \end{cases} \\
\overline{\Lambda}_k &= \prod_{l=0}^{k} \overline{\lambda}_l
\end{aligned} \tag{3.10}$$

where $|.|$ denotes the matrix determinant and $\sqrt{\Sigma_0}$ is the positive definite square root of Σ_0. Let $\mathcal{G}_k$ denote the complete filtration generated by random variables $\{x_0, \ldots, x_k, r_0, \ldots, r_k, y_0, \ldots, y_k\}$. Now, if we define a new probability measure P by setting the restriction of the Radon-Nikodym derivative $\frac{dP}{d\overline{P}}|_{\mathcal{G}_k}$ to $\overline{\Lambda}_k$, then, under the new probability measure P, $\{w_k \in \mathbb{R}^n\}$ and $\{v_k \in \mathbb{R}^m\}$, $k \in \mathbb{N}$ defined as

$$w_k \triangleq B^{-1}(r_k)(x_k - A(r_k)x_{k-1}), \tag{3.11}$$

$$v_k \triangleq D^{-1}(r_k)(y_k - C(r_k)x_k) \tag{3.12}$$

are sequences of i.i.d. Gaussian random variables with zero-mean and covariance I_n and I_m respectively.[4] Moreover, the distribution of $\{r_k\}$ remains unchanged.[5] Note that, under both P and $\overline{P}$, the modal state $\{r_k\}$, $k \in \mathbb{N}$ has a semi-martingale representation

$$r_{k+1} = \Pi^T r_k + m_{k+1} \tag{3.13}$$

where m_k is a $\mathcal{G}_k$ martingale increment. The probability measure P is the nominal measure under which the expectations like the one in Eqn. 3.6 are taken. The expectations under the reference probability measure $\overline{P}$, which are shown by $\overline{E}$, can be taken much more easily than the ones under P thanks to the independence properties. By Theorem A.3, the two expectations can be related in the context of Eqn. 3.6 as

$$E[\exp\{\theta\Psi_{0,k}(\zeta)\}|\mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k \exp\{\theta\Psi_{0,k}(\zeta)\}|\mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k|\mathcal{Y}_k]}. \tag{3.14}$$

[4] The proof of this fact follows the same lines as the proof presented for linear Gauss-Markov systems in Lemma A.7.

[5] Proof of this fact is very similar to those given in [39] and [38].

Since the denominator of the right hand side in Eqn. 3.14 is independent of ζ, the cumulative risk-sensitive estimate $\hat{x}_{k|k}^{CRS}$ is alternatively given as

$$\hat{x}_{k|k}^{CRS} = \arg\min_{\zeta \in \mathbb{R}^n} \overline{E}[\overline{\Lambda}_k \exp\{\theta \Psi_{0,k}(\zeta)\} \,|\mathcal{Y}_k]. \tag{3.15}$$

3.3 IMM Filter Derivation and Approximation Identification

The aim of this section is twofold. First, it prepares the reader for the quantities and methodology involved in the derivations of the main result of the chapter given in Sec. 3.4. Secondly, it identifies the approximations made by the IMM algorithm under the nominal probability measure, in the reference probability domain. This identification then leads us to introduce the same type of approximations in the derivation of the cumulative risk-sensitive filter to avoid ever-growing memory and computation requirements of the optimal solution.

The usage of reference probability methods in the derivation of IMM-type filters is not new. For example, in [43], a hybrid filter is derived for the case where not only the process $\{x_k\}$ but also the modulating Markov chain $\{r_k\}$ is (directly) measured. Also, the case where only the (direct) noisy measurements of the Markov chain $\{r_k\}$ are available is considered in [44] which yields a finite dimensional filter. In none of these references, is what the approximations of the IMM filter correspond to in the reference probability domain examined. Being able to make these approximations in the reference probability domain would give one the ability to derive novel hybrid filters in the areas where the reference probability method is applicable.

In IMM filtering, the aim is to calculate the MMSE recursive estimate $\hat{x}_{k|k}^{MS}$ defined as

$$\hat{x}_{k|k}^{MS} \triangleq E\left[x_k | \mathcal{Y}_k\right]. \tag{3.16}$$

Using Theorem A.3, we can write

$$E\left[x_k|\mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k x_k|\mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k|\mathcal{Y}_k\right]}. \tag{3.17}$$

Therefore, instead of calculating $E\left[x_k|\mathcal{Y}_k\right]$, one can calculate the unnormalized expectation $\overline{E}\left[\overline{\Lambda}_k x_k|\mathcal{Y}_k\right]$ which is defined using the probability measure $\overline{P}$ under which $\{x_k\}$ and $\{y_k\}$ are i.i.d. sequences. The results of the two procedures would be equivalent after suitably normalizing the unnormalized expectation $\overline{E}\left[\overline{\Lambda}_k x_k|\mathcal{Y}_k\right]$. At this point, we define the unnormalized density functions $\alpha_k^j(x)$ as

$$\alpha_k^j(x)dx = \overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle \mathcal{I}_{\{x_k\in dx\}}\Big|\mathcal{Y}_k\right] \tag{3.18}$$

where the function $\mathcal{I}_A(\omega)$ defined as

$$\mathcal{I}_A(\omega) \triangleq \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \tag{3.19}$$

denotes the indicator function of the set A and the notation $\langle r_k, e_j\rangle$ stands for the inner product $r_k^T e_j$ which is equal to the jth element of r_k.

Using a simple reasoning, we can see that, if $f : \mathbb{R}^n \to \mathbb{R}$ is any test function (i.e., measurable function with compact support), the following equality is satisfied.

$$\overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle f(x_k)\Big|\mathcal{Y}_k\right] = \int f(x)\alpha_k^j(x)dx. \tag{3.20}$$

Using this and the fact that

$$\sum_{j=1}^{N}\langle r_k, e_j\rangle = 1, \tag{3.21}$$

we can write the unnormalized estimate $\overline{E}\left[\overline{\Lambda}_k x_k|\mathcal{Y}_k\right]$ as

$$\begin{aligned} \overline{E}\left[\overline{\Lambda}_k x_k|\mathcal{Y}_k\right] &= \sum_{j=1}^{N}\overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle x_k|\mathcal{Y}_k\right] &(3.22)\\ &= \sum_{j=1}^{N}\int x\alpha_k^j(x)dx = \int x\sum_{j=1}^{N}\alpha_k^j(x)dx. &(3.23)\end{aligned}$$

The same reasoning as in Eqn. 3.23 yields $\overline{E}[\overline{\Lambda}_k|\mathcal{Y}_k]$ as

$$\overline{E}[\overline{\Lambda}_k|\mathcal{Y}_k] = \sum_{j=1}^{N} \overline{E}[\overline{\Lambda}_k\langle r_k, e_j\rangle|\mathcal{Y}_k] = \sum_{j=1}^{N} \int \alpha_k^j(\xi)d\xi. \tag{3.24}$$

Combining the results of Eqn. 3.23 and Eqn. 3.24,

$$E[x_k|\mathcal{Y}_k] = \sum_{j=1}^{N} \int x\beta_k^j(x)dx \tag{3.25}$$

where

$$\beta_k^j(x) = \frac{\alpha_k^j(x)}{\sum_{l=1}^{N}\int \alpha_k^l(\xi)d\xi}. \tag{3.26}$$

Therefore, the set of density functions $\{\alpha_k^j(.)\}_{j=1}^N$ can be interpreted as an "information state" for the problem [45].

3.3.1 Recursion

The following theorem gives a recursion for the density $\alpha_k^j(x)$.

Theorem 3.1 *The density functions $\alpha_k^j(x)$, $k \geq 1$ satisfy the following recursion.*

$$\alpha_k^j(x) = \frac{\psi(D_j^{-1}(y_k - C_j x))}{|B_j||D_j|\psi(y_k)} \int \phi(B_j^{-1}(x - A_j z)) \sum_{i=1}^{N} \pi_{ij}\alpha_{k-1}^i(z)dz \tag{3.27}$$

where $A_j \triangleq A(e_j)$, $B_j \triangleq B(e_j)$, $C_j \triangleq C(e_j)$, and $D_j \triangleq D(e_j)$ for $j = 1, \ldots, N$.

Proof Let $g : \mathbb{R}^n \to \mathbb{R}$ be any test function. Then,

$$\begin{aligned}
\int g(x)\alpha_k^j(x)dx &= \overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle g(x_k)|\mathcal{Y}_k\right] && (3.28)\\
&= \overline{E}\left[\overline{\Lambda}_{k-1}\overline{\lambda}_k\langle r_k, e_j\rangle g(x_k)|\mathcal{Y}_k\right] && (3.29)\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B^{-1}(r_k)(x_k - A(r_k)x_{k-1}))}{|B(r_k)|\phi(x_k)} \\
&\quad \times\frac{\psi(D^{-1}(r_k)(y_k - C(r_k)x_k))}{|D(r_k)|\psi(y_k)}\langle r_k, e_j\rangle g(x_k)\Big|\mathcal{Y}_k\Big] && (3.30)\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\
&\quad \times\langle r_k, e_j\rangle g(x_k)\Big|\mathcal{Y}_k\Big] && (3.31)
\end{aligned}$$

$$
\begin{aligned}
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\
&\quad \times\langle \Pi^T r_{k-1} + m_k, e_j\rangle g(x_k)\Big|\mathcal{Y}_k\Big] \qquad (3.32)\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\
&\quad \times\langle \Pi^T r_{k-1}, e_j\rangle g(x_k)\Big|\mathcal{Y}_k\Big] \\
&+ \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\
&\quad \times\langle m_k, e_j\rangle g(x_k)\Big|\mathcal{Y}_k\Big]. \qquad (3.33)
\end{aligned}
$$

The second expectation on the right hand side of Eqn. 3.33 is zero due to facts that m_k is a $\mathcal{G}_k$-martingale increment and that under the probability measure $\overline{P}$, the process $\{r_k\}$, and hence the process $\{m_k\}$, is independent of the processes $\{x_k\}$ and $\{y_k\}$. Now using the identity

$$
\langle \Pi^T r_{k-1}, e_j\rangle = \sum_{i=1}^{N}\pi_{ij}\langle r_{k-1}, e_i\rangle, \qquad (3.34)
$$

Eqn 3.33 becomes

$$
\begin{aligned}
\int g(x)\alpha_k^j(x)dx &= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \\
&\quad \times\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)}\langle r_{k-1}, e_i\rangle g(x_k)\Big|\mathcal{Y}_k\Big] \qquad (3.35)\\
&= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1}, e_i\rangle\overline{E}\Big[\frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \\
&\quad \times\frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)}g(x)\Big|x_{k-1}, \mathcal{Y}_k\Big]\Big|\mathcal{Y}_k\Big] \qquad (3.36)
\end{aligned}
$$

The inner expectation in Eqn. 3.36 can easily be taken as follows due to the independence properties of the sequence $\{x_k\}$ under $\overline{P}$.

$$
\begin{aligned}
\int g(x)\alpha_k^j(x)dx &= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1}, e_i\rangle\int\frac{\phi(B_j^{-1}(x - A_j x_{k-1}))}{|B_j|\phi(x)} \\
&\quad \times\frac{\psi(D_j^{-1}(y_k - C_j x))}{|D_j|\psi(y_k)}g(x)\phi(x)dx\Big|\mathcal{Y}_k\Big] \qquad (3.37)\\
&= \sum_{i=1}^{N}\pi_{ij}\int\int\frac{\phi(B_j^{-1}(x - A_j z))}{|B_j|}\frac{\psi(D_j^{-1}(y_k - C_j x))}{|D_j|\psi(y_k)}
\end{aligned}
$$

$$\times g(x)dx\alpha_{k-1}^{i}(z)dz \tag{3.38}$$

$$= \int g(x)\Big[\frac{\psi(D_j^{-1}(y_k - C_j x))}{|B_j||D_j|\psi(y_k)} \int \phi(B_j^{-1}(x - A_j z)) \times \sum_{i=1}^{N} \pi_{ij}\alpha_{k-1}^{i}(z)dz\Big]dx. \tag{3.39}$$

Since this equality is satisfied for all test functions $g(.)$, the recursion in Eqn. 3.27 is satisfied. □

Note that this result is different than (a special case of) the recursion given in [43]. This is because [43] takes the system dynamics equation as

$$x_{k+1} = A(r_k)x_k + B(r_k)w_{k+1}, \tag{3.40}$$

which is different than Eqn. 3.1, and measurement equation as in Eqn. 3.2. This, at the end, causes a slight change in the information pattern and yields a different recursion. Our selection follows the convention of the IMM filter derivation in [13] where r_k denotes the model (mode) in effect during the sampling period ending at time k.

3.3.2 Initial Densities

Let $f : \mathbb{R}^n \to \mathbb{R}$ be any test function. Then, the initial densities $\alpha_0^j(x)$ can be calculated as follows.

$$\int f(x)\alpha_0^j(x)dx = \overline{E}\Big[\overline{\Lambda}_0\langle r_0, e_j\rangle f(x_0)|\mathcal{Y}_0\Big] \tag{3.41}$$

$$= \overline{E}\Big[\overline{\lambda}_0\langle r_0, e_j\rangle f(x_0)|\mathcal{Y}_0\Big] \tag{3.42}$$

$$= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0 - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)} \frac{\psi(D^{-1}(r_0)(y_0 - C(r_0)x_0))}{|D(r_0)|\psi(y_0)} \times f(x_0)\langle r_0, e_j\rangle\Big|\mathcal{Y}_0\Big] \tag{3.43}$$

$$= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0 - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)} \frac{\psi(D_j^{-1}(y_0 - C_j x_0))}{|D_j|\psi(y_0)}\langle r_0, e_j\rangle \times f(x_0)\Big|\mathcal{Y}_0\Big]$$

$$= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0 - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)} \frac{\psi(D_j^{-1}(y_0 - C_j x_0))}{|D_j|\psi(y_0)}$$

$$\times f(x_0)\Big|\mathcal{Y}_0\Big]\overline{E}\left[\langle r_0, e_j\rangle\right] \tag{3.44}$$

$$= \int f(x)\pi_0^j \frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x)} \frac{\psi(D_j^{-1}(y_0 - C_j x))}{|D_j|\psi(y_0)} \times \phi(x)dx \tag{3.45}$$

$$= \int f(x)\pi_0^j \frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|} \frac{\psi(D_j^{-1}(y_0 - C_j x))}{|D_j|\psi(y_0)} dx.$$

Since the equality holds for all test function $f(.)$, the initial density $\alpha_0^j(x)$ is given as

$$\alpha_0^j(x) = \pi_0^j \frac{\psi(D_j^{-1}(y_0 - C_j x))}{|D_j|\psi(y_0)} \frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|} \tag{3.46}$$

$$= \frac{\pi_0^j}{\psi(y_0)}\mathcal{N}(y_0; C_j x, D_j D_j^T)\mathcal{N}(x; \bar{x}_0, \Sigma_0) \tag{3.47}$$

$$= \frac{\pi_0^j}{\psi(y_0)}\mathcal{N}(y_0; C_j \bar{x}_0, S_0^j)\mathcal{N}(x; \hat{x}_{0|0}^j, \Sigma_{0|0}^j) \tag{3.48}$$

$$= \bar{c}_0^j \mathcal{N}(x; \hat{x}_{0|0}^j, \Sigma_{0|0}^j). \tag{3.49}$$

where

$$S_0^j \triangleq C_j \Sigma_0 C_j^T + D_j D_j^T, \tag{3.50}$$

$$\hat{x}_{0|0}^j \triangleq \bar{x}_0 + \Sigma_0 C_j^T (S_0^j)^{-1}(y_0 - C_j \bar{x}_0), \tag{3.51}$$

$$\Sigma_{0|0}^j \triangleq \Sigma_0 - \Sigma_0 C_j^T (S_0^j)^{-1} C_j \Sigma_0, \tag{3.52}$$

$$\bar{c}_0^j = \frac{\pi_0^j}{\psi(y_0)}\mathcal{N}(y_0; C_j \bar{x}_0, S_0^j). \tag{3.53}$$

Here, while going from Eqn. 3.47 to Eqn. 3.48, we used the result of Appendix B.2.

3.3.3 Approximation

By Eqn. 3.49, the densities $\alpha_0^j(x)$ are of the form given as

$$\alpha_0^j(x) = \bar{c}_0^j \mathcal{N}(x; \hat{x}_{0|0}^j, \Sigma_{0|0}^j). \tag{3.54}$$

Considering this and the recursion in Eqn. 3.27, we see that the density $\alpha_k^j(x)$ must be a sum of N^k unnormalized Gaussian densities. Therefore, the number of statistics to be kept increases exponentially. At this point, the IMM

approximation mechanism comes into picture. IMM algorithm, at each time step k, keeps a single Gaussian for the normalized densities $p(x_k|r_k = e_j, \mathcal{Y}_k)$. It achieves this by approximating, at each time step, the normalized density $p(x_{k-1}|r_k = e_j, \mathcal{Y}_{k-1})$ which is actually a Gaussian mixture with N components by a single Gaussian, i.e.,

$$\begin{aligned} p(x_{k-1}|r_k = e_j, \mathcal{Y}_{k-1}) &= \sum_{i=1}^{N} \underbrace{P(r_{k-1} = e_i|r_k = e_j, \mathcal{Y}_{k-1})}_{\triangleq \mu^{ij}_{k-1|k-1}} \\ &\quad \times \underbrace{p(x_{k-1}|r_{k-1} = e_i, r_k = e_j, \mathcal{Y}_{k-1})}_{=p(x_{k-1}|r_{k-1}=e_i, \mathcal{Y}_{k-1})} \qquad (3.55) \\ &= \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \mathcal{N}(x_{k-1}; \hat{x}^i_{k-1|k-1}, \Sigma^i_{k-1|k-1}) \qquad (3.56) \\ &\approx \mathcal{N}(x_{k-1}, \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1}) \qquad (3.57) \end{aligned}$$

where

$$\begin{aligned} \hat{x}^{0j}_{k-1|k-1} &\triangleq \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \hat{x}^i_{k-1|k-1}, \qquad (3.58) \\ \Sigma^{0j}_{k-1|k-1} &\triangleq \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \Big[P^i_{k-1|k-1} \\ &\quad + (\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})(\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})^T \Big]. \qquad (3.59) \end{aligned}$$

Lemma 3.1 *The normalized density $p(x_{k-1}|r_k = e_j, \mathcal{Y}_{k-1})$ of the IMM filter satisfies the following equality in terms of unnormalized densities $\{\alpha^i_{k-1}(x)\}_{i=1}^N$.*

$$p(x_{k-1}|r_k = e_j, \mathcal{Y}_{k-1}) = \frac{\sum_{i=1}^N \pi_{ij} \alpha^i_{k-1}(x_{k-1})}{\sum_{i=1}^N \pi_{ij} \int \alpha^i_{k-1}(\xi) d\xi}. \qquad (3.60)$$

Proof Note that from Bayes theorem

$$p(x_{k-1}|r_k = e_j, \mathcal{Y}_{k-1}) = \frac{p(x_{k-1}, r_k = e_j|\mathcal{Y}_{k-1})}{P(r_k = e_j|\mathcal{Y}_{k-1})}. \qquad (3.61)$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be any test function. Then,

$$\begin{aligned} \int f(x) p(x|r_k = e_j, \mathcal{Y}_{k-1}) dx &= \frac{\int f(x) p(x, r_k = e_j|\mathcal{Y}_{k-1}) dx}{P(r_k = e_j|\mathcal{Y}_{k-1})} \qquad (3.62) \\ &= \frac{E\left[f(x_{k-1}) \langle r_k, e_j \rangle | \mathcal{Y}_{k-1}\right]}{E\left[\langle r_k, e_j \rangle | \mathcal{Y}_{k-1}\right]} \qquad (3.63) \end{aligned}$$

$$= \frac{\overline{E}\left[\overline{\Lambda}_{k-1} f(x_{k-1})\langle r_k, e_j\rangle | \mathcal{Y}_{k-1}\right]}{\overline{E}\left[\overline{\Lambda}_{k-1}\langle r_k, e_j\rangle | \mathcal{Y}_{k-1}\right]} \tag{3.64}$$

$$= \frac{\overline{E}\left[\overline{\Lambda}_{k-1} f(x_{k-1})\langle \Pi^T r_{k-1}, e_j\rangle | \mathcal{Y}_{k-1}\right]}{\overline{E}\left[\overline{\Lambda}_{k-1}\langle \Pi^T r_{k-1}, e_j\rangle | \mathcal{Y}_{k-1}\right]}. \tag{3.65}$$

Using the fact that

$$\langle \Pi^T r_{k-1}, e_j\rangle = \sum_{i=1}^{N} \pi_{ij}\langle r_{k-1}, e_i\rangle, \tag{3.66}$$

Eqn. 3.65 will read,

$$\begin{aligned}
\int f(x) p(x|r_k = e_j, \mathcal{Y}_{k-1}) dx &= \frac{\sum_{i=1}^{N} \pi_{ij} \overline{E}\left[\overline{\Lambda}_{k-1} f(x_{k-1})\langle r_{k-1}, e_i\rangle | \mathcal{Y}_k\right]}{\sum_{i=1}^{N} \pi_{ij} \overline{E}\left[\overline{\Lambda}_{k-1}\langle r_{k-1}, e_i\rangle | \mathcal{Y}_{k-1}\right]} \\
&= \frac{\sum_{i=1}^{N} \pi_{ij} \int f(x)\alpha_{k-1}^i(x) dx}{\sum_{i=1}^{N} \pi_{ij} \int \alpha_{k-1}^i(\xi) d\xi} \qquad (3.67) \\
&= \frac{\int f(x) \sum_{i=1}^{N} \pi_{ij}\alpha_{k-1}^i(x) dx}{\sum_{i=1}^{N} \pi_{ij} \int \alpha_{k-1}^i(\xi) d\xi} \qquad (3.68) \\
&= \int f(x) \frac{\sum_{i=1}^{N} \pi_{ij}\alpha_{k-1}^i(x)}{\sum_{i=1}^{N} \pi_{ij} \int \alpha_{k-1}^i(\xi) d\xi} dx. \qquad (3.69)
\end{aligned}$$

Since the equality is satisfied for any test function $f(.)$, we conclude that Eqn. 3.60 is satisfied. □

Theorem 3.2 *Assuming that the unnormalized densities satisfy the equation*

$$\alpha_{k-1}^j(x) = \bar{c}_{k-1}^j \mathcal{N}(x; \hat{x}_{k-1|k-1}^j, \Sigma_{k-1|k-1}^j) \tag{3.70}$$

where

$$\bar{c}_{k-1}^j \triangleq \int \alpha_{k-1}^j(\xi) d\xi, \tag{3.71}$$

the normalized IMM approximation corresponds to making the approximation given as

$$\sum_{i=1}^{N} \pi_{ij}\alpha_{k-1}^j(x) \approx c_{k-1}^j \mathcal{N}(x; \hat{x}_{k-1|k-1}^{0j}, \Sigma_{k-1|k-1}^{0j}) \tag{3.72}$$

in the reference probability domain (i.e., in the recursion of Eqn. 3.27) where

$$c_{k-1}^j \triangleq \sum_{i=1}^{N} \pi_{ij}\bar{c}_{k-1}^i, \tag{3.73}$$

$$\hat{x}_{k-1|k-1}^{0j} \triangleq \sum_{i=1}^{N} \frac{\pi_{ij}\bar{c}_{k-1}^i}{c_{k-1}^j} \hat{x}_{k-1|k-1}^i, \tag{3.74}$$

$$\Sigma^{0j}_{k-1|k-1} \triangleq \sum_{i=1}^{N} \frac{\pi_{ij}\bar{c}^i_{k-1}}{c^j_{k-1}} \Big[P^i_{k-1|k-1} \tag{3.75}$$

$$+(\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})(\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})^T \Big]. \tag{3.76}$$

Proof In the light of Lemma 3.1, we have

$$p(x_{k-1}|r_k = e_j, \mathcal{Y}_k) = \frac{1}{c^j_{k-1}} \sum_{i=1}^{N} \pi_{ij}\alpha^i_{k-1}(x_{k-1}) \tag{3.77}$$

$$= \sum_{i=1}^{N} \frac{\pi_{ij}\bar{c}^i_{k-1}}{c^j_{k-1}} \mathcal{N}(x_{k-1}; \hat{x}^i_{k-1|k-1}, \Sigma^i_{k-1|k-1}) \tag{3.78}$$

$$\approx \mathcal{N}(x; \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1}). \tag{3.79}$$

The results of the theorem are now obvious considering the approximation in Eqn. 3.57. Note that, initially, $\alpha^j_0(x)$ satisfies the assumption in Eqn. 3.70. When we make the approximation in Eqn. 3.72, considering the recursion given in Eqn. 3.27, the density $\alpha^j_k(x)$ will always satisfy the assumption of Eqn. 3.70. Therefore, the assumption is not a restriction for the application of the theorem at any time instant k. □

Substituting the approximation into the recursion in Eqn. 3.27, we get

$$\alpha^j_k(x) = \frac{c^j_{k-1}\psi(D_j^{-1}(y_k - C_j x))}{|B_j||D_j|\psi(y_k)} \int \phi(B_j^{-1}(x - A_j z)) \times \mathcal{N}(z; \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1})dz \tag{3.80}$$

$$= \frac{c^j_{k-1}\psi(D_j^{-1}(y_k - C_j x))}{|D_j|\psi(y_k)} \int \frac{\phi(B_j^{-1}(x - A_j z))}{|B_j|} \times \mathcal{N}(z; \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1})dz \tag{3.81}$$

$$= \frac{c^j_{k-1}\psi(D_j^{-1}(y_k - C_j x))}{|D_j|\psi(y_k)} \int \mathcal{N}(x; A_j z, B_j B_j^T) \times \mathcal{N}(z; \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1})dz \tag{3.82}$$

$$= \frac{c^j_{k-1}\psi(D_j^{-1}(y_k - C_j x))}{|D_j|\psi(y_k)} \mathcal{N}(x; \hat{x}^j_{k|k-1}, \Sigma^j_{k|k-1}) \tag{3.83}$$

$$= \frac{c^j_{k-1}}{\psi(y_k)} \mathcal{N}(y_k; C_j x, D_j D_j^T) \mathcal{N}(x; \hat{x}^j_{k|k-1}, \Sigma^j_{k|k-1}) \tag{3.84}$$

where

$$\hat{x}^j_{k|k-1} = A_j \hat{x}^{0j}_{k-1|k-1}, \tag{3.85}$$

$$\Sigma^j_{k|k-1} = A_j \Sigma^{0j}_{k-1|k-1} A_j^T + B_j B_j^T. \tag{3.86}$$

Using the result on the multiplication of the Gaussian densities given in App. B.2, we get

$$\alpha_k^j(x) = \frac{c_{k-1}^j \Delta_j(y_k)}{\psi(y_k)} \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \tag{3.87}$$

where

$$\begin{aligned}
\Delta_j(y_k) &\triangleq \mathcal{N}(y_k; C\hat{x}_{k|k-1}^j, S_k^j), && (3.88)\\
S_k^j &= C_j \Sigma_{k|k-1}^j C_j^T + D_j D_j^T, && (3.89)\\
\Sigma_{k|k}^j &= \Sigma_{k|k-1}^j - \Sigma_{k|k-1}^j C_j^T (S_k^j)^{-1} C_j \Sigma_{k|k-1}^j, && (3.90)\\
\hat{x}_{k|k}^j &= \hat{x}_{k|k-1}^j + \Sigma_{k|k-1}^j C_j^T (S_k^j)^{-1} (y_k - C_j \hat{x}_{k|k-1}^j). && (3.91)
\end{aligned}$$

3.3.4 Final Estimate Calculation

The normalized densities $\beta_k^j(x)$ can now be easily calculated as

$$\beta_k^j(x) = \frac{\alpha_k^j(x)}{\sum_{l=1}^N \int \alpha_k^l(\xi) d\xi} = \mu_k^j \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \tag{3.92}$$

where

$$\mu_k^j = \frac{c_{k-1}^j \Delta_j(y_k)}{\sum_{l=1}^N c_{k-1}^l \Delta_l(y_k)}. \tag{3.93}$$

Then, the IMM estimate $\hat{x}_{k|k}^{MS}$ is given as

$$\hat{x}_{k|k}^{MS} = \sum_{j=1}^N \mu_k^j \hat{x}_{k|k}^j. \tag{3.94}$$

Remark 3.1 *It is important to note that, for the calculation of the mixed and the final IMM estimates, all that matters about the coefficients $\bar{c}_k^j$ (or c_k^j) is their relative magnitudes (and not their absolute magnitudes). Therefore, at any time-step k, one can multiply the the coefficients $\bar{c}_k^j$ (or c_k^j) by a common constant number without affecting the output estimate. This will be the case for the cumulative risk-sensitive multiple model filter derived in the next section.*

One cycle of the IMM algorithm whose required steps are derived above is illustrated in Fig. 3.1.

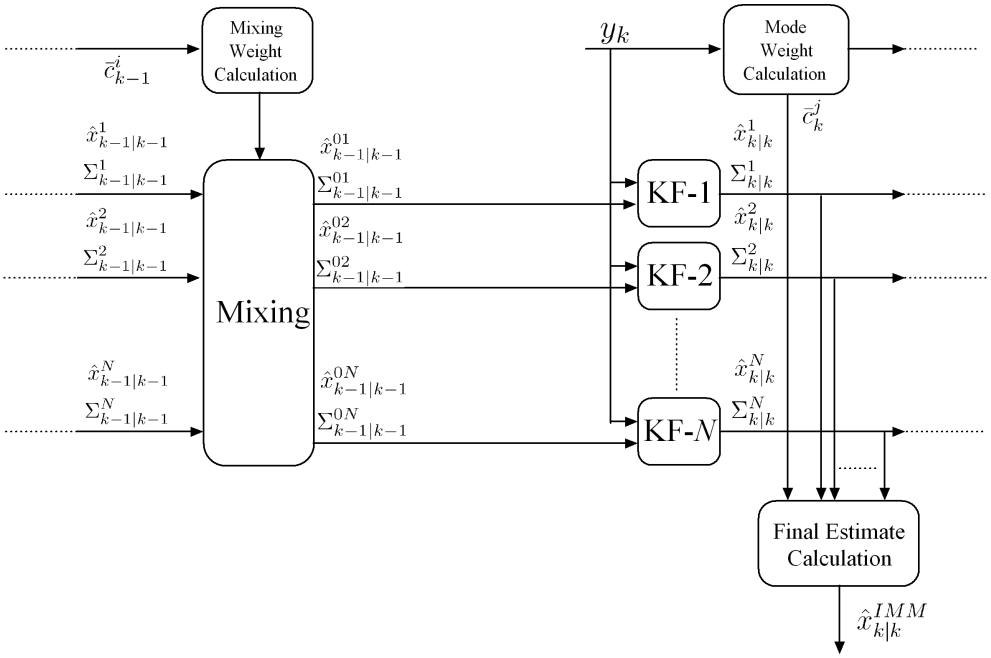

Figure 3.1: One cycle of IMM algorithm with N models. KF is the abbreviation of Kalman filter.

3.4 Cumulative Risk-Sensitive Multiple-Model Filter

In the following, we are going to make a similar derivation for the risk-sensitive multiple-model filter, which is called as CRS-IMM (where the abbreviation "CRS" stands for "cumulative risk-sensitive"), using the reference probability method. For this purpose, we define the unnormalized density function

$$\gamma^j_k(x)dx \triangleq \overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\}\mathcal{I}_{\{x_k\in dx\}}\middle|\mathcal{Y}_k\right]. \tag{3.95}$$

Using a simple argument, one can show that, if $f:\mathbb{R}^n\to\mathbb{R}$ is any test function,

$$\overline{E}\left[\overline{\Lambda}_k\langle r_k, e_j\rangle \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\} f(x_k)\middle|\mathcal{Y}_k\right] = \int f(x)\gamma^j_k(x)dx. \tag{3.96}$$

Notation: In this section, due to length of the formulas, we will use the following abbreviations.

$$\exp(+,x,\bar{x},\Sigma) \triangleq \exp\left\{\frac{1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})\right\}, \tag{3.97}$$

$$\exp(-,x,\bar{x},\Sigma) \triangleq \exp\left\{-\frac{1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})\right\}. \tag{3.98}$$

3.4.1 Recursion

The following theorem gives a recursion for the densities $\gamma_k^j(.)$.

Theorem 3.3 *The densities $\gamma_k^j(x)$, $k \geq 1$ satisfy the following recursion.*

$$\begin{aligned}\gamma_k^j(x) &= \frac{\psi(D_j^{-1}(y_k - C_j x))}{|B_j||D_j|\psi(y_k)} \int \phi(B_j^{-1}(x - A_j z)) \\ &\quad \times \exp\left(+, z, \hat{x}_{k-1|k-1}^{CRS}, \frac{1}{\theta} Q_{k-1}^{-1}\right) \sum_{i=1}^{N} \pi_{ij} \gamma_{k-1}^i(z) dz. \qquad (3.99)\end{aligned}$$

Proof Let $f : \mathbb{R}^n \to \mathbb{R}$ be any test function. Then,

$$\begin{aligned}\int f(x)\gamma_k^j(x)dx &= \overline{E}\left[\overline{\Lambda}_k \langle r_k, e_j\rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} f(x_k) \middle| \mathcal{Y}_k\right] \qquad (3.100) \\ &= \overline{E}\Big[\overline{\Lambda}_{k-1} \langle r_k, e_j\rangle f(x_k) \frac{\phi(B^{-1}(r_k)(x_k - A(r_k)x_{k-1}))}{|B(r_k)|\phi(x_k)} \\ &\quad \times \frac{\psi(D^{-1}(r_k)(y_k - C(r_k)x_k))}{|D(r_k)|\psi(y_k)} \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} \Big| \mathcal{Y}_k\Big] \\ &= \overline{E}\Big[\overline{\Lambda}_{k-1} \frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\ &\quad \times \langle r_k, e_j\rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} f(x_k) \Big| \mathcal{Y}_k\Big] \qquad (3.101) \\ &= \overline{E}\Big[\overline{\Lambda}_{k-1} \frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\ &\quad \times \langle \Pi^T r_{k-1} + m_k, e_j\rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} f(x_k) \Big| \mathcal{Y}_k\Big] \\ &= \overline{E}\Big[\overline{\Lambda}_{k-1} \frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\ &\quad \times \langle \Pi^T r_{k-1}, e_j\rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} f(x_k) \Big| \mathcal{Y}_k\Big] \\ &+ \overline{E}\Big[\overline{\Lambda}_{k-1} \frac{\phi(B_j^{-1}(x_k - A_j x_{k-1}))}{|B_j|\phi(x_k)} \frac{\psi(D_j^{-1}(y_k - C_j x_k))}{|D_j|\psi(y_k)} \\ &\quad \times \langle m_k, e_j\rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} f(x_k) \Big| \mathcal{Y}_k\Big]. \qquad (3.102)\end{aligned}$$

The second expectation in the summation in Eqn. 3.102 is zero due to facts that m_k is a $\mathcal{G}_k$-martingale increment and that under the probability measure $\overline{P}$, the process $\{r_k\}$, and hence the process $\{m_k\}$, is independent of the processes $\{x_k\}$ and $\{y_k\}$. Now using the identity

$$\langle \Pi^T r_{k-1}, e_j\rangle = \sum_{i=1}^{N} \pi_{ij} \langle r_{k-1}, e_i\rangle, \qquad (3.103)$$

Eqn 3.102 becomes,

$$\begin{aligned}
\int f(x)\gamma_k^j(x)dx &= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1},e_i\rangle\frac{\phi(B_j^{-1}(x_k-A_jx_{k-1}))}{|B_j|\phi(x_k)} \\
&\quad\times\frac{\psi(D_j^{-1}(y_k-C_jx_k))}{|D_j|\psi(y_k)}\exp\left\{\theta\hat{\Psi}_{0,k-1}\right\}f(x_k)\Big|\mathcal{Y}_k\Big] \\
&= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1},e_i\rangle\exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \\
&\quad\times\frac{\phi(B_j^{-1}(x_k-A_jx_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k-C_jx_k))}{|D_j|\psi(y_k)} \\
&\quad\times\exp\left(+,x_{k-1},\hat{x}_{k-1|k-1}^{CRS},\frac{1}{\theta}Q_{k-1}^{-1}\right)f(x_k)\Big|\mathcal{Y}_k\Big] \qquad (3.104) \\
&= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1},e_i\rangle\exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \\
&\quad\times\exp\left(+,x_{k-1},\hat{x}_{k-1|k-1}^{CRS},\frac{1}{\theta}Q_{k-1}^{-1}\right) \\
&\quad\times\overline{E}\Big[\frac{\phi(B_j^{-1}(x_k-A_jx_{k-1}))}{|B_j|\phi(x_k)}\frac{\psi(D_j^{-1}(y_k-C_jx_k))}{|D_j|\psi(y_k)} \\
&\quad\times f(x_k)\Big|x_{k-1},\mathcal{Y}_k\Big]\Big|\mathcal{Y}_k\Big] \qquad (3.105)
\end{aligned}$$

The inner expectation in Eqn. 3.105 can easily be taken as follows due to the independence properties of the sequence $\{x_k\}$ under $\overline{P}$.

$$\begin{aligned}
f(x)\gamma_k^j(x)dx &= \sum_{i=1}^{N}\pi_{ij}\overline{E}\Big[\overline{\Lambda}_{k-1}\langle r_{k-1},e_i\rangle\exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \\
&\quad\times\exp\left(+,x_{k-1},\hat{x}_{k-1|k-1}^{CRS},\frac{1}{\theta}Q_{k-1}^{-1}\right) \\
&\quad\times\int\frac{\phi(B_j^{-1}(x-A_jx_{k-1}))}{|B_j|\phi(x)}\frac{\psi(D_j^{-1}(y_k-C_jx_k))}{|D_j|\psi(y_k)}f(x) \\
&\quad\times\phi(x)dx\Big|\mathcal{Y}_{k-1}\Big] \qquad (3.106) \\
&= \sum_{i=1}^{N}\pi_{ij}\int\exp\left(+,z,\hat{x}_{k-1|k-1}^{CRS},\frac{1}{\theta}Q_{k-1}^{-1}\right) \\
&\quad\times\int\phi(B_j^{-1}(x-A_jz))\frac{\psi(D_j^{-1}(y_k-C_jx))}{|B_j||D_j|\psi(y_k)}f(x)dx \\
&\quad\times\gamma_{k-1}^i(z)dz \\
&= \int f(x)\frac{\psi(D_j^{-1}(y_k-C_jx))}{|B_j||D_j|\psi(y_k)}\int\phi(B_j^{-1}(x-A_jz))
\end{aligned}$$

$$\times \exp\left(+, z, \hat{x}^{CRS}_{k-1|k-1}, \frac{1}{\theta}Q^{-1}_{k-1}\right) \sum_{i=1}^{N} \pi_{ij}\gamma^i_{k-1}(z)dzdx.$$

Since this equality is satisfied for all test functions $f(.)$, the recursion in Eqn. 3.99 is satisfied. □

3.4.2 Initial Densities

The initial densities $\gamma^j_0(x)$ satisfy

$$\begin{aligned}
\int f(x)\gamma^j_0(x)dx &= \overline{E}\left[\overline{\Lambda}_0\langle r_0, e_j\rangle \exp\left\{\theta\hat{\Psi}_{0,-1}\right\} f(x_0)\middle|\mathcal{Y}_0\right] && (3.107)\\
&= \overline{E}\left[\overline{\Lambda}_0\langle r_0, e_j\rangle f(x_0)\middle|\mathcal{Y}_0\right] && (3.108)\\
&= \overline{E}\left[\overline{\lambda}_0\langle r_0, e_j\rangle f(x_0)\middle|\mathcal{Y}_0\right]. && (3.109)
\end{aligned}$$

The right hand side of Eqn. 3.109 is the same as that of Eqn. 3.42 giving the result below.

$$\gamma^j_0(x) = \bar{c}^j_0\mathcal{N}(x; \hat{x}^j_{0|0}, \Sigma^j_{0|0}) \tag{3.110}$$

where

$$\begin{aligned}
\hat{x}^j_{0|0} &\triangleq \bar{x}_0 + \Sigma_0 C_j^T (S^j_0)^{-1}(y_0 - C_j\bar{x}_0), && (3.111)\\
\Sigma^j_{0|0} &\triangleq \Sigma_0 - \Sigma_0 C_j^T (S^j_0)^{-1} C_j \Sigma_0, && (3.112)\\
\bar{c}^j_0 &= \frac{\pi^j_0}{\psi(y_0)}\mathcal{N}(y_0; C_j\bar{x}_0, S^j_0), && (3.113)\\
S^j_0 &\triangleq C_j\Sigma_0 C_j^T + D_j D_j^T. && (3.114)
\end{aligned}$$

3.4.3 Approximation

The form of initial density $\gamma^j_0(x)$ in Eqn. 3.110 and the recursion in Eqn. 3.99 implies that $\gamma^j_k(x)$ is the sum of N^k unnormalized Gaussian densities. Therefore, the number of statistics to be kept increases exponentially in the optimal filter. In order to have a filter which can be implemented with finite resources, we are going to use the same kind of approximation here as in the IMM filter. The approximation we make is therefore given as

$$\sum_{i=1}^{N} \pi_{ij}\gamma^j_{k-1}(x) \approx c^j_{k-1}\mathcal{N}(x; \hat{x}^{0j}_{k-1|k-1}, \Sigma^{0j}_{k-1|k-1}) \tag{3.115}$$

where

$$c_{k-1}^{j} \triangleq \sum_{i=1}^{N} \pi_{ij} \bar{c}_{k-1}^{i}, \tag{3.116}$$

$$\bar{c}_{k-1}^{j} \triangleq \int \gamma_{k-1}^{j}(\xi) d\xi, \tag{3.117}$$

$$\hat{x}_{k-1|k-1}^{0j} \triangleq \sum_{i=1}^{N} \frac{\pi_{ij} \bar{c}_{k-1}^{i}}{c_{k-1}^{j}} \hat{x}_{k-1|k-1}^{i}, \tag{3.118}$$

$$\Sigma_{k-1|k-1}^{0j} \triangleq \sum_{i=1}^{N} \frac{\pi_{ij} \bar{c}_{k-1}^{i}}{c_{k-1}^{j}} \Big[P_{k-1|k-1}^{i} + (\hat{x}_{k-1|k-1}^{i} - \hat{x}_{k-1|k-1}^{0j})(\hat{x}_{k-1|k-1}^{i} - \hat{x}_{k-1|k-1}^{0j})^{T} \Big]. \tag{3.119}$$

This approximation is of the same form as the approximation of the IMM filter and it causes the densities $\gamma_k^j(x)$ to be unnormalized Gaussian densities in the form given below.

$$\gamma_k^j(x) = \bar{c}_k^j \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j). \tag{3.120}$$

Therefore; the resulting filter is finite-dimensional and requires finite number of statistics to be kept. In the following, the recursions for the required statistics $\hat{x}_{k|k}^j$, $\Sigma_{k|k}^j$ and $\bar{c}_k^j$ are obtained. Substituting the approximation in Eqn. 3.115 into the recursion in Eqn. 3.99, we get

$$\begin{aligned} \gamma_k^j(x) &= c_{k-1}^j \frac{\psi(D_j^{-1}(y_k - C_j x))}{|B_j||D_j|\psi(y_k)} \int \phi(B_j^{-1}(x - A_j z)) \\ &\quad \times \exp\left(+, z, \hat{x}_{k-1|k-1}^{CRS}, \frac{1}{\theta} Q_{k-1}^{-1}\right) \mathcal{N}(z; \hat{x}_{k-1|k-1}^{0j}, \Sigma_{k-1|k-1}^{0j}) dz \\ &= c_{k-1}^j \frac{\psi(D_j^{-1}(y_k - C_j x))}{\sqrt{|2\pi\Sigma_{k-1|k-1}^{0j}|}|B_j||D_j|\psi(y_k)} \\ &\quad \times \int \phi(B_j^{-1}(x - A_j z)) \exp\left(+, z, \hat{x}_{k-1|k-1}^{CRS}, \frac{1}{\theta} Q_{k-1}^{-1}\right) \\ &\quad \times \exp\left(-, z, \hat{x}_{k-1|k-1}^{0j}, \Sigma_{k-1|k-1}^{0j}\right) dz. \end{aligned} \tag{3.121}$$

Using the result on the multiplication of exponentials given in App. B.1, Eqn. 3.121 turns into

$$\gamma_k^j(x) = c_{k-1}^j \frac{\sqrt{|2\pi\Sigma_{k-1|k-1}^{0j}|}\psi(D_j^{-1}(y_k - C_j x))}{\sqrt{|2\pi\Sigma_{k-1|k-1}^{0j}|}|B_j||D_j|\psi(y_k)}$$

$$
\begin{aligned}
& \times \int \phi(B_j^{-1}(x - A_j z)) \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \\
& \times \mathcal{N}(z; \underline{\hat{x}}^{0j}_{k-1|k-1}, \underline{\Sigma}^{0j}_{k-1|k-1}) dz \qquad (3.122) \\
= \; & c^j_{k-1} \frac{\sqrt{|\underline{\Sigma}^{0j}_{k-1|k-1}|}\psi(D_j^{-1}(y_k - C_j x))}{\sqrt{|\Sigma^{0j}_{k-1|k-1}||B_j||D_j|}\psi(y_k)} \\
& \times \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \\
& \times \int \phi(B_j^{-1}(x - A_j z)) \mathcal{N}(z; \underline{\hat{x}}^{0j}_{k-1|k-1}, \underline{\Sigma}^{0j}_{k-1|k-1}) dz \qquad (3.123) \\
= \; & c^j_{k-1} \frac{\sqrt{|\underline{\Sigma}^{0j}_{k-1|k-1}|}\psi(D_j^{-1}(y_k - C_j x))}{\sqrt{|\Sigma^{0j}_{k-1|k-1}||D_j|}\psi(y_k)} \\
& \times \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \\
& \times \int \mathcal{N}(x, A_j z, B_j B_j^T) \mathcal{N}(z; \underline{\hat{x}}^{0j}_{k-1|k-1}, \underline{\Sigma}^{0j}_{k-1|k-1}) dz \qquad (3.124)
\end{aligned}
$$

where

$$
\begin{aligned}
S^{0j}_{k-1|k-1} &= \frac{1}{\theta} Q^{-1}_{k-1} - \Sigma^{0j}_{k-1|k-1}, & (3.125) \\
\underline{\Sigma}^{0j}_{k-1|k-1} &= \left[(\Sigma^{0j}_{k-1|k-1})^{-1} - \theta Q_{k-1}\right]^{-1}, & (3.126) \\
\underline{\hat{x}}^{0j}_{k-1|k-1} &= \underline{\Sigma}^{0j}_{k-1|k-1}\left[(\Sigma^{0j}_{k-1|k-1})^{-1}\hat{x}^{0j}_{k-1|k-1} - \theta Q_{k-1}\hat{x}^{CRS}_{k-1|k-1}\right]. & (3.127)
\end{aligned}
$$

Applying the result of App. B.4 onto the integral in Eqn. 3.124, we obtain

$$
\begin{aligned}
\gamma^j_k(x) &= c^j_{k-1} \frac{\sqrt{|\underline{\Sigma}^{0j}_{k-1|k-1}|}\psi(D_j^{-1}(y_k - C_j x))}{\sqrt{|\Sigma^{0j}_{k-1|k-1}||D_j|}\psi(y_k)} \\
& \times \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \\
& \times \mathcal{N}(x; \hat{x}^j_{k|k-1}, \Sigma^j_{k|k-1}) \qquad (3.128) \\
&= c^j_{k-1} \frac{\sqrt{|\underline{\Sigma}^{0j}_{k-1|k-1}|}}{\sqrt{|\Sigma^{0j}_{k-1|k-1}|}\psi(y_k)} \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \\
& \times \mathcal{N}(y_k; C_j x, D_j D_j^T) \mathcal{N}(x; \hat{x}^j_{k|k-1}, \Sigma^j_{k|k-1}) \qquad (3.129)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{x}^j_{k|k-1} &= A_j \underline{\hat{x}}^{0j}_{k-1|k-1}, & (3.130) \\
\Sigma^j_{k|k-1} &= A_j \underline{\Sigma}^{0j}_{k-1|k-1} A_j^T + B_j B_j^T. & (3.131)
\end{aligned}
$$

The result of Appendix B.2 can be applied to Eqn. 3.129 to yield

$$\begin{aligned}
\gamma_k^j(x) &= c_{k-1}^j \frac{\sqrt{|\Sigma_{k-1|k-1}^{0j}|}}{\sqrt{|\Sigma_{k-1|k-1}^{0j}|\psi(y_k)}} \mathcal{N}(y_k; C_j x, D_j D_j^T) \\
&\quad \times \exp\left(+, \hat{x}_{k-1|k-1}^{CRS}, \hat{x}_{k-1|k-1}^{0j}, S_{k-1|k-1}^{0j}\right) \\
&\quad \times \mathcal{N}(y_k; C_j \hat{x}_{k|k-1}^j, S_{k|k-1}^j) \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \qquad (3.132) \\
&= \bar{c}_k^j \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \qquad (3.133)
\end{aligned}$$

where

$$\begin{aligned}
\bar{c}_k^j &= \frac{c_{k-1}^j \sqrt{|\Sigma_{k-1|k-1}^{0j}|}}{\sqrt{|\Sigma_{k-1|k-1}^{0j}|\psi(y_k)}} \mathcal{N}(y_k; C_j \hat{x}_{k|k-1}^j, S_{k|k-1}^j) \\
&\quad \times \exp\left(+, \hat{x}_{k-1|k-1}^{CRS}, \hat{x}_{k-1|k-1}^{0j}, S_{k-1|k-1}^{0j}\right), \qquad (3.134) \\
S_{k|k-1}^j &= C_j \Sigma_{k|k-1}^j C_j^T + D_j D_j^T, \qquad (3.135) \\
\hat{x}_{k|k}^j &= \hat{x}_{k|k-1}^j + \Sigma_{k|k-1}^j C_j^T (y_k - C_j \hat{x}_{k|k-1}^j), \qquad (3.136) \\
\Sigma_{k|k}^j &= \Sigma_{k|k-1}^j - \Sigma_{k|k-1}^j C_j^T (S_{k|k-1}^j)^{-1} C_j \Sigma_{k|k-1}^j. \qquad (3.137)
\end{aligned}$$

3.4.4 Final Estimate Calculation

Remembering the cumulative risk-sensitive estimate formula given in Eqn. 3.15,

$$\begin{aligned}
\hat{x}_{k|k}^{CRS} &= \arg\min_{\zeta \in \mathbb{R}^n} \overline{E}\left[\overline{\Lambda}_k \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\} \exp\left\{\frac{\theta}{2}(x_k - \zeta)^T Q_k (x_k - \zeta)\right\} \middle| \mathcal{Y}_k\right] \\
&= \arg\min_{\zeta \in \mathbb{R}^n} \sum_{j=1}^N \overline{E}\left[\overline{\Lambda}_k \langle r_k, e_j \rangle \exp\left\{\theta \hat{\Psi}_{0,k-1}\right\}\right. \\
&\quad \left. \times \exp\left\{\frac{\theta}{2}(x_k - \zeta)^T Q_k (x_k - \zeta)\right\} \middle| \mathcal{Y}_k\right] \qquad (3.138) \\
&= \arg\min_{\zeta \in \mathbb{R}^n} \sum_{j=1}^N \int \exp\left\{\frac{\theta}{2}(x - \zeta)^T Q_k (x - \zeta)\right\} \gamma_k^j(x) dx. \qquad (3.139)
\end{aligned}$$

At this stage of the problem, in order to calculate the final estimate, two different approximation schemes which yield very different results can be used. Here, we examine both of the approaches and call the resulting CRS-IMM variants as CRS-IMM1 and CRS-IMM2.

Approximation Scheme-1: The first scheme is to make the following approximation.

$$\sum_{j=1}^{N} \gamma_k^j(x) = \sum_{j=1}^{N} \bar{c}_k^j \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) \approx \bar{c}_k \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}) \tag{3.140}$$

where

$$\bar{c}_k \triangleq \sum_{j=1}^{N} \bar{c}_k^j, \tag{3.141}$$

$$\hat{x}_{k|k} = \sum_{j=1}^{N} \frac{\bar{c}_k^j}{\bar{c}_k} \hat{x}_{k|k}^j, \tag{3.142}$$

$$\Sigma_{k|k} = \sum_{j=1}^{N} \frac{\bar{c}_k^j}{\bar{c}_k} \left[\Sigma_{k|k}^j + (\hat{x}_{k|k}^j - \hat{x}_{k|k})(\hat{x}_{k|k}^j - \hat{x}_{k|k})^T \right]. \tag{3.143}$$

The final estimate $\hat{x}_{k|k}^{CRS}$ is then given by

$$\hat{x}_{k|k}^{CRS} = \arg\min_{\zeta \in \mathbb{R}^n} \sum_{j=1}^{N} \int \exp\left\{ \frac{\theta}{2} (x-\zeta)^T Q_k (x-\zeta) \right\} \gamma_k^j(x) dx \tag{3.144}$$

$$= \arg\min_{\zeta \in \mathbb{R}^n} \int \exp\left\{ \frac{\theta}{2} (x-\zeta)^T Q_k (x-\zeta) \right\} \sum_{j=1}^{N} \gamma_k^j(x) dx \tag{3.145}$$

$$\approx \arg\min_{\zeta \in \mathbb{R}^n} \bar{c}_k \int \exp\left\{ \frac{\theta}{2} (x-\zeta)^T Q_k (x-\zeta) \right\} \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}) dx$$

$$= \arg\min_{\zeta \in \mathbb{R}^n} \bar{c}_k \frac{\sqrt{|M_k|}}{\sqrt{|\Sigma_{k|k}|}} \exp\left\{ \frac{1}{2} (\zeta - \hat{x}_{k|k})^T N_k (\zeta - \hat{x}_{k|k}) \right\} \tag{3.146}$$

$$= \hat{x}_{k|k} \tag{3.147}$$

where

$$M_k = \left[\Sigma_{k|k}^{-1} - \theta Q_k \right]^{-1}, \tag{3.148}$$

$$N_k = \left[\frac{1}{\theta} Q_k^{-1} - \Sigma_{k|k} \right]^{-1}. \tag{3.149}$$

As a result, using the first approximation scheme, the final estimate is obtained as

$$\hat{x}_{k|k}^{CRS} = \sum_{j=1}^{N} \frac{\bar{c}_k^j}{\bar{c}_k} \hat{x}_{k|k}^j. \tag{3.150}$$

Approximation Scheme-2: Substituting the densities $\gamma_k^j(x)$ into Eqn. 3.139, we obtain

$$\begin{aligned}\hat{x}_{k|k}^{CRS} &= \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \bar{c}_k^j \int \exp\left\{\frac{\theta}{2}(x-\zeta)^T Q_k (x-\zeta)\right\} \mathcal{N}(x; \hat{x}_{k|k}^j, \Sigma_{k|k}^j) dx \\ &= \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \frac{\bar{c}_k^j}{\sqrt{|2\pi\Sigma_{k|k}^j|}} \int \exp\left\{\frac{\theta}{2}(x-\zeta)^T Q_k (x-\zeta)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(x-\hat{x}_{k|k}^j)^T (\Sigma_{k|k}^j)^{-1} (x-\hat{x}_{k|k}^j)\right\} dx.\end{aligned}$$

After taking the integrals using the result of Appendix B.3, we get

$$\begin{aligned}\hat{x}_{k|k}^{CRS} &= \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \frac{\bar{c}_k^j \sqrt{|T_k^j|}}{\sqrt{|\Sigma_{k|k}^j|}} \exp\left\{\frac{1}{2}(\zeta-\hat{x}_{k|k}^j)^T U_k^j (\zeta-\hat{x}_{k|k}^j)\right\} \qquad (3.151) \\ &= \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \frac{\bar{c}_k^j \sqrt{|\frac{1}{\theta}Q_k^{-1}||U_k^j||\Sigma_{k|k}^j|}}{\sqrt{|\Sigma_{k|k}^j|}} \exp\left\{\frac{1}{2}(\zeta-\hat{x}_{k|k}^j)^T U_k^j (\zeta-\hat{x}_{k|k}^j)\right\} \\ &= \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \frac{\bar{c}_k^j \sqrt{|U_k^j|}}{\sqrt{|\theta Q_k|}} \exp\left\{\frac{1}{2}(\zeta-\hat{x}_{k|k}^j)^T U_k^j (\zeta-\hat{x}_{k|k}^j)\right\} \qquad (3.152)\end{aligned}$$

where

$$T_k^j = \left[(\Sigma_{k|k}^j)^{-1} - \theta Q_k\right]^{-1}, \qquad (3.153)$$

$$U_k^j = \left[\frac{1}{\theta}Q_k^{-1} - \Sigma_{k|k}^j\right]^{-1}. \qquad (3.154)$$

The summation of weighted exponentials in Eqn. 3.152 is similar to the case encountered Chapter 2. We here use the same approach by replacing the exponentials by their first-order Taylor series expansion i.e., we let

$$\exp(x) \approx 1 + x. \qquad (3.155)$$

Then, the second type risk-sensitive final estimate $\hat{x}_{k|k}^{CRS}$ is given as

$$\hat{x}_{k|k}^{CRS} = \arg\min_{\zeta\in\mathbb{R}^n} \sum_{j=1}^{N} \frac{\bar{c}_k^j \sqrt{|U_k^j|}}{2\sqrt{|\theta Q_k|}} (\zeta-\hat{x}_{k|k}^j)^T U_k^j (\zeta-\hat{x}_{k|k}^j).$$

Taking the gradient with respect to ζ and equating to zero, we obtain

$$\hat{x}_{k|k}^{CRS} = \left[\sum_{j=1}^{N} \bar{c}_k^j \sqrt{|U_k^j|} U_k^j\right]^{-1} \left[\sum_{j=1}^{N} \bar{c}_k^j \sqrt{|U_k^j|} U_k^j \hat{x}_{k|k}^j\right].$$

Summary: Here, we summarize the derived cumulative risk-sensitive multiple-model estimation algorithm (both variants). The algorithm, like the IMM filter, keeps an unnormalized Gaussian density for each density $\gamma_k^j(x)$. At each time step, it has to store the means $\{\hat{x}_{k|k}^j\}_{j=1}^N$, covariances $\{\Sigma_{k|k}^j\}_{j=1}^N$, the weighting coefficients $\{\bar{c}_k^j\}_{j=1}^N$ and a single final output estimate $\hat{x}_{k|k}^{CRS}$. The storage requirement for the final estimate is a distinct characteristic of the cumulative risk-sensitive filter and is not required in the IMM filter.

Assuming that the required statistic data from the previous time step (i.e., time $k-1$) are available (i.e., the means $\{\hat{x}_{k-1|k-1}^i\}_{i=1}^N$, covariances $\{\Sigma_{k-1|k-1}^i\}_{i=1}^N$, the weighting coefficients $\{\bar{c}_k^i\}_{i=1}^N$ and the final output estimate $\hat{x}_{k-1|k-1}^{CRS}$ are available.), the algorithm has the following steps.

- **Mixing:** The previous state estimates $\{\hat{x}_{k-1|k-1}^i\}_{i=1}^N$ and covariances $\{\Sigma_{k-1|k-1}^i\}_{i=1}^N$ are mixed to form new mixed estimates $\{\hat{x}_{k-1|k-1}^{0j}\}_{j=1}^N$ and covariances $\{\Sigma_{k-1|k-1}^{0j}\}_{j=1}^N$ as follows.

$$\begin{aligned}\hat{x}_{k-1|k-1}^{0j} &= \sum_{i=1}^{N} \frac{\pi_{ij}\bar{c}_{k-1}^i}{c_{k-1}^j}\hat{x}_{k-1|k-1}^i, \qquad (3.156)\\ \Sigma_{k-1|k-1}^{0j} &= \sum_{i=1}^{N} \frac{\pi_{ij}\bar{c}_{k-1}^i}{c_{k-1}^j}\Big[\Sigma_{k-1|k-1}^i \\ &\qquad +(\hat{x}_{k-1|k-1}^i - \hat{x}_{k-1|k-1}^{0j})(\hat{x}_{k-1|k-1}^i - \hat{x}_{k-1|k-1}^{0j})^T\Big]\end{aligned}$$

  where

$$c_{k-1}^j \triangleq \sum_{i=1}^{N} \pi_{ij}\bar{c}_{k-1}^i. \qquad (3.157)$$

- **Modification on Mixed Estimates and Covariances:** The mixed state estimates $\{\hat{x}_{k-1|k-1}^{0j}\}_{j=1}^N$ and covariances $\{\Sigma_{k-1|k-1}^{0j}\}_{j=1}^N$ are modified to obtain modified mixed estimates $\{\underline{\hat{x}}_{k-1|k-1}^{0j}\}_{j=1}^N$ and covariances $\{\underline{\Sigma}_{k-1|k-1}^{0j}\}_{j=1}^N$ as follows.

$$\begin{aligned}\underline{\Sigma}_{k-1|k-1}^{0j} &= \Big[(\Sigma_{k-1|k-1}^{0j})^{-1} - \theta Q_{k-1}\Big]^{-1}, \qquad (3.158)\\ \underline{\hat{x}}_{k-1|k-1}^{0j} &= \underline{\Sigma}_{k-1|k-1}^{0j} \\ &\qquad \times \Big[(\Sigma_{k-1|k-1}^{0j})^{-1}\hat{x}_{k-1|k-1}^{0j} - \theta Q_{k-1}\hat{x}_{k-1|k-1}^{CRS}\Big]. \qquad (3.159)\end{aligned}$$

- **Kalman Filtering:** The modified mixed estimates $\{\underline{\hat{x}}^{0j}_{k-1|k-1}\}_{j=1}^{N}$ and covariances $\{\underline{\Sigma}^{0j}_{k-1|k-1}\}_{j=1}^{N}$ are updated using corresponding mode-matched Kalman filters to obtain estimates $\{\hat{x}^{j}_{k|k}\}_{j=1}^{N}$ and covariances $\{\Sigma^{j}_{k|k}\}_{j=1}^{N}$.

  - Prediction Update

$$\hat{x}^{j}_{k|k-1} = A_j \underline{\hat{x}}^{0j}_{k-1|k-1}, \tag{3.160}$$

$$\Sigma^{j}_{k|k-1} = A_j \underline{\Sigma}^{0j}_{k-1|k-1} A_j^T + B_j B_j^T. \tag{3.161}$$

  - Measurement Update

$$\hat{x}^{j}_{k|k} = \hat{x}^{j}_{k|k-1} + \Sigma^{j}_{k|k-1} C_j^T (y_k - C_j \hat{x}^{j}_{k|k-1}), \tag{3.162}$$

$$\Sigma^{j}_{k|k} = \Sigma^{j}_{k|k-1} - \Sigma^{j}_{k|k-1} C_j^T (S^{j}_{k|k-1})^{-1} C_j \Sigma^{j}_{k|k-1} \tag{3.163}$$

    where

$$S^{j}_{k|k-1} = C_j \Sigma^{j}_{k|k-1} C_j^T + D_j D_j^T. \tag{3.164}$$

- **Obtaining the Coefficients $\{\bar{c}^{j}_{k}\}_{j=1}^{N}$:** The coefficients $\{\bar{c}^{j}_{k}\}_{j=1}^{N}$ are calculated as follows

$$\begin{aligned} \bar{c}^{j}_{k} = & \frac{c^{j}_{k-1}\sqrt{|\underline{\Sigma}^{0j}_{k-1|k-1}|}}{\sqrt{|\Sigma^{0j}_{k-1|k-1}|}\psi(y_k)} \mathcal{N}(y_k; C_j \hat{x}^{j}_{k|k-1}, S^{j}_{k|k-1}) \\ & \times \exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \end{aligned} \tag{3.165}$$

  where

$$S^{0j}_{k-1|k-1} = \frac{1}{\theta} Q^{-1}_{k-1} - \Sigma^{0j}_{k-1|k-1}. \tag{3.166}$$

- **Output Estimate Calculation:** The output estimate can be calculated in two different ways yielding two variants of CRS-IMM as

  - CRS-IMM1: Final estimate $\hat{x}^{CRS}_{k|k}$ is given as

$$\hat{x}^{CRS}_{k|k} = \sum_{j=1}^{N} \frac{\bar{c}^{j}_{k}}{\bar{c}_k} \hat{x}^{j}_{k|k} \tag{3.167}$$

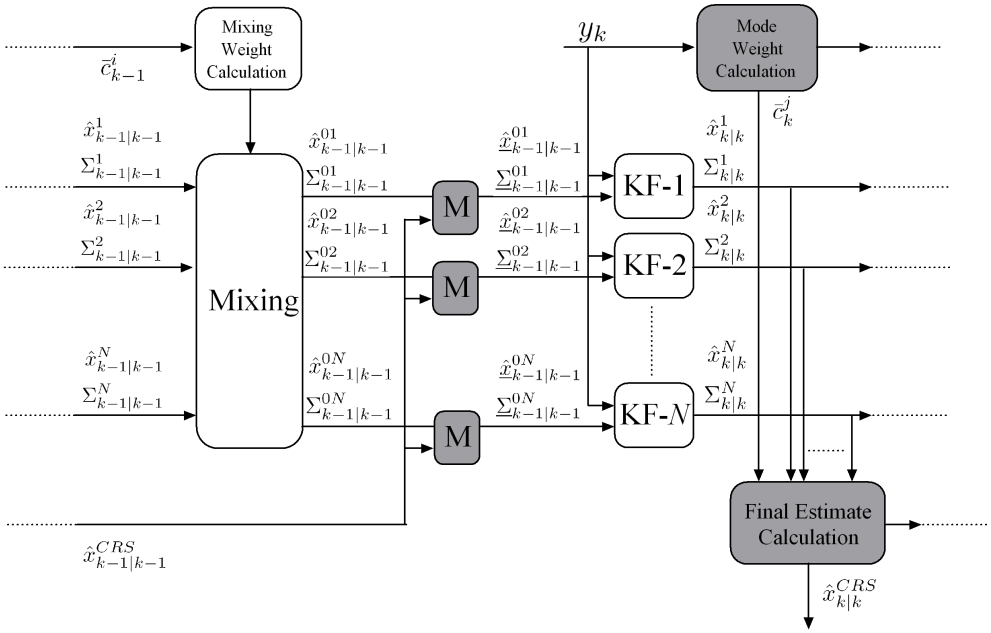

Figure 3.2: One cycle of CRS-IMM algorithm with N models. M and KF are the abbreviations of modification and Kalman filter respectively. The shaded regions in the figure signify the parts of the CRS-IMM algorithm different from the IMM filter.

where

$$\bar{c}_k \triangleq \sum_{j=1}^{N} \bar{c}_k^j. \tag{3.168}$$

– CRS-IMM2: Final estimate $\hat{x}_{k|k}^{CRS}$ is given as

$$\hat{x}_{k|k}^{CRS} = \left[\sum_{j=1}^{N} \bar{c}_k^j \sqrt{|U_k^j|} U_k^j\right]^{-1} \left[\sum_{j=1}^{N} \bar{c}_k^j \sqrt{|U_k^j|} U_k^j \hat{x}_{k|k}^j\right] \tag{3.169}$$

where

$$U_k^j = \left[\frac{1}{\theta} Q_k^{-1} - \Sigma_{k|k}^j\right]^{-1}. \tag{3.170}$$

A flow-chart of one cycle of the CRS-IMM algorithm is shown in Fig. 3.2 where the parts of the algorithm which are different from the IMM filter are denoted with shaded boxes.

3.5 Properties and Implementation Issues of the CRS-IMM Filter

This section highlights some properties of the CRS-IMM algorithm. Furthermore, some implementation issues are also considered.

3.5.1 Properties

3.5.1.1 Comparison with IMM Filter

The overall structure of the CRS-IMM is similar to IMM filter except for the modification step. In this step each mixed estimate $\hat{x}^{0j}_{k-1|k-1}$ with mixed covariance $\Sigma^{0j}_{k-1|k-1}$ is updated with the risk-sensitive final estimate $\hat{x}^{CRS}_{k-1|k-1}$ of time $k-1$. This modification process is equivalent to the measurement update of a Kalman filter with equivalent measurement

$$\hat{x}^{CRS}_{k-1|k-1} = \hat{x}^{0j}_{k-1|k-1} + q_{k-1} \tag{3.171}$$

where the generalized random variable q_{k-1} has the negative definite covariance $-\frac{1}{\theta}Q^{-1}_{k-1}$. This covariance subtraction resembles, in a way, the information decorrelation approach of the fusion systems in target tracking [46].

Another remarkable difference of CRS-IMM from IMM is that the final output of the CRS-IMM affects the state and coefficient update equations. In IMM filter this final estimate is calculated only for output purposes [13].

3.5.1.2 Case $N = 1$

When there is only one model in the JMLS, the mixing and the final output calculations do not modify the estimates i.e.,

$$\hat{x}^{01}_{k-1|k-1} = \hat{x}^{1}_{k-1|k-1}, \tag{3.172}$$

$$\hat{x}^{CRS}_{k|k} = \hat{x}^{1}_{k|k}. \tag{3.173}$$

Moreover, since

$$\hat{x}^{CRS}_{k-1|k-1} = \hat{x}^{1}_{k-1|k-1} = \hat{x}^{01}_{k-1|k-1}, \tag{3.174}$$

the modification on the mixed quantities modifies only the mixed covariance and not the mixed estimate. The CRS-IMM filter (both variants), at the end, becomes the same filter as the one given in [26] and [47] which are derived for linear Gauss-Markov systems.[6]

When $N \neq 1$, it must be emphasized that the CRS-IMM filter modifies both the mixed estimates and the mixed covariances. Therefore, although the filter reduces to the risk-sensitive filter for linear Gauss-Markov systems when $N = 1$, it is not a combination of these filters when $N \neq 1$ unlike the case of IMM which is a combination of Kalman filters (when $N \neq 1$).

3.5.1.3 Case $\theta \to 0$

A general property of the risk-sensitive filters is that they reduce to MMSE filters when the risk-sensitive parameter θ tends to zero. For example, the risk-sensitive filter derived in [26] and [47] for linear Gauss-Markov systems reduces to the Kalman filter when $\theta \to 0$.[7]

When $\theta \to 0$, the modification step in CRS-IMM filter does not make any modification on the mixed estimate and covariances i.e.,

$$\underline{\hat{x}}^{0j}_{k-1|k-1} \rightarrow \hat{x}^{0j}_{k-1|k-1}, \tag{3.175}$$

$$\underline{\Sigma}^{0j}_{k-1|k-1} \rightarrow \Sigma^{0j}_{k-1|k-1}. \tag{3.176}$$

Since $(S^{0j}_{k-1|k-1})^{-1} \to \theta Q_{k-1}$,

$$\exp\left(+, \hat{x}^{CRS}_{k-1|k-1}, \hat{x}^{0j}_{k-1|k-1}, S^{0j}_{k-1|k-1}\right) \longrightarrow 1. \tag{3.177}$$

The facts given in Eqn. 3.176 and Eqn. 3.177 make the coefficients $\bar{c}^j_k$ (or c^j_k) equal to those in the IMM filter. This shows that the cumulative risk-sensitive estimate $\hat{x}^{CRS}_{k|k}$ of CRS-IMM1 is equal to IMM estimate $\hat{x}^{MS}_{k|k}$. Observing that when $\theta \to 0$, U^j_k tends to θQ_k which is independent of j, the risk-sensitive state estimate $\hat{x}^{CRS}_{k|k}$ of CRS-IMM2 also becomes equal to the IMM estimate. As a result, when θ goes to zero, both of the CRS-IMM filter variants reduce to the IMM filter.

[6] This filter is also derived using the reference probability method in App. A.3.
[7] This fact is also apparent in the derivations of App. A.

3.5.2 Implementation Issues

3.5.2.1 Convergence

The derivation of the filter requires the inequalities

$$(\Sigma^{0j}_{k-1|k-1})^{-1} > \theta Q_{k-1} \quad \text{for} \quad j = 1, 2, \ldots, N, \tag{3.178}$$

$$(\Sigma^{j}_{k|k})^{-1} > \theta Q_{k} \quad \text{for} \quad j = 1, 2, \ldots, N \tag{3.179}$$

be satisfied. Thus, the algorithm is convergent if the risk-sensitive parameter θ is sufficiently small.

3.5.2.2 Systems with Arbitrary Noise Covariances

In the derivation of the CRS-IMM algorithm, the process and measurement noise covariances have been taken to be identity matrices. In many applications, these covariance matrices are non-identity positive definite matrices. If one is given with non-identity covariances $M_k > 0$ and $N_k > 0$ corresponding to w_k and v_k respectively then, the system representation given by Eqn. 3.1 and Eqn. 3.2, can be written as

$$x_{k+1} = A(r_{k+1})x_k + B'(r_{k+1})w'_{k+1}, \tag{3.180}$$

$$y_k = C(r_k)x_k + D'(r_k)v'_k \tag{3.181}$$

where

$$B'(r_k) \triangleq B(r_k)\sqrt{M_k}, \quad D'(r_k) \triangleq D(r_k)\sqrt{N_k} \tag{3.182}$$

and $\sqrt{M_k}$ and $\sqrt{N_k}$ are positive definite square roots of M_k and N_k respectively. Moreover, the new noise terms w'_k and v'_k are Gaussian noise terms with identity covariances. This solution amounts to replacing every instance of $B_jB_j^T$ and $D_jD_j^T$ in the final estimator formulas with $B_jM_kB_j^T$ and $D_jN_kD_j^T$ respectively.

3.5.2.3 Systems with Non-Invertible and/or Non-Square Noise Gain Matrices

In the derivation, the noise gain matrices $B(r_k)$ and $D(r_k)$ have been assumed to be invertible. In practice, these matrices may not even be square. In this

section, we show that this is not a restriction for the final formulas of CRS-IMM algorithm to apply. Suppose that $B(r_k)$ ($D(r_k)$) is not square but its columns are linearly independent. Then, one can always augment the matrix $B(r_k)$ ($D(r_k)$) with appropriate number of column vectors of the form ϵb_i (ϵd_i) where b_i's (d_i's) are unity-norm vectors which are linearly independent with the columns of $B(r_k)$ ($D(r_k)$) and $\epsilon > 0$ is a small real number. These extra columns would correspond to dummy noise terms added to process noise w_k (measurement noise v_k). Then, making the same derivation and taking the limit as $\epsilon \to 0$ in the final formulas, one can see that the same formulas for the CRS-IMM algorithm are satisfied.

When the columns of the matrix $B(r_k)$ ($D(r_k)$) are not initially linearly independent, one can always find a matrix $B'(r_k)$ ($D'(r_k)$) whose columns are linearly independent and a Gaussian random variable w'_k (v'_k) with identity covariance such that the distribution of $B(r_k)w_k$ ($D(r_k)v_k$) is the same as that of $B'(r_k)w'_k$ ($D'(r_k)v'_k$). Then, the above arguments apply with $B'(r_k)$ ($D'(r_k)$) instead of $B(r_k)$ ($D(r_k)$).

3.5.2.4 Numerical Issues

It has been noted in Remark 3.1 that only the relative (i.e., not absolute) magnitudes of the coefficients $\bar{c}_k^j$ (or c_k^j) are important for IMM final and mixed estimate calculations. This is also the case for the CRS-IMM mixed and final estimate calculations. At any time step k, the coefficients $\bar{c}_k^j$ (or c_k^j) can be multiplied by a common positive constant without changing the estimated quantities. In the cases where the coefficients get too small or too big to handle in computer, one can make some normalization on them accordingly without affecting the performance. Moreover, the term $\psi(y_k)$ given in the coefficient update equations, which can cause a division by zero in the computer, can be discarded safely.

3.6 Simulation Results

In this section, the performance of the CRS-IMM algorithm will be observed and compared to that of the IMM algorithm. For this purpose, we consider a similar target-tracking scenario to the one used in Chapter 2.[8] The target dynamics in one-dimension is given as

$$\begin{bmatrix} p_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} p_k \\ v_k \end{bmatrix}}_{x_k} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} [a_k + w_k] \tag{3.183}$$

where p_k, v_k and a_k denote the target position, velocity and acceleration respectively. The initial state x_0 is normally distributed with mean $\bar{x}_0$ and covariance P_0 which are given as,

$$\bar{x}_0 = \begin{bmatrix} 80000 \\ 400 \end{bmatrix}, \qquad P_0 = \begin{bmatrix} 10000 & 1000 \\ 1000 & 10000 \end{bmatrix}. \tag{3.184}$$

The acceleration process a_k is a finite-state Markov chain with states in the set $\{0, 10, -10\}$. The transition probability matrix for the finite-state Markov chain is

$$\Pi = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} \tag{3.185}$$

which corresponds to a moderately maneuvering target. The white process noise $w_k \sim \mathcal{N}(w_k; 0, \sigma_w^2 = 2^2)$ represents small acceleration changes. It is assumed that only the positions are measured, i.e.,

$$y_k = p_k + \nu_k \tag{3.186}$$

where the terms $\nu_k \sim \mathcal{N}(\nu_k; 0, 100^2)$ stands for the normally distributed white measurement noise. The sampling period T is taken to be 10secs.

The IMM and CRS-IMM algorithms are run on the artificially generated measurements of the system defined above for 1000 Monte-Carlo runs. For

[8] Note that this example is a slightly modified version of the one given in [1].

Table 3.1: Average RMS Errors (per sample) of the IMM, CRS-IMM1 and CRS-IMM2 algorithms.

RMS-Error per sample	IMM	CRS-IMM1	CRS-IMM2
Position (m)	99.687	99.619	99.628
Velocity (m/sec)	43.896	35.364	35.374

the CRS-IMM filter variants, the risk sensitive parameter θ has been taken as 7×10^{-5} and the weighting matrices Q_k have been set to identity matrix I_2 for all k.

The algorithms are assumed not to know all of the characteristics of the system.[9] The measurement covariance and the true transition probability matrix given in Eqn. 3.185 are assumed to be known.[10] The possible acceleration values and the process noise variance are not known by the algorithms. To cover all possible target accelerations, the algorithms choose the acceleration values as $\{0, 20, -20\}$[11] and the process noise variance as $\sigma_w^2 = 4^2$.

In Fig. 3.3 and Fig. 3.4, the RMS position and velocity errors of the IMM and CRS-IMM1 algorithms are presented. The errors of the CRS-IMM2 are not shown on the figures since they are visually indistinguishable from those of CRS-IMM1. The average RMS position and velocity errors per measurement sample are given in Table 3.1. The RMS position errors of the filters are very near to each other and around 100m's. The values are very close to the position measurement standard deviation. Therefore, for only position estimation, using these filters can give only few meters extra accuracy in this

[9] Since the comparisons are made in terms of RMS errors, it is clear that when all the parameters of the system are known to the algorithms, the IMM will beat CRS-IMM since it minimizes expected quadratic estimation error. CRS-IMM, on the other hand, tries to minimize higher order moments of the error as well at the expense of increasing the RMS-errors.

[10] Note that the transition probability matrices are generally unknown in practical systems and applications. We have made this assumption here just to make it clear that the errors in the resulting simulations are caused by other uncertainties and not by the unknown transition probabilities. The unknown transition probabilities did not cause much change in the simulation results presented.

[11] The models of the related JMLS corresponding to acceleration values 20 and -20 have deterministic input values. IMM filtering with these models is straightforward using the standard Kalman filtering with deterministic inputs which modifies only the prediction step.

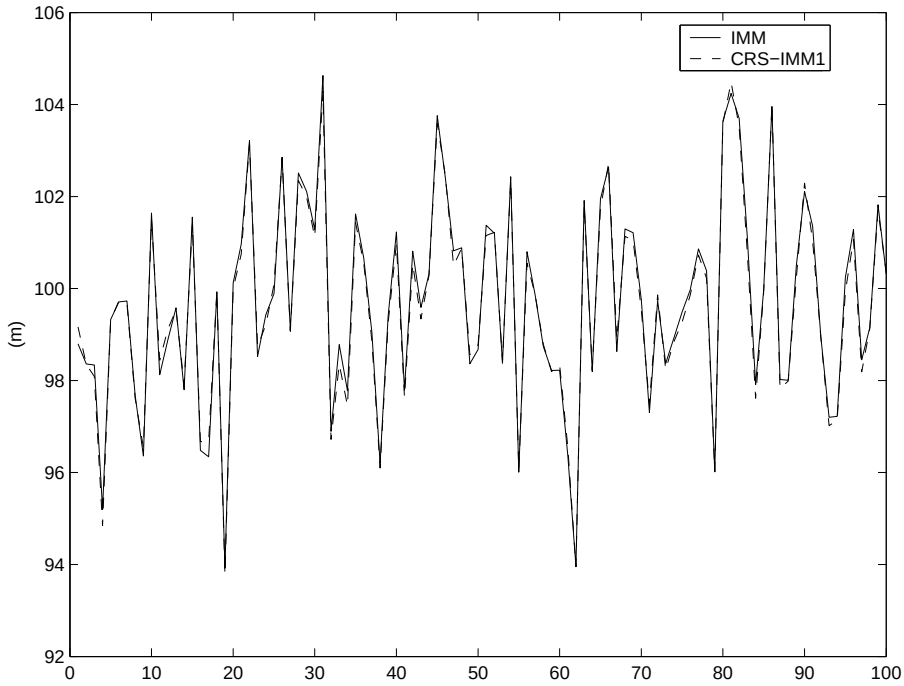

Figure 3.3: RMS position errors of the IMM and CRS-IMM1 algorithms.

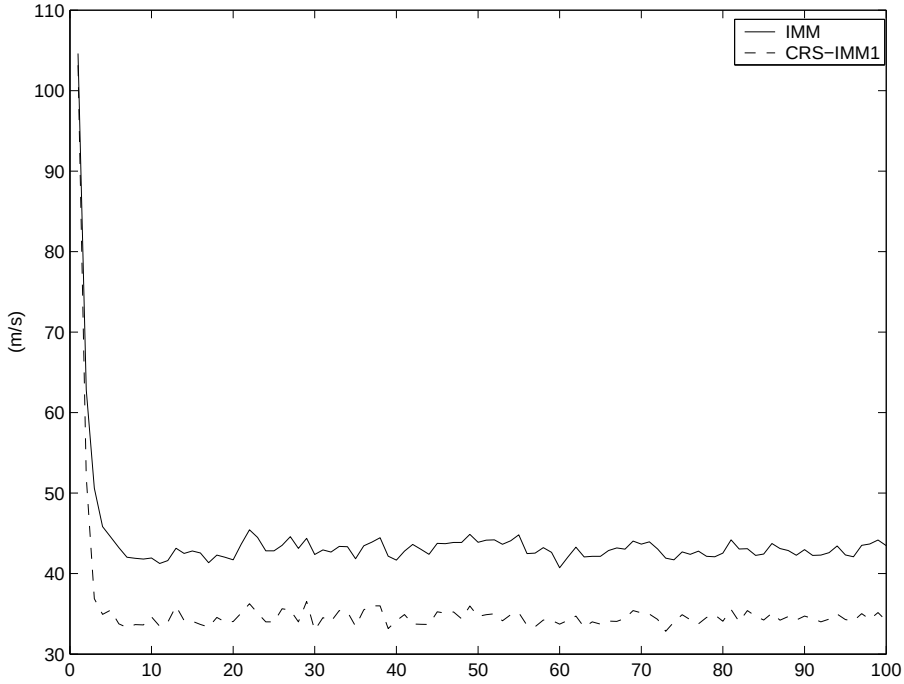

Figure 3.4: RMS velocity errors of the IMM and CRS-IMM1 algorithms.

uncertain parameter scenario. Using a simpler estimator $\hat{p}_k = y_k$ would be more efficient with similar position error variance. For velocity estimation, the RMS error curves are quite different showing the advantage of using CRS-IMM. Here, we compare the velocity estimation performances of the filters with a very simple estimator which gives the velocity estimates (with one sample delay) as

$$\hat{v}_k = \frac{y_{k+1} - y_k}{T}. \tag{3.187}$$

We can calculate the error variance of this simple estimator as

$$\begin{aligned}
E(\hat{v}_k - v_k)^2 &= E\left[(\frac{y_{k+1} - y_k}{T} - v_k)^2\right] && (3.188)\\
&= E\left[(\frac{p_{k+1} - p_k + \nu_{k+1} - \nu_k - Tv_k}{T})^2\right] \\
&= E\left[(\frac{\frac{T^2}{2}a_k + \frac{T^2}{2}w_k + \nu_{k+1} - \nu_k}{T})^2\right] && (3.189)\\
&= E\left[(\frac{T}{2}a_k + \frac{T}{2}w_k + \frac{\nu_{k+1} - \nu_k}{T})^2\right] && (3.190)\\
&= E\left[(5a_k + 5w_k + 0.1\nu_{k+1} - 0.1\nu_k)^2\right] && (3.191)\\
&= 25E[a_k^2] + 25\sigma_w^2 + 0.02\sigma_\nu^2. && (3.192)
\end{aligned}$$

Assuming that Markov chain a_k has reached steady-state, $E[a_k^2] = \frac{2}{3}10^2$ and substituting $\sigma_w^2 = 4$ and $\sigma_\nu^2 = 100^2$, we obtain $E(\hat{v}_k - v_k)^2 = 1.96 \times 10^3 = 44.34^2$. Comparing this to the average RMS velocity errors per sample, we see that, IMM filter,[12] under unknown acceleration and process noise values, has the velocity error standard deviation only about 0.5m/sec below that of the simpler estimator. The reduction in the case of CRS-IMM is approximately 10m/sec which is twenty times more than that of IMM. Consequently, for accurate velocity estimation, CRS-IMM might be a better choice in the cases where the system parameters are highly uncertain.

The average RMS position and velocity errors per measurement sample of the algorithms for different θ values are plotted in Fig. 3.5 and Fig. 3.6 respectively. Note that the error curves for CRS-IMM1 and CRS-IMM2 are

[12] Note that IMM filters can be designed differently and different designs might yield different results. See [13] for design issues and [14] for a survey of available variants.

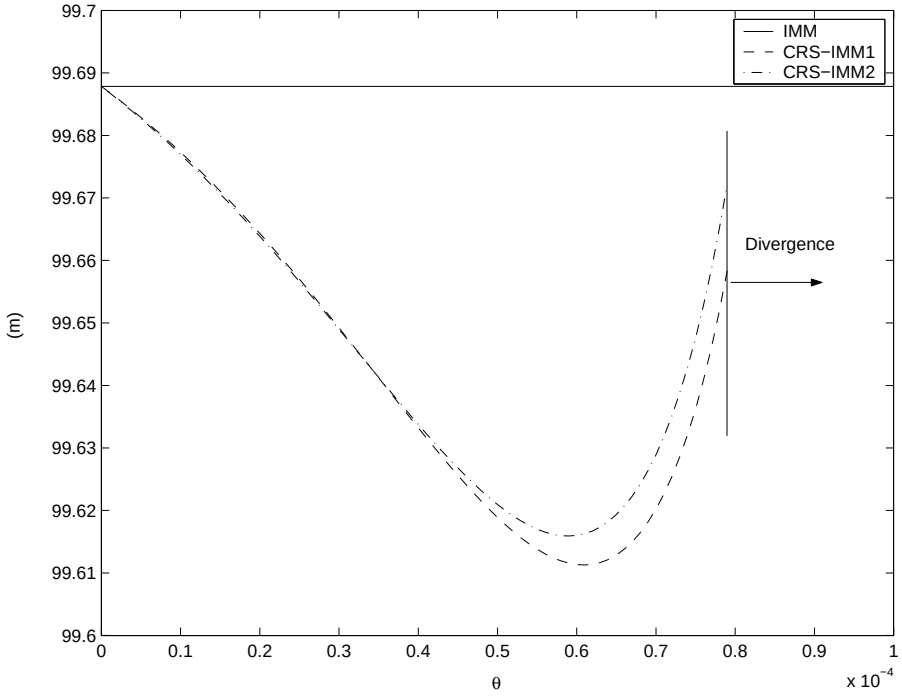

Figure 3.5: Average RMS position errors (per sample) of the IMM, CRS-IMM1 and CRS-IMM2 algorithms for different θ values.

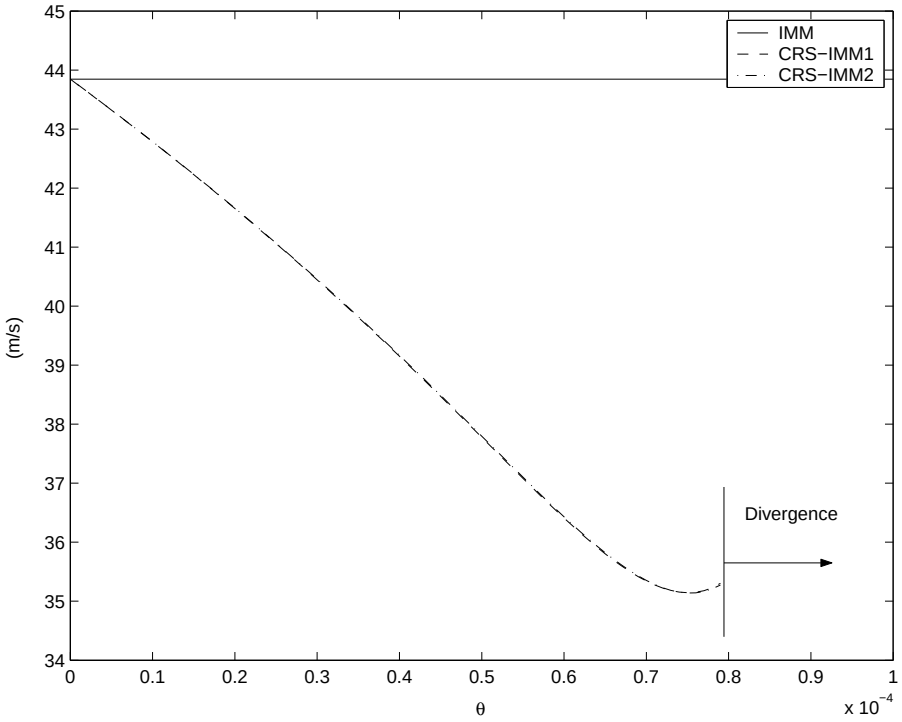

Figure 3.6: Average RMS velocity errors (per sample) of the IMM, CRS-IMM1 and CRS-IMM2 algorithms for different θ values.

indistinguishable (on top of each other) in Fig. 3.6. In both figures, the average RMS errors of the algorithms decrease (with increasing θ) almost linearly below that of the IMM algorithm until the value of the risk-sensitive parameter approaches the divergence limit which is about 8×10^{-5} for this example. The errors then begin to increase when θ gets closer to the divergence limit. The figures also suggest that better performance characteristics than those shown in Figures 3.3 and 3.4 can be obtained if better θ selection mechanisms are used.

3.7 Conclusion

A cumulative risk-sensitive multiple-model filter which can be thought of as a generalization of the well-known IMM filter is proposed. The filter, which (approximately) minimizes expected exponential of the cumulative quadratic estimation error, reduces to the IMM filter when the risk sensitive parameter θ tends to zero and to the risk-sensitive filter for linear Gauss-Markov systems when the number of models in the JMLS is unity. These cases show that the algorithm is a unifying framework under which the concepts of

1. risk-sensitive multiple-model filtering,
2. MMSE multiple-model (IMM) filtering,
3. risk-sensitive filtering for linear Gauss-Markov systems,
4. Kalman filtering

are combined. The proposed algorithm does not require any uncertainty description and therefore it can be especially useful in applications where the modeling uncertainty descriptions are not known.

CHAPTER 4

ONLINE ESTIMATION OF THE TRANSITION PROBABILITIES OF JMLSS USING RECURSIVE KULLBACK-LEIBLER METHOD

4.1 Introduction

In almost all existing state estimation methods for JMLSs, the transition probabilities of the underlying finite-state Markov chain are assumed to be known a priori. Even in the cases when they are not known, some diagonally dominant probability transition matrix is generally used. This unfortunate reality results, in part, from the limitedness of the related early literature. The uncertainty caused by the unknown transition probabilities associated with a JMLS attracted the attention of the researchers as early as 1973 when a Bayesian solution which involves a numerical integration to the problem of transition probability estimation was presented in [48]. In this work, Sawaragi et al. considered also the adaptive estimation of the state using these estimates in a system having interrupted measurements, which actually was a special case of a JMLS. At that time, this kind of systems was quite popular, and later, Tugnait and Haddad investigated the asymptotic behavior of this Bayesian solution when the unknown transition probability matrix can take values from a finite set in [49]. For the same case (where the unknown transition probability matrix can take values from a finite set), after Tugnait and Haddad presented a

similar solution for linear systems with Markovian jump noise parameters [50], Tugnait proposed an approximate maximum likelihood approach for general JMLSs [51]. Even when the transition probability matrix belongs to a finite set, it is concluded in [51] that the standard maximum likelihood estimation is not computationally feasible. Quite later, an (approximate) expectation-maximization procedure which maximizes a lower bound on the log-likelihood is given in [52] for switching state-space models[1] where the states of uncoupled state-space systems which have constant dynamics are switched in the output. Therefore, until 2004, when Jilkov and Li has addressed the problem of online MMSE estimation of the transition probabilities [1], the literature seemed to lack a complete approach of transition probability identification for JMLSs where transition probabilities belong to a continuous valued set (rather than being in a finite set).

In [1], the authors proposed some (conditionally) optimal (although infeasible to implement with its ever growing memory requirements) and sub-optimal MMSE algorithms for the online estimation of the transition probabilities as well as the base and modal states associated with a JMLS. In addition, [54] solves the problem of state estimation of JMLSs with unknown transition probabilities using Bayesian sampling based on the results of [1]. Although, to the author's knowledge, there is no previous attempt to solve the complete problem (in the sense of previous paragraph) until [1], a literature is already available for a similar problem associated with hidden Markov models (HMM)[2, 55, 56, 57, 58]. The aim of this chapter is to adapt one of these existing methods in the HMM literature, namely, the recursive Kullback-Leibler (RKL) estimation procedure [2, 55] for transition probabilities to the case of the JMLSs. Our method is fundamentally different from the algorithms given in [1] in that we consider the transition probabilities as deterministic quantities (whereas [1] considers them to be random variables).

The chapter is organized as follows. In Sec. 4.2, the problem definition

[1] This type of systems, as introduced in [52], are different than JMLSs although some researchers use a similar phrase "switching systems" for JMLSs (e.g. [53]).

and the method of solution are given. The RKL algorithm is applied to the case of JMLSs in Sec. 4.3. The performance of the algorithm is illustrated on two examples with Monte Carlo runs in Sec. 4.4. The chapter is finalized with conclusions in Sec. 4.5.

4.2 Problem Definition and Solution Methodology

The following JMLS model is considered

$$x_{k+1} = A(r_{k+1})x_k + B(r_{k+1})w_{k+1}, \tag{4.1}$$

$$y_k = C(r_k)x_k + D(r_k)v_k \tag{4.2}$$

where

- $\{x_k\}$ is the continuous-valued base-state sequence with initial distribution

$$x_0 \sim \mathcal{N}(x_0; \bar{x}_0, \Sigma_0), \tag{4.3}$$

  where the notation $\mathcal{N}(x; \bar{x}, \Sigma)$ stands for a Gaussian probability density function for dummy variable x which has a mean $\bar{x}$ and covariance Σ.

- $\{r_k\}$ is the unknown discrete-valued modal-state sequence,

- $\{y_k\}$ is the noisy observation sequence,

- $\{w_k\}$ is a white process noise sequence with distribution,

$$w_k \sim \mathcal{N}(w_k; 0, Q_k), \tag{4.4}$$

- $\{v_k\}$ is a white measurement noise sequence independent from the process noise w_k with distribution

$$v_k \sim \mathcal{N}(v_k; 0, R_k). \tag{4.5}$$

The discrete-valued modal-state $r_k \in \{1, 2, \ldots, N\}$ is assumed to be a first-order finite-state homogenous Markov chain with fixed but unknown transition

probability matrix $\Pi = [\pi_{ij}]$. The basic variables w_k, v_k, x_0 and the modal-state sequence r_k are assumed to be mutually independent for all k. The time-varying matrices $A(r_k)$, $B(r_k)$, $C(r_k)$, and $D(r_k)$ are assumed to be known for each value of r_k.

Notes about the Notation: In the following,

- The capital letters with superscripts will denote sequences with the superscript corresponding to the index of the last element i.e.,

$$Y^n \triangleq \{y_1, y_2, \ldots, y_n\}, \tag{4.6}$$

$$\widehat{\Theta}^n \triangleq \{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n\}. \tag{4.7}$$

- The estimates are shown by hats over the letters.

- The iterations for the estimated quantities are shown by subscripts or by parenthesized superscripts for quantities already subscripted (e.g., $\hat{\theta}_n$ or $\hat{\pi}_{ij}^{(n)}$).

- The individual elements of vector or matrix quantities are denoted by subscripted parentheses (e.g., When x is a vector, its ith element is shown by $(x)_i$. When Σ is a matrix, the element of it corresponding to its ith row and jth column is denoted as $(\Sigma)_{i,j}$.)

- All the probability density functions involved in the equations are assumed to exist.

- The parametric probability density functions and expected value operations with θ being the vector of parameters are denoted using the parameter value as a given condition (e.g., $p(x|\theta)$ and $E[x|\theta]$) although the parameters are assumed to be deterministic quantities. A more complicated example is shown in Eqn. 4.8.

$$E\Big[p(x|\theta)\Big|\theta'\Big] \triangleq \int p(x|\theta)p(x|\theta')dx. \tag{4.8}$$

- Note that when the sequences of random variables (e.g., Y^n) appear in the given conditions of expectations or densities, they should more rigorously denote the σ-algebras generated by them (e.g., $\mathcal{Y}_n \triangleq \sigma(Y^n)$). However, in this chapter, for the sake of simplicity, we will not differentiate between these two quantities.

We are to find an online update mechanism for estimating the unknown fixed transition probabilities of the JMLS defined above which can work coupled with a conventional state estimator such as GPB or IMM algorithm.

4.2.1 Solution Methodology

The RKL approach introduced, in general, by [59] and applied later to HMMs by [2], achieves the estimation of the unknown parameters by minimizing recursively the Kullback-Leibler divergence [27] (called as relative entropy by information theorists)

$$C_n(\theta) = E\Big[\log \frac{p(Y^n|\theta_0)}{p(Y^n|\theta)}\Big|\theta_0\Big] \triangleq \int \log\left(\frac{p(Y^n|\theta_0)}{p(Y^n|\theta)}\right) p(Y^n|\theta_0)dY^n \tag{4.9}$$

between the likelihood function $p(Y^n|\theta)$ of the unknown parameters θ and the true likelihood $p(Y^n|\theta_0)$. The parameter estimate $\hat{\theta}_n^{RKL}$ is therefore given by

$$\hat{\theta}_n^{RKL} = \arg\min_{\theta} C_n(\theta). \tag{4.10}$$

Minimizing the cost function given in Eqn. 4.9 is equivalent to maximizing the reward function $J_n(\theta)$ given as

$$J_n(\theta) = E[\log p(Y^n|\theta)|\theta_0]. \tag{4.11}$$

The likelihood function $p(Y^n|\theta)$ is extremely difficult (if not impossible) to calculate for the JMLSs like the case in many other applications. Therefore, we are going to consider Y^n as the known (incomplete) part of the complete data $K^n = \{Y^n, X^n, R^n\}$. The same incomplete data approach is used in [18] and [12] to calculate off-line and online base-state estimates using the

expectation maximization procedure respectively. We can obtain using the Bayes rule that

$$p(Y^n|\theta) = \frac{p(Y^n, X^n, R^n|\theta)}{p(X^n, R^n|Y^n, \theta)} \tag{4.12}$$

which gives

$$\log p(Y^n|\theta) = \log p(Y^n, X^n, R^n|\theta) - \log p(X^n, R^n|Y^n, \theta). \tag{4.13}$$

Taking expected values of both sides given Y^n at a fixed parameter value θ',

$$\begin{aligned} \log p(Y^n|\theta) &= E[\log p(Y^n, X^n, R^n|\theta)|Y^n, \theta'] \\ &- E[\log p(X^n, R^n|Y^n, \theta)|Y^n, \theta'] \end{aligned} \tag{4.14}$$

where the left hand side of Eqn. 4.13 is taken immediately out of the expectation since the likelihood $p(Y^n|\theta)$ is deterministic with the given information in the expectation. Defining the functions $Q_n(\theta, \theta', Y^n)$ and $P_n(\theta, \theta', Y^n)$ as

$$Q_n(\theta, \theta', Y^n) \triangleq E[\log p(Y^n, X^n, R^n|\theta)|Y^n, \theta'], \tag{4.15}$$

$$P_n(\theta, \theta', Y^n) \triangleq E[\log p(X^n, R^n|Y^n, \theta)|Y^n, \theta'], \tag{4.16}$$

we can express $\log p(Y^n|\theta)$ as follows.

$$\log p(Y^n|\theta) = Q_n(\theta, \theta', Y^n) - P_n(\theta, \theta', Y^n). \tag{4.17}$$

The reward function $J_n(\theta)$ is then given by

$$J_n(\theta) = E[Q_n(\theta, \theta', Y^n)|\theta_0] - E[P_n(\theta, \theta', Y^n)|\theta_0] \tag{4.18}$$

$$\triangleq \overline{Q}_n(\theta, \theta') - \overline{P}_n(\theta, \theta'). \tag{4.19}$$

Note that by Jensen's inequality [27],

$$P_n(\theta, \theta', Y^n) \leq P_n(\theta', \theta', Y^n) \quad \forall Y^n \tag{4.20}$$

and therefore,

$$\overline{P}_n(\theta, \theta') \leq \overline{P}_n(\theta', \theta'). \tag{4.21}$$

Using this, we can say that

$$\begin{aligned} J_n(\theta') &= \overline{Q}_n(\theta', \theta') - \overline{P}_n(\theta', \theta') && (4.22) \\ &\leq \overline{Q}_n(\theta', \theta') - \overline{P}_n(\theta, \theta') && (4.23) \\ &= J_n(\theta) + \overline{Q}_n(\theta', \theta') - \overline{Q}_n(\theta, \theta') && (4.24) \end{aligned}$$

which results in the following fact.

$$\overline{Q}_n(\theta, \theta') \geq \overline{Q}_n(\theta', \theta') \text{ implies } J_n(\theta) \geq J_n(\theta').$$

An iterative but off-line algorithm maximizing the reward function $J_n(\theta)$ is therefore given as

$$\hat{\theta}_{k+1} = \arg\max_{\theta} \overline{Q}_n(\theta, \hat{\theta}_k). \tag{4.25}$$

Recognizing that the calculation of the function $\overline{Q}_n(\theta, \hat{\theta}_n)$=$E[Q_n(\theta, \hat{\theta}_n, Y^n)|\theta_0]$ requires the knowledge of the true parameter value θ_0, we necessarily must choose a suboptimal recursion. The generally adopted solution to this type of problems is to use $Q_n(\theta, \hat{\theta}_n, Y^n)$ instead of its ensemble average. Therefore, an approximate iterative algorithm for maximizing $J_n(\theta)$ is

$$\hat{\theta}_{k+1} = \arg\max_{\theta} Q_n(\theta, \hat{\theta}_k, Y^n). \tag{4.26}$$

The sequential version of this algorithm becomes

$$\hat{\theta}_{n+1} = \arg\max_{\theta} Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1}) \tag{4.27}$$

which avoids the reprocessing of all the accumulating data when a new estimate $\hat{\theta}_n$ is obtained. In the cases where the analytical maximization of $Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ is not possible, one can use the so-called stochastic approximation type algorithms [28, 60]. These algorithms were introduced to find the zeros of the functions which are unknown but whose noise corrupted observations can be obtained [61]. The stochastic approximation algorithm that we will use is intended to find the zeros of the gradient of the reward function, i.e., the zeros of $\frac{\partial}{\partial\theta} Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$. This type of algorithms is generally called in the field as the stochastic gradient algorithms. The recursion of the algorithm is as follows.

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \epsilon_n G_{n+1}(\hat{\theta}_n, \widehat{\Theta}^n, Y^{n+1}) \tag{4.28}$$

where ϵ_n is a sequence of small scalar positive gains and $G_{n+1}(\hat{\theta}_n, \widehat{\Theta}^n, Y^{n+1})$ is a (possibly noisy) measurement of the gradient of $Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ with respect to the unknown parameter θ evaluated at the last parameter estimate $\hat{\theta}_n$ i.e.,

$$G_{n+1}(\hat{\theta}_n, \widehat{\Theta}^n, Y^{n+1}) = \frac{\partial}{\partial\theta} Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})\Big|_{\theta=\hat{\theta}_n} + \Delta W_{n+1}(\widehat{\Theta}^n, Y^{n+1}) \quad (4.29)$$

where $\Delta W_{n+1}(\widehat{\Theta}^n, Y^{n+1})$ is the parameter dependent noise term in the measurement of the gradient. The almost-sure and some other weaker types of convergence for the algorithm to the true parameter value θ_0 is guaranteed if some conditions on the step-size sequence ϵ_n and the measurement noise $\Delta W_{n+1}(\widehat{\Theta}^n, Y^{n+1})$ are satisfied [28, 60]. In the cases where constraints exist on the parameters θ, the updated parameters are usually projected onto the constraint surface i.e.,

$$\hat{\theta}_{n+1} = \mathcal{P}\{\hat{\theta}_n + \epsilon_n G_{n+1}(\hat{\theta}_n, \widehat{\Theta}^n, Y^{n+1})\} \quad (4.30)$$

where $\mathcal{P}$ denotes the projection operator related with the constraint surface (manifold) which returns the nearest element in the constraint surface to the updated parameter value. When the unknown parameter vector θ is composed of the elements of the probability transition matrix of the JMLS, i.e.,

$$\theta \;=\; [\pi_{11}, \pi_{12}, \ldots, \pi_{1N}, \pi_{21}, \pi_{22}, \ldots, \pi_{2N}, \ldots\ldots\ldots, \pi_{N1}, \pi_{N2}, \ldots, \pi_{NN}]^T$$

or in a more compact form,

$$(\theta)_{N(i-1)+j} = \pi_{ij} \quad 1 \leq i, j \leq N, \quad (4.31)$$

the following constraint set is in effect.

$$\sum_{j=1}^{N} \pi_{ij} = 1 \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad \pi_{ij} \geq 0 \quad (4.32)$$

for $1 \leq i, j \leq N$. These constraints define N standard N-simplices. At the end of each parameter update, each updated set of probabilities $\{\pi_{ij}\}_{j=1}^{N}$ must be projected onto the standard simplex $\mathcal{S}_i$ defined by the constraints

$$\sum_{j=1}^{N} \pi_{ij} = 1 \quad \text{and} \quad \pi_{ij} \geq 0 \quad \text{for} \quad j = 1, \ldots, N. \quad (4.33)$$

The problem of finding the projection of a vector onto a standard simplex can be formulated as a standard quadratic programming problem and the projection can be found in at most N steps. Throughout the chapter, we are going to call the projection operator of a standard N-simplex as $\mathcal{P}$ and, with an abuse of notation, we are going to denote the projected probability value π_{ij}^* obtained after the set of updated probabilities $\{\pi_{ij}\}_{j=1}^N$ is projected onto the corresponding simplex $\mathcal{S}_i$ with

$$\pi_{ij}^* = \mathcal{P}\{\pi_{ij}\}. \tag{4.34}$$

A description of the recursive method used for obtaining the projections for our algorithm is given in App. C. See, for example, Sec. 2.6.2 of [62] for a general solution of the problem of obtaining projections onto a simplex.

4.3 Derivation of the Algorithm

In this section, we are going to adapt the RKL procedure described in Sec. 4.2 to the problem of estimating the transition probabilities associated with a JMLS. For this purpose, we maximize the function $Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ which is given as

$$Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1}) = E\left[\log p(K^{n+1}|\theta) \,\middle|\, Y^{n+1}, \widehat{\Theta}^n\right] \tag{4.35}$$

where the likelihood $p(K^{n+1}|\theta)$ of the complete data set K^{n+1} for a JMLS has the following expansion.

$$\begin{aligned} p(K^{n+1}|\theta) &\triangleq p(Y^{n+1}, X^{n+1}, R^{n+1}|\theta) && (4.36)\\ &= p(y_{n+1}|x_{n+1}, r_{n+1})p(x_{n+1}|x_n, r_{n+1})P(r_{n+1}|r_n, \theta) \\ &\quad \times p(Y^n, X^n, R^n|\theta). && (4.37) \end{aligned}$$

The log-likelihood $\log p(K^{n+1}|\theta)$ is simply

$$\begin{aligned} \log p(K^{n+1}|\theta) &= \log\Big(p(y_{n+1}|x_{n+1}, r_{n+1})p(x_{n+1}|x_n, r_{n+1})P(r_{n+1}|r_n, \theta)\Big) \\ &\quad + \log p(Y^n, X^n, R^n|\theta). && (4.38) \end{aligned}$$

Defining the the log terms L_1 and L_2 as

$$\begin{aligned} L_1(y_{n+1}, x_{n+1}, x_n, r_{n+1}, r_n|\theta) &\triangleq \log\Big(p(y_{n+1}|x_{n+1}, r_{n+1})p(x_{n+1}|x_n, r_{n+1}) \\ &\quad \times P(r_{n+1}|r_n, \theta)\Big), \qquad (4.39) \\ L_2(Y^n, X^n, R^n|\theta) &\triangleq \log p(Y^n, X^n, R^n|\theta), \qquad (4.40) \end{aligned}$$

the function $Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ is written as,

$$Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1}) = \mathbb{E}_1 + \mathbb{E}_2 \qquad (4.41)$$

where

$$\begin{aligned} \mathbb{E}_1 &\triangleq E\left[L_1(y_{n+1}, x_{n+1}, x_n, r_{n+1}, r_n|\theta)\Big|Y^{n+1}, \widehat{\Theta}^n\right], \qquad (4.42) \\ \mathbb{E}_2 &\triangleq E\left[L_2(Y^n, X^n, R^n|\theta)\Big|Y^{n+1}, \widehat{\Theta}^n\right]. \qquad (4.43) \end{aligned}$$

4.3.1 Calculation of the Expectation $\mathbb{E}_1$

The expectation $\mathbb{E}_1$ is by definition,

$$\begin{aligned} \mathbb{E}_1 &= \int\int \sum_i \sum_j L_1(y_{n+1}, x_{n+1}, x_n, j, i|\theta) \\ &\quad \times p(x_{n+1}, x_n, r_{n+1} = j, r_n = i|Y^{n+1}, \widehat{\Theta}^n) dx_{n+1} dx_n \qquad (4.44) \end{aligned}$$

where L_1 can be written as follows.

$$\begin{aligned} L_1(y_{n+1}, x_{n+1}, x_n, j, i|\theta) &= \log p(y_{n+1}|x_{n+1}, r_{n+1} = j) \\ &\quad + \log p(x_{n+1}|x_n, r_{n+1} = j) + \log \pi_{ij}. \qquad (4.45) \end{aligned}$$

The integrals in Eqn. 4.44 can be expanded as the sum of three terms as

$$\begin{aligned} \mathbb{E}_1 &= \int \sum_j \log\left(p(y_{n+1}|x_{n+1}, r_{n+1} = j)\right) \\ &\quad \times p(x_{n+1}, r_{n+1} = j|Y^{n+1}, \widehat{\Theta}^n) dx_{n+1} \\ &\quad + \int\int \sum_j \log\left(p(x_{n+1}|x_n, r_{n+1} = j)\right) \\ &\quad \times p(x_{n+1}, x_n, r_{n+1} = j|Y^{n+1}, \widehat{\Theta}^n) dx_n dx_{n+1} \\ &\quad + \sum_i \sum_j \log(\pi_{ij}) P(r_{n+1} = j, r_n = i|Y^{n+1}, \widehat{\Theta}^n). \qquad (4.46) \end{aligned}$$

In the sequential identification process, we are only interested in the derivative of the reward function $Q_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ with respect to the unknown parameters π_{ij} which constitute the parameter vector θ. Only the third term on the right hand side of Eqn. 4.46 contributes to the related derivative which gives the result,

$$\frac{\partial \mathbb{E}_1}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_i \sum_j \log(\pi_{ij}) P(r_{n+1} = j, r_n = i | Y^{n+1}, \widehat{\Theta}^n). \tag{4.47}$$

Then, it easily follows that,

$$\left(\frac{\partial \mathbb{E}_1}{\partial \theta} \right)_{N(i-1)+j} \triangleq \frac{\partial \mathbb{E}_1}{\partial \pi_{ij}} = \frac{1}{\pi_{ij}} P(r_{n+1} = j, r_n = i | Y^{n+1}, \widehat{\Theta}^n) \tag{4.48}$$

for $1 \leq i, j \leq N$.

4.3.2 Calculation of the Expectation $\mathbb{E}_2$

The second expectation $\mathbb{E}_2$ is by definition,

$$\mathbb{E}_2 = \sum_{R^n} \int \log(p(Y^n, X^n, R^n|\theta)) p(X^n, R^n | Y^{n+1}, \widehat{\Theta}^n) dX^n. \tag{4.49}$$

The density function $p(X^n, R^n | Y^{n+1}, \widehat{\Theta}^n)$ can be written using Bayes rule as

$$\begin{aligned} p(X^n, R^n | Y^{n+1}, \widehat{\Theta}^n) &= \frac{p(y_{n+1} | X^n, R^n, Y^n, \widehat{\Theta}^n)}{p(y_{n+1} | Y^n, \widehat{\Theta}^n)} p(X^n, R^n | Y^n, \widehat{\Theta}^n) & (4.50) \\ &= \frac{p(y_{n+1} | x_n, r_n, \hat{\theta}_n)}{p(y_{n+1} | Y^n, \widehat{\Theta}^n)} p(X^n, R^n | Y^n, \widehat{\Theta}^{n-1}) & (4.51) \end{aligned}$$

where Eqn. 4.51 is written by dropping the redundant terms in the given conditions of Eqn. 4.50. Substituting Eqn. 4.51 into Eqn. 4.49,

$$\begin{aligned} \mathbb{E}_2 &= \sum_{R^n} \int \log(p(Y^n, X^n, R^n|\theta)) \frac{p(y_{n+1} | x_n, r_n, \hat{\theta}_n)}{p(y_{n+1} | Y^n, \widehat{\Theta}^n)} p(X^n, R^n | Y^n, \widehat{\Theta}^{n-1}) dX^n \\ &= \sum_{R^n} \int \log(p(Y^n, X^n, R^n|\theta)) p(X^n, R^n | Y^n, \widehat{\Theta}^{n-1}) dX^n \\ &\quad + \sum_{R^n} \int \log(p(Y^n, X^n, R^n|\theta)) \left(\frac{p(y_{n+1} | x_n, r_n, \hat{\theta}_n)}{p(y_{n+1} | Y^n, \widehat{\Theta}^n)} - 1 \right) & (4.52) \\ &\quad \times p(X^n, R^n | Y^n, \widehat{\Theta}^{n-1}) dX^n \\ &= Q_n(\theta, \widehat{\Theta}^{n-1}, Y^n) + \Delta Q_n(y_{n+1}, \theta, \widehat{\Theta}^n) & (4.53) \end{aligned}$$

where

$$\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n) \triangleq \sum_{R^n}\int \log\left(p(Y^n,X^n,R^n|\theta)\right)\left(\frac{p(y_{n+1}|x_n,r_n,\hat{\theta}_n)}{p(y_{n+1}|Y^n,\widehat{\Theta}^n)}-1\right) \times p(X^n,R^n|Y^n,\widehat{\Theta}^{n-1})dX^n.$$

In the stochastic approximation algorithm, we are interested in the derivative of the expectation $\mathbb{E}_2$ evaluated at the current model value $\hat{\theta}_n$ which is

$$\frac{\partial \mathbb{E}_2}{\partial\theta}\Big|_{\theta=\hat{\theta}_n} = \frac{\partial Q_n(\theta,\widehat{\Theta}^{n-1},Y^n)}{\partial\theta}\Big|_{\theta=\hat{\theta}_n} + \frac{\partial}{\partial\theta}\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)\Big|_{\theta=\hat{\theta}_n}. \quad (4.54)$$

The function $Q_n(\theta,\widehat{\Theta}^{n-1},Y^n)$ is assumed to be maximized by the previous iteration and therefore, the first term on the right hand side of Eqn. 4.54 is approximately zero, i.e.,

$$\frac{\partial Q_n(\theta,\widehat{\Theta}^{n-1},Y^n)}{\partial\theta}\Big|_{\theta=\hat{\theta}_n} \approx 0. \quad (4.55)$$

The assumption in Eqn.4.55, which is rarely true, is a generally adopted one in practice to generate recursive identification algorithms [63].

Defining the sequence ΔM_n as

$$\Delta M_n \triangleq \frac{\partial}{\partial\theta}\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)\Big|_{\theta=\hat{\theta}_n}, \quad (4.56)$$

the partial derivative $\frac{\partial\mathbb{E}_2}{\partial\theta}$ evaluated at the current parameter estimate $\hat{\theta}_n$ is approximately given as

$$\frac{\partial \mathbb{E}_2}{\partial\theta}\Big|_{\theta=\hat{\theta}_n} \approx \Delta M_n. \quad (4.57)$$

Remark 4.1 *Note here that,*

$$E[\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)|Y^n,\widehat{\Theta}^n]$$

$$\triangleq \int \Delta Q_n(y_{n+1})p(y_{n+1}|Y^n,\widehat{\Theta}^n)dy_{n+1} \quad (4.58)$$

$$= \sum_{R^n}\int \log\left(p(Y^n,X^n,R^n|\theta)\right) \times \underbrace{\int p(y_{n+1}|x_n,r_n,\hat{\theta}_n) - p(y_{n+1}|Y^n,\widehat{\Theta}^n)dy_{n+1}}_{=0} \times p(X^n,R^n|Y^n,\widehat{\Theta}^{n-1})dX^n = 0. \quad (4.59)$$

Actually, more can be said about the sequence ΔQ_n in the following way,

$$E[\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)|\Delta Q^{n-1}] \\ = E\Big[\underbrace{E[\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)|Y^n,\widehat{\Theta}^n]}_{=0}\Big|\Delta Q^{n-1}\Big] = 0 \quad (4.60)$$

almost surely where $\Delta Q^{n-1} \triangleq \{\Delta Q_1, \Delta Q_2, \ldots, \Delta Q_{n-1}\}$. The first equality in Eqn. 4.60 can be written due to the fact that ΔQ^{n-1} is $\sigma(Y^n)$-measurable where $\sigma(Y^n)$ denotes the σ-algebra generated by Y^n. As a result, the random sequence ΔQ_n is a martingale difference.

Under the assumption that the partial derivative with respect to the unknown parameters θ and the integration corresponding to expectation can be interchanged,

$$\begin{aligned} E[\Delta M_n|\Delta M^{n-1}] &= E\Big[\frac{\partial}{\partial\theta}\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)\Big|_{\theta=\hat{\theta}_n}\Big|\Delta M^{n-1}\Big] \\ &= \frac{\partial}{\partial\theta}E\Big[\underbrace{E[\Delta Q_n(y_{n+1},\theta,\widehat{\Theta}^n)|Y^n,\widehat{\Theta}^n]}_{=0}\Big|\Delta M^{n-1}\Big]\Big|_{\theta=\hat{\theta}_n} \\ &= 0 \quad a.s. \end{aligned}$$

Consequently, the sequence ΔM_n is also a martingale difference.

4.3.3 RKL Recursions

Combining Eqn. 4.48 and Eqn. 4.57, the elements of the gradient of the reward function $Q_{n+1}(\theta,\widehat{\Theta}^n,Y^{n+1})$ is written as

$$\Big(\frac{\partial}{\partial\theta}Q_{n+1}(\theta,\widehat{\Theta}^n,Y^{n+1})\Big)_{N(i-1)+j} \\ = \frac{P(r_{n+1}=j,r_n=i|Y^{n+1},\widehat{\Theta}^n)}{\pi_{ij}} + (\Delta M_n)_{N(i-1)+j} \quad (4.61)$$

for $1 \leq i,j \leq N$. Thus, the first term on the right hand side of Eqn. 4.61 can be used as a measurement (in a martingale difference noise) of the left hand side, i.e.,

$$\Big(G_{n+1}(\theta,\widehat{\Theta}^n,Y^{n+1})\Big)_{N(i-1)+j} = \frac{1}{\pi_{ij}}P(r_{n+1}=j,r_n=i|Y^{n+1},\widehat{\Theta}^n). \quad (4.62)$$

The martingale difference characteristics of the noise term ΔM_n (See Remark 4.1) is a nice property which is generally used in proving convergence of this type of algorithms [28]. Nevertheless, the complexity involved in the algorithm makes it difficult to reach a direct conclusion on the algorithm's convergence using the existing results. We are therefore going to address the convergence issue in this document only through simulation results.

The noisy gradient measurement $G_{n+1}(\theta, \widehat{\Theta}^n, Y^{n+1})$ evaluated at the current parameter estimate $\hat{\theta}_n$ is finally given as

$$\Big(G_{n+1}(\hat{\theta}_n, \widehat{\Theta}^n, Y^{n+1})\Big)_{N(i-1)+j} = \frac{1}{\hat{\pi}_{ij}^{(n)}} P(r_{n+1} = j, r_n = i | Y^{n+1}, \widehat{\Theta}^n) \quad (4.63)$$

for $1 \leq i, j \leq N$ where $\hat{\pi}_{ij}^{(n)} = (\hat{\theta}_n)_{N(i-1)+j}$.

Remark 4.2 *Note here that the partial derivatives given in Eqn. 4.63 are always nonnegative. This stems from the fact that the constraints are not considered in the cost function. The constraints, therefore, will play an essential role for the stability of the algorithm.*

Adding the projection operation related with the constraints to each recursion of the algorithm, the update rule for the transition probability estimate $\hat{\pi}_{ij}^{(n)}$ is given as

$$\hat{\pi}_{ij}^{(n+1)} = \mathcal{P}\Big\{\hat{\pi}_{ij}^{(n)} + \epsilon_n \frac{P(r_{n+1} = j, r_n = i | Y^{n+1}, \widehat{\Theta}^n)}{\hat{\pi}_{ij}^{(n)}}\Big\} \quad (4.64)$$

for $1 \leq i, j \leq N$.

Remark 4.3 *Defining the updated non-projected parameters $\tilde{\pi}_{ij}^{(n+1)}$ as*

$$\tilde{\pi}_{ij}^{(n+1)} \triangleq \hat{\pi}_{ij}^{(n)} + \epsilon_n \frac{1}{\hat{\pi}_{ij}^{(n)}} P(r_{n+1} = j, r_n = i | Y^{n+1}, \widehat{\Theta}^n), \quad (4.65)$$

we see that they are always non-negative and satisfy the inequality

$$\sum_{j=1}^{N} \tilde{\pi}_{ij}^{(n+1)} \geq 1 \quad \textit{for} \quad i = 1, 2, \ldots, N. \quad (4.66)$$

We give a simple method of projection for the set of non-negative quantities $\Big\{\tilde{\pi}_{ij}^{(n+1)}\Big\}_{j=1}^{N}$ satisfying this inequality in App. C.

4.3.4 Calculation of the Probabilities

In order to be able to calculate the probabilities $P(r_{n+1}=j, r_n=i|Y^{n+1}, \widehat{\Theta}^n)$, we are going to assume that an online conventional state estimator like IMM or GPB algorithm which uses the probability transition matrix estimates $\widehat{\Theta}^n$ of the RKL algorithm is also run on the measurements Y^n . The conventional state estimator is required to supply the mode-conditioned state estimates

$$\hat{x}^i_{n|n} \triangleq E[x_n|Y^n, r_n=i, \widehat{\Theta}^n] \quad \text{for} \quad i=1\ldots N, \tag{4.67}$$

covariances

$$P^i_{n|n} \triangleq E[(x_n-\hat{x}^i_{n|n})(x_n-\hat{x}^i_{n|n})^T|Y^n, r_n=i, \widehat{\Theta}^n] \tag{4.68}$$

for $i=1\ldots N$ and the mode probabilities

$$\mu_i(n) = P\{r_n=i|Y^n, \widehat{\Theta}^n\}. \tag{4.69}$$

Using these quantities, we can calculate the required probabilities approximately as

$$\begin{aligned}
P(r_{n+1}=j, r_n=i|Y^{n+1}, \widehat{\Theta}^n) &= \frac{p(y_{n+1}|r_{n+1}=j, r_n=i, Y^n, \widehat{\Theta}^n)}{p(y_{n+1}|Y^n, \widehat{\Theta}^n)} \\
&\quad \times P(r_{n+1}=j|r_n=i, \widehat{\Theta}^n) P(r_n=i|Y^n, \widehat{\Theta}^n) \\
&\approx \frac{p(y_{n+1}|r_{n+1}=j, \hat{x}^i_{n|n}, P^i_{n|n})}{p(y_{n+1}|Y^n, \widehat{\Theta}^n)} \\
&\quad \times P(r_{n+1}=j|r_n=i, \widehat{\Theta}^n) P(r_n=i|Y^n, \widehat{\Theta}^n) \\
&= \frac{p(y_{n+1}|r_{n+1}=j, \hat{x}^i_{n|n}, P^i_{n|n}) \hat{\pi}^{(n)}_{ij} \mu_i(n)}{p(y_{n+1}|Y^n, \widehat{\Theta}^n)} \qquad (4.70) \\
&= \frac{\mathcal{N}(y_{n+1}; \hat{y}^{ij}_{n+1|n}, S^{ij}_{n+1|n}) \hat{\pi}^{(n)}_{ij} \mu_i(n)}{\sum_j \sum_i \mathcal{N}(y_{n+1}; \hat{y}^{ij}_{n+1|n}, S^{ij}_{n+1|n}) \hat{\pi}^{(n)}_{ij} \mu_i(n)} \qquad (4.71)
\end{aligned}$$

where

$$\begin{aligned}
\hat{y}^{ij}_{n+1|n} &\triangleq C(j)A(j)\hat{x}^i_{n|n} \\
S^{ij}_{n+1|n} &\triangleq C(j)\big(A(j)P^i_{n|n}A^T(j) + B(j)Q_kB(j)^T\big)C^T(j) + D(j)R_kD(j)^T.
\end{aligned}$$

Substituting Eqn. 4.71 into Eqn. 4.64, the update rule for the transition probabilities becomes

$$\hat{\pi}_{ij}^{(n+1)} = \mathcal{P}\Big\{\hat{\pi}_{ij}^{(n)} + \epsilon_n \frac{\mathcal{N}(y_{n+1}; \hat{y}_{n+1|n}^{ij}, S_{n+1|n}^{ij})\mu_i(n)}{\sum_j \sum_i \mathcal{N}(y_{n+1}; \hat{y}_{n+1|n}^{ij}, S_{n+1|n}^{ij})\hat{\pi}_{ij}^{(n)}\mu_i(n)}\Big\} \tag{4.72}$$

where the projection operator $\mathcal{P}$ is described in Sec. 4.2.1.

4.3.5 Selection of the Step-Size Sequence ϵ_n

The choice of the step-size sequence ϵ_n is an important issue for obtaining a satisfactory performance with this type of algorithms. The sufficient conditions on the step-sizes for convergence are

$$\sum_{n=0}^{\infty} \epsilon_n = \infty \quad \text{and} \quad \sum_{n=0}^{\infty} \epsilon_n^2 < \infty \tag{4.73}$$

which guarantee a decrease in the value of step-size ϵ_n which is fast enough to allow sufficient averaging of the noisy gradient measurements and slow enough to avoid premature convergence of the algorithm. In the cases where one needs the algorithm to track small changes in the parameter values, constant and sufficiently small step-sizes are used [28, 60]. Some adaptive step-size sequence selection mechanisms exist in the literature [60] but we are going to choose constant step-sizes during the simulations for the sake of simplicity.

4.4 Simulation Results

This section evaluates the performance of the recursive Kullback-Leibler method by means of computer simulations. For this purpose, we are going to use two examples which illustrate the capabilities of the RKL algorithm we have derived in the previous section.

4.4.1 Example 1

In order to evaluate the performance of our method, in this sub-section, we are going to use the same example and methodology as in [1] which is related

with a system with failures. Consider the scalar dynamical system given by

$$x_{k+1} = x_k + w_k, \tag{4.74}$$

$$y_k = (r_k - 1)x_k + (100 - 90(r_k - 1))v_k \tag{4.75}$$

where $x_0 \sim \mathcal{N}(x_0; 0, 20^2)$, $w_k \sim \mathcal{N}(w_k; 0, 2^2)$, $v_k \sim \mathcal{N}(v_k; 0, 1)$ with x_0, w_k and v_k being mutually independent for $k = 1, 2, \ldots$. The model sequence $r_k \in \{1, 2\}$ is a first-order, two-state, homogeneous Markov process with probability transition matrix $\Pi = [\pi_{ij}]$ given as

$$\Pi = \begin{bmatrix} 0.6 & 0.4 \\ 0.85 & 0.15 \end{bmatrix}. \tag{4.76}$$

Note that this system corresponds to a system with frequent measurement failures with the modal-state $r_k = 1$ corresponding to the case of the failure. Three IMM algorithms were implemented to estimate the state of this system.

- The first IMM algorithm is an exact one which uses the true probability transition matrix of the system given in Eqn. 4.76.

- The second IMM algorithm is a non-adaptive one which uses a typical probability transition matrix with $\pi_{11} = \pi_{22} = 0.9$.

- The third IMM algorithm is an adaptive one which estimates the probability transition matrix of the underlying Markov chain using the recursive Kullback-Leibler method with the constant step-size sequence $\epsilon_n = 0.02$.

The algorithms use equal initial mode-probabilities (i.e., $\mu_i(0) = 0.5$ for $i = 1, 2$) and therefore assume no knowledge of the initial state distribution for the finite-state Markov chain. In order to obtain ensemble averages, 1000 Monte Carlo runs are performed. In each run,

- The underlying Markov chain is simulated using the true probability transition matrix given in Eqn. 4.76. The true mode-states are obtained.

The initial state distribution of the Markov chain is randomly created assuming that all possible state distributions are equally likely.

- The true base-state of the system described by Eqn. 4.74 is simulated using an artificially generated process noise sequence $\{w_k\}$.
- The measurement sequence of the system described by Eqn. 4.75 is obtained using the true mode-state sequence, base-state sequence and artificially generated measurement noise sequence $\{v_k\}$.
- The IMM algorithms are executed using the measurement sequence.
- The absolute base-state estimation errors of the algorithms are calculated.

The average estimation performance of the recursive Kullback-Leibler method is shown in Fig. 4.1. Note that there exist very small biases in the estimates like the case in [1]. These seem to be due to the approximations in the derivation process of the algorithm and seem to be acceptable. The mean absolute base-state estimation errors of the three IMM filters are shown in Fig. 4.2. The errors of the exact IMM are at the minimum level as expected. The adaptive algorithm fed by the recursive Kullback-Leibler estimation process appears to beat the non-adaptive IMM filter significantly.

4.4.2 Example 2

As a second example, a hypothetical three-model scalar JMLS (i.e., $r_k \in \{1, 2, 3\}$) is used. The parameters of the system are given as

- $A(1) = 0.8$, $A(2) = 0.9$, $A(3) = 1$,
- $B(i) = 1$ and $D(i) = 1$ for $i = 1, 2, 3$,
- $C(1) = 1$, $C(2) = 2$, $C(3) = 4$,
- $x_0 \sim \mathcal{N}(x_0; 0, 2^2)$,

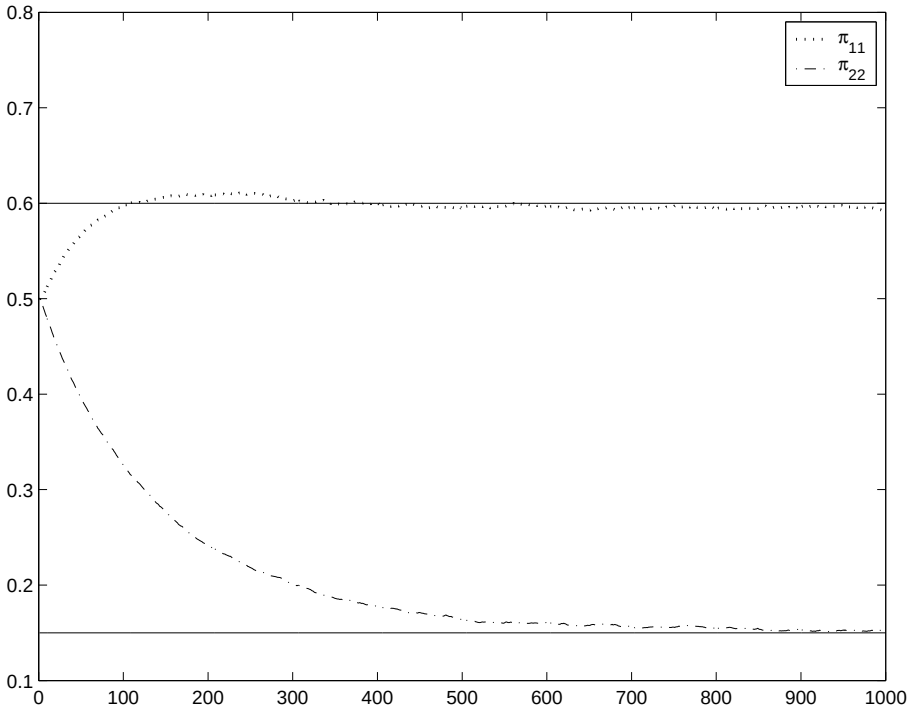

Figure 4.1: The average transition probability estimation performance of the recursive Kullback-Leibler method.

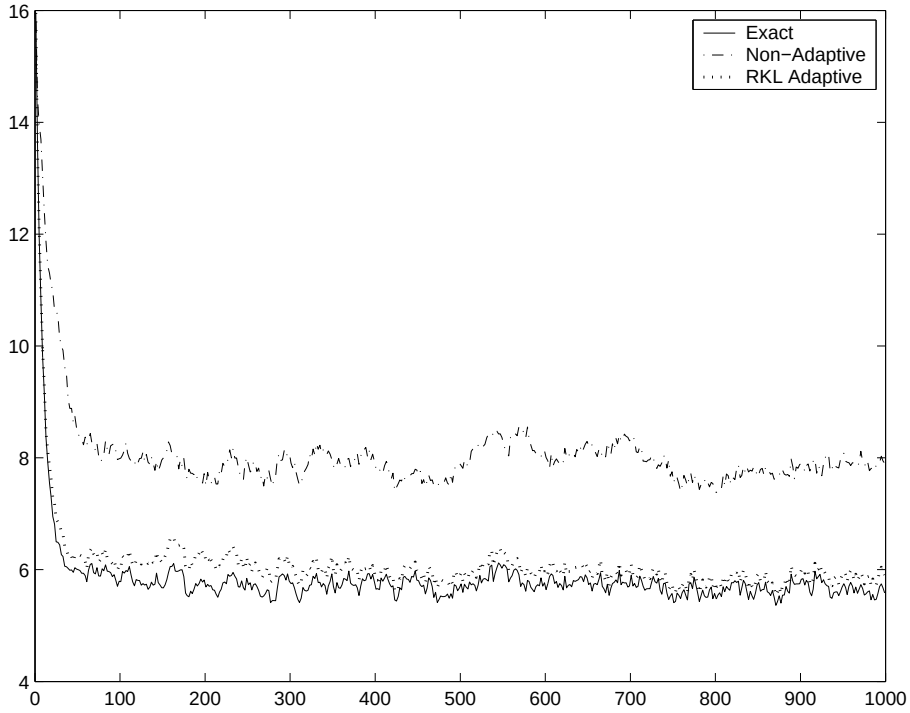

Figure 4.2: The mean-absolute base-state estimation errors of the IMM algorithms.

- $w_k \sim \mathcal{N}(w_k; 0, 2^2)$, $v_k \sim \mathcal{N}(v_k; 0, 1)$.

and the true transition probability associated with the mode sequence r_k is taken as

$$\Pi = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.25 & 0.5 & 0.25 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}. \tag{4.77}$$

The three IMM algorithms which have the same features as in the first example are executed. The constant probability transition matrix of the non-adaptive IMM filter is selected as

$$\Pi_{\text{ Non-Adaptive}} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}. \tag{4.78}$$

The initial probability transition matrix estimate $\widehat{\Pi}_0$ for the RKL algorithm of the adaptive IMM algorithm is taken as

$$\widehat{\Pi}_0 = \begin{bmatrix} 0.33 & 0.33 & 0.34 \\ 0.33 & 0.33 & 0.34 \\ 0.33 & 0.33 & 0.34 \end{bmatrix} \tag{4.79}$$

and the step-size sequence of the RKL algorithm is set to $\epsilon_n = 0.002$. The transition matrix estimation performance of the RKL algorithm is illustrated in Fig.4.3. Small biases again exist in the estimates but the convergence trend is evident.

The mean-absolute base-state estimation errors of the three IMM filters are shown in Fig. 4.4. The errors of the adaptive algorithm fed by the recursive Kullback-Leibler estimation process are very close to those of the exact IMM after a transient period.

4.5 Conclusions

A new method for the estimation of the transition probabilities of the JMLSs is given. Although the derivation of the method is quite complicated, the result-

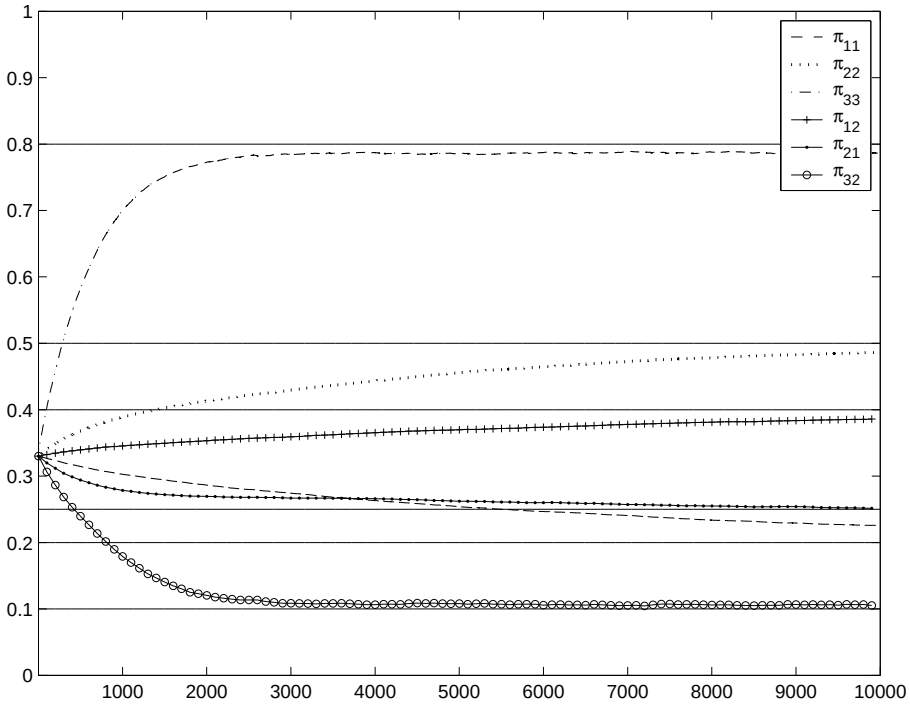

Figure 4.3: The average transition probability estimation performance of the recursive Kullback-Leibler method.

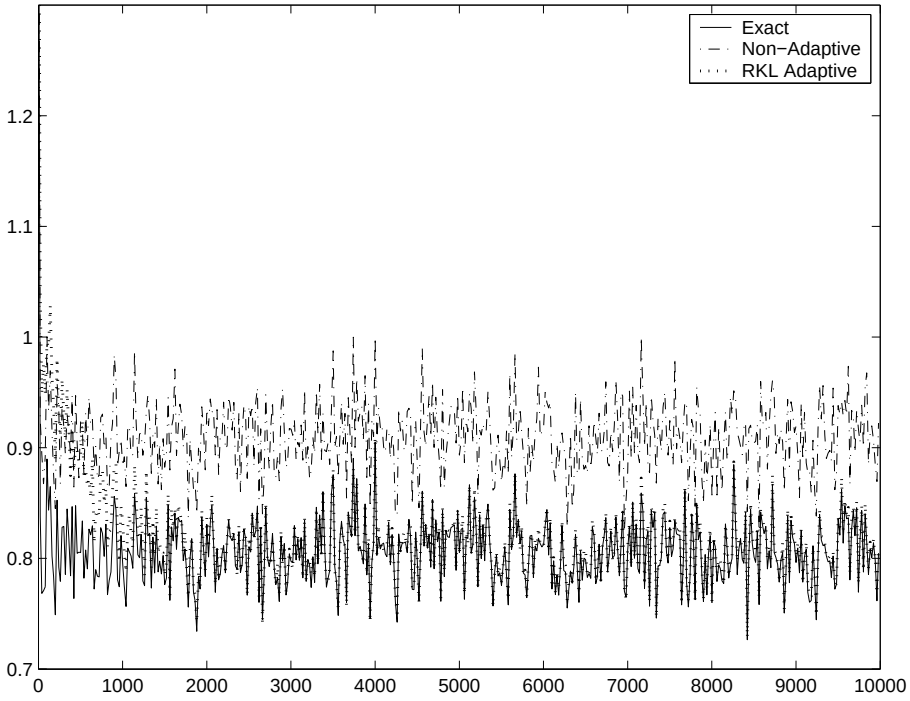

Figure 4.4: Mean-absolute base-state estimation errors of the IMM algorithms.

ing recursions are easy to implement using the outputs of an online multiple-model state estimation algorithm like IMM or GPB approaches. The simulation results show that the algorithm can be an efficient alternative to those described in [1].

CHAPTER 5

MAXIMUM LIKELIHOOD ESTIMATION OF THE TRANSITION PROBABILITIES OF JMLSS

5.1 Introduction

As mentioned in Sec. 4.1, the maximum likelihood (ML) algorithm was one of the first approaches applied to the transition probability estimation of JMLSs. However, the existing algorithms suffers from the following drawbacks:

- They consider only the cases where the unknown transition probability matrices are in a finite set [51].
- They don't apply to JMLSs [52].

In addition to these, while stating the infeasibility of the ML approach for the JMLSs, [51] did not have the comfortable approximation scheme of the IMM filter at hand in 1982. The profitable place that the IMM filter occupies on the performance vs. computation curve makes it a useful candidate to use in the approximations of the infeasible ML algorithms.

In this chapter, using the motivation above, we propose a maximum likelihood estimation based solution to the transition probability estimation problem. Our algorithm is new in that it assumes that the unknown transition probability matrix takes values in a continuous valued set[1] and it is derived

[1] Specifically, this set is composed of the N-simplices considered in Chapter 4.

specifically for the JMLSs. The algorithm can be used either online or offline. We use the reference probability method to derive the algorithm which enables us to identify the approximations required to form a feasible method easily.[2]

The chapter is organized as follows. In Sec. 5.2, the problem definition and the related measure change parameters are given. Application of the ML algorithm to JMLSs is presented in Sec. 5.3 which concludes that the algorithm should calculate the estimates of the number of jumps between the states of the Markov chain recursively. The presentation of the main results is made in Sec. 5.4 where the related recursive calculation procedure is derived. The formulas for final transition probability estimates utilizing the estimates of the number of jumps are given in Sec. 5.5. Sec. 5.6 presents a brief summary of the derived algorithm along with some modifications and implementation issues. Moreover, achieving the base state estimation using the derived algorithm without any additional base-state estimators is also outlined in Sec. 5.6. The transition probability estimation performance of the algorithm is illustrated in Sec. 5.7 using the same examples as the ones in Chapter 4. Also, comparison with the RKL method of Chapter 4 is made in Sec. 5.7 on one of the examples. The chapter is finalized with conclusions in Sec. 5.8.

5.2 Problem Definition

The following JMLS model is considered.

$$x_{k+1} = A(r_{k+1})x_k + B(r_{k+1})w_{k+1}, \tag{5.1}$$

$$y_k = C(r_k)x_k + D(r_k)v_k \tag{5.2}$$

where

- $\{x_k \in \mathbb{R}^n\}$ is the continuous-valued base-state sequence with initial distribution

$$x_0 \sim \mathcal{N}(x_0; \bar{x}_0, \Sigma_0) \tag{5.3}$$

[2] We again note the requirement for the unfamiliar reader to consult App. A for the related background about the reference probability method.

where the notation $\mathcal{N}(x; \bar{x}, \Sigma)$ stands for a Gaussian probability density function for dummy variable x which has a mean $\bar{x}$ and covariance Σ.

- $\{r_k\}$ is the unknown discrete-valued modal-state sequence,
- $\{y_k \in \mathbb{R}^m\}$ is the noisy observation sequence,
- $\{w_k \in \mathbb{R}^n\}$ is a white process noise sequence with distribution

$$w_k \sim \mathcal{N}(w_k; 0, I_n) \tag{5.4}$$

where I_n denotes the identity matrix of size $n \times n$.

- $\{v_k \in \mathbb{R}^m\}$ is a white measurement noise sequence independent from the process noise w_k with distribution

$$v_k \sim \mathcal{N}(v_k; 0, I_m). \tag{5.5}$$

The discrete-valued modal-state $r_k \in \{e_1, e_2, \ldots, e_N\}$ is assumed to be a first-order finite-state homogenous Markov chain with fixed but unknown transition probability matrix $\Pi = [\pi_{ij}]$. Here, the variable $e_j \in \mathbb{R}^N$ denotes the canonical unit vector with unity at the jth position and zeros elsewhere. The basic variables w_k, v_k, x_0 and the modal-state sequence r_k are assumed to be mutually independent for all k. The time-varying matrices $A(r_k)$, $B(r_k)$, $C(r_k)$, and $D(r_k)$ are assumed to be known for each value of r_k. Moreover, the matrices $B(r_k)$ and $D(r_k)$ are assumed to be invertible. This is a requirement of the derivation using the reference probability method.[3]

The aim of the chapter is to find a possibly approximate online ML estimator for the unknown fixed transition probabilities of the JMLS defined above.

5.2.1 Change of Measure

In the derivations given in the subsequent sections, it is initially assumed that we are in an ideal probability space $(\Omega, \mathcal{F}, \overline{P})$ where, under the probability measure $\overline{P}$

[3] See Sec. 3.5.2.3 for how this restriction can be bypassed easily.

- $\{x_k\}$, $k \in \mathbb{N}$ is a sequence of i.i.d. random variables which are Gaussian distributed with zero mean and covariance I_n. Call their density function as $\phi(x) = \mathcal{N}(x; 0, I_n)$.

- $\{y_k\}$, $k \in \mathbb{N}$ is a sequence of i.i.d. random variables which are Gaussian distributed with zero mean and covariance I_m. Call their density function as $\psi(x) = \mathcal{N}(x; 0, I_m)$.

- $\{r_k\}$, $k \in \mathbb{N}$ is a first-order finite-state homogenous Markov chain with transition probability matrix $\Pi = [\pi_{ij}]$ and initial distribution $\pi_0 = [\ \pi_0^1 \quad \pi_0^2 \quad \cdots \quad \pi_0^N\ ]$.

We define the sequence of random variables $\{\overline{\lambda}_l\}$ and $\{\overline{\Lambda}_k\}$, $k, l \in \mathbb{N}$ as

$$\overline{\lambda}_l = \begin{cases} \frac{\phi(\sqrt{\Sigma_0}^{-1}(x_l - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_l)} \frac{\psi(D^{-1}(r_l)(y_l - C(r_l)x_l))}{|D(r_l)|\psi(y_l)} & l = 0 \\ \frac{\phi(B^{-1}(r_l)(x_l - A(r_l)x_{l-1}))}{|B(r_l)|\phi(x_l)} \frac{\psi(D^{-1}(r_l)(y_l - C(r_l)x_l))}{|D(r_l)|\psi(y_l)} & l > 0 \end{cases}, \tag{5.6}$$

$$\overline{\Lambda}_k = \prod_{l=0}^{k} \overline{\lambda}_l \tag{5.7}$$

where $\sqrt{\Sigma_0}$ is the positive definite square root of Σ_0 and $|.|$ denotes the matrix determinant. Let $\mathcal{G}_k$ denote the complete filtration generated by random variables $\{x_0, \ldots, x_k, r_0, \ldots, r_k, y_0, \ldots, y_k\}$. We define a new probability measure P by setting the restriction of the Radon-Nikodym derivative $\frac{dP}{d\overline{P}}|_{\mathcal{G}_k}$ to $\overline{\Lambda}_k$. Under the new probability measure P, the processes $\{w_k \in \mathbb{R}^n\}$ and $\{v_k \in \mathbb{R}^m\}$, $k \in \mathbb{N}$ defined as

$$w_k \triangleq B^{-1}(r_k)(x_k - A(r_k)x_{k-1}), \tag{5.8}$$

$$v_k \triangleq D^{-1}(r_k)(y_k - C(r_k)x_k) \tag{5.9}$$

are sequences of i.i.d. Gaussian random variables with zero-mean and covariance I_n and I_m respectively[4] and the distribution of $\{r_k\}$ remains unchanged.[5] Note that, under both P and $\overline{P}$, the modal state $\{r_k \in \{e_1, e_2, \ldots, e_N\}\}$, $k \in \mathbb{N}$

[4] The proof of this fact follows the same lines as the proof presented for linear Gauss-Markov systems in Lemma A.7.

[5] Proofs of this fact is very similar to those given in [39] and [38].

has a semi-martingale representation

$$r_{k+1} = \Pi^T r_k + m_{k+1} \tag{5.10}$$

where m_k is a $\mathcal{G}_k$ martingale increment. The probability measure P is the nominal measure under which the results are required. The expectations under the reference probability measure $\overline{P}$, which are shown by $\overline{E}$, can be taken much more easily than the ones under P thanks to the independence properties.

5.3 Maximum Likelihood Estimation for Transition Probabilities

Maximum likelihood estimation is a powerful tool whose properties are well appreciated in the parameter estimation community. Let $\{P_\theta, \theta \in \Theta\}$ be a family of parameterized probability measures, where θ denotes the parameter vector, on a measurable space $(\Omega, \mathcal{F})$ all absolutely continuous with respect to a fixed probability measure P_0. The σ-field generated by the measurements $\{y_0, \ldots, y_k\}$ is called as $\mathcal{Y}_k \in \mathcal{F}$. The likelihood function for obtaining an estimate of θ based on the information available in $\mathcal{Y}_k$ is defined as

$$L_k(\theta) \triangleq E_0\Big[\frac{dP_\theta}{dP_0}\Big|\mathcal{Y}_k\Big] \tag{5.11}$$

where E_0 denotes the expectation with respect to the probability measure P_0 and $\frac{dP_\theta}{dP_0}$ is the Radon-Nikodym derivative of the probability measure P_θ with respect to P_0. The maximum likelihood estimate θ_k^{ML} is then given as

$$\theta_k^{ML} = \arg\max_{\theta\in\Theta} L_k(\theta) \tag{5.12}$$

In many problems, like the case in ours, the calculation of $L_k(\theta)$ in Eqn. 5.11 is impossible or computationally costly. In this case, the well-known expectation maximization (EM) algorithm comes into picture to provide an iterative solution which has the following steps with increasing number of measurements:

- Step-1: Set $k = 1$ and choose $\hat{\theta}_0$.

- Step-2: Set $l = 0$ and set $\hat{\theta}_{k,0} = \hat{\theta}_{k-1}$.

- Step-3: (E-Step) Compute $Q_k(\theta, \hat{\theta}_{k,l})$ as

$$Q_k(\theta, \hat{\theta}_{k,l}) = E_{\hat{\theta}_{k,l}}\Big[\log \frac{dP_\theta}{dP_{\hat{\theta}_{k,l}}}\Big|\mathcal{Y}_k\Big]. \tag{5.13}$$

- Step-4: (M-Step) Set $\hat{\theta}_{k,l+1}$ as

$$\hat{\theta}_{k,l+1} \in \arg\max_{\theta\in\Theta} Q_k(\theta, \hat{\theta}_{k,l}). \tag{5.14}$$

- Step-5:

  - If the stopping criterion is satisfied, set $\hat{\theta}_k = \hat{\theta}_{k,l+1}$. Set k to $k+1$, go to Step-2.
  - If the stopping criterion is not satisfied, set l to $l+1$ and go to Step-3.

Since our algorithm is required to be able to work in an online fashion, we are going to use a predetermined finite number ($L > 0$) of iterations with respect to l in our algorithm. Here, we illustrate the case L=1, which gives the worst performance among these algorithms. The steps of the resulting algorithm turns out to be:

- Step-1: Set $k = 1$ and choose $\hat{\theta}_0$.

- Step-2: (E-Step) Compute $Q_k(\theta, \hat{\theta}_{k-1})$ as

$$Q_k(\theta, \hat{\theta}_{k-1}) = E_{\hat{\theta}_{k-1}}\Big[\log \frac{dP_\theta}{dP_{\hat{\theta}_{k-1}}}\Big|\mathcal{Y}_k\Big]. \tag{5.15}$$

- Step-4: (M-Step) Set $\hat{\theta}_k$ as

$$\hat{\theta}_k \in \arg\max_{\theta\in\Theta} Q_k(\theta, \hat{\theta}_{k-1}). \tag{5.16}$$

- Step-5: Set k to $k+1$, go to Step-2.

In our problem, like the case in Chapter 4, the unknown parameter vector θ is composed of the transition probabilities π_{ij} as follows.

$$\theta \;=\; [\pi_{11}, \pi_{12}, \ldots, \pi_{1N}, \pi_{21}, \pi_{22}, \ldots, \pi_{2N}, \ldots\ldots, \pi_{N1}, \pi_{N2}, \ldots, \pi_{NN}]^T.$$

In [38], the transition probability estimates $\hat{\pi}_{ij}^{(k)}$ of the hidden Markov models (HMMs) are shown to be calculated using the above framework as

$$\hat{\pi}_{ij}^{(k)} = \frac{\hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1})}{\sum_{j=1}^{N} \hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1})} \quad \text{for} \quad 1 \leq i, j \leq N \tag{5.17}$$

where

$$\mathcal{J}_{uv}^{(k)} \;\triangleq\; \sum_{l=1}^{k} \langle r_l, e_v \rangle \langle r_{l-1}, e_u \rangle, \tag{5.18}$$

$$\hat{\mathcal{J}}_{uv}^{(k)}(\hat{\theta}_{k-1}) \;\triangleq\; E_{\hat{\theta}_{k-1}}[\mathcal{J}_{uv}^{(k)} | \mathcal{Y}_k]. \tag{5.19}$$

The quantity $\mathcal{J}_{uv}^{(k)}$ is the number of jumps of the Markov chain from state e_u to state e_v until time k and $\hat{\mathcal{J}}_{uv}^{(k)}(\theta)$ is the estimate of it based on the parameter θ using the information contained in $\mathcal{Y}_k$. Using a very similar derivation as in [38], it can be shown that the formula (5.17) is valid for the case of JMLSs.

In the next section, we are going to show how the estimates $\hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1})$ of the number of jumps $\mathcal{J}_{ij}^{(k)}$ can be calculated recursively.

5.4 Recursive Estimation of Number of Jumps $\mathcal{J}_{ij}^{(k)}$

In the previous section, it has been seen that the maximum likelihood estimation of the transition probabilities requires the calculation of the estimated number of jumps $\hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1})$ defined as

$$\hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1}) \triangleq E_{\hat{\theta}_{k-1}}[\mathcal{J}_{ij}^{(k)} | \mathcal{Y}_k]. \tag{5.20}$$

In this section, we are going to find recursive update formulas for the estimated number of jumps using the reference probability method. Using Theorem A.3,

we can see that

$$\hat{\mathcal{J}}_{ij}^{(k)}(\hat{\theta}_{k-1}) = \frac{\overline{E}_{\hat{\theta}_{k-1}}[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} | \mathcal{Y}_k]}{\overline{E}_{\hat{\theta}_{k-1}}[\overline{\Lambda}_k | \mathcal{Y}_k]}. \tag{5.21}$$

For the recursive calculation of $\hat{\mathcal{J}}_{ij}^{(k)}$, the following density definitions are required.

$$\eta_k^{ij-l}(x|\theta)dx \triangleq \overline{E}_\theta[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} I_{\{x_k \in dx\}} \langle r_k, e_l \rangle | \mathcal{Y}_k], \tag{5.22}$$

$$\alpha_k^l(x|\theta)dx \triangleq \overline{E}_\theta[\overline{\Lambda}_k I_{\{x_k \in dx\}} \langle r_k, e_l \rangle | \mathcal{Y}_k] \tag{5.23}$$

where the function $\mathcal{I}_A(\omega)$ defined as

$$\mathcal{I}_A(\omega) \triangleq \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \tag{5.24}$$

denotes the indicator function of the set A and the notation $\langle r_k, e_j \rangle$ stands for the the inner product $r_k^T e_j$ which is equal to the jth element of r_k. Note that, the densities $\alpha_k^j(x|\theta)$ can be easily calculated (approximated) by using a multiple model estimation algorithms such as IMM or GPB2 filters.[6] Using a simple reasoning, we can see that, if $f : \mathbb{R}^n \to \mathbb{R}$ is any test function (i.e., measurable function with compact support), the following equalities are satisfied.

$$\overline{E}_\theta[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} f(x_k) \langle r_k, e_l \rangle | \mathcal{Y}_k] = \int f(x) \eta_k^{ij-l}(x|\theta)dx, \tag{5.25}$$

$$\overline{E}_\theta[\overline{\Lambda}_k f(x_k) \langle r_k, e_l \rangle | \mathcal{Y}_k] = \int f(x) \alpha_k^l(x|\theta)dx. \tag{5.26}$$

Making use of Eqn. 5.25 and Eqn. 5.26 we can write

$$\overline{E}_\theta[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} | \mathcal{Y}_k] = \sum_{l=1}^{N} \int \eta_k^{ij-l}(x|\theta)dx, \tag{5.27}$$

$$\overline{E}_\theta[\overline{\Lambda}_k | \mathcal{Y}_k] = \sum_{l=1}^{N} \int \alpha_k^l(x|\theta)dx. \tag{5.28}$$

These, considering Eqn. 5.21, show that the densities $\{\eta_k^{ij-l}\}$, $1 \leq i, j, l \leq N$ and $\{\alpha_k^l\}$, $1 \leq l \leq N$ form a sufficient statistics for the estimation of the

[6] The derivation of the IMM filter presented in Chapter 3 illustrates the recursion of these densities.

transition probabilities. Noting that

$$k\alpha_k^l(x|\theta) = \sum_{i=1}^{N}\sum_{j=1}^{N}\eta_k^{ij-l}(x|\theta), \tag{5.29}$$

we can conclude that the set of densities $\{\eta_k^{ij-l}\}$, $1 \le i,j,l \le N$ is alone a sufficient statistics. The densities $\{\alpha_k^l\}$, $1 \le l \le N$ will only be used for notational simplicity.

5.4.1 Recursion

Theorem 5.1 *The densities $\eta_k^{ij-l}(x|\theta)$ satisfy the following recursion.*

$$\begin{aligned}\eta_k^{ij-l}(x|\theta) &= \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l|\psi(y_k)}\int \phi(B_l^{-1}(x - A_l z)) \\ &\quad \times\Big[\sum_{m=1}^{N}\pi_{ml}\eta_{k-1}^{ij-m}(z|\theta) + \langle e_l, e_j\rangle\pi_{ij}\alpha_{k-1}^{i}(z|\theta)\Big]dz \end{aligned}\tag{5.30}$$

where $A_l \triangleq A(e_l)$, $B_l \triangleq B(e_l)$, $C_l \triangleq C(e_l)$, and $D_l \triangleq D(e_l)$ for $l = 1,\ldots,N$.

Proof Let $f : \mathbb{R}^n \to \mathbb{R}$ be any test function, then

$$\begin{aligned}\int \eta_k^{ij-l}(x|\theta)f(x)dx &= \overline{E}_\theta[\overline{\Lambda}_k\mathcal{J}_{ij}^{(k)}f(x_k)\langle r_k, e_l\rangle|\mathcal{Y}_k] &(5.31)\\ &= \overline{E}_\theta\Big[\overline{\Lambda}_{k-1}\frac{\phi(B^{-1}(r_k)(x_k - A(r_k)x_{k-1}))}{|B(r_k)|\phi(x_k)} \\ &\quad \times\frac{\psi(D^{-1}(r_k)(y_k - C(r_k)x_k))}{|D(r_k)|\psi(y_k)}\mathcal{J}_{ij}^{(k)}f(x_k)\langle r_k, e_l\rangle\Big|\mathcal{Y}_k\Big] \\ &= \overline{E}_\theta\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)}\frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)} \\ &\quad \times\mathcal{J}_{ij}^{(k)}f(x_k)\langle r_k, e_l\rangle\Big|\mathcal{Y}_k\Big] &(5.32)\end{aligned}$$

Using the definition of $\mathcal{J}_{ij}^{(k)}$, we can write

$$\mathcal{J}_{ij}^{(k)} = \mathcal{J}_{ij}^{(k-1)} + \langle r_k, e_j\rangle\langle r_{k-1}, e_i\rangle. \tag{5.33}$$

Substituting this into Eqn. 5.32, we get

$$\begin{aligned}&\int \eta_k^{ij-l}(x|\theta)f(x)dx \\ &\qquad = \overline{E}_\theta\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)}\frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}\end{aligned}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k)\langle r_k, e_l\rangle \Big| \mathcal{Y}_k \Big]$$

$$+\overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times f(x_k)\langle r_k, e_j\rangle\langle r_{k-1}, e_i\rangle\langle r_k, e_l\rangle \Big| \mathcal{Y}_k \Big] \tag{5.34}$$

$$= \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k)\langle r_k, e_l\rangle \Big| \mathcal{Y}_k \Big]$$

$$+\langle e_l, e_j\rangle \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times f(x_k)\langle r_k, e_j\rangle\langle r_{k-1}, e_i\rangle \Big| \mathcal{Y}_k \Big] \tag{5.35}$$

$$= \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k)\langle \Pi^T r_{k-1}, e_l\rangle \Big| \mathcal{Y}_k \Big]$$

$$+\overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k)\langle m_k, e_l\rangle \Big| \mathcal{Y}_k \Big]$$

$$+\langle e_l, e_j\rangle \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times f(x_k)\langle \Pi^T r_{k-1}, e_j\rangle\langle r_{k-1}, e_i\rangle \Big| \mathcal{Y}_k \Big]$$

$$+\langle e_l, e_j\rangle \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times f(x_k)\langle m_k, e_j\rangle\langle r_{k-1}, e_i\rangle \Big| \mathcal{Y}_k \Big]. \tag{5.36}$$

The second and the fourth expectations on the right hand side of Eqn. 5.36 are zero due to facts that m_k is a $\mathcal{G}_k$-martingale increment and that under the probability measure $\overline{P}$, the process $\{r_k\}$, and hence the process $\{m_k\}$, is independent of the processes $\{x_k\}$ and $\{y_k\}$. Now using the identity

$$\langle \Pi^T r_{k-1}, e_j\rangle = \sum_{i=1}^{N} \pi_{ij}\langle r_{k-1}, e_i\rangle, \tag{5.37}$$

Eqn 5.36 becomes

$$\int \eta_k^{ij-l}(x|\theta) f(x) dx \tag{5.38}$$

$$= \sum_{m=1}^{N} \pi_{ml} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l|\phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l|\psi(y_k)}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k) \langle r_{k-1}, e_m \rangle \Big| \mathcal{Y}_k \Big]$$

$$+ \langle e_l, e_j \rangle \pi_{ij} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l| \phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l| \psi(y_k)}$$

$$\times f(x_k) \langle r_{k-1}, e_i \rangle \Big| \mathcal{Y}_k \Big] \tag{5.39}$$

$$= \sum_{m=1}^{N} \pi_{ml} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \overline{E}_\theta \Big[\frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l| \phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l| \psi(y_k)}$$

$$\times \mathcal{J}_{ij}^{(k-1)} f(x_k) \Big| x_{k-1}, \mathcal{Y}_k \Big] \langle r_{k-1}, e_m \rangle \Big| \mathcal{Y}_k \Big]$$

$$+ \langle e_l, e_j \rangle \pi_{ij} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \overline{E}_\theta \Big[\frac{\phi(B_l^{-1}(x_k - A_l x_{k-1}))}{|B_l| \phi(x_k)} \frac{\psi(D_l^{-1}(y_k - C_l x_k))}{|D_l| \psi(y_k)}$$

$$\times f(x_k) \Big| x_{k-1}, \mathcal{Y}_k \Big] \langle r_{k-1}, e_i \rangle \Big| \mathcal{Y}_k \Big] \tag{5.40}$$

The inner expectations in Eqn. 5.40 can easily be taken as follows due to the independence properties of the sequence $\{x_k\}$ under $\overline{P}$.

$$\int \eta_k^{ij-l}(x|\theta) f(x) dx \tag{5.41}$$

$$= \sum_{m=1}^{N} \pi_{ml} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \mathcal{J}_{ij}^{(k-1)} \int \frac{\phi(B_l^{-1}(x - A_l x_{k-1}))}{|B_l| \phi(x)} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)}$$

$$\times f(x) \phi(x) dx \langle r_{k-1}, e_m \rangle \Big| \mathcal{Y}_k \Big]$$

$$+ \langle e_l, e_j \rangle \pi_{ij} \overline{E}_\theta \Big[\overline{\Lambda}_{k-1} \int \frac{\phi(B_l^{-1}(x - A_l x_{k-1}))}{|B_l| \phi(x)} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)}$$

$$\times f(x) \phi(x) dx \langle r_{k-1}, e_i \rangle \Big| \mathcal{Y}_k \Big] \tag{5.42}$$

$$= \sum_{m=1}^{N} \pi_{ml} \int \int \frac{\phi(B_l^{-1}(x - A_l z))}{|B_l|} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)} f(x) dx \tag{5.43}$$

$$\times \eta_{k-1}^{ij-m}(z|\theta) dz$$

$$+ \langle e_l, e_j \rangle \pi_{ij} \int \int \frac{\phi(B_l^{-1}(x - A_l z))}{|B_l|} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)} f(x) dx$$

$$\times \alpha_{k-1}^{i}(z|\theta) dz \tag{5.44}$$

$$= \int f(x) \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l| \psi(y_k)} \int \phi(B_l^{-1}(x - A_l z))$$

$$\times \sum_{m=1}^{N} \pi_{ml} \eta_{k-1}^{ij-m}(z|\theta) dz dx$$

$$+ \int f(x) \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l| \psi(y_k)} \int \phi(B_l^{-1}(x - A_l z))$$

$$\times \langle e_l, e_j \rangle \pi_{ij} \alpha_{k-1}^{i}(z|\theta) dz dx \tag{5.45}$$

$$= \int f(x) \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l|\psi(y_k)} \int \phi(B_l^{-1}(x - A_l z))$$
$$\times \Big[\sum_{m=1}^{N} \pi_{ml} \eta_{k-1}^{ij-m}(z|\theta) + \langle e_l, e_j \rangle \pi_{ij} \alpha_{k-1}^{i}(z|\theta) \Big] dz dx. \quad (5.46)$$

Since this equation is satisfied for all test functions, the recursion in Eqn. 5.30 holds. □

Remark 5.1 *Since the recursion in Eqn. 5.30 is linear, once the initial densities* $\{\alpha_0^i(x|\theta)\}$ *are (unnormalized) Gaussian densities (which is the case in our model), the densities* $\{\eta_k^{ij-l}(x|\theta)\}$ *become unnormalized Gaussian mixtures with ever increasing number of components as* k *increases.*

Corollary 1 *The densities* $\eta_k^{ij-l}(x|\theta)$ *satisfy the recursion*

$$\eta_k^{ij-l}(x|\theta) = \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l|\psi(y_k)} \int \phi(B_l^{-1}(x - A_l z))$$
$$\times \Big[\sum_{m=1}^{N} \pi_{ml} \eta_{k-1}^{ij-m}(z|\theta) + \frac{\langle e_l, e_j \rangle \pi_{ij}}{k-1} \sum_{u=1}^{N} \sum_{v=1}^{N} \eta_{k-1}^{uv-i}(z|\theta) \Big] dz \quad (5.47)$$

for $k > 1$*, which is in terms of only the densities* $\{\eta_k^{ij-l}\}$*,* $1 \leq i, j, l \leq N$*.*

Proof Proof is given simply by the substitution of the relationship given by Eqn. 5.29 into the recursion of Eqn. 5.30. □

The recursion in Eqn. 5.47 is the main result of this chapter which results in a maximum likelihood estimator for the transition probabilities associated with JMLSs.

5.4.2 Initial Densities

Since the initial number of jumps $\mathcal{J}_{ij}^{(0)}$ is given as

$$\mathcal{J}_{ij}^{(0)} \triangleq \sum_{l=1}^{0} \langle r_l, e_j \rangle \langle r_{l-1}, e_i \rangle = 0, \quad (5.48)$$

the initial unnormalized densities $\{\eta_0^{ij-l}(x|\theta)\}$, $1 \leq i, j, l \leq N$ are all equal to zero, i.e.,

$$\eta_0^{ij-l}(x|\theta) = 0 \quad \forall x \quad \text{for} \quad 1 \leq i, j, l \leq N. \quad (5.49)$$

Remark 5.2 *An important point about the recursion of Eqn. 5.47 is that the right hand side has* $\frac{0}{0}$ *indeterminate form when* $k = 1$*. Therefore, for* $k = 1$*, the recursion in Eqn. 5.30 is still to be used.*

5.4.3 Approximation

For the transition probability estimation at time k, i.e., for the calculation of $\hat{\theta}_k$, we need the densities $\{\eta_k^{ij-l}(x|\hat{\theta}_{k-1})\}$, $1 \leq i, j, l \leq N$ due to the identity in Eqn. 5.17. Hence, we need to calculate the densities $\{\eta_k^{ij-l}(x|\hat{\theta}_{k-1})\}$ recursively. The recursion in Eqn. 5.47, in order to calculate these densities, requires the densities $\{\eta_{k-1}^{ij-l}(x|\hat{\theta}_{k-1})\}$, $1 \leq i, j, l \leq N$. However, at time $k-1$, the densities $\{\eta_{k-1}^{ij-l}(x|\hat{\theta}_{k-1})\}$, $1 \leq i, j, l \leq N$ are not available. Instead, the densities $\{\eta_{k-1}^{ij-l}(x|\hat{\theta}_{k-2})\}$, $1 \leq i, j, l \leq N$ are available since these densities must have been used for computing the transition probability estimates at time $k-1$. This problem can be iterated back until the initial time $k = 0$ which makes the reprocessing of all measurements necessary. In order to obtain an online recursive estimation mechanism, we therefore make the approximations

$$\eta_{k-1}^{ij-l}(x|\hat{\theta}_{k-1}) \;\approx\; \eta_{k-1}^{ij-l}(x|\hat{\theta}_{k-2}) \quad \text{for} \quad 1 \leq i, j, l \leq N \tag{5.50}$$

for all k. This assumption basically amounts to assuming that the parameter estimates do not change much between consecutive time instants.

In MMSE multiple model estimation, it is well-known that the the number of statistics to be kept for calculating the densities $\alpha_k^l(x|\theta)$ grows exponentially [13].[7] The optimal recursion given in Eqn. 5.30, shows that the situation is worse for the densities $\eta_k^{ij-l}(x|\theta)$ due to the fact that the recursions are excited by the densities $\alpha_{k-1}^l(x|\theta)$. Even if the densities $\alpha_{k-1}^l(x|\theta)$ are approximated by single Gaussian densities as the case in the IMM or GPB2 filters, the number of statistics to be kept for densities $\eta_k^{ij-l}(x|\theta)$ still grows exponentially (due to the exponentially increasing number of unnormalized Gaussian components in mixture). This extreme growing is observed more clearly in the recursion of

[7] This fact is also apparent in the derivation of the IMM filter in Sec. 3.3.

Eqn. 5.47. This situation, therefore, makes additional approximations necessary. Here, we are going to make IMM-type approximations[8] to keep a single Gaussian density for each $\eta_k^{ij-l}(x|\theta)$. Assuming that each previous density $\eta_{k-1}^{ij-l}(x|\theta)$ is approximated by a single Gaussian, i.e.,

$$\eta_{k-1}^{ij-l}(x|\theta) \approx d_{k-1}^{ij-l}\mathcal{N}(x; \hat{x}_{k-1|k-1}^{ij-l}, \Sigma_{k-1|k-1}^{ij-l}), \tag{5.51}$$

the Gaussian mixture $[\sum_{m=1}^{N}\pi_{ml}\eta_{k-1}^{ij-m}(z|\theta) + \frac{\langle e_l,e_j\rangle\pi_{ij}}{k-1}\sum_{u=1}^{N}\sum_{v=1}^{N}\eta_{k-1}^{uv-i}(z|\theta)]$ is approximated by a single Gaussian, i.e.,

$$\Big[\sum_{m=1}^{N}\pi_{ml}\eta_{k-1}^{ij-m}(z|\theta) + \frac{\langle e_l,e_j\rangle\pi_{ij}}{k-1}\sum_{u=1}^{N}\sum_{v=1}^{N}\eta_{k-1}^{uv-i}(z|\theta)\Big] \approx d_{k-1}^{ij-0l}\mathcal{N}(x; \hat{x}_{k-1|k-1}^{ij-0l}, \Sigma_{k-1|k-1}^{ij-0l}) \tag{5.52}$$

where

$$d_{k-1}^{ij-0l} = \sum_{m=1}^{N}\pi_{ml}d_{k-1}^{ij-m} + \frac{\langle e_l,e_j\rangle\pi_{ij}}{k-1}\sum_{u=1}^{N}\sum_{v=1}^{N}d_{k-1}^{uv-i}, \tag{5.53}$$

$$\hat{x}_{k-1|k-1}^{ij-0l} = \sum_{m=1}^{N}\frac{\pi_{ml}d_{k-1}^{ij-m}}{d_{k-1}^{ij-0l}}\hat{x}_{k-1|k-1}^{ij-m} + \frac{\langle e_l,e_j\rangle}{k-1}\frac{\pi_{ij}}{d_{k-1}^{ij-0l}}\sum_{u=1}^{N}\sum_{v=1}^{N}d_{k-1}^{uv-i}\hat{x}_{k-1|k-1}^{uv-i}, \tag{5.54}$$

$$\begin{aligned}\Sigma_{k-1|k-1}^{ij-0l} = \sum_{m=1}^{N}\frac{\pi_{ml}d_{k-1}^{ij-m}}{d_{k-1}^{ij-0l}}\Big[&\Sigma_{k-1|k-1}^{ij-m} \\ &+(\hat{x}_{k-1|k-1}^{ij-m} - \hat{x}_{k-1|k-1}^{ij-0l})(\hat{x}_{k-1|k-1}^{ij-m} - \hat{x}_{k-1|k-1}^{ij-0l})^T\Big] \\ +\frac{\langle e_l,e_j\rangle}{k-1}\frac{\pi_{ij}}{d_{k-1}^{ij-0l}}\sum_{u=1}^{N}\sum_{v=1}^{N}d_{k-1}^{uv-i}\Big[&\Sigma_{k-1|k-1}^{uv-i} \\ &+(\hat{x}_{k-1|k-1}^{uv-i} - \hat{x}_{k-1|k-1}^{ij-0l})(\hat{x}_{k-1|k-1}^{uv-i} - \hat{x}_{k-1|k-1}^{ij-0l})^T\Big].\end{aligned} \tag{5.55}$$

Note that this approximation process is similar to the mixing stage of the IMM filter [13].[9] With these approximations, the densities $\eta_k^{ij}(x|\theta)$ are automatically kept in the form

$$\eta_k^{ij-l}(x|\theta) = d_k^{ij-l}\mathcal{N}(x; \hat{x}_{k|k}^{ij-l}, \Sigma_{k|k}^{ij-l}). \tag{5.56}$$

When the approximation in Eqn. 5.52 is substituted into the recursion of Eqn. 5.47, we obtain

$$\eta_k^{ij-l}(x|\theta) = \frac{\psi(D_l^{-1}(y_k - C_l x))}{|B_l||D_l|\psi(y_k)}$$

[8] See the derivation of the IMM filter in Sec. 3.3 for IMM-type approximations.

[9] Mixing stage is also detailed in the derivation of the IMM filter in Sec. 3.3.

$$
\begin{aligned}
& \times \int \phi(B_l^{-1}(x - A_l z)) d_{k-1}^{ij-0l} \mathcal{N}(z; \hat{x}_{k-1|k-1}^{ij-0l}, \Sigma_{k-1|k-1}^{ij-0l}) dz \\
= \; & d_{k-1}^{ij-0l} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)} \\
& \times \int \mathcal{N}(x; A_l z, B_l B_l^T) \mathcal{N}(z; \hat{x}_{k-1|k-1}^{ij-0l}, \Sigma_{k-1|k-1}^{ij-0l}) dz \qquad (5.57) \\
= \; & d_{k-1}^{ij-0l} \frac{\psi(D_l^{-1}(y_k - C_l x))}{|D_l| \psi(y_k)} \mathcal{N}(x; \hat{x}_{k|k-1}^{ij-l}, \Sigma_{k|k-1}^{ij-l}) \qquad (5.58) \\
= \; & \frac{d_{k-1}^{ij-0l}}{\psi(y_k)} \mathcal{N}(y_k; C_l x, D_l D_l^T) \mathcal{N}(x; \hat{x}_{k|k-1}^{ij-l}, \Sigma_{k|k-1}^{ij-l}) \qquad (5.59)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{x}_{k|k-1}^{ij-l} &= A_l \hat{x}_{k-1|k-1}^{ij-0l}, \qquad (5.60) \\
\Sigma_{k|k-1}^{ij-l} &= A_l \Sigma_{k-1|k-1}^{ij-0l} A_l^T + B_l B_l^T. \qquad (5.61)
\end{aligned}
$$

Here, while passing from Eqn. 5.57 to 5.58, we used the well-known result on the integral of Gaussian densities which is given in App. B.4. Using the result on the multiplication of Gaussian densities in App. B.2, Eqn. 5.59 can be written as

$$
\begin{aligned}
\eta_k^{ij-l}(x|\theta) &= \frac{d_{k-1}^{ij-0l}}{\psi(y_k)} \mathcal{N}(y_k; C_l \hat{x}_{k|k-1}^{ij-l}, S_k^{ij-l}) \mathcal{N}(x; \hat{x}_{k|k}^{ij-l}, \Sigma_{k|k}^{ij-l}) \qquad (5.62) \\
&= d_k^{ij-l} \mathcal{N}(x; \hat{x}_{k|k}^{ij-l}, \Sigma_{k|k}^{ij-l}) \qquad (5.63)
\end{aligned}
$$

where

$$
\begin{aligned}
S_k^{ij-l} &= C_l \Sigma_{k|k-1}^{ij-l} C_l^T + D_l D_l^T, \qquad (5.64) \\
\hat{x}_{k|k}^{ij-l} &= \hat{x}_{k|k-1}^{ij-l} + \Sigma_{k|k-1}^{ij-l} C_l^T (S_k^{ij-l})^{-1} (y_k - C_l \hat{x}_{k|k-1}^{ij-l}), \qquad (5.65) \\
\Sigma_{k|k}^{ij-l} &= \Sigma_{k|k-1}^{ij-l} - \Sigma_{k|k-1}^{ij-l} C_l^T (S_k^{ij-l})^{-1} C_l \Sigma_{k|k-1}^{ij-l}, \qquad (5.66) \\
d_k^{ij-l} &= \frac{d_{k-1}^{ij-0l}}{\psi(y_k)} \mathcal{N}(y_k; C_l \hat{x}_{k|k-1}^{ij-l}, S_k^{ij-l}). \qquad (5.67)
\end{aligned}
$$

5.4.4 Initialization

The update formulas given in the previous section assume that the previous density functions $\{\eta_{k-1}^{ij-l}(x|\theta)\}$, $1 \leq i, j, l \leq N$ at each estimation step are represented by a single unnormalized Gaussian. The initial densities $\{\eta_0^{ij-l}(x|\theta)\}$,

$1 \leq i, j, l \leq N$ given by Eqn. 5.49 can be expressed as unnormalized Gaussians with arbitrary means and covariances and with zero weights. However, this causes a $\frac{0}{0}$ indeterminate form on the right hand side of recursion of Eqn. 5.47 for $k = 1$. This can be avoided by using the recursion of Eqn. 5.30 which requires the densities $\{\alpha_0^i(x|\theta)\}_{i=1}^N$. The initial densities $\alpha_0^j(x|\theta)$ can be calculated as follows.[10]

$$\begin{aligned}
\int f(x)\alpha_0^j(x|\theta)dx &= \overline{E}\Big[\overline{\Lambda}_0\langle r_0, e_j\rangle f(x_0)|\mathcal{Y}_0\Big] && (5.68)\\
&= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)}\frac{\psi(D^{-1}(r_0)(y_0-C(r_0)x_0))}{|D(r_0)|\psi(y_0)} \\
&\quad \times f(x_0)\langle r_0, e_j\rangle\Big|\mathcal{Y}_0\Big] && (5.69)\\
&= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)}\frac{\psi(D_j^{-1}(y_0-C_jx_0))}{|D_j|\psi(y_0)} \\
&\quad \times \langle r_0, e_j\rangle f(x_0)\Big|\mathcal{Y}_0\Big] && (5.70)\\
&= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)}\frac{\psi(D_j^{-1}(y_0-C_jx_0))}{|D_j|\psi(y_0)}f(x_0)\Big|\mathcal{Y}_0\Big] \\
&\quad \times \overline{E}\left[\langle r_0, e_j\rangle\right] && (5.71)\\
&= \int f(x)\pi_0^j\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x)}\frac{\psi(D_j^{-1}(y_0-C_jx))}{|D_j|\psi(y_0)}\phi(x)dx \\
&= \int f(x)\pi_0^j\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|}\frac{\psi(D_j^{-1}(y_0-C_jx))}{|D_j|\psi(y_0)}dx.
\end{aligned}$$

Since the equality holds for all test functions $f(.)$, the initial density $\alpha_0^j(x|\theta)$ is given as

$$\begin{aligned}
\alpha_0^j(x|\theta) &= \pi_0^j\frac{\psi(D_j^{-1}(y_0-C_jx))}{|D_j|\psi(y_0)}\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|} && (5.72)\\
&= \frac{\pi_0^j}{\psi(y_0)}\mathcal{N}(y_0; C_jx, D_jD_j^T)\mathcal{N}(x; \bar{x}_0, \Sigma_0) && (5.73)\\
&= \frac{\pi_0^j}{\psi(y_0)}\mathcal{N}(y_0; C_j\bar{x}_0, S_0^j)\mathcal{N}(x; \hat{x}_{0|0}^j, \Sigma_{0|0}^j) && (5.74)\\
&= \bar{c}_0^j\mathcal{N}(x; \hat{x}_{0|0}^j, \Sigma_{0|0}^j). && (5.75)
\end{aligned}$$

where

$$S_0^j \triangleq C_j\Sigma_0C_j^T + D_jD_j^T, \quad (5.76)$$

[10] This calculation was also done in Sec. 3.3 and is repeated here for convenience.

$$\hat{x}^j_{0|0} \triangleq \bar{x}_0 + \Sigma_0 C_j^T (S_0^j)^{-1}(y_0 - C_j \bar{x}_0), \tag{5.77}$$

$$\Sigma^j_{0|0} \triangleq \Sigma_0 - \Sigma_0 C_j^T (S_0^j)^{-1} C_j \Sigma_0, \tag{5.78}$$

$$\bar{c}_0^j = \frac{\pi_0^j}{\psi(y_0)} \mathcal{N}(y_0; C_j \bar{x}_0, S_0^j). \tag{5.79}$$

Here, while passing from Eqn. 5.73 to 5.74, the result on the multiplication of Gaussian densities, which is given in App. B.2, is used. The recursion in Eqn. 5.30 for $k = 1$ gives the densities $\{\eta_1^{ij-l}(x|\theta)\}$, $1 \leq i, j, l \leq N$ as

$$\eta_1^{ij-l}(x|\theta) = d_1^{ij-l} \mathcal{N}(x; \hat{x}^{ij-l}_{1|1}, \Sigma^{ij-l}_{1|1}) \tag{5.80}$$

where

$$d_1^{ij-l} = \begin{cases} \frac{\bar{c}_0^i \pi_{ij}}{\psi(y_1)} \mathcal{N}(y_1; C_j \hat{x}^{ij-j}_{1|0}, S_1^{ij-j}) & l = j \\ 0 & \text{otherwise} \end{cases}, \tag{5.81}$$

$$\hat{x}^{ij-l}_{1|1} = \begin{cases} \hat{x}^{ij-j}_{1|0} + \Sigma^{ij-j}_{1|0} C_j^T (S_1^{ij-j})^{-1}(y_1 - C_j \hat{x}^{ij-j}_{1|0}) & l = j \\ \text{Arbitrary} & \text{otherwise} \end{cases}, \tag{5.82}$$

$$\Sigma^{ij-l}_{1|1} = \begin{cases} \Sigma^{ij-j}_{1|0} - \Sigma^{ij-j}_{1|0} C_j^T (S_1^{ij-j})^{-1} C_j \Sigma^{ij-j}_{1|0} & l = j \\ \text{Arbitrary} & \text{otherwise} \end{cases}. \tag{5.83}$$

Here, the estimates $\hat{x}^{ij-j}_{1|0}$, covariances $\Sigma^{ij-j}_{1|0}$ and S_1^{ij-j} are given as

$$\hat{x}^{ij-j}_{1|0} = A_j \hat{x}^i_{0|0}, \tag{5.84}$$

$$\Sigma^{ij-j}_{1|0} = A_j \Sigma^i_{0|0} A_j^T + B_j B_j^T, \tag{5.85}$$

$$S_1^{ij-j} = C_j \Sigma^{ij-j}_{1|0} C_j^T + D_j D_j^T. \tag{5.86}$$

Note that some of the estimates $\hat{x}^{ij-l}_{1|1}$ and covariances $\Sigma^{ij-l}_{1|1}$ can be given arbitrary values. Their values do not affect the subsequent iterations of the algorithm since their corresponding mode weights d_1^{ij-l} are initialized as zero.

5.5 Calculation of Transition Probability Estimates

In this section, we give the formulas for the transition probability estimate $\hat{\pi}_{ij}^{(k)}$ assuming the the densities $\{\eta_k^{ij-l}(x|\hat{\theta}_{k-1})\}$ have been calculated. Substituting

Eqn. 5.21 into Eqn. 5.17, we obtain

$$\hat{\pi}_{ij}^{(k)} = \frac{\overline{E}_{\hat{\theta}_{k-1}}[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} | \mathcal{Y}_k]}{\sum_{j=1}^{N} \overline{E}_{\hat{\theta}_{k-1}}[\overline{\Lambda}_k \mathcal{J}_{ij}^{(k)} | \mathcal{Y}_k]} \quad \text{for} \quad 1 \leq i, j \leq N. \tag{5.87}$$

When we combine this with Eqn. 5.27, we get

$$\hat{\pi}_{ij}^{(k)} = \frac{\sum_{l=1}^{N} \int \eta_k^{ij-l}(x|\hat{\theta}_{k-1})dx}{\sum_{j=1}^{N}\sum_{l=1}^{N} \int \eta_k^{ij-l}(x|\hat{\theta}_{k-1})dx} \quad \text{for} \quad 1 \leq i, j \leq N. \tag{5.88}$$

Since the densities $\eta_k^{ij-l}(x|\hat{\theta}_{k-1})$ are approximated by a single weighted Gaussian density, we have

$$\int \eta_k^{ij-l}(x|\hat{\theta}_{k-1})dx = d_k^{ij-l} \tag{5.89}$$

which gives

$$\hat{\pi}_{ij}^{(k)} = \frac{\sum_{l=1}^{N} d_k^{ij-l}}{\sum_{j=1}^{N}\sum_{l=1}^{N} d_k^{ij-l}} \quad \text{for} \quad 1 \leq i, j \leq N. \tag{5.90}$$

Remark 5.3 *It is important to note that, for the calculation of the transition probability estimates, all that matters about the coefficients d_k^{ij-l} (or d_k^{ij-0l}) is their relative magnitudes (and not their absolute magnitudes). Therefore, at any time-step k, one can multiply the the coefficients d_k^{ij-l} (or d_k^{ij-0l}) by a common constant number without affecting the output estimate. In the cases where the coefficients get too small or too big to handle in computer, one can make some normalization on them accordingly without affecting the performance. Moreover, the term $\psi(y_k)$ given in the coefficient update equations, which can cause a division by zero in the computer, can be discarded safely since it is a common factor to all of the coefficients.*

5.6 Summary, Modifications and Base-State Estimation

In this section, we summarize the algorithm derived in the previous sections.

- Initialization at $k = 1$:

– Set the coefficients (mode weights) $\{d_1^{ij-l}\}$, estimates $\{\hat{x}_{1|1}^{ij-l}\}$ and covariances $\{\Sigma_{1|1}^{ij-l}\}$ for $1 \leq i, j, l \leq N$ as

$$d_1^{ij-l} = \begin{cases} \frac{\bar{c}_0^i \hat{\pi}_{ij}^{(0)}}{\psi(y_1)} \mathcal{N}(y_1; C_j \hat{x}_{1|0}^{ij-j}, S_1^{ij-j}) & l = j \\ 0 & \text{otherwise} \end{cases}, \tag{5.91}$$

$$\hat{x}_{1|1}^{ij-l} = \begin{cases} \hat{x}_{1|0}^{ij-j} + \Sigma_{1|0}^{ij-j} C_j^T (S_1^{ij-j})^{-1} (y_1 - C_j \hat{x}_{1|0}^{ij-j}) & l = j \\ \text{Arbitrary} & \text{otherwise} \end{cases} \tag{5.92}$$

$$\Sigma_{1|1}^{ij-l} = \begin{cases} \Sigma_{1|0}^{ij-j} - \Sigma_{1|0}^{ij-j} C_j^T (S_1^{ij-j})^{-1} C_j \Sigma_{1|0}^{ij-j} & l = j \\ \text{Arbitrary} & \text{otherwise} \end{cases}. \tag{5.93}$$

where

$$\hat{x}_{1|0}^{ij-j} = A_j \hat{x}_{0|0}^i, \tag{5.94}$$

$$\Sigma_{1|0}^{ij-j} = A_j \Sigma_{0|0}^i A_j^T + B_j B_j^T, \tag{5.95}$$

$$S_1^{ij-j} = C_j \Sigma_{1|0}^{ij-j} C_j^T + D_j D_j^T. \tag{5.96}$$

Note that the quantities $\bar{c}_0^i$, $\hat{x}_{0|0}^i$ and $\Sigma_{0|0}^i$ required for Eqns. from 5.91 to 5.96 are given by

$$S_0^j \triangleq C_j \Sigma_0 C_j^T + D_j D_j^T, \tag{5.97}$$

$$\hat{x}_{0|0}^j \triangleq \bar{x}_0 + \Sigma_0 C_j^T (S_0^j)^{-1} (y_0 - C_j \bar{x}_0), \tag{5.98}$$

$$\Sigma_{0|0}^j \triangleq \Sigma_0 - \Sigma_0 C_j^T (S_0^j)^{-1} C_j \Sigma_0, \tag{5.99}$$

$$\bar{c}_0^j = \frac{\pi_0^j}{\psi(y_0)} \mathcal{N}(y_0; C_j \bar{x}_0, S_0^j). \tag{5.100}$$

– Calculate the transition probability estimates as

$$\hat{\pi}_{ij}^{(1)} = \frac{\sum_{l=1}^N d_1^{ij-l}}{\sum_{j=1}^N \sum_{l=1}^N d_1^{ij-l}} \quad \text{for} \quad 1 \leq i, j \leq N. \tag{5.101}$$

- Update for $k > 1$:

– Mixing: Calculate the mixed coefficients $\{d_{k-1}^{ij-0l}\}$, estimates $\{\hat{x}_{k-1|k-1}^{ij-0l}\}$ and covariances $\{\Sigma_{k-1|k-1}^{ij-0l}\}$ for $1 \leq i, j, l \leq N$ as

$$d_{k-1}^{ij-0l} = \sum_{m=1}^N \hat{\pi}_{ml}^{(k-1)} d_{k-1}^{ij-m}$$

$$+\frac{\langle e_l, e_j\rangle \hat{\pi}_{ij}^{(k-1)}}{k-1}\sum_{u=1}^{N}\sum_{v=1}^{N} d_{k-1}^{uv-i}, \tag{5.102}$$

$$\begin{aligned}\hat{x}_{k-1|k-1}^{ij-0l} &= \sum_{m=1}^{N}\frac{\hat{\pi}_{ml}^{(k-1)} d_{k-1}^{ij-m}}{d_{k-1}^{ij-0l}}\hat{x}_{k-1|k-1}^{ij-m} \\ &\quad +\frac{\langle e_l, e_j\rangle}{k-1}\frac{\hat{\pi}_{ij}^{(k-1)}}{d_{k-1}^{ij-0l}}\sum_{u=1}^{N}\sum_{v=1}^{N} d_{k-1}^{uv-i}\hat{x}_{k-1|k-1}^{uv-i}, \end{aligned}\tag{5.103}$$

$$\begin{aligned}\Sigma_{k-1|k-1}^{ij-0l} &= \sum_{m=1}^{N}\frac{\hat{\pi}_{ml}^{(k-1)} d_{k-1}^{ij-m}}{d_{k-1}^{ij-0l}}\Big[\Sigma_{k-1|k-1}^{ij-m} \\ &\quad +(\hat{x}_{k-1|k-1}^{ij-m}-\hat{x}_{k-1|k-1}^{ij-0l})(\hat{x}_{k-1|k-1}^{ij-m}-\hat{x}_{k-1|k-1}^{ij-0l})^T\Big] \\ &\quad +\frac{\langle e_l, e_j\rangle}{k-1}\frac{\hat{\pi}_{ij}^{(k-1)}}{d_{k-1}^{ij-0l}}\sum_{u=1}^{N}\sum_{v=1}^{N} d_{k-1}^{uv-i}\Big[\Sigma_{k-1|k-1}^{uv-i} \\ &\quad +(\hat{x}_{k-1|k-1}^{uv-i}-\hat{x}_{k-1|k-1}^{ij-0l})(\hat{x}_{k-1|k-1}^{uv-i}-\hat{x}_{k-1|k-1}^{ij-0l})^T\Big]. \end{aligned}\tag{5.104}$$

– Kalman Filtering: Using the mixed coefficients $\{d_{k-1}^{ij-0l}\}$, estimates $\{\hat{x}_{k-1|k-1}^{ij-0l}\}$ and covariances $\{\Sigma_{k-1|k-1}^{ij-0l}\}$ for $1 \le i, j, l \le N$ as the initial conditions, run Kalman filters to obtain updated mode weights $\{d_{k-1}^{ij-l}\}$, estimates $\{\hat{x}_{k|k}^{ij-0l}\}$ and covariances $\{\Sigma_{k|k}^{ij-0l}\}$ for $1 \le i, j, l \le N$. The required Kalman filtering equations are given as

  * Prediction Update:

$$\hat{x}_{k|k-1}^{ij-l} = A_l \hat{x}_{k-1|k-1}^{ij-0l}, \tag{5.105}$$

$$\Sigma_{k|k-1}^{ij-l} = A_l \Sigma_{k-1|k-1}^{ij-0l} A_l^T + B_l B_l^T. \tag{5.106}$$

  * Measurement Update:

$$\begin{aligned}\hat{x}_{k|k}^{ij-l} &= \hat{x}_{k|k-1}^{ij-l} + \Sigma_{k|k-1}^{ij-l} C_l^T (S_k^{ij-l})^{-1}(y_k - C_l \hat{x}_{k|k-1}^{ij-l}), \\ \Sigma_{k|k}^{ij-l} &= \Sigma_{k|k-1}^{ij-l} - \Sigma_{k|k-1}^{ij-l} C_l^T (S_k^{ij-l})^{-1} C_l \Sigma_{k|k-1}^{ij-l}, \end{aligned}\tag{5.107}$$

$$d_k^{ij-l} = \frac{d_{k-1}^{ij-l}}{\psi(y_k)}\mathcal{N}(y_k; C_l \hat{x}_{k|k-1}^{ij-l}, S_k^{ij-l}), \tag{5.108}$$

$$S_k^{ij-l} = C_l \Sigma_{k|k-1}^{ij-l} C_l^T + D_l D_l^T. \tag{5.109}$$

– Estimate Calculation: The new transition probability estimates are calculated as

$$\hat{\pi}_{ij}^{(k)} = \frac{\sum_{l=1}^{N} d_k^{ij-l}}{\sum_{j=1}^{N}\sum_{l=1}^{N} d_k^{ij-l}} \quad \text{for} \quad 1 \le i, j \le N. \tag{5.110}$$

5.6.1 Modifications

The algorithm summarized above calculates the expected number of jumps at each step k using the previously estimated transition probabilities $\{\hat{\pi}_{ij}^{(k-1)}\}_{i,j=1}^{N}$ and then it updates the transition probability estimates with new ones $\{\hat{\pi}_{ij}^{(k)}\}_{i,j=1}^{N}$. The algorithm therefore works as/in a closed estimation loop. Since there are insufficient number of jumps observed between the states in the first few sampling instants to estimate the transition probabilities, most probability estimates erratically jump to either zero or unity in the unmodified version of the algorithm. It is therefore desirable not to close the estimation loop for a few samples at the initial part of the estimation. During this period, the algorithm uses the initial transition probabilities and does not update the transition probability estimates. After a sufficient number of sampling periods, the loop can be safely closed to update the transition probability estimates.

It is observed during the simulations that, the ever decaying characteristics of the coefficient $\frac{1}{k-1}$ in Eqns. 5.102 to 5.104 reduces the convergence speed significantly. Therefore, it is desirable to limit this decay. For this purpose, the coefficient $\frac{1}{k-1}$ is replaced by $\max\{\frac{1}{k-1}, \epsilon_D\}$ where $\epsilon_D \ll 1$ is called as the decay limit factor. This modification makes the algorithm converge faster and track the possible transition rate changes at the expense of making it slightly more prone to noise.

5.6.2 Joint Base-State Estimation

The approximate ML algorithm derived above can provide mode-conditioned state estimates and covariances along with overall approximate MMSE estimate and covariance as a by-product. Therefore, if the base-state estimates and covariances of the JMLS are also required, one need not execute an additional state estimator (like IMM of GPB2) which uses the online calculated transition probabilities. The mode-conditioned state estimates $\hat{x}_{k|k}^{m}$, covariances $\Sigma_{k|k}^{m}$ and mode-probabilities μ_k^m defined as

$$\hat{x}_{k|k}^{m} \triangleq E\{x_k|\mathcal{Y}_k, r_k = m, \widehat{\Theta}^{k-1}\}, \tag{5.111}$$

$$\Sigma^m_{k|k} \triangleq E\{(x_k - \hat{x}^m_{k|k})(x_k - \hat{x}^m_{k|k})^T | \mathcal{Y}_k, r_k = m, \widehat{\Theta}^{k-1}\}, \quad (5.112)$$

$$\mu^m_k \triangleq P\{r_k = m | \mathcal{Y}_k, \widehat{\Theta}^{k-1}\} \quad (5.113)$$

can be calculated using the ML estimator states $\hat{x}^{ij-l}_{k|k}$, covariances $\Sigma^{ij-l}_{k|k}$ and mode weights d^{ij-l}_k using the following formulas:

$$\hat{x}^m_{k|k} = \frac{1}{\sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k} \sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k \hat{x}^{ij-m}_{k|k}, \quad (5.114)$$

$$\Sigma^m_{k|k} = \frac{1}{\sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k} \sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k \Big[\Sigma^{ij-m}_{k|k} + (\hat{x}^{ij-m}_{k|k} - \hat{x}^m_{k|k})(\hat{x}^{ij-m}_{k|k} - \hat{x}^m_{k|k})^T\Big], \quad (5.115)$$

$$\mu^m_k = \frac{\sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k}{\sum_{m=1}^N \sum_{i=1}^N \sum_{j=1}^N d^{ij-m}_k}. \quad (5.116)$$

These formulas can be obtained easily as a direct consequence of the relationship given in Eqn. 5.29. The overall MMSE state estimates $\hat{x}_{k|k}$ and covariances $\Sigma_{k|k}$ defined as

$$\hat{x}_{k|k} \triangleq E\{x_k | \mathcal{Y}_k, \widehat{\Theta}^{k-1}\}, \quad (5.117)$$

$$\Sigma_{k|k} \triangleq E\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T | \mathcal{Y}_k, \widehat{\Theta}^{k-1}\} \quad (5.118)$$

can then be calculated using IMM (or GPB2) final output calculation formulas [13].

5.7 Simulation Results

This section illustrates the performance of the maximum likelihood estimator on the simulated scenarios used in Chapter 4. Here the properties of the examples are repeated for convenience. The first scenario is taken from [1] where the following system model is considered.

$$x_{k+1} = x_k + w_k, \quad (5.119)$$

$$y_k = (r_k - 1)x_k + (100 - 90(r_k - 1))v_k \quad (5.120)$$

where

$$x_0 \sim \mathcal{N}(x_0; 0, 20^2), \tag{5.121}$$

$$w_k \sim \mathcal{N}(w_k; 0, 2^2), \tag{5.122}$$

$$v_k \sim \mathcal{N}(v_k; 0, 1) \tag{5.123}$$

with x_0, w_k and v_k being mutually independent for $k = 1, 2, \ldots$. The model sequence $r_k \in \{1, 2\}$ is a first-order, two-state, homogeneous Markov process with probability transition matrix $\Pi = [\pi_{ij}]$ given as

$$\Pi = \begin{bmatrix} 0.6 & 0.4 \\ 0.85 & 0.15 \end{bmatrix}. \tag{5.124}$$

Note that this system corresponds to a system with frequent measurement failures with the modal-state $r_k = 1$ corresponding to the case of the failure. The modified ML algorithm is run on the simulated measurements of this system with initial transition probabilities $\hat{\pi}_{11}^{(0)} = \hat{\pi}_{22}^{(0)} = 0.5$ and the transition probabilities are estimated for 1000 Monte-Carlo runs. The length of open loop estimation period and the decay limit factor are selected as 20 and $\epsilon_D = 0.01$ respectively. The average estimation performance of the modified ML algorithm is given in Fig. 5.1.

Note that during open loop period which corresponds to the first twenty sampling instants, the initial transition probability estimates are not changed. Once the estimation loop is closed, the transition probability estimates jump to values which are near to their corresponding true values and then they continue to converge towards the true values.

The second example contains a hypothetical scalar jump Markov linear system which has three models. The parameters of the system are given as

- $A(1) = 0.8$, $A(2) = 0.9$, $A(3) = 1$
- $B(i) = 1$ for $i = 1, 2, 3$.
- $C(1) = 1$, $C(2) = 2$, $C(3) = 4$

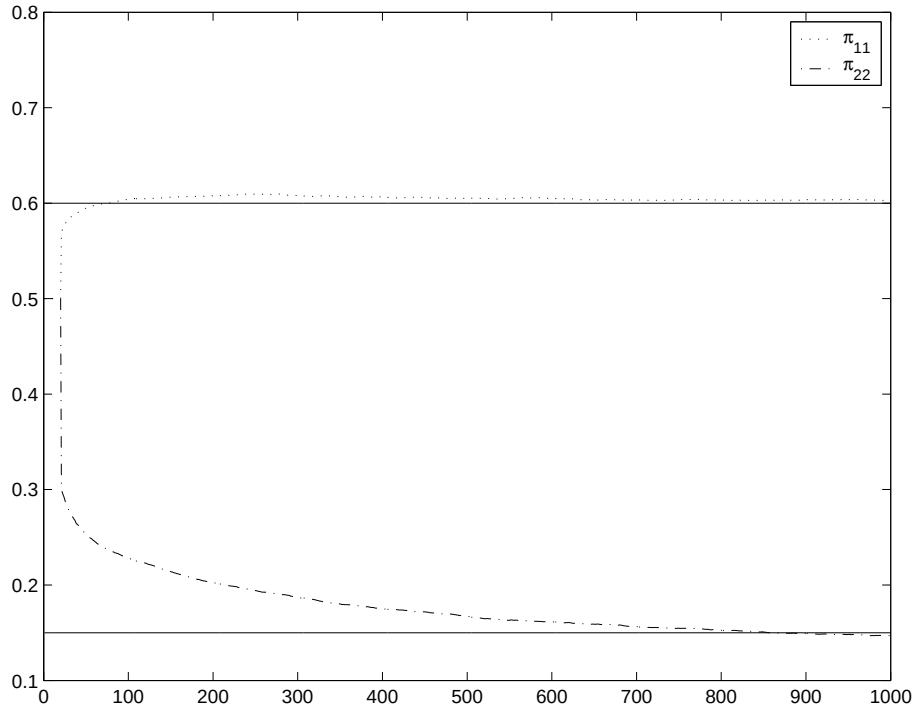

Figure 5.1: Average transition probability estimation performance of the modified maximum likelihood estimation algorithm.

- $D(i) = 1$ for $i = 1, 2, 3$.
- $x_0 \sim \mathcal{N}(x_0; 0, 2^2)$
- $w_k \sim \mathcal{N}(w_k; 0, 2^2)$
- $v_k \sim \mathcal{N}(v_k; 0, 1)$

and the true transition probability associated with the mode sequence r_k is taken as

$$\Pi = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.25 & 0.5 & 0.25 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}. \tag{5.125}$$

The initial probability transition matrix estimate $\widehat{\Pi}_0$ is selected as

$$\widehat{\Pi}_0 = \begin{bmatrix} 0.33 & 0.33 & 0.34 \\ 0.33 & 0.33 & 0.34 \\ 0.33 & 0.33 & 0.34 \end{bmatrix}. \tag{5.126}$$

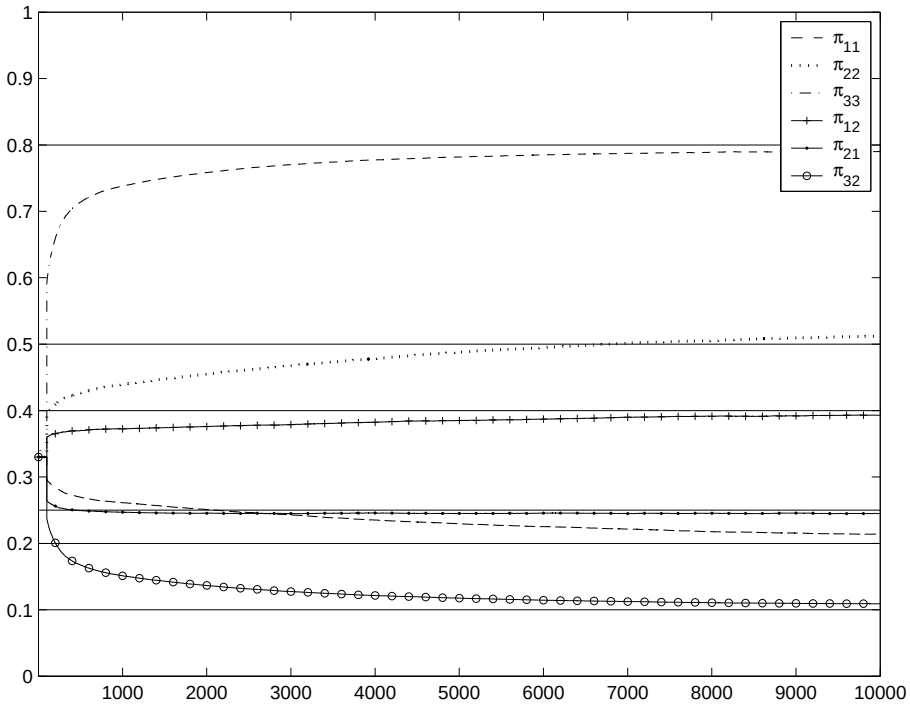

Figure 5.2: Average transition probability estimation performance of the modified maximum likelihood estimation algorithm.

The modified algorithm works with open estimation loop for the first 100 sampling instants and the decay limit factor is selected as $\epsilon_D = 0.001$. The average estimation performance for 1000 Monte-Carlo runs are shown in Fig. 5.2.

In this plot, there exists some bias in the estimated probabilities like the case in [1]. These can be attributed to the approximations involved in the derivation and they seem to be negligible.

5.7.1 Comparison with RKL Algorithm

In this subsection, we compare the estimation performance of the ML method to that of the Recursive Kullback-Leibler (RKL) method presented in Chapter 4. For this purpose, we plot the estimation performances of the methods on the same plot for the first example given above. For the RKL method, the constant step-size sequence is selected as $\epsilon_n = 0.02$ like the case in Chapter 4. The average transition probability estimation and typical single-run estimation

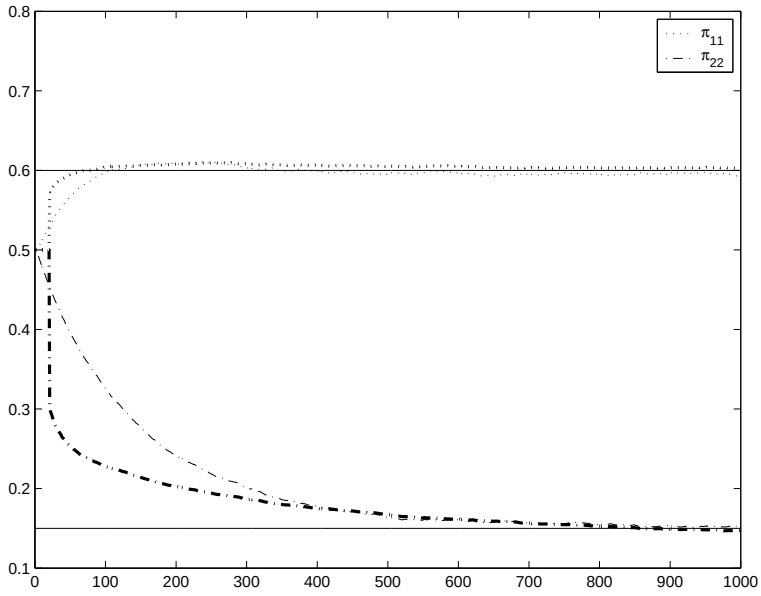

Figure 5.3: Average transition probability estimation performances of the modified maximum likelihood estimation algorithm and the RKL algorithm. The thicker and thinner lines denote the ML and RKL estimation results respectively.

performances of the methods are shown on Figs. 5.3 and 5.4 respectively.

According to the average performance plot, the initial convergence rate of ML algorithm is much faster than that of the RKL algorithm. However, in the long terms, the estimation performances get similar to each other. This situation suggests the usage of ML algorithm at the beginning of the estimation for fast convergence and then switching to the RKL algorithm with low computation for the rest of the estimation. The single-run comparison is also worth mentioning. Fig. 5.4 shows that the probability estimate trajectories of ML are less noise-prone than that of RKL. The noise susceptibility of the RKL method can be reduced by decreasing the step-size ϵ_n but this, in turn, would further reduce the convergence speed of the RKL method significantly.

5.8 Conclusions

This chapter proposed an approximate ML estimator for the transition probabilities associated with JMLSs. The estimator, which requires the mode

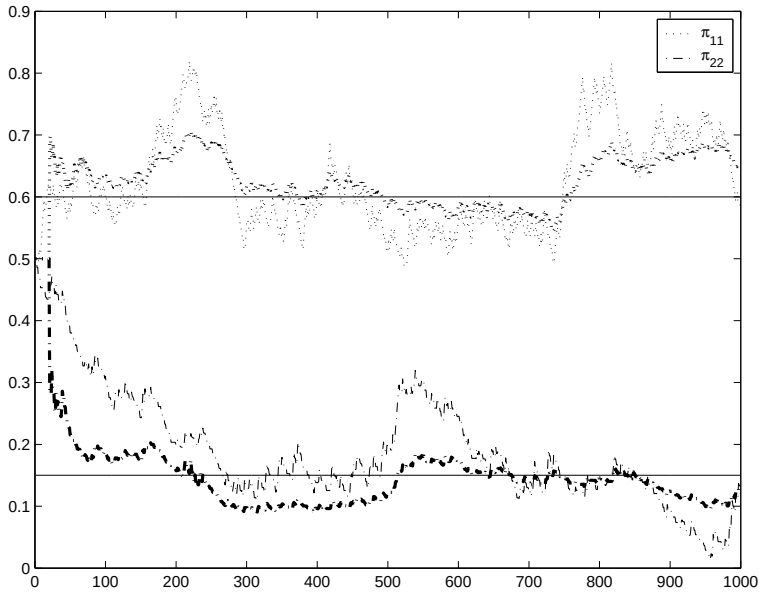

Figure 5.4: Single-run transition probability estimation performances of the modified maximum likelihood estimation algorithm and the RKL algorithm. The thicker and thinner lines denote the ML and RKL estimation results respectively.

weights of a N^3 component IMM filter to calculate the transition probabilities, can also supply the base-state estimates and covariances as a by-product. The simulation studies show that the estimator is faster (at least in the initial phases of the estimation) and less noise-prone than the RKL algorithm of Chapter 4.

CHAPTER 6

MIXED MULTIPLE MODEL ESTIMATION ALGORITHMS

6.1 Introduction

Multiple-model estimation techniques are mainly proposed for solving the problem of state estimation in the cases where the model uncertainties can be covered by a finite number of models. However, the computational cost of obtaining the optimal minimum mean-square error (MMSE) estimate of the state in multiple-model configurations increases exponentially in time [13].[1] Therefore; approximations are necessarily made to obtain suboptimal but computationally cheaper estimates. The most well-known examples of these suboptimal approaches are the Generalized Pseudo Bayesian (GPB) [7, 8] and the Interacting Multiple Model (IMM) [11, 13] algorithms. In these algorithms, the multiple component Gaussian mixtures are approximated by a single Gaussian, matching the mean and covariance of the densities. In this chapter, the effect of this approximation applied to the input of an optimal Bayesian filter is investigated. The first and second moments of the error between the estimate resulting from the approximation (which is the estimate of a Kalman filter) and MMSE-optimal estimate which is obtained from Bayesian density recursions are examined. An analytical expression for the covariance of the resulting error (due to the approximation) in the filter estimate which can be calculated before filtering, is found. This measure is then used as a tool for

[1] This fact was also observed in the IMM filter derivation of Sec. 3.3.

generating a mixed IMM-GPB2 algorithm.

In the literature, GPB2 and IMM algorithms are always used by selecting one algorithm or another for all times (that estimates are required). However, it is possible to generate hybrid algorithms by applying GPB2 or IMM interchangeably based on some criteria for each sampling time t_k. In other words, applying IMM for sampling time t_k (for all models of JMLS) and then if some criteria are satisfied applying GPB2 for time t_{k+1} (for all models of the JMLS) is possible. This application, once a useful criterion is selected, is straightforward. What is achieved in this chapter is to show that it is possible to extend the above mentioned hybrid scheme to a mixed scheme where at each time step t_k, the algorithm can select a different estimation procedure (either GPB2 or IMM) for each model of the JMLS. The parallel structures of these algorithms, which are composed of multiple Kalman filters, make this extension possible. For the criterion required for the selection of the algorithms is provided by the approximation analysis mentioned above. The resulting mixed algorithm can combine IMM and GPB2 algorithms while achieving the performance of GPB2 with less computational load.

The outline of the chapter is as follows. In Section 6.2, the problem definition is made and the optimal solution in terms of Bayesian density updates is given. The estimate after the approximation is obtained in Section 6.3. Section 6.4 is composed of the investigation of the first two moments of the error between optimal and the approximate estimates. We proceed with a brief review of the IMM and GPB2 algorithms in Sec.6.5 which is followed by the introduction of the mixed IMM-GPB2 algorithm in Section 6.6. In Section 6.7, the simulation results are presented for the mixed IMM-GPB2 algorithm. The chapter is finalized with conclusions in Section 6.8.

6.2 Problem Definition and Optimal Solution

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let x_k, y_k, w_k and v_k be random variables defined on this space. We assume that the probability density functions

for all of the defined random variables exist and that all the random variables are integrable. Suppose we are given the following discrete-time state-space representation

$$x_{k+1} = Ax_k + Bw_k, \tag{6.1}$$

$$y_k = Cx_k + Dv_k \tag{6.2}$$

where w_k and v_k are uncorrelated, white Gaussian noises with zero mean and covariance Q and R respectively. The prior density $p(x_0)$ for the initial state x_0 is the Gaussian mixture which is given below

$$p(x_0) = \sum_{i=1}^{N} p_i \mathcal{N}(x_0; \bar{x}_i, \Sigma_i) \tag{6.3}$$

where the notation $\mathcal{N}(x; \bar{x}, \Sigma)$ denotes the multivariate normal density with dummy variable x, mean $\bar{x}$ and covariance Σ. The component probabilities $\{p_i\}_{i=1}^{N}$ sum up to unity, i.e., $\sum_{i=1}^{N} p_i = 1$.

6.2.1 Problem Definition

We are interested in the optimal MMSE estimate $\hat{x}_{k|k}$ and its covariance $\Sigma_{k|k}$ given as

$$\hat{x}_{k|k} = E[x_k|Y^k], \tag{6.4}$$

$$\Sigma_{k|k} = E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T|Y^k] \tag{6.5}$$

where

$$Y^k \triangleq \{y_0, y_1, \ldots, y_k\}.^2 \tag{6.6}$$

In the remaining parts of the chapter, the prediction and measurement updates for the Bayesian (or Kalman) filters will be examined separately. In each case, we assume that we are at an intermediate stage of a recursive estimation process and our input density (information state to be updated) is an N-component Gaussian mixture. Therefore; in prediction (measurement) update,

[2] In this chapter, we do not discriminate between Y^k and $\mathcal{Y}_k$ which denotes the σ-algebra generated by the random variables $y_0, y_1, \ldots, y_k$ for the sake of simplicity.

it is assumed that the input density $p(x_{k-1}|Y^{k-1})$ ($p(x_k|Y^{k-1})$) is a Gaussian mixture. In a nonrecursive framework, the input densities correspond to prior density of the estimatee at the corresponding level of estimation.

6.2.2 Optimal Solution

The optimal solution to the problem stated above is well-known and given for example in [64] or [65]. In the following, we repeat these facts with their basic proofs.

6.2.2.1 Prediction Update

The general Bayesian density update equation for MMSE prediction update is given as follows [45].

$$p(x_k|Y^{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Y^{k-1})dx_{k-1}. \tag{6.7}$$

Let $p(x_{k-1}|Y^{k-1})$ be the N-component Gaussian mixture given below

$$p(x_{k-1}|Y^{k-1}) = \sum_{i=1}^{N} p_i \mathcal{N}(x_{k-1}; \bar{x}_i, \Sigma_i). \tag{6.8}$$

Substituting Eqn. 6.8 into Eqn. 6.7, we obtain

$$p(x_k|Y^{k-1}) = \int p(x_k|x_{k-1}) \sum_{i=1}^{N} p_i \mathcal{N}(x_{k-1}; \bar{x}_i, \Sigma_i)dx_{k-1} \tag{6.9}$$

$$= \sum_{i=1}^{N} p_i \int p(x_k|x_{k-1})\mathcal{N}(x_{k-1}; \bar{x}_i, \Sigma_i)dx_{k-1}. \tag{6.10}$$

Noting the system dynamics given in Eqn. 6.1, the conditional probability density function $p(x_k|x_{k-1})$ is given as

$$p(x_k|x_{k-1}) = \mathcal{N}(x_k; Ax_{k-1}, BQB^T). \tag{6.11}$$

Each integral in Eqn. 6.10 therefore can be evaluated easily using the Kalman filter prediction update formulas. Then,

$$p(x_k|Y^{k-1}) = \sum_{i=1}^{N} p_i \mathcal{N}(x_k; \bar{x}_i^-, \Sigma_i^-) \tag{6.12}$$

where

$$\bar{x}_i^- = A\bar{x}_i, \tag{6.13}$$

$$\Sigma_i^- = A\Sigma_i A^T + BQB^T \tag{6.14}$$

for $i = 1, 2$. The optimal MMSE prediction $\hat{x}^{op}_{k|k-1}$ and its covariance $\Sigma^{op}_{k|k-1}$ is therefore given by

$$\hat{x}^{op}_{k|k-1} = \sum_{i=1}^{N} p_i \bar{x}_i^-, \tag{6.15}$$

$$\Sigma^{op}_{k|k-1} = \sum_{i=1}^{N} p_i \left[\Sigma_i^- + (\bar{x}_i^- - \hat{x}^{op}_{k|k-1})(\bar{x}_i^- - \hat{x}^{op}_{k|k-1})^T\right]. \tag{6.16}$$

6.2.2.2 Measurement Update

The general Bayesian density update equation for MMSE measurement update is given as follows [45].

$$p(x_k|Y^k) = \frac{p(y_k|x_k)p(x_k|Y^{k-1})}{\int p(y_k|x_k)p(x_k|Y^{k-1})dx_k}. \tag{6.17}$$

Let $p(x_k|Y^{k-1})$ be the N-component Gaussian mixture given below.

$$p(x_k|Y^{k-1}) = \sum_{i=1}^{N} p_i \mathcal{N}(x_k; \bar{x}_i, \Sigma_i). \tag{6.18}$$

Substituting Eqn. 6.18 into Eqn. 6.17, we get

$$p(x_k|Y^k) = \frac{p(y_k|x_k)\sum_{i=1}^{N} p_i \mathcal{N}(x_k; \bar{x}_i, \Sigma_i)}{\int p(y_k|x_k)\sum_{i=1}^{N} p_i \mathcal{N}(x_k; \bar{x}_i, \Sigma_i)dx_k} \tag{6.19}$$

$$= \frac{\sum_{i=1}^{N} p_i p(y_k|x_k)\mathcal{N}(x_k; \bar{x}_i, \Sigma_i)}{\sum_{i=1}^{N} p_i \int p(y_k|x_k)\mathcal{N}(x_k; \bar{x}_i, \Sigma_i)dx_k}. \tag{6.20}$$

Using the Kalman filter measurement update formulas, the following expression can be written for each term in the numerator of the right hand side of Eqn. 6.20.

$$p(y_k|x_k)\mathcal{N}(x_k; \bar{x}_i, \Sigma_i) = \mathcal{N}(x_k; \bar{x}_i^+, \Sigma_i^+) \int p(y_k|x_k)\mathcal{N}(x_k; \bar{x}_i, \Sigma_i)dx_k \tag{6.21}$$

where

$$\bar{x}_i^+ = \bar{x}_i + K_i(y_k - C\bar{x}_i), \tag{6.22}$$

$$\Sigma_i^+ = \Sigma_i - K_i S_i K_i^T = \Sigma_i - \Sigma_i C^T S_i^{-1} C\Sigma_i \tag{6.23}$$

with

$$K_i = \Sigma_i C^T S_i^{-1} = \Sigma_i^+ C^T R^{-1}, \tag{6.24}$$

$$S_i = C\Sigma_i C^T + DRD^T. \tag{6.25}$$

Noting the measurement equation given in Eqn. 6.2, the conditional density $p(y_k|x_k)$ is given as

$$p(y_k|x_k) = \mathcal{N}(y_k; Cx_k, R). \tag{6.26}$$

Using this, each integral term in the denominator of Eqn. 6.20 (and also the integral in Eqn. 6.21) can be calculated as

$$\int p(y_k|x_k)\mathcal{N}(x_k; \bar{x}_i, \Sigma_i)dx_k = \mathcal{N}(y_k; \bar{y}_i, S_i) \tag{6.27}$$

where $\bar{y}_i = C\bar{x}_i$ and the definition of S_i is given in Eqn. 6.25. Using these results, the posterior density $p(x_k|Y^k)$ becomes

$$p(x_k|Y^k) = \frac{\sum_{i=1}^N p_i \mathcal{N}(y_k; \bar{y}_i, S_i)\mathcal{N}(x_k; \bar{x}_i^+, \Sigma_i^+)}{\sum_{i=1}^N p_i \mathcal{N}(y_k; \bar{y}_i, S_i)}. \tag{6.28}$$

Defining new measurement dependent probabilities p_i^+ as

$$p_i^+ \triangleq \frac{p_i \mathcal{N}(y_k; \bar{y}_i, S_i)}{\sum_{j=1}^N p_j \mathcal{N}(y_k; \bar{y}_j, S_j)}, \tag{6.29}$$

the posterior density $p(x_k|Y^k)$ is given as

$$p(x_k|Y^k) = \sum_{i=1}^N p_i^+ \mathcal{N}(x_k; \bar{x}_i^+, \Sigma_i^+). \tag{6.30}$$

Calculating, the mean and the covariance of this density gives us the following optimal filtered estimate $\hat{x}_{k|k}^{op}$ and covariance $\Sigma_{k|k}^{op}$ as

$$\hat{x}_{k|k}^{op} = \sum_{i=1}^N p_i^+ \bar{x}_i^+, \tag{6.31}$$

$$\Sigma_{k|k}^{op} = \sum_{i=1}^N p_i^+ \left[\Sigma_i^+ + (\bar{x}_i^+ - \hat{x}_{k|k}^{op})(\bar{x}_i^+ - \hat{x}_{k|k}^{op})^T\right]. \tag{6.32}$$

6.3 Approximate Solution

In this section, we are going to approximate the Gaussian mixtures which are input to the prediction and measurement updates of the optimal Bayesian recursions in the previous section by a single Gaussian whose mean and covariance are matched to those of the Gaussian mixture. Therefore, we will assume in each update, the input Gaussian mixture

$$p(x) = \sum_{i=1}^{N} p_i \mathcal{N}(x, \bar{x}_i, \Sigma_i) \tag{6.33}$$

is approximated by

$$p_{app}(x) = \mathcal{N}(x, \bar{x}_{app}, \Sigma_{app}) \tag{6.34}$$

where

$$\bar{x}_{app} \triangleq \sum_{i=1}^{N} p_i \bar{x}_i, \tag{6.35}$$

$$\Sigma_{app} \triangleq \sum_{i=1}^{N} p_i \left[\Sigma_i + (\bar{x}_i - \bar{x}_{app})(\bar{x}_i - \bar{x}_{app})^T\right]. \tag{6.36}$$

The Bayesian recursions after this approximation turn into simply the standard Kalman filter updates.

6.3.1 Prediction Update

Now, assuming that the input Gaussian mixture

$$p(x_{k-1}|Y^{k-1}) = \sum_{i=1}^{N} p_i \mathcal{N}(x_{k-1}, \bar{x}_i, \Sigma_i) \tag{6.37}$$

is approximated by a single Gaussian

$$p(x_{k-1}|Y^{k-1}) \approx \mathcal{N}(x_{k-1}; \bar{x}_{app}, \Sigma_{app}) \tag{6.38}$$

where $\bar{x}_{app}$ and Σ_{app} are given in Eqn. 6.35 and Eqn. 6.36 respectively. Then the suboptimal prediction resulting from this approximation is

$$\hat{x}^{sub}_{k|k-1} = A\bar{x}_{app}, \tag{6.39}$$

$$\Sigma^{sub}_{k|k-1} = A\Sigma_{app}A^T + BQB^T. \tag{6.40}$$

Substituting $\bar{x}_{app}$ given in Eqn. 6.35 into Eqn. 6.39

$$\hat{x}^{sub}_{k|k-1} = A\sum_{i=1}^{N} p_i \bar{x}_i = \sum_{i=1}^{N} p_i A\bar{x}_i = \sum_{i=1}^{N} p_i \bar{x}_i^- = \hat{x}^{op}_{k|k-1} \tag{6.41}$$

Substituting Σ_{app} in Eqn. 6.36 into Eqn. 6.40, we obtain

$$\begin{aligned}
\Sigma^{sub}_{k|k-1} &= A\sum_{i=1}^{N} p_i \left[\Sigma_i + (\bar{x}_i - \bar{x}_{app})(\bar{x}_i - \bar{x}_{app})^T\right] A^T + BQB^T & (6.42)\\
&= \sum_{i=1}^{N} p_i \left[A\Sigma_i A^T + BQB^T + A(\bar{x}_i - \bar{x}_{app})(\bar{x}_i - \bar{x}_{app})^T A^T\right] & (6.43)\\
&= \sum_{i=1}^{N} p_i \left[\Sigma_i^- + (\bar{x}_i^- - A\bar{x}_{app})(\bar{x}_i^- - A\bar{x}_{app})^T\right] & (6.44)\\
&= \sum_{i=1}^{N} p_i \left[\Sigma_i^- + (\bar{x}_i^- - \hat{x}^{sub}_{k|k-1})(\bar{x}_i^- - \hat{x}^{sub}_{k|k-1})^T\right] & (6.45)\\
&= \sum_{i=1}^{N} p_i \left[\Sigma_i^- + (\bar{x}_i^- - \hat{x}^{op}_{k|k-1})(\bar{x}_i^- - \hat{x}^{op}_{k|k-1})^T\right] = \Sigma^{op}_{k|k-1}. & (6.46)
\end{aligned}$$

These two results prove that the Kalman filter prediction update yields the same first two moments as the Bayesian density recursion under the moment matched single Gaussian approximation. In other words, the single Gaussian approximation causes no change in the MMSE estimate and covariance. This invariance property of the Kalman filter prediction update is already known and is the main motive underlying the IMM-type hypothesis merging (instead of GPB2-type merging) [13].

6.3.2 Measurement Update

Assuming that the input Gaussian mixture

$$p(x_k|Y^{k-1}) = \sum_{i=1}^{N} p_i \mathcal{N}(x_k, \bar{x}_i, \Sigma_i) \tag{6.47}$$

is approximated by a single Gaussian

$$p(x_k|Y^{k-1}) \approx \mathcal{N}(x_k; \bar{x}_{app}, \Sigma_{app}) \tag{6.48}$$

where $\bar{x}_{app}$ and Σ_{app} are given in Eqn. 6.35 and Eqn. 6.36 respectively. Then, the Kalman filter measurement update gives the suboptimal estimate and covariance as

$$\hat{x}^{sub}_{k|k} = \bar{x}_{app} + K_{sub}(y_k - C\bar{x}_{app}), \tag{6.49}$$

$$\Sigma^{sub}_{k|k} = \Sigma_{app} - K_{sub}S_{sub}K^T_{sub} = \Sigma_{app} - \Sigma_{app}C^T S^{-1}_{sub}C\Sigma_{app} \tag{6.50}$$

where

$$K_{sub} = \Sigma_{app}C^T S^{-1}_{sub} = \Sigma^{sub}_{k|k}C^T R^{-1}, \tag{6.51}$$

$$S_{sub} = C\Sigma_{app}C^T + DRD^T = \sum_{i=1}^{N} p_i \left[S_i + (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^T\right] \tag{6.52}$$

$$\bar{y} = \sum_{i=1}^{N} p_i \bar{y}_i. \tag{6.53}$$

Unfortunately, these equations show that the optimal estimate and covariance given in Eqn. 6.31 and Eqn. 6.32 respectively are different than the suboptimal ones above. Therefore, under the moment-matched single Gaussian approximation, using the standard Kalman filter measurement update (unlike the prediction update) causes our final estimate to deviate from the optimal MMSE one.

This type of approximation appears in IMM and GPB filtering and possibly in other multiple-model estimation applications. Specifically, this approximation is made during the mixing process[3] of the IMM filter at the input of each component Kalman filter and it is the main approximation that makes the IMM filter different from the GPB2 filter. Up to now, the difference between the two filters was analyzed only by means of Monte-Carlo runs [11] and to the author's knowledge, this chapter's work is the first one which tries to quantify the deviation analytically. In the next section, the first two moments of the error caused by the moment-matched single Gaussian approximation will be examined.

[3] See the derivation of the IMM filter in Sec. 3.3 for an illustration of the mixing process.

6.4 Error Analysis

In the previous section, it was shown that the Kalman filter measurement update results in erroneous estimate and covariance under moment-matched single Gaussian approximation unlike the prediction update whose resulting estimate (i.e, prediction) is invariant under the same approximation. Defining the error Δ caused by the measurement update[4] as

$$\Delta \triangleq \hat{x}_{k|k}^{sub} - \hat{x}_{k|k}^{op}, \tag{6.54}$$

in this section, we are interested in the conditional expected value

$$\overline{\Delta} \triangleq E_{y_k}[\Delta|Y^{k-1}] \tag{6.55}$$

and the conditional covariance

$$\Sigma_\Delta \triangleq E_{y_k}[(\Delta - \overline{\Delta})(\Delta - \overline{\Delta})^T|Y^{k-1}] \tag{6.56}$$

of this error. Note that the expectations given above are to be taken with respect to y_k which has a conditional density given as

$$p(y_k|Y^{k-1}) = \int p(y_k|x_k)p(x_k|Y^{k-1})dx_k = \sum_{j=1}^{N} p_j \mathcal{N}(y_k; \bar{y}_j, S_j). \tag{6.57}$$

Therefore; the density with respect to which the expectations are to be taken is also a Gaussian mixture.

6.4.1 Calculation of the Mean of Δ

In order to find $\overline{\Delta}$, we need the expected values of the the optimal and sub-optimal estimates.

$$\begin{aligned} E_{y_k}[\hat{x}_{k|k}^{sub}|Y^{k-1}] &= E_{y_k}[\bar{x}_{app} + K_{sub}\left(y_k - C\bar{x}_{app}\right)|Y^{k-1}] && (6.58)\\ &= \bar{x}_{app} + K_{sub}\left(E_{y_k}[y_k|Y^{k-1}] - C\bar{x}_{app}\right) && (6.59)\\ &= \bar{x}_{app} + K_{sub}\left(\sum_{i=1}^{N} p_i\bar{y}_i - C\bar{x}_{app}\right). && (6.60) \end{aligned}$$

[4] In an IMM framework, this error may correspond to the error caused in the filtered estimates of individual (component) Kalman filters by the moment-matched single Gaussian approximation made in the mixing process at the input of the filters.

Replacing $C\bar{x}_{app}$ by $\sum_{i=1}^{N} p_i \bar{y}_i$,

$$\boxed{E_{y_k}[\hat{x}_{k|k}^{sub}|Y^{k-1}] = \bar{x}_{app}.} \tag{6.61}$$

Note that this is just a rephrasing of the fact that

$$E_{y_k}[\hat{x}_{k|k}|Y^{k-1}] = E_{y_k}[E[x_k|Y^k]|Y^{k-1}] = E[x_k|Y^{k-1}] = \hat{x}_{k|k-1}. \tag{6.62}$$

In a more technical language, this is the manifestation of the fact that, for an integrable random variable ξ, the sequence of random variables ξ_n defined as

$$\xi_n \triangleq E[\xi|\mathcal{F}_n] \tag{6.63}$$

is a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n=1}^{\infty}$[66]. In our case, $\xi = x_k$ and $\mathcal{F}_n$ is the σ-field generated by Y^n (i.e., $\mathcal{Y}_n$). Another interpretation is that the Kalman filter and the optimal Bayesian density measurement updates are unbiased estimators. Using the same fact, we can easily conclude that

$$E_{y_k}[\hat{x}_{k|k}^{op}|Y^{k-1}] = \bar{x}_{app}. \tag{6.64}$$

However, we are going to prove this fact the long way as well to gain some insight about the quantities we are dealing with.

$$\begin{aligned} E_{y_k}[\hat{x}_{k|k}^{op}|Y^{k-1}] &= E_{y_k}\left[\sum_{i=1}^{N} p_i^+ \bar{x}_i^+ \middle| Y^{k-1}\right] && (6.65)\\ &= E_{y_k}\left[\sum_{i=1}^{N} p_i^+ \left[\bar{x}_i + K_i(y_k - C\bar{x}_i)\right] \middle| Y^{k-1}\right] && (6.66)\\ &= \sum_{i=1}^{N} E_{y_k}[p_i^+|Y^{k-1}]\left[I - K_i C\right]\bar{x}_i + K_i E_{y_k}[p_i^+ y_k|Y^{k-1}]. \end{aligned}$$

Therefore; for the evaluation of this expected value, we need other intermediate expected values $E_{y_k}[p_i^+|Y^{k-1}]$ and $E_{y_k}[p_i^+ y_k|Y^{k-1}]$. Note that

$$p_i^+ \triangleq \frac{p_i \mathcal{N}(y_k; \bar{y}_i, S_i)}{\sum_{j=1}^{N} p_j \mathcal{N}(y_k; \bar{y}_j, S_j)} = \frac{p_i \mathcal{N}(y_k; \bar{y}_i, S_i)}{p(y_k|Y^{k-1})}. \tag{6.67}$$

Then,

$$E_{y_k}[p_i^+|Y^{k-1}] \triangleq \int p_i^+ p(y_k|Y^{k-1}) dy_k = p_i \int \mathcal{N}(y_k; \bar{y}_i, S_i) dy_k = p_i. \tag{6.68}$$

As a result,

$$\boxed{E_{y_k}[p_i^+|Y^{k-1}] = p_i.} \tag{6.69}$$

In the same way,

$$E_{y_k}[p_i^+ y_k|Y^{k-1}] \triangleq \int p_i^+ y_k p(y_k|Y^{k-1})dy_k = p_i \int y_k \mathcal{N}(y_k; \bar{y}_i, S_i)dy_k = p_i \bar{y}_i.$$

Consequently,

$$\boxed{E_{y_k}[p_i^+ y_k|Y^{k-1}] = p_i \bar{y}_i.} \tag{6.70}$$

Using these expected values, we calculate the expected value of the optimal filtered estimate $\hat{x}_{k|k}^{op}$ as

$$E_{y_k}[\hat{x}_{k|k}^{op}|Y^{k-1}] = \sum_{i=1}^{N} p_i \left[I - K_i C\right] \bar{x}_i + p_i K_i \bar{y}_i = \sum_{i=1}^{N} p_i \bar{x}_i = \bar{x}_{app}. \tag{6.71}$$

Having calculated the required expected values of the sub-optimal and optimal estimates, the mean of Δ is given as

$$\overline{\Delta} \triangleq E_{y_k}[\Delta|Y^{k-1}] = 0 \tag{6.72}$$

which means that the moment-matched single Gaussian approximation does not cause any bias in the estimate.

6.4.2 Calculation of the Covariance of Δ

The covariance Σ_Δ of Δ then can be calculated by

$$\Sigma_\Delta = E_{y_k}[\Delta\Delta^T|Y^{k-1}] \tag{6.73}$$

where

$$\begin{aligned}
\Delta &\triangleq \hat{x}_{k|k}^{sub} - \hat{x}_{k|k}^{op} = \bar{x}_{app} + K_{sub}(y_k - C\bar{x}_{app}) - \textstyle\sum_{i=1}^{N} p_i^+ \bar{x}_i^+ && (6.74)\\
&= \bar{x}_{app} + K_{sub}(y_k - \bar{y}) - \textstyle\sum_{i=1}^{N} p_i^+ \left[\bar{x}_i + K_i\left(y_k - C\bar{x}_i\right)\right] && (6.75)\\
&= \bar{x}_{app} + K_{sub}(y_k - \bar{y}) - \textstyle\sum_{i=1}^{N} p_i^+ \left[\bar{x}_i + K_i\left(y_k - \bar{y}_i\right)\right]. && (6.76)
\end{aligned}$$

Since the expression is too long, we are going to assign the terms to some auxiliary variables. Define

$$T_1 \triangleq \bar{x}_{app}, \tag{6.77}$$

$$T_2 \triangleq K_{sub}(y_k - \bar{y}), \tag{6.78}$$

$$T_3 \triangleq \sum_{i=1}^{N} p_i^+ \bar{x}_i, \tag{6.79}$$

$$T_4 \triangleq \sum_{i=1}^{N} p_i^+ K_i \left(y_k - \bar{y}_i\right). \tag{6.80}$$

Then,

$$\Sigma_\Delta = E_{y_k}[(T_1 + T_2 - T_3 - T_4)\,(T_1 + T_2 - T_3 - T_4)^T \,\big|\, Y^{k-1}]. \tag{6.81}$$

Thus, the covariance calculation requires the expected values of the form

$$E_{y_k}\left[T_i T_j^T | Y^{k-1}\right] \quad \text{for} \quad i,j = 1,\ldots,4. \tag{6.82}$$

The evaluation of these expected values involves second order marginal and cross moments of the posterior probabilities p_i^+ and y_k with respect to the density $p(y_k|Y^{k-1})$. These expected values are extremely difficult (if not impossible) to evaluate analytically, and therefore; some approximations has to be made. The evaluation of the second marginal moment of the probability p_i^+ shown below illustrates the approximation used to calculate these integrals.

$$E_{y_k}[(p_i^+)^2|Y^{k-1}] \triangleq \int (p_i^+)^2 p(y_k|Y^{k-1}) dy_k \tag{6.83}$$

$$= \int \frac{p_i^2 \mathcal{N}^2(y_k;\bar{y}_i,S_i)}{\sum_{j=1}^{N} p_j \mathcal{N}(y_k;\bar{y}_j,S_j)} dy_k \tag{6.84}$$

$$= p_i \int \frac{p_i \mathcal{N}(y_k;\bar{y}_i,S_i)}{\sum_{j=1}^{N} p_j \mathcal{N}(y_k;\bar{y}_j,S_j)} \mathcal{N}(y_k;\bar{y}_i,S_i) dy_k. \tag{6.85}$$

At this point, we see that the numerator of the integrand is the square of a Gaussian density which decreases quite fast. Due to this, the integration is effectively around the mean value $\bar{y}_i$. Assuming that the means $\bar{y}_i$ of the Gaussian components are separated sufficiently, in the effective integration range, the Gaussian mixture in the denominator can be approximated as

$$\sum_{j=1}^{N} p_j \mathcal{N}(y_k;\bar{y}_j,S_j) \approx p_i \mathcal{N}(y_k;\bar{y}_i,S_i). \tag{6.86}$$

This corresponds to assuming that the measurement dependent posterior probability p_i^+ is approximately unity around the mean $\bar{y}_i$ within the 2σ covariance

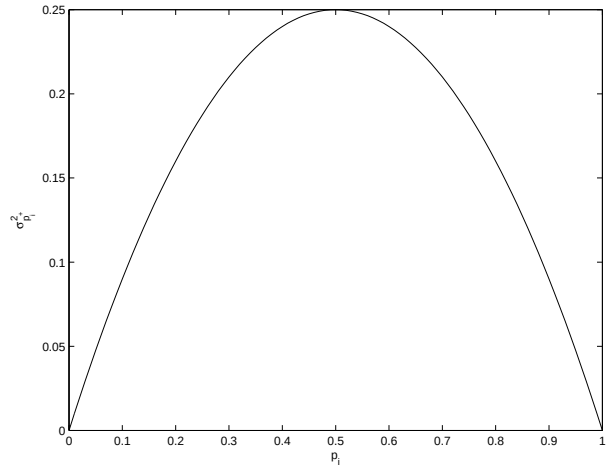

Figure 6.1: Variance of the probability p_i^+ vs. p_i

ellipse. Many other approximation schemes such as Taylor series expansions etc. are also possible but they turn out to yield negative variance values for the probabilities p_i^+ in some extreme cases. After the approximation, we get

$$E_{y_k}[(p_i^+)^2|Y^{k-1}] \approx p_i \int \mathcal{N}(y_k; \bar{y}_i, S_i) dy_k = p_i. \tag{6.87}$$

Using this identity, we can calculate the variance $\sigma^2_{p_i^+}$ of the probability p_i^+ as

$$\sigma^2_{p_i^+} = E_{y_k}[(p_i^+)^2|Y^{k-1}] - E^2_{y_k}[p_i^+|Y^{k-1}] \approx p_i - p_i^2. \tag{6.88}$$

This variance vs. the prior probability p_i is shown in Fig.6.1.

With the help of the same approximation, the expected values defined in Eqn. 6.82 can be evaluated easily.[5] After some involved algebraic manipulations which are presented in App. D, the covariance of the error resulting from moment-matched single Gaussian approximation is found to be

$$\Sigma_\Delta = \sum_{i=1}^{N} p_i \Sigma_i C^T S_i^{-1} C \Sigma_i - \Sigma_{app} C^T S_{sub}^{-1} C \Sigma_{app} + \sum_{i=1}^{N} p_i \bar{x}_i \bar{x}_i^T - \bar{x}_{app} \bar{x}_{app}^T \tag{6.89}$$

where

$$\bar{x}_{app} = \sum_{i=1}^{N} p_i \bar{x}_i, \tag{6.90}$$

$$\bar{y} = C \bar{x}_{app}, \tag{6.91}$$

[5] Evaluation of these expectations are investigated in full detail in App. D.

$$S_{sub} = \sum_{i=1}^{N} p_i \left[S_i + (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^T \right], \tag{6.92}$$

$$\Sigma_{app} = \sum_{i=1}^{N} p_i \left[\Sigma_i + (\bar{x}_i - \bar{x}_{app})(\bar{x}_i - \bar{x}_{app})^T \right]. \tag{6.93}$$

6.5 Review of the IMM and GPB2 Algorithms

In this section, we are going to present a brief review of IMM and GPB2 algorithms. In this way, the approximations made in both of the algorithms and the differences will be illustrated. We consider the following JMLS model.

$$x_k = A_{r_k} x_{k-1} + B_{r_k} w_k, \tag{6.94}$$

$$y_k = C_{r_k} x_k + D_{r_k} v_k. \tag{6.95}$$

The parameter matrices $\{A_i, B_i, C_i, D_i\}$ are assumed to be known for $i \in \{1, \ldots, N\}$ and $r_k \in \{1, \ldots, N\}$ is a finite-state Markov chain with initial distribution $\pi_0 = \left[\pi_0^1, \pi_0^2 \ldots, \pi_0^N\right]$ and probability transition matrix $\Pi = [\pi_{ij}]_{i,j=1}^N$. The noise processes w_k and v_k are white, uncorrelated and normally distributed with zero mean and covariances Q and R respectively. The following descriptions of the IMM and GPB2 algorithms have been adapted from [13].

6.5.1 Steps of the GPB2 Algorithm

At each time step, an N-model GPB2 algorithm keeps N filtered estimates $\{\hat{x}_{k|k}^j\}_{j=1}^N$, covariances $\{\Sigma_{k|k}^j\}_{j=1}^N$ and mode probabilities $\{\mu_k^j\}_{j=1}^N$. At each cycle, given the previously kept filtered estimates $\{\hat{x}_{k-1|k-1}^j\}_{j=1}^N$, covariances $\{\Sigma_{k-1|k-1}^j\}_{j=1}^N$ and the mode probabilities $\{\mu_{k-1}^j\}_{j=1}^N$, the algorithm calculates the updated values of these quantities using the following steps:

- **Mode-Matched Kalman Filtering:** The algorithm takes each previous filtered estimate $\hat{x}_{k-1|k-1}^i$ and covariance $\Sigma_{k-1|k-1}^i$ and executes N Kalman filters each matched to a different model. All N Kalman filters use the filtered estimate $\hat{x}_{k-1|k-1}^i$ and covariance $\Sigma_{k-1|k-1}^i$ as their initial state and covariance. Since there are N previous filtered estimates, a

total of N^2 Kalman filters are executed. At the end of the filtering, new filtered estimates $\hat{x}^{ij}_{k|k}$ and covariances $\Sigma^{ij}_{k|k}$ are obtained. The filtering equations are given as follows.

$$\begin{aligned}
\hat{x}^{ij}_{k|k-1} &= A_j \hat{x}^{i}_{k-1|k-1}, && (6.96)\\
\Sigma^{ij}_{k|k-1} &= A_j \Sigma^{i}_{k-1|k-1} A_j^T + B_j Q B_j^T, && (6.97)\\
S^{ij}_{k} &= C_j \Sigma^{ij}_{k|k-1} C_j^T + D_j R D_j^T, && (6.98)\\
K^{ij}_{k} &= \Sigma^{ij}_{k|k-1} C_j^T \left(S^{ij}_{k}\right)^{-1}, && (6.99)\\
\hat{x}^{ij}_{k|k} &= \hat{x}^{ij}_{k|k-1} + K^{ij}_{k}(y_k - C_j \hat{x}^{ij}_{k|k-1}), && (6.100)\\
\Sigma^{ij}_{k|k} &= \Sigma^{ij}_{k|k-1} - K^{ij}_{k} S^{ij}_{k} \left(K^{ij}_{k}\right)^T. && (6.101)
\end{aligned}$$

- **Calculation of the Merging Probabilities:** At this step, the merging probabilities $\{\mu^{ij}_{k|k-1}\}_{i=1}^N$ are calculated for each j. These probabilities are used to merge the estimates $\{\hat{x}^{ij}_{k|k}\}_{i=1}^N$ and covariances $\{\Sigma^{ij}_{k|k}\}_{i=1}^N$ for each j. The probabilities are calculated as follows,

$$\mu^{ij}_{k|k-1} = \frac{1}{c_j} \Delta^{ij}_k \pi_{ij} \mu^{i}_{k-1} \qquad (6.102)$$

where

$$\begin{aligned}
c_j &= \sum_{i=1}^{N} \Delta^{ij}_k \pi_{ij} \mu^{i}_{k-1}, && (6.103)\\
\Delta^{ij}_k &= \mathcal{N}(y_k; C_j \hat{x}^{ij}_{k|k-1}, S^{ij}_k). && (6.104)
\end{aligned}$$

- **Merging:** N filtered estimates $\{\hat{x}^{ij}_{k|k}\}_{i=1}^N$ and covariances $\{\Sigma^{ij}_{k|k}\}_{i=1}^N$ are merged for each j and the new filtered estimates $\{\hat{x}^{j}_{k|k}\}_{j=1}^N$ and covariances $\{\Sigma^{j}_{k|k}\}_{j=1}^N$ are obtained. The merging is done as follows.

$$\begin{aligned}
\hat{x}^{j}_{k|k} &= \sum_{i=1}^{N} \mu^{ij}_{k-1|k} \hat{x}^{ij}_{k|k}, && (6.105)\\
\Sigma^{j}_{k|k} &= \sum_{i=1}^{N} \mu^{ij}_{k-1|k} \left[\Sigma^{ij}_{k|k} + (\hat{x}^{ij}_{k|k} - \hat{x}^{j}_{k|k})(\hat{x}^{ij}_{k|k} - \hat{x}^{j}_{k|k})^T\right]. && (6.106)
\end{aligned}$$

- **Mode Probability Update:** The previous mode probabilities $\{\mu^{i}_{k-1}\}_{i=1}^N$ are updated to obtain the new mode probabilities $\{\mu^{j}_{k}\}_{j=1}^N$ as follows.

$$\mu^{j}_{k} = \frac{1}{c} \sum_{i=1}^{N} \Delta^{ij}_k \pi_{ij} \mu^{i}_{k-1} \qquad (6.107)$$

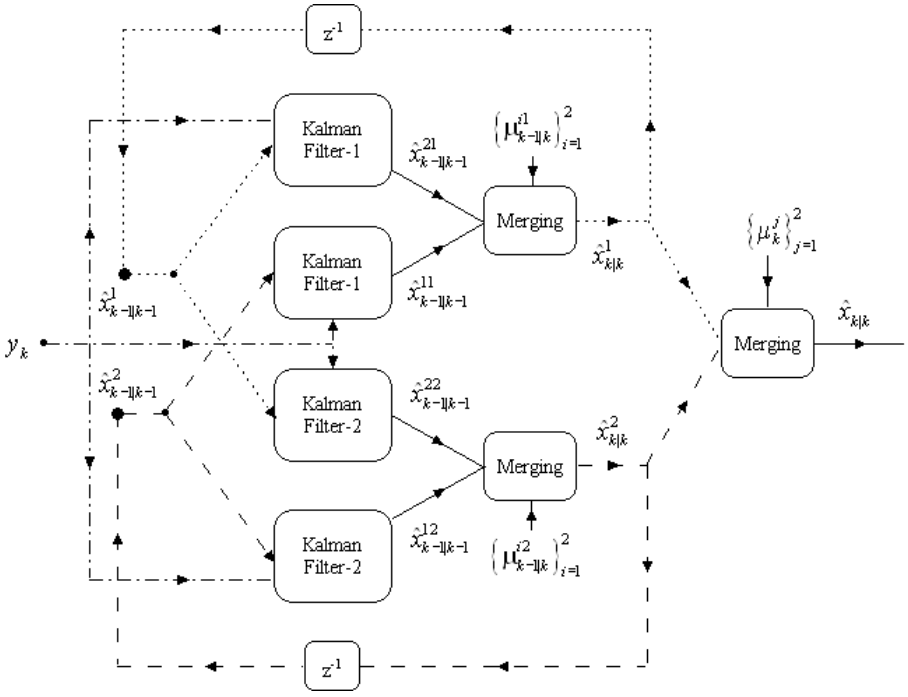

Figure 6.2: Block diagram of a two-model GPB2 algorithm.

where

$$c = \sum_{j=1}^{N} \sum_{i=1}^{N} \Delta_k^{ij} \pi_{ij} \mu_{k-1}^i. \tag{6.108}$$

- **Output Estimate and Covariance Calculation by Merging:** The output estimate and covariance are calculated by merging the filtered estimates $\{\hat{x}_{k|k}^j\}_{j=1}^N$ and covariances $\{\Sigma_{k|k}^j\}_{j=1}^N$. The merging is done using the updated mode probabilities $\{\mu_k^j\}_{j=1}^N$ as follows.

$$\hat{x}_{k|k} = \sum_{j=1}^{N} \mu_k^j \hat{x}_{k|k}^j, \tag{6.109}$$

$$\Sigma_{k|k} = \sum_{j=1}^{N} \mu_k^j \left[\Sigma_{k|k}^j + (\hat{x}_{k|k}^j - \hat{x}_{k|k})(\hat{x}_{k|k}^j - \hat{x}_{k|k})^T \right]. \tag{6.110}$$

The steps of the GPB2 algorithm related with the filtered state estimates are summarized for a two-model GPB2 algorithm in Fig. 6.2.

6.5.2 Steps of the IMM Algorithm

At each time step, an N-model IMM algorithm keeps N filtered estimates $\{\hat{x}^j_{k|k}\}^N_{j=1}$, covariances $\{\Sigma^j_{k|k}\}^N_{j=1}$ and mode probabilities $\{\mu^j_k\}^N_{j=1}$. At each cycle, given the previously kept filtered estimates $\{\hat{x}^j_{k-1|k-1}\}^N_{j=1}$, covariances $\{\Sigma^j_{k-1|k-1}\}^N_{j=1}$ and the mode probabilities $\{\mu^j_{k-1}\}^N_{j=1}$, the algorithm calculates the updated values of these quantities using the following steps:

- **Calculation of Mixing Probabilities:** In contrast to GPB2, IMM algorithm merges the previous filtered estimates $\{\hat{x}^j_{k-1|k-1}\}^N_{j=1}$ and covariances $\{\Sigma^j_{k-1|k-1}\}^N_{j=1}$ to obtain N-different initial estimates $\{\hat{x}^{0j}_{k-1|k-1}\}^N_{j=1}$ and the covariances $\{\Sigma^{0j}_{k-1|k-1}\}^N_{j=1}$ for the mode-matched Kalman filters. This process is called as *mixing* and the merging probabilities used for this purpose are called as the *mixing probabilities*. The mixing probabilities are calculated as follows.

$$\mu^{ij}_{k-1|k-1} = \frac{1}{\bar{c}_j}\pi_{ij}\mu^i_{k-1} \tag{6.111}$$

  where

$$\bar{c}_j = \sum_{i=1}^{N} \pi_{ij}\mu^i_{k-1}. \tag{6.112}$$

- **Mixing:** The previous filtered estimates $\{\hat{x}^j_{k-1|k-1}\}^N_{j=1}$ and covariances $\{\Sigma^j_{k-1|k-1}\}^N_{j=1}$ are merged to obtain N-different initial estimates $\{\hat{x}^{0j}_{k-1|k-1}\}^N_{j=1}$ and the covariances $\{\Sigma^{0j}_{k-1|k-1}\}^N_{j=1}$ as follows.

$$\hat{x}^{0j}_{k-1|k-1} = \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1}\hat{x}^i_{k-1|k-1}, \tag{6.113}$$

$$\Sigma^{0j}_{k-1|k-1} = \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1}\Big[\Sigma^i_{k-1|k-1} + (\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})(\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})^T\Big]. \tag{6.114}$$

  Note that the mixing process described above represents the moment-matched single Gaussian approximation.

- **Mode-Matched Kalman Filtering:** The algorithm takes each initial estimate $\hat{x}^{0j}_{k-1|k-1}$ and covariance $\Sigma^{0j}_{k-1|k-1}$ and executes N Kalman filters

each matched to a different model. All N Kalman filters uses its corresponding initial estimate $\hat{x}^{0j}_{k-1|k-1}$ and covariance $\Sigma^{0j}_{k-1|k-1}$ as their initial state and covariance. Since there are N initial conditions, N Kalman filters are executed. At the end of this filtering the new filtered estimates $\hat{x}^{j}_{k|k}$ and covariances $\hat{x}^{j}_{k|k}$ are obtained. The filtering equations are given as follows.

$$\begin{aligned}
\hat{x}^{j}_{k|k-1} &= A_j \hat{x}^{0j}_{k-1|k-1}, && (6.115)\\
\Sigma^{j}_{k|k-1} &= A_j \Sigma^{0j}_{k-1|k-1} A_j^T + B_j Q B_j^T, && (6.116)\\
S^{j}_{k} &= C_j \Sigma^{j}_{k|k-1} C_j^T + D_j R D_j^T, && (6.117)\\
K^{j}_{k} &= \Sigma^{j}_{k|k-1} C_j^T \left(S^{j}_{k}\right)^{-1}, && (6.118)\\
\hat{x}^{j}_{k|k} &= \hat{x}^{j}_{k|k-1} + K^{j}_{k}(y_k - C_j \hat{x}^{j}_{k|k-1}), && (6.119)\\
\Sigma^{j}_{k|k} &= \Sigma^{j}_{k|k-1} - K^{j}_{k} S^{j}_{k} \left(K^{j}_{k}\right)^T. && (6.120)
\end{aligned}$$

- **Mode Probability Update:** The previous mode probabilities $\{\mu^{i}_{k-1}\}_{i=1}^{N}$ are updated to obtain the new mode probabilities $\{\mu^{j}_{k}\}_{j=1}^{N}$ as follows.

$$\mu^{j}_{k} = \frac{1}{\bar{c}} \Delta^{j}_{k} \sum_{i=1}^{N} \pi_{ij} \mu^{i}_{k-1} \qquad (6.121)$$

where

$$\begin{aligned}
\bar{c} &= \sum_{j=1}^{N} \Delta^{j}_{k} \sum_{i=1}^{N} \pi_{ij} \mu^{i}_{k-1}, && (6.122)\\
\Delta^{j}_{k} &= \mathcal{N}(y_k; C_j \hat{x}^{j}_{k|k-1}, S^{j}_{k}). && (6.123)
\end{aligned}$$

- **Output Estimate and Covariance Calculation by Merging:** The output estimate and covariance are calculated by merging the filtered estimates $\{\hat{x}^{j}_{k|k}\}_{j=1}^{N}$ and covariances $\{\Sigma^{j}_{k|k}\}_{j=1}^{N}$. The merging is done using the updated mode probabilities $\{\mu^{j}_{k}\}_{j=1}^{N}$ as follows.

$$\begin{aligned}
\hat{x}_{k|k} &= \sum_{j=1}^{N} \mu^{j}_{k} \hat{x}^{j}_{k|k}, && (6.124)\\
\Sigma_{k|k} &= \sum_{j=1}^{N} \mu^{j}_{k} \left[\Sigma^{j}_{k|k} + (\hat{x}^{j}_{k|k} - \hat{x}_{k|k})(\hat{x}^{j}_{k|k} - \hat{x}_{k|k})^T\right]. && (6.125)
\end{aligned}$$

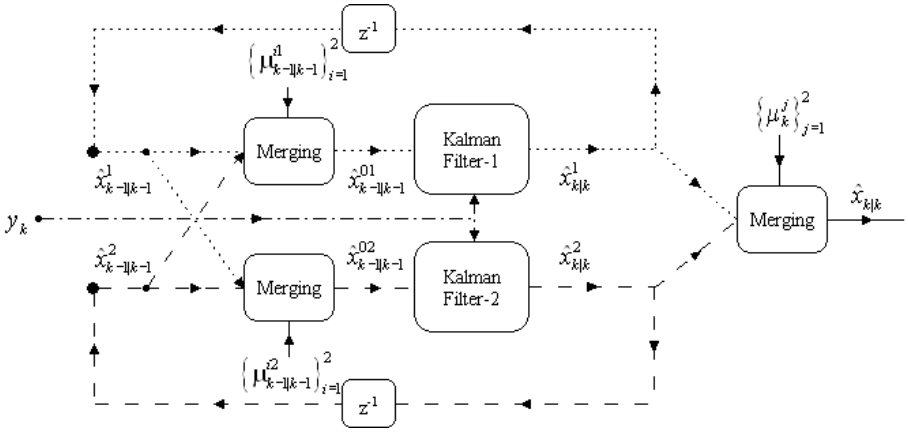

Figure 6.3: Block diagram of a two-model IMM algorithm.

The steps of the IMM algorithm related with the filtered state estimates are summarized for a two-model IMM algorithm in Fig. 6.3.

6.6 Efficient Mixed IMM-GPB2 Algorithm

As observed in the discussion above, the moment-matched single Gaussian approximation is made at many steps of both IMM and GPB2 algorithms. In this section, we are going to concentrate on specifically the ones made in the mixing step of the IMM algorithm because they are the main approximations differentiating an IMM filter from a GPB2 filter. In GPB2 algorithm, for each model j, each input estimate $\hat{x}^i_{k-1|k-1}$ is passed through the Kalman filter matched to model j and the resulting estimates are merged to form the filtered estimate corresponding to that model (i.e., $\hat{x}^j_{k|k}$). In contrast to this, in the IMM filter, the input estimates $\{\hat{x}^i_{k-1|k-1}\}_{i=1}^N$ are mixed (merged) first to form a single initial estimate (for each model), and this initial estimate is then input to the Kalman filter matched to the model j to obtain filtered estimate corresponding to that model (i.e., $\hat{x}^j_{k|k}$). This initial mixing process in the IMM algorithm is the main reason for the increase in the state estimation errors of the IMM filter relative to the GPB2 filter. Comparing these processes with our error analysis, we see that our covariance formula calculated in Sec. 6.4 can be used to calculate the covariance of the error between the filtered

estimates (of each model) that would be obtained using IMM-type merging (first mixing then filtering) and those that would be obtained by GPB2-type merging (first filtering then merging). Therefore, at the beginning of each algorithm cycle, we can decide, by observing the error covariance, whether to make IMM-type or GPB2-type merging for each model. For the models which would result in large errors with the IMM-type merging, GPB2-type merging can be selected to decrease the errors. This idea is the main motivation for our mixed IMM-GPB2 filter.

In the following, we are going to present the steps of our Mixed IMM-GPB2 algorithm.

6.6.1 Steps of the Mixed IMM-GPB2 Algorithm

At each time step, an N-model mixed IMM-GPB2 algorithm keeps N filtered estimates $\{\hat{x}^j_{k|k}\}^N_{j=1}$, covariances $\{\Sigma^j_{k|k}\}^N_{j=1}$ and mode probabilities $\{\mu^j_k\}^N_{j=1}$. At the beginning of each cycle, the algorithm calculates a statistics γ^j_k for each model to decide whether to make IMM-type or GPB-type merging for that model. Given the previously kept filtered estimates $\{\hat{x}^j_{k-1|k-1}\}^N_{j=1}$, covariances $\{\Sigma^j_{k-1|k-1}\}^N_{j=1}$ and the mode probabilities $\{\mu^j_{k-1}\}^N_{j=1}$, the algorithm calculates the updated values of these quantities using the following steps:

- **Calculation of Mixing Probabilities:** The mixing probabilities are required for IMM-type merging (if any) and for the calculation of the statistics γ^j_k. The mixing probabilities are calculated as follows (same as the ones in IMM).

$$\mu^{ij}_{k-1|k-1} = \frac{1}{\bar{c}_j}\pi_{ij}\mu^i_{k-1} \tag{6.126}$$

  where

$$\bar{c}_j = \sum_{i=1}^{N} \pi_{ij}\mu^i_{k-1}. \tag{6.127}$$

- **Calculation of Predicted Estimates and Merged Predicted Estimates:** This step is required only for the calculation of the statistics.

The computational load in this step can be alleviated using the simplifications mentioned in Sec. 6.6.2. The predicted estimates and covariances are calculated as follows.

$$\hat{x}^{ij}_{k|k-1} = A_j \hat{x}^{i}_{k-1|k-1}, \tag{6.128}$$

$$\Sigma^{ij}_{k|k-1} = A_j \Sigma^{i}_{k-1|k-1} A_j^T + B_j Q B_j^T, \tag{6.129}$$

$$S^{ij}_{k} = C_j \Sigma^{ij}_{k|k-1} C_j^T + D_j R D_j^T. \tag{6.130}$$

The merged predicted estimates and covariances for each model are calculated as follows.

$$\hat{x}^{j}_{app} = \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \hat{x}^{ij}_{k|k-1}, \tag{6.131}$$

$$\begin{aligned} \Sigma^{j}_{app} = & \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \Big[\Sigma^{ij}_{k|k-1} \\ & + (\hat{x}^{ij}_{k|k-1} - \hat{x}^{j}_{app})(\hat{x}^{ij}_{k|k-1} - \hat{x}_{app})^T \Big], \end{aligned} \tag{6.132}$$

$$S^{j}_{sub} = C_j \Sigma^{j}_{app} C_j^T + D_j R D_j^T. \tag{6.133}$$

- **Error Covariance Calculation** At this step, the algorithm calculates for each model j, the covariance of the error Σ_j that would be induced in the filtered estimate $\hat{x}^{j}_{k|k}$ if one uses IMM-type merging instead of GPB2-type merging. This covariance is given using our analysis and calculated error covariance formula in Sec. 6.4 as follows.

$$\begin{aligned} \Sigma_j = & \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \Sigma^{ij}_{k|k-1} C_j^T \left(S^{ij}_{k}\right)^{-1} C_j \Sigma^{ij}_{k|k-1} \\ & - \Sigma^{j}_{app} C_j^T \left(S^{j}_{sub}\right)^{-1} C_j \left(\Sigma^{j}_{app}\right)^T \\ & + \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \hat{x}^{ij}_{k|k-1} \left(\hat{x}^{ij}_{k|k-1}\right)^T - \hat{x}^{j}_{app} \left(\hat{x}^{j}_{app}\right)^T. \end{aligned} \tag{6.134}$$

At this stage, the algorithm has to determine, for each model j, whether the matrix Σ_j is "big" enough to switch to GPB-type merging for that model. This can be done using a matrix norm to obtain a statistics γ^j_k and a simple thresholding. Also, the application specific heuristics can do well for this purpose. In fact, in Sec. 6.7, we are going use such a

heuristics for the simulation. From this point on, we are going to assume that the required test has already been completed and a decision on whether to use IMM-type or GPB2-type merging has already been made for every model.

- **For every model j**

  If IMM-type merging is used

  1. **Mixing:** The previous filtered estimates $\{\hat{x}^j_{k-1|k-1}\}_{j=1}^N$ and covariances $\{\Sigma^j_{k-1|k-1}\}_{j=1}^N$ are merged to obtain N-different initial estimates $\{\hat{x}^{0j}_{k-1|k-1}\}_{j=1}^N$ and the covariances $\{\Sigma^{0j}_{k-1|k-1}\}_{j=1}^N$ as follows.

$$\hat{x}^{0j}_{k-1|k-1} = \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \hat{x}^i_{k-1|k-1}, \tag{6.135}$$

$$\Sigma^{0j}_{k-1|k-1} = \sum_{i=1}^{N} \mu^{ij}_{k-1|k-1} \Big[\Sigma^i_{k-1|k-1} + (\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})(\hat{x}^i_{k-1|k-1} - \hat{x}^{0j}_{k-1|k-1})^T\Big]. \tag{6.136}$$

  2. **Mode-Matched Kalman Filtering:** The algorithm takes each initial estimate $\hat{x}^{0j}_{k-1|k-1}$ and covariance $\Sigma^{0j}_{k-1|k-1}$ and executes N Kalman filters each matched to a different model. All N Kalman filters uses its corresponding initial estimate $\hat{x}^{0j}_{k-1|k-1}$ and covariance $\Sigma^{0j}_{k-1|k-1}$ as their initial state and covariance. Since there are N initial conditions, N Kalman filters are executed. At the end of this filtering the new filtered estimates $\hat{x}^j_{k|k}$ and covariances $\hat{x}^j_{k|k}$ are obtained. The filtering equations are given as follows.

$$\hat{x}^j_{k|k-1} = A_j \hat{x}^{0j}_{k-1|k-1}, \tag{6.137}$$

$$\Sigma^j_{k|k-1} = A_j \Sigma^{0j}_{k-1|k-1} A_j^T + B_j Q B_j^T, \tag{6.138}$$

$$S^j_k = C_j \Sigma^j_{k|k-1} C_j^T + D_j R D_j^T, \tag{6.139}$$

$$K^j_k = \Sigma^j_{k|k-1} C_j^T \left(S^j_k\right)^{-1}, \tag{6.140}$$

$$\hat{x}^j_{k|k} = \hat{x}^j_{k|k-1} + K^j_k (y_k - C_j \hat{x}^j_{k|k-1}), \tag{6.141}$$

$$\Sigma^j_{k|k} = \Sigma^j_{k|k-1} - K^j_k S^j_k \left(K^j_k\right)^T. \tag{6.142}$$

3. **Likelihood Calculation:** The likelihood Δ_k^j of the current measurement y_k is calculated as follows

$$\Delta_k^j = \mathcal{N}(y_k; C_j \hat{x}_{k|k-1}^j, S_k^j). \tag{6.143}$$

If GPB2-type merging is used

1. **Kalman Filter Measurement Updates** Since the Kalman filter prediction updates have been already done for statistics calculations, in this step, only measurement updates are required. These updates are done as follows.

$$\begin{aligned} K_k^{ij} &= \Sigma_{k|k-1}^{ij} C_j^T \left(S_k^{ij}\right)^{-1}, & (6.144)\\ \hat{x}_{k|k}^{ij} &= \hat{x}_{k|k-1}^{ij} + K_k^{ij}(y_k - C_j \hat{x}_{k|k-1}^{ij}), & (6.145)\\ \Sigma_{k|k}^{ij} &= \Sigma_{k|k-1}^{ij} - K_k^{ij} S_k^{ij} \left(K_k^{ij}\right)^T. & (6.146) \end{aligned}$$

2. **Calculation of the Merging Probabilities:** At this step, the merging probabilities $\{\mu_{k|k-1}^{ij}\}_{i=1}^N$ are calculated for each j. These probabilities are used to merge the estimates $\{\hat{x}_{k|k}^{ij}\}_{i=1}^N$ and covariances $\{\Sigma_{k|k}^{ij}\}_{i=1}^N$ for each j. The probabilities are calculated as follows.

$$\mu_{k|k-1}^{ij} = \frac{1}{c_j} \Delta_k^{ij} \pi_{ij} \mu_{k-1}^i \tag{6.147}$$

where

$$\begin{aligned} c_j &= \sum_{i=1}^N \Delta_k^{ij} \pi_{ij} \mu_{k-1}^i, & (6.148)\\ \Delta_k^{ij} &= \mathcal{N}(y_k; C_j \hat{x}_{k|k-1}^{ij}, S_k^{ij}). & (6.149) \end{aligned}$$

3. **Merging:** N filtered estimates $\{\hat{x}_{k|k}^{ij}\}_{i=1}^N$ and covariances $\{\Sigma_{k|k}^{ij}\}_{i=1}^N$ are merged for each j and the new filtered estimates $\{\hat{x}_{k|k}^j\}_{j=1}^N$ and covariances $\{\Sigma_{k|k}^j\}_{j=1}^N$ are obtained. The merging is done as follows.

$$\begin{aligned} \hat{x}_{k|k}^j &= \sum_{i=1}^N \mu_{k-1|k}^{ij} \hat{x}_{k|k}^{ij}, & (6.150)\\ \Sigma_{k|k}^j &= \sum_{i=1}^N \mu_{k-1|k}^{ij} \left[\Sigma_{k|k}^{ij} + (\hat{x}_{k|k}^{ij} - \hat{x}_{k|k}^j)(\hat{x}_{k|k}^{ij} - \hat{x}_{k|k}^j)^T\right]. \end{aligned}$$

4. **Likelihood Calculation:** The likelihood Δ_k^j of the current measurement y_k is calculated as follows.

$$\Delta_k^j = \sum_{i=1}^{N} \Delta_k^{ij} \mu_{k-1|k-1}^{ij}. \tag{6.151}$$

- **Mode Probability Update:** The previous mode probabilities $\{\mu_{k-1}^i\}_{i=1}^N$ are updated to obtain the new mode probabilities $\{\mu_k^j\}_{j=1}^N$ as follows.

$$\mu_k^j = \frac{1}{\bar{c}} \Delta_k^j \sum_{i=1}^{N} \pi_{ij} \mu_{k-1}^i \tag{6.152}$$

where

$$\bar{c} = \sum_{j=1}^{N} \Delta_k^j \sum_{i=1}^{N} \pi_{ij} \mu_{k-1}^i, \tag{6.153}$$

$$\Delta_k^j = \mathcal{N}(y_k; C_j \hat{x}_{k|k-1}^j, S_k^j). \tag{6.154}$$

- **Output Estimate and Covariance Calculation by Merging:** The output estimate and covariance are calculated by merging the filtered estimates $\{\hat{x}_{k|k}^j\}_{j=1}^N$ and covariances $\{\Sigma_{k|k}^j\}_{j=1}^N$. The merging is done using the updated mode probabilities $\{\mu_k^j\}_{j=1}^N$ as follows.

$$\hat{x}_{k|k} = \sum_{j=1}^{N} \mu_k^j \hat{x}_{k|k}^j, \tag{6.155}$$

$$\Sigma_{k|k} = \sum_{j=1}^{N} \mu_k^j \left[\Sigma_{k|k}^j + (\hat{x}_{k|k}^j - \hat{x}_{k|k})(\hat{x}_{k|k}^j - \hat{x}_{k|k})^T \right]. \tag{6.156}$$

6.6.2 Possible Simplifications

Note that the calculations required for the statistics calculations are quite complicated. The following are some simplification suggestions that can be applied without reducing the performance of the algorithm substantially.

1. The predicted quantities $\hat{x}_{k|k-1}^{ij}$, $\Sigma_{k|k-1}^{ij}$, S_k^{ij} can be replaced with their equivalents in the previous sampling period i.e., with $\hat{x}_{k-1|k-2}^{ij}$, $\Sigma_{k-1|k-2}^{ij}$, S_{k-1}^{ij} respectively. In this way, some or all of them might have been calculated already during the Kalman filtering. Also, in that case, the

multiplications in the form $\Sigma_{k-1|k}^{ij} C_j^T \left(S_{k-1}^{ij}\right)^{-1}$ might have been calculated during Kalman gain calculations. Obviously, if this simplification is made, then, while using the GPB2-type merging, one has to calculate the required predictions as well.

2. The calculation of the statistics can be done only at every N sampling periods and between the periods, the results of the last statistics calculation can be used.

3. Note that the statistics calculations need not require all the elements of the matrices Σ_j. For most of the case, for example, one might try to use only the diagonal elements of the matrices Σ_j for statistics calculation. In that case, the computations required for the calculation of the matrices Σ_j can be reduced significantly.

6.7 Simulation Results

In this section, the performance of the mixed IMM-GPB2 algorithm will be observed and compared to those of the IMM and GPB2 algorithms. For this purpose, we consider a simplified example of a moving target whose acceleration evolves according to a finite-state Markov chain. This example,which was also used in the simulation of Chapters 2 and 3, is repeated here for the sake of completeness.[6] The target dynamics in one-dimension is given as

$$\underbrace{\begin{bmatrix} p_k \\ v_k \end{bmatrix}}_{x_k} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{k-1} \\ v_{k-1} \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} [a_k + w_k] \tag{6.157}$$

where p_k, v_k and a_k denote the target position, velocity and acceleration respectively. The initial state x_0 is normally distributed with mean $\bar{x}_0$ and covariance Σ_0 which are given as

$$\bar{x}_0 = \begin{bmatrix} 80000 \\ 400 \end{bmatrix} \quad \text{and} \quad \Sigma_0 = \begin{bmatrix} 10000 & 1000 \\ 1000 & 10000 \end{bmatrix}. \tag{6.158}$$

[6] Note that this example is a slightly modified version of one of the examples given in [1].

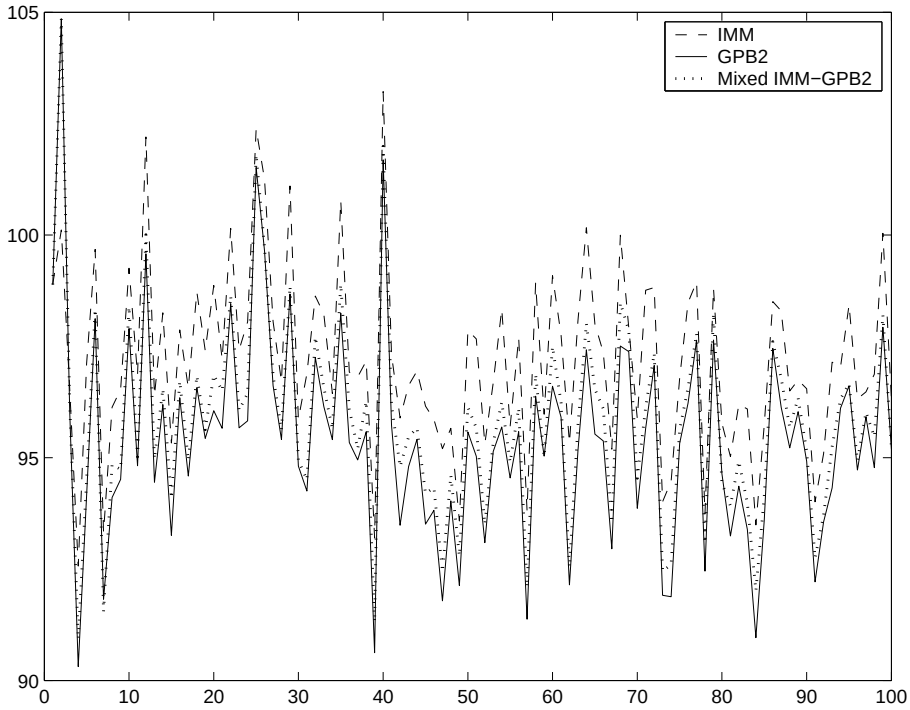

Figure 6.4: RMS position errors of the IMM, GPB2 and mixed IMM-GPB2 algorithms

The acceleration process a_k is a finite-state Markov chain with states in the set $\{0, 20, -20\}$. The initial probability distribution for the states is given as $\pi_0 = [0.8, 0.1, 0.1]$. The transition probability matrix for the finite-state Markov chain is

$$\Pi = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \tag{6.159}$$

which corresponds to a highly maneuvering target. The white process noise $w_k \sim \mathcal{N}(w_k; 0, 2^2)$ represents small acceleration changes. It is assumed that only the positions are measured, i.e.,

$$y_k = p_k + \nu_k \tag{6.160}$$

where the terms $\nu_k \sim \mathcal{N}(w_k; 0, 100^2)$ stands for the normally distributed white measurement noise. The sampling period T is taken to be 10secs.

Using the measurements coming from this system, we execute IMM, GPB2 and our mixed IMM-GPB2 algorithm. The mixed IMM-GPB2 algorithm is

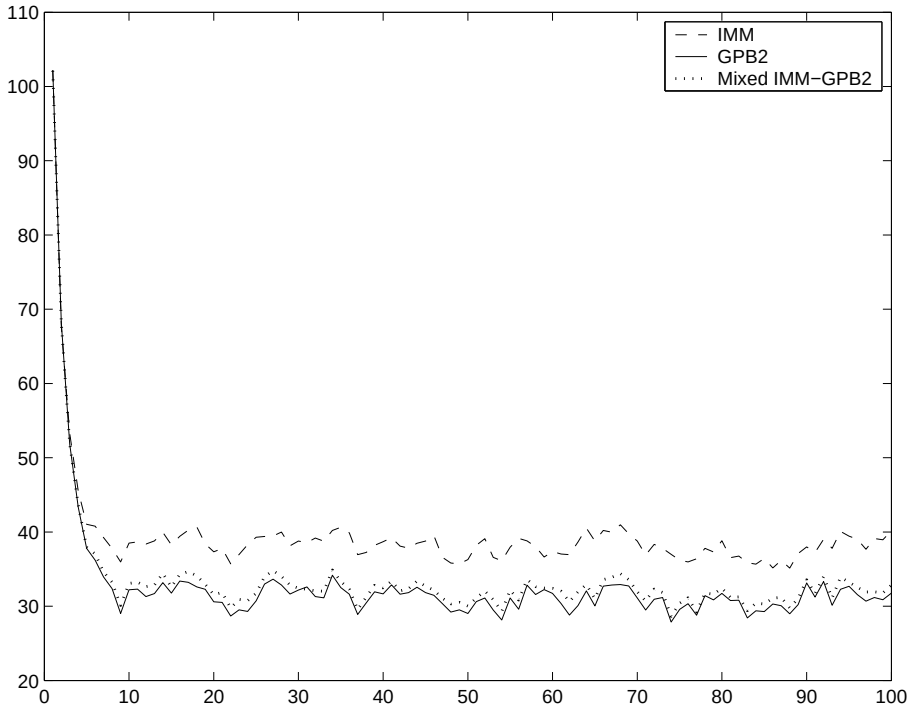

Figure 6.5: RMS velocity errors of the IMM, GPB2 and mixed IMM-GPB2 algorithms

implemented using the velocity error standard deviation as the merging decision statistics. This means that we only use the elements corresponding to second row and second column of the 2×2 matrices Σ_j for the statistics calculation. Realizing also that the matrices $C_j = [1\ 0]$ for $j = 1, 2, 3$, this reduces the computations required for the statistics calculations to an almost negligible level. We have taken the statistics threshold γ_{thresh} as 3 (m/sec) which means that if the statistics γ_k^j for model j is smaller than 3, the IMM-type merging will be used for that model. Otherwise, GPB-type merging is applied. RMS position errors resulting from 1000 Monte-Carlo runs are presented in Fig. 6.4. The corresponding velocity errors are shown in Fig. 6.5. As observed from the figures, the GPB2 shows the best performance in both of the cases as expected. The errors of the Mixed IMM-GPB2 algorithm are below those of the IMM filter and are very near to those of the GPB2 algorithm. The IMM and GPB2 algorithms uses 3 and 9 Kalman filters for each measurement respectively. The mixed IMM-GPB2 algorithm proves to use an

Table 6.1: The average number of Kalman filters and RMS errors obtained with different γ_{thresh} values.

γ_{thresh}	Average number of Kalman filters used per measurement	Average RMS position error (m) per measurement	Average RMS velocity error (m/sec) per measurement
0.5	7.6	95.39	32.69
1	6.9	95.40	32.71
3	5.2	95.83	33.56
4	4.7	96.15	35.48
5	4.2	96.56	36.90
7	3	97.20	39.33
GPB2	9	95.38	32.69
IMM	3	97.20	39.33

average of 5.2 Kalman filters per each measurement. This shows that, to reach the performance of the GPB2 algorithm, one can use much less Kalman filters compared to the number of Kalman filters required for the GPB2 algorithm. The average position RMS errors for the IMM filter and the GPB2 filter per measurement are 97.20 and 95.38 respectively. The corresponding RMS error value for the mixed IMM-GPB2 filter is 95.83. The average velocity RMS errors for the IMM and GPB2 algorithms per measurement are given as 39.33 and 32.69 respectively. The corresponding RMS error for the mixed IMM-GPB2 algorithm is 33.56 which is a significant reduction in the error relative to the IMM filter towards the performance of GPB2 algorithm. The results obtained using different values of the statistics threshold γ_{thresh} are shown in Table 6.1. The error characteristics of the mixed IMM-GPB2 algorithm seem to reach a virtually indistinguishable level from that of GPB2 with only 7.6 Kalman filters on the average per measurement.

6.8 Conclusions

In this chapter, the error caused by the moment-matched single Gaussian approximation of the Gaussian mixtures applied at the input of the optimal Bayesian filter is examined analytically. The prediction update of the filter is shown to be invariant (in the MMSE sense) under the approximation. The

measurement update, on the other hand, is shown to cause some difference (i.e., error) between the optimal and the approximated (sub-optimal) estimates. The resulting error proves to be of zero mean. An analytical formula to calculate its covariance approximately is given.

The calculated covariance is then used to obtain a mixed IMM-GPB2 algorithm which combines the IMM and GPB2 algorithms in which the single Gaussian approximations are abundant. The resulting algorithm turns out to reach the performance of the GPB2 filter with less number of Kalman filters (per measurement) than the GPB2 algorithm. Although only the case of IMM and GPB2 is considered in this study, our belief is that the same methodology can be used for combining different order GPB methods as well.

CHAPTER 7

DISCUSSION AND FUTURE WORK

In this thesis, an example framework has been presented to make improvements over the state of the art of multiple model state estimation. In the first category of improvements, risk-sensitive estimation problem has been investigated for JMLSs. While the case of instantaneous cost function is quite simple and requires the modification of only the output estimate calculation step of IMM filter, the case of cumulative cost function turns out to be quite complicated. The analysis of the cumulative case results in a unifying framework which combines

- Risk-sensitive multiple model filter
- IMM filter
- Risk-sensitive filter for linear Gauss-Markov systems
- Kalman filter

and, in author's opinion, is the most impressive contribution of this thesis. Moreover, the investigation of the IMM filter and its approximations using the reference probability method has the potential of contributing to future research involving JMLSs.

In the second category of improvements, two online transition probability estimators and one mixed multiple model state estimation algorithm are presented. Although the transition probability estimators extend the previously applied algorithms to HMMs to the case of JMLS, the derivations involved are quite original due to the fact that Markov chain is buried one more layer deeper

under the measurements in JMLSs than in HMMs. Especially the derivation presented for the maximum likelihood estimator shows the potential of the reference probability method to handle complex systems.

In author's opinion, the work on mixed multiple model estimation is the most straightforward part of this thesis. This is caused, in part, by the non-complexity of the analysis involved in the examination of the single Gaussian approximation. It is, however, still important to note that the idea of mixing two different multiple model state estimation procedures efficiently is an interesting and new problem and it might actually be the only way for obtaining high performance MMSE algorithms (which approach optimality) with low computations.

In addition to the results presented, this thesis also shows the directions for a future study in the search of better (and better) multiple model estimation schemes. In order to obtain improvements of the first type, the main methodology applied here suggests the application of other estimation schemes or criteria to the case of JMLSs. In this regard, it is interesting to see that, although the JMLSs are actually nonlinear systems, the effect of the nonlinear control community on the field of multiple model estimation is hardly felt. The reason for this might be the doubly stochastic nature of the problem. Consequently, using these facts as a motivation, the application of estimation and control literature on deterministic nonlinear systems to the (stochastic or deterministic) multiple model estimation problem seems to have a future. In fact, it was one of the author's intentions at some part of the thesis research to apply the sliding mode observer theory to the case of JMLSs.

The second category of improvements suggested in this thesis obviously point to the fact that the highly researched area of HMMs already gives a starting point for the system identification problems associated with JMLSs. There are many methods in the HMM literature, some of which has already been referenced here, which can be generalized to JMLSs with no more efforts than made in this thesis. Moreover, the concept of a mixed multiple model estimator issues a new challenge in the area of multiple model estimation

for efficiently combining the different aspects of different estimators to obtain brand-new algorithms which outperform the old ones in the performance vs. computation curve. In this regard, the implications of this idea in general estimation theory, under the title of "estimator fusion" might also deserve some attention.

APPENDIX A

REFERENCE PROBABILITY METHOD

This appendix summarizes the reference probability method which is the state of the art of taking expected values easily. We only examine the case of discrete-time stochastic processes to evade the curse of continuous-time specialties which involve high-level measure theory. The appendix is organized in three sections. In Sec. A.1, the necessary theorems and lemmas for changing measures which is the crucial part of the reference probability method is given. Then, the derivations of the famous Kalman filter and the risk-sensitive filter for linear Gauss-Markov systems are made in Secs. A.2 and A.3 respectively using the reference probability method. This appendix has been adapted mainly from [39]. Some of the theorems, lemmas and examples have been taken directly. The proofs, derivations and details have been elaborated by the comments of the author.

A.1 Background

This section gives the necessary theorems and lemmas with illustrating examples which will be used in the following sections.

Remark A.1 *Given any non-negative random variable x on a probability space $(\Omega, \mathcal{F}, P)$ with finite $E(x) = \int_\Omega x dP = 1$, one can define another probability measure $\overline{P}$ of $\mathcal{F}$ by setting for every $F \in \mathcal{F}$:*

$$\overline{P}(F) = \int_F x dP. \tag{A.1}$$

Clearly, if $P(F) = 0$ *then* $\overline{P}(F) = 0$ *for any* $F \in \mathcal{F}$ *and we say that* $\overline{P}$ *is absolutely continuous with respect to* P *and denote it by* $\overline{P} \ll P$.

In Remark A.1, we state that, if a random variable with appropriate conditions in a probability measure P is given, then using it, we can define a new probability measure $\overline{P}$ which is absolutely continuous with respect to the original probability measure P. The Radon-Nikodym theorem claims the converse i.e., given a probability measure $\overline{P}$ which is absolutely continuous with respect to the original probability measure P, it claims that the random variable x satisfying Eqn. A.1 exists.

Theorem A.1 (Radon-Nikodym) *Let* $(\Omega, \mathcal{F})$ *be a measurable space,* μ *a* σ*-finite measure and* $\overline{\mu}$ *a signed measure (i.e.,* $\overline{\mu} = \overline{\mu}_1 - \overline{\mu}_2$*, where at least one of the measures* $\overline{\mu}_1$ *and* $\overline{\mu}_2$ *is finite) such that for each* $F \in \mathcal{F}$*,* $\mu(F) = 0$ *implies* $\overline{\mu}(F) = 0$*, i.e.,* $\overline{\mu} \ll \mu$*. Then there exists an* $\mathcal{F}$*-measurable function* Λ *with values in the extended real line* $[-\infty, +\infty]$ *such that*

$$\overline{\mu}(C) = \int_C \Lambda(\omega) d\mu(\omega) \tag{A.2}$$

for all $C \in \mathcal{F}$*. The function* Λ *is unique up to sets of* μ*-measure zero; i.e., if* $h(.)$ *is another* $\mathcal{F}$*-measurable function such that* $\overline{\mu}(C) = \int_C h(\omega) d\mu(\omega)$ *for all* $C \in \mathcal{F}$*, then* $\mu\{\omega : \Lambda(\omega) \neq h(\omega)\} = 0$*. If* $\overline{\mu}$ *is a positive measure, then* Λ *has its values in* $[0, +\infty]$*. We write*

$$\frac{d\overline{\mu}}{d\mu}\Big|_{\mathcal{F}} = \Lambda. \tag{A.3}$$

In the case of probability measures, the Radon-Nikodym theorem reads as follows. If P and $\overline{P}$ are two probability measures on $(\Omega, \mathcal{F})$ such that for each $B \in \mathcal{F}$, $P(B) = 0$ implies $\overline{P}(B) = 0$ ($\overline{P} \ll P$), then there exists a nonnegative random variable Λ, such that

$$\overline{P}(C) = \int_C \Lambda dP \quad \text{for all} \quad C \in \mathcal{F}. \tag{A.4}$$

We write

$$\frac{d\overline{P}}{dP}\Big|_{\mathcal{F}} = \Lambda. \tag{A.5}$$

Taking $C = \Omega$ we see that

$$\overline{P}(\Omega) = 1 = \int_{\Omega} \Lambda dP = E[\Lambda] \tag{A.6}$$

so that $\overline{P}$ is a probability measure if and only if Λ is nonnegative and $E[\Lambda] = 1$. If those conditions are satisfied, Λ is called the density of $\overline{P}$ with respect to P, or the *Radon-Nikodym derivative* of $\overline{P}$ with respect to P.

Radon-Nikodym derivatives might look, at a first glance, too abstract to be useful to an engineer. However, the probability density functions used extensively in engineering literature are indeed Radon-Nikodym derivatives. This can be explained as follows: Suppose that x is a vector random variable of dimension n and suppose that the probability density function $f_x(x)$ of x with respect to probability measure P exists. Then, we know that, for any Lebesgue measurable set $A \in \mathbb{R}^n$,

$$P(x \in A) = P_x(A) = \int_{\{x \in A\}} dP = \int_A f_x(x) dx \tag{A.7}$$

where $P_x(A)$ is the probability measure induced by the random variable x and $\{x \in A\}$ is the set defined as

$$\{x \in A\} \triangleq \{\omega \in \Omega : x(\omega) \in A\}. \tag{A.8}$$

Due to Eqn. A.7, the density function $f_x(x)$ is the Radon-Nikodym derivative of the probability measure P_x with respect to the Lebesgue measure.

Example A.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space on which are defined the random variables $Y_1, Y_2, \ldots, Y_n$. Let $\mathcal{F}_n = \sigma\{Y_1, Y_2, \ldots, Y_n\}$. Let $\overline{P}$ be another probability measure on $\mathcal{F}$. Suppose that under P and $\overline{P}$ the random variables $Y_1, Y_2, \ldots, Y_N$ have joint densities $f_n(.)$ and $\overline{f}_n(.)$ respectively, with respect to n-dimensional Lebesgue measure. Then the Radon-Nikodym derivative*

$$\frac{d\overline{P}}{dP}\Big|_{\mathcal{F}_n} = \frac{\overline{f}_n(Y_1, Y_2, \ldots, Y_n)}{f_n(Y_1, Y_2, \ldots, Y_n)} \tag{A.9}$$

is the likelihood ratio of the two probability measures in the presence of a sample of observations $\{Y_1, Y_2, \ldots, Y_n\}$.

Let $x \in L_1$ (i.e., $E|x| < \infty$) be a non-negative random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. The probability space $(\Omega, \mathcal{G}, P)$ is a coarsening of the original one and x is, in general, not measurable with respect to $\mathcal{G}$. Now, we seek a $\mathcal{G}$-measurable random variable, which we denote temporarily by $x_{\mathcal{G}}$ that assumes, on average, the same values as x. That is, we seek an integrable random variable $x_{\mathcal{G}}$ such that $x_{\mathcal{G}}$ is $\mathcal{G}$-measurable and

$$\int_A x_{\mathcal{G}} dP = \int_A x dP \quad \forall A \in \mathcal{G}. \tag{A.10}$$

Defining the set function $Q(A) \triangleq \int_A x dP$, we see that Q is a measure and it is absolutely continuous with respect to P. Then, the Radon-Nikodym theorem guarantees the existence of a $\mathcal{G}$-measurable random variable $E[x|\mathcal{G}]$ which is uniquely determined up to a null set such that

$$\int_A x dP = \int_A E[x|\mathcal{G}] dP \quad \forall A \in \mathcal{G}. \tag{A.11}$$

The random variable $E[x|\mathcal{G}]$ is called as the conditional expectation of the random variable x given the σ-field $\mathcal{G}$.

Theorem A.2 (Conditional Bayes' Theorem) *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{G} \subset \mathcal{F}$is a sub-$\sigma$-field. Suppose $\overline{P}$ is another probability measure absolutely continuous with respect to P ($\overline{P} \ll P$) and with a Radon-Nikodym derivative*

$$\frac{d\overline{P}}{dP} = \Lambda. \tag{A.12}$$

Then if ϕ is any integrable $\mathcal{F}$-measurable random variable,

$$\overline{E}[\phi|\mathcal{G}] = \begin{cases} \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} & \text{if} \quad E[\Lambda|\mathcal{G}] > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.13}$$

where $\overline{E}$ and E denotes the expectations with respect to probability measures $\overline{P}$ and P respectively.

Another useful version of Theorem A.2 is the following theorem.

Theorem A.3 *Suppose $(\Omega, \mathcal{F}, P)$ is a probability measure with a filtration $\{\mathcal{F}_t, t \geq 0\}$. Suppose $\overline{P}$ is another probability measure absolutely continuous*

with respect to P ($\overline{P} \ll P$) *on* $\mathcal{F}$ *and with a Radon-Nikodym derivative*

$$\frac{d\overline{P}}{dP} = \Lambda. \tag{A.14}$$

If we define the martingale

$$\Lambda_t \triangleq E[\Lambda|\mathcal{F}_t] \tag{A.15}$$

then, if ϕ_t *is any* $\mathcal{F}_t$*-adapted process,*

$$\overline{E}[\phi_t|\mathcal{F}_s] = \begin{cases} \frac{E[\Lambda_t\phi_t|\mathcal{F}_s]}{E[\Lambda_t|\mathcal{F}_s]} & \text{if} \quad E[\Lambda_t|\mathcal{F}_s] > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{A.16}$$

Example A.2 *Let* $\{x_n\}$ *be a sequence of random variables on some probability space* $(\Omega, \mathcal{F}, P)$*. Consider the filtration* $\{\mathcal{F}_n = \sigma\{x_1, \ldots, x_n\}\}$*. Assume that the (one step predicted) probability density functions* $\phi_{n|n-1}(x_n) = p(x_n|\mathcal{F}_{n-1})$ *exist and are positive. Suppose that we wish to define a new probability measure* $\overline{P}$ *on* $(\Omega, \bigvee \mathcal{F}_n)$ *such that* x_n *are independent identically distributed (i.i.d.) with positive probability density function* α*. Let* $\lambda_0 = 1$ *and for* $k \geq 1$

$$\lambda_k = \frac{\alpha(x_k)}{\phi_{k|k-1}(x_k)}, \tag{A.17}$$

$$\Lambda_n = \prod_{k=0}^{n} \lambda_k, \tag{A.18}$$

and we define the probability measure $\overline{P}$ *to satisfy*

$$\frac{d\overline{P}}{dP}(\omega)\Big|_{\mathcal{F}_n} = \Lambda_n(\omega). \tag{A.19}$$

Lemma A.1 *The sequence of random variables* $\{\Lambda_n\}$*,* $n \geq 0$ *is an* $\{\mathcal{F}_n, P\}$*-martingale with* P*-mean 1. Moreover, under* $\overline{P}$*,* $\{x_n\}$ *is a sequence of i.i.d. random variables with probability density function* $\alpha(.)$*.*

Proof We have to show that

$$E[\Lambda_n|\mathcal{F}_{n-1}] = \Lambda_{n-1}. \tag{A.20}$$

However, $\Lambda_n = \Lambda_{n-1}\lambda_n$ and since Λ_{n-1} is $\mathcal{F}_{n-1}$-measurable, we must show that $E[\lambda_n|\mathcal{F}_{n-1}] = 1$. In view of the definition of λ_n

$$E[\lambda_n|\mathcal{F}_{n-1}] = E\left[\frac{\alpha(x_n)}{\phi_{n|n-1}(x_n)}\middle|\mathcal{F}_{n-1}\right] = \int_{\mathbb{R}} \frac{\alpha(x_n)}{\phi_{n|n-1}(x_n)}\phi_{n|n-1}(x_n)dx_n = 1.$$

Therefore; the sequence $\{\Lambda_n\}$ is a $\{\mathcal{F}_n, P\}$-martingale. Moreover,

$$E[\lambda_1] = E[\Lambda_1] = E\left[E[\Lambda_2|\mathcal{F}_1]\right] = E[\Lambda_2] = \cdots = E[\Lambda_n] = \cdots = 1. \tag{A.21}$$

Let f be any integrable real-valued "test" function (a measurable function with compact support). Using Theorem A.3, we get

$$\overline{E}[f(x_n)|\mathcal{F}_{n-1}] = \frac{E[f(x_n)\Lambda_n|\mathcal{F}_{n-1}]}{E[\Lambda_n|\mathcal{F}_{n-1}]} = \frac{\Lambda_{n-1}E[f(x_n)\lambda_n|\mathcal{F}_{n-1}]}{\Lambda_{n-1}E[\lambda_n|\mathcal{F}_{n-1}]}. \tag{A.22}$$

Since $E[\lambda_n|\mathcal{F}_{n-1}] = 1$,

$$\begin{aligned}
\overline{E}[f(x_n)|\mathcal{F}_{n-1}] &= E[f(x_n)\lambda_n|\mathcal{F}_{n-1}] && \text{(A.23)}\\
&= \int_{\mathbb{R}} f(x_n)\frac{\alpha(x_n)}{\phi_{n|n-1}(x_n)}\phi_{n|n-1}(x_n)dx_n && \text{(A.24)}\\
&= \int_{\mathbb{R}} f(x_n)\alpha(x_n)dx_n. && \text{(A.25)}
\end{aligned}$$

Since Eqn. A.25 is true for all test functions $f(.)$, $\{x_n\}$ is a sequence of i.i.d. random variables with probability density functions $\alpha(.)$. □

It makes sense here to mention that, under $\overline{P}$, the one step prediction density $\overline{\phi}_k(x_k) \triangleq \overline{p}(x_k|\mathcal{F}_{k-1})$of x_k is equal to $\alpha(x_k)$ because $\{x_k\}$ is an i.i.d. sequence. Therefore, while changing measures we equate the Radon-Nikodym derivative $\frac{d\overline{P}}{dP}(\omega)|_{\mathcal{F}_n}$ to $\Lambda_n = \prod_{k=0}^{n}\lambda_k$ where

$$\lambda_k = \frac{\overline{\phi}_{k|k-1}(x_k)}{\phi_{k|k-1}(x_k)}. \tag{A.26}$$

In other words, λ_k is equal to the ratio of the one step predicted densities of x_k under the two probability measures.

Example A.3 (Change of measure for linear systems: Scalar case) *In reference probability methods, initially, all processes are defined on an "ideal" probability space $(\Omega, \mathcal{F}, \overline{P})$; then a new probability measure P is defined so that a model which has the required characteristics will hold.*

Suppose we would like to define a probability measure P under which we have a system with states $x_k \in \mathbb{R}$ which satisfies the linear dynamics

$$x_{k+1} = ax_k + bw_{k+1}. \tag{A.27}$$

where $\{w_k\}$, $k \in \mathbb{N}$ *is a sequence of independent normal random variables having zero-mean and unity variance. We assume that* $x_0 \sim \mathcal{N}(0, b^2)$. *Let* $\{\mathcal{F}_k = \sigma\{x_0, x_1, \ldots, x_k\}\}$ *for* $k \in \mathbb{N}$ *be a complete filtration; that is,* $\mathcal{F}_0$ *contains all the P-null events.*

We initially begin with the ideal probability measure $\overline{P}$, *under which* $\{x_k\}$, $k \in \mathbb{N}$ *is an i.i.d. sequence with density function* $\phi = \mathcal{N}(0, 1)$. *For each* $l = 0, 1, \ldots$, *we define*

$$\overline{\lambda}_l = \frac{\phi(b^{-1}(x_l - ax_{l-1}))}{b\phi(x_l)}, \tag{A.28}$$

$$\overline{\Lambda}_k = \prod_{l=0}^{k} \overline{\lambda}_l. \tag{A.29}$$

Lemma A.2 *The process* $\{\overline{\Lambda}_k\}$, $k \in \mathbb{N}$ *is a* $\{\mathcal{F}_k, \overline{P}\}$*-martingale.*

Proof Since $\overline{\Lambda}_{k-1}$ is $\mathcal{F}_{k-1}$-measurable,

$$\overline{E}[\overline{\Lambda}_k|\mathcal{F}_{k-1}] = \overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}]. \tag{A.30}$$

So that it is enough to show that $\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = 1$.

$$\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = \overline{E}\left[\frac{\phi(b^{-1}(x_k - ax_{k-1}))}{b\phi(x_k)}\middle|\mathcal{F}_{k-1}\right] \tag{A.31}$$

$$= \int \frac{\phi(b^{-1}(x_k - ax_{k-1}))}{b\phi(x_k)}\phi(x_k)dx_k \tag{A.32}$$

$$= \int b^{-1}\phi(b^{-1}(x_k - ax_{k-1}))dx_k. \tag{A.33}$$

Making the change of variable $u = b^{-1}(x_k - ax_{k-1})$ ($du = b^{-1}dx_k$), we obtain

$$\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = \int \phi(u)du = 1. \tag{A.34}$$

□

We then define P on $\{\Omega, \mathcal{F}\}$ by setting the restriction of the Radon-Nikodym derivative $\frac{dP}{d\overline{P}}$ to $\mathcal{F}_k$ equal to $\overline{\Lambda}_k$. Under the newly defined probability measure P, we have the following lemma.

Lemma A.3 *On* $\{\Omega, \mathcal{F}\}$ *and under* P, $\{w_k\}$, $k \in \mathbb{N}$ *is a sequence of i.i.d. Gaussian random variables with zero mean and unity variance where*

$$w_k \triangleq b^{-1}(x_k - ax_{k-1}). \tag{A.35}$$

Proof Suppose $f : \mathbb{R} \to \mathbb{R}$ is a "test" function (i.e., a measurable function with compact support). Then with E and $\overline{E}$ denoting the expectations under P and $\overline{P}$ respectively, using Theorem A.3,

$$E[f(w_k)|\mathcal{F}_{k-1}] = \frac{\overline{E}[\overline{\Lambda}_k f(w_k)|\mathcal{F}_{k-1}]}{\overline{E}[\overline{\Lambda}_k|\mathcal{F}_{k-1}]} \tag{A.36}$$

$$= \frac{\overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k f(w_k)|\mathcal{F}_{k-1}]}{\overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}]}. \tag{A.37}$$

Since $\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = 1$,

$$\begin{aligned} E[f(w_k)|\mathcal{F}_{k-1}] &= \overline{E}[\overline{\lambda}_k f(w_k)|\mathcal{F}_{k-1}] \qquad (A.38)\\ &= \overline{E}\left[\frac{\phi(b^{-1}(x_k - ax_{k-1}))}{b\phi(x_k)} f(b^{-1}(x_k - ax_{k-1}))|\mathcal{F}_{k-1}\right] \\ &= \int \frac{\phi(b^{-1}(x_k - ax_{k-1}))}{b\phi(x_k)} f(b^{-1}(x_k - ax_{k-1}))\phi(x_k)dx_k \\ &= \int b^{-1}\phi(b^{-1}(x_k - ax_{k-1}))f(b^{-1}(x_k - ax_{k-1}))dx_k. \end{aligned}$$

Using the change of variables $u = b^{-1}(x_k - ax_{k-1})$, we have

$$E[f(w_k)|\mathcal{F}_{k-1}] = \int f(u)\phi(u)du. \tag{A.39}$$

Since Eqn. A.39 is true for all test functions $f(.)$, the lemma is proved. □

What this lemma states is that, under the newly defined probability measure P, the state process $\{x_k\}$ satisfies the dynamics given in Eqn. A.27. Therefore, the change of measure obtained using the Radon-Nikodym derivatives $\overline{\Lambda}_k$ passes us from an ideal probability measure $\overline{P}$ to the probability measure P under which we require our results.

Example A.4 (Change of measure for linear systems: Vector case) *Suppose we would like to define a probability measure P under which we have a system with states $x_k \in \mathbb{R}^n$ which satisfies the linear dynamics*

$$x_{k+1} = Ax_k + Bw_{k+1} \tag{A.40}$$

where A and B are $n \times n$ matrices and B is assumed to be invertible. We assume that $x_0 \sim \mathcal{N}(0, BB^T)$. Let $\{\mathcal{F}_k = \sigma\{x_0, x_1, \ldots, x_k\}\}$ for $k \in \mathbb{N}$ be a complete filtration; that is, $\mathcal{F}_0$ contains all the P-null events.

We initially begin with the ideal probability measure $\overline{P}$, under which $\{x_k\}$, $k \in \mathbb{N}$ is an i.i.d. sequence with density function $\phi(x) = \mathcal{N}(x; 0, I_n)$. For each $l = 0, 1, \ldots$, we define

$$\overline{\lambda}_l = \frac{\phi(B^{-1}(x_l - Ax_{l-1}))}{|B|\phi(x_l)}, \tag{A.41}$$

$$\overline{\Lambda}_k = \prod_{l=0}^{k} \overline{\lambda}_l \tag{A.42}$$

where $|B|$ denotes the determinant of the matrix B.

Lemma A.4 *The process $\{\overline{\Lambda}_k\}$, $k \in \mathbb{N}$ is a $\{\mathcal{F}_k, \overline{P}\}$-martingale.*

Proof Since $\overline{\Lambda}_{k-1}$ is $\mathcal{F}_{k-1}$-measurable,

$$\overline{E}[\overline{\Lambda}_k|\mathcal{F}_{k-1}] = \overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}]. \tag{A.43}$$

So that it is enough to show that $\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = 1$.

$$\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = \overline{E}\left[\frac{\phi(B^{-1}(x_k - Ax_{k-1}))}{|B|\phi(x_k)}\middle|\mathcal{F}_{k-1}\right] \tag{A.44}$$

$$= \int \frac{\phi(B^{-1}(x_k - Ax_{k-1}))}{|B|\phi(x_k)}\phi(x_k)dx_k \tag{A.45}$$

$$= \int |B|^{-1}\phi(B^{-1}(x_k - Ax_{k-1}))dx_k. \tag{A.46}$$

Making the change of variable $u = B^{-1}(x_k - Ax_{k-1})$ ($du = |B|^{-1}dx_k$), we get

$$\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = \int \phi(u)du = 1. \tag{A.47}$$

□

We then define P on $\{\Omega, \mathcal{F}\}$ by setting the restriction of the Radon-Nikodym derivative $\frac{dP}{d\overline{P}}$ to $\mathcal{F}_k$ equal to $\overline{\Lambda}_k$. Under the newly defined probability measure P, we have the following lemma.

Lemma A.5 *On $\{\Omega, \mathcal{F}\}$ and under P, $\{w_k \in \mathbb{R}^n\}$, $k \in \mathbb{N}$ is a sequence of i.i.d. Gaussian random variables with zero mean and covariance I_n where*

$$w_k \triangleq B^{-1}(x_k - Ax_{k-1}). \tag{A.48}$$

Proof *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a "test" function (i.e., a measurable function with compact support). Then with E and $\overline{E}$ denoting the expectations under P*

and $\overline{P}$ *respectively, using theorem A.3,*

$$E[f(w_k)|\mathcal{F}_{k-1}] = \frac{\overline{E}[\overline{\Lambda}_k f(w_k)|\mathcal{F}_{k-1}]}{\overline{E}[\overline{\Lambda}_k|\mathcal{F}_{k-1}]} \tag{A.49}$$

$$= \frac{\overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k f(w_k)|\mathcal{F}_{k-1}]}{\overline{\Lambda}_{k-1}\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}]}. \tag{A.50}$$

Since $\overline{E}[\overline{\lambda}_k|\mathcal{F}_{k-1}] = 1$,

$$\begin{aligned} E[f(w_k)|\mathcal{F}_{k-1}] &= \overline{E}[\overline{\lambda}_k f(w_k)|\mathcal{F}_{k-1}] \qquad (A.51)\\ &= \overline{E}\left[\frac{\phi(B^{-1}(x_k - Ax_{k-1}))}{|B|\phi(x_k)} f(B^{-1}(x_k - Ax_{k-1}))|\mathcal{F}_{k-1}\right] \\ &= \int \frac{\phi(B^{-1}(x_k - Ax_{k-1}))}{|B|\phi(x_k)} f(B^{-1}(x_k - Ax_{k-1}))\phi(x_k)dx_k \\ &= \int |B|^{-1}\phi(B^{-1}(x_k - Ax_{k-1}))f(B^{-1}(x_k - Ax_{k-1}))dx_k. \end{aligned}$$

Using the change of variables $u = B^{-1}(x_k - Ax_{k-1})$, *we have*

$$E[f(w_k)|\mathcal{F}_{k-1}] = \int f(u)\phi(u)du. \tag{A.52}$$

Since Eqn. A.52 is true for all test functions $f(.)$, *the lemma is proved.* □

Similar to the scalar case presented in Example A.4, under the newly defined probability measure P, the state process $\{x_k\}$ satisfies the dynamics given in Eqn. A.40. Therefore, the change of measure obtained using the Radon-Nikodym derivatives $\overline{\Lambda}_k$ again passes us from an ideal probability measure $\overline{P}$ to the probability measure P under which we require our results.

A.2 Derivation of Kalman Filter

Let $(\Omega, \mathcal{F}, P)$ be a probability space (under which we required our results) upon which $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are normally distributed with means 0 and covariance matrices I_n and I_m respectively. Assume that B_k and D_k are $n \times n$ and $m \times m$ nonsingular matrices respectively. Let $x_0 \in \mathbb{R}^n$ be a Gaussian random variable with mean $\bar{x}_0$ and covariance Σ_0.

We consider the linear Gauss-Markov system for which the state and observations satisfy the following equations.

$$x_{k+1} = A_{k+1}x_k + B_{k+1}w_{k+1}, \tag{A.53}$$

$$y_k = C_k x_k + D_k v_k. \tag{A.54}$$

Let $\{\mathcal{F}_k\}$, $k \in \mathbb{N}$ and $\{\mathcal{Y}_k\}$, $k \in \mathbb{N}$ be the complete filtrations generated by $\{x_0, \ldots, x_k\}$ and by $\{y_0, \ldots, y_k\}$ respectively. The minimum mean square error (MMSE) estimate $\hat{x}_{k|k}^{MS}$ of the state of this system is defined as

$$\hat{x}_{k|k}^{MS} = \arg\min_{\xi \in \mathbb{R}^n} E\left[\frac{1}{2}(x_k - \xi)^T Q_k (x_k - \xi)\Big|\mathcal{Y}_k\right] \tag{A.55}$$

where $\{Q_k\}$, $k \in \mathbb{N}$ is a sequence of positive definite matrices. Solution to this problem is given basically by the conditional mean of the state x_k given the information in $\mathcal{Y}_k$. Kalman filter is the well-known estimator which calculates these MMSE estimates and the related covariances given as

$$\hat{x}_{k|k}^{MS} = E[x_k|\mathcal{Y}_k], \tag{A.56}$$

$$\Sigma_{k|k}^{MS} = E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T|\mathcal{Y}_k]. \tag{A.57}$$

In this section, using the reference probability method, we are going to derive Kalman filter recursions.

Note that the above quantities requires the densities $p(x_k|\mathcal{Y}_k)$ which are propagated using nonlinear Bayesian recursions. In the following, we are going to show that the same calculations are possible using some unnormalized densities (which do not integrate to unity).

A.2.1 Change of Measure

Initially, we suppose all processes are defined on an "ideal" probability space $(\Omega, \mathcal{F}, \overline{P})$; then under the probability measure P, to be defined, the model described by Eqn. A.53 and A.54 will hold.

Suppose that, under $\overline{P}$,

- $\{x_k\}$, $k \in \mathbb{N}$ is a sequence independent and identically distributed (i.i.d.) random variables which are Gaussian distributed with zero mean and covariance I_n. Call their density function as $\phi(x) = \mathcal{N}(x; 0, I_n)$.

- $\{y_k\}$, $k \in \mathbb{N}$ is a sequence i.i.d. random variables which are Gaussian distributed with zero mean and covariance I_m. Call their density function as $\psi(y) = \mathcal{N}(y; 0, I_m)$.

In order to obtain the required Radon-Nikodym derivative, we define the sequence of random variables $\{\overline{\lambda}_l\}$ and $\{\overline{\Lambda}_k\}$, $k, l \in \mathbb{N}$ as

$$\overline{\lambda}_l = \begin{cases} \frac{\phi(\sqrt{\Sigma_0}^{-1}(x_l - \bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_l)} \frac{\psi(D_l^{-1}(y_l - C_l x_l))}{|D_l|\psi(y_l)} & l = 0 \\ \frac{\phi(B_l^{-1}(x_l - A_l x_{l-1}))}{|B_l|\phi(x_l)} \frac{\psi(D_l^{-1}(y_l - C_l x_l))}{|D_l|\psi(y_l)} & l > 0 \end{cases}, \tag{A.58}$$

$$\overline{\Lambda}_k = \prod_{l=0}^{k} \overline{\lambda}_l. \tag{A.59}$$

Let $\mathcal{G}_k$ be the complete σ-field generated by the random variables $\{x_0, \ldots, x_k, y_0, \ldots, y_k\}$.

Lemma A.6 *The process $\{\overline{\Lambda}_k\}$, $k \in \mathbb{N}$ is a $\{\mathcal{G}_k, \overline{P}\}$-martingale.*

Proof Noting that $\overline{\Lambda}_{k-1}$ is $\mathcal{G}_{k-1}$-measurable,

$$\overline{E}[\overline{\Lambda}_k | \mathcal{G}_{k-1}] = \overline{\Lambda}_{k-1} \overline{E}[\overline{\lambda}_k | \mathcal{G}_{k-1}]. \tag{A.60}$$

Therefore, it is enough to prove that $\overline{E}[\overline{\lambda}_k | \mathcal{G}_{k-1}] = 1$.

$$\begin{aligned} \overline{E}[\overline{\lambda}_k | \mathcal{G}_{k-1}] &= \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \Big| \mathcal{G}_{k-1}\Big] \\ &= \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \\ &\quad \times \overline{E}\Big[\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \Big| \mathcal{G}_{k-1}, x_k\Big] \Big| \mathcal{G}_{k-1}\Big]. \end{aligned} \tag{A.61}$$

We can calculate the inside expectation as

$$\begin{aligned} \overline{E}\Big[\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \Big| \mathcal{G}_{k-1}, x_k\Big] &= \int \frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \psi(y_k) dy_k \\ &= \int \psi(u) du = 1. \end{aligned} \tag{A.62}$$

Substituting this result into Eqn. A.61

$$\overline{E}[\overline{\lambda}_k | \mathcal{G}_{k-1}] = \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \Big| \mathcal{G}_{k-1}\Big] \tag{A.63}$$

$$= \int \frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \phi(x_k) dx_k \tag{A.64}$$

$$= \int \phi(u) du = 1. \tag{A.65}$$

This result proves the lemma. □

We now define a new probability measure on $(\Omega, \mathcal{F})$ by setting the restriction of the Radon-Nikodym derivative $\frac{dP}{d\overline{P}}$ to $\mathcal{G}_k$ equal to $\overline{\Lambda}_k$, that is,

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \overline{\Lambda}_k. \tag{A.66}$$

Lemma A.7 *On $\{\Omega, \mathcal{F}\}$ and under P, $\{w_k \in \mathbb{R}^n\}$ and $\{v_k \in \mathbb{R}^m\}$, $k \in \mathbb{N}$ defined as*

$$w_k \triangleq B_k^{-1}(x_k - A_k x_{k-1}), \tag{A.67}$$

$$v_k \triangleq D_k^{-1}(y_k - C_k x_k) \tag{A.68}$$

are sequences of i.i.d. Gaussian random variables with zero mean and covariance I_n and I_m respectively.

Proof Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$ be test functions. Then with E and $\overline{E}$ denoting the expectation under P and $\overline{P}$ respectively, using Theorem A.3, we obtain

$$E[f(w_k)g(v_k)|\mathcal{G}_{k-1}] = \frac{\overline{E}[\overline{\Lambda}_k f(w_k)g(v_k)|\mathcal{G}_{k-1}]}{\overline{E}[\overline{\Lambda}_k|\mathcal{G}_{k-1}]} \tag{A.69}$$

$$= \overline{E}[\overline{\lambda}_k f(w_k)g(v_k)|\mathcal{G}_{k-1}]. \tag{A.70}$$

Substituting $\overline{\lambda}_k$, w_k, and v_k into Eqn. A.70, we get

$$\begin{aligned} E[f(w_k)g(v_k)|\mathcal{G}_{k-1}] &= \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)}\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \\ &\quad \times f(B_k^{-1}(x_k - A_k x_{k-1}))g(D_k^{-1}(y_k - C_k x_k))\Big|\mathcal{G}_{k-1}\Big] \\ &= \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} f(B_k^{-1}(x_k - A_k x_{k-1})) \\ &\quad \times \overline{E}\Big[\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{D_k\psi(y_k)} \\ &\quad \times g(D_k^{-1}(y_k - C_k x_k))\Big|\mathcal{G}_{k-1}, x_k\Big]\Big|\mathcal{G}_{k-1}\Big]. \end{aligned} \tag{A.71}$$

We can calculate the inner expectation as

$$\overline{E}\Big[\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} g(D_k^{-1}(y_k - C_k x_k))\Big|\mathcal{G}_{k-1}, x_k\Big]$$

$$= \int \frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} g(D_k^{-1}(y_k - C_k x_k))\psi(y_k)dy_k \tag{A.72}$$

$$= \int |D_k|^{-1}\psi(D_k^{-1}(y_k - C_k x_k))g(D_k^{-1}(y_k - C_k x_k))dy_k. \tag{A.73}$$

After a change of variable $u = D_{k+1}^{-1}(y_{k+1} - C_{k+1}x_{k+1})$, the result becomes

$$\overline{E}\Big[\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} g(D_k^{-1}(y_k - C_k x_k))\Big|\mathcal{G}_{k-1}, x_k\Big] = \int g(u)\psi(u)du. \quad \text{(A.74)}$$

which is a constant (independent of x_k). Substituting this result into Eqn. A.71, we obtain

$$\begin{aligned} E[f(w_k)g(v_k)|\mathcal{G}_{k-1}] &= \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} f(B_k^{-1}(x_k - A_k x_{k-1})) \\ &\quad \times \int g(u)\psi(u)du\Big|\mathcal{G}_{k-1}\Big] \quad \text{(A.75)} \\ &= \int g(u)\psi(u)du \int \frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \\ &\quad \times f(B_k^{-1}(x_k - A_k x_{k-1}))\phi(x_k)dx_k \quad \text{(A.76)} \\ &= \int g(u)\psi(u)du \int f(v)\phi(v)dv. \quad \text{(A.77)} \end{aligned}$$

This result proves the lemma. □

A.2.2 Recursive Estimation

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a "test" function. Using Theorem A.3, we get

$$E[g(x_k)|\mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k g(x_k)|\mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k|\mathcal{Y}_k]}. \quad \text{(A.78)}$$

The unnormalized conditional expectation $\overline{E}[\overline{\Lambda}_k g(x_k)|\mathcal{Y}_k]$ in the numerator is the critical part. This is called generally as a measure-valued process and for obtaining it, we can define the unnormalized density $\alpha_k(x)$ as

$$\alpha_k(x)dx \triangleq E[\overline{\Lambda}_k \mathcal{I}_{\{x_k \in dx\}}|\mathcal{Y}_k]. \quad \text{(A.79)}$$

where the function $\mathcal{I}_A(\omega)$ defined as

$$\mathcal{I}_A(\omega) \triangleq \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \quad \text{(A.80)}$$

denotes the indicator function of the set A. The density $\alpha_k(x)$ can be interpreted as an information state for the solution of the problem under probability measure $\overline{P}$ and it can be shown for any test function $f : \mathbb{R}^n \to \mathbb{R}$ that

$$\overline{E}[\overline{\Lambda}_k f(x_k)|\mathcal{Y}_k] = \int f(x)\alpha_k(x)dx. \quad \text{(A.81)}$$

We can define, in the same way, the normalized conditional density $p_k(x)$ such that

$$E[f(x_k)|\mathcal{Y}_k] = \int f(x)p_k(x)dx. \tag{A.82}$$

Now, using Eqn. A.78, we see that

$$E[g(x_k)|\mathcal{Y}_k] = \int g(x)p_k(x)dx = \frac{\int g(x)\alpha_k(x)dx}{\int \alpha_k(x)dx} = \int g(x)\frac{\alpha_k(x)}{\int \alpha_k(z)dz}dx.$$

Thus, we obtain the result

$$p_k(x) = \frac{\alpha_k(x)}{\int \alpha_k(z)dz}. \tag{A.83}$$

Notice that the MMSE estimates and covariances given in Eqns. A.56 and A.57 are the mean and covariance of the density $p_k(x)$ which can be calculated using the unnormalized density $\alpha_k(x)$. In the following, a recursive relationship will be found for the unnormalized density $\alpha_k(x)$.

Theorem A.4 *The following recursive relationship holds for the unnormalized densities* $\alpha_k(x)$

$$\alpha_k(x) = \frac{\psi(D_k^{-1}(y_k - C_k x))}{|D_k||B_k|\psi(y_k)} \int \phi(B_k^{-1}(x - A_k z))\alpha_{k-1}(z)dz. \tag{A.84}$$

Proof Let $g : \mathbb{R}^n \to \mathbb{R}$ be any test function, then

$$\begin{aligned}
\int g(x)\alpha_k(x)dx &= \overline{E}[\overline{\Lambda}_k g(x_k)|\mathcal{Y}_k] && \text{(A.85)}\\
&= \overline{E}[\overline{\Lambda}_{k-1}\overline{\lambda}_k g(x_k)|\mathcal{Y}_k] && \text{(A.86)}\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)}\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \\
&\quad \times g(x_k)\Big|\mathcal{Y}_k\Big] && \text{(A.87)}\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)}\frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \\
&\quad \times g(x_k)\Big|x_{k-1}, \mathcal{Y}_k\Big]\Big|\mathcal{Y}_k\Big] && \text{(A.88)}
\end{aligned}$$

The inner expectation in Eqn. A.88 can easily be taken as follows due to the independence properties of the sequence $\{x_k\}$ under $\overline{P}$.

$$\int g(x)\alpha_k(x)dx = \overline{E}\Big[\overline{\Lambda}_{k-1}\int \frac{\phi(B_k^{-1}(x - A_k x_{k-1}))\psi(D_k^{-1}(y_k - C_k x))}{|B_k||D_k|\phi(x)\psi(y_k)}$$

$$
\begin{aligned}
& \times g(x)\phi(x)dx\Big|\mathcal{Y}_k\Big] \\
= \; & \overline{E}\Big[\overline{\Lambda}_{k-1}\int \frac{\phi(B_k^{-1}(x-A_kx_{k-1}))\psi(D_k^{-1}(y_k-C_kx))}{|B_k||D_k|\psi(y_k)} \\
& \times g(x)dx\Big|\mathcal{Y}_k\Big] && \text{(A.89)} \\
= \; & \overline{E}\Big[\overline{\Lambda}_{k-1}\int \frac{\phi(B_k^{-1}(x-A_kx_{k-1}))\psi(D_k^{-1}(y_k-C_kx))}{|B_k||D_k|\psi(y_k)} \\
& \times g(x)dx\Big|\mathcal{Y}_{k-1}\Big] && \text{(A.90)} \\
= \; & \int\int \frac{\phi(B_k^{-1}(x-A_kz))\psi(D_k^{-1}(y_k-C_kx))}{|B_k||D_k|\psi(y_k)} \\
& \times g(x)\alpha_{k-1}(z)dxdz && \text{(A.91)} \\
= \; & \int g(x)\frac{\psi(D_k^{-1}(y_k-C_kx))}{|B_k||D_k|\psi(y_k)} \\
& \times \int \phi(B_k^{-1}(x-A_kz))\alpha_{k-1}(z)dzdx. && \text{(A.92)}
\end{aligned}
$$

Since this equality holds for all test functions $g(.)$, the theorem is proved. □

The initial densities $\alpha_0(x)$ can be calculated using the test function $h : \mathbb{R}^n \to \mathbb{R}^n$ as follows.

$$
\begin{aligned}
\int h(x)\alpha_0(x)dx &= \overline{E}[\overline{\Lambda}_0 h(x_0)|\mathcal{Y}_0] && \text{(A.93)} \\
&= \overline{E}[\overline{\lambda}_0 h(x_0)|\mathcal{Y}_0] && \text{(A.94)} \\
&= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)}\frac{\psi(D_0^{-1}(y_0-C_0x_0))}{|D_0|\psi(y_0)}h(x_0)\Big|\mathcal{Y}_0\Big] \\
&= \int \frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x)}\frac{\psi(D_0^{-1}(y_0-C_0x))}{|D_0|\psi(y_0)}h(x)\phi(x)dx \\
&= \int h(x)\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|}\frac{\psi(D_0^{-1}(y_0-C_0x))}{|D_0|\psi(y_0)}dx. && \text{(A.95)}
\end{aligned}
$$

Since the equality holds for all test function $h(.)$, the initial density $\alpha_0(x)$ is given as

$$
\begin{aligned}
\alpha_0(x) &= \frac{\psi(D_0^{-1}(y_0-C_0x))}{|D_0|\psi(y_0)}\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|} && \text{(A.96)} \\
&= \frac{1}{\psi(y_0)}\mathcal{N}(y_0; C_0x, D_0D_0^T)\mathcal{N}(x; \bar{x}_0, \Sigma_0) && \text{(A.97)} \\
&= \frac{1}{\psi(y_0)}\mathcal{N}(y_0; C_0\bar{x}_0, S_0)\mathcal{N}(x; \hat{x}_{0|0}, \Sigma_{0|0}) && \text{(A.98)} \\
&= \bar{c}_0\mathcal{N}(x; \hat{x}_{0|0}, \Sigma_{0|0}) && \text{(A.99)}
\end{aligned}
$$

where

$$S_0 \triangleq C_0\Sigma_0 C_0^T + D_0 D_0^T, \tag{A.100}$$

$$\hat{x}_{0|0} \triangleq \bar{x}_0 + \Sigma_0 C_0^T (S_0)^{-1}(y_0 - C_0\bar{x}_0), \tag{A.101}$$

$$\Sigma_{0|0} \triangleq \Sigma_0 - \Sigma_0 C_0^T S_0^{-1} C_0 \Sigma_0, \tag{A.102}$$

$$\bar{c}_0 = \frac{1}{\psi(y_0)}\mathcal{N}(y_0; C_0\bar{x}_0, S_0). \tag{A.103}$$

Here, while going from Eqn. A.97 to Eqn. A.98, we used the result of App. B.2. Note that the initial density $\alpha_0(x)$ is an unnormalized Gaussian. Since the recursion given in Eqn. A.84 is linear, the unnormalized densities $\alpha_k(x)$ are all unnormalized Gaussian densities. Suppose that the density $\alpha_k(x)$ is given as

$$\alpha_k(x) = \bar{c}_k \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}). \tag{A.104}$$

Using this, we can find the normalized densities $p_k(x)$ as

$$p_k(x) = \frac{\alpha_k(x)}{\int \alpha_k(z)dz} = \frac{\alpha_k(x)}{\bar{c}_k} = \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}). \tag{A.105}$$

Therefore, the quantities $\hat{x}_{k|k}$ and $\Sigma_{k|k}$ are actually minimum mean square error estimates and covariances defined as

$$\hat{x}_{k|k} \triangleq E[x_k|\mathcal{Y}_k] = \hat{x}_{k|k}^{MS}, \tag{A.106}$$

$$\Sigma_{k|k} \triangleq E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T|\mathcal{Y}_k] = \Sigma_{k|k}^{MS}. \tag{A.107}$$

Now, we are now going to find recursive relations for the estimates $\hat{x}_{k|k}$, covariances $\Sigma_{k|k}$ and the weighting coefficients $\bar{c}_k$ using the recursion of Eqn. A.84. Using the facts that

$$\frac{\phi(B_k^{-1}(x - A_k z))}{|B_k|} = \mathcal{N}(x; A_k z, B_k B_k^T), \tag{A.108}$$

$$\frac{\psi(D_k^{-1}(y_k - C_k x))}{|D_k|} = \mathcal{N}(y_k; C_k x, D_k D_k^T), \tag{A.109}$$

we can write the recursion of Eqn. A.84 as

$$\alpha_k(x) = \frac{\mathcal{N}(y_k; C_k x, D_k D_k^T)}{\psi(y_k)} \int \mathcal{N}(x; A_k z, B_k B_k^T)\alpha_{k-1}(z)dz. \tag{A.110}$$

Substituting the expression of $\alpha_{k-1}(z)$ implied by Eqn. A.104 into Eqn. A.110, we get

$$\begin{aligned}\alpha_k(x) &= \frac{\bar{c}_{k-1}}{\psi(y_k)}\mathcal{N}(y_k; C_k x, D_k D_k^T)\int \mathcal{N}(x; A_k z, B_k B_k^T) \\ &\quad \times \mathcal{N}(z; \hat{x}_{k-1|k-1}, \Sigma_{k-1|k-1})dz. \end{aligned} \tag{A.111}$$

Using the result on the integral of Gaussian densities given in App. B.4 on the integral on the right hand side of Eqn. A.111, we obtain

$$\alpha_k(x) = \frac{\bar{c}_{k-1}}{\psi(y_k)}\mathcal{N}(y_k; C_k x, D_k D_k^T)\mathcal{N}(x; \hat{x}_{k|k-1}, \Sigma_{k|k-1}) \tag{A.112}$$

where

$$\hat{x}_{k|k-1} \triangleq A_k \hat{x}_{k-1|k-1}, \tag{A.113}$$

$$\Sigma_{k|k-1} \triangleq A_k \Sigma_{k-1|k-1} A_k^T + B_k B_k^T, \tag{A.114}$$

which are called as the Kalman prediction update equations. Making use of the result on multiplication of Gaussian densities given in App. B.2 on the multiplication of Eqn. A.112, we get

$$\alpha_k(x) = \bar{c}_k \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}) \tag{A.115}$$

where

$$\bar{c}_k = \frac{\bar{c}_{k-1}}{\psi(y_k)}\mathcal{N}(y_k; C_k \hat{x}_{k|k-1}, S_k), \tag{A.116}$$

$$S_k = C_k \Sigma_{k|k-1} C_k^T + D_k D_k^T, \tag{A.117}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \Sigma_{k|k-1} C_k^T S_k^{-1}(y_k - C_k \hat{x}_{k|k-1}), \tag{A.118}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1} C_k^T S_k^{-1} C_k \Sigma_{k|k-1}, \tag{A.119}$$

which are called as the Kalman measurement update equations. Note that the weighting coefficients $\bar{c}_k$ are not required for the recursions of the estimate $\hat{x}_{k|k}$ and the covariance $\Sigma_{k|k}$. Therefore, one need not propagate the weighting coefficients in the execution of the Kalman filter.

Although the derivation presented above is straightforward, it makes one recognize the key elements in these types of derivations. The main points can be summarized in two items given as

- Integral of the unnormalized Gaussian density multiplied by another Gaussian density represents Kalman filter prediction update.
- Multiplication of a density with a Gaussian density represents Kalman filter measurement update.

These are obviously the generalized versions of the Bayesian density recursions. The above observations are highly useful when interpreting the complex density recursions involved in Chapters 3 and 5.

It is also appropriate here to emphasize that the reference probability used here can be applied to non-Gaussian noise cases as long as the absolute continuity condition of the measures is not violated. However, in that case, obtaining the recursive forms for the algorithms would be much more difficult because the multiplication and the integral of the non-Gaussian densities will not be evaluated and manipulated as easily as done above.

A.3 Derivation of the Risk-Sensitive Filter for Linear Gauss-Markov Systems

In this section, we again consider the linear Gauss-Markov system defined in App. A.2. The MMSE estimate $\hat{x}_{k|k}^{MS}$ of the state of this system was defined as

$$\hat{x}_{k|k}^{MS} = \arg\min_{\xi\in\mathbb{R}^n} E\left[\frac{1}{2}(x_k-\xi)^T Q_k (x_k-\xi)\Big|\mathcal{Y}_k\right] \tag{A.120}$$

An alternative definition is given as follows.

$$\hat{x}_{k|k}^{MS} = \arg\min_{\xi\in\mathbb{R}^n} E\left[\Phi_{0,k}(\xi)|\mathcal{Y}_k\right] \tag{A.121}$$

where

$$\Phi_{0,k}(\xi) \triangleq \hat{\Phi}_{0,k-1} + \frac{1}{2}(x_k-\xi)^T Q_k (x_k-\xi), \tag{A.122}$$

$$\hat{\Phi}_{m,n} \triangleq \sum_{l=m}^{n} \frac{1}{2}(x_l - \hat{x}_{l|l}^{MS})^T Q_l (x_l - \hat{x}_{l|l}^{MS}). \tag{A.123}$$

Note that the two problems defined in Eqn. A.120 and Eqn. A.121 are equivalent to each other.[1] A risk-sensitive generalization of the problem given in Eqn. A.55 is to find $\hat{x}_{k|k}^{RS}$ as

$$\hat{x}_{k|k}^{RS} = \arg\min_{\zeta\in\mathbb{R}^n} E\left[\exp\left\{\frac{\theta}{2}(x_k-\zeta)^T Q_k(x_k-\zeta)\right\}\Big|\mathcal{Y}_k\right]. \tag{A.124}$$

This problem has been solved in [25] and the resulting filter turns out to be the same as the Kalman filter. The generalization of the second problem in Eqn. A.121 is to find $\hat{x}_{k|k}^{RS}$ as

$$\hat{x}_{k|k}^{RS} = \arg\min_{\zeta\in\mathbb{R}^n} E\left[\theta\exp\left\{\theta\Psi_{0,k}(\zeta)\right\}\Big|\mathcal{Y}_k\right] \tag{A.125}$$

where

$$\Psi_{0,k}(\zeta) \triangleq \hat{\Psi}_{0,k-1} + \frac{1}{2}(x_k-\zeta)^T Q_k(x_k-\zeta), \tag{A.126}$$

$$\hat{\Psi}_{m,n} \triangleq \sum_{l=m}^{n}\frac{1}{2}(x_l-\hat{x}_{l|l}^{RS})^T Q_l(x_l-\hat{x}_{l|l}^{RS}). \tag{A.127}$$

This risk sensitive estimation problem is much harder than its MS equivalent in Eqn. A.121 due to the fact that the expectation cannot be distributed over the individual terms after the exponential operation in Eqn. A.125 due to the highly dependent characteristics of the state process. This problem has been solved in [40] using the reference probability method, and here, we are going to re-derive the resulting filter using the change of measure given in App. A.2.1. Our change of measure makes both the measurement and state sequences i.i.d. whereas the change of measure in [40] makes only the measurement sequence i.i.d.. Note that this difference of change of measure requires the whole derivation to be repeated.

A.3.1 Derivation of the Filter

Using Theorem A.3, we can write the expectation in Eqn. A.125 as

$$E\left[\theta\exp\left\{\theta\Psi_{0,k}(\zeta)\right\}|\mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k\theta\exp\left\{\theta\Psi_{0,k}(\zeta)\right\}|\mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k|\mathcal{Y}_k\right]} \tag{A.128}$$

[1] This equivalence stems from the fact that expectation can be distributed over summation in Eqn. A.121.

where $\overline{E}$ and E denotes the expectations with respect to probability measures $\overline{P}$ and P respectively. Since the denominator of the right hand side of Eqn. A.128 is independent of ζ, the risk-sensitive estimate $\hat{x}^{RS}_{k|k}$ is equivalently given as

$$\hat{x}^{RS}_{k|k} = \arg\min_{\zeta\in\mathbb{R}^n} \overline{E}\left[\overline{\Lambda}_k \theta \exp\left\{\theta\Psi_{0,k}(\zeta)\right\}|\mathcal{Y}_k\right]. \tag{A.129}$$

At this point, we define the unnormalized density function $\gamma_k(x)$ as

$$\gamma_k(x)dx = \overline{E}\left[\overline{\Lambda}_k \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\}\mathcal{I}_{\{x_k\in dx\}}\Big|\mathcal{Y}_k\right]. \tag{A.130}$$

This density function can be interpreted as an "information state" for the problem [45]. Using a simple reasoning, we can see that, if $f(.)$ is any test function the following equality is satisfied.

$$\overline{E}\left[\overline{\Lambda}_k \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\} f(x_k)\Big|\mathcal{Y}_k\right] = \int f(x)\gamma_k(x)dx. \tag{A.131}$$

Using this, we can write the risk-sensitive estimate as

$$\begin{aligned}
\hat{x}^{RS}_{k|k} &= \arg\min_{\zeta\in\mathbb{R}^n} \overline{E}\left[\overline{\Lambda}_k \theta \exp\left\{\theta\Psi_{0,k}(\zeta)\right\}|\mathcal{Y}_k\right] &\text{(A.132)}\\
&= \arg\min_{\zeta\in\mathbb{R}^n} \overline{E}\left[\overline{\Lambda}_k \theta \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\}\exp\left\{\frac{\theta}{2}(x_k-\zeta)^T Q_k (x_k-\zeta)\right\}\Big|\mathcal{Y}_k\right]\\
&= \arg\min_{\zeta\in\mathbb{R}^n} \int \exp\left\{\frac{\theta}{2}(x-\zeta)^T Q_k (x-\zeta)\right\}\gamma_k(x)dx. &\text{(A.133)}
\end{aligned}$$

Notation: In the following part, due to length of the formulas, we will use the following abbreviations.

$$\exp(+,x,\bar{x},\Sigma) \triangleq \exp\left\{\frac{1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})\right\}, \tag{A.134}$$

$$\exp(-,x,\bar{x},\Sigma) \triangleq \exp\left\{-\frac{1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})\right\}. \tag{A.135}$$

The following theorem gives a recursion for the density $\gamma_k(x)$.

Theorem A.5 *The unnormalized density $\gamma_k(x)$ satisfies the following recursion.*

$$\begin{aligned}
\gamma_k(x) &= \frac{\psi(D_k^{-1}(y_k - C_k x))}{|B_k||D_k|\psi(y_k)} \int \phi(B_k^{-1}(x - A_k z))\\
&\quad \times \exp\left\{+, z, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q^{-1}_{k-1}\right\}\gamma_{k-1}(z)dz. &\text{(A.136)}
\end{aligned}$$

Proof Let $f : \mathbb{R}^n \to \mathbb{R}$ be any test function. Then

$$\begin{aligned}
\int f(x)\gamma_k(x)dx &\triangleq \overline{E}\left[\overline{\Lambda}_k \exp\left\{\theta\hat{\Psi}_{0,k-1}\right\} f(x_k)\Big|\mathcal{Y}_k\right] && \text{(A.137)}\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1}\overline{\lambda}_k \exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \exp\left\{+, x_{k-1}, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\\
&\quad \times f(x_k)\Big|\mathcal{Y}_k\Big] && \text{(A.138)}\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1} \exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)}\\
&\quad \times \frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)} \exp\left\{+, x_{k-1}, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\\
&\quad \times f(x_k)\Big|\mathcal{Y}_k\Big] && \text{(A.139)}\\
&= \overline{E}\Big[\overline{\Lambda}_{k-1} \exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \exp\left\{+, x_{k-1}, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\\
&\quad \times \overline{E}\Big[\frac{\phi(B_k^{-1}(x_k - A_k x_{k-1}))}{|B_k|\phi(x_k)} \frac{\psi(D_k^{-1}(y_k - C_k x_k))}{|D_k|\psi(y_k)}\\
&\quad \times f(x_k)\Big|x_{k-1}, \mathcal{Y}_k\Big]\Big|\mathcal{Y}_k\Big] && \text{(A.140)}
\end{aligned}$$

The inner expectation in Eqn. A.140 can easily be taken as follows due to the independence properties of the sequence $\{x_k\}$ under $\overline{P}$.

$$\begin{aligned}
\int f(x)\gamma_k(x)dx &= \overline{E}\Big[\overline{\Lambda}_{k-1} \exp\left\{\theta\hat{\Psi}_{0,k-2}\right\} \exp\left\{+, x_{k-1}, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\\
&\quad \times \int \frac{\phi(B_k^{-1}(x - A_k x_{k-1}))}{|B_k|\phi(x)} \frac{\psi(D_k^{-1}(y_k - C_k x))}{|D_k|\psi(y_k)}\\
&\quad \times f(x)\phi(x)dx\Big|\mathcal{Y}_{k-1}\Big] && \text{(A.141)}\\
&= \int \exp\left\{+, z, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\gamma_{k-1}(z)\\
&\quad \times \int \frac{\phi(B_k^{-1}(x - A_k z))}{|B_k|} \frac{\psi(D_k^{-1}(y_k - C_k x))}{|D_k|\psi(y_k)} f(x)dxdz\\
&= \int f(x)\Big[\frac{\psi(D_k^{-1}(y_k - C_k x))}{|B_k||D_k|\psi(y_k)} \int \phi(B_k^{-1}(x - A_k z))\\
&\quad \times \exp\left\{+, z, \hat{x}^{RS}_{k-1|k-1}, \frac{1}{\theta}Q_{k-1}\right\}\gamma_{k-1}(z)dz\Big]dx. && \text{(A.142)}
\end{aligned}$$

Since the equality is satisfied for any test function $f(.)$, the recursion of Eqn. A.136 holds. □

The initial density $\gamma_0(x)$ can be calculated as

$$\overline{E}\left[\overline{\Lambda}_0 \exp\left\{\theta\hat{\Psi}_{0,-1}\right\} f(x_0)\Big|\mathcal{Y}_0\right] = \overline{E}\left[\overline{\Lambda}_0 f(x_0)\Big|\mathcal{Y}_0\right] \quad \text{(A.143)}$$

$$
\begin{aligned}
&= \overline{E}\left[\overline{\lambda}_0 f(x_0)\Big|\mathcal{Y}_0\right] && \text{(A.144)}\\
&= \overline{E}\Big[\frac{\phi(\sqrt{\Sigma_0}^{-1}(x_0-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x_0)}\\
&\quad \times\frac{\psi(D_0^{-1}(y_0-C_0x_0))}{|D_0|\psi(y_0)}f(x_0)\Big|\mathcal{Y}_0\Big]\\
&= \int f(x)\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|\phi(x)}\\
&\quad \times\frac{\psi(D_0^{-1}(y_0-C_0x))}{|D_0|\psi(y_0)}\phi(x)dx. && \text{(A.145)}
\end{aligned}
$$

Thus

$$
\gamma_0(x) = \frac{\psi(D_0^{-1}(y_0-C_0x))}{|D_0|\psi(y_0)}\frac{\phi(\sqrt{\Sigma_0}^{-1}(x-\bar{x}_0))}{|\sqrt{\Sigma_0}|} \qquad \text{(A.146)}
$$

Note that this equation is the same as the formula for $\alpha_0(x)$ given in Eqn. A.96. Therefore $\gamma_0(x)$ is given as

$$
\gamma_0(x) = \bar{d}_0\mathcal{N}(x;\hat{x}_{0|0},\Sigma_{0|0}) \qquad \text{(A.147)}
$$

where

$$
\begin{aligned}
S_0 &\triangleq C_0\Sigma_0C_0^T + D_0D_0^T, && \text{(A.148)}\\
\hat{x}_{0|0} &\triangleq \bar{x}_0 + \Sigma_0C_0^T(S_0)^{-1}(y_0-C_0\bar{x}_0), && \text{(A.149)}\\
\Sigma_{0|0} &\triangleq \Sigma_0 - \Sigma_0C_0^TS_0^{-1}C_0\Sigma_0, && \text{(A.150)}\\
\bar{d}_0 &= \frac{1}{\psi(y_0)}\mathcal{N}(y_0;C_0\bar{x}_0,S_0). && \text{(A.151)}
\end{aligned}
$$

A careful examination of the recursion given in Eqn. A.136 reveals that all $\gamma_k(x)$ are in the form of $\gamma_0(x)$, i.e.,

$$
\gamma_k(x) = \bar{d}_k\mathcal{N}(x;\hat{x}_{k|k},\Sigma_{k|k}) \qquad \text{(A.152)}
$$

Substituting this form into Eqn. A.133, and using the result of App. B.3, we obtain the risk-sensitive state estimate $\hat{x}^{RS}_{k|k}$ as

$$
\begin{aligned}
\hat{x}^{RS}_{k|k} &= \arg\min_{\zeta\in\mathbb{R}^n}\bar{d}_k\int\exp\left\{\frac{\theta}{2}(x-\zeta)^TQ_k(x-\zeta)\right\}\mathcal{N}(x;\hat{x}_{k|k},\Sigma_{k|k})dx\\
&= \arg\min_{\zeta\in\mathbb{R}^n}\frac{\bar{d}_k}{\sqrt{|2\pi\Sigma_{k|k}|}}\int\exp\left\{\frac{\theta}{2}(x-\zeta)^TQ_k(x-\zeta)\right\}
\end{aligned}
$$

$$\times \exp\left\{-\frac{1}{2}(x-\hat{x}_{k|k})^T\Sigma_{k|k}^{-1}(x-\hat{x}_{k|k})\right\}dx \tag{A.153}$$

$$= \arg\min_{\zeta\in\mathbb{R}^n} \frac{\bar{d}_k\sqrt{|2\pi U|}}{\sqrt{|2\pi\Sigma_{k|k}|}} \exp\left\{\frac{1}{2}(\zeta-\hat{x}_{k|k})^T V(\zeta-\hat{x}_{k|k})\right\} \tag{A.154}$$

$$= \hat{x}_{k|k} \tag{A.155}$$

where

$$U \triangleq \left[\frac{1}{\theta}Q_k^{-1}-\Sigma_{k|k}\right]^{-1} \tag{A.156}$$

$$V \triangleq \left[\Sigma_{k|k}^{-1}-\theta Q_k\right]^{-1} \tag{A.157}$$

Hence, the mean $\hat{x}_{k|k}$ of the unnormalized Gaussian density $\gamma_k(x)$ is the risk-sensitive estimate $\hat{x}_{k|k}^{RS}$. This shows us that, all we need to obtain is a recursive relationship between consecutive $\hat{x}_{k|k}$ values. Now, assuming that we know the quantities $\bar{d}_{k-1}$, $\hat{x}_{k-1|k-1}$ and $\Sigma_{k-1|k-1}$, we obtain $\bar{d}_k$, $\hat{x}_{k|k}$ and $\Sigma_{k|k}$ using the recursion in Eqn. A.136. Note that the initial quantities $\bar{d}_0$, $\hat{x}_{0|0}$ and $\Sigma_{0|0}$ are already given by Eqn. A.151, A.149 and A.150 respectively. Substituting Eqn. A.152 into the recursion in Eqn. A.136, and noting that $\hat{x}_{k-1|k-1}^{RS} = \hat{x}_{k-1|k-1}$, we obtain

$$\begin{aligned}
\gamma_k(x) &= \frac{\bar{d}_{k-1}\psi(D_k^{-1}(y_k-C_kx))}{\sqrt{|2\pi\Sigma_{k-1|k-1}|}|D_k|\psi(y_k)}\int \mathcal{N}(x;A_kz,B_kB_k^T) \\
&\quad \times \exp\left\{+,z,\hat{x}_{k-1|k-1},\frac{1}{\theta}Q_{k-1}^{-1}\right\}\exp\left\{-,z,\hat{x}_{k-1|k-1},\Sigma_{k-1|k-1}\right\}dz \\
&= \frac{\bar{d}_{k-1}\psi(D_k^{-1}(y_k-C_kx))}{\sqrt{|2\pi\Sigma_{k-1|k-1}|}|D_k|\psi(y_k)}\int \mathcal{N}(x;A_kz,B_kB_k^T) \\
&\quad \times \exp\left\{-\frac{1}{2}(z-\hat{x}_{k-1|k-1})^T\left[\Sigma_{k-1|k-1}^{-1}-\theta Q_{k-1}\right](z-\hat{x}_{k-1|k-1})\right\}dz \\
&= \frac{\bar{d}_{k-1}\psi(D_k^{-1}(y_k-C_kx))}{\sqrt{|\Sigma_{k-1|k-1}||\Sigma_{k-1|k-1}^{-1}-\theta Q_{k-1}|}|D_k|\psi(y_k)}\int \mathcal{N}(x;A_kz,B_kB_k^T) \\
&\quad \times \mathcal{N}(z;\hat{x}_{k-1|k-1},[\Sigma_{k-1|k-1}^{-1}-\theta Q_{k-1}]^{-1})dz.
\end{aligned} \tag{A.158}$$

Note that while writing Eqn. A.158, it is assumed that $[\Sigma_{k-1|k-1}^{-1}-\theta Q_{k-1}] > 0$. Using the result of App. B.4, the integral in the Eqn. A.158 can be taken as follows

$$\gamma_k(x) = \frac{\bar{d}_{k-1}\psi(D_k^{-1}(y_k-C_kx))}{\sqrt{|\Sigma_{k-1|k-1}||\Sigma_{k-1|k-1}^{-1}-\theta Q_{k-1}|}|D_k|\psi(y_k)}\mathcal{N}(x;A_k\hat{x}_{k-1|k-1},\Sigma_{k|k-1})$$

$$= \frac{\bar{d}_{k-1}}{\sqrt{|\Sigma_{k-1|k-1}||\Sigma^{-1}_{k-1|k-1} - \theta Q_{k-1}|\psi(y_k)}} \mathcal{N}(y_k; C_k x, D_k D_k^T) \times \mathcal{N}(x; A_k \hat{x}_{k-1|k-1}, \Sigma_{k|k-1}) \quad \text{(A.159)}$$

where

$$\Sigma_{k|k-1} = A_k \left[\Sigma^{-1}_{k-1|k-1} - \theta Q_{k-1}\right]^{-1} A_k^T + B_k B_k^T. \quad \text{(A.160)}$$

Using the result of App. B.2 on the multiplication of the Gaussian densities in Eqn. A.159, we obtain

$$\gamma_k(x) = \frac{\bar{d}_{k-1}}{\sqrt{|\Sigma_{k-1|k-1}||\Sigma^{-1}_{k-1|k-1} - \theta Q_{k-1}|\psi(y_k)}} \mathcal{N}(y_k; C_k A_k \hat{x}_{k-1|k-1}, S_k) \times \mathcal{N}(x; \hat{x}_{k|k}, \Sigma_{k|k}) \quad \text{(A.161)}$$

where

$$S_k = C_k \Sigma_{k|k-1} C_k^T + D_k D_k^T, \quad \text{(A.162)}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1} C_k^T S_k^{-1} C_k \Sigma_{k|k-1}, \quad \text{(A.163)}$$

$$\hat{x}_{k|k} = A_k \hat{x}_{k-1|k-1} + \Sigma_{k|k-1} C_k^T S_k^{-1} \left(y_k - C_k A_k \hat{x}_{k-1|k-1}\right). \quad \text{(A.164)}$$

Summarizing the results,

$$\hat{x}^{RS}_{k|k} = A_k \hat{x}^{RS}_{k-1|k-1} + \Sigma_{k|k-1} C_k^T S_k^{-1} \left(y_k - C_k A_k \hat{x}^{RS}_{k-1|k-1}\right) \quad \text{(A.165)}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1} C_k^T S_k^{-1} C_k \Sigma_{k|k-1} \quad \text{(A.166)}$$

$$\Sigma_{k|k-1} = A_k \left[\Sigma^{-1}_{k-1} - \theta Q\right]^{-1} A_k^T + B_k B_k^T \quad \text{(A.167)}$$

$$S_k = C_k \Sigma_{k|k-1} C_k^T + D_k D_k^T \quad \text{(A.168)}$$

Note also that, the recursion for the constants $\bar{d}_k$ are given as follows,

$$\bar{d}_k = \frac{\bar{d}_{k-1}}{\sqrt{|\Sigma_{k-1|k-1}||\Sigma^{-1}_{k-1|k-1} - \theta Q_{k-1}|\psi(y_k)}} \mathcal{N}(y_k; C_k A_k \hat{x}_{k-1|k-1}, S_k)$$

Furthermore, the recursions described by Eqn. A.165, A.166, A.167 and A.168 can be written in prediction and measurement update forms (as in the case of Kalman filter) as

- Prediction Update

$$\hat{x}^{RS}_{k|k-1} = A_k \hat{x}^{RS}_{k-1|k-1}, \tag{A.169}$$

$$\Sigma_{k|k-1} = A_k \left[\Sigma^{-1}_{k-1|k-1} - \theta Q_{k-1}\right]^{-1} A_k^T + B_k B_k^T. \tag{A.170}$$

- Measurement Update

$$\hat{x}^{RS}_{k|k} = \hat{x}^{RS}_{k|k-1} + \Sigma_{k|k-1} C_k^T S_k^{-1} \left(y_k - C_k \hat{x}^{RS}_{k|k-1}\right), \tag{A.171}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1} C_k^T S_k^{-1} C_k \Sigma_{k|k-1}, \tag{A.172}$$

$$S_k = C_k \Sigma_{k|k-1} C_k^T + D_k D_k^T. \tag{A.173}$$

These updates show that the only difference of the risk-sensitive filter from the Kalman filter is in the covariance prediction update given in Eqn. A.170. The Eqn. A.170 also makes it clear that when $\theta = 0$, the risk-sensitive filter reduces to the Kalman filter. It is important here to see that the recursion of the weighting coefficients $\bar{d}_k$ is not required for the recursion of estimates and covariances. Therefore one need not propagate the weighting coefficients during the execution.

APPENDIX B

RESULTS ON MULTIPLICATION AND INTEGRAL OF EXPONENTIALS AND GAUSSIAN DENSITIES

This appendix gives brief results on multiplication and integral of exponentials and Gaussian densities which are used in some of the derivations of the thesis.

B.1 Multiplication of Two Exponentials

Define the multiplication $\mathcal{M}$ as

$$\begin{aligned}\mathcal{M} &\triangleq \exp\left\{-\frac{1}{2}(y-Cx)^T R^{-1}(y-Cx)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})\right\}. \end{aligned} \tag{B.1}$$

Using some completion to square argument, one can see that

$$\begin{aligned}\mathcal{M} &= \exp\left\{-\frac{1}{2}(y-C\bar{x})^T \left[C\Sigma C^T + R\right]^{-1}(y-C\bar{x})\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(x-g)^T F^{-1}(x-g)\right\} \end{aligned} \tag{B.2}$$

where

$$F \triangleq \left[\Sigma^{-1} + C^T R^{-1} C\right]^{-1} \tag{B.3}$$

$$= \Sigma - \Sigma C^T \left[C\Sigma C^T + R\right]^{-1} C\Sigma, \tag{B.4}$$

$$g \triangleq F\left[\Sigma^{-1}\bar{x} + C^T R^{-1} y\right] \tag{B.5}$$

$$= \bar{x} + \Sigma C^T \left[C\Sigma C^T + R\right]^{-1}(y - C\bar{x}). \tag{B.6}$$

B.2 Multiplication of Two Gaussian Densities

Using the standard Kalman filter measurement updates, we can show easily that, the following equation is satisfied.

$$\mathcal{N}(y; Cx, R)\mathcal{N}(x, \bar{x}, \Sigma) = \mathcal{N}(y, C\bar{x}, S)\mathcal{N}(x, g, F) \tag{B.7}$$

where

$$\begin{aligned} S &= C\Sigma C^T + R, & \text{(B.8)} \\ g &= \bar{x} + \Sigma C^T S^{-1}(y - C\bar{x}), & \text{(B.9)} \\ F &= \Sigma - \Sigma C^T S^{-1} C\Sigma. & \text{(B.10)} \end{aligned}$$

B.3 Integral of Two Exponentials

The integral $\mathcal{I}(\xi, \bar{x})$ defined as

$$\begin{aligned} \mathcal{I}(\xi, \bar{x}) \triangleq & \int \exp\left\{\frac{1}{2}(\xi - Ax)^T Q(\xi - Ax)\right\} \\ & \times \exp\left\{-\frac{1}{2}(x - \bar{x})^T \Sigma^{-1}(x - \bar{x})\right\} dx \end{aligned} \tag{B.11}$$

satisfies,

$$\mathcal{I}(\xi, \bar{x}) = \sqrt{|2\pi M|} \exp\left\{\frac{1}{2}(\xi - A\bar{x})^T S(\xi - A\bar{x})\right\} \tag{B.12}$$

where

$$\begin{aligned} S &\triangleq \left[Q^{-1} - A\Sigma A^T\right]^{-1} > 0, & \text{(B.13)} \\ M &\triangleq \left[\Sigma^{-1} - A^T Q A\right]^{-1} > 0. & \text{(B.14)} \end{aligned}$$

For the convergence of the integral, the matrices S and M must be positive definite, which are actually equivalent conditions.

B.4 Integral of Two Gaussian Densities

Using the standard Kalman filter prediction updates, we can show that the following equation is satisfied.

$$\int \mathcal{N}(x; Az, Q)\mathcal{N}(z; \bar{x}, \Sigma)dz = \mathcal{N}(x, A\bar{x}, A\Sigma A^T + Q)$$

where Q and Σ are positive definite matrices.

APPENDIX C

RECURSIVE PROJECTION ALGORITHM USED IN CHAPTER 4

The following operation finds the projection $\bar{x}$ of the N-vector x given as $x = [\ x_1 \quad x_2 \quad \cdots \quad x_N\ ]^T$ which satisfies the inequalities (See Remark 4.3)

$$\sum_{i=1}^{N} x_i \geq 1 \quad \text{and} \quad x_i \geq 0 \quad \text{for} \quad i = 1, 2, \ldots, N \tag{C.1}$$

onto the standard N-simplex defined by its elements in at most N recursions.

$$\bar{x} = \text{Project}(x, N) \tag{C.2}$$

where the recursive function Project(.,.) is defined as follows:

function p =Project(x, N)
  $\epsilon = \left(\sum_{i=1}^{N} x_i - 1\right)/N$
  if $\epsilon \leq \min_i x_i$
    $p = x - \epsilon \bar{1}_N$
  else
    $j = \arg\min_i x_i$
    $y = \left[x\left((1:(j-1))\right)^T x\left((j+1):N\right)^T\right]^T$
    m =Project$(y, N-1)$
    $p(1:(j-1)) = m(1:(j-1))$
    $p(j) = 0$
    $p((j+1):N) = m(j:(N-1))$
  end

Here, $\bar{1}_N = [\ 1 \quad 1 \quad \cdots \quad 1\ ]^T$ denotes the N-vector composed of ones and the notation $x(i:j)$ stands for

$$x(i:j) \triangleq \begin{cases} [\ x_i \quad \cdots \quad x_j\ ]^T & \text{for} \quad j \geq i \\ [.]^T & \text{otherwise} \end{cases} \tag{C.3}$$

where the notation [.] denotes an empty vector.

The proof of the fact that this operation obtains the projection of x onto the standard N-simplex can be made easily by formulating the problem as a standard quadratic programming problem and is not given here. See [62] for a general solution which does not require the conditions in Eqn. C.1.

APPENDIX D

DERIVATION OF THE ERROR COVARIANCE OF CHAPTER 6

In Chapter 6, the error covariance formula given in Eqn. 6.89 is given without proof for the sake of simplicity. In this appendix, we give a brief derivation of that result.

After defining the quantities

$$T_1 \triangleq \bar{x}_{app}, \tag{D.1}$$

$$T_2 \triangleq K_{sub}(y_k - \bar{y}), \tag{D.2}$$

$$T_3 \triangleq \sum_{i=1}^{N} p_i^+ \bar{x}_i, \tag{D.3}$$

$$T_4 \triangleq \sum_{i=1}^{N} p_i^+ K_i \left(y_k - \bar{y}_i\right), \tag{D.4}$$

it was noted in Chapter 6 that the covariance Σ_Δ is given as

$$\Sigma_\Delta = E_{y_k}[(T_1 + T_2 - T_3 - T_4)\,(T_1 + T_2 - T_3 - T_4)^T \,\big|\, Y^{k-1}]. \tag{D.5}$$

Thus, the covariance calculation requires the expected values of the form

$$E_{y_k}\left[T_i T_j^T | Y^{k-1}\right] \quad \text{for} \quad i, j = 1, \ldots, 4. \tag{D.6}$$

The approximation required to evaluate these expectations was suggested in Chapter 6 as:

- Assuming that the means $\bar{y}_i$ of the Gaussian components of $p(y_k|Y^{k-1})$ are sufficiently separated or equivalently

- Assuming that the measurement dependent posterior probability p_i^+ is approximately unity around the mean $\bar{y}_i$ within the 2σ covariance ellipse.

Using this approximation, it was shown in Chapter 6 that the following equation is satisfied.

$$E_{y_k}[(p_i^+)^2|Y^{k-1}] = p_i. \tag{D.7}$$

Similarly, using the same approximation, one can easily conclude that $p_i^+ p_j^+ \approx 0$ for all y_k values when $i \neq j$. Using these two facts, we can write the following result.

$$E_{y_k}[p_i^+ p_j^+|Y^{k-1}] = \begin{cases} p_i, & i = j \\ 0, & i \neq j \end{cases}. \tag{D.8}$$

Other basic expectations which can be evaluated similarly are listed below with their results:

- $E_{y_k}[p_i^+ p_j^+ y_k|Y^{k-1}] = \begin{cases} p_i\bar{y}_i, & i = j \\ 0, & i \neq j \end{cases}$,
- $E_{y_k}[p_i^+ p_j^+ (y_k - \bar{y}_i)(y_k - \bar{y}_j)^T|Y^{k-1}] = \begin{cases} p_i S_i, & i = j \\ 0, & i \neq j \end{cases}$.

Note that the property $p_i^+ p_j^+ \approx 0$ when $i \neq j$ is so useful that it turns the double summations appearing in the expectations like $E_{y_k}\left[T_3 T_3^T|Y^{k-1}\right]$, $E_{y_k}\left[T_3 T_4^T|Y^{k-1}\right]$, $E_{y_k}\left[T_4 T_4^T|Y^{k-1}\right]$, etc. into single summations.

D.1 Calculation of the Terms $E_{y_k}\left[T_i T_j^T|Y^{k-1}\right]$

Using the expectations calculated in Chapter 6 and the basic expectation results given in the previous part of the appendix, in this section, the expectations $E_{y_k}\left[T_i T_j^T|Y^{k-1}\right]$ are evaluated one by one. Only the case $i \leq j$ is investigated since

$$E_{y_k}\left[T_i T_j^T|Y^{k-1}\right] = \left(E_{y_k}\left[T_j T_i^T|Y^{k-1}\right]\right)^T. \tag{D.9}$$

- **Calculation of $E_{y_k}\left[T_1T_1^T|Y^{k-1}\right]$:**

$$E_{y_k}\left[T_1T_1^T|Y^{k-1}\right] = T_1T_1^T = \bar{x}_{app}\bar{x}_{app}^T. \tag{D.10}$$

- **Calculation of $E_{y_k}\left[T_1T_2^T|Y^{k-1}\right]$:**

$$E_{y_k}\left[T_1T_2^T|Y^{k-1}\right] = T_1E_{y_k}\left[T_2^T|Y^{k-1}\right] = \bar{x}_{app}0 = 0. \tag{D.11}$$

- **Calculation of $E_{y_k}\left[T_1T_3^T|Y^{k-1}\right]$:**

$$\begin{aligned} E_{y_k}\left[T_1T_3^T|Y^{k-1}\right] &= T_1E_{y_k}\left[T_3^T|Y^{k-1}\right] \\ &= \bar{x}_{app}\sum_{i=1}^{N} E_{y_k}\left[p_i^+|Y^{k-1}\right]\bar{x}_i^T \\ &= \bar{x}_{app}\sum_{i=1}^{N} p_i\bar{x}_i^T = \bar{x}_{app}\bar{x}_{app}^T. \end{aligned} \tag{D.12}$$

- **Calculation of $E_{y_k}\left[T_1T_4^T|Y^{k-1}\right]$:**

$$\begin{aligned} E_{y_k}\left[T_1T_4^T|Y^{k-1}\right] &= T_1E_{y_k}\left[T_4^T|Y^{k-1}\right] \\ &= \bar{x}_{app}\sum_{i=1}^{N}\left(E_{y_k}\left[p_i^+y_k|Y^{k-1}\right] - E_{y_k}\left[p_i^+|Y^{k-1}\right]\bar{y}_i\right)K_i^T \\ &= \bar{x}_{app}\sum_{i=1}^{N}\left(p_i\bar{y}_i - p_i\bar{y}_i\right)K_i^T = 0. \end{aligned} \tag{D.13}$$

- **Calculation of $E_{y_k}\left[T_2T_2^T|Y^{k-1}\right]$:**

$$\begin{aligned} E_{y_k}\left[T_2T_2^T|Y^{k-1}\right] &= K_{sub}E_{y_k}\left[(y_k-\bar{y})(y_k-\bar{y})^T|Y^{k-1}\right]K_{sub}^T \\ &= K_{sub}S_{sub}K_{sub}^T = \Sigma_{app}C^TS_{sub}^{-1}C\Sigma_{app}. \end{aligned} \tag{D.14}$$

- **Calculation of $E_{y_k}\left[T_2T_3^T|Y^{k-1}\right]$:**

$$\begin{aligned} E_{y_k}\left[T_2T_3^T|Y^{k-1}\right] &= K_{sub}\sum_{i=1}^{N}\left(E_{y_k}\left[p_i^+y_k|Y^{k-1}\right] - E_{y_k}\left[p_i^+|Y^{k-1}\right]\bar{y}\right)\bar{x}_i^T \\ &= K_{sub}\left[\sum_{i=1}^{N} p_i\bar{y}_i\bar{x}_i^T - \sum_{i=1}^{N} p_i\bar{y}\bar{x}_i^T\right] \\ &= K_{sub}\sum_{i=1}^{N} p_i(\bar{y}_i-\bar{y})\bar{x}_i^T. \end{aligned} \tag{D.15}$$

- **Calculation of $E_{y_k}\left[T_2T_4^T|Y^{k-1}\right]$:**

$$\begin{aligned}
E_{y_k}\left[T_2T_4^T|Y^{k-1}\right] &= K_{sub}\sum_{i=1}^{N} E_{y_k}\left[p_i^+(y_k-\bar{y})(y_k-\bar{y}_i)^T|Y^{k-1}\right]K_i^T \\
&= K_{sub}\sum_{i=1}^{N} E_{y_k}\left[p_i^+(y_k-\bar{y}_i)(y_k-\bar{y}_i)^T|Y^{k-1}\right]K_i^T \\
&\quad -K_{sub}\sum_{i=1}^{N} E_{y_k}\left[p_i^+(\bar{y}_i-\bar{y})(y_k-\bar{y}_i)^T|Y^{k-1}\right]K_i^T \\
&= K_{sub}\sum_{i=1}^{N} p_iS_iK_i^T \\
&\quad -K_{sub}\sum_{i=1}^{N}(\bar{y}_i-\bar{y})E_{y_k}\left[p_i^+(y_k-\bar{y}_i)^T|Y^{k-1}\right]K_i^T \\
&= K_{sub}\sum_{i=1}^{N} p_iS_iK_i^T = K_{sub}C\sum_{i=1}^{N} p_i\Sigma_i \\
&= K_{sub}C\left[\Sigma_{app}-\sum_{i=1}^{N} p_i(\bar{x}_i-\bar{x}_{app})(\bar{x}_i-\bar{x}_{app})^T\right] \\
&= K_{sub}C\left[\Sigma_{app}-\sum_{i=1}^{N} p_i(\bar{x}_i-\bar{x}_{app})\bar{x}_i^T\right] \\
&= \Sigma_{app}C^TS_{sub}^{-1}C\Sigma_{app}-K_{sub}\sum_{i=1}^{N} p_i(\bar{y}_i-\bar{y})\bar{x}_i^T.
\end{aligned}$$

- **Calculation of $E_{y_k}\left[T_3T_3^T|Y^{k-1}\right]$:**

$$E_{y_k}\left[T_3T_3^T|Y^{k-1}\right] = \sum_{i=1}^{N} E_{y_k}\left[(p_i^+)^2|Y^{k-1}\right]\bar{x}_i\bar{x}_i^T = \sum_{i=1}^{N} p_i\bar{x}_i\bar{x}_i^T. \quad \text{(D.16)}$$

- **Calculation of $E_{y_k}\left[T_3T_4^T|Y^{k-1}\right]$:**

$$\begin{aligned}
E_{y_k}\left[T_3T_4^T|Y^{k-1}\right] &= \sum_{i=1}^{N}\bar{x}_iE_{y_k}\left[(p_i^+)^2(y_k-\bar{y}_i)|Y^{k-1}\right]K_i^T \\
&= \sum_{i=1}^{N}\bar{x}_i(p_i\bar{y}_i-p_i\bar{y}_i)K_i^T = 0. \qquad \text{(D.17)}
\end{aligned}$$

- **Calculation of $E_{y_k}\left[T_4T_4^T|Y^{k-1}\right]$:**

$$\begin{aligned}
E_{y_k}\left[T_4T_4^T|Y^{k-1}\right] &= \sum_{i=1}^{N} K_iE_{y_k}\left[(p_i^+)^2(y_k-\bar{y}_i)(y_k-\bar{y}_i)^T|Y^{k-1}\right]K_i^T \\
&= \sum_{i=1}^{N} p_iK_iS_iK_i^T = \sum_{i=1}^{N} p_i\Sigma_iC^TS_i^{-1}C\Sigma_i. \qquad \text{(D.18)}
\end{aligned}$$

D.2 Calculation of the Covariance Σ_Δ

In this section, in order to evaluate the expectation

$$\Sigma_\Delta = E_{y_k}[(T_1 + T_2 - T_3 - T_4)\,(T_1 + T_2 - T_3 - T_4)^T \big| Y^{k-1}], \tag{D.19}$$

we combine the results of the previous section as

$$\begin{aligned}
\Sigma_\Delta &= \bar{x}_{app}\bar{x}_{app}^T - \bar{x}_{app}\bar{x}_{app}^T + \Sigma_{app}C^T S_{sub}^{-1} C\Sigma_{app} - K_{sub}\sum_{i=1}^{N} p_i(\bar{y}_i - \bar{y})\bar{x}_i^T \\
&\quad -\Sigma_{app}C^T S_{sub}^{-1} C\Sigma_{app} + K_{sub}\sum_{i=1}^{N} p_i(\bar{y}_i - \bar{y})\bar{x}_i^T - \bar{x}_{app}\bar{x}_{app}^T \\
&\quad -\sum_{i=1}^{N} p_i\bar{x}_i(\bar{y}_i - \bar{y})^T K_{sub}^T + \sum_{i=1}^{N} p_i\bar{x}_i\bar{x}_i^T - \Sigma_{app}C^T S_{sub}^{-1} C\Sigma_{app} \\
&\quad +\sum_{i=1}^{N} p_i\bar{x}_i(\bar{y}_i - \bar{y})^T K_{sub}^T + \sum_{i=1}^{N} p_i\Sigma_i C^T S_i^{-1} C\Sigma_i \qquad \text{(D.20)} \\
&= \sum_{i=1}^{N} p_i\Sigma_i C^T S_i^{-1} C\Sigma_i - \Sigma_{app}C^T S_{sub}^{-1} C\Sigma_{app} + \sum_{i=1}^{N} p_i\bar{x}_i\bar{x}_i^T - \bar{x}_{app}\bar{x}_{app}^T
\end{aligned}$$

which is the same as Eqn. 6.89. Note that while writing the right hand side of Eqn. D.20, we used the fact

$$E_{y_k}\left[T_i T_j^T | Y^{k-1}\right] = \left(E_{y_k}\left[T_j T_i^T | Y^{k-1}\right]\right)^T \tag{D.21}$$

extensively.

REFERENCES

[1] V. P. Jilkov and X. R. Li, "Online Bayesian estimation of transition probabilities for Markovian jump systems," *IEEE Trans. Signal Processing*, vol. 52, no. 6, pp. 1620–1630, June 2004.

[2] V. Krishnamurthy and J. B. Moore, "Online estimation of hidden Markov model parameters based on the Kullback-Leibler information measure," *IEEE Trans. Signal Processing*, vol. 41, no. 8, pp. 2557–2573, Aug. 1993.

[3] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 34–45, Mar. 1960.

[4] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*. New York: Technology Press and Wiley, 1949.

[5] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 146–181, Mar. 1974.

[6] M. Green and D. J. N. Limebeer, *Linear Robust Control*. New Jersey: Prentice Hall, 1995.

[7] G. A. Ackerson and K. S. Fu, "On state estimation in switching environments," *IEEE Trans. Automat. Contr.*, vol. 15, no. 1, pp. 10–17, Aug. 1970.

[8] J. K. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Trans. Automat. Contr.*, vol. 27, no. 5, pp. 1054–1065, Oct. 1982.

[9] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, vol. 13, no. 4, pp. 429–434, July 1977.

[10] J. K. Tugnait and A. H. Haddad, "A detection estimation scheme for state estimation in switching environment," *Automatica*, vol. 15, no. 4, pp. 477–481, July 1979.

[11] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markov switching coefficients," *IEEE Trans. Automat. Contr.*, vol. 33, no. 8, pp. 780–783, Aug. 1988.

[12] L. A. Johnston and V. Krishnamurthy, "An improvement to the interacting multiple model (IMM) algorithm," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2909–2923, Dec. 2001.

[13] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques, and Software.* Norwell, MA: Artech House, 1993.

[14] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.

[15] O. L. V. Costa, "Linear minimum mean square error estimation for discrete-time Markovian jump linear systems," *IEEE Trans. Automat. Contr.*, vol. 39, no. 8, pp. 1685–1689, Aug. 1994.

[16] O. L. V. Costa and S. Guerra, "Stationary filter for linear minimum mean square error estimator of discrete-time Markovian jump systems," *IEEE Trans. Automat. Contr.*, vol. 47, no. 8, pp. 1351–1356, Aug. 2002.

[17] Q. Zhang, "Optimal filtering of discrete-time hybrid systems," *Journal of Optimization Theory and Applications*, vol. 100, no. 1, pp. 123–144, Jan. 1999.

[18] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of jump Markov linear systems," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2139–2156, Aug. 1999.

[19] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc.-F: Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[20] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 613–624, Mar. 1999.

[21] A. Doucet, A. Logothetis, and V. Krishnamurthy, "Stochastic sampling algorithms for state estimation of jump Markov linear systems," *IEEE Trans. Automat. Contr.*, vol. 45, no. 1, pp. 188–202, Jan. 2000.

[22] Y. Boers and J. N. Driessen, "Interacting multiple model particle filter," *IEE P-Radar Son. Nav.*, vol. 150, no. 5, pp. 344–349, Oct. 2003.

[23] D. H. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic games," *IEEE Trans. Automat. Contr.*, vol. 18, no. 2, pp. 124–131, Apr. 1973.

[24] P. Whittle, "Risk-sensitive linear/quadratic/Gaussian control," *Adv. Applied Probability*, vol. 13, pp. 746–777, 1981.

[25] J. B. Moore, R. J. Elliott, and S. Dey, "Risk-sensitive generalizations of minimum variance estimation and control," *Journal of Mathematical Systems, Estimation and Control*, vol. 7, no. 1, pp. 1–15, 1997.

[26] J. L. Speyer, C. Fan, and R. N. Banavar, "Optimal stochastic estimation with exponential cost criteria," in *Proc. 31st Conf. Decision Contr.*, vol. 2, 1992, pp. 2293–2298.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[28] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.

[29] X. R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Trans. Automat. Contr.*, vol. 41, no. 4, pp. 478–493, Apr. 1996.

[30] C. E. de Souza and M. D. Fragoso, "$\mathcal{H}_\infty$ filtering for Markovian jump linear systems," *International Journal of Systems Science*, vol. 33, no. 11, pp. 909–915, 2002.

[31] ——, "Robust $\mathcal{H}_\infty$ filtering for uncertain Markovian jump linear systems," *Int. J. of Robust and Nonlinear Control*, vol. 12, pp. 435–466, 2002.

[32] ——, "$\mathcal{H}_\infty$ filtering for discrete-time linear systems with Markovian jumping parameters," *IEE Proc.-F: Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[33] C. E. de Souza, "A mode-independent $\mathcal{H}_\infty$ filter design for discrete-time Markovian jump linear systems," in *Proceedings of the* 42*nd Conference on Decision and Control*, Hawai, USA, 2003.

[34] M. S. Mahmoud and P. Shi, "Optimal guaranteed cost filtering for Markovian jump discrete-time systems," *Mathematical Problems In Engineering*, vol. 2004, no. 1, pp. 33–48, 2004.

[35] P. Shi, E. K. Boukas, and R. K. Agarwal, "Kalman filtering for continuous-time uncertain systems with Markovian jumping parameters," *IEEE Trans. Automat. Contr.*, vol. 44, no. 8, pp. 1592–1597, Aug. 1999.

[36] O. L. V. Costa and S. Guerra, "Robust linear filtering for discrete-time hybrid Markov linear systems," *International Journal of Control*, vol. 75, no. 10, pp. 712–727, 2002.

[37] R. K. Boel, M. R. James, and I. R. Petersen, "Robustness and risk-sensitive filtering," *IEEE Trans. Automat. Contr.*, vol. 47, no. 3, pp. 451–461, Mar. 2002.

[38] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. NY, USA: Springer-Verlag, 1994.

[39] L. Aggoun and R. Elliott, *Measure Theory and Filtering: Introduction with Applications*. United Kingdom: Cambridge University Press, 2004.

[40] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Trans. Automat. Contr.*, vol. 42, no. 11, pp. 1587–1591, Nov. 1997.

[41] ——, "Finite-dimensional risk-sensitive filters and smoothers for discrete-time nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 44, no. 6, pp. 1234–1239, June 1999.

[42] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*. Philadelphia: SIAM, 1999.

[43] R. J. Elliott, F. Dufour, and D. D. Sworder, "Exact hybrid filters in discrete time," *IEEE Trans. Automat. Contr.*, vol. 41, no. 12, pp. 1807–1810, Dec. 1996.

[44] V. Krishnamurthy and J. Evans, "Finite-dimensional filters for passive tracking of Markov jump linear systems," *Automatica*, vol. 34, no. 6, pp. 765–770, June 1998.

[45] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Englewood Cliffs, New Jersey: Prentice Hall, 1986.

[46] M. E. Liggins II, C. Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proc. IEEE*, vol. 85, no. 1, pp. 95–107, Jan. 1997.

[47] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," in *Proceedings of the American Control Conference*, Seattle, Washington, USA, 1995.

[48] Y. Sawaragi, T. Katayama, and S. Fujishige, "Adaptive estimation for a linear system with interrupted observation," *IEEE Trans. Automat. Contr.*, vol. 18, no. 2, pp. 152–154, Apr. 1973.

[49] J. K. Tugnait and A. H. Haddad, "State estimation under uncertain observations with unknown statistics," *IEEE Trans. Automat. Contr.*, vol. 24, no. 2, pp. 201–210, Apr. 1979.

[50] ——, "Adaptive estimation in linear systems with unknown markovian noise statistics," *IEEE Trans. Inform. Theory*, vol. 26, no. 1, pp. 66–78, Jan. 1980.

[51] J. K. Tugnait, "Adaptive estimation and identification for discrete systems with markov jump parameters," *IEEE Trans. Automat. Contr.*, vol. 27, no. 5, pp. 1054–1065, Oct. 1982.

[52] Z. Ghahramani and G. E. Hinton, "Variational learning for switching state-space models," *Neural Computation*, vol. 12, no. 4, pp. 831–864, Apr. 2000.

[53] E. K. Boukas, *Stochastic Switching System: Analysis and Design*. Boston: Birkhäuser, 2006.

[54] V. P. Jilkov, X. R. Li, and D. S. Angelova, "Estimation of Markovian jump systems with unknown transition probabilities through Bayesian sampling," in *NMA '02: Revised Papers from the 5th International Conference on Numerical Methods and Applications*. London, UK: Springer-Verlag, 2003, pp. 307–315.

[55] I. B. Collings, V. Krishnamurthy, and J. B. Moore, "Online identification of hidden Markov models via recursive prediction error techniques," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3535–3539, Dec. 1994.

[56] J. J. Ford and J. B. Moore, "Adaptive estimation of HMM transition probabilities," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1374–1385, May 1998.

[57] J. C. Stiller and G. Radons, "Online estimation of hidden Markov models," *IEEE Signal Processing Lett.*, vol. 6, no. 8, pp. 213–215, Aug. 1999.

[58] V. Krishnamurthy and G. G. Yin, "Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 458–476, Feb. 2002.

[59] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 9, pp. 1652–1654, Sept. 1990.

[60] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1990.

[61] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, vol. 22, pp. 400–407, 1951.

[62] A. S. Poznyak and K. Najim, *Learning Automata and Stochastic Optimization*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1997.

[63] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[64] H. W. Sorenson and D. L. Alspach, "Recursive bayesian estimation using gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[65] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[66] Z. Brzeźniak and T. Zastawniak, *Basic Stochastic Processes: A Course Through Exercises*. London, UK: Springer-Verlag, 1999.

VITA

Umut Orguner was born in İzmir, Turkey in 1977. He received his B.S. and M.S. degrees both in Electrical Engineering from Middle East Technical University (METU), Ankara, Turkey in 1999 and 2002 respectively. Since 1999, he is a research and teaching assistant in the Department of Electrical and Electronics Engineering at METU. During his graduate studies, he completed several university-industry collaboration projects under the support of Tübitak-Bilten (Information Technologies and Electronics Research Institute of Scientific and Technological Research Council of Turkey) and Aselsan Inc. on maneuvering, multi-target, multi-sensor target tracking in video and radar. His research interests include estimation theory, multiple-model estimation, target tracking and information fusion.

His publications are as follows:

Journal Papers

- U. Orguner and M. Demirekler, "An online sequential algorithm for the estimation of transition probabilities for jump Markov linear systems," *Automatica*, vol. 42, no. 10, pp. 1735–1744, Oct. 2006.
- B. Akın, U. Orguner, A. Ersak and M. Ehsani, "A simple derivative-free non-linear state observer for sensorless AC drives," *IEEE Transactions on Mechatronics*, vol. 11, no. 5, pp. 634–643, Oct. 2006.
- B. Akın, U. Orguner, H.A. Toliyat and M. Rayner, "Low order PWM inverter harmonics contributions to the inverter fed IM fault diagnosis," To appear in *IEEE Transactions on Industrial Electronics*.
- U. Orguner and M. Demirekler, "Risk-sensitive filtering for jump Markov linear systems," Submitted to *Automatica*.
- U. Orguner and M. Demirekler, "Analysis of the effects of the single Gaussian approximation of Gaussian mixtures in Bayesian filtering with

applications to mixed multiple-model estimation algorithms," Submitted to *International Journal of Control*.

- U. Orguner and M. Demirekler, "Maximum likelihood estimation of transition probabilities of jump Markov linear systems," To be submitted to *IEEE Transactions on Signal Processing*.

Conference Papers

- Ö. Salor, M. Demirekler and U. Orguner, "Kalman filter approach for pitch determination of speech signals," Proceedings of Speech and Computer 2006 (SPECOM'2006), St. Petersburg, Russia, Jun. 2006.

- B. Akın, U. Orguner, A. Ersak and M. Ehsani, "A comparative study on nonlinear state estimators applied to sensorless AC drives: MRAS and Kalman filter," Proceedings of the 30th Annual Conference of IEEE Industrial Electronics Society 2004 (IECON 2004), vol. 3, pp. 2148-2153, Nov. 2004.

- B. Akın, U. Orguner and A. Ersak, "A comparative study on Kalman filtering techniques designed for state estimation of AC drive systems," Proceedings of the IEEE International Conference on Mechatronics 2004 (ICM 2004), vol. 3, pp. 439-445, Istanbul, Turkey, Jun. 2004.

- B. Akın, U. Orguner and A. Ersak, "An experimental performance test of a derivative-free nonlinear state observer designed for sensorless AC drives," Proceedings of the IEEE International Conference on Mechatronics 2004 (ICM 2004), vol. 3, pp. 432-438, Istanbul, Turkey, Jun. 2004.

- B. Akın, U. Orguner and A. Ersak, "State estimation of induction motor using unscented Kalman filter," Proceedings of the IEEE International Conference on Control Applications 2003 (CCA 2003), vol. 2, pp. 915-919, Istanbul, Turkey, Jun. 2003.

M.S. Thesis

- U. Orguner, "Evaluation of Two New Methodologies for Maneuvering Target Tracking," Middle East Technical University, Ankara, Turkey, Sep. 2002.