EXPECTATION PROPAGATION FOR STATE ESTIMATION WITH DISCRETE-VALUED HIDDEN RANDOM VARIABLES

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

BY

ELİF SARITAŞ

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

FEBRUARY 2023

Approval of the thesis:

EXPECTATION PROPAGATION FOR STATE ESTIMATION WITH DISCRETE-VALUED HIDDEN RANDOM VARIABLES

submitted by **ELİF SARITAŞ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. Umut Orguner
Supervisor, **Electrical and Electronics Engineering** _____

Examining Committee Members:

Assoc. Prof. Dr. Emre Özkan
Electrical and Electronics Engineering Department, METU _____

Prof. Dr. Umut Orguner
Electrical and Electronics Engineering Department, METU _____

Assoc. Prof. Dr. Cem Tekin
Electrical and Electronics Engineering Dept., Bilkent Uni. _____

Assoc. Prof. Dr. Mustafa Mert Ankaralı
Electrical and Electronics Engineering Department, METU _____

Assist. Prof. Dr. Gökhan Soysal
Electrical and Electronics Engineering Dept., Ankara Uni. _____

Date: **21.02.2023**

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, Surname: Elif Sarıtaş

Signature :

ABSTRACT

EXPECTATION PROPAGATION FOR STATE ESTIMATION WITH DISCRETE-VALUED HIDDEN RANDOM VARIABLES

Sarıtaş, Elif
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Umut Orguner

February 2023, 114 pages

In this thesis, the expectation propagation (EP) approach of Minka is considered for the estimation problems in dynamical systems with discrete hidden random variables where optimal posteriors are usually intractable. The concept of context adjustment is introduced to avoid/alleviate indefinite covariance problems encountered in standard EP implementations in a systematic way. Additionally, the moment projection (M-projection) problem involving pseudo-Gaussian likelihoods as factors is solved to be used in the backward pass of the proposed smoothers.

The first type of estimation problem of interest investigates the so-called jump Markov linear systems (JMLS), where the state dynamics and/or measurement relation jumps between different alternatives based on the state of a Markov chain. This type of system model is extensively used in applications such as target tracking, fault detection and isolation, and machine learning. In the thesis, filtering and smoothing algorithms are derived using EP with context adjustment for JMLSs, and their relation to the existing methods in the literature is discussed. The simulation results on several scenarios show that the proposed algorithms have similar or better performance compared with the alternative methods.

The second type of problem considered in the thesis is the state estimation under measurement origin uncertainty. This problem, also known as data association or correspondence problem, frequently appears in applications such as sensor fusion and target tracking with imperfect sensors. A fixed-interval smoother based on EP with context adjustment is presented for the data association in multi-target tracking problem. Moreover, the suggested smoother is adapted to the filtering problem through a sliding-window mechanism. The proposed methods are compared to their alternatives with a discussion of their benefits and shortcomings.

Keywords: Expectation propagation, jump Markov linear systems, switching dynamical systems, state estimation, data association, multi-target tracking, smoothing, filtering.

ÖZ

AYRIK DEĞERLİ GİZLİ RASTGELE DEĞİŞKEN İÇEREN DURUM KESTİRİMİ İÇİN BEKLENTİ YAYILIMI

Sarıtaş, Elif
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Umut Orguner

Şubat 2023 , 114 sayfa

Bu tezde, en iyi sonsal çözümün genellikle hesaplanamaz olduğu ayrık değerli gizli rastgele değişken içeren dinamik sistemlerde durum kestirimi problemleri, Minka tarafından önerilen beklenti yayılımı (EP) yaklaşımıyla ele alınmıştır. EP uygulamalarında sistematik olarak karşılaşılan belirsiz ortak değişinti matrisi problemini önlemek için bağlam ayarlama kavramı ortaya atılmıltır. Ayrıca, önerilen düzleştime algoritmalarının geri yönde geçişlerinde kullanılmak üzere, sözde-Gauss olabilirlik fonksiyonu formunda çarpan içeren moment izdüşüm problemi çözülmüştür.

İrdelenen kestirim problemlerinden ilki, durum dinamiğinin veya ölçüm ilişkisinin bir Markov zincirinin durumuna bağlı olarak çeşitli seçenekler arasında atladığı Markov atlamalı doğrusal sistemlerde(JMLS) durum kestirimi problemidir. Bu tip sistem modeli, hedef izleme, hata tespiti ile ayrıştırılması ve makine öğrenmesi gibi alanlarda yoğunlukla kullanılmaktadır. Bu çalışmada, JMLS'ler için EP'ye dayanan özgün süzme ve düzleştirme algoritmaları türetilip literatürdeki diğer metotlarla ilişkileri araştırılmıştır. Çeşitli senaryolarla yürütülen benzetim çalışmalarının sonuçları, önerilen metotların benzerleriyle aynı oranda ya da onlardan daha başarılı olduğunu

ortaya koymuştur.

Çalışmada ele alınan diğer problem ise ölçüm - kaynak belirsizliği altında durum kestirimidir. Veri eşleme ya da örtüşme problemi olarak da bilinen bu problemle, duyucu tümleştirme ve hedef takibi gibi yetersiz veri üreten duyucu içeren uygulamalarda sıklıkla karşılaşılmaktadır. Bu tez çalışmasında, çoklu hedep takibinde veri eşleme problemi için, sabit aralıklı bir düzleştirici türetilmiştir. Ayrıca önerilen düzleştirici, kayar pencere aracılığıyla süzgeç olarak da uyarlanmıştır. Önerilen yöntemler benzerleri ile kıyaslanmış ve faydaları ile eksiklikleri üzerinde durulmuştur.

Anahtar Kelimeler: Beklenti yayılımı, Markov atlamalı doğrusal sistemler, anahtarlamalı dinamik sistemler, durum kestirimi, veri eşleme, çoklu hedef takibi, süzme, düzleştirme.

That’s all folks!

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Umut Orguner for his continuous guidance throughout this long journey. I also sincerely appreciate Emre Özkan, Cem Tekin, Mustafa Mert Ankaralı, and Gökhan Soysal for their insightful suggestions and constructive criticism.

I am genuinely thankful that I had the chance to meet Afşar Saranlı, Arzu Koç, Ece Şenan Güran Schmidt, Filiz Cerit, Mustafa Mert Ankaralı, and Mübeccel Demirekler all of whom have always been very sincere and helpful with me.

With their continuous support, my friends Cumhur Çakmak, Ayşe Deniz Duyul, Beril Beşbinar, Selim Özgen, Kenan Ahıska, Ezgi Kumru, and Fırat Kumru were particularly helpful during rough times, and I would like to recognize Hafize Zambak, Fürüzan Kumru, and Erdal Kumru for their great consideration. I am very thankful to each and every one of them.

My sisters Şeri and Selen Naz Sarıtaş always bring joy into my life, and this time was no exception. I am sincerely indebted to them. I also owe a debt of gratitude to my parents Selime and Osman Sarıtaş for their unwavering trust and support that enabled me to make it this far.

Lastly, I feel extremely lucky to have Murat Kumru by my side throughout this roller coaster ride. Without his encouragement and profound belief in me, this thesis would not have been possible. You are my rock and I am eagerly looking forward to exploring together what the future holds for us.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xvi

LIST OF FIGURES xviii

LIST OF ABBREVIATIONS xx

CHAPTERS

1 INTRODUCTION 1

1.1 Main Contributions 4

1.2 Outline of the Thesis 4

2 EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT . . . 7

2.1 Introduction 7

2.2 Variational Inference 8

2.3 Expectation Propagation 10

2.3.1 Idea of Context Adjustment 11

2.3.2 Related Ideas in the Literature 13

2.4 M-Projection with Pseudo-Likelihoods 14

2.5 Summary of the Chapter 15

3 EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT FOR SMOOTHING AND FILTERING OF JUMP MARKOV LINEAR SYSTEMS 17

3.1 Introduction 17

3.2 Problem Definition 19

3.3 Expectation Propagation with Context Adjustment for Jump Markov Linear Systems 21

3.3.1 Assumed Forms of the Factors 21

3.3.2 Derivation of the Updates 23

3.3.2.1 Update of the Forward Factor $\rho_n^f(\mathbf{x}_n, r_n)$ 23

3.3.2.2 Update for the Backward Factor $q_n^b(\mathbf{x}_n, r_n)$ 24

3.3.2.3 Computing the Final State Estimates 27

3.3.3 Selection of the Adjustment Exponents 27

3.3.3.1 Selection of γ_n^f 27

3.3.3.2 Selection of γ_n^b 28

3.3.4 Pseudo-Code of the Algorithm 29

3.3.5 Computational Complexity 33

3.4 Extension to Filtering Problem 33

3.4.1 Pseudo Code of the Filtering Algorithm 34

3.5 Simulation Results 34

3.5.1 Test Scenarios 34

3.5.1.1 Model-Match Scenario 35

3.5.1.2 Model-Mismatch Scenario 35

3.5.1.3 Barber's Scenario 36

3.5.2 Results . . . 37

3.5.2.1 Results for the Smoothing Algorithms . . . 37

3.5.2.2 Results for the Filtering Algorithms . . . 39

3.6 Conclusion . . . 40

4 EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT FOR TARGET TRACKING UNDER MEASUREMENT ORIGIN UNCERTAINTY . . . 43

4.1 Introduction . . . 43

4.2 Problem Definition . . . 45

4.3 Expectation Propagation with Context Adjustment for Data Association Problem in Multi-Target Tracking . . . 48

4.3.1 Assumed Forms of the Factors . . . 49

4.3.2 Derivation of the Updates . . . 50

4.3.2.1 Update of the Forward Factor $\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)$ . . . 50

4.3.2.2 Update of the Backward Factor $q_n^b(\mathbf{X}_n)$ . . . 51

4.3.2.3 Computing the Final State Estimates . . . 53

4.3.3 Selection of the Adjustment Exponents . . . 54

4.3.3.1 Selection of γ_n^f . . . 54

4.3.3.2 Selection of γ_n^b . . . 54

4.3.4 Pseudo-Code of the Algorithm . . . 54

4.3.5 Computational Complexity . . . 57

4.4 Extension to Filtering Problem . . . 58

4.4.1 Pseudo Code of the Filtering Algorithm . . . 58

4.5 Simulation Results . . . 58

4.5.1 Test Scenarios 58
4.5.1.1 Nearly Constant Velocity Scenario 59
4.5.1.2 U-Turn Scenario 60
4.5.2 Performance Metrics 61
4.5.3 Results 62
4.5.3.1 Results for the Smoothing Algorithms 62
4.5.3.2 Results for the Filtering Algorithms 67
4.6 Conclusion 68
5 CONCLUSION 75
5.1 Future Work 76
REFERENCES 79
APPENDICES
A M-PROJECTION INVOLVING PSEUDO-LIKELIHOODS 87
A.1 Minimization with respect to the measurement $\mathbf{y}$ 88
A.2 Minimization with respect to the measurement noise covariance $\mathbf{R}$ . . 88
A.3 Minimization with respect to the measurement matrix $\mathbf{C}$ 90
B DERIVATIONS USED IN THE FACTOR UPDATES IN JMLS 93
B.1 Derivations for the Forward Factor, $\rho_n^f(\mathbf{x}_n, r_n)$ 93
B.2 Derivations for the Backward Factor $q_n^b(\mathbf{x}_n, r_n)$ 95
C DERIVATIONS USED IN THE FACTOR UPDATES FOR TARGET TRACKING UNDER MEASUREMENT ORIGIN UNCERTAINTY 97
C.1 Update of the Forward Factor $q_n^f(\mathbf{X}_n, \mathbf{r}_n)$ 97
C.2 Update of the Backward Factor $q_n^b(\mathbf{X}_n)$ 99

D GAUSSIAN IDENTITIES 103

D.1 Gaussian Algebra 103

D.2 Moment Matching between a Mixture of Scaled Gaussians and a Scaled Gaussian 104

E ABEL'S THEOREM [1] 105

F PROOF OF (3.9) AND (3.10) 107

G PROOF OF (A.17) 109

H PROOF OF (A.19) 111

CURRICULUM VITAE 113

LIST OF TABLES

TABLES

Table 3.1 The mode accuracies of the smoothers. 39

Table 3.2 The mode accuracies of the filters. 40

Table 4.1 The simulation parameters of the tests 59

Table 4.2 The simulation parameters of the U-Turn Test 61

Table 4.3 The smoothed mean OSPA and mean errors for Test-1. 64

Table 4.4 The smoothed convergence score and mean errors for Test-1 after the divergent tracks are left out. 64

Table 4.5 The smoothed mean OSPA and mean errors for Test-2. 65

Table 4.6 The smoothed convergence score and mean errors for Test-2 after the divergent tracks are left out. 65

Table 4.7 The smoothed mean OSPA and mean errors for Test-3. 66

Table 4.8 The smoothed convergence score and mean errors for Test-3 after the divergent tracks are left out. 66

Table 4.9 The smoothed mean OSPA and mean errors for U-Turn Test. 67

Table 4.10 The smoothed divergence score and mean errors for U-Turn Test after the divergent tracks are left out. 67

Table 4.11 The filtered mean OSPA and mean errors for Test-1. 68

Table 4.12 The filtered convergence score and mean errors for Test-1. 69

Table 4.13 The filtered mean OSPA and mean errors for Test-2. 69

Table 4.14 The filtered convergence score and mean errors for Test-2. 69

Table 4.15 The filtered mean OSPA and mean errors for Test-3. 70

Table 4.16 The filtered convergence score and mean errors for Test-3. 70

Table 4.17 The filtered mean OSPA and mean errors for U-Turn Test. 70

Table 4.18 The filtered convergence score and mean errors for U-Turn Test. . . 71

LIST OF FIGURES

FIGURES

Figure 1.1 A simple hybrid model adopted from [2]. 2

Figure 3.1 Factor graph for the exact model. 20

Figure 3.2 The approximate factor graph. 21

Figure 3.3 Filtering with EP when the window size is 4. 34

Figure 3.4 Ten sample trajectories the target can follow in the model-mismatch scenario. The trajectories correspond to 10 turn-rates uniformly spaced in the interval $[-\pi/50, \pi/50]$ rad/s. The magenta dot shows the start of the target trajectories. 36

Figure 3.5 The smoother errors for different scenarios. The figures on the left and right indicate the RMS and the median error curves, respectively. 41

Figure 3.6 The filter errors for different scenarios. The figures on the left and right indicate the RMS and the median error curves, respectively. . . 42

Figure 4.1 Factor graph for the exact model. 48

Figure 4.2 The approximate factor graph for the data association problem . 49

Figure 4.3 The trajectories of the U-Turn Test. The big circles indicate the starting points of the trajectories. 60

Figure 4.4 Example realizations from the tests for each smoother. The ones on the left and right are examples from Test-3 when $K = 4$ and U-Turn Test, respectively. The measurements are shown with dots whose shading gets darker as time progresses. 72

Figure 4.5 Example realizations from the tests for each filter. The ones on the left and right are examples from Test-3 when $K = 4$ and U-Turn Test, respectively. The measurements are shown with dots whose shading gets darker as time progresses. 73

LIST OF ABBREVIATIONS

2-D	Two-dimensional Euclidean space
EC	expectation correction
EM	expectation maximization
EP	expectation propagation
EPwCA	expectation propagation with context adjustment
GNN	global nearest neighbor
GPB	generalized pseudo-Bayesian
IMM	interacting multiple model
JMS	jump Markov system
JMLS	jump Markov linear system
JPDA	joint probabilistic data association
KL	Kullback-Leibler
LBP	loopy belief propagation
M-projection	moment projection
MTT	multi-target tracking
OSPA	optimal sub-pattern assignment metric
PDA	probabilistic data association
PMed.	positional median
PMHT	probabilistic multi-hypothesis tracking
PRMS	positional root-mean-square
RTS	Rauch-Tung-Striebel
RMS	root-mean square
TSS	track switch score
VMed.	velocity median

VMP	variational message passing
VRMS	velocity root-mean-square

CHAPTER 1

INTRODUCTION

With the ever-increasing computational capabilities and the developments in sensor technology, Bayesian state estimation [3] is an indispensable part of engineering problems involving dynamical systems. It has been widely applied and deeply studied in various areas such as target tracking [4] [5], robotics [6] [7], biology [8], econometrics [9] [10], machine learning [11] [12], and so forth.

Despite the advances in this field, some problems still remain open for further research due to their complex nature. One class of such problems is inference in hybrid systems, that is, the systems containing discrete-valued hidden random values in addition to a continuous base variable. This type of random variable is frequently incorporated into probabilistic models to describe the phenomena that directly or indirectly affect the observations, such as switching system or measurement dynamics, model outliers [13], and fault diagnosis [14]. Though the hybrid probabilistic representation of a problem can be handy, finding the optimal solution is an $\mathcal{NP}$-hard problem [2].

Consider the following simple example of a hybrid probabilistic model in Fig. 1.1 adopted from [2] where $r_1, \ldots, r_N$ are discrete variables and $x_1, \ldots, x_N$ are the continuous base states to be estimated. Assuming that the base states are conditionally-Gaussian and the discrete-valued variables are binary, we are interested in the following marginalized density.

$$p(x_n) = \sum_{\{r_j\}_{j=1}^{N}} \int_{\substack{\{x_i\}_{i=1}^{N} \\ i \neq n}} p(x_1, \ldots, x_N, r_1, \ldots, r_N) \tag{1.1}$$

where $n \in \{1, \ldots, N\}$. When $n = 1$, the density is written as

$$p(x_1) = \sum_{r_1} p(x_1 \mid r_1) p(r_1), \tag{1.2}$$

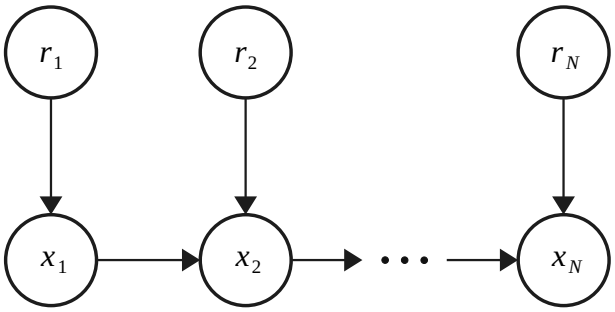


Figure 1.1: A simple hybrid model adopted from [2].

where we make use of the chain rule. Even in this simple case, the marginalized density turns into a mixture of Gaussians with two modes. Taking it one step further and writing the distribution for $n = 2$ gives

$$p(x_2) = \sum_{r_2} p(r_2) \int_{x_1} p(x_2 \mid r_2, x_1) \sum_{r_1} p(x_1 \mid r_1) p(r_1), \tag{1.3}$$

which yields a mixture of Gaussians with 2^2 components. If one repeats the same procedure up to time n, the marginalized density for x_n, i.e., $p(x_n)$, will be a Gaussian mixture with 2^n number of modes. Each of these modes corresponds to a hypothesis, essentially a description of one possible history for the discrete state, r_n. Thus, the time and space required to compute the estimates grow exponentially, indicating that the optimal solution for state estimation in hybrid systems is intractable. There exist various methods in the literature to maintain the number of hypotheses at a manageable level; however, they give sub-optimal solutions. The pruning method discards the hypotheses/mixture modes with low probability to reduce the number of stored hypotheses, whereas merging/collapsing approximates the mixture of Gaussians with a lower number of components. In the probabilistic graphical modeling framework, one way to reduce the computational burden is to remove some of the links between the nodes, which implies eliminating some of the conditional dependencies in the model [11]. Message passing algorithms, such as loopy belief propagation [15], expectation propagation(EP) [16] are applied on these approximate graphs to obtain approximate solutions to inference in hybrid networks.

In this thesis, we will resort to EP to achieve better approximate solutions to some state estimation problems having discrete-valued random variables. To this end, we

will propose an improved version of EP that resolves EP's numerical stability issues arising in implementations. Our approach is twofold. Firstly, we will introduce an additional parameter to EP that controls the information flow in the probabilistic graph. We will name this new version of EP as expectation propagation with context adjustment (EPwCA). Secondly, we will make use of the solution for the M-projection (moment projection) problem that approximates a Gaussian density with a pseudo-likelihood function. To the best of our knowledge, this problem has not been solved before; using its results, we will suggest a new form of factor in the EP factor graphs that reflect the exact probabilistic model more truthfully.

Combining the new EP approach with the better approximated probabilistic models, we will examine two different state estimation problems with discrete-valued hidden random variables in their models in this study. We first tackle the state estimation problem in the jump-Markov linear systems (JMLSs), also known as switching linear dynamical systems, that are widely used to model the changes in the behavior of the systems. Such changes may be due to operational reasons, or they can stem from system failures or environmental disturbances. The discrete-valued random variable is included in the probabilistic models to represent the mode of operation. They have been used in various fields for example econometrics [17], classification of human motion [18], music transcription [19], and maneuvering target tracking [4]. We will derive a smoother and a filter for JMLSs in the framework of EP. The second problem we study in this thesis is the state estimation under measurement origin uncertainty, also known as the data association or correspondence problem in the literature [2]. The discrete random variable is used as an association/correspondence variable that determines the source/origin of the measurements. Some example applications of this problem from various fields are identity resolution and word alignment [20] problems in text processing, image registration in computer vision [2], and genome correspondence problem [21] in biology. In this thesis, we will address the target tracking problem under the measurement origin uncertainty, which has been a long-studied topic in the tracking community [5], [4]. When a set of measurements with no origin information is received at a single scan, their source should be identified before inferring the state. Using the proposed variant of EP, we will derive a fixed-interval smoother and a filter.

1.1 Main Contributions

Our contributions in this thesis are as follows.

- We introduce the concept of context adjustment to overcome the indefinite covariance matrix problem encountered in EP implementations. We name this version of EP as Expectation Propagation with Context Adjustment (EPwCA), and it is first proposed in
  E. Sarıtaş and U. Orguner, "Expectation Propagation with Context Adjustment for Smoothing of Jump Markov Linear Systems," submitted to *IEEE Transactions on Aerospace and Electronic Systems* on May 29, 2022, and revised on January 3, 2023.
- We solve the M-projection problem that involves the pseudo-likelihood function as the approximating function, which is not previously attempted in the literature. Our approach is first presented in the submitted paper, as well.
- Using EPwCA and the results of the M-projection with pseudo-likelihoods, we derive a fixed-interval smoother for state estimation in JMLSs. This work constitutes the main body of the submitted article.
- We extend the work on the state estimation problem for JMLSs using EPwCA to the filtering problem by implementing as a fixed-lag smoother.
- We apply the same approach to the target tracking problem under the measurement origin uncertainty and obtain a novel smoother and a filter.

1.2 Outline of the Thesis

The rest of this thesis is structured as follows. Chapter 2 gives an overview of variational Bayes approaches focusing on EP. The numerical instabilities inherent to the EP algorithm are discussed, and our improvement to circumvent these issues is presented. Additionally, the M-projection problem, which assumes that the approximating functions have the form of pseudo-likelihoods, is solved. This theory is used in the solutions of the following problems. Chapter 3 addresses the smoothing and filtering

problems in JMLSs using the proposed method, namely, EPwCA, and investigates in detail the derived algorithms' performance by comparing them to their alternatives in different scenarios. Chapter 4 tackles the target tracking problem under the measurement origin uncertainty in the framework of EPwCA. A fixed-interval smoother and a filter are derived in this chapter. A discussion on the benefits and shortcomings of the proposed techniques together with a performance comparison with the existing methods are presented as well. Chapter 5 concludes this thesis.

CHAPTER 2

EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT

2.1 Introduction

In Bayesian inference, the purpose is to compute the posterior density for the latent (hidden) variable x given the observation set D. Using the Bayes rule, this posterior can be written as

$$p(x \mid D) = \frac{p(x, D)}{\int p(x, D)dx}. \tag{2.1}$$

The closed-form solution to this problem exists only in limited settings. In fact, the integral in (2.1) cannot be evaluated in many practical problems. High dimensional latent space or intractable analytical expressions may cause this problem. Therefore, it is inevitable to resort to approximate methods.

The approximate methods related to inference problems can be divided into two main groups: sampling-based methods and variational inference methods. Sampling-based techniques are called Monte Carlo methods in literature and rely on the idea of generating random samples in significant amounts from a density to get an approximation of the integral involving it. Though it works well in any shape of density, it brings a computational burden and the assessment of convergence is quite difficult [11]. There exists a vast literature on this approach and it is applied to many different inference problems [22], [23].

The second type of approximation, variational inference, is at the center of interest in this thesis, and we will explore it in more detail in the following sections. This chapter is organized as follows. In Section 2.2, a brief background on variational inference is given. Then, we overview EP and introduce the idea of context adjustment in

Section 2.3. Section 2.4 proposes a novel method for M-projection involving pseudo-likelihoods that is to be used in the solutions to the expectation propagation problems in the following chapters.

2.2 Variational Inference

The variational inference technique is an analytical approximation method built upon the calculus of variations. An approximate joint distribution is determined using tools of this field, which minimizes a divergence measure.

As expressed in (2.1), the marginalization of the joint distribution over the latent variable hinders the computation of the exact posterior. To simplify the problem, the posterior is approximated with another distribution as

$$p(x \mid D) = \frac{p(x, D)}{\int p(x, D)dx} \approx q(x). \tag{2.2}$$

The simplification can be achieved through this variational distribution, $q(x)$, in two ways. First, we can assume a specific form of density for $q(x)$. By restricting the family of densities $q(x)$ can attain, the computation of the integral can be made tractable. Second, we may consider a particularly factorized approximate joint distribution so that a less complex Bayesian network is acquired. This factorization may be in the form that disjoint groups of latent variables constitute factors, which is known as mean-field approximation [11]. This approach allows local treatments by easing the computational burden at the expense of losing the dependence information between some variables.

Once the form of the approximate distribution is determined, its parameters are computed by minimizing a cost function. Thus, one should also decide on the cost aside from the assumptions made. Since the purpose is to make the approximate joint density as close to the true one as possible, the cost will be a functional that is termed as divergence in statistics. Various divergence measures exist in Bayesian inference literature [24], and each works well in certain problems. One of the most commonly used divergences is the reverse (exclusive) Kullback-Leibler (KL) divergence, defined

as

$$\mathrm{KL}(q(x) \parallel p(x)) = \int_x q(x) \log \frac{q(x)}{p(x)} dx. \tag{2.3}$$

When $p(\cdot)$ is a Gaussian mixture and $q(\cdot)$ is selected as a Gaussian density, then minimizing the reverse KL divergence with respect to $q(\cdot)$ will give an optimal $q(\cdot)$ which is close to the high-probability mode of the mixture, $p(\cdot)$ [25]. A different version of KL divergence, called forward (inclusive) KL divergence, is

$$\mathrm{KL}\left(p(x) \parallel q(x)\right) = \int_x p(x) \log \frac{p(x)}{q(x)} dx, \tag{2.4}$$

that is in use in many techniques such as EP [16]. In contrast to the reverse one, given that $p(\cdot)$ is a Gaussian mixture, and $q(\cdot)$ is assumed to be a Gaussian density, the optimal solution to the minimization of the forward KL divergence with respect to $q(\cdot)$ is a Gaussian density that matches the moments (mean and covariance) of $p(\cdot)$. A more general version of these two is called α-divergence, defined by

$$D_\alpha(p(x) \parallel q(x)) = \frac{\int_x \left(\alpha p(x) + (1-\alpha) q(x) - p(x)^\alpha q(x)^{1-\alpha}\right) dx}{\alpha(1-\alpha)}. \tag{2.5}$$

The first two divergences we define also belong to the α-divergence family, where the first case corresponds to $\alpha \to 0$, and in the latter one is for $\alpha \to 1$.

When the reverse KL divergence is utilized, the problem of inferring the posterior density via the variational Bayes approach turns into the following

$$\min \ \mathrm{KL}\left(q(x) \parallel p(x \mid D)\right) = \int q(x) \log \frac{q(x)}{p(x \mid D)} dx, \tag{2.6}$$

where $q(x)$ is factorized as

$$q(x) = \prod_i q_i(x_i), \tag{2.7}$$

and

$$\int q_i(x_i) dx_i = 1 \ \ \forall i. \tag{2.8}$$

Note that $\mathrm{KL}\left(q(x) \parallel p(x \mid D)\right) = 0$, when the approximate distribution, $q(x)$, achieves the true posterior, $p(x \mid D)$. The optimal solution for this problem is given as [11]

$$q_j^*(x_j) = \frac{\exp\left(\mathrm{E}_{i \neq j}\left[\ln p(x, D)\right]\right)}{\int \exp\left(\mathrm{E}_{i \neq j}\left[\ln p(x, D)\right]\right) dx_j}. \tag{2.9}$$

It should be emphasized that the solution for the j^{th} factor depends on the others through the expectation operation. Hence, the solution is attained by iteratively updating each factor.

Calculating the expectation in (2.9) by hand is a tedious process that is not only time-consuming but also prone to errors. Especially in large networks, this requires an exhaustive effort. Winn et al. [26] proposed a message-passing algorithm in the variational Bayes framework to overcome this problem. The technique known as variational message passing (VMP) is based on the fact that the variables affecting the expectation in (2.9) are solely the ones that lie in the Markov blanket [11] of the node j, and the nodes that are not in the blanket contribute only to the constant term. Therefore, local manipulations are made possible through message exchange between the neighboring nodes, i.e., between parent and child nodes. Another approach that employs a message-passing mechanism is expectation propagation, which is explained in detail in the following section.

2.3 Expectation Propagation

Expectation propagation (EP) is an approximate variational inference method that has gathered much attention and applied in various problems [16], [27], e.g., classification [28], regression [29], reinforcement learning [30]; in communication applications, e.g., signal detection [31, 32], turbo equalization [33], phase retrieval [34]; and in system identification [35]. It makes use of the forward KL divergence in contrast to VMP. The usage of this measure has two consequences. The first one is that if the approximate distribution is assumed to be in the exponential family, then the minimization of the forward KL divergence reduces to a simple moment-matching problem. The other outcome of using this divergence is the necessity of taking expectation under $p(x)$ which is the reason we apply variational inference techniques in the first place. The procedure that EP follows to overcome this problem is explained in this section.

In this framework, suppose that we would like to approximate an intractable target

distribution $p(x)$ that can be factorized as follows up to a constant.

$$p(x) \propto \prod_{j=1}^{J} p_j(x) \tag{2.10}$$

EP iteratively approximates $p(x)$ with $q(x)$, which is factorized in the same way as $p(x)$ as follows.

$$q(x) \propto \prod_{j=1}^{J} q_j(x), \tag{2.11}$$

where the factor $q_j(x)$ is the approximation of the jth true factor, $p_j(x)$. When the factor $q_i(x)$ is to be updated, one first obtains the distribution

$$q_{\setminus i}(x) \propto \prod_{\substack{j=1 \\ j \neq i}}^{J} q_j(x) \propto \frac{q(x)}{q_i(x)}. \tag{2.12}$$

The distribution $q_{\setminus i}(x)$ is called the *cavity distribution* [36] which represents the context in which the factor $q_i(x)$ is to be updated. The next step is to calculate the so-called *tilted distribution* [36] $\bar{q}_i(x)$ for the ith factor given as

$$\bar{q}_i(x) \propto p_i(x) q_{\setminus i}(x). \tag{2.13}$$

The ith factor $q_i(x)$ is then found by minimizing the (forward) KL divergence from $q(x)$ to the tilted distribution $\bar{q}_i(x)$ as follows.

$$q_i^{\text{new}}(\cdot) = \arg\min_{q_i(\cdot)} \mathrm{KL}\left(\bar{q}_i(\cdot) \parallel q(\cdot)\right) \tag{2.14}$$

This procedure is repeated by visiting all factors sequentially until convergence.

Although it has been applied successfully in various problems, one drawback of EP is the absence of a guarantee for convergence [37]. There have been suggestions to improve convergence properties in the literature. Damping, being one of these techniques, performs updates using weighted combinations of old and new messages [38], and double-loop algorithm guarantees to converge to a local minimum of a Bethe free energy with a slower convergence rate [39].

2.3.1 Idea of Context Adjustment

The minimization in (2.14) is called *M-projection*, or moment projection, in the literature [40] and amounts to matching the moments, i.e., the sufficient statistics,

for distributions from the exponential family. Specifically, the density $q(x)$ obtained at the end of the minimization is the M-projection of the tilted distribution $\bar{q}_i(x)$. Denoting the M-projection as $\mathcal{P}_{\mathrm{M}}\{\cdot\}$ we can write (2.14) as

$$q_i^{\text{new}}(x) = \frac{\mathcal{P}_{\mathrm{M}}\{\bar{q}_i(x)\}}{q_{\setminus i}(x)}. \tag{2.15}$$

By using the definition of the tilted distribution $\bar{q}_i(x)$ in (2.13) we can obtain

$$q_i^{\text{new}}(x) = \frac{\mathcal{P}_{\mathrm{M}}\{p_i(x)q_{\setminus i}(x)\}}{q_{\setminus i}(x)}. \tag{2.16}$$

The division by the cavity distribution $q_{\setminus i}(x)$ above might cause the approximating factor $q_i(x)$ not to be a proper density function, e.g., $q_i(x)$ might turn out to be a scaled-Gaussian term with an indefinite covariance matrix. Having such covariances is not a problem for inference, per se, as it is only the final approximate distribution $q(x)$, which should be a proper density function. Nevertheless, such factors can sometimes cause numerical instabilities that might result in EP having a poor performance. One such case was reported in [41, p. 180] for inference in JMLSs, which is the dynamic system under consideration in this study.[1] Hence, it is crucial to have a structured way to avoid such problems, which was our primary motivation behind the idea of context adjustment explained below.

In this work, we propose to adjust the effect of the cavity distribution $q_{\setminus i}(x)$ in the calculation of the ith factor $q_i(x)$ as follows.

$$q_i^{\text{new}}(x) = \frac{\mathcal{P}_{\mathrm{M}}\{p_i(x)q_{\setminus i}^{\gamma_i}(x)\}}{q_{\setminus i}^{\gamma_i}(x)}, \tag{2.17}$$

where $0 \leq \gamma_i \leq 1$ is the adjustment exponent. When the exponent is set to unity, i.e., $\gamma_i = 1$, the update rule (2.17) reduces to the standard EP update rule in (2.16). By reducing the exponent γ_i towards zero, one can effectively adjust the effect of the context represented by the cavity distribution in the update of $q_i(x)$. When $\gamma_i = 0$, the context is completely removed and the ith updated factor becomes the M-projection of the true factor $p_i(x)$. If we assume that the true factor $p_i(x)$ is a proper density function, then we can easily see that there exists a specific adjustment factor value $\bar{\gamma}_i$ (which can be a positive value if it exists or can be trivially selected as $\bar{\gamma}_i = 0$) for which the update (2.17) will yield a proper density function $q_i^{\text{new}}(x)$ for $\gamma_i \leq \bar{\gamma}_i$

[1] We had similar observations in our EP implementations in the early stages of the current work.

even when the standard EP update in (2.16) will not. In the following parts of this paper, we will call the usage of the exponents as in (2.17) as *context adjustment* and the corresponding expectation propagation algorithm as EP with Context Adjustment (EPwCA).

2.3.2 Related Ideas in the Literature

The most related idea in the literature to context adjustment is the Power-EP described in [42] where the following update rule is proposed for the ith factor $q_i(x)$.

$$q_i^{\text{new}}(x) = \left(\frac{\mathcal{P}_{\text{M}}\{p_i^{1/\eta_i}(x) q_{\setminus i}(x)\}}{q_{\setminus i}(x)} \right)^{\eta_i}, \tag{2.18}$$

where $\eta_i \in \mathbb{R}\setminus\{0\}$ is the adjustment exponent. At first glance, the update rule of the Power-EP might seem to be the same as the context adjustment update rule (2.17) (with $\eta_i = \gamma_i$). however, this is not the case since

$$\left(\mathcal{P}_{\text{M}}\{p_i^{1/\eta_i}(x) q_{\setminus i}(x)\}\right)^{\eta_i} \neq \mathcal{P}_{\text{M}}\{p_i(x) q_{\setminus i}^{\eta_i}(x)\} \tag{2.19}$$

in general.

Another idea related to the powers of the densities in the context of EP is called *damping* [36,38], which is the name given to slowing down the EP updates to improve convergence properties as follows.

$$q_i^{\text{new}}(x) = \left(q_i^{\text{old}}(x)\right)^{(1-\delta)} \left(\frac{\mathcal{P}_{\text{M}}\{p_i(x) q_{\setminus i}(x)\}}{q_{\setminus i}(x)} \right)^{\delta}, \tag{2.20}$$

where $0 < \delta \leq 1$ is the damping coefficient. Note that damping does not change the fixed point of the standard EP. Hence, EP and EP with damping would provide the same results if convergence is achieved for both algorithms. On the other hand, context adjustment changes the fixed point of the standard EP, as the Power-EP, by sacrificing the context to some extent to improve the numerical and convergence properties. It is worth mentioning here that the standard EP with damping might still suffer from numerical and convergence issues as reported in [41, p. 180] for JMLSs.

2.4 M-Projection with Pseudo-Likelihoods

In this section, we introduce a solution to the M-projection problem, when the approximating density is in the likelihood form. We will make use of the theory developed here when solving the expectation problem in the next chapters.

In Bayesian estimation problems, it is common to have factors in the form of a likelihood $p(\mathbf{y}|\mathbf{x})$ where $\mathbf{y}$ is the measurement and $\mathbf{x}$ is the unknown quantity to be estimated. The likelihood factors, $p(\mathbf{y}|\mathbf{x})$ do not have to be normalizable and hence they are not proper densities (with respect to $\mathbf{x}$). In this section, we consider M-projection when the approximating density $q(\mathbf{x})$ contains a likelihood of the form

$$p(y|x) = \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{x}, \mathbf{R}) \tag{2.21}$$

where $\mathbf{x} \in \mathbb{R}^{d_x}$ is the unknown quantity to be estimated, $\mathbf{y} \in \mathbb{R}^l$ is the measurement where $l \leq d_x$, $\mathbf{C} \in \mathbb{R}^{l \times d_x}$ is the measurement matrix which is assumed to be full row-rank, and $\mathbf{R} \in \mathbb{R}^{l \times l}$ is the positive-definite measurement noise covariance matrix. the notation $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P})$ denotes a Gaussian distribution of the random variable $\mathbf{x}$ with mean $\hat{\mathbf{x}}$ and covariance $\mathbf{P}$.

To the best of our knowledge, the M-projection problem involving such factors has not been investigated in the literature before. The M-projection problem we consider is given as

$$\{\mathbf{y}^*, \mathbf{C}^*, \mathbf{R}^*\} = \arg \min_{\{\mathbf{y}, \mathbf{C}, \mathbf{R}\}} \mathrm{KL}(p(\cdot)||q(\cdot)), \tag{2.22}$$

where $p(\cdot)$ is an arbitrary probability density function with mean $\bar{\mathbf{x}}$ and covariance $\overline{\mathbf{P}}$, and the density $q(\mathbf{x})$ is defined as

$$q(\mathbf{x}) \propto \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{x}, \mathbf{R})\, \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P}), \tag{2.23}$$

where $\mathbf{y} \in \mathbb{R}^\ell$, $\mathbf{C} \in \mathbb{R}^{\ell \times d_x}$ and $\mathbf{R} \in \mathbb{R}^{\ell \times \ell}$ with $\ell \leq d_x$. We also assume that $\mathbf{C}$ is full row-rank and $\mathbf{R}$ is a positive definite matrix. In Appendix A, we show that analytical solutions for $\mathbf{y}$ and $\mathbf{R}$ exist, yet they depend on $\mathbf{C}$, for which we cannot find an analytical solution. Therefore, we seek a sub-optimal solution whose derivation details are presented in Appendix A and obtain the following expressions.

$$\mathbf{C}^* = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_L \end{bmatrix}^{\mathrm{T}}, \tag{2.24a}$$

$$\mathbf{R}^* = \mathrm{diag}\left(\frac{\mathbf{e}_1^{\mathrm{T}}\mathbf{P}\mathbf{e}_1}{\lambda_1 - 1}, \ldots, \frac{\mathbf{e}_L^{\mathrm{T}}\mathbf{P}\mathbf{e}_L}{\lambda_L - 1}\right), \tag{2.24b}$$

$$\mathbf{y}^* = \begin{bmatrix} \frac{\mathbf{e}_1^{\mathrm{T}}(\lambda_1\bar{\mathbf{x}} - \hat{\mathbf{x}})}{\lambda_1 - 1} & \cdots & \frac{\mathbf{e}_L^{\mathrm{T}}(\lambda_L\bar{\mathbf{x}} - \hat{\mathbf{x}})}{\lambda_L - 1} \end{bmatrix}^{\mathrm{T}}, \tag{2.24c}$$

where $\{\mathbf{e}_i, \lambda_i\}$ for $i = 1, \ldots, d_x$ are the eigenvalue-eigenvector pairs that solve the following generalized eigenvalue problem

$$\mathbf{P}\mathbf{e}_i = \lambda_i \overline{\mathbf{P}}\mathbf{e}_i. \tag{2.25}$$

2.5 Summary of the Chapter

This chapter reviews the variational Bayes approach and EP, in particular. We discuss the practical issues in implementing the EP that hinder its application to various problems. We introduce the idea of context adjustment to mitigate the numerical issues of EP. For the first time in literature, we derive the analytical sub-optimal solution to the M-projection that uses likelihood factors. In the following chapters, Chapter 3 and Chapter 4, we will address the state estimation problem in JMLSs and the target tracking problem under measurement origin uncertainty, respectively, by applying EPwCA and integrating the pseudo-likelihoods factors into our solutions using the results of Section 2.4.

CHAPTER 3

EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT FOR SMOOTHING AND FILTERING OF JUMP MARKOV LINEAR SYSTEMS

3.1 Introduction

Bayesian inference for dynamic systems is usually performed in two stages: time/prediction update and measurement update/correction [3] which requires the knowledge of state transition density and measurement likelihood, respectively. In some simple cases, these densities can be derived from the knowledge of state and measurement models. However, in many practical cases, the state and measurement models are either not known, or they are uncertain or they change during the operation of the system, and hence they also need to be estimated along with the state of the system. In such cases, it is a common strategy to deal with this model uncertainty by using a fixed number of alternative state and/or measurement models the system can switch between during its operation. The dynamic systems switching between different modes of operation and hence having switching mathematical models are called *switching dynamical systems* . Since the actual (model) switching times of the system are not known, it is common to mathematically model the system mode as a discrete random process, usually with Markovian behavior, i.e., as a Markov chain. Switching dynamical systems whose mode evolves according to a Markov chain are called *jump Markov systems* (JMSs). In this paper, we are interested in the smoothing of jump Markov linear systems (JMLSs).

The Bayesian state estimation of JMSs involves the consideration of alternative mode histories of the system, which results in posteriors in the form of mixtures. The optimal Bayesian estimator has to consider all possible mode histories of the sys-

tem whose number grows exponentially as the time progresses, leading to posterior mixtures with an intractable number of components to store and process. Consequently, sub-optimal approaches are followed to keep the number of mixture components under control while preserving the estimation performances. Although there exist pruning-based algorithms that reduce mixtures by discarding the mixture components with low weights in the literature [43], most of the popular JMS filters are based on the idea of merging/collapsing mixture components via moment matching. The second-order generalized pseudo-Bayesian (GPB2) [44] and interacting multiple model (IMM) [45] filters are widely-utilized merging and moment-matching based approaches, which keep mixtures with R components as summary statistics at each time step where R is the number of mode states in the JMS. GPB2 and IMM filters execute R^2 and R filters, respectively, at each time step to update their summary statistics. Though the IMM filter can be derived as an approximation of the GPB2 filter, it can perform almost equally well in many practical scenarios [4, p. 466].

The difficulties encountered in the filtering of JMSs mentioned above carry over into the smoothing problem for JMSs. A two-filter formula [46] based fixed interval smoother was proposed for JMSs in [47]. The main disadvantage of this approach is that it requires the invertibility of the JMS in time for the backward filter. As a result, more recent approaches to smoothing of JMSs utilize the Rauch-Tung-Striebel (RTS) formulae [48] to derive the smoothers [49–53], i.e., after forward filtering, they run backward smoothers starting from the last filtered estimates. The smoothers proposed in [49], [50] and [51] run R^2 smoothers at each time step to update R smoothed mixture components in their backward smoothing pass. From the perspective of the number of smoothers run per time step, these studies can be interpreted as GPB2 type smoothers. Among these works, the powerful expectation correction approach proposed by Barber in [51] can keep an arbitrary number of mixture components as the summary statistics at each time step in the backward pass. The most recent works on fixed-interval smoothing of JMSs were presented in [52] and [53] where IMM type smoothers running R smoothers at each time step were proposed. In this work, we propose a fixed interval smoother for jump Markov linear systems based on *expectation propagation* which is an approximate variational inference technique [11], [40].

In the literature, there have been few attempts to solve the inference problem in JMSs

using the EP framework. The approximations on densities employed by EP usually amount to moment-matching [16,27]. Considering that the dominating filters, namely GPB2 [44] and IMM [45] filters, for JMSs both utilize moment-matching as opposed to their, now obsolete, pruning-based alternatives [54, 55], the application of EP to the problem of smoothing for JMSs would constitute a promising and potentially significant line of research. In spite of this, there have been few attempts to solve the inference problem in JMSs using the EP framework in the literature. In [31], Qi and Minka apply EP to a linear system switching according to a white mode sequence among multiple measurement models for solving the signal detection problem in flat-fading channels. Heskes and Zoeter derived an EP smoother for jump Markov linear systems in [39, 56]. However, it has been reported by Barber in [41, p. 180] that the EP smoother in [39,56] suffers from significant numerical and convergence problems.

In this chapter, we propose a fixed-interval smoother for JMLSs by incorporating two major modifications on the standard EP formulation to overcome its numerical and convergence problems. The first modification we propose is the adjustment of the context in which the factors are updated so that one avoids indefinite covariances in EP iterations. The second modification proposed is the change of the backward factors into Gaussian pseudo-likelihoods, for which the problem of M-projection [40] is solved for the first time in the literature to the best of the authors' knowledge.

The rest of the chapter is organized as follows. In Section 3.2, we give the problem definition. In Section 3.3, the proposed EP smoother is derived. We present an extension of the proposed solution to the filtering problem in Section 3.4. Section 3.5 presents numerical results comparing the proposed smoother with the alternative methods on scenarios in which standard EP has numerical and convergence problems. The chapter is concluded in Section 3.6.

3.2 Problem Definition

We consider a JMLS with the base state at time n denoted as $\mathbf{x}_n \in \mathbb{R}^{d_x}$, $n = 0, \ldots, N$. The discrete mode state is shown by r_n, and it takes values in the set $\{1 \ldots, R\}$ where R stands for the number of modes. The mode state follows a homogeneous Markov

chain with the transition probability matrix $\mathbf{\Pi} \in \mathbb{R}^{R\times R} = [\pi_i^j]$ where π_i^j denotes the probability of going from the ith mode to the jth mode. The observation at time n is denoted by $\mathbf{y}_n \in \mathbb{R}^{d_y}$, and $\mathbf{y}_{0:n}$ is the set of all measurements up to and including time n.

The JMLS mathematical model considered in this study is given as follows.

$$\mathbf{x}_n, r_n|\mathbf{x}_{n-1}, r_{n-1} \sim \pi_{r_{n-1}}^{r_n} \mathcal{N}\left(\mathbf{x}_n; \mathbf{A}_n^{r_n}\mathbf{x}_{n-1}, \mathbf{Q}_n^{r_n}\right), \tag{3.1a}$$

$$\mathbf{x}_0, r_0 \sim \pi_{0|-1}^{r_0} \mathcal{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|-1}^{r_0}, \mathbf{\Sigma}_{0|-1}^{r_0}), \tag{3.1b}$$

$$\mathbf{y}_n|\mathbf{x}_n, r_n \sim \mathcal{N}\left(\mathbf{y}_n; \mathbf{C}_n^{r_n}\mathbf{x}_n, \mathbf{R}_n^{r_n}\right), \tag{3.1c}$$

$$r_n|r_{n-1} \sim \pi_{r_{n-1}}^{r_n}, \tag{3.1d}$$

where $\mathbf{A}_n^{r_n} \in \mathbb{R}^{d_x\times d_x}$ is the state transition matrix; $\mathbf{C}_n^{r_n} \in \mathbb{R}^{d_y\times d_x}$ is the measurement matrix; $\mathbf{Q}_n^{r_n} \in \mathbb{R}^{d_x\times d_x}$ and $\mathbf{R}_n^{r_n} \in \mathbb{R}^{d_y\times d_y}$ are the positive-definite covariance matrices of the process noise and the measurement noise, respectively. In general, all of these matrices depend on the mode state r_n. The factor graph corresponding to this model is given in Fig. 3.1 where the base and mode states are placed in a single node at each time step. In this work, our purpose is to derive a numerically stable fixed interval

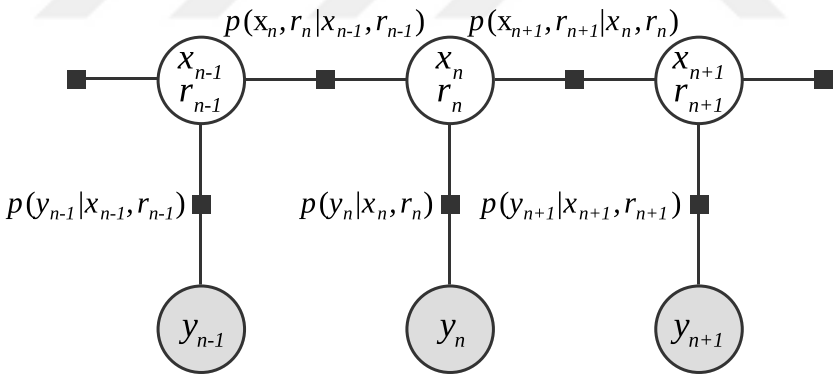


Figure 3.1: Factor graph for the exact model.

smoother based on EP for the aforementioned JMLS model to find the approximate smoothed posterior distributions given as

$$p(\mathbf{x}_n, r_n \mid \mathbf{y}_{0:N}) = \pi_{n|N}^{r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{x}_{n|N}^{r_n}, \mathbf{\Sigma}_{n|N}^{r_n}), \tag{3.2}$$

for $n = 0, \ldots, N$.

3.3 Expectation Propagation with Context Adjustment for Jump Markov Linear Systems

In order to calculate an approximate smoothed posterior in the form of (3.2), we make the following approximation

$$p(\mathbf{x}_n, r_n \mid \mathbf{x}_{n-1}, r_{n-1}) \approx q_n^f(\mathbf{x}_n, r_n) q_{n-1}^b(\mathbf{x}_{n-1}, r_{n-1}), \tag{3.3}$$

where $q_n^f(\mathbf{x}_n, r_n)$ and $q_n^b(\mathbf{x}_n, r_n)$ are the forward and backward factors, respectively. The approximate factor graph is illustrated in Fig. 3.2. The approximate smoothed

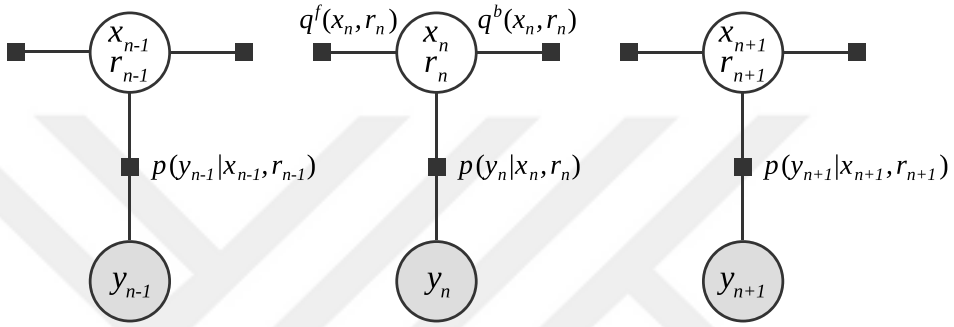


Figure 3.2: The approximate factor graph.

posterior is then given as

$$q_n(\mathbf{x}_n, r_n) \propto q_n^b(\mathbf{x}_n, r_n) p(\mathbf{y}_n \mid \mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n). \tag{3.4}$$

Comparing this factorized approximation with the exact two-filter factorization [46] given as

$$p(\mathbf{x}_n, r_n | \mathbf{y}_{0:N}) \propto p(\mathbf{y}_{n+1:N} | \mathbf{x}_n, r_n) p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n}), \tag{3.5}$$

we can see that the backward factor, $q_n^b(\mathbf{x}_n, r_n)$, approximates the likelihood, $p(\mathbf{y}_{n+1:N} | \mathbf{x}_n, r_n)$, while the remaining terms, $p(\mathbf{y}_n | \mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n)$, approximate the filtered posterior, $p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n})$.

3.3.1 Assumed Forms of the Factors

We assume the following scaled exponential form for the forward factor $q_n^f(\mathbf{x}_n, r_n)$.

$$q_n^f(\mathbf{x}_n, r_n) \triangleq \pi_n^{f,r_n} \exp\left(-\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n}(\cdot) \right), \tag{3.6}$$

where $\boldsymbol{\mu}_n^{f,r_n} \in \mathbb{R}^{d_x}$, $\boldsymbol{\Phi}_n^{f,r_n} \in \mathbb{R}^{d_x \times d_x}$ and we denote long quadratic forms $\mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x}$ as $\mathbf{x}^{\mathrm{T}}\mathbf{P}(\cdot)$ for the sake of brevity. The weights $\pi_n^{f,r_n} \in \mathbb{R}_{\geq 0}$, $r_n = 1, \ldots, R$, are non-negative and they satisfy $\sum_{r_n} \pi_n^{f,r_n} = 1$. We now define the function $\rho_n^f(\mathbf{x}_n, r_n)$ as

$$\rho_n^f(\mathbf{x}_n, r_n) \triangleq \frac{p(\mathbf{y}_n|\mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n)}{\int_{\mathbf{x}_n} \sum_{r_n} p(\mathbf{y}_n|\mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n)}, \tag{3.7}$$

where we assumed that the function $p(\mathbf{y}_n|\mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n)$ can be normalized to a proper density function since it approximates the filtered posterior $p(\mathbf{x}_n, r_n|\mathbf{y}_{0:n})$. A sufficient condition for $p(\mathbf{y}_n|\mathbf{x}_n, r_n) q_n^f(\mathbf{x}_n, r_n)$ to be normalizable is given as

$$\boldsymbol{\Phi}_n^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{C}_n^{r_n} > 0. \tag{3.8}$$

When this condition holds, we can write

$$\rho_n^f(\mathbf{x}_n, r_n) = \alpha_n^{f,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}), \tag{3.9}$$

where

$$\begin{aligned} \alpha_n^{f,r_n} \propto \frac{\pi_n^{f,r_n} \sqrt{|\mathbf{P}_n^{f,r_n}|}}{\sqrt{|\mathbf{R}_n^{r_n}|}} \exp\Bigg(-\frac{1}{2}\Big(& \mathbf{y}_n^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n \\ & + (\boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n} \\ & - (\mathbf{m}_n^{f,r_n})^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1} \mathbf{m}_n^{f,r_n} \Big) \Bigg), \end{aligned} \tag{3.10a}$$

$$\mathbf{m}_n^{f,r_n} = \mathbf{P}_n^{f,r_n} (\boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n), \tag{3.10b}$$

$$\mathbf{P}_n^{f,r_n} = (\boldsymbol{\Phi}_n^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{C}_n^{r_n})^{-1}, \tag{3.10c}$$

with $\sum_{r_n} \alpha_n^{f,r_n} = 1$ (See Appendix F for a proof of (3.9) and (3.10)). It should be clear from the expressions above that the triple $\{\alpha_n^{f,r_n}, \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}\}$ can be uniquely identified from the triple $\{\pi_n^{f,r_n}, \boldsymbol{\mu}_n^{f,r_n}, \boldsymbol{\Phi}_n^{f,r_n}\}$ and vice versa. As a result, the optimization with respect to the factor $q_n^f(\mathbf{x}_n, r_n)$ can be made equivalently with respect to $\rho_n^f(\mathbf{x}_n, r_n)$ instead, which is what we are going to do in the following.

Noting once again that the backward factor $q_n^b(\mathbf{x}_n, r_n)$ approximates the likelihood $p(\mathbf{y}_{n+1:N}|\mathbf{x}_n, r_n)$, the assumed form for $q_n^b(\mathbf{x}_n, r_n)$ will be a scaled Gaussian pseudo-likelihood function given as follows.

$$q_n^b(\mathbf{x}_n, r_n) \triangleq \pi_n^{b,r_n} \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n} \mathbf{x}_n, \mathbf{R}_n^{b,r_n}\right), \tag{3.11}$$

where $\mathbf{y}_n^{b,r_n} \in \mathbb{R}^{\ell_n}$, $\mathbf{C}_n^{b,r_n} \in \mathbb{R}^{\ell_n \times d_x}$, and $\mathbf{R}_n^{b,r_n} \in \mathbb{R}^{\ell_n \times \ell_n}$ are the quantities to be determined by EP. The weights $\pi_n^{b,r_n} \in \mathbb{R}_{\geq 0}$, $r_n = 1, \ldots, R$, are non-negative and they satisfy $\sum_{r_n} \pi_n^{b,r_n} = 1$. Note that although we assume that $\mathbf{R}_n^{b,r_n} > 0$, the factor $q_n^b(\mathbf{x}_n, r_n)$ above does not have to be a proper density function.

Note that we define the weights π_n^{f,r_n}, α_n^{f,r_n} and π_n^{b,r_n} of the factors $q_n^f(\cdot,\cdot)$, $\rho_n^f(\cdot,\cdot)$ and $q_n^b(\cdot,\cdot)$, respectively, such that they all sum to unity (over r_n). Theoretically, such a restriction is actually not necessary since only the ratios of these weights for different r_n carry information. However, when the weights are not normalized over a long horizon in a practical implementation, there appears the risk of numerical issues due to underflow/overflow, which was the reason for us adopting this restriction.

3.3.2 Derivation of the Updates

3.3.2.1 Update of the Forward Factor $\rho_n^f(\mathbf{x}_n, r_n)$

The expectation propagation problem for the forward factor is

$$\rho_n^{f,\text{new}}(\cdot,\cdot) = \arg\min_{\rho_n^f(\cdot,\cdot)} \mathrm{KL}\left(\bar{\psi}_n^f(\cdot,\cdot) \parallel \psi_n^f(\cdot,\cdot)\right), \tag{3.12}$$

where $\bar{\psi}_n^f(\cdot,\cdot)$ and $\psi_n^f(\cdot,\cdot)$ are given as

$$\begin{aligned}\bar{\psi}_n^f(\mathbf{x}_n, r_n) \propto{}& \left[q_n^b(\mathbf{x}_n, r_n)\right]^{\gamma_n^f} p(\mathbf{y}_n \mid \mathbf{x}_n, r_n) \\ &\times \sum_{r_{n-1}} \int_{\mathbf{x}_{n-1}} \Big[p(\mathbf{x}_n, r_n \mid \mathbf{x}_{n-1}, r_{n-1}) \\ &\times \rho_{n-1}^f(\mathbf{x}_{n-1}, r_{n-1})\Big], \end{aligned} \tag{3.13a}$$

$$\psi_n^f(\mathbf{x}_n, r_n) \propto \left[q_n^b(\mathbf{x}_n, r_n)\right]^{\gamma_n^f} \rho_n^f(\mathbf{x}_n, r_n), \tag{3.13b}$$

where the context adjustment is applied on the cavity distribution $q_n^b(\mathbf{x}_n, r_n)$ with the adjustment exponent γ_n^f. Note that the terms on the right-hand side of (3.13a) next to the (adjusted) cavity distribution $q_n^b(\mathbf{x}_n, r_n)$ represent the filtered posterior $p(\mathbf{x}_n, r_n \mid \mathbf{y}_{0:n})$ approximated by replacing $p(\mathbf{x}_{n-1}, r_{n-1} \mid \mathbf{y}_{0:n-1})$ with $\rho_{n-1}^f(\mathbf{x}_{n-1}, r_{n-1})$. After tedious calculations, the densities $\bar{\psi}_n^f(\mathbf{x}_n, r_n)$ and $\psi_n^f(\mathbf{x}_n, r_n)$ turn out to be

$$\bar{\psi}_n^f(\mathbf{x}_n, r_n) \propto \bar{\beta}_n^{f,r_n} \sum_{r_{n-1}} \beta_{n,r_{n-1}}^{f,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_{n,r_{n-1}}^{f,r_n}, \mathbf{V}_{n,r_{n-1}}^{f,r_n}), \tag{3.14a}$$

$$\psi_n^f(\mathbf{x}_n, r_n) \propto \beta_n^{f,r_n} \alpha_n^{f,r_n} \mathcal{N}(\mathbf{x}_n, \mathbf{v}_n^{f,r_n}, \mathbf{V}_n^{f,r_n}), \tag{3.14b}$$

where the parameters of the densities are derived in Appendix B.1. Note that $\mathbf{v}_n^{f,r_n}$ and $\mathbf{V}_n^{f,r_n}$ depend on the decision variables $\mathbf{m}_n^{f,r_n}$ and $\mathbf{P}_n^{f,r_n}$, i.e., the mode-conditioned mean and covariance appearing in $\rho_n^f(\mathbf{x}_n, r_n)$.

Using the result of Appendix D.2, the optimization problem in (3.12) can be solved by setting

$$\alpha_n^{f,r_n} \propto \bar{\beta}_n^{f,r_n} / \beta_n^{f,r_n}, \tag{3.15a}$$

$$\begin{aligned}\mathbf{m}_n^{f,r_n} = &\mathbf{P}_n^{f,r_n} \big((\overline{\mathbf{V}}_n^{f,r_n})^{-1} \bar{\mathbf{v}}_n^{f,r_n} \\ &- \gamma_n^f (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} (\mathbf{R}_n^{b,r_n})^{-1} \mathbf{y}_n^{b,r_n} \big),\end{aligned} \tag{3.15b}$$

$$\mathbf{P}_n^{f,r_n} = \big((\overline{\mathbf{V}}_n^{f,r_n})^{-1} - \gamma_n^f (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} (\mathbf{R}_n^{b,r_n})^{-1} \mathbf{C}_n^{b,r_n} \big)^{-1}, \tag{3.15c}$$

where

$$\bar{\mathbf{v}}_n^{f,r_n} \triangleq \sum_{r_{n-1}} \beta_{n,r_{n-1}}^{f,r_n} \mathbf{v}_{n,r_{n-1}}^{f,r_n}, \tag{3.16a}$$

$$\overline{\mathbf{V}}_n^{f,r_n} \triangleq \sum_{r_{n-1}} \beta_{n,r_{n-1}}^{f,r_n} \big[\mathbf{V}_{n,r_{n-1}}^{f,r_n} + (\mathbf{v}_{n,r_{n-1}}^{f,r_n} - \bar{\mathbf{v}}_n^{f,r_n})(\cdot)^{\mathrm{T}} \big]. \tag{3.16b}$$

3.3.2.2 Update for the Backward Factor $q_n^b(\mathbf{x}_n, r_n)$

The expectation propagation problem for the backward factor is

$$q_n^{b,\text{new}}(\cdot,\cdot) = \underset{q_n^b(\cdot,\cdot)}{\operatorname{argmin}} \, \mathrm{KL}\left(\bar{\psi}_n^b(\cdot,\cdot) \parallel \psi_n^b(\cdot,\cdot) \right), \tag{3.17}$$

where $\bar{\psi}_n^b(\cdot,\cdot)$ and $\psi_n^b(\cdot,\cdot)$ are given as

$$\begin{aligned}\bar{\psi}_n^b(\mathbf{x}_n, r_n) \propto &\left[\rho_n^f(\mathbf{x}_n, r_n) \right]^{\gamma_n^b} \\ &\times \sum_{r_{n+1}} \int_{\mathbf{x}_{n+1}} \Big[p(\mathbf{x}_{n+1}, r_{n+1} \mid \mathbf{x}_n, r_n) \\ &\times p(\mathbf{y}_{n+1} \mid \mathbf{x}_{n+1}, r_{n+1}) q_{n+1}^b(\mathbf{x}_{n+1}, r_{n+1}) \Big],\end{aligned} \tag{3.18a}$$

$$\psi_n^b(\mathbf{x}_n, r_n) \propto \left[\rho_n^f(\mathbf{x}_n, r_n) \right]^{\gamma_n^b} q_n^b(\mathbf{x}_n, r_n), \tag{3.18b}$$

where the context adjustment is applied on the cavity distribution $\rho_n^f(\mathbf{x}_n, r_n)$ with the adjustment exponent γ_n^b. Note that the terms on the right-hand side of (3.18a) next

to the (adjusted) cavity distribution $\rho_n^f(\mathbf{x}_n, r_n)$ represent the likelihood $p(\mathbf{y}_{n+1:N} \mid \mathbf{x}_n, r_n)$ approximated by replacing $p(\mathbf{y}_{n+2:N} \mid \mathbf{x}_{n+1}, r_{n+1})$ with $q_{n+1}^b(x_{n+1}, r_{n+1})$. After tedious calculations, the density $\bar{\psi}_n^b(\mathbf{x}_n, r_n)$ turns out to be

$$\bar{\psi}_n^b(\mathbf{x}_n, r_n) \propto \bar{\beta}_n^{b,r_n} \sum_{r_{n+1}} \beta_{n,r_n}^{b,r_{n+1}} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_{n,r_n}^{b,r_{n+1}}, \mathbf{V}_{n,r_n}^{b,r_{n+1}}), \tag{3.19a}$$

$$\approx \bar{\beta}_n^{b,r_n} \mathcal{N}(\mathbf{x}_n, \bar{\mathbf{v}}_n^{b,r_n}, \overline{\mathbf{V}}_n^{b,r_n}), \tag{3.19b}$$

where the parameters of the components are derived in Appendix B.2. The mean and the covariance of $\bar{\psi}_n^b(\mathbf{x}_n, r_n)$ are given as

$$\bar{\mathbf{v}}_n^{b,r_n} \triangleq \sum_{r_{n+1}} \beta_{n,r_n}^{b,r_{n+1}} \mathbf{v}_{n,r_n}^{b,r_{n+1}}, \tag{3.20a}$$

$$\overline{\mathbf{V}}_n^{b,r_n} \triangleq \sum_{r_{n+1}} \beta_{n,r_n}^{b,r_{n+1}}, \left[\mathbf{V}_{n,r_n}^{b,r_{n+1}} + (\mathbf{v}_{n,r_n}^{b,r_{n+1}} - \bar{\mathbf{v}}_n^{b,r_n})(\cdot)^{\mathrm{T}}\right]. \tag{3.20b}$$

The density $\psi_n^b(\mathbf{x}_n, r_n)$ in (3.18b) is given as follows.

$$\psi_n^b(\mathbf{x}_n, r_n) \propto (\alpha_n^{f,r_n})^{\gamma_n^b} \pi_n^{b,r_n} |\mathbf{P}_n^{f,r_n}|^{\frac{1-\gamma_n^b}{2}}$$

$$\times \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n}\mathbf{x}_n, \mathbf{R}_n^{b,r_n}\right) \mathcal{N}\left(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b}\right) \tag{3.21a}$$

$$\propto \beta_n^{b,r_n} \pi_n^{b,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_n^{b,r_n} \mathbf{V}_n^{b,r_n}) \tag{3.21b}$$

where the parameters are derived in Appendix B.2. In order to obtain the unknown parameters of the pseudo-likelihood, i.e., $\mathbf{y}_n^{b,r_n}$, $\mathbf{C}_n^{b,r_n}$, and $\mathbf{R}_n^{b,r_n}$ for $r_n = 1, \ldots, R$, we use the results of Appendix A, where we solve the problem of M-projection involving pseudo likelihood terms. Note that for the weights β_n^{b,r_n} to have the same units for $r_n = 1, \ldots, R$, the dimensions $\ell_n^{r_n}$ of the pseudo-measurements, $\mathbf{y}_n^{b,r_n}$, should be the same for $r_n = 1, \ldots, R$, i.e., $\ell_n^{r_n} = \ell_n$.

The first step for finding $\mathbf{y}_n^{b,r_n}$, $\mathbf{C}_n^{b,r_n}$, and $\mathbf{R}_n^{b,r_n}$ for $r_n = 1, \ldots, R$, is to solve the following R generalized eigenvalue problems.

$$\frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b} \mathbf{e}_{n,i}^{r_n} = \lambda_{n,i}^{r_n} \overline{\mathbf{V}}_n^{b,r_n} \mathbf{e}_{n,i}^{r_n} \tag{3.22}$$

for $r_n = 1, \ldots, R$, $i = 1, \ldots, d_x$. Let $L_n^{r_n}$ denote the number of generalized eigenvalues satisfying $\lambda_{n,i}^{r_n} > 1$, i.e., $L_n^{r_n} \triangleq \#\big(\{\lambda_{n,i}^{r_n}, i = 1, \ldots, d_x | \lambda_{n,i}^{r_n} > 1\}\big)$, where the operator $\#(\cdot)$ gives the cardinality of the argument set. Set

$$\ell_n \triangleq \min_{r_n} L_n^{r_n}. \tag{3.23}$$

If $\ell_n = 0$, this means that we cannot solve the optimization problem in (3.17) with $\mathbf{R}_n^{b,r_n} > 0$ for $r_n = 1, \ldots, R$. Hence, in this case, we set

$$q_n^b(\mathbf{x}_n, r_n) = 1 \tag{3.24}$$

for $\mathbf{x}_n \in \mathbb{R}^{d_x}$ and $r_n \in \{1, \ldots, R\}$. This effectively means that the smoothed estimates at time n turn into filtered estimates and hence we do not allow information from the future measurements to affect the smoothed estimates before and at the time n. Note that one can try to avoid this situation to some extent by reducing the context adjustment exponent γ_n^b.

If $\ell_n > 0$, we find the ℓ_n generalized eigenvalue-eigenvector pairs, among the ones belonging to generalized eigenvalues satisfying $\lambda_{n,i}^{r_n} > 1$, maximizing either the generalized eigenvalues $\lambda_{n,i}^{r_n}$ or the following score.[1]

$$s_{n,i}^{r_n} \triangleq \log \lambda_{n,i}^{r_n} + \frac{1}{\lambda_{n,i}^{r_n}} + \gamma_n^b \frac{((\mathbf{e}_{n,i}^{r_n})^{\mathrm{T}}(\mathbf{m}_n^{f,r_n} - \bar{\mathbf{v}}_n^{b,r_n}))^2}{(\mathbf{e}_{n,i}^{r_n})^{\mathrm{T}} \mathbf{P}_n^{f,r_n} \mathbf{e}_{n,i}^{r_n}}, \tag{3.25}$$

for $r_n = 1, \ldots, R$. Since the eigenvalues with the corresponding eigenvectors can be ordered arbitrarily, let us denote these eigenvalue-eigenvector pairs as $\{\lambda_{n,i}^{r_n}, \mathbf{e}_{n,i}^{r_n}\}_{i=1}^{\ell_n}$ without loss of generality. We then set the unknowns $\mathbf{y}_n^{b,r_n}$, $\mathbf{C}_n^{b,r_n}$, and $\mathbf{R}_n^{b,r_n}$ as follows.

$$\mathbf{y}_n^{b,r_n} = \begin{bmatrix} \frac{(\mathbf{e}_{n,1}^{r_n})^{\mathrm{T}} \tilde{\mathbf{v}}_{n,1}^{b,r_n}}{\lambda_{n,1}^{r_n} - 1} & \cdots & \frac{(\mathbf{e}_{n,\ell_n}^{r_n})^{\mathrm{T}} \tilde{\mathbf{v}}_{n,\ell_n}^{b,r_n}}{\lambda_{n,\ell_n}^{r_n} - 1} \end{bmatrix}^{\mathrm{T}}, \tag{3.26a}$$

$$\mathbf{C}_n^{b,r_n} = \begin{bmatrix} \mathbf{e}_{n,1}^{r_n} & \mathbf{e}_{n,2}^{r_n} & \cdots & \mathbf{e}_{n,\ell_n}^{r_n} \end{bmatrix}^{\mathrm{T}}, \tag{3.26b}$$

$$\mathbf{R}_n^{b,r_n} = \operatorname{diag}\left(\frac{(\mathbf{e}_{n,1}^{r_n})^{\mathrm{T}} \mathbf{P}_n^{f,r_n} \mathbf{e}_{n,1}^{r_n}}{\gamma_n^b (\lambda_{n,1}^{r_n} - 1)}, \ldots, \frac{(\mathbf{e}_{n,\ell_n}^{r_n})^{\mathrm{T}} \mathbf{P}_n^{f,r_n} \mathbf{e}_{n,\ell_n}^{r_n}}{\gamma_n^b (\lambda_{n,\ell_n}^{r_n} - 1)} \right). \tag{3.26c}$$

for $r_n = 1, \ldots, R$, where

$$\tilde{\mathbf{v}}_{n,i}^{b,r_n} \triangleq \lambda_{n,i}^{r_n} \bar{\mathbf{v}}_n^{b,r_n} - \mathbf{m}_n^{f,r_n}. \tag{3.27}$$

As the final step of the update, we set

$$\pi_n^{b,r_n} \propto \bar{\beta}_n^{b,r_n} / \beta_n^{b,r_n} \tag{3.28}$$

for $r_n = 1, \ldots, R$.

[1] We have not seen a significant distinction between maximizing the generalized eigenvalues $\lambda_{n,i}^{r_n}$ or the score in the simulations conducted in this work. Hence, in the following parts, we only present the results which were obtained by maximizing the generalized eigenvalues.

3.3.2.3 Computing the Final State Estimates

The approximate smoothed distribution $q_n(\mathbf{x}_n, r_n)$ in (3.4) when $n < N$ is given as

$$q_n(\mathbf{x}_n, r_n) = \frac{q_n^b(\mathbf{x}_n, r_n)\rho_n^f(\mathbf{x}_n, r_n)}{\int_{\mathbf{x}_n}\sum_{r_n} q_n^b(\mathbf{x}_n, r_n)\rho_n^f(\mathbf{x}_n, r_n)} \tag{3.29a}$$

$$= \pi_{n|N}^{r_n}\mathcal{N}(\mathbf{x}_n; \mathbf{x}_{n|N}^{r_n}, \mathbf{\Sigma}_{n|N}^{r_n}), \tag{3.29b}$$

where

$$\pi_{n|N}^{r_n} \propto \alpha_n^{f,r_n}\pi_n^{b,r_n}\mathcal{N}(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n}\mathbf{m}_n^{f,r_n}, \mathbf{S}_n^{r_n}), \tag{3.30a}$$

$$\mathbf{x}_{n|N}^{r_n} = \mathbf{\Sigma}_{n|N}^{r_n}\big((\mathbf{P}_n^{f,r_n})^{-1}\mathbf{m}_n^{f,r_n} + (\mathbf{C}_n^{b,r_n})^{\mathrm{T}}(\mathbf{R}_n^{b,r_n})^{-1}\mathbf{y}_n^{b,r_n}\big), \tag{3.30b}$$

$$\mathbf{\Sigma}_{n|N}^{r_n} = \big((\mathbf{P}_n^{f,r_n})^{-1} + (\mathbf{C}_n^{b,r_n})^{\mathrm{T}}(\mathbf{R}_n^{b,r_n})^{-1}\mathbf{C}_n^{b,r_n}\big)^{-1}, \tag{3.30c}$$

for $r_n = 1, \ldots, R$, where

$$\mathbf{S}_n^{r_n} \triangleq \mathbf{C}_n^{b,r_n}\mathbf{P}_n^{f,r_n}(\mathbf{C}_n^{b,r_n})^{\mathrm{T}} + \mathbf{R}_n^{b,r_n}, \tag{3.31}$$

for $r_n = 1, \ldots, R$. Since there is no backward factor at time $n = N$, we have $q_N(\mathbf{x}_N, r_N) = \rho_N^f(\mathbf{x}_N, r_N)$ yielding,

$$\pi_{N|N}^{r_N} \propto \alpha_N^{f,r_N}, \qquad \mathbf{x}_{N|N}^{r_N} = \mathbf{m}_N^{f,r_N}, \qquad \mathbf{\Sigma}_{N|N}^{r_N} = \mathbf{P}_N^{f,r_N}, \tag{3.32}$$

for $r_N = 1, \ldots, R$.

3.3.3 Selection of the Adjustment Exponents

3.3.3.1 Selection of γ_n^f

We propose the selection of the forward adjustment exponent, γ_n^f, in the EP iteration as follows. We initialize all of the exponents at unity, i.e., $\gamma_n^f = 1$, $n = 1, \ldots, N-1$, after the initialization of the factors $\rho_n^f(\cdot,\cdot)$, $n = 1, \ldots, N$. Suppose that, at some iteration of EPwCA, we would like to update the (parameters of the) forward factor $\rho_n^f(\cdot,\cdot)$ as described in Section 3.3.2.1. We then perform the covariance subtraction in (3.15c) to calculate the matrix

$$(\mathbf{P}_n^{f,r_n})^{-1} = (\overline{\mathbf{V}}_n^{f,r_n})^{-1} - \gamma_n^f(\mathbf{C}_n^{b,r_n})^{\mathrm{T}}(\mathbf{R}_n^{b,r_n})^{-1}\mathbf{C}_n^{b,r_n}, \tag{3.33}$$

for all of the modes $r_n \in \{1, \ldots, R\}$. If the matrices $(\mathbf{P}_n^{f,r_n})^{-1}$, $r_n = 1, \ldots, R$, satisfy both of the conditions

$$\min \operatorname{eig}(\mathbf{P}_n^{f,r_n})^{-1} \geq \epsilon_1, \tag{3.34a}$$

$$\min \operatorname{eig} \overline{\mathbf{V}}_n^{f,r_n} (\mathbf{P}_n^{f,r_n})^{-1} \geq \epsilon_2, \tag{3.34b}$$

for $r_n = 1, \ldots, R$, where $\epsilon_1 > 0$ and $\epsilon_2 \geq 0$ are thresholds to be selected, then we keep the value of γ_n^f the same. Otherwise, i.e., if at least one of the matrices $(\mathbf{P}_n^{f,r_n})^{-1}$, $r_n = 1, \ldots, R$, violate either of the conditions (3.34), then we reduce the exponent γ_n^f as follows.

$$\gamma_n^f \leftarrow \max\{0, \gamma_n^f - \Delta\gamma^f\}, \tag{3.35}$$

where $0 < \Delta\gamma^f \leq 1$ is a predefined constant exponent decrement. After this exponent reduction, we repeat the update of $\rho_n^f(\cdot, \cdot)$ for all of the modes $r_n = 1, \ldots, R$. The condition in (3.34a) guarantees the positive definiteness and the invertibility of the matrix $(\mathbf{P}_n^{f,r_n})^{-1}$ and the design threshold ϵ_1 should be selected such that inverse of the matrix $(\mathbf{P}_n^{f,r_n})^{-1}$ can be found without any numerical problems. The condition (3.34b) ensures that the information after the subtraction (represented by $(\mathbf{P}_n^{f,r_n})^{-1}$) remains above a predefined percentage of the initial information (represented by $(\overline{\mathbf{V}}_n^{f,r_n})^{-1}$). Selecting e.g., $\epsilon_2 = 0.01$ ensures that the information after the subtraction is always going to remain above 1% of the initial information (represented by $(\overline{\mathbf{V}}_n^{f,r_n})^{-1}$). Note that selecting $\epsilon_2 = 0$ effectively removes this condition.

The exponent reduction can be applied multiple times if the matrices $(\mathbf{P}_n^{f,r_n})^{-1}$, $r_n = 1, \ldots, R$ continue to violate the conditions (3.34). With the reduction rule above, the minimum value of γ_n^f is zero at which both of the conditions (3.34) are guaranteed to be satisfied for $r_n = 1, \ldots, R$. Note that one can choose the exponent decrement $\Delta\gamma^f$ as a function of discrete-time n if there is a need to do so.

3.3.3.2 Selection of γ_n^b

Note that the update procedure for (the parameters of) the backward factor $q_n^b(\cdot, \cdot)$ described in Section 3.3.2.2 works for any value of $0 < \gamma_n^b \leq 1$. However, it is possible that the dimension ℓ_n of the pseudo-measurement $\mathbf{y}_n^{b,r_n}$, $r_n \in \{1, \ldots, R\}$

turns out to be zero, i.e., no information is transferred from the future measurements to time n, which effectively resets the smoothing process at time n. By reducing the value of the adjustment exponent, γ_n^b, one can try to avoid this problem to some extent. However, setting γ_n^b to a value too close to zero might significantly increase the state covariances' condition number, leading to numerical problems. In this work, we set $\gamma_n^b = 1$, $n = 0, \ldots, N-1$, in all of the simulations.

3.3.4 Pseudo-Code of the Algorithm

We initialize the EPwCA factors as follows. Note that the parameters of the first forward factor $\rho_0^f(\mathbf{x}_0, r_0)$ can be found by updating the initial distribution $p(\mathbf{x}_0, r_0)$ with the measurement $\mathbf{y}_0$ as

$$\alpha_0^{f,r_0} \propto \pi_{0|-1}^{r_0} \mathcal{N}(\mathbf{y}_0; \mathbf{C}_0^{r_0}, \mathbf{S}_0^{r_0}), \tag{3.36a}$$

$$\mathbf{m}_0^{f,r_0} = \mathbf{P}_0^{f,r_0}((\mathbf{\Sigma}_{0|-1}^{r_0})^{-1}\mathbf{x}_{0|-1}^{r_0} + (\mathbf{C}_0^{r_0})^{\mathrm{T}}(\mathbf{R}_0^{r_0})^{-1}\mathbf{y}_0), \tag{3.36b}$$

$$\mathbf{P}_0^{f,r_0} = ((\mathbf{\Sigma}_{0|-1}^{r_0})^{-1} + (\mathbf{C}_0^{r_0})^{\mathrm{T}}(\mathbf{R}_0^{r_0})^{-1}\mathbf{C}_0^{r_0})^{-1}, \tag{3.36c}$$

for $r_0 = 1, \ldots, R$, where

$$\mathbf{S}_0^{r_0} \triangleq \mathbf{C}_0^{r_0}\mathbf{\Sigma}_{0|-1}^{r_0}(\mathbf{C}_0^{r_0})^{\mathrm{T}} + \mathbf{R}_0^{r_0} \tag{3.37}$$

for $r_0 = 1, \ldots, R$. After initializing the first forward factor $\rho_0^f(\mathbf{x}_0, r_0)$ like this, it is possible to initialize the forward factors by making a forward pass to calculate $\rho_n^f(\cdot, \cdot)$, $n = 1, \ldots, N$, after setting $\gamma_n^f = 0$, $n = 1, \ldots, N$. Note that this first forward pass without the effects of the backward factors (since we have $\gamma_n^f = 0$) is equivalent to GPB2 filter [44]. We can then initialize the backward factors by making a backward pass after setting $\gamma_n^b = 1$, $n = 1, \ldots, N$, and $q_N^b(\mathbf{x}_N, r_N) = 1$ for $r_N = 1, \ldots, R$.

After this initialization, the factors of EPwCA can be updated in any order one likes until a convergence criterion is satisfied. We observed in our experiments that consecutive forward and backward passes similar to an RTS smoother [48] would yield satisfactory results. A pseudo-code of the proposed update mechanism can be found in Algorithm 1. The proposed algorithm uses the forward and backward factor updates given in Algorithm 2 and Algorithm 3, respectively, which utilize the corresponding adjustment exponent update mechanisms described in Section 3.3.3. The

iterations in Algorithm 1 stop, i.e., the algorithm is assumed to have converged, if the root-mean-square value (over the smoothing interval) of the differences between the smoothed state estimates in consecutive iterations becomes lower than or equal to the user-defined threshold Γ_{conv} or the number of iterations goes above $J_{\max}$, which is the maximum number of iterations allowed.

In the forward pass of Algorithm 1, we have also added damping after the forward factor update with the damping coefficient δ. Letting the old and the updated nth forward factor statistics be denoted as $\{\bar{\alpha}_n^{f,r_n}, \bar{\mathbf{m}}_n^{f,r_n}, \overline{\mathbf{P}}_n^{f,r_n}\}_{r_n=1}^R$ and $\{\alpha_n^{f,r_n}, \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}\}_{r_n=1}^R$, respectively, the damping procedure described by (2.20) is given as follows.

$$\begin{aligned}
\tilde{\alpha}_n^{f,r_n} \propto & (\bar{\alpha}_n^{f,r_n})^{1-\delta}(\alpha_n^{f,r_n})^{\delta}|\overline{\mathbf{P}}_n^{f,r_n}|^{\frac{\delta}{2}}|\mathbf{P}_n^{f,r_n}|^{\frac{1-\delta}{2}} \\
& \times \mathcal{N}\left(\bar{\mathbf{m}}_n^{f,r_n}; \mathbf{m}_n^{f,r_n}, \frac{\overline{\mathbf{P}}_n^{f,r_n}}{1-\delta} + \frac{\mathbf{P}_n^{f,r_n}}{\delta}\right) && \text{(3.38a)} \\
\widetilde{\mathbf{P}}_n^{f,r_n} = & \big((1-\delta)\overline{\mathbf{P}}_n^{f,r_n})^{-1} + \delta(\mathbf{P}_n^{f,r_n})^{-1}\big)^{-1} && \text{(3.38b)} \\
\tilde{\mathbf{m}}_n^{f,r_n} = & \widetilde{\mathbf{P}}_n^{f,r_n}\big((1-\delta)\overline{\mathbf{P}}_n^{f,r_n})^{-1}\bar{\mathbf{m}}_n^{f,r_n} + \delta(\mathbf{P}_n^{f,r_n})^{-1}\mathbf{m}_n^{f,r_n}\big) && \text{(3.38c)}
\end{aligned}$$

for $r_n = 1, \ldots, R$.

Algorithm 1 Pseudo Code for EPwCA Smoother

Input: $\{\pi_{0|-1}^r, \mathbf{x}_{0|-1}^r, \mathbf{\Sigma}_{0|-1}^r\}_{r=1}^R$; $\{\mathbf{y}_n\}_{n=0}^N$; Γ_{conv}; $J_{\max}$; δ.

Output: $\{\pi_{n|N}^r, \mathbf{x}_{n|N}^r, \mathbf{\Sigma}_{n|N}^r\}_{r=1}^R$ for $n = 1, \ldots, N$.

$\gamma_n^f = 0$, $n = 1, \ldots, N$.
Calculate $\{\alpha_0^{f,r}, \mathbf{m}_0^{f,r}, \mathbf{P}_0^{f,r}\}_{r=1}^R$.
for $n = 1 : N$ **do**
    Run Alg. 2 to get $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^R$.
end for
$\gamma_n^f = 1$, $n = 1, \ldots, N-1$.
$\gamma_n^b = 1$, $n = 0, \ldots, N-1$.
for $n = N-1 : -1 : 1$ **do**
    Run Alg. 3 to get $\{\pi_n^{b,r}, \mathbf{y}_n^{b,r}, \mathbf{C}_n^{b,r}, \mathbf{R}_n^{b,r}\}_{r=1}^R$.
end for
Find the final estimates $\{\pi_{n|N}^r, \mathbf{x}_{n|N}^r, \mathbf{\Sigma}_{n|N}^r\}_{r=1}^R$ for $n = 0, \ldots, N$.
$j = 1$
$e_{\text{conv}} = \infty$

while $e_{\text{conv}} > \Gamma_{\text{conv}}$ & $j \leq J_{\max}$ **do**
  $\mathbf{x}_{n|N}^{r,\text{old}} = \mathbf{x}_{n|N}^{r}$, $r = 1, \ldots, R$; $n = 0, \ldots, N$.
  for $n = 1 : N$ **do**
    if $\delta < 1$ **then** ▷ Damping is applied
      Save $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^{R}$ as $\{\bar{\alpha}_n^{f,r}, \bar{\mathbf{m}}_n^{f,r}, \overline{\mathbf{P}}_n^{f,r}\}_{r=1}^{R}$
    end if
    Run Alg. 2 to get $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^{R}$.
    if $\delta < 1$ **then** ▷ Damping is applied
      Calculate $\{\tilde{\alpha}_n^{f,r}, \tilde{\mathbf{m}}_n^{f,r}, \widetilde{\mathbf{P}}_n^{f,r}\}_{r=1}^{R}$ using (3.38).
      Set $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^{R}$ to $\{\tilde{\alpha}_n^{f,r}, \tilde{\mathbf{m}}_n^{f,r}, \widetilde{\mathbf{P}}_n^{f,r}\}_{r=1}^{R}$
    end if
  end for
  for $n = N - 1 : -1 : 1$ **do**
    Run Alg. 3 to get $\{\pi_n^{b,r}, \mathbf{y}_n^{b,r}, \mathbf{C}_n^{b,r}, \mathbf{R}_n^{b,r}\}_{r=1}^{R}$.
  end for
  Find the final estimates $\{\pi_{n|N}^{r}, \mathbf{x}_{n|N}^{r}, \mathbf{\Sigma}_{n|N}^{r}\}_{r=1}^{R}$ for $n = 0, \ldots, N$.
  $e_{\text{conv}} = \left(\frac{1}{d_x N} \sum_{n=0}^{N} \sum_{r=1}^{R} \|\mathbf{x}_{n|N}^{r,\text{old}} - \mathbf{x}_{n|N}^{r}\|^2\right)^{0.5}$
  $j \leftarrow j + 1$
end while

Algorithm 2 Pseudo Code for Forward Factor Update

Input: γ_n^f, $\Delta\gamma^f$, ϵ_1, ϵ_2.
Input: $\{\alpha_{n-1}^{f,r}, \mathbf{m}_{n-1}^{f,r}, \mathbf{P}_{n-1}^{f,r}\}_{r=1}^{R}$, $\mathbf{y}_n$.
Input: $\{\pi_n^{b,r}, \mathbf{y}_n^{b,r}, \mathbf{C}_n^{b,r}, \mathbf{R}_n^{b,r}\}_{r=1}^{R}$. ▷ If $\gamma_n^f \neq 0$
Output: $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^{R}$,$\gamma_n^f$.

gammaReductionFlag $= 1$
while gammaReductionFlag $= 1$ **do**
  gammaReductionFlag $= 0$
  for $r = 1 : R$ **do**
    Calculate $\bar{\beta}_n^{f,r}, \bar{\mathbf{v}}_n^{f,r}, \overline{\mathbf{V}}_n^{f,r}$.

Calculate $\mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}$.
Calculate unnormalized $\alpha_n^{f,r}$.
if $\min \text{eig } (\mathbf{P}_n^{f,r})^{-1} \leq \epsilon_1$ or
$\min \text{eig } \overline{\mathbf{V}}_n^{f,r} (\mathbf{P}_n^{f,r})^{-1} \leq \epsilon_2$ **then**
$\text{gammaReductionFlag} = 1$
end if
end for
Normalize $\alpha_n^{f,r}$, $r = 1, \ldots, R$.
if $\text{gammaReductionFlag} = 1$ **then**
$\gamma_n^f \leftarrow \max\{0, \gamma_n^f - \Delta\gamma^f\}$
end if
end while

Algorithm 3 Pseudo Code for Backward Factor Update

Input: γ_n^b, ϵ_3.
Input: $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^R$, $\mathbf{y}_{n+1}$.
Input: $\{\pi_{n+1}^{b,r}, \mathbf{y}_{n+1}^{b,r}, \mathbf{C}_{n+1}^{b,r}, \mathbf{R}_{n+1}^{b,r}\}_{r=1}^R$. ▷ If $n \leq N-2$
Output: $\{\pi_n^{b,r}, \mathbf{y}_n^{b,r}, \mathbf{C}_n^{b,r}, \mathbf{R}_n^{b,r}\}_{r=1}^R$.

for $r = 1 : R$ **do**
Calculate $\bar{\beta}_n^{b,r}, \bar{\mathbf{v}}_n^{b,r}, \overline{\mathbf{V}}_n^{b,r}$.
Solve the generalized eigenvalue problem

$$\frac{\mathbf{P}_n^{f,r}}{\gamma_n^b} \mathbf{e}_{n,i}^r = \lambda_{n,i}^r \overline{\mathbf{V}}_n^{b,r} \mathbf{e}_{n,i}^r$$

Order the generalized eigenvalue-eigenvector pairs in decreasing generalized eigenvalue order.
$L_n^r = \#\big(\{\lambda_{n,i}^r, i = 1, \ldots, d_x | \lambda_{n,i}^r > 1 + \epsilon_3\}\big)$
end for
$\ell_n = \min_r L_n^r$
for $r = 1 : R$ **do**
Choose the first ℓ_n generalized eigenvalue-eigenvector pairs for the rth mode.
Calculate $\mathbf{y}_n^{b,r}, \mathbf{C}_n^{b,r}, \mathbf{R}_n^{b,r}$.
Calculate unnormalized $\pi_n^{b,r}$.

12: **end for**

13: Normalize $\pi_n^{b,r}$, $r = 1, \ldots, R$.

3.3.5 Computational Complexity

The computational complexity of EPwCA can be investigated by calculating the computational complexity of its three sub-blocks: forward factor update, backward factor update and smoothed posterior distribution calculation, whose computational complexities are denoted as F, B, and E, respectively. Once we know F, B, and E, the overall computational complexity of EPwCA is $\mathcal{O}(J_{\max} N(F + B + E))$.

The computational complexity of a forward factor update is dominated by R^2 Kalman filters (having the measurement update in information form) with computational complexity of $\mathcal{O}(d_x^3)$, which might have to be executed at most $1/\Delta\gamma^f$ times. Therefore, the computational load of a single forward pass is $F = \mathcal{O}(R^2 d_x^3/\Delta\gamma^f)$. The computational complexity of a backward factor update is dominated by R^2 Kalman smoothers (in information form) with computational complexity of $\mathcal{O}(d_x^3)$. Note that R generalized eigenvalue problems (with Hermitian matrices) and R naive sorting algorithms have the computational complexities $\mathcal{O}(Rd_x^3)$ and $\mathcal{O}(Rd_x^2)$, respectively, which are both below $\mathcal{O}(R^2 d_x^3)$. Hence, we have $B = \mathcal{O}(R^2 d_x^3)$. The computational complexity of calculating a smoothed posterior distribution involves R Kalman filter measurement updates (in information form). As a result, we have $E = \mathcal{O}(Rd_x^3)$.

3.4 Extension to Filtering Problem

We tackle the problem of filtering with EPwCA by implementing a fixed-lag form of the smoother and recording the estimate at the last instant in the moving window as the filtered output.

To clarify this approach, we resort to the example illustrated in Fig. 3.3, where the window size is selected as $\tau = 4$. In order to obtain the filtered estimate at time n, we apply EPwCA on the window that extends from $n - 3$ to n, which is indicated by the blue box. By running EPwCA on this interval, the estimate at time n is regarded

as the filtering output, whereas the estimates corresponding to the remaining instants in the interval are the smoothed ones. Hence, the green circles denote the filtered estimates, and the yellow boxes cover the fixed-lag smoothing outcomes.

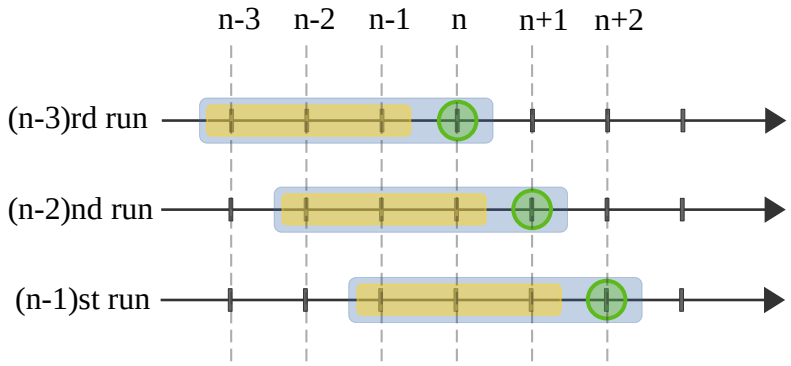


Figure 3.3: Filtering with EP when the window size is 4.

3.4.1 Pseudo Code of the Filtering Algorithm

Algorithm 4 Pseudo Code for EPwCA Filter

Input: $\{\pi^r_{0|-1}, \mathbf{x}^r_{0|-1}, \mathbf{\Sigma}^r_{0|-1}\}_{r=1}^R$; $\{\mathbf{y}_n\}_{n=0}^N$; τ; Γ_{conv}; $J_{\max}$; δ.

Output: $\{\pi^r_{n|n}, \mathbf{x}^r_{n|n}, \mathbf{\Sigma}^r_{n|n}\}_{r=1}^R$ for $n = 1, \ldots, N$.

1: $\gamma^f_n = 0$, $n = 1, \ldots, N$.

2: **for** $n = 1 : N$ **do**

3: Run Alg. 1 using $\{\mathbf{y}_j\}_{j=i}^n$ to get $\{\pi^r_{j|n}, \mathbf{x}^r_{j|n}, \mathbf{\Sigma}^r_{j|n}\}_{r=1}^R$ for $i = \max(1, n - \tau)$.

4: Save $\{\pi^r_{n|n}, \mathbf{x}^r_{n|n}, \mathbf{\Sigma}^r_{n|n}\}_{r=1}^R$ as the filter output.

5: Use $\{\alpha^{f,r}_{i+1}, \mathbf{m}^{f,r}_{i+1}, \mathbf{P}^{f,r}_{i+1}\}_{r=1}^R$ to initialize the Alg. 2 in the next run.

6: **end for**

3.5 Simulation Results

3.5.1 Test Scenarios

The proposed EPwCA smoother and filter are tested on three different scenarios, which are detailed below.

3.5.1.1 Model-Match Scenario

In this scenario we consider a target with the state vector $\mathbf{x}_n \triangleq [p_n^{\mathrm{x}}\ p_n^{\mathrm{y}}\ v_n^{\mathrm{x}}\ v_n^{\mathrm{y}}]^{\mathrm{T}}$, $n = 0, \ldots, 100$, which is composed of 2D position and velocity of the target. The target moves according to a JMLS composed of $R = 2$ nearly constant velocity models with different process noise standard deviations. The JMLS parameters are given as

$$\mathbf{A}^1 = \mathbf{A}^2 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2, \tag{3.39a}$$

$$\mathbf{Q}^r = (\sigma_w^r)^2 \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \otimes \mathbf{I}_2, \quad r = 1, 2, \tag{3.39b}$$

$$\mathbf{C}^1 = \mathbf{C}^2 = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}, \tag{3.39c}$$

$$\mathbf{R}^1 = \mathbf{R}^2 = \sigma_v^2 \mathbf{I}_2, \tag{3.39d}$$

where $T = 1\,\mathrm{s}$, $\sigma_w^1 = 0.1\,\mathrm{m/s}^{1.5}$, $\sigma_w^2 = 10\,\mathrm{m/s}^{1.5}$, $\sigma_v = 10\,\mathrm{m}^2$. The parameters of the initial distribution $p(\mathbf{x}_0, r_0)$ are set to be $\pi_{0|-1}^{r_0} = 0.5$ and

$$\hat{\mathbf{x}}_{0|-1}^{r_0} = [0\ 0\ 100\ 100]^{\mathrm{T}}, \qquad \mathbf{\Sigma}_{0|-1}^{r_0} = 100^2 \mathbf{I}_4, \tag{3.40}$$

for $r_0 = 1, 2$. The transition probability matrix $\mathbf{\Pi}$ is set such that $\pi_1^1 = \pi_2^2 = 0.9$. The smoothing algorithms use the same JMLS model as the one given above. Hence we have a complete model-match between the smoother models and the true target motion, hence the name model-match scenario.

3.5.1.2 Model-Mismatch Scenario

In this scenario our aim is to examine the performance of the proposed smoother on a more realistic set of target trajectories than those used in the model-match case. For this purpose, we consider a target with the same state $\mathbf{x}_n$, $n = 0, \ldots, 150$, as the one in the model-match scenario. The target first moves with constant velocity for 50 seconds, then makes a coordinated turn for 50 seconds with a randomly selected turn-rate and then finishes its trajectory by moving again with constant velocity for 50 seconds. The turn-rate of the target is distributed uniformly over the

[2] The covariance matrix $\mathbf{Q}^r$ in (3.39b) of the process noise is obtained by the discretization of the continuous-time nearly constant velocity model [4, p. 270], [57, eq. (15)]

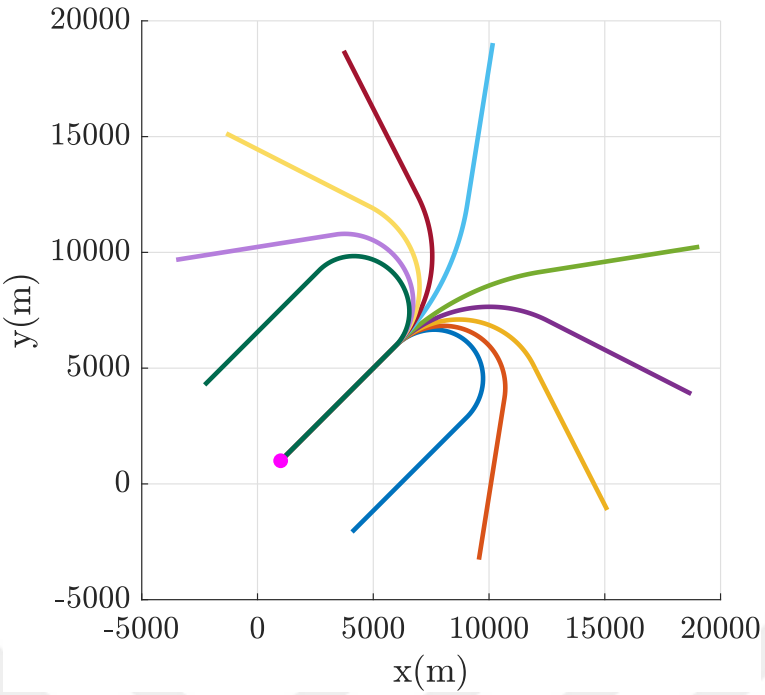


Figure 3.4: Ten sample trajectories the target can follow in the model-mismatch scenario. The trajectories correspond to 10 turn-rates uniformly spaced in the interval $[-\pi/50, \pi/50]$ rad/s. The magenta dot shows the start of the target trajectories.

interval $[-\pi/50, \pi/50]$ rad/s. Ten sample trajectories, corresponding to 10 turn-rates uniformly spaced over the interval $[-\pi/50, \pi/50]$ rad/s, the target can follow are shown in Fig. 3.4. The JMLS model used in the smoothers is the same as the one in the model-match scenario except that the initial mode-conditioned state estimates are selected as $\hat{\mathbf{x}}_{0|-1}^{r_0} = [1000\ 1000\ 100\ 100]^{\mathrm{T}}$ for $r_0 = 1, 2$. As a result, there is a significant mismatch between the true target motion and the motion model (i.e., the parameters $\mathbf{A}^r$, $\mathbf{Q}^r$, $r = 1, 2$) used in the smoothers, hence the name model-mismatch scenario.

3.5.1.3 Barber's Scenario

This scenario is taken from [51], where Barber calls it the hard problem. In this scenario, the state vector $\mathbf{x}_n \in \mathbb{R}^{d_x}$, $n = 0, \ldots, 100$, is observed via scalar observations at each time instant, i.e., $d_y = 1$. There are two modes, i.e., $R = 2$, with the transition

probability matrix $\mathbf{\Pi}$ set such that $\pi_1^1 = \pi_2^2 = 2/3$. [3] The JMLS parameters are given as

$$\mathbf{A}^r = 0.9999 \times \texttt{orth}(\texttt{randn}(d_x)), \qquad \mathbf{Q}^r = \sigma_w^2 \mathbf{I}_{d_x}, \tag{3.41a}$$

$$\mathbf{C}^r = \texttt{randn}(d_y, d_x), \qquad \mathbf{R}^r = \sigma_v^2 \mathbf{I}_{n_y}, \tag{3.41b}$$

for $r = 1, 2$, where $\texttt{randn}(\cdot)$ is the Matlab command creating a matrix of random samples generated from a standard normal distribution (with the desired size), and the Matlab command $\texttt{orth}(\cdot)$ generates an orthonormal basis for the range space of the argument matrix. The parameters in (3.41) are selected as $d_x = 30, \sigma_w = 0.1, \sigma_v = \sqrt{30}$. The parameters of the initial distribution $p(\mathbf{x}_0, r_0)$ are set as $\pi_{0|-1}^{r_0} = 0.5$ and

$$\hat{\mathbf{x}}_{0|-1}^{r_0} = 10 \times \texttt{randn}(d_x, 1), \qquad \mathbf{\Sigma}_{0|-1}^{r_0} = \mathbf{I}_{d_x}, \tag{3.42}$$

for $r_0 = 1, 2$. The smoothers use the same model as the one given above.

3.5.2 Results

3.5.2.1 Results for the Smoothing Algorithms

We implemented the following smoothers for performance comparison:

- **Lopez & Danes [53]:** The smoothing algorithm proposed by Lopez & Danes [53].
- **Barber [41]:** The expectation correction (EC) smoother proposed by Barber in [41, 51]. We present two sets of results of the EC smoother, denoted as Barber-I and Barber-II, obtained when it keeps R (one component for each mode) and $2R$ (2 components for each mode) component mixtures for the smoothed posterior at each time step.
- **Nadarajah et al. [52]:** The smoothing algorithm proposed by Nadarajah et al. [52].

[3] Barber selects $\pi_1^1 = \pi_2^2 = 0.5$ for the hard scenario in [51]. However, such a selection makes the Markov chain white, i.e., the Markov chain turns into a memoryless (i.i.d.) sequence of Bernoulli random variables, which we did not want. Making the selection $\pi_1^1 = \pi_2^2 = 2/3$ instead has been observed to cause only minor changes in the results to be presented while preserving some weak memory.

- **EPwCAS:** The proposed smoother in this work with the parameters $J_{\max} = 50$, $\Gamma_{\text{conv}} = 10^{-4}$, $\epsilon_1 = 10^{-10}$, $\epsilon_2 = 10^{-2}$, $\epsilon_3 = 10^{-6}$, $\Delta\gamma_f = 10^{-1}$. Note that standard EP had numerical and convergence problems in all three scenarios above, and hence we are not able to present its results in this work.

We conducted 500 Monte Carlo runs for each scenario. The performance of the alternative methods is assessed using the root-mean-square (RMS) and median errors. Additionally, we define a new metric which is the ratio of correct decisions made by a smoother to the total number of decisions made. We refer to this ratio as mode accuracy in this study. The mode with the highest mode probability at each time instant is considered to be the mode decided at that instant.

In the results of the model-match scenario depicted in Figs. 3.5a and 3.5b, EPwCAS exhibits the best performance among all smoothers by beating Lopez & Danes by a slight margin. In the results for the model-mismatch scenario shown in Figs. 3.5c and 3.5d, when the target moves with constant velocity, EPwCAS obtains the second-best results after Barber-II. On the other hand, Lopez & Danes has marginally lower RMS and median position errors than EPwCAS when the target maneuvers. Note that the results for all methods, especially those for velocity estimation, are very similar during the maneuver. Barber-I and Nadarajah et al. demonstrate unexpectedly poor performance during the non-maneuvering parts of the scenario. The results for the Barber's scenario shown in Figs. 3.5e and 3.5f reveal the superiority of EPwCAS over Lopez & Danes and Barber-I quite clearly. The RMS errors of EPwCASEPwCA smoother are nearly half of those methods. Since Barber-II keeps two mixture components for each mode, unlike the other smoothers, it can achieve much lower RMS errors than the other methods. Nevertheless, the median errors of EPwCAS are the same as those of Barber-II. Aligned with the error values, the mode accuracies in Table 3.1 suggests that EPwCAS reaches the highest accuracy value among all the smoothers in the first two scenarios and falls behind only Barber-II which stores more mixture components. Overall, EPwCAS can be said to have similar or better performance compared to the smoother, which keeps the same summary statistics in the literature.

Table 3.1: The mode accuracies of the smoothers.

	Model-match	Model-mismatch	Barber's
Lopez & Danes	0.81	0.89	0.67
Barber-I	0.82	0.90	0.82
Barber-II	0.83	0.91	**0.99**
Nadarajah et al.	0.75	0.82	0.54
EPwCAS	**0.85**	**0.92**	0.96

3.5.2.2 Results for the Filtering Algorithms

To compare the filtering performance of EPwCA, we have implemented three alternative methods,

- **GPB2 [44]:** A well-known filter that runs R^2 Kalman filter storing R^2 number of hypotheses before merging.
- **IMM [45]:** It is the de facto filter for JMS. By incorporating a mixing procedure at each cycle, the number of Kalman filters and the number of hypotheses stored are kept at R.
- **Ma et al. [58]:** A filter that approximates the joint distribution of the mode and base state by variational Bayes inference.
- **EPwCAF:** The proposed filter that uses the following parameter set: $J_{\max} = 2$, $\Gamma_{\text{conv}} = 10^{-2}$, $\epsilon_1 = 10^{-10}$, $\epsilon_2 = 10^{-2}$, $\epsilon_3 = 10^{-6}$, $\Delta\gamma_f = 10^{-1}$. The window size of the EPwCAF is set to 3. We apply no damping in model-match and model-mismatch scenarios, while for Barber's scenario, damping is carried out with a damping coefficient of $\delta = 0.5$.

In the model-match scenario, our algorithm performs slightly better than its alternatives in both position and velocity RMS and median errors, as shown in Fig. 3.6a and 3.6b. When the target follows the nearly-constant velocity model in the model-mismatch scenario, the advantage of EPwCA as a filter becomes more distinguishable in Fig. 3.6c and 3.6d. Although the differences between position RMS and velocity

Table 3.2: The mode accuracies of the filters.

	Model-match	Model-Mismatch	Barber's
IMM	0.68	0.83	0.64
GPB2	0.69	**0.85**	0.84
Ma et al.	0.64	0.76	0.56
EPwCAF	**0.75**	**0.85**	**0.91**

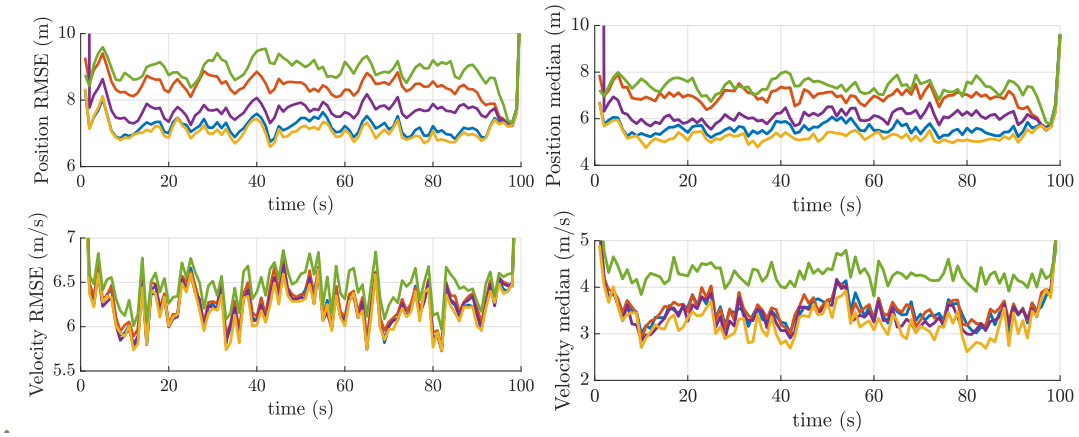
errors of the IMM, GPB2, and EPwCAF are almost similar in the interval of maneuvering, the overall improvement provided by EPwCAF is much greater. On the other hand, though Ma et al. attains the lowest errors in terms of velocity, it falls short in estimating the position. The results for Barber's scenario in Fig. 3.6e and 3.6f draw a similar picture to the smoothing performance of EPwCA, that is, the EPwCAF surpasses the others in the filtering problem, as well. In all scenarios, the errors at the last time instant increase due to the absence of a backward factor, which, in fact, makes the smoothed estimates turn into filtered ones.

Only in the Barber's scenario, EPwCAF appears to perform better than EPwCAS, which is due to errors propagating in the backward direction in the backward pass, especially when an incorrect mode decision is made.

The mode accuracies of the filters are given in Table 3.2 for each scenario. EPwCAF attains the best accuracy values in all cases. Especially in Barber's scenario, it outperforms the others.

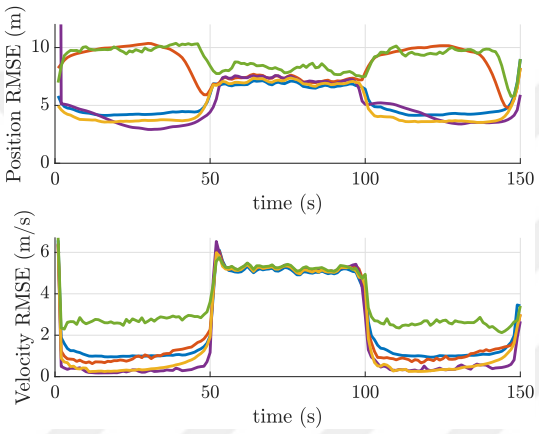
3.6 Conclusion

Fixed-interval smoothing and filtering problems for JMLSs have been studied in an EP framework. Numerical studies have shown that EPwCA and pseudo-likelihood backward factors improve poor numerical and convergence properties of EP for JMLSs reported in the literature. The proposed smoother and the filter demonstrate similar or better performance compared to their alternatives.

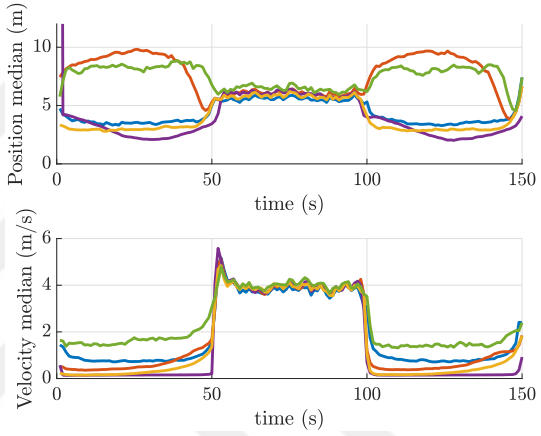


(a) RMS errors for model-match scenario.

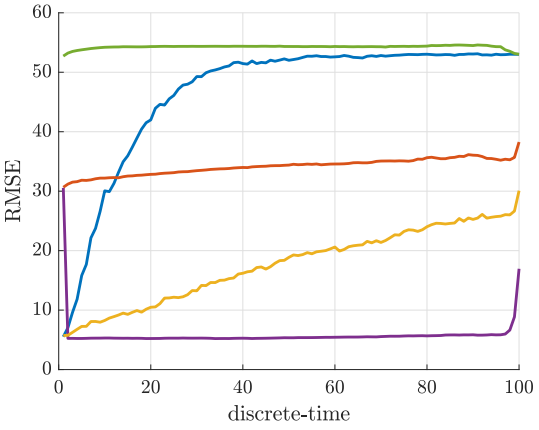
(b) Median errors for model-match scenario.



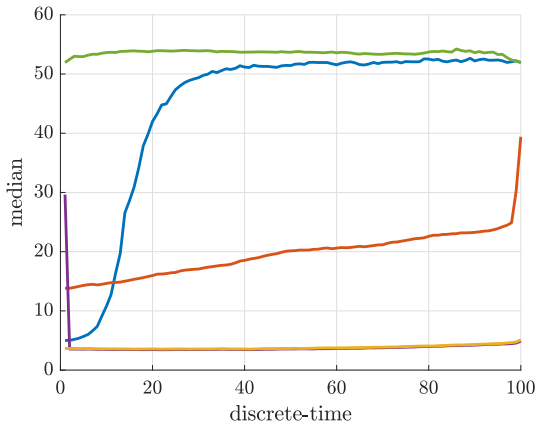
(c) RMS errors for model-mismatch scenario.



(d) Median errors for model-mismatch scenario.



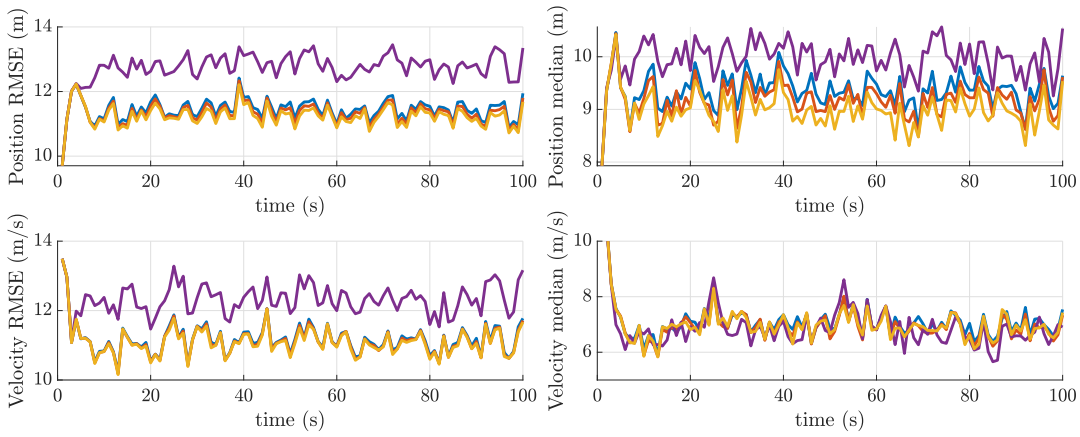
(e) RMS errors for Barber's scenario.



(f) Median errors for Barber's scenario.

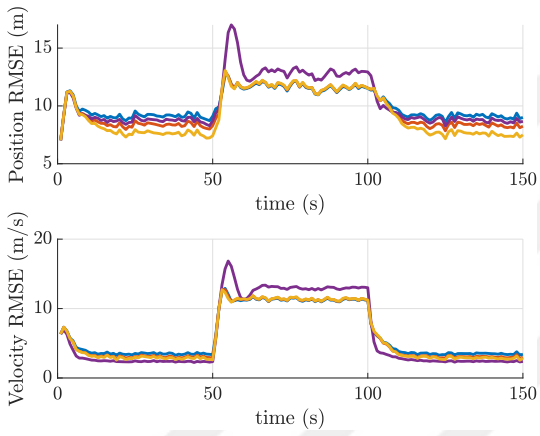


Figure 3.5: The smoother errors for different scenarios. The figures on the left and right indicate the RMS and the median error curves, respectively.

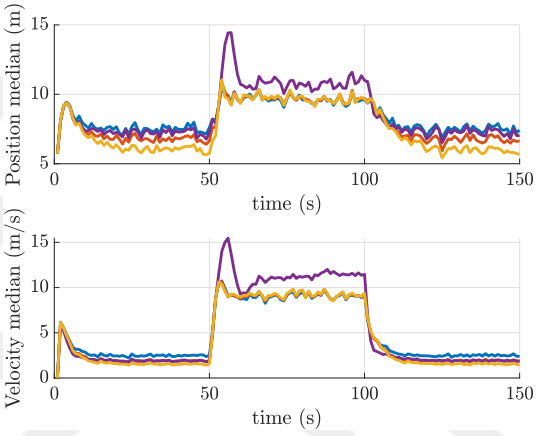


(a) RMS errors for model-match scenario.

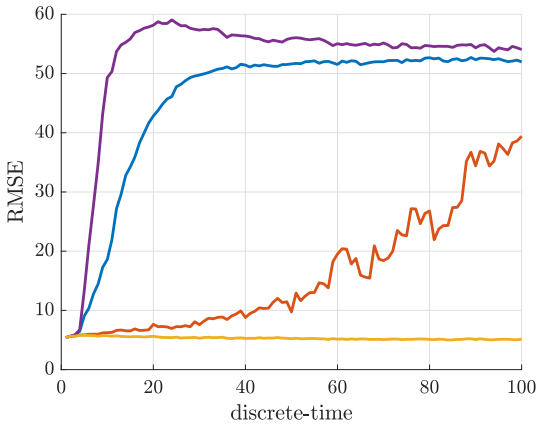
(b) Median errors for model-match scenario.



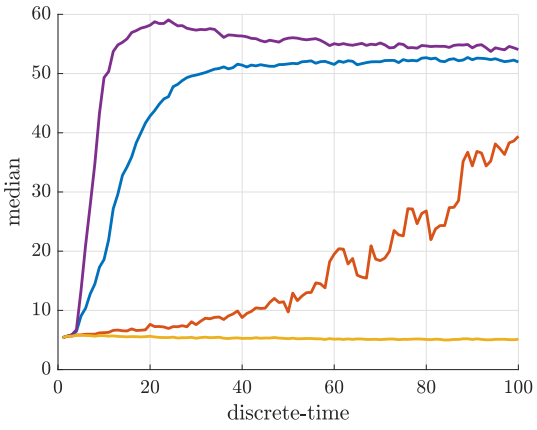
(c) RMS errors for model-mismatch scenario.



(d) Median errors for model-mismatch scenario.



(e) RMS errors for Barber's scenario.



(f) Median errors for Barber's scenario.



Figure 3.6: The filter errors for different scenarios. The figures on the left and right indicate the RMS and the median error curves, respectively.

CHAPTER 4

EXPECTATION PROPAGATION WITH CONTEXT ADJUSTMENT FOR TARGET TRACKING UNDER MEASUREMENT ORIGIN UNCERTAINTY

4.1 Introduction

Target tracking refers to the act of inferring the kinematic properties of targets by processing noisy measurements received from sensors. In real-life problems, mostly there are multiple targets to be tracked, and sensors provide a batch of noisy measurements in a single scan that contain no information about their origin. This measurement batch includes clutter as well, which is spurious measurements due to false alarms and/or reflections from other objects that are out of interest, in addition to the target-originated measurements. It is also highly likely that a measurement belonging to a target is missing due to misdetection.

The target tracking literature [5], [59] is predominantly based on the Bayesian inference methods [3]. A general Bayesian recursion that is to be used to estimate the state of a target would be composed of two stages: prediction update and measurement update. In the latter, the predicted density is updated with one or multiple measurements generated by the target of interest. Thus, to be able to apply Bayesian inference techniques to the target tracking problems, it is crucial to resolve the measurement origin uncertainty. This problem of identifying the source of measurements and associating them to the targets of interest is referred to as the *data association* problem in the literature [5], [59]. In this context, an association hypothesis is an assignment from targets to measurements. In an optimal Bayesian solution, we must consider all possible assignment hypotheses for all time scans. This is a challenging effort because even in a single target tracking case, the number of association hypotheses increases

exponentially in time according to $(M+1)^n$, where M is the number of observations at a scan, and n is the time index. We also note that this number grows even faster in the multi-target case. Therefore, the optimal Bayesian solution for the data association problem is intractable, and we resort to other approaches that yield approximate, sub-optimal solutions.

The data association problem has been examined in depth over the years, and there exist numerous filters in the literature, which handle this problem in different ways. The nearest neighbor approach [5], [4] is a simple method that chooses the measurement closest to the predicted measurement for each target in a probabilistic sense. Therefore, it makes a hard decision. The global nearest neighbor (GNN) [5], [4] being its extension to the multi-target tracking case, selects the hypothesis among all others that yields the highest probability. Another method is the probabilistic data association (PDA) [60] which is a soft decision mechanism. In PDA, there is only one target, and all association hypotheses are weighted out with their probability of association to form an equivalent measurement. Then, the predicted state of the target is updated by using the equivalent measurement. Hence, the PDA algorithm makes use of all the measurements. Its MTT variant, namely, joint probabilistic data association (JPDA) [61], forms the hypotheses considering the existence of other targets in addition to the measurements. A well-known technique applied for MTT is multi-hypotheses tracking (MHT) [62]. With this approach, a subset of hypotheses is kept at each scan, and observations at the following scans are hoped to resolve the data association problem. However, as time progresses, this results in a combinatorial increase in the number of hypotheses. In order to control the growth of this number, a couple of techniques such as pruning, clustering, and merging are applied.

The smoother studies on data association are more limited in the literature. A popular smoothing algorithm is probabilistic multi-hypothesis tracking (PMHT) [63]. In this approach, assuming that the number of objects is fixed and known, associating more than one measurement is allowed to a single target (at each time instant), leading to a soft association. The solution is achieved by making use of expectation maximization (EM) [64] and RTS smoother formulae [65]. As addressed by the authors of the original PMHT paper in [66], this method has the problems of non-adaptivity, narcissism, and hospitality to clutters, in addition to some implementation issues. There are

many versions of PMHT in the literature that aim at alleviating these issues [66] and extend this approach to other problems [67], [68]. In contrast to PMHT, the smoother proposed in [69] abides by the point target model where a target can generate at most one measurement, and each measurement can belong to only one target at a time instant. In this work, Rahmathullah et al. utilize loopy-belief propagation (LBP) [15] and EM to compute association probabilities and the state estimates of the objects, and perform RTS smoother to obtain the smoothed posteriors. As the algorithm suffers from initialization issues, they suggest incorporating deterministic annealing [70] and homotopy methods [71], as well. To our knowledge, there has been no attempt to solve the data association problem using EP up till now. In this thesis, we will derive smoother equations for data association under clutter for MTT based on EP. Specifically, we will use the proposed EPwCA approach that yields a numerically stable algorithm. Moreover, we will implement it as a fixed-lag smoother to achieve filtering results and examine its performance as a filter.

This chapter is structured as follows. The problem definition is given in Section 4.2. Then, we derive the smoother equation of EPwCA smoother in Section 4.3 and describe how it can be utilized as a filter in Section 4.4. In Section 4.5, we illustrate the simulation results, and this chapter is concluded in Section 4.6.

4.2 Problem Definition

We consider a data association problem where the number of objects, denoted by K, is fixed and known[1], i.e., no object spawns or dies throughout the scenario. The base state of the kth object at time n is shown by $\mathbf{x}_n^k \in \mathbb{R}^{d_x}$, whose prior distribution is $p(\mathbf{x}_0^k) = \mathcal{N}(\mathbf{x}_0^k; \mathbf{x}_{0|-1}^k, \mathbf{\Sigma}_{0|-1}^k)$. At each time, we receive a set of measurements whose sources are unknown. This set, defined as $\mathbf{Y}_n = \{\mathbf{y}_n^1, \mathbf{y}_n^2, \ldots, \mathbf{y}_n^{m_n}\}$ at time n where m_n is the number of measurements, is a collection of clutter, false alarm, and possibly the true measurements that originate from the objects of interest. To describe an association hypothesis, we define a discrete association vector, $\mathbf{r}_n \in \mathbb{R}^K$. That is,

[1] In target tracking literature, it is a common practice to assume that the number of objects is fixed and known [59], [5], as estimating it is a challenging problem. Under this assumption, most of the filters employ track maintenance mechanisms so that the number of targets is revealed to the filter. In smoothing problems, this assumption is even more viable, as the data is collected in advance.

$\mathbf{r}_n = [\ r_n^1 \quad r_n^2 \quad \cdots \quad r_n^K\ ]^{\mathrm{T}}$, and when kth object is misdetected, $r_n^k = 0$ and $r_n^k = j$ corresponds to the case of assigning the jth measurement to the kth object. In a single instant n, the set of valid hypotheses is described by the set $\mathcal{R}_n = \{\mathbf{r}_n^1, \ldots, \mathbf{r}_n^{R_n}\}$ where R_n is the cardinality of this set, i.e., it is the number of hypotheses. While forming a valid association vector, we make two assumptions: (1) an object is either detected or misdetected, and (2) a measurement cannot be assigned to more than one object. $\mathbf{y}_n^{r_n^k} \in \mathbb{R}^{d_y}$ stands for the measurement received from the kth object.

We consider the following probabilistic model in this study for the kth target.

$$p(\mathbf{x}_n^k \mid \mathbf{x}_{n-1}^k) = \mathcal{N}\left(\mathbf{x}_n; \mathbf{A}\mathbf{x}_{n-1}^k, \mathbf{Q}\right), \tag{4.1a}$$

$$p(\mathbf{y}_n^{r_n^k} \mid \mathbf{x}_n^k) = \mathcal{N}\big(\mathbf{y}_n^{r_n^k}; \mathbf{C}\mathbf{x}_n^k, \mathbf{R}\big), \tag{4.1b}$$

for $r_n^k \neq 0$ where $\mathbf{A} \in \mathbb{R}^{d_x \times d_x}$ is the state transition matrix; $\mathbf{C} \in \mathbb{R}^{d_y \times d_x}$ is the measurement matrix; $\mathbf{Q} \in \mathbb{R}^{d_x \times d_x}$ and $\mathbf{R} \in \mathbb{R}^{d_y \times d_y}$ are the positive-definite covariance matrices of the process noise and the measurement noise, respectively. The process noise and the measurement noise are independent. We assume that the objects move independently and share the same motion dynamics[2]. Therefore, we write the state transition density in (4.1a) for an augmented base state, $\mathbf{X}_n \in \mathbb{R}^{d_X \times d_X}$, containing the base states of all objects as follows.

$$p(\mathbf{X}_n \mid \mathbf{X}_{n-1}) = \mathcal{N}\left(\mathbf{X}_n; \overline{\mathbf{A}}\mathbf{X}_{n-1}, \overline{\mathbf{Q}}\right), \tag{4.2}$$

where

$$\mathbf{X}_n = [\ (\mathbf{x}_n^1)^{\mathrm{T}} \quad (\mathbf{x}_n^2)^{\mathrm{T}} \quad \cdots \quad (\mathbf{x}_n^K)^{\mathrm{T}}\ ]^{\mathrm{T}}, \tag{4.3a}$$

$$\overline{\mathbf{A}} = \mathrm{I}_K \otimes \mathbf{A}, \tag{4.3b}$$

$$\overline{\mathbf{Q}} = \mathrm{I}_K \otimes \mathbf{Q}, \tag{4.3c}$$

where the sign $\otimes$ denotes the Kronecker product.

A target is detected with probability $P_D \in [0, 1]$. We denote the number of measurements which are assigned to an object by an association hypothesis as by m_n^o, hence it is equal to the number of detected objects, that is $m_n^o \triangleq \#\Big(\left\{r_n^k, 1 \leq k \leq K \mid r_n^k > 0\right\}\Big)$.

[2] One can also assume a time-varying and/or object-dependent probabilistic model, the theory developed in this chapter appertains to those models, as well.

The number of false alarms is Poisson distributed with the probability mass function $P_{FA}(\cdot)$ defined as

$$P_{FA}(m) = \mathcal{PO}(m; \lambda_c) = \frac{(\beta_{FA}V)^m e^{-\beta_{FA}V}}{m!}, \tag{4.4}$$

where λ_c is the average number of false alarms in the surveillance region and it is given as $\lambda_c = \beta_{FA}V$ where β_{FA} is the constant clutter intensity and V is the surveillance region. The spatial probability density function of the false alarms is denoted by $f_c(\cdot)$. In this study, we assume that it is a uniform distribution over the surveillance region, that is,

$$f_c(\mathbf{y}_n) = \frac{1}{V}. \tag{4.5}$$

Given the number of measurements m_n at time n, we can derive the prior probability mass function of an association vector $\mathbf{r}_n$ as

$$\begin{aligned}
p(\mathbf{r}_n \mid m_n) =& P_D^{m_n^o}(1-P_D)^{K-m_n^o} P_{FA}(m_n - m_n^o) \frac{1}{\binom{m_n}{m_n^o} m_n^o!} && \text{(4.6a)}\\
=& P_D^{m_n^0}(1-P_D)^{K-m_n^o} \frac{(\beta_{FA}V)^{m_n-m_n^o} e^{-\beta_{FA}V}}{(m_n-m_n^o)!} \frac{1}{\frac{m_n!}{(m_n-m_n^o)!m_n^o!} m_n^o!} && \text{(4.6b)}\\
\propto& P_D^{m_n^o}(1-P_D)^{K-m_n^o}(\beta_{FA}V)^{m_n-m_n^o} && \text{(4.6c)}\\
=& P_D^{m_n^o}(1-P_D)^{K-m_n^o}(\beta_{FA}V)^{m_n-K+K-m_n^o} && \text{(4.6d)}\\
\propto& P_D^{m_n^o}(1-P_D)^{K-m_n^o}(\beta_{FA}V)^{K-m_n^o}. && \text{(4.6e)}
\end{aligned}$$

The measurement likelihood function of $\mathbf{Y}_n$ given the augmented state vector, the association vector, and the number of measurements is

$$\begin{aligned}
p(\mathbf{Y}_n \mid \mathbf{r}_n, m_n, \mathbf{X}_n) =& \prod_{\substack{0\le k\le K\\ 0<r_n^k}} p(\mathbf{y}_n^{r_n^k} \mid \mathbf{x}_n, r_n^k) \prod_{\substack{1\le j\le m_n\\ 0\le k\le K\\ r_n^k\neq j}} f_c(\mathbf{y}_n^j), && \text{(4.7a)}\\
=& \prod_{\substack{0\le k\le K\\ 0<r_n^k}} p(\mathbf{y}_n^{r_n^k} \mid \mathbf{x}_n, r_n^k) \left(\frac{1}{V}\right)^{m_n-m_n^o}, && \text{(4.7b)}\\
\propto& \left(\frac{1}{V}\right)^{K-m_n^o} \prod_{\substack{0\le k\le K\\ 0<r_n^k}} p(\mathbf{y}_n^{r_n^k} \mid \mathbf{x}_n, r_n^k). && \text{(4.7c)}
\end{aligned}$$

In the rest of this chapter, we omit the number of measurements, m_n, from the conditional densities for the sake of brevity, as it is a measured, and hence known, variable at each time step.

The factor graph corresponding to the exact probabilistic model is given in Fig. 4.1. It is not feasible to compute the optimal smoothed posterior of the augmented state $\mathbf{X}_n$, therefore, we aim at obtaining an approximate solution to the problem using EPwCA.

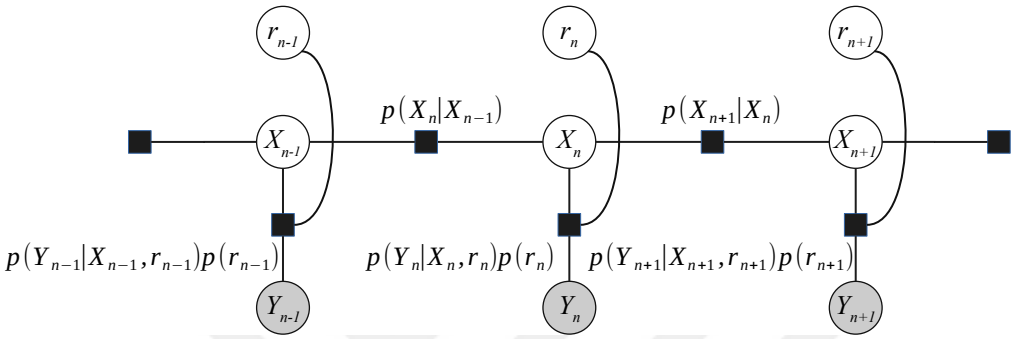


Figure 4.1: Factor graph for the exact model.

In this chapter, we will obtain the approximate smoothed posterior distributions given as

$$p(\mathbf{X}_n \mid \mathbf{Y}_{0:N}) = \mathcal{N}(\mathbf{X}_n; \mathbf{X}_{n|N}, \mathbf{\Sigma}_{n|N}) \tag{4.8}$$

for $n = 0, \ldots, N$, where $\mathbf{Y}_{0:N} = \{\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_N\}$. To this end, we will derive a fixed interval smoother using EPwCA in Section 4.3. Moreover, we will apply our algorithm as a fixed-lag smoother, as described in Section 4.4 to obtain the approximate filtered distributions written as

$$p(\mathbf{X}_n \mid \mathbf{Y}_{0:n}) = \mathcal{N}(\mathbf{X}_n; \mathbf{X}_{n|n}, \mathbf{\Sigma}_{n|n}). \tag{4.9}$$

4.3 Expectation Propagation with Context Adjustment for Data Association Problem in Multi-Target Tracking

We make the following approximation to compute the approximate posterior in (4.8)

$$p(\mathbf{X}_n \mid \mathbf{X}_{n-1}) \approx q_n^f(\mathbf{X}_n) q_{n-1}^b(\mathbf{X}_{n-1}), \tag{4.10}$$

where $q_n^f(\mathbf{X}_n)$ and $q_{n-1}^b(\mathbf{X}_{n-1})$ are the forward and backward factors, respectively. The corresponding approximate factor graph is depicted in Fig. 4.2. Using the approximate factors, we calculate the approximate smoothed posteriors as

$$q(\mathbf{X}_n) \propto \sum_{r_n} q_n^f(\mathbf{X}_n) p(\mathbf{Y}_n \mid \mathbf{X}_n, \mathbf{r}_n) p(\mathbf{r}_n) q_n^b(\mathbf{X}_n). \tag{4.11}$$

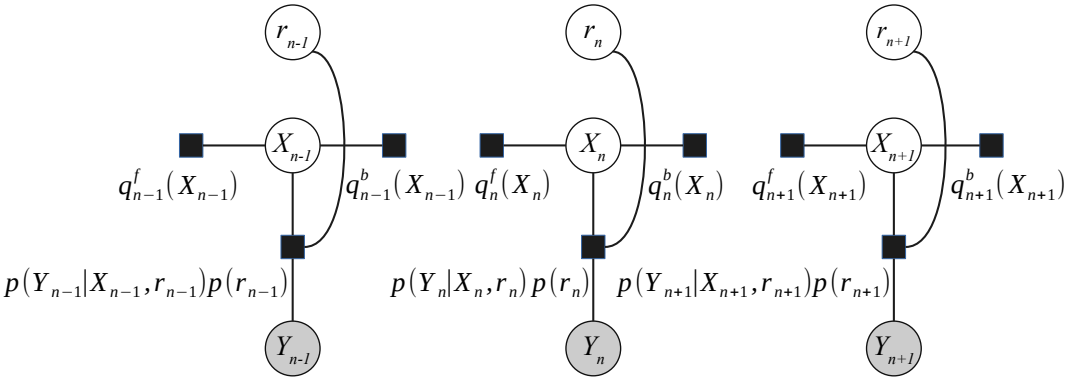


Figure 4.2: The approximate factor graph for the data association problem

Similar to the JMLS problem in Section 3.3, the backward factor approximates the likelihood function involving future measurements, i.e., $p(\mathbf{Y}_{n+1:N} \mid \mathbf{X}_n)$, whereas the filtered posterior is approximated by marginalizing out the remaining terms, i.e., $q_n^f(\mathbf{X}_n)p(\mathbf{Y}_n \mid \mathbf{X}_n, \mathbf{r}_n)p(\mathbf{r}_n)$, with respect to the association vector, $\mathbf{r}_n$.

4.3.1 Assumed Forms of the Factors

We assume that the forward factor, $q_n^f(\mathbf{X}_n)$, has an exponential form

$$q_n^f(\mathbf{X}_n) \triangleq \exp\left(-\frac{1}{2}(\mathbf{X}_n - \boldsymbol{\mu}_n^f)^{\mathrm{T}} \boldsymbol{\Phi}_n^f(\cdot) \right), \tag{4.12}$$

where $\boldsymbol{\mu}_n^f \in \mathbb{R}^{d_X}$, $\boldsymbol{\Phi}_n^f \in \mathbb{R}^{d_X \times d_X}$. We make use of the function $\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)$ that is now defined for the augmented state as follows

$$\rho_n^f(\mathbf{X}_n, \mathbf{r}_n) \triangleq \frac{p(\mathbf{Y}_n|\mathbf{X}_n, \mathbf{r}_n)p(\mathbf{r}_n)q_n^f(\mathbf{X}_n)}{\int_{\mathbf{X}_n} \sum_{\mathbf{r}_n} p(\mathbf{Y}_n|\mathbf{X}_n, \mathbf{r}_n)p(\mathbf{r}_n)q_n^f(\mathbf{X}_n)}, \tag{4.13}$$

where we assume that $p(\mathbf{Y}_n|\mathbf{X}_n, \mathbf{r}_n)p(\mathbf{r}_n)q_n^f(\mathbf{X}_n)$ is normalizable, the sufficient conditions for which are

$$\boldsymbol{\Phi}_n^f + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C} > 0, \tag{4.14a}$$

$$\mathbf{\Phi}_n^f > 0. \tag{4.14b}$$

Under these conditions, we can write

$$\rho_n^f(\mathbf{X}_n, \mathbf{r}_n) = \alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}), \tag{4.15}$$

where

$$\alpha_{n,\mathbf{r}_n}^f \triangleq \prod_K \alpha_{n,\mathbf{r}_n}^{f,k}, \tag{4.16a}$$

$$\alpha_{n,\mathbf{r}_n}^{f,k} \triangleq \begin{cases} (1-P_D)\beta_{FA}, & \mathbf{r}_n^k = 0 \\ \frac{\sqrt{|\mathbf{P}_{n,\mathbf{r}_n}^{f,k}|}}{\sqrt{|\mathbf{R}|}} \exp\Bigg(-\frac{1}{2}\Big((\mathbf{y}_n^{\mathbf{r}_n^k})^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{y}_n^{\mathbf{r}_n^k} \\ \qquad + (\boldsymbol{\mu}_n^{f,k})^{\mathrm{T}} \mathbf{\Phi}_n^{f,k} \boldsymbol{\mu}_n^{f,k}, & \mathbf{r}_n^k \neq 0, \end{cases} \tag{4.16b}$$

$$\mathbf{m}_{n,\mathbf{r}_n}^f \triangleq \left[\begin{array}{cccc} \left(\mathbf{m}_{n,\mathbf{r}_n}^{f,1}\right)^{\mathrm{T}} & \left(\mathbf{m}_{n,\mathbf{r}_n}^{f,2}\right)^{\mathrm{T}} & \cdots & \left(\mathbf{m}_{n,\mathbf{r}_n}^{f,K}\right)^{\mathrm{T}} \end{array} \right]^{\mathrm{T}}, \tag{4.16c}$$

$$\mathbf{m}_{n,\mathbf{r}_n}^{f,k} \triangleq \begin{cases} \boldsymbol{\mu}_n^{f,k}, & \mathbf{r}_n^k = 0 \\ \mathbf{P}_{n,\mathbf{r}_n}^{f,k} \left((\mathbf{\Phi}_n^{f,k})^{-1} \boldsymbol{\mu}_n^{f,k} + \mathbf{C}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{y}_n^{\mathbf{r}_n^k} \right), & \mathbf{r}_n^k \neq 0, \end{cases} \tag{4.16d}$$

$$\mathbf{P}_{n,\mathbf{r}_n}^f \triangleq \text{blkdiag} \left[\begin{array}{cccc} \mathbf{P}_{n,\mathbf{r}_n}^{f,1}, & \mathbf{P}_{n,\mathbf{r}_n}^{f,2}, & \cdots, & \mathbf{P}_{n,\mathbf{r}_n}^{f,K} \end{array} \right], \tag{4.16e}$$

$$\mathbf{P}_{n,\mathbf{r}_n}^{f,k} \triangleq \begin{cases} \mathbf{\Phi}_n^{f,k}, & \mathbf{r}_n^k = 0 \\ \left((\mathbf{\Phi}_n^{f,k})^{-1} + \mathbf{C}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{C} \right)^{-1}, & \mathbf{r}_n^k \neq 0 \end{cases}. \tag{4.16f}$$

Note that we can uniquely determine the parameters $\{\alpha_{n,\mathbf{r}_n}^f, \mathbf{m}_{n,\mathbf{r}_n}^f, \mathbf{P}_{n,\mathbf{r}_n}^f\}$ from $\{\boldsymbol{\mu}_n^f, \mathbf{\Phi}_n^k\}$. Therefore, we use $\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)$ instead of $q_n^f(\mathbf{X}_n)$ in the optimization as we do in the JMLS problem. Similarly, we assume that the backward factor is the following pseudo likelihood function

$$q_n^b(\mathbf{X}_n) \triangleq \mathcal{N}(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \mathbf{R}_n^b), \tag{4.17}$$

where $\mathbf{y}_n^b \in \mathbb{R}^{L_n}$, $\mathbf{C}_n^b \in \mathbb{R}^{L_n \times d_X}$, and $\mathbf{R}_n^b \in \mathbb{R}^{L_n \times L_n}$ are the variables to be found by EPwCA.

4.3.2 Derivation of the Updates

4.3.2.1 Update of the Forward Factor $\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)$

The expectation propagation problem for the forward factor is

$$\rho_n^{new,f}(\cdot,\cdot) = \underset{\rho_n^f(\cdot,\cdot)}{\arg\min}\, \mathrm{KL}\left(\bar{\psi}_n^f(\cdot,\cdot) \parallel \psi_n^f(\cdot,\cdot) \right), \tag{4.18}$$

where the densities $\bar{\psi}_n^f(\cdot,\cdot)$ and $\psi_n^f(\cdot,\cdot)$ are defined as

$$\bar{\psi}_n^f(\mathbf{X}_n,\mathbf{r}_n) \propto \left[q_n^b(\mathbf{X}_n)\right]^{\gamma_n^f} p(\mathbf{Y}_n \mid \mathbf{X}_n,\mathbf{r}_n)p(\mathbf{r}_n) \sum_{\mathbf{r}_{n-1}} \int_{\mathbf{X}_{n-1}} \Big[p(\mathbf{X}_n \mid \mathbf{X}_{n-1}) \times \rho_{n-1}^f(\mathbf{X}_{n-1},\mathbf{r}_{n-1})\Big], \tag{4.19a}$$

$$\psi_n^f(\mathbf{X}_n,\mathbf{r}_n) \propto \left[q_n^b(\mathbf{X}_n)\right]^{\gamma_n^f} \rho_n^f(\mathbf{X}_n,\mathbf{r}_n). \tag{4.19b}$$

Note that we apply the context adjustment on the cavity distribution $q_n^b(\mathbf{X}_n)$ above with the adjustment exponent γ_n^f. Using the results of the derivations in Appendix C.1, we write the densities as

$$\bar{\psi}_n^f(\mathbf{X}_n,\mathbf{r}_n) \propto \bar{\beta}_n^{f,\mathbf{r}_n} \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}), \tag{4.20a}$$

$$\psi_n^f(\mathbf{X}_n,\mathbf{r}_n) \propto \beta_n^{f,\mathbf{r}_n} \alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n, \mathbf{v}_n^{f,\mathbf{r}_n}, \mathbf{V}_n^{f,\mathbf{r}_n}). \tag{4.20b}$$

Then, using the Appendix D.2, the solution of the M-projection problem in (4.18) is given by

$$\alpha_n^{f,\mathbf{r}_n} \propto \bar{\beta}_n^{f,\mathbf{r}_n} / \beta_n^{f,\mathbf{r}_n}, \tag{4.21a}$$

$$\mathbf{m}_n^{f,\mathbf{r}_n} = \mathbf{P}_n^{f,\mathbf{r}_n}\big((\overline{\mathbf{V}}_n^{f,\mathbf{r}_n})^{-1}\bar{\mathbf{v}}_n^{f,\mathbf{r}_n} - \gamma_n^f(\mathbf{C}_n^b)^{\mathrm{T}}(\mathbf{R}_n^b)^{-1}\mathbf{y}_n^b\big), \tag{4.21b}$$

$$\mathbf{P}_n^{f,\mathbf{r}_n} = \big((\overline{\mathbf{V}}_n^{f,\mathbf{r}_n})^{-1} - \gamma_n^f(\mathbf{C}_n^b)^{\mathrm{T}}(\mathbf{R}_n^b)^{-1}\mathbf{C}_n^b\big)^{-1}, \tag{4.21c}$$

where

$$\bar{\mathbf{v}}_n^{f,\mathbf{r}_n} \triangleq \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \tag{4.22a}$$

$$\overline{\mathbf{V}}_n^{f,\mathbf{r}_n} \triangleq \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \big[\mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} + (\mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} - \bar{\mathbf{v}}_n^{f,\mathbf{r}_n})(\cdot)^{\mathrm{T}}\big]. \tag{4.22b}$$

4.3.2.2 Update of the Backward Factor $q_n^b(\mathbf{X}_n)$

The variational inference problem for the backward factor is defined as

$$q_n^{b,new}(\cdot) = \underset{q_n^b(\cdot)}{\operatorname{argmin}} \, \mathrm{KL}\left(\bar{\psi}_n^b(\cdot) \parallel \psi_n^b(\cdot)\right), \tag{4.23}$$

where

$$\bar{\psi}_n^b(\mathbf{X}_n) \propto \sum_{\mathbf{r}_n} \left[\rho_n^f(\mathbf{X}_n,\mathbf{r}_n)\right]^{\gamma_n^b} \sum_{\mathbf{r}_{n+1}} \int_{\mathbf{X}_{n+1}} \Big[p(\mathbf{X}_{n+1} \mid \mathbf{X}_n)p(\mathbf{Y}_{n+1} \mid \mathbf{X}_{n+1},\mathbf{r}_{n+1})$$

$$\times\, q_{n+1}^{b}\left(\mathbf{X}_{n+1}\right) p(\mathbf{r}_{n+1})\Big], \tag{4.24a}$$

$$\psi_{n}^{b}(\mathbf{X}_{n}) \propto q_{n}^{b}(\mathbf{X}_{n}) \sum_{\mathbf{r}_{n}} \left[\rho_{n}^{f}(\mathbf{X}_{n}, \mathbf{r}_{n})\right]^{\gamma_{n}^{b}}, \tag{4.24b}$$

where the context adjustment is applied on $\rho_{n}^{f}(\mathbf{X}_{n}, \mathbf{r}_{n})$ with the context adjustment exponent γ_{n}^{b}. After tedious calculations presented in Appendix C.2, these densities are written as

$$\bar{\psi}_{n}^{b}(\mathbf{X}_{n}) \propto \mathcal{N}(\mathbf{X}_{n}; \bar{\mathbf{v}}_{n}^{b}, \overline{\mathbf{V}}_{n}^{b}), \tag{4.25a}$$

$$\psi_{n}^{b}(\mathbf{X}_{n}) \propto \mathcal{N}(\mathbf{y}_{n}^{b}; \mathbf{C}_{n}^{b}\mathbf{X}_{n}, \mathbf{R}_{n}^{b})\, \mathcal{N}(\mathbf{X}_{n}; \mathbf{v}_{n}^{b}, \mathbf{V}_{n}^{b}). \tag{4.25b}$$

In order to solve the M-projection problem in (4.23) we make use of the results of Appendix A as in Section 3.3.2.2. Then, the generalized eigenvalue problem we solve becomes

$$\frac{\mathbf{V}_{n}^{b}}{\gamma_{n}^{b}}\mathbf{e}_{n,i} = \lambda_{n,i}\overline{\mathbf{V}}_{n}^{b}\mathbf{e}_{n,i} \quad i = 1, \ldots, d_{X}. \tag{4.26}$$

We denote the number of generalized eigenvalues that satisfy $\lambda_{n,i} > 1$ using L_n, that is, $L_n \triangleq \#\big(\{\lambda_{n,i}, i = 1, \ldots, d_X | \lambda_{n,i} > 1\}\big)$. If $L_n = 0$, then we set $q_{n}^{b}(\mathbf{X}_{n}) = 1$ implying that (4.23) has no solution satisfying $\mathbf{R}_{n}^{b} > 0$. In such a case, by setting $q_{n}^{b}(\mathbf{X}_{n}) = 1$, the smoothed estimates at time n are rendered as the filtered estimates. In this a case, one can alternatively reduce the context adjustment exponent to find a solution.

When $L_n > 0$, among the generalized eigenvalue-eigenvectors with $\lambda_{n,i} > 1$, one can either choose the pairs with the maximum eigenvalues or the ones that maximize the following score.

$$s_{n,i} \triangleq \log \lambda_{n,i} + \frac{1}{\lambda_{n,i}} + \gamma_{n}^{b} \frac{(\mathbf{e}_{n,i}^{\mathrm{T}}(\mathbf{v}_{n}^{b} - \bar{\mathbf{v}}_{n}^{b}))^{2}}{\mathbf{e}_{n,i}^{\mathrm{T}}\mathbf{V}_{n}^{b}\mathbf{e}_{n,i}}, \tag{4.27}$$

Once the generalized eigenvalue-eigenvector pairs are selected, the solution is given as

$$\mathbf{y}_{n}^{b} = \Big[\ \tfrac{(\mathbf{e}_{n,1})^{\mathrm{T}}\tilde{\mathbf{v}}_{n,1}^{b}}{\lambda_{n,1}-1} \quad \cdots \quad \tfrac{(\mathbf{e}_{n,L_n})^{\mathrm{T}}\tilde{\mathbf{v}}_{n,L_n}^{b}}{\lambda_{n,L_n}-1}\ \Big]^{\mathrm{T}}, \tag{4.28a}$$

$$\mathbf{C}_{n}^{b} = \Big[\ \mathbf{e}_{n,1} \quad \mathbf{e}_{n,2} \quad \cdots \quad \mathbf{e}_{n,L_n}\ \Big]^{\mathrm{T}}, \tag{4.28b}$$

$$\mathbf{R}_{n}^{b} = \mathrm{diag}\left(\frac{(\mathbf{e}_{n,1})^{\mathrm{T}}\mathbf{V}_{n}^{b}\mathbf{e}_{n,1}}{\lambda_{n,1}-1}, \ldots, \frac{(\mathbf{e}_{n,L_n})^{\mathrm{T}}\mathbf{V}_{n}^{b}\mathbf{e}_{n,L_n}}{\lambda_{n,L_n}-1}\right), \tag{4.28c}$$

where

$$\tilde{\mathbf{v}}_{n,i}^{b} \triangleq \lambda_{n,i}\bar{\mathbf{v}}_{n}^{b} - \mathbf{v}_{n}^{b}. \tag{4.29}$$

4.3.2.3 Computing the Final State Estimates

The approximate smoothed posterior in (4.11) when $n < N$ is computed as

$$\begin{aligned} q_n(\mathbf{X}_n) &= \sum_{\mathbf{r}_n} \frac{q_n^b(\mathbf{X}_n)\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)}{\int_{\mathbf{X}_n}\sum_{\mathbf{r}_n} q_n^b(\mathbf{X}_n)\rho_n^f(\mathbf{X}_n, \mathbf{r}_n)}, && \text{(4.30a)} \\ &= \sum_{\mathbf{r}_n} \pi_{n|N}^{\mathbf{r}_n}\mathcal{N}(\mathbf{X}_n; \mathbf{X}_{n|N}^{\mathbf{r}_n}, \mathbf{\Sigma}_{n|N}^{\mathbf{r}_n}), && \text{(4.30b)} \\ &= \mathcal{N}(\mathbf{X}_n; \mathbf{X}_{n|N}, \mathbf{\Sigma}_{n|N}), && \text{(4.30c)} \end{aligned}$$

where

$$\begin{aligned} \bar{\pi}_{n|N}^{\mathbf{r}_n} &\propto \alpha_n^{f,\mathbf{r}_n}\mathcal{N}(\mathbf{y}_n^b; \mathbf{C}_n^b\mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{S}_n^{\mathbf{r}_n}), && \text{(4.31a)} \\ \pi_{n|N}^{\mathbf{r}_n} &\triangleq \frac{\bar{\pi}_{n|N}^{\mathbf{r}_n}}{\sum_{r_n}\bar{\pi}_{n|N}^{\mathbf{r}_n}}, && \text{(4.31b)} \\ \mathbf{S}_n^{\mathbf{r}_n} &\triangleq \mathbf{C}_n^b\mathbf{P}_n^{f,\mathbf{r}_n}(\mathbf{C}_n^b)^{\mathrm{T}} + \mathbf{R}_n^b, && \text{(4.31c)} \\ \mathbf{X}_{n|N}^{\mathbf{r}_n} &= \mathbf{\Sigma}_{n|N}^{\mathbf{r}_n}\big((\mathbf{P}_n^{f,\mathbf{r}_n})^{-1}\mathbf{m}_n^{f,\mathbf{r}_n} + (\mathbf{C}_n^b)^{\mathrm{T}}(\mathbf{R}_n^b)^{-1}\mathbf{y}_n^b\big), && \text{(4.31d)} \\ \mathbf{\Sigma}_{n|N}^{\mathbf{r}_n} &= \big((\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} + (\mathbf{C}_n^b)^{\mathrm{T}}(\mathbf{R}_n^b)^{-1}\mathbf{C}_n^b\big)^{-1}, && \text{(4.31e)} \\ \mathbf{X}_{n|N} &= \sum_{r_n}\pi_{n|N}^{\mathbf{r}_n}\mathbf{X}_{n|N}^{\mathbf{r}_n}, && \text{(4.31f)} \\ \mathbf{\Sigma}_{n|N} &= \sum_{\mathbf{r}_n}\pi_{n|N}^{\mathbf{r}_n}\big(\mathbf{\Sigma}_{n|N}^{\mathbf{r}_n} + (\mathbf{X}_{n|N}^{\mathbf{r}_n} - \mathbf{X}_{n|N})(\mathbf{X}_{n|N}^{\mathbf{r}_n} - \mathbf{X}_{n|N})^{\mathrm{T}}\big). && \text{(4.31g)} \end{aligned}$$

The smoothed posterior density for kth object is computed as

$$\begin{aligned} p(\mathbf{x}_n^k \mid \mathbf{Y}_{0:N}) &= \mathcal{N}(\mathbf{x}_n^k; \mathbf{x}_{n|N}^k\mathbf{\Sigma}_{n|N}^k), && \text{(4.32a)} \\ &= \int_{\{\mathbf{x}_n^j\}_{\substack{j=1 \\ j\neq k}}^{K}} \mathcal{N}(\mathbf{X}_n; \mathbf{X}_{n|N}\mathbf{\Sigma}_{n|N}), && \text{(4.32b)} \end{aligned}$$

for $k = 1, \ldots, K$. Then, using Lemma 2,

$$\begin{aligned} \mathbf{x}_{n|N}^k &= [\mathbf{X}_{n|N}]_{i:j}, && \text{(4.33a)} \\ \mathbf{\Sigma}_{n|N}^k &= [\mathbf{\Sigma}_{n|N}]_{i:j\times i:j} && \text{(4.33b)} \end{aligned}$$

where $i = 1 + (k-1) \cdot d_x$ and $j = k \cdot d_x$, for $k = 1, \ldots, K$, and $[\cdot]_{i:j}$ indicates the sub-vector containing the elements from ith index to the jth. Similarly, $[\cdot]_{i:j \times i:j}$ selects the sub-block which is formed by the elements from ith rows and columns to the jth.

As we have no backward factor at $n = N$, the parameters of the mixture components in (4.30b) become

$$\pi_{N|N}^{\mathbf{r}_N} \propto \alpha_N^{f,\mathbf{r}_N}, \qquad \mathbf{x}_{N|N}^{\mathbf{r}_N} = \mathbf{m}_N^{f,\mathbf{r}_N}, \qquad \mathbf{\Sigma}_{N|N}^{\mathbf{r}_N} = \mathbf{P}_N^{f,\mathbf{r}_N}, \tag{4.34}$$

for $\mathbf{r}_N = 1, \ldots, R_N$.

4.3.3 Selection of the Adjustment Exponents

4.3.3.1 Selection of γ_n^f

The forward adjustment exponent, γ_n^f, is selected as described in Section 3.3.3.1 for $n = 1, \ldots, N$.

4.3.3.2 Selection of γ_n^b

The selection procedure of γ_n^b is as presented in Sec. 3.3.3.2.

4.3.4 Pseudo-Code of the Algorithm

We initialize the EPwCA factors as follows. We first form the initial distribution for the augmented state vector as follows.

$$p(\mathbf{X}_0) = \mathcal{N}(\mathbf{X}_0; \mathbf{X}_{0|-1}, \mathbf{\Sigma}_{0|-1}), \tag{4.35}$$

where the augmented quantities are defined as

$$\mathbf{X}_0 = [\ (\mathbf{x}_0^1)^{\mathrm{T}} \quad \cdots \quad (\mathbf{x}_0^K)^{\mathrm{T}}\ ]^{\mathrm{T}}, \tag{4.36a}$$

$$\mathbf{X}_{0|-1} = [\ (\mathbf{x}_{0|-1}^1)^{\mathrm{T}} \quad \cdots \quad (\mathbf{x}_{0|-1}^K)^{\mathrm{T}}\ ]^{\mathrm{T}}, \tag{4.36b}$$

$$\boldsymbol{\Sigma}_{0|-1} \triangleq \text{blkdiag}\left[\ \boldsymbol{\Sigma}^1_{0|-1},\ \ \cdots,\ \ \boldsymbol{\Sigma}^K_{0|-1}\ \right]. \tag{4.36c}$$

(4.36d)

We compute the parameters of the first forward factor $\rho^f_0(\mathbf{x}_0, r_0)$ by performing a measurement update on the initial distribution $p(\mathbf{X}_0)$ using the augmented measurement quantities given in (C.3) for $n = 0$ and $\forall k$ satisfying $\mathbf{r}^k_0 > 0$ for all $\mathbf{r}_0$ as

$$\alpha^{f,\mathbf{r}_0}_0 \propto \mathcal{N}(\bar{\mathbf{y}}^{f,\mathbf{r}_0}_0; \overline{\mathbf{C}}^{f,\mathbf{r}_0}_0, \mathbf{S}^{\mathbf{r}_0}_0), \tag{4.37a}$$

$$\mathbf{m}^{f,\mathbf{r}_0}_0 = \mathbf{P}^{f,\mathbf{r}_0}_0(\boldsymbol{\Sigma}^{-1}_{0|-1}\mathbf{X}_{0|-1} + (\overline{\mathbf{C}}^{f,\mathbf{r}_0}_0)^{\mathrm{T}}(\overline{\mathbf{R}}^{f,\mathbf{r}_0}_0)^{-1}\bar{\mathbf{y}}^{f,\mathbf{r}_0}_0), \tag{4.37b}$$

$$\mathbf{P}^{f,\mathbf{r}_0}_0 = ((\boldsymbol{\Sigma}_{0|-1})^{-1} + (\overline{\mathbf{C}}^{f,\mathbf{r}_0}_0)^{\mathrm{T}}(\overline{\mathbf{R}}^{f,\mathbf{r}_0}_0)^{-1}\overline{\mathbf{C}}^{f,\mathbf{r}_0}_0)^{-1}, \tag{4.37c}$$

where

$$\mathbf{S}^{\mathbf{r}_0}_0 \triangleq \overline{\mathbf{C}}^{\mathbf{r}_0}_0 \boldsymbol{\Sigma}_{0|-1}(\overline{\mathbf{C}}^{\mathbf{r}_0}_0)^{\mathrm{T}} + \overline{\mathbf{R}}^{\mathbf{r}_0}_0. \tag{4.38}$$

As outlined in Section 3.3.4, a forward pass with $\gamma^f_n = 0$, $n = 1, \ldots, N$ is performed to initialize the forward factors $\rho^f_n(\cdot, \cdot)$, $n = 1, \ldots, N$. Then, we perform forward and backward passes successively until a convergence criterion is met. The pseudo-code of the proposed smoother algorithm is given in Algorithm 5. The forward and backward pass algorithms are described in Algorithm 6 and Algorithm 7, respectively. In contrast to the JMLS problem, damping is not included in Algorithm 6 in this case, as no need has arisen during our simulations. If one needs to do otherwise, they can refer to (3.38).

Algorithm 5 Pseudo Code for EPwCA Smoother

Input: $, \mathbf{X}_{0|-1}, \boldsymbol{\Sigma}_{0|-1}$; $\{\mathbf{Y}_n\}^N_{n=0}$; Γ_{conv}; $J_{\max}$; δ, $\{\mathcal{R}_n\}^N_{n=0}$.

Output: $\{\mathbf{x}_{n|N}, \boldsymbol{\Sigma}_{n|N}\}^N_{n=0}$.

1: $\gamma^f_n = 0, n = 1, \ldots, N$.

2: Calculate $\{\alpha^{f,\mathbf{r}}_0, \mathbf{m}^{f,\mathbf{r}}_0, \mathbf{P}^{f,\mathbf{r}}_0\}^R_{\mathbf{r}=1}$.

3: **for** $n = 1 : N$ **do**

4: Run Alg. 6 to get $\{\alpha^{f,\mathbf{r}}_n, \mathbf{m}^{f,\mathbf{r}}_n, \mathbf{P}^{f,\mathbf{r}}_n\}^R_{\mathbf{r}=1}$.

5: **end for**

6: $\gamma^f_n = 1, n = 1, \ldots, N-1$.

7: $\gamma^b_n = 1, n = 0, \ldots, N-1$.

8: **for** $n = N-1 : -1 : 1$ **do**

Run Alg. 7 to get $\{\mathbf{y}_n^b, \mathbf{C}_n^b, \mathbf{R}_n^b\}$.
end for
Find the final estimates $\{\mathbf{X}_{n|N}, \mathbf{\Sigma}_{n|N}\}$ for $n = 0, \ldots, N$.
$j = 1$
$e_{\text{conv}} = \infty$
while $e_{\text{conv}} > \Gamma_{\text{conv}}$ & $j \leq J_{\max}$ **do**
$\mathbf{x}_{n|N}^{\text{old}} = \mathbf{x}_{n|N}$; $n = 0, \ldots, N$.
for $n = 1 : N$ **do**
Run Alg. 6 to get $\{\alpha_n^{f,\mathbf{r}}, \mathbf{m}_n^{f,\mathbf{r}}, \mathbf{P}_n^{f,\mathbf{r}}\}_{\mathbf{r}=1}^R$.
end for
for $n = N - 1 : -1 : 1$ **do**
Run Alg. 7 to get $\{\mathbf{y}_n^b, \mathbf{C}_n^b, \mathbf{R}_n^b\}$.
end for
Find the final estimates $\{\mathbf{x}_{n|N}, \mathbf{\Sigma}_{n|N}\}$ for $n = 0, \ldots, N$.
$e_{\text{conv}} = \left(\frac{1}{d_X N} \sum_{n=0}^N \|\mathbf{x}_{n|N}^{\text{old}} - \mathbf{x}_{n|N}\|^2\right)^{0.5}$
$j \leftarrow j + 1$
end while

Algorithm 6 Pseudo Code for Forward Factor Update

Input: γ_n^f, $\Delta\gamma^f$, ϵ_1, ϵ_2.
Input: $\{\alpha_{n-1}^{f,\mathbf{r}}, \mathbf{m}_{n-1}^{f,\mathbf{r}}, \mathbf{P}_{n-1}^{f,\mathbf{r}}\}_{\mathbf{r}=1}^R$, $\mathbf{Y}_n$, $\mathcal{R}_n$.
Input: $\mathbf{y}_n^b, \mathbf{C}_n^b, \mathbf{R}_n^b$. ▷ If $\gamma_n^f \neq 0$
Output: $\{\alpha_n^{f,r}, \mathbf{m}_n^{f,r}, \mathbf{P}_n^{f,r}\}_{r=1}^R$,$\gamma_n^f$.
gammaReductionFlag $= 1$
while gammaReductionFlag $= 1$ **do**
gammaReductionFlag $= 0$
for $\mathbf{r} = 1 : R$ **do**
Calculate $\bar{\beta}_n^{f,\mathbf{r}}, \bar{\mathbf{v}}_n^{f,\mathbf{r}}, \overline{\mathbf{V}}_n^{f,\mathbf{r}}$.
Calculate $\mathbf{m}_n^{f,\mathbf{r}}, \mathbf{P}_n^{f,\mathbf{r}}$.
Calculate unnormalized $\alpha_n^{f,\mathbf{r}}$.
if $\min \text{eig } (\mathbf{P}_n^{f,\mathbf{r}})^{-1} \leq \epsilon_1$ or
$\min \text{eig } \overline{\mathbf{V}}_n^{f,\mathbf{r}} (\mathbf{P}_n^{f,\mathbf{r}})^{-1} \leq \epsilon_2$ **then**
gammaReductionFlag $= 1$

11: **end if**

12: **end for**

13: Normalize $\alpha_n^{f,\mathbf{r}}$, $r = 1, \ldots, R$.

14: **if** $\mathrm{gammaReductionFlag} = 1$ **then**

15: $\gamma_n^f \leftarrow \max\{0, \gamma_n^f - \Delta\gamma^f\}$

16: **end if**

17: **end while**

Algorithm 7 Pseudo Code for Backward Factor Update

Input: γ_n^b, ϵ_3.

Input: $\{\alpha_n^{f,\mathbf{r}}, \mathbf{m}_n^{f,\mathbf{r}}, \mathbf{P}_n^{f,\mathbf{r}}\}_{\mathbf{r}=1}^R$, $\mathbf{Y}_{n+1}$, $\mathcal{R}_{n+1}$.

Input: $\mathbf{y}_{n+1}^b, \mathbf{C}_{n+1}^b, \mathbf{R}_{n+1}^b$. ▷ If $n \leq N-2$

Output: $\mathbf{y}_n^b, \mathbf{C}_n^b, \mathbf{R}_n^b$.

1: Calculate $\bar{\mathbf{v}}_n^b$, $\overline{\mathbf{V}}_n^b$.

2: Calculate $\mathbf{v}_n^b$, $\mathbf{V}_n^b$.

3: Solve the generalized eigenvalue problem

$$\mathbf{V}_n^b \mathbf{e}_{n,i} = \lambda_{n,i} \overline{\mathbf{V}}_n^b \mathbf{e}_{n,i}$$

4: Order the generalized eigenvalue-eigenvector pairs in decreasing generalized eigenvalue order.

5: $L_n = \#\big(\{\lambda_{n,i}, i = 1, \ldots, d_X | \lambda_{n,i} > 1 + \epsilon_3\}\big)$

6: Choose the first L_n generalized eigenvalue-eigenvector pairs.

7: Calculate $\mathbf{y}_n^b, \mathbf{C}_n^b, \mathbf{R}_n^b$.

4.3.5 Computational Complexity

In this section, we analyze the computational complexity of EPwCA Smoother. Similar to Section 3.3.5, the algorithm is partitioned into three sub-blocks: the forward pass, the backward pass, and the computation of the state estimates, with computational loads F, B, and E, respectively. The overall computational load of EPwCA smoother is $\mathcal{O}(J_{max} N(F + B + E))$.

The computational loads of the sub-blocks are $F = \mathcal{O}\left(R_n R_{n-1} d_X^3 / \Delta\gamma^f\right)\big)$, $B = \mathcal{O}(R_n R_{n+1} d_X^3)$, and $E = \mathcal{O}(R_n d_X^3)$, as they are all dominated by the Kalman filter

or smoother operations. A downside of the proposed smoother is that the number of hypotheses, R_n at time n, is determined by permuting the measurements and no-detection case with the targets, therefore the computational cost increases combinatorially fast. Thus, it can be necessary to incorporate some mechanisms, such as gating, in order to control the number of hypotheses, when the computational capacity is limited or when tracking under heavy-clutter.

4.4 Extension to Filtering Problem

We apply the same filtering mechanism explained in Section 3.4.

4.4.1 Pseudo Code of the Filtering Algorithm

Algorithm 8 Pseudo Code for EPwCA Filter

Input: $\mathbf{X}_{0|-1}, \mathbf{\Sigma}_{0|-1}$; $\{\mathbf{Y}_n\}_{n=0}^{N}$; τ; Γ_{conv}; $J_{\max}$; δ, $\{\mathcal{R}_n\}_{n=0}^{N}$.

Output: $\mathbf{X}_{n|n}, \mathbf{\Sigma}_{n|n}$ for $n = 1, \ldots, N$.

1: $\gamma_n^f = 0$, $n = 1, \ldots, N$.
2: **for** $n = 1 : N$ **do**
3: Run Alg. 1 using $\{\mathbf{Y}_j\}_{j=i}^{n}$ and $\{\mathcal{R}_j\}_{j=i}^{n}$ to get $\{\mathbf{X}_{j|n}, \mathbf{\Sigma}_{j|n}\}$ for $i = \max(1, n-\tau)$.
4: Save $\{\mathbf{X}_{n|n}, \mathbf{\Sigma}_{n|n}\}$ as the filter output.
5: Use $\{\alpha_{i+1}^{f,\mathbf{r}}, \mathbf{m}_{i+1}^{f,\mathbf{r}}, \mathbf{P}_{i+1}^{f,\mathbf{r}}\}$ to initialize the Alg. 2 in the next run.
6: **end for**

4.5 Simulation Results

4.5.1 Test Scenarios

The suggested smoother and filter are tested on two scenarios.

Table 4.1: The simulation parameters of the tests

	K	P_D	β_{FA} $(\times 10^{-4})$
Test 1	2	0.9, 0.8, 0.7	1.5
Test 2	2	0.9	1.5, 3, 4.5
Test 3	2, 4, 6	0.9	1.5

4.5.1.1 Nearly Constant Velocity Scenario

We use a nearly-constant velocity model in 2-D with different parameter sets to test the proposed smoother and filter algorithms. The base state vector for the kth target is $\mathbf{x}_n^k \triangleq [p_n^{\mathrm{x}}\ p_n^{\mathrm{y}}\ v_n^{\mathrm{x}}\ v_n^{\mathrm{y}}]^{\mathrm{T}}$, $n = 0, \ldots, 10$. The model is

$$\mathbf{A} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2, \tag{4.39a}$$

$$\mathbf{Q} = (\sigma_w)^2 \begin{bmatrix} T^4/4 & T^3/2 \\ T^3/2 & T^2 \end{bmatrix} \otimes \mathbf{I}_2, \tag{4.39b}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}, \tag{4.39c}$$

$$\mathbf{R} = \sigma_v^2 \mathbf{I}_2, \tag{4.39d}$$

where $T = 0.5\,\mathrm{s}$, $\sigma_w = 8\,\mathrm{m/s}^2$, $\sigma_v = \sqrt{5}\,\mathrm{m}$. The parameters of the initial distribution for each target state are given as $p(\mathbf{x}_0^k) = \mathcal{N}(\mathbf{x}_0^k; \boldsymbol{\mu}_{0|-1}^k, \boldsymbol{\Sigma}_{0|-1}^k)$ with

$$\boldsymbol{\Sigma}_{0|-1}^k = \begin{bmatrix} 25 & 0 \\ 0 & 0.01 \end{bmatrix} \otimes \mathbf{I}_2. \tag{4.40}$$

The position entries of $\boldsymbol{\mu}_{0|-1}^k$ are selected randomly according to a uniform distribution from the surveillance region of $(0, 100) \times (0, 100)$, while the velocity components are drawn from a uniform distribution between $[-5, 5]$ m/s. The algorithms use the same model as the one given above. Table 4.1 shows the simulation parameters of different tests.

4.5.1.2 U-Turn Scenario

Inspired by [72], this scenario is composed of constant velocity motion and coordinated turn motion. The two targets approach each other with a constant velocity motion along the y-axis, and after making a coordinated turn, they move side by side along the x-axis with the constant velocity. The targets split with a second coordinated turn and move away from each other. In conclusion, two U-shaped trajectories that are quite close to each other at some interval are formed, which are illustrated in 4.3.

The starting points of the targets are $(0, 3000)$ and $(0, -3000)$. During the constant velocity motion, the speed of each target is 100 m/s. The separation between the targets during the second interval of constant velocity motion is 50m. Each constant velocity motion lasts 20 s, whereas the duration of each interval of coordinated turn is 15 s. The turn-rate of the coordinated turn is 0.1 rad/s.

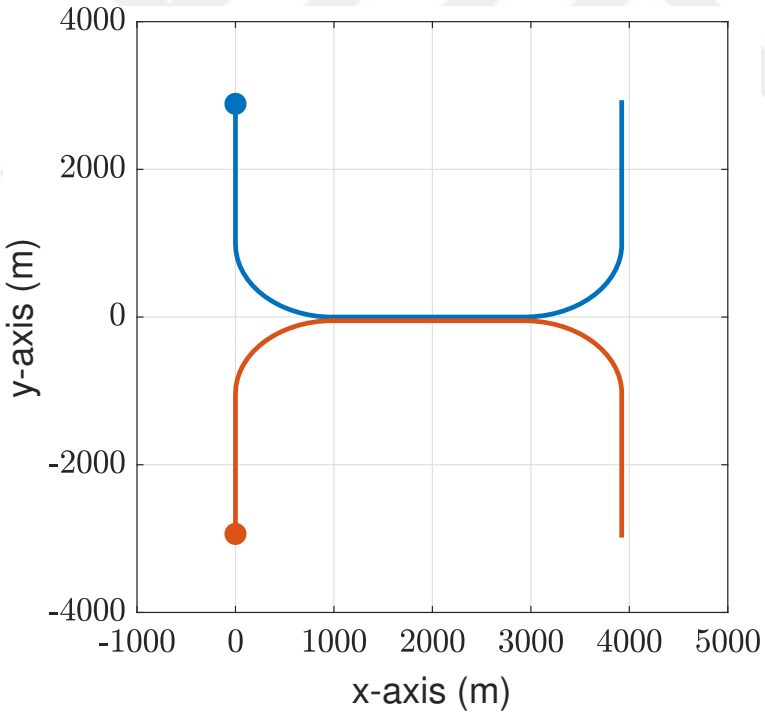


Figure 4.3: The trajectories of the U-Turn Test. The big circles indicate the starting points of the trajectories.

We carry out only one test having the parameters given in Table 4.2 due to the poor

Table 4.2: The simulation parameters of the U-Turn Test

	K	P_D	β_{FA}
U-Turn Test	2	0.999	$1 \cdot 10^{-7}$

performance of the alternative methods even in such an optimistic setting.

4.5.2 Performance Metrics

We consider the mean RMS and median errors that are averaged over targets as performance metrics. In addition to these, we also make use of the optimal sub-pattern assignment (OSPA) [73] that is available as a built-in function on MATLAB. The definition of OSPA is given as

$$D(\mathcal{X}_n, \mathcal{Y}_n) = (d_{loc}^p + d_{card}^p + d_{lab}^p)^{\frac{1}{p}} \tag{4.41}$$

where $\mathcal{X}_n$ and $\mathcal{Y}_n$ are set of ground truths and tracks at time n, respectively, with the cardinalities, l, and k. d_{loc} is the localization error between the assigned pairs that is given by

$$d_{loc} = \left[\frac{1}{k} \left(\min_{\pi_k \in \Pi} \sum_{i=1}^{l} d_c^p(x_{n,i}, y_{n,\pi(i)}) \right) \right]^{\frac{1}{p}}, \tag{4.42}$$

where $d_c(x, y)$ is the saturated distance, that is,

$$d_c(x, y) = \min(c, d(x, y)), \tag{4.43}$$

where c is the cut-off distance. The cardinality error, d_{card}, penalizes the unequal number of tracks and ground truths according to the following formula

$$d_{card} = \left(c^p \frac{|k - l|}{k} \right)^{\frac{1}{p}}. \tag{4.44}$$

Therefore, it implicitly accounts for the missed-targets and/or false tracks. However, this study assumes that the number of targets is fixed and known; thus, this error component is always zero. Finally, the last error component in (4.41) stands for the labeling errors, and it is defined as

$$d_{lab} = \left[\frac{1}{k} \left(\min_{\pi_k \in \Pi} \sum_{i=1}^{l} \alpha^p \gamma(L(x_{n,i}), L(y_{n,\pi(i)})) \right) \right]^{\frac{1}{p}}, \tag{4.45}$$

where $L(\cdot)$ retrieves the labels of the tracks and ground truths. If the labels match the correct assignment, then the function γ returns 0; otherwise, it gives 1. α is the penalty for incorrect assignments and should satisfy the condition $\alpha \leq c$.

In this study, we used the Euclidian distance as the distance measure $d(\cdot, \cdot)$. The order, p, is selected as 2. The cut-off distance and labelling penalty are assigned to $c = 50$m and $\alpha = 50$, respectively. The assignment decisions between the ground truths and the tracks are obtained using the GNN method.

Another metric we make use of is the ratio of the number of Monte Carlo runs in which a switch of track identity occurs to the total number of Monte Carlo runs. We name this ratio as track switch score (TSS). Furthermore, in order to have an insight into the algorithms' performance in terms of track convergence, we keep a convergence score for each method, which is calculated as the ratio of the number of converged tracks to the total number of tracks. A track is said to have diverged if the error between a track and the corresponding true trajectory exceeds a given threshold for 20% of the track's duration. Additionally, new RMS and median errors are computed after discarding the diverged tracks. The convergence threshold is set to 5 times the standard deviation of the prior density of the velocity.

4.5.3 Results

4.5.3.1 Results for the Smoothing Algorithms

We have implemented the following smoothers for comparison:

- **Rahmathullah et al. [69]**: The smoother proposed by Rahmathullah et al. [69].
- **PMHT [63]**: The smoothing algorithm proposed by Streit & Luginbuhl [63].
- **EPwCAS**: The proposed smoother in this work with parameters $J_{\text{max}} = 20$, $\Gamma_{\text{conv}} = 10^{-4}$, $\epsilon_1 = 10^{-10}$, $\epsilon_2 = 10^{-2}$, $\epsilon_3 = 10^{-6}$, $\Delta\gamma_f = 10^{-1}$. No damping is applied in this problem. Furthermore, we include the gating mechanism with a probability of 0.99 in order to ease the computational burden. However, the performance of the alternative methods degrades when gating is in operation,

therefore, gating is not employed with the alternatives.

The performance assessment is carried over 100 Monte Carlo runs for each test. One example of the estimates of the Test-3 with $K = 4$ is illustrated on the right-hand side of Fig. 4.4 where EPwCAS exhibits the best result. The example for the U-Turn Test in Fig. 4.4 emphasizes the performance difference between the smoothers. The EPwCAF is able to follow the targets, while the others fail during the first maneuver due to poor initialization.

In Tables 4.3 and 4.4, we observe that the proposed smoother, EPwCAS, performs better in all metrics than Rahmathullah and PMHT. Decreasing the probability of detection, P_D, results in a dramatic increase in median errors of Rahmathullah and PMHT, while they increase only slightly in the case of EPwCAS. Additionally, EPwCAS performs better in terms of the convergence score. Tables 4.5 and 4.6 illustrate the performances with respect to the increasing false alarm rates. EPwCAS outperforms other methods by achieving nearly half of the error and OSPA values of the other methods. The increasing false alarm rate does not affect significantly the metrics related to the track convergence, on the other hand, the decrease in the scores of Rahmathullah and PMHT is much faster. Similar to the previous tests, the OSPA value and the convergence score of the EPwCAS in Tables 4.7 and 4.6 indicate its superiority over the alternative techniques. In all these tests, EPwCAS scores lower TSS. Though TSSs of the alternative methods are close to the EPwCA, they actually suffer from track coalescence which does not affect this metric. Finally, it is evident from Tables 4.9 and 4.10 that Rahmathullah and PMHT fail to follow the maneuvering targets, while EPwCAS is able to track the targets fully with a considerably low RMS and median errors.

In conclusion, the proposed smoother demonstrates better performance in all metrics. However, its use in scenarios with much higher clutter density and a greater number of targets is limited due to its computational burden even when the gating is applied. On the other hand, though other methods are not computationally heavy algorithms, they are quite sensitive to initialization. Moreover, the mechanisms that are applied to overcome this issue come with their overheads. Therefore, their computational times become much longer compared to EPwCAS in scenarios covered in this thesis.

Table 4.3: The smoothed mean OSPA and mean errors for Test-1.

P_D		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
0.9	Rahmathullah et al.	15.59	37.99	13.32	4.42	5.71	0.03
	PMHT	12.27	32.40	11.23	3.04	5.31	0.05
	EPwCAS	**4.67**	**16.44**	**6.61**	**1.98**	**3.45**	**0**
0.8	Rahmathullah et al.	21.76	44.08	16.16	19.09	9.94	0.06
	PMHT	19.24	41.88	13.89	18.79	9.10	0.07
	EPwCAS	**6.15**	**22.29**	**7.97**	**2.28**	**3.74**	**0.01**
0.7	Rahmathullah et al.	29.62	59.99	19.62	40.75	14.70	0.06
	PMHT	23.95	48.08	15.59	27.75	11.53	0.08
	EPwCAS	**8.63**	**25.12**	**9.83**	**2.78**	**4.32**	**0.04**

Table 4.4: The smoothed convergence score and mean errors for Test-1 after the divergent tracks are left out.

P_D		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
0.9	Rahmathullah et al.	0.73	20.61	1.63	8.63	2.62
	PMHT	0.78	19.19	1.65	7.22	2.87
	EPwCAS	**0.93**	**12.52**	**1.27**	**5.44**	**2.18**
0.8	Rahmathullah et al.	0.59	23.01	1.96	8.09	3.00
	PMHT	0.59	26.84	2.01	8.98	3.19
	EPwCAS	**0.90**	**17.36**	**1.35**	**5.85**	**2.30**
0.7	Rahmathullah et al.	0.41	31.28	2.70	10.78	3.83
	PMHT	0.46	25.43	2.34	8.34	3.46
	EPwCAS	**0.84**	**17.69**	**1.52**	**6.37**	**2.49**

Table 4.5: The smoothed mean OSPA and mean errors for Test-2.

β_{FA} $(\times 10^{-4})$		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
1.5	Rahmathullah et al.	15.59	37.99	13.32	4.42	5.71	0.03
	PMHT	12.27	32.40	11.23	3.04	5.31	0.05
	EPwCAS	**4.67**	**16.44**	**6.61**	**1.98**	**3.45**	**0**
3	Rahmathullah et al.	16.61	37.82	15.64	11.47	7.72	0.05
	PMHT	14.04	34.18	11.60	9.16	7.25	0.04
	EPwCAS	**6.11**	**16.50**	**7.30**	**2.05**	**3.69**	**0.03**
4.5	Rahmathullah et al.	19.19	43.27	16.98	15.76	9.71	0
	PMHT	18.76	40.23	13.88	20.62	9.41	0.03
	EPwCAS	**6.31**	**17.93**	**7.77**	**2.28**	**3.75**	**0**

Table 4.6: The smoothed convergence score and mean errors for Test-2 after the divergent tracks are left out.

β_{FA} $(\times 10^{-4})$		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
1.5	Rahmathullah et al.	0.73	20.61	1.63	8.63	2.62
	PMHT	0.78	19.19	1.65	7.22	2.87
	EPwCAS	**0.93**	**12.52**	**1.27**	**5.44**	**2.18**
3	Rahmathullah et al.	0.66	25.82	1.69	11.23	2.73
	PMHT	0.70	20.23	1.80	6.50	2.94
	EPwCAS	**0.88**	**10.12**	**1.23**	**3.63**	**2.08**
4.5	Rahmathullah et al.	0.61	29.35	2.24	9.83	3.31
	PMHT	0.56	23.54	2.40	9.24	3.50
	EPwCAS	**0.89**	**13.09**	**1.29**	**5.13**	**2.27**

Table 4.7: The smoothed mean OSPA and mean errors for Test-3.

K		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
2	Rahmathullah et al.	15.59	37.99	13.32	4.42	5.71	0.03
	PMHT	12.27	32.40	11.23	3.04	5.31	0.05
	EPwCAS	**4.67**	**16.44**	**6.61**	**1.98**	**3.45**	**0**
4	Rahmathullah et al.	17.29	41.75	14.27	31.45	11.60	0.15
	PMHT	14.80	38.99	12.10	26.17	9.46	0.10
	EPwCAS	**5.89**	**23.39**	**7.81**	**2.37**	**3.89**	**0.08**
6	Rahmathullah et al.	17.04	48.95	16.97	44.07	15.00	0.37
	PMHT	20.58	46.41	14.66	38.11	13.08	0.33
	EPwCAS	**5.53**	**24.10**	**8.82**	**4.28**	**4.60**	**0.20**

Table 4.8: The smoothed convergence score and mean errors for Test-3 after the divergent tracks are left out.

K		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
2	Rahmathullah et al.	0.73	20.61	1.63	8.63	2.62
	PMHT	0.78	19.19	1.65	7.22	2.87
	EPwCAS	**0.93**	**12.52**	**1.27**	**5.44**	**2.18**
4	Rahmathullah et al.	0.61	27.72	6.51	9.72	4.59
	PMHT	0.69	21.01	2.36	7.68	3.64
	EPwCAS	**0.88**	**16.34**	**1.47**	**5.73**	**2.48**
6	Rahmathullah et al.	0.50	34.33	19.33	12.23	7.26
	PMHT	0.58	29.44	14.82	9.52	5.90
	EPwCAS	**0.86**	**19.37**	**1.78**	**6.51**	**2.79**

Table 4.9: The smoothed mean OSPA and mean errors for U-Turn Test.

	OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
Rahmathullah et al.	45.56	3405.4	561.79	3373.6	512.03	0
PMHT	53.50	3265.2	118.71	3268.5	118.74	0
EPwCAS	**0.57**	**0.61**	**3.29**	**0.55**	**3.02**	0

Table 4.10: The smoothed divergence score and mean errors for U-Turn Test after the divergent tracks are left out.

	score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
Rahmathullah et al.	0	-	-	-	-
PMHT	0	-	-	-	-
EPwCAS	**1**	**0.43**	**0.35**	**2.30**	**1.90**

4.5.3.2 Results for the Filtering Algorithms

In order to carry out the performance analysis following filters are implemented.

- **JPDA [61]**: A widely-used algorithm that stores a single hypothesis at the end of each time step. It performs the measurement update using an equivalent measurement formed by merging the association hypotheses in a single scan.
- **EPwCAF**: The proposed filter in this work with a window size of 3. The other parameters are $J_{\max} = 2$, $\Gamma_{\text{conv}} = 10^{-2}$, $\epsilon_1 = 10^{-10}$, $\epsilon_2 = 10^{-2}$, $\epsilon_3 = 10^{-6}$, $\Delta\gamma_f = 10^{-1}$. No damping is applied.

We performed 100 Monte Carlo runs for each test in Table 4.1. The right-hand side of Fig. 4.5, in which the filters produce very similar estimates, illustrates an example run for Test-3 with $K = 4$. On the other hand, JPDA mistakenly exchanges the target identities during the second maneuver in the realization of U-Turn Test in Fig.4.5, whereas EPwCAF successfully follows and correctly distinguishes the targets.

Each test reveals an almost similar behavior in regard to the changing parameters.

Table 4.11: The filtered mean OSPA and mean errors for Test-1.

P_D		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
0.9	JPDA	**5.00**	20.90	11.03	4.09	7.94	0
	EPwCAF	5.88	**18.96**	**8.59**	**2.70**	**5.24**	0
0.8	JPDA	**7.01**	31.02	13.15	4.96	8.83	0.02
	EPwCAF	8.56	**28.32**	**10.52**	**3.13**	**5.77**	**0**
0.7	JPDA	**9.99**	34.72	14.19	6.94	9.99	**0.01**
	EPwCAF	10.66	**30.08**	**11.30**	**3.75**	**6.44**	0.02

Tables 4.11-4.14 suggest that there are only slight differences between EPwCAF and JPDA. The OSPA value, score, and errors of the convergent runs of JPDA are better than those of EPwCAF. In contrast, EPwCAF attains lower error values compared to JPDA before discarding the divergent estimates. In Tables 4.15 and 4.16, though JPDA achieves better performance values, the results of the two filters are very close as opposed to the previous tests. On the other hand, the performance difference between the two methods is rather striking in U-Turn Test. EPwCAF surpasses JPDA by following the maneuvering targets in all runs, while JPDA is able to generate a track that is not diverging in only one instance due to track switches. The poor performance stems from the target identity switches during the maneuvers. That explains the considerable gap between the errors of EPwCAF and JPDA in Tables 4.17 and 4.10. We note that the performance of the proposed algorithm can be improved further by increasing fixed-lag smoothing the window size.

As mentioned in Section 3.5, the filter results can come out better than the smoother results, as a faulty assignment decision made at an instant can propagate to the previous time instants and degrade the performance in the smoother.

4.6 Conclusion

In this chapter, a fixed-interval smoother and a filter have been derived for the data association problem in MTT. The concept of context adjustment and the idea of using

Table 4.12: The filtered convergence score and mean errors for Test-1.

P_D		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
0.9	JPDA	**0.96**	**10.92**	**1.82**	**5.74**	**3.59**
	EPwCAF	0.92	12.29	1.85	5.89	3.64
0.8	JPDA	**0.93**	19.23	2.08	7.18	3.83
	EPwCAF	0.85	**18.22**	**2.01**	**6.92**	**3.81**
0.7	JPDA	**0.86**	20.63	2.54	7.64	4.20
	EPwCAF	0.82	**18.34**	**2.37**	**7.37**	**4.14**

Table 4.13: The filtered mean OSPA and mean errors for Test-2.

β_{FA} $(\times 10^{-4})$		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
1.5	JPDA	**5.00**	20.90	11.03	4.09	7.94	0
	EPwCAF	5.88	**18.96**	**8.59**	**2.70**	**5.24**	0
3	JPDA	**6.34**	19.38	11.07	4.41	8.30	0.02
	EPwCAF	7.67	**18.42**	**8.46**	**2.62**	**5.09**	0.02
4.5	JPDA	**6.72**	21.28	11.04	5.09	8.66	0
	EPwCAF	7.63	**18.73**	**8.62**	**2.83**	**5.54**	0

Table 4.14: The filtered convergence score and mean errors for Test-2.

β_{FA} $(\times 10^{-4})$		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
1.5	JPDA	**0.96**	**10.92**	**1.82**	**5.74**	**3.59**
	EPwCAF	0.92	12.29	1.85	5.89	3.64
3	JPDA	**0.91**	11.71	1.82	**5.13**	3.39
	EPwCAF	0.85	**10.93**	**1.77**	5.14	3.39
4.5	JPDA	**0.91**	14.37	2.04	**5.98**	3.88
	EPwCAF	0.88	**14.01**	**1.95**	6.01	**3.76**

Table 4.15: The filtered mean OSPA and mean errors for Test-3.

K		OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
2	JPDA	**5.00**	20.90	11.03	4.09	7.94	0
	EPwCAF	5.88	**18.96**	**8.59**	**2.70**	**5.24**	0
4	JPDA	**5.45**	45.08	17.37	6.85	12.51	0.12
	EPwCAF	7.30	**23.49**	**9.28**	**2.71**	**5.32**	**0.05**
6	JPDA	**6.26**	60.98	23.22	11.83	16.91	0.27
	EPwCAF	8.04	**20.86**	**8.64**	**2.73**	**5.27**	**0.26**

Table 4.16: The filtered convergence score and mean errors for Test-3.

K		score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
2	JPDA	**0.96**	**10.92**	**1.82**	**5.74**	**3.59**
	EPwCAF	0.92	12.29	1.85	5.89	3.64
4	JPDA	**0.90**	**15.90**	**2.18**	**6.27**	**4.01**
	EPwCAF	0.87	17.02	2.21	6.50	4.03
6	JPDA	**0.87**	15.65	**2.52**	**6.30**	**4.42**
	EPwCAF	0.84	**14.97**	2.69	6.43	4.43

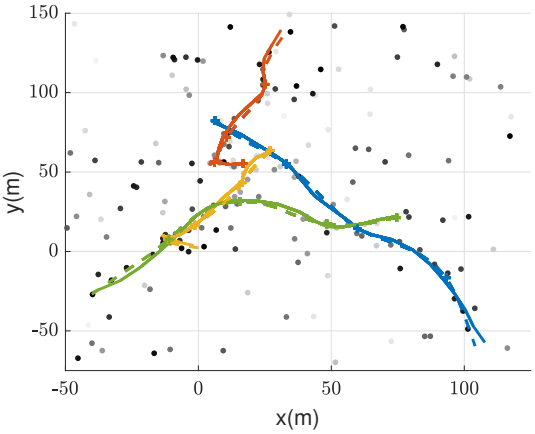
Table 4.17: The filtered mean OSPA and mean errors for U-Turn Test.

	OSPA	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)	TSS
JPDA	27.46	991.89	62.57	1012.4	89.91	0.96
EPwCAF	**0.60**	**0.46**	**3.36**	**0.37**	**2.82**	**0**

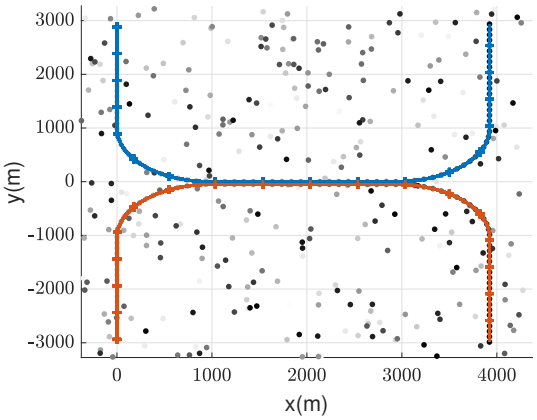
Table 4.18: The filtered convergence score and mean errors for U-Turn Test.

	score	PRMS(m)	VRMS(m/s)	PMed.(m)	VMed.(m/s)
JPDA	0.01	8.37	8.37	4.44	4.44
EPwCAF	**1**	**0.46**	**0.37**	**3.36**	**2.82**

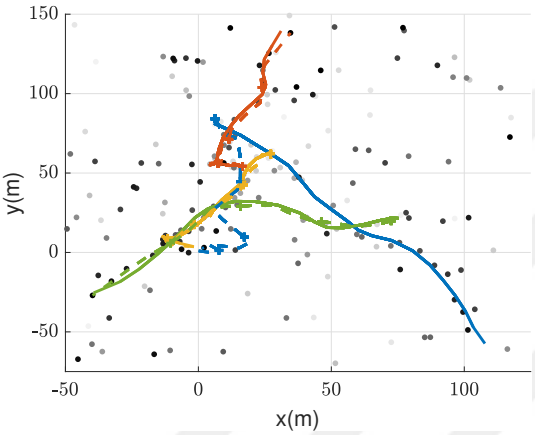
pseudo-likelihoods introduced in Chapter 2 have been incorporated into the derivation process. The performances of the proposed algorithms have been examined by comparing them to their alternatives in different environments. We have shown that the proposed smoother significantly improves the estimation results together with the assignment decisions, and the proposed filter is on par with its alternative. The alternative smoothers suffer from poor initialization, whereas EPwCA smoother has a greater computational cost.



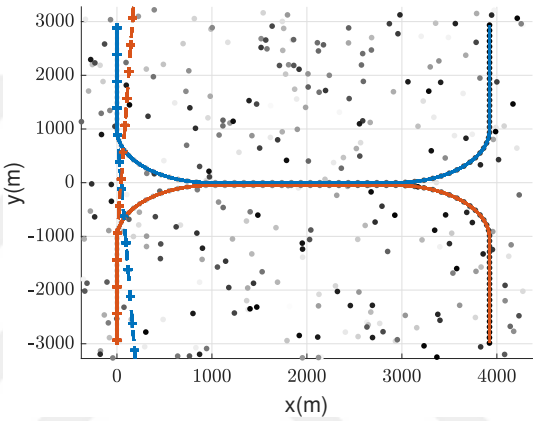
(a) EPwCAS



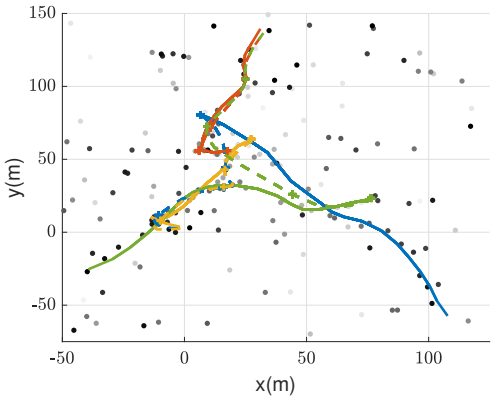
(b) EPwCAS



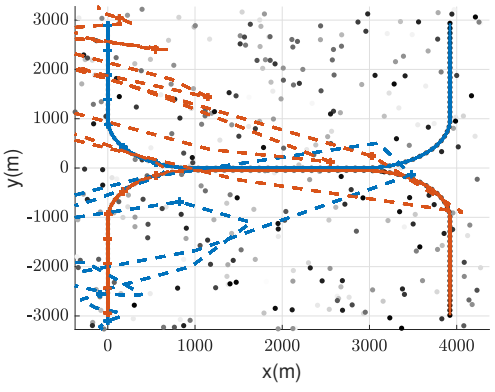
(c) PMHT



(d) PMHT



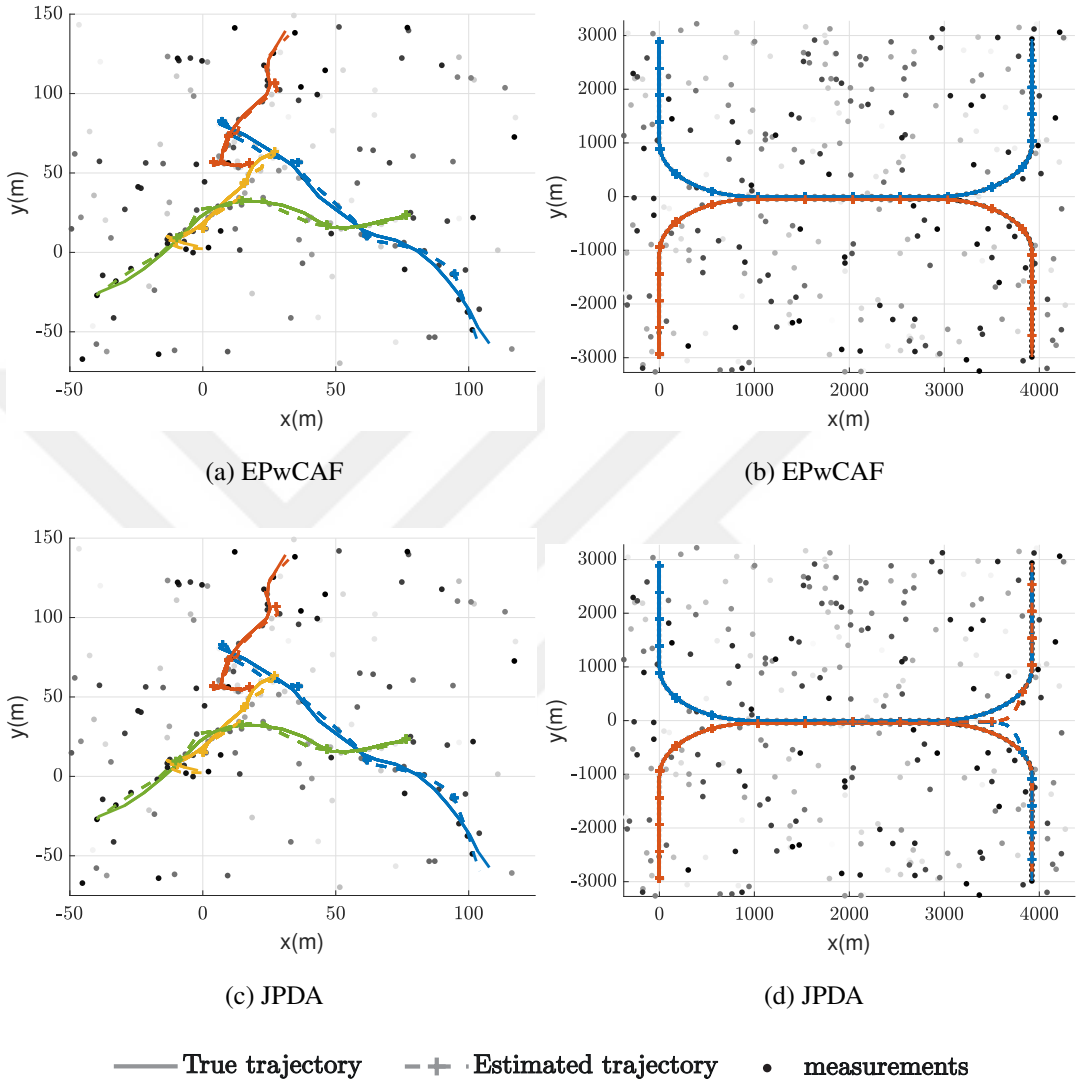
(e) Rahmathullah et al.



(f) Rahmathullah et al.

True trajectory Estimated trajectory measurements

Figure 4.4: Example realizations from the tests for each smoother. The ones on the left and right are examples from Test-3 when $K = 4$ and U-Turn Test, respectively. The measurements are shown with dots whose shading gets darker as time progresses.



(a) EPwCAF (b) EPwCAF

(c) JPDA (d) JPDA

Figure 4.5: Example realizations from the tests for each filter. The ones on the left and right are examples from Test-3 when $K = 4$ and U-Turn Test, respectively. The measurements are shown with dots whose shading gets darker as time progresses.

CHAPTER 5

CONCLUSION

In this thesis, we have addressed the state estimation problem with discrete-valued hidden random variables in which the optimal Bayesian solution is intractable due to the exponential increase in computational complexity.

To derive a sub-optimal solution, we investigated the problem in the EP framework, which can exhibit numerical issues intrinsically, preventing its usage in real-life problems. To overcome these difficulties and improve its convergence properties, we have introduced the concept of context adjustment that allows us to control the amount of information leaking through the cavity distribution in the EP problem. We have named this novel version of EP with Context Adjustment. Additionally, we have proposed the pseudo-likelihoods as a new form of backward factors in the EP so that the probabilistic nature of the problem is better reflected in the approximate model. To this end, we have derived the analytical sub-optimal optimization solution of the M-projection problem for the first time in the literature, which can also be applicable to other estimation problems as well. During the simulations, we observed that the proposed novelties have ameliorated the performance of the EP by alleviating the numerical issues in the problems where standard EP fails to converge. However, the theoretical analysis of these improvements in the convergence properties of EP deserves further study.

Using the EPwCA and the new form of backward factors, we have tackled two problems involving discrete hidden random variables: the state estimation in JMLSs, and the target tracking problem under the measurement origin uncertainty. Though EP was considered as an inference method in JMLSs before, those studies lacked a comprehensive analysis of performance and failed to present a convergent algorithm. In

this study, we have derived a fixed-interval smoother based on EPwCA for JMLSs, which eliminates the convergence problems. Detailed simulations with different scenarios have shown that the proposed smoother is either on par with the alternative methods or surpasses them. Furthermore, we have attained an EP-based filter by applying the EPwCA on a sliding window, which performed better in challenging scenarios than the popular filters in the literature.

The second problem we have examined in the framework of EP is the target tracking problem under the measurement origin uncertainty. To our knowledge, this problem has not been studied in this framework before this thesis, and we have exploited the EPwCA to obtain a fixed-interval smoother and its filter extension. We have conducted a detailed examination of the performance of the proposed algorithms by changing the scenario parameters and comparing them to their alternatives. We have observed that both the EPwCA smoother and the EPwCA filter are more robust to these changes, exhibiting less track coalescence in general and having significantly lower RMS and median errors. On the other hand, the proposed methods are computationally expensive to apply in cases with a high number of targets and false alarms.

5.1 Future Work

A list of topics open for further research in the light of this thesis is given below.

- We could provide only empirical evidence obtained from the simulations about the improved convergence properties of EPwCA compared to standard EP. An interesting future study direction might involve a theoretical analysis of why and how the contributions of the current study resulted in improvements in convergence, which might get us one step closer to fully understanding the convergence properties of EP and its variants.

- The use of EPwCA in JMLSs can be extended to other applications, such as the estimation problem in JMLSs as in [51] and clustering [11], that involve mixtures instead of single Gaussians.

- One can consider investigating the augmented switching linear dynamical sys-

tems in the framework of EPwCA, in which the transition probabilities depend on the continuous base state as opposed to the independent transition probabilities in standard JMLSs.

- Our simulations regarding the data association problem were limited to scenarios with moderate clutter intensity and a moderate number of targets due to computational burden. Examining the ways of reducing the computational complexity that will enable the application of EPwCA to cases with higher clutter intensity with a larger number of targets is open for further research.

- Finally, one can consider adapting the proposed data association algorithms to the tracking problems where the number of targets is unknown. To do so, additional binary-valued hidden random variables that represent the presence of each target at each time step can be introduced into the probabilistic model, which would bring further combinatoric challenges to the estimation problem.

REFERENCES

[1] J. S. Abel, *Localization using range differences*. PhD thesis, Stanford University, 1989.

[2] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[3] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. John Wiley & Sons, 2001.

[5] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[6] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.

[7] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[8] J. D. Murray, *Mathematical Biology: I. An Introduction*. Springer, 2002.

[9] E. Greenberg, *Introduction to Bayesian Econometrics*. Cambridge University Press, 2012.

[10] G. Koop, *Bayesian Econometrics*. John Wiley & Sons, 2003.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[13] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*. University of California, Berkeley, 2002.

[14] P. J. Mosterman and G. Biswas, "Diagnosis of continuous valued systems in transient operating regions," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 29, no. 6, pp. 554–565, 1999.

[15] K. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," *arXiv preprint arXiv:1301.6725*, 2013.

[16] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 362–369, 2001.

[17] J. Piger, "Econometrics: Models of regime changes," in *Complex systems in finance and econometrics*, pp. 190–202, Springer, 2009.

[18] V. Pavlovic, J. M. Rehg, and J. MacCormick, "Learning switching linear models of human motion," *Advances in Neural Information Processing Systems*, vol. 13, 2000.

[19] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 679–694, 2006.

[20] S. Lacoste-Julien, B. Taskar, D. Klein, and M. Jordan, "Word alignment via quadratic assignment," 2006.

[21] M. Kellis, N. Patterson, B. Birren, B. Berger, and E. S. Lander, "Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery," *Journal of Computational Biology*, vol. 11, no. 2-3, pp. 319–355, 2004.

[22] A. Doucet, A. Smith, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics, Springer New York, 2001.

[23] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, pp. 5–43, Jan 2003.

[24] T. Minka, "Divergence measures and message passing," tech. rep., Microsoft Research, Cambridge UK, 2005.

[25] C. W. Fox and S. J. Roberts, "A tutorial on variational Bayesian inference," *Artificial Intelligence Review*, vol. 38, no. 2, pp. 85–95, 2012.

[26] J. Winn and C. M. Bishop, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, no. Apr, pp. 661–694, 2005.

[27] T. P. Minka, *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.

[28] D. Hernández-Lobato and J. M. Hernández-Lobato, "Scalable Gaussian process classification via expectation propagation," in *Artificial Intelligence and Statistics*, pp. 168–176, 2016.

[29] T. Bui, D. Hernández-Lobato, J. Hernandez-Lobato, Y. Li, and R. Turner, "Deep Gaussian processes for regression using approximate expectation propagation," in *International Conference on Machine Learning*, pp. 1472–1481, 2016.

[30] T. Furmston and D. Barber, "Variational methods for reinforcement learning," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 241–248, 2010.

[31] Y. Qi and T. P. Minka, "Window-based expectation propagation for adaptive signal detection in flat-fading channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 348–355, 2007.

[32] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, 2014.

[33] I. Santos, J. J. Murillo-Fuentes, R. Boloix-Tortosa, E. Arias-de Reyna, and P. M. Olmos, "Expectation propagation as turbo equalizer in ISI channels," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 360–370, 2016.

[34] J. Ma, R. Dudeja, J. Xu, A. Maleki, and X. Wang, "Spectral method for phase retrieval: An expectation propagation perspective," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 1332–1355, 2021.

[35] M. Deisenroth and S. Mohamed, "Expectation propagation in Gaussian process dynamical systems," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[36] A. Vehtari, A. Gelman, T. Sivula, P. Jylänki, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. P. Robert, "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data.," *Journal of Machine Learning Research*, vol. 21, pp. 17–1, 2020.

[37] P. Jylänki, J. Vanhatalo, and A. Vehtari, "Robust Gaussian process regression with a Student-t likelihood," *Journal of Machine Learning Research*, vol. 12, no. 99, pp. 3227–3257, 2011.

[38] T. P. Minka and J. D. Lafferty, "Expectation-propogation for the generative aspect model," *CoRR*, vol. abs/1301.0588, 2013.

[39] T. Heskes and O. Zoeter, "Expectation propogation for approximate inference in dynamic Bayesian networks," *CoRR*, vol. abs/1301.0572, 2013.

[40] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.

[41] D. Barber, *Bayesian Time Series Models*, ch. Approximate inference in switching linear dynamical systems using Gaussian mixtures, pp. 166—181. Cambridge University Press, 2011.

[42] T. Minka, "Power EP," Tech. Rep. MSR-TR-2004-149, Microsoft Research, Cambridge UK, Jan. 2004.

[43] X. Rong Li and V. Jilkov, "Survey of maneuvering target tracking. Part V. Multiple-model methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1255–1321, 2005.

[44] C.-B. Chang and M. Athans, "State estimation for discrete systems with switching parameters," *IEEE Transactions on Aerospace and Electronic Systems*, no. 3, pp. 418–425, 1978.

[45] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for

systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[46] G. Kitagawa, "The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother," *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 4, pp. 605–623, 1994.

[47] R. Helmick, W. Blair, and S. Hoffman, "Fixed-interval smoothing for Markovian switching systems," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1845–1855, 1995.

[48] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[49] C.-J. Kim, "Dynamic linear models with Markov-switching," *Journal of Econometrics*, vol. 60, no. 1-2, pp. 1–22, 1994.

[50] W. Koch, "Fixed-interval retrodiction approach to Bayesian IMM-MHT for maneuvering multiple targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 2–14, 2000.

[51] D. Barber, "Expectation correction for smoothing in switching linear Gaussian state space models," *Journal of Machine Learning Research*, vol. 7, pp. 2515–2540, 2006.

[52] N. Nadarajah, R. Tharmarasa, M. McDonald, and T. Kirubarajan, "IMM forward filtering and backward smoothing for maneuvering target tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2673–2678, 2012.

[53] R. Lopez and P. Danes, "Low-complexity IMM smoothing for jump Markov nonlinear systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1261–1272, 2017.

[54] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, vol. 13, pp. 429–434, July 1977.

[55] J. Tugnait and A. Haddad, "A detection-estimation scheme for state estimation in switching environments," *Automatica*, vol. 15, pp. 477–481, July 1979.

[56] O. Zoeter and T. Heskes, *Bayesian Time Series Models*, ch. Expectation propagation and generalized EP methods for inference in switching linear dynamical systems, pp. 141—165. Cambridge University Press, 2011.

[57] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[58] Y. Ma, S. Zhao, and B. Huang, "Multiple-model state estimation based on variational Bayesian inference," *IEEE Transactions on Automatic Control*, vol. 64, pp. 1679–1685, Apr. 2019.

[59] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell, *Bayesian Multiple Target Tracking*. Artech House, Dec. 2013.

[60] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.

[61] K.-C. Chang and Y. Bar-Shalom, "Joint probabilistic data association for multitarget tracking with possibly unresolved measurements and maneuvers," *IEEE Transactions on Automatic Control*, vol. 29, no. 7, pp. 585–594, 1984.

[62] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[63] R. L. Streit and T. E. Luginbuhl, "Probabilistic multi-hypothesis tracking," tech. rep., Naval Underwater Systems Center Newport RI, 1995.

[64] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[65] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[66] P. Willett, Y. Ruan, and R. Streit, "PMHT: Problems and some solutions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 738–754, 2002.

[67] D. F. Crouse, M. Guerriero, and P. Willett, "A critical look at the PMHT," *Journal of Advances in Information Fusion*, vol. 4, no. 2, pp. 93–116, 2009.

[68] M. Wieneke and W. Koch, "A PMHT approach for extended objects and object groups," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349–2370, 2012.

[69] A. S. Rahmathullah, R. Selvan, and L. Svensson, "A batch algorithm for estimating trajectories of point targets using expectation maximization," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4792–4804, 2016.

[70] N. Ueda and R. Nakano, "Deterministic annealing EM algorithm," *Neural networks*, vol. 11, no. 2, pp. 271–282, 1998.

[71] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

[72] H. A. Blom and E. A. Bloem, "Combining IMM and JPDA for tracking multiple maneuvering targets in clutter," in *Proceedings of the Fifth International Conference on Information Fusion*, vol. 1, pp. 705–712, IEEE, 2002.

[73] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.

APPENDIX A

M-PROJECTION INVOLVING PSEUDO-LIKELIHOODS

In this appendix, we solve the following minimization problem

$$\{\mathbf{y}^*, \mathbf{C}^*, \mathbf{R}^*\} = \arg \min_{\{\mathbf{y},\mathbf{C},\mathbf{R}\}} \mathrm{KL}(p(\cdot)||q(\cdot)), \tag{A.1}$$

where $p(\cdot)$ is an arbitrary probability density function with known mean $\bar{\mathbf{x}}$ and covariance $\overline{\mathbf{P}}$, and the density $q(\mathbf{x})$ is defined as

$$q(\mathbf{x}) \propto \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{x}, \mathbf{R})\, \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P}), \tag{A.2}$$

where $\mathbf{y} \in \mathbb{R}^{\ell}$, $\mathbf{C} \in \mathbb{R}^{\ell \times d_x}$ and $\mathbf{R} \in \mathbb{R}^{\ell \times \ell}$ with $\ell \leq d_x$ are unknowns to be determined, and the mean $\hat{\mathbf{x}}$ and covariance $\overline{\mathbf{P}}$ are given. We can write $q(\mathbf{x})$ as

$$q(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \tilde{\mathbf{x}}, \widetilde{\mathbf{P}}), \tag{A.3}$$

where

$$\widetilde{\mathbf{P}} \triangleq (\mathbf{P}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C})^{-1}, \tag{A.4a}$$

$$\tilde{\mathbf{x}} \triangleq \widetilde{\mathbf{P}}(\mathbf{P}^{-1}\hat{\mathbf{x}} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}). \tag{A.4b}$$

We assume that $\mathbf{C}$ has full row-rank and the matrix $\mathbf{R}$ is positive definite. We can calculate the KL divergence between from $q(\mathbf{x})$ to $p(\mathbf{x})$ as

$$\mathrm{KL}(p(\cdot)||q(\cdot)) \stackrel{+}{=} -\int p(\mathbf{x}) \log \mathcal{N}(\mathbf{x}; \tilde{\mathbf{x}}, \widetilde{\mathbf{P}}) \mathrm{d}\mathbf{x} \tag{A.5a}$$

$$\stackrel{+}{=} \frac{1}{2} \log |\widetilde{\mathbf{P}}| + \frac{1}{2} \mathrm{tr}(\widetilde{\mathbf{P}}^{-1}\overline{\mathbf{P}}) + \frac{1}{2}(\tilde{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}} \widetilde{\mathbf{P}}^{-1}(\cdot) \tag{A.5b}$$

where the sign $\stackrel{+}{=}$ denotes equality up to an additive constant with respect to the optimized variables. The cost function above is denoted in the following as J, i.e.,

$$J = \frac{1}{2} \log |\widetilde{\mathbf{P}}| + \frac{1}{2} \mathrm{tr}(\widetilde{\mathbf{P}}^{-1}\overline{\mathbf{P}}) + \frac{1}{2}(\tilde{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}} \widetilde{\mathbf{P}}^{-1}(\cdot), \tag{A.6}$$

where for the sake of brevity we drop the arguments of the cost.

A.1 Minimization with respect to the measurement $\mathbf{y}$

We can compute the derivative of the cost (A.6) with respect to the measurement $\mathbf{y}$ as follows.

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{y}} =& \frac{\partial}{\partial \mathbf{y}}(\tilde{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot), && \text{(A.7a)}\\
=& \frac{\partial}{\partial \mathbf{y}}(\widetilde{\mathbf{P}}(\widetilde{\mathbf{P}}^{-1}\hat{\mathbf{x}} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}) - \bar{\mathbf{x}})^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot), && \text{(A.7b)}\\
=& 2\mathbf{R}^{-1}\mathbf{C}\widetilde{\mathbf{P}}\widetilde{\mathbf{P}}^{-1}(\widetilde{\mathbf{P}}(\mathbf{P}^{-1}\hat{\mathbf{x}} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}) - \bar{\mathbf{x}}), && \text{(A.7c)}\\
=& 2\mathbf{R}^{-1}\mathbf{C}(\widetilde{\mathbf{P}}(\mathbf{P}^{-1}\hat{\mathbf{x}} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}) - \bar{\mathbf{x}}). && \text{(A.7d)}
\end{aligned}$$

Equating the right hand side of (A.7d) to zero, multiplying both sides with $\mathbf{R}$ from the left, and then solving for $\mathbf{y}$ would give the optimal measurement $\mathbf{y}^*$ as

$$\mathbf{y}^* = \mathbf{R}\big(\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\big)^{-1}\mathbf{C}\big(\bar{\mathbf{x}} - \widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}}\big). \tag{A.8}$$

A.2 Minimization with respect to the measurement noise covariance $\mathbf{R}$

We first substitute $\mathbf{y}^*$ in (A.8) into the quadratic term in J.

$$\begin{aligned}
&(\tilde{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot) = (\widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}} - \bar{\mathbf{x}} + \widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^*)^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot),\\
=&\big(\widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}} - \bar{\mathbf{x}} + \widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}(\bar{\mathbf{x}} - \widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}})\big)^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot), && \text{(A.9a)}\\
=&\big(\big(\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\widetilde{\mathbf{P}}\big)(\mathbf{P}^{-1}\hat{\mathbf{x}} - \widetilde{\mathbf{P}}^{-1}\bar{\mathbf{x}})\big)^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\big(\,\cdot\,\big). && \text{(A.9b)}
\end{aligned}$$

Applying Theorem 4 in Appendix E on the right hand side of (A.9b) gives

$$\begin{aligned}
&\big(\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\widetilde{\mathbf{P}}\big)(\mathbf{P}^{-1}\hat{\mathbf{x}} - \widetilde{\mathbf{P}}^{-1}\bar{\mathbf{x}})\\
=&\big(\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}(\widetilde{\mathbf{P}}^{-1} - \widetilde{\mathbf{P}}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1})\widetilde{\mathbf{P}}\big)(\mathbf{P}^{-1}\hat{\mathbf{x}} - \widetilde{\mathbf{P}}^{-1}\bar{\mathbf{x}}), && \text{(A.10a)}\\
=&\,\mathbf{U}(\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}(\mathbf{P}^{-1}\hat{\mathbf{x}} - \widetilde{\mathbf{P}}^{-1}\bar{\mathbf{x}}) && \text{(A.10b)}\\
=&\,\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}(\hat{\mathbf{x}} - \bar{\mathbf{x}}) && \text{(A.10c)}
\end{aligned}$$

where $\mathbf{U} \in \mathbb{R}^{d_x \times (d_x - \ell)}$ is a matrix such that $\mathbf{C}\mathbf{U} = 0$ and the augmented square matrix $[\mathbf{C}^{\mathrm{T}}\ \ \widetilde{\mathbf{P}}^{-1}\mathbf{U}]$ is invertible. The equality (A.10c) is written by using the facts

$$\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\mathbf{U} = \mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U}, \qquad \mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\bar{\mathbf{x}} = \mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\bar{\mathbf{x}}, \tag{A.11}$$

which can be derived from the condition $\mathbf{C}\mathbf{U} = 0$. We insert (A.10c) back into (A.9b) to get

$$\begin{aligned}
&\big(\widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}} - \bar{\mathbf{x}} + \widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^{*}\big)^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\big(\,\cdot\,\big) \\
&=(\hat{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\widetilde{\mathbf{P}}^{-1}(\cdot), && \text{(A.12a)} \\
&=(\hat{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}(\cdot), && \text{(A.12b)} \\
&=(\hat{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{P}^{-1}(\cdot) - (\hat{\mathbf{x}} - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}(\cdot), && \text{(A.12c)}
\end{aligned}$$

where we made use of Theorem 4 in Appendix E once again. We observe that (A.12c) is independent of $\mathbf{R}$, hence only the first two terms of J in (A.6) have to be optimized with respect to $\mathbf{R}$. Since $\mathbf{R}$ is invertible, optimization with respect to $\mathbf{R}$ is equivalent to optimization with respect to $\mathbf{R}^{-1}$. Taking the derivative of the first two terms of J on the right-hand side of (A.6) with respect to $\mathbf{R}^{-1}$, we get

$$\begin{aligned}
2\frac{\partial J}{\partial \mathbf{R}^{-1}} =& \frac{\partial \log|\widetilde{\mathbf{P}}|}{\partial \mathbf{R}^{-1}} + \frac{\partial \mathrm{tr}\big(\widetilde{\mathbf{P}}^{-1}\overline{\mathbf{P}}\big)}{\partial \mathbf{R}^{-1}}, && \text{(A.13a)} \\
=& -\frac{\partial \log|\mathbf{P}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}|}{\partial \mathbf{R}^{-1}} + \frac{\partial \mathrm{tr}\big((\mathbf{P}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C})\overline{\mathbf{P}}\big)}{\partial \mathbf{R}^{-1}}, && \text{(A.13b)} \\
=& -\frac{\partial \log|\mathbf{P}^{-1}|}{\partial \mathbf{R}^{-1}} - \frac{\partial \log|\mathbf{I} + \mathbf{P}^{1/2}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\mathbf{P}^{1/2}|}{\partial \mathbf{R}^{-1}} + \frac{\partial \mathrm{tr}\big(\mathbf{P}^{-1}\overline{\mathbf{P}}\big)}{\partial \mathbf{R}^{-1}} \\
&+ \frac{\partial \mathrm{tr}\big(\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\overline{\mathbf{P}}\big)}{\partial \mathbf{R}^{-1}} && \text{(A.13c)} \\
\overset{+}{=}& -\frac{\partial \log|\mathbf{I} + \mathbf{R}^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}|}{\partial \mathbf{R}^{-1}} + \frac{\partial \mathrm{tr}\big(\mathbf{R}^{-1}\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\big)}{\partial \mathbf{R}^{-1}}, && \text{(A.13d)} \\
=& -\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{I} + \mathbf{R}^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1} + \mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}. && \text{(A.13e)}
\end{aligned}$$

We equate (A.13e) to zero and solve for $\mathbf{R}^{-1}$ to get

$$\mathbf{I} + (\mathbf{R}^{*})^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} = (\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}, \tag{A.14}$$

which is equivalent to

$$(\mathbf{R}^{*})^{-1} = (\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1} - (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}. \tag{A.15}$$

Assuming that the right hand side of (A.15) is invertible, we get

$$\mathbf{R}^{*} = \big((\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1} - (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\big)^{-1}. \tag{A.16}$$

Substituting (A.15) into (A.8) results in the following expression for $\mathbf{y}^{*}$ after tedious algebra.

$$\mathbf{y}^{*} = \mathbf{R}^{*}\big((\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\bar{\mathbf{x}} - (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\hat{\mathbf{x}}\big), \tag{A.17}$$

a proof of which is given in Appendix G.

A.3 Minimization with respect to the measurement matrix $\mathbf{C}$

By substituting the result (A.12c) into the cost J in (A.6), we can write

$$2J \stackrel{+}{=} \log|\widetilde{\mathbf{P}}| + \mathrm{tr}(\widetilde{\mathbf{P}}^{-1}\overline{\mathbf{P}}) - (\hat{\mathbf{x}}-\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}(\cdot) \tag{A.18a}$$

$$\stackrel{+}{=} -\log|\mathbf{I}+\mathbf{R}^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}| + \mathrm{tr}(\mathbf{R}^{-1}\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}) - (\hat{\mathbf{x}}-\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}(\cdot) \tag{A.18b}$$

where we used the definition of $\widetilde{\mathbf{P}}$ in (A.4a) and the determinant equality

$$|\widetilde{\mathbf{P}}||\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}+\mathbf{R}| = |\mathbf{P}||\mathbf{R}|, \tag{A.19}$$

which implies

$$|\widetilde{\mathbf{P}}| \propto |\mathbf{R}^{-1}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}+\mathbf{R})|^{-1} = |\mathbf{I}+\mathbf{R}^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}|^{-1}. \tag{A.20}$$

Substituting $\mathbf{R} \leftarrow \mathbf{R}^*$ into (A.18b) we obtain

$$\begin{aligned} 2J \stackrel{+}{=} &\log|(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}| + \ell - \mathrm{tr}((\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}) \\ &- (\hat{\mathbf{x}}-\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}(\cdot) \end{aligned} \tag{A.21}$$

Unfortunately, it is difficult (if not impossible) to find an analytical solution for $\mathbf{C}$ using the cost function above. Hence we are going to follow a sub-optimal approach below.

Suppose, for the time being, that $\mathbf{C}$ has only a single row, i.e., the measurement $\mathbf{y}$ is a scalar. Note that in this case the measurement covariance $\mathbf{R}$ becomes a scalar as well. Let us then call $\mathbf{C}$ as $\mathbf{c}$ and $\mathbf{R}$ as $\mathbf{r}$ in this case. It should be clear from the cost (A.21) that it is only the direction of $\mathbf{c}$ that matters for the optimization and the scale can be assigned arbitrarily, i.e., the cost for $\alpha\mathbf{c}$ is the same as the cost for $\mathbf{c}$ for any $\alpha \neq 0$. Hence, we can select $\mathbf{c}$ such that $\mathbf{c}\mathbf{P}\mathbf{c}^{\mathrm{T}} = 1$ without loss of generality. The optimal measurement information $(\mathbf{r}^*)^{-1}$ (i.e., the inverse of the measurement variance) found in the previous part would then become

$$(\mathbf{r}^*)^{-1} = \frac{1}{\mathbf{c}\overline{\mathbf{P}}\mathbf{c}^{\mathrm{T}}} - \frac{1}{\mathbf{c}\mathbf{P}\mathbf{c}^{\mathrm{T}}}. \tag{A.22}$$

We can now try to maximize the information in the measurement by solving the following optimization problem.

$$\mathbf{c}^* = \arg\max_{\mathbf{c}\mathbf{P}\mathbf{c}^{\mathrm{T}}=1} (\mathbf{r}^*)^{-1}. \tag{A.23}$$

It is now not difficult to see that optimal solution $\mathbf{c}^*$ for this optimization problem should be transpose of the normalized generalized eigenvector corresponding to the largest generalized eigenvalue λ^* for the following generalized eigenvalue problem.

$$\mathbf{P}\mathbf{e}_i = \lambda_i \overline{\mathbf{P}}\mathbf{e}_i, \tag{A.24}$$

for $i = 1, \ldots, d_x$. The optimal value of $(\mathbf{r}^*)^{-1}$ would be

$$(\mathbf{r}^*)^{-1} = \frac{\lambda^* - 1}{\mathbf{c}^*\mathbf{P}(\mathbf{c}^*)^{\mathrm{T}}}. \tag{A.25}$$

We see clearly that for the optimal measurement noise variance $\mathbf{r}$ to be positive, we need the largest generalized eigenvalue λ^* to be larger than unity.[1]

In view of the analysis above, we propose here to form the measurement matrix $\mathbf{C}$ out of as many linearly independent generalized eigenvectors (for the generalized eigenvalue problem in (A.24)) with generalized eigenvalues larger than unity as possible.

Suppose now that the generalized eigenvalue problem in (A.24) has been solved and $0 \leq L \leq d_x$ generalized eigenvalues (repetitions allowed) $\lambda_i > 1, i = 1, \ldots, L$, and the corresponding linearly independent normalized generalized eigenvectors $\mathbf{e}_i$ have been found. When we set $\mathbf{C}$ as

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_L \end{bmatrix}^{\mathrm{T}}, \tag{A.26}$$

by the properties of the symmetric generalized eigenvalue problem in (A.24), both $\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}$ and $\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}$ are diagonal matrices given as

$$\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}} = \mathrm{diag}(\mathbf{e}_1^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{e}_1, \ldots, \mathbf{e}_L^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{e}_L), \tag{A.27a}$$

$$\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} = \mathrm{diag}\left(\mathbf{e}_1^{\mathrm{T}}\mathbf{P}\mathbf{e}_1, \ldots, \mathbf{e}_L^{\mathrm{T}}\mathbf{P}\mathbf{e}_L\right). \tag{A.27b}$$

Substituting $\mathbf{C}$ in (A.26) into the cost J in (A.21), we get

$$2J \stackrel{+}{=} \sum_{i=1}^{L} \left(1 + \log \frac{\mathbf{e}_i^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{e}_i}{\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i} - \frac{\mathbf{e}_i^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{e}_i}{\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i} - \frac{(\mathbf{e}_i^{\mathrm{T}}(\hat{\mathbf{x}} - \bar{\mathbf{x}}))^2}{\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i}\right) \tag{A.28}$$

Now using the fact that $\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i = \lambda_i \mathbf{e}_i^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{e}_i$, which can be derived from (A.24), we get

$$2J \stackrel{+}{=} \sum_{i=1}^{L} \left(1 - \log \lambda_i - \frac{1}{\lambda_i} - \frac{(\mathbf{e}_i^{\mathrm{T}}(\hat{\mathbf{x}} - \bar{\mathbf{x}}))^2}{\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i}\right). \tag{A.29}$$

[1] Note that since the matrices involved in the generalized eigenvalue problem in (A.24) are symmetric and positive definite, the generalized eigenvalues λ_i, $i = 1, \ldots, d_x$, are all real and positive. Furthermore, the corresponding generalized eigenvectors $\mathbf{e}_i$, $i = 1, \ldots, d_x$, can be selected to be orthogonal with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \mathbf{x}^{\mathrm{T}}\overline{\mathbf{P}}\mathbf{y}$.

It turns out that each term in the summation above is non-positive. Hence, each additional row of $\mathbf{C}$ in (A.26) reduces the cost J or keeps it constant. Moreover, if one desires to use less rows than L, one can always select the generalized eigenvectors $\mathbf{e}_i$ (corresponding to $\lambda_i > 1$) with the largest scores s_i, defined as

$$s_i \triangleq \log \lambda_i + \frac{1}{\lambda_i} + \frac{(\mathbf{e}_i^{\mathrm{T}}(\hat{\mathbf{x}} - \bar{\mathbf{x}}))^2}{\mathbf{e}_i^{\mathrm{T}}\mathbf{P}\mathbf{e}_i}, \tag{A.30}$$

in order to minimize the cost J the most with the allowed number of rows for $\mathbf{C}$. Substituting $\mathbf{C}$ in (A.26) into $\mathbf{R}^*$ in (A.16), and $\mathbf{y}^*$ in (A.17), we obtain

$$\mathbf{R}^* = \mathrm{diag}\left(\frac{\mathbf{e}_1^{\mathrm{T}}\mathbf{P}\mathbf{e}_1}{\lambda_1 - 1}, \ldots, \frac{\mathbf{e}_L^{\mathrm{T}}\mathbf{P}\mathbf{e}_L}{\lambda_L - 1}\right), \tag{A.31a}$$

$$\mathbf{y}^* = \left[\begin{array}{ccc} \frac{\mathbf{e}_1^{\mathrm{T}}(\lambda_1\bar{\mathbf{x}} - \hat{\mathbf{x}})}{\lambda_1 - 1} & \cdots & \frac{\mathbf{e}_L^{\mathrm{T}}(\lambda_L\bar{\mathbf{x}} - \hat{\mathbf{x}})}{\lambda_L - 1} \end{array}\right]^{\mathrm{T}}, \tag{A.31b}$$

where it should be clear that $\mathbf{R}^* > 0$ when $\lambda_i > 1$ for $i = 1, \ldots, L$.

APPENDIX B

DERIVATIONS USED IN THE FACTOR UPDATES IN JMLS

B.1 Derivations for the Forward Factor, $\rho_n^f(\mathbf{x}_n, r_n)$

When $\gamma_n^f > 0$, the tilted density $\bar{\psi}_n^f(x_n, r_n)$ in (3.13a) can be written as follows.

$$\begin{aligned}\bar{\psi}_n^f(\mathbf{x}_n, r_n) \propto & \left[\pi_n^{b,r_n} \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n}\mathbf{x}_n, \mathbf{R}_n^{b,r_n}\right)\right]^{\gamma_n^f} \mathcal{N}(\mathbf{y}_n; \mathbf{C}_n^{r_n}\mathbf{x}_n, \mathbf{R}_n^{r_n}) \\ & \times \sum_{r_{n-1}} \int_{\mathbf{x}_{n-1}} \left[\pi_{r_{n-1}}^{r_n} \alpha_{n-1}^{f,r_{n-1}} \mathcal{N}(\mathbf{x}_n; \mathbf{A}_n^{r_n}\mathbf{x}_{n-1}, \mathbf{Q}_n^{r_n}) \right. \\ & \left. \times \mathcal{N}\left(\mathbf{x}_{n-1}; \mathbf{m}_{n-1}^{f,r_{n-1}}, \mathbf{P}_{n-1}^{f,r_{n-1}}\right)\right], \qquad \text{(B.1a)} \\ \propto & (\pi_n^{b,r_n})^{\gamma_n^f} |\mathbf{R}_n^{b,r_n}|^{\frac{1-\gamma_n^f}{2}} \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n}\mathbf{x}_n, \frac{\mathbf{R}_n^{b,r_n}}{\gamma_n^f}\right) \mathcal{N}(\mathbf{y}_n; \mathbf{C}_n^{r_n}\mathbf{x}_n, \mathbf{R}_n^{r_n}) \\ & \times \sum_{r_{n-1}} \left[\pi_{r_{n-1}}^{r_n} \alpha_{n-1}^{f,r_{n-1}} \mathcal{N}\left(\mathbf{x}_n; \mathbf{m}_{n,r_{n-1}}^{f,r_n}, \mathbf{P}_{n,r_{n-1}}^{f,r_n}\right)\right], \qquad \text{(B.1b)}\end{aligned}$$

where

$$\mathbf{m}_{n,r_{n-1}}^{f,r_n} \triangleq \mathbf{A}_n^{r_n} \mathbf{m}_{n-1}^{f,r_{n-1}} \qquad \text{(B.2a)}$$

$$\mathbf{P}_{n,r_{n-1}}^{f,r_n} \triangleq \mathbf{A}_n^{r_n} \mathbf{P}_{n-1}^{f,r_{n-1}} (\mathbf{A}_n^{r_n})^{\mathrm{T}} + \mathbf{Q}_n^{r_n}. \qquad \text{(B.2b)}$$

We now define the augmented quantities below.

$$\bar{\mathbf{y}}_n^{r_n} \triangleq \begin{bmatrix} \mathbf{y}_n^{\mathrm{T}} & (\mathbf{y}_n^{b,r_n})^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \qquad \text{(B.3a)}$$

$$\overline{\mathbf{C}}_n^{r_n} \triangleq \begin{bmatrix} (\mathbf{C}_n^{r_n})^{\mathrm{T}} & (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \qquad \text{(B.3b)}$$

$$\overline{\mathbf{R}}_n^{r_n} \triangleq \begin{bmatrix} \mathbf{R}_n^{r_n} & 0 \\ 0 & \frac{1}{\gamma_n^f}\mathbf{R}_n^{b,r_n} \end{bmatrix}, \qquad \text{(B.3c)}$$

which enables us to write

$$\bar{\psi}_n^f(\mathbf{x}_n, r_n) \propto (\pi_n^{b,r_n})^{\gamma_n^f} |\mathbf{R}_n^{b,r_n}|^{\frac{1-\gamma_n^f}{2}} \mathcal{N}\left(\bar{\mathbf{y}}_n^{r_n}; \overline{\mathbf{C}}_n^{r_n}\mathbf{x}_n, \overline{\mathbf{R}}_n^{r_n}\right) \sum_{r_{n-1}} \left[\pi_{r_{n-1}}^{r_n} \alpha_{n-1}^{f,r_{n-1}}\right.$$

$$\times \mathcal{N}(\mathbf{x}_n; \mathbf{m}_{n,r_{n-1}}^{f,r_n}, \mathbf{P}_{n,r_{n-1}}^{f,r_n})\Big], \tag{B.4a}$$

$$= \sum_{r_{n-1}} \bar{\beta}_{n,r_{n-1}}^{f,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_{n,r_{n-1}}^{f,r_n}, \mathbf{V}_{n,r_{n-1}}^{f,r_n}), \tag{B.4b}$$

$$= \bar{\beta}_n^{f,r_n} \sum_{r_{n-1}} \beta_{n,r_{n-1}}^{f,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_{n,r_{n-1}}^{f,r_n}, \mathbf{V}_{n,r_{n-1}}^{f,r_n}), \tag{B.4c}$$

where

$$\bar{\beta}_n^{f,r_n} \triangleq \sum_{r_{n-1}} \bar{\beta}_{n,r_{n-1}}^{f,r_n} \tag{B.5a}$$

$$\beta_{n,r_{n-1}}^{f,r_n} \triangleq \bar{\beta}_{n,r_{n-1}}^{f,r_n} / \bar{\beta}_n^{f,r_n}, \tag{B.5b}$$

$$\bar{\beta}_{n,r_{n-1}}^{f,r_n} \triangleq (\pi_n^{b,r_n})^{\gamma_n^f} |\mathbf{R}_n^{b,r_n}|^{\frac{1-\gamma_n^f}{2}} \alpha_{n-1}^{f,r_{n-1}} \pi_{r_{n-1}}^{r_n} \mathcal{N}\left(\bar{\mathbf{y}}_n^{r_n}; \overline{\mathbf{C}}_n^{r_n} \mathbf{m}_{n,r_{n-1}}^{f,r_n}, \mathbf{S}_{n,r_{n-1}}^{f,r_n}\right), \tag{B.5c}$$

$$\mathbf{S}_{n,r_{n-1}}^{f,r_n} \triangleq \overline{\mathbf{C}}_n^{r_n} \mathbf{P}_{n,r_{n-1}}^{f,r_n} (\overline{\mathbf{C}}_n^{r_n})^{\mathrm{T}} + \overline{\mathbf{R}}_n^{r_n}, \tag{B.5d}$$

$$\mathbf{V}_{n,r_{n-1}}^{f,r_n} \triangleq \left((\mathbf{P}_{n,r_{n-1}}^{f,r_n})^{-1} + (\overline{\mathbf{C}}_n^{r_n})^{\mathrm{T}} (\overline{\mathbf{R}}_n^{r_n})^{-1} \overline{\mathbf{C}}_n^{r_n}\right)^{-1}, \tag{B.5e}$$

$$\mathbf{v}_{n,r_{n-1}}^{f,r_n} \triangleq \mathbf{V}_{n,r_{n-1}}^{f,r_n} \left((\mathbf{P}_{n,r_{n-1}}^{f,r_n})^{-1} \mathbf{m}_{n,r_{n-1}}^{f,r_n} + (\overline{\mathbf{C}}_n^{r_n})^{\mathrm{T}} (\overline{\mathbf{R}}_n^{r_n})^{-1} \bar{\mathbf{y}}_n^{r_n}\right). \tag{B.5f}$$

When $\gamma_n^f > 0$, the density $\psi_n^f(\mathbf{x}_n, r_n)$ in (3.13b) is given as

$$\psi_n^f(\mathbf{x}_n, r_n) \propto (\pi_n^{b,r_n})^{\gamma_n^f} |\mathbf{R}_n^{b,r_n}|^{\frac{1-\gamma_n^f}{2}} \alpha_n^{f,r_n} \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n} \mathbf{x}_n, \frac{\mathbf{R}_n^{b,r_n}}{\gamma_n^f}\right)$$
$$\times \mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}), \tag{B.6a}$$

$$= \beta_n^{f,r_n} \alpha_n^{f,r_n} \mathcal{N}(\mathbf{x}_n, \mathbf{v}_n^{f,r_n}, \mathbf{V}_n^{f,r_n}), \tag{B.6b}$$

where

$$\beta_n^{f,r_n} \triangleq (\pi_n^{b,r_n})^{\gamma_n^f} |\mathbf{R}_n^{b,r_n}|^{\frac{1-\gamma_n^f}{2}} \mathcal{N}\left(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n} \mathbf{m}_n^{f,r_n}, \mathbf{S}_n^{f,r_n}\right), \tag{B.7a}$$

$$\mathbf{S}_n^{f,r_n} \triangleq \mathbf{C}_n^{b,r_n} \mathbf{P}_n^{f,r_n} (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} + \frac{\mathbf{R}_n^{b,r_n}}{\gamma_n^f}, \tag{B.7b}$$

$$\mathbf{V}_n^{f,r_n} \triangleq \left((\mathbf{P}_n^{f,r_n})^{-1} + \gamma_n^f (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} (\mathbf{R}_n^{b,r_n})^{-1} \mathbf{C}_n^{b,r_n}\right)^{-1}, \tag{B.7c}$$

$$\mathbf{v}_n^{f,r_n} \triangleq \mathbf{V}_n^{f,r_n} \left((\mathbf{P}_n^{f,r_n})^{-1} \mathbf{m}_n^{f,r_n} + \gamma_n^f (\mathbf{C}_n^{b,r_n})^{\mathrm{T}} (\mathbf{R}_n^{b,r_n})^{-1} \mathbf{y}_n^{b,r_n}\right). \tag{B.7d}$$

When $\gamma_n^f = 0$, the tilted density $\bar{\psi}_n^f(\cdot, \cdot)$ in (3.13a) turns out to be in the same form as (B.4c) with the parameters between (B.5c)–(B.5f) replaced with the following.

$$\bar{\beta}_{n,r_{n-1}}^{f,r_n} \triangleq \alpha_{n-1}^{f,r_{n-1}} \pi_{r_{n-1}}^{r_n} \mathcal{N}(\mathbf{y}_n; \mathbf{C}_n^{r_n} \mathbf{m}_{n,r_{n-1}}^{f,r_n}, \mathbf{S}_{n,r_{n-1}}^{f,r_n}), \tag{B.8a}$$

$$\mathbf{S}_{n,r_{n-1}}^{f,r_n} \triangleq \mathbf{C}_n^{r_n} \mathbf{P}_{n,r_{n-1}}^{f,r_n} (\mathbf{C}_n^{r_n})^{\mathrm{T}} + \mathbf{R}_n^{r_n} \tag{B.8b}$$

$$\mathbf{V}_{n,r_{n-1}}^{f,r_n} \triangleq \left((\mathbf{P}_{n,r_{n-1}}^{f,r_n})^{-1} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{C}_n^{r_n}\right)^{-1}, \tag{B.8c}$$

$$\mathbf{v}_{n,r_{n-1}}^{f,r_n} \triangleq \mathbf{V}_{n,r_{n-1}}^{f,r_n}\big((\mathbf{P}_{n,r_{n-1}}^{f,r_n})^{-1}\mathbf{m}_{n,r_{n-1}}^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}}(\mathbf{R}_n^{r_n})^{-1}\mathbf{y}_n\big). \tag{B.8d}$$

When $\gamma_n^f = 0$, the density $\psi_n^f(\mathbf{x}_n, r_n)$ in (3.13b) becomes

$$\psi_n^f(\mathbf{x}_n, r_n) = \rho_n^f(\mathbf{x}_n, r_n) = \alpha_{n,r_n}^f \mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}) \tag{B.9}$$

B.2 Derivations for the Backward Factor $q_n^b(\mathbf{x}_n, r_n)$

When $\gamma_n^b > 0$, the tilted density $\bar{\psi}_n^b(x_n, r_n)$ in (3.18a) can be written as follows.

$$\begin{aligned}
\bar{\psi}_n^b(\mathbf{x}_n, r_n) \propto & \left[\alpha_n^{f,r_n}\mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n})\right]^{\gamma_n^b} \sum_{r_{n+1}} \int_{\mathbf{x}_{n+1}} \Big[\pi_{r_n}^{r_{n+1}}\pi_{n+1}^{b,r_{n+1}} \\
& \times \mathcal{N}(\mathbf{x}_{n+1}; \mathbf{A}_{n+1}^{r_{n+1}}\mathbf{x}_n, \mathbf{Q}_{n+1}^{r_{n+1}})\mathcal{N}(\bar{\mathbf{y}}_{n+1}^{r_{n+1}}; \overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{x}_{n+1}, \overline{\mathbf{R}}_{n+1}^{r_{n+1}})\Big] && \text{(B.10a)} \\
\propto & \sum_{r_{n+1}} \Big[(\alpha_n^{f,r_n})^{\gamma_n^b}\pi_{r_n}^{r_{n+1}}\pi_{n+1}^{b,r_{n+1}} \left|\mathbf{P}_n^{f,r_n}\right|^{\frac{1-\gamma_n^b}{2}} \mathcal{N}\left(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b}\right) \\
& \times \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{r_{n+1}}; \overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}}\mathbf{x}_n, \mathbf{S}_{n+1}^{b,r_{n+1}})\Big] && \text{(B.10b)} \\
= & \bar{\beta}_n^{b,r_n} \sum_{r_{n+1}} \beta_{n,r_n}^{b,r_{n+1}} \mathcal{N}(\mathbf{x}_n; \mathbf{v}_{n,r_n}^{b,r_{n+1}}, \mathbf{V}_{n,r_n}^{b,r_{n+1}}) && \text{(B.10c)}
\end{aligned}$$

where the augmented quantities $\bar{\mathbf{y}}_{n+1}^{r_{n+1}}$, $\overline{\mathbf{C}}_{n+1}^{r_{n+1}}$, $\overline{\mathbf{R}}_{n+1}^{r_{n+1}}$ are the same as those defined in (B.3) except that one needs to set $\gamma_{n+1}^f = 1$ in $\overline{\mathbf{R}}_{n+1}^{r_{n+1}}$ and we have

$$\begin{aligned}
\mathbf{S}_{n+1}^{b,r_{n+1}} \triangleq & \overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{Q}_{n+1}^{r_{n+1}}(\overline{\mathbf{C}}_{n+1}^{r_{n+1}})^{\mathrm{T}} + \overline{\mathbf{R}}_{n+1}^{r_{n+1}}, && \text{(B.11a)} \\
\bar{\beta}_n^{b,r_n} \triangleq & \sum_{r_{n+1}} \bar{\beta}_{n,r_n}^{b,r_{n+1}} && \text{(B.11b)} \\
\beta_{n,r_n}^{b,r_{n+1}} \triangleq & \bar{\beta}_{n,r_n}^{b,r_{n+1}} / \bar{\beta}_n^{b,r_n} && \text{(B.11c)} \\
\bar{\beta}_{n,r_n}^{b,r_{n+1}} \triangleq & (\alpha_n^{f,r_n})^{\gamma_n^b}\pi_{r_n}^{r_{n+1}}\pi_{n+1}^{b,r_{n+1}} \left|\mathbf{P}_n^{f,r_n}\right|^{\frac{1-\gamma_n^b}{2}} \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{r_{n+1}}; \overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}}\mathbf{m}_n^{f,r_n}, \overline{\mathbf{S}}_{n+1}^{r_{n+1}}), && \text{(B.11d)} \\
\overline{\mathbf{S}}_{n+1}^{b,r_{n+1}} \triangleq & \mathbf{S}_{n+1}^{b,r_{n+1}} + \overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}}\frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b}(\overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}})^{\mathrm{T}}, && \text{(B.11e)} \\
\mathbf{V}_{n,r_n}^{b,r_{n+1}} \triangleq & \big(\gamma_n^b(\mathbf{P}_n^{f,r_n})^{-1} + (\overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}})^{\mathrm{T}}(\mathbf{S}_{n+1}^{b,r_{n+1}})^{-1}\overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}}\big)^{-1}, && \text{(B.11f)} \\
\mathbf{v}_{n,r_n}^{b,r_{n+1}} \triangleq & \mathbf{V}_{n,r_n}^{b,r_{n+1}}\big(\gamma_n^b(\mathbf{P}_n^{f,r_n})^{-1}\mathbf{m}_n^{f,r_n} + (\overline{\mathbf{C}}_{n+1}^{r_{n+1}}\mathbf{A}_{n+1}^{r_{n+1}})^{\mathrm{T}}(\mathbf{S}_{n+1}^{b,r_{n+1}})^{-1}\bar{\mathbf{y}}_{n+1}^{r_{n+1}}\big)^{-1}. && \text{(B.11g)}
\end{aligned}$$

The density $\psi_n^b(\mathbf{x}_n, r_n)$ in (3.18b) is given as

$$\psi_n^b(\mathbf{x}_n, r_n) \propto \pi_n^{b,r_n}\mathcal{N}(\mathbf{y}_n^{b,r_n}; \mathbf{C}_n^{b,r_n}\mathbf{x}_n, \mathbf{R}_n^{b,r_n})\left[\alpha_n^{f,r_n}\mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n})\right]^{\gamma_n^b} \tag{B.12a}$$

$$\propto(\alpha_n^{f,r_n})^{\gamma_n^b}\pi_n^{b,r_n}|\mathbf{P}_n^{f,r_n}|^{\frac{1-\gamma_n^b}{2}}\mathcal{N}\left(\mathbf{y}_n^{b,r_n};\mathbf{C}_n^{b,r_n}\mathbf{x}_n,\mathbf{R}_n^{b,r_n}\right)$$

$$\times\mathcal{N}\left(\mathbf{x}_n;\mathbf{m}_n^{f,r_n},\frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b}\right) \tag{B.12b}$$

$$\propto\beta_n^{b,r_n}\pi_n^{b,r_n}\mathcal{N}(\mathbf{x}_n;\mathbf{v}_n^{b,r_n}\mathbf{V}_n^{b,r_n}), \tag{B.12c}$$

where

$$\beta_n^{b,r_n}=(\alpha_n^{f,r_n})^{\gamma_n^b}|\mathbf{P}_n^{f,r_n}|^{\frac{1-\gamma_n^b}{2}}\mathcal{N}\left(\mathbf{y}_n^{b,r_n};\mathbf{C}_n^{b,r_n}\mathbf{m}_n^{f,r_n},\mathbf{S}_n^{b,r_n}\right), \tag{B.13a}$$

$$\mathbf{S}_n^{b,r_n}\triangleq\mathbf{C}_n^{b,r_n}\frac{\mathbf{P}_n^{f,r_n}}{\gamma_n^b}(\mathbf{C}_n^{b,r_n})^{\mathrm{T}}+\mathbf{R}_n^{b,r_n}, \tag{B.13b}$$

$$\mathbf{V}_n^{b,r_n}=\left(\gamma_n^b(\mathbf{P}_n^{f,r_n})^{-1}+(\mathbf{C}_n^{b,r_n})^{\mathrm{T}}(\mathbf{R}_n^{b,r_n})^{-1}\mathbf{C}_n^{b,r_n}\right)^{-1}, \tag{B.13c}$$

$$\mathbf{v}_n^{b,r_n}=\mathbf{V}_n^{b,r_n}\left(\gamma_n^b(\mathbf{P}_n^{f,r_n})^{-1}\mathbf{m}_n^{f,r_n}+(\mathbf{C}_n^{b,r_n})^{\mathrm{T}}(\mathbf{R}_n^{b,r_n})^{-1}\mathbf{y}_n^{b,r_n}\right). \tag{B.13d}$$

APPENDIX C

DERIVATIONS USED IN THE FACTOR UPDATES FOR TARGET TRACKING UNDER MEASUREMENT ORIGIN UNCERTAINTY

C.1 Update of the Forward Factor $q_n^f(\mathbf{X}_n, \mathbf{r}_n)$

When $\gamma_n^f > 0$, we rewrite the tilted density in (4.19a) using the definitions in (4.16) as

$$\bar{\psi}_n^f(\mathbf{X}_n, \mathbf{r}_n) \propto \left[\mathcal{N}\left(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \mathbf{R}_n^b\right)\right]^{\gamma_n^f} p(\mathbf{Y}_n \mid \mathbf{X}_n, \mathbf{r}_n) p(\mathbf{r}_n) \sum_{\mathbf{r}_{n-1}} \int_{\mathbf{X}_{n-1}} \Big[\alpha_{n-1}^{f,\mathbf{r}_{n-1}}$$

$$\times \mathcal{N}(\mathbf{X}_n; \overline{\mathbf{A}}\mathbf{X}_{n-1}, \overline{\mathbf{Q}})\, \mathcal{N}(\mathbf{X}_{n-1}; \mathbf{m}_{n-1}^{f,\mathbf{r}_{n-1}}, \mathbf{P}_{n-1}^{f,\mathbf{r}_{n-1}})\Big], \tag{C.1a}$$

$$\propto |\mathbf{R}|^{\frac{1-\gamma_n^f}{2}} \mathcal{N}\left(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \frac{\mathbf{R}_n^b}{\gamma_n^f}\right) p(\mathbf{Y}_n \mid \mathbf{X}_n, \mathbf{r}_n) p(\mathbf{r}_n)$$

$$\times \sum_{\mathbf{r}_{n-1}} \left[\alpha_{n-1}^{f,\mathbf{r}_{n-1}} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_{n,\mathbf{r}_n}^{f,\mathbf{r}_{n-1}}, \mathbf{P}_{n,\mathbf{r}_n}^{f,\mathbf{r}_{n-1}})\right], \tag{C.1b}$$

where

$$\mathbf{m}_{n,\mathbf{r}_n}^{f,\mathbf{r}_{n-1}} \triangleq \overline{\mathbf{A}} \mathbf{m}_{n-1}^{f,\mathbf{r}_{n-1}}, \tag{C.2a}$$

$$\mathbf{P}_{n,\mathbf{r}_n}^{f,\mathbf{r}_{n-1}} \triangleq \overline{\mathbf{A}} \mathbf{P}_{n-1}^{f,\mathbf{r}_{n-1}} \overline{\mathbf{A}}^{\mathrm{T}} + \overline{\mathbf{Q}}. \tag{C.2b}$$

We again form the augmented quantities which are defined below,

$$\overline{\mathbf{y}}_n^{\mathbf{r}_n} \triangleq \left[\begin{array}{cccc} \left(\mathbf{y}_n^b\right)^{\mathrm{T}} & \left(\overline{\mathbf{y}}_n^{\mathbf{r}_n,1}\right)^{\mathrm{T}} & \cdots & \left(\overline{\mathbf{y}}_n^{\mathbf{r}_n,K}\right)^{\mathrm{T}} \end{array}\right],^{\mathrm{T}} \tag{C.3a}$$

$$\overline{\mathbf{y}}_n^{\mathbf{r}_n,k} \triangleq \begin{cases} [\,], & \mathbf{r}_n^k = 0 \\ \mathbf{y}_n^{\mathbf{r}_n^k}, & \mathbf{r}_n^k \neq 0 \end{cases}, \tag{C.3b}$$

$$\overline{\mathbf{C}}_n^{\mathbf{r}_n} \triangleq \left[\begin{array}{cccc} \left(\mathbf{C}_n^b\right)^{\mathrm{T}} & \left(\overline{\mathbf{C}}_n^{\mathbf{r}_n,1}\right)^{\mathrm{T}} & \cdots & \left(\overline{\mathbf{C}}_n^{\mathbf{r}_n,K}\right)^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}, \tag{C.3c}$$

$$\overline{\mathbf{C}}_n^{\mathbf{r}_n,k} \triangleq \begin{cases} [\,], & \mathbf{r}_n^k = 0 \\ \left[\begin{array}{ccc} 0_{1\times(k-1)\cdot d_x} & \mathbf{C} & 0_{1\times(K-k)\cdot d_x} \end{array}\right], & \mathbf{r}_n^k \neq 0 \end{cases}, \tag{C.3d}$$

$$\overline{\mathbf{R}}_n^{\mathbf{r}_n} \triangleq \text{blkdiag}\left[\ \tfrac{1}{\gamma_n^f}\mathbf{R}_n^b,\ \ \overline{\mathbf{R}}_n^{\mathbf{r}_n,1},\ \ \cdots,\ \ \overline{\mathbf{R}}_n^{\mathbf{r}_n,K}\ \right], \tag{C.3e}$$

$$\overline{\mathbf{R}}_n^{\mathbf{r}_n,k} \triangleq \begin{cases} [\ ], & \mathbf{r}_n^k = 0 \\ \mathbf{R}, & \mathbf{r}_n^k \neq 0 \end{cases}, \tag{C.3f}$$

so that we can write

$$\begin{aligned}\bar{\psi}_n^f(\mathbf{X}_n, \mathbf{r}_n) \propto & |\mathbf{R}_n^b|^{\frac{1-\gamma_n^f}{2}} \mathcal{N}\left(\bar{\mathbf{y}}_n^{\mathbf{r}_n}; \overline{\mathbf{C}}_n^{\mathbf{r}_n}\mathbf{X}_n, \overline{\mathbf{R}}_n^{\mathbf{r}_n}\right) p(\mathbf{r}_n) \sum_{\mathbf{r}_{n-1}} \Big[\alpha_{n-1}^{f,\mathbf{r}_{n-1}} \\ & \times \mathcal{N}(\mathbf{X}_n; \mathbf{m}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \mathbf{P}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n})\Big], && \text{(C.4a)} \\ = & \sum_{\mathbf{r}_{n-1}} \bar{\beta}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}), && \text{(C.4b)} \\ = & \bar{\beta}_n^{f,\mathbf{r}_n} \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}) && \text{(C.4c)} \\ \approx & \bar{\beta}_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n, \bar{\mathbf{v}}_n^{f,\mathbf{r}_n}, \overline{\mathbf{V}}_n^{f,\mathbf{r}_n}), && \text{(C.4d)}\end{aligned}$$

where

$$\begin{aligned}\bar{\beta}_n^{f,\mathbf{r}_n} \triangleq & \sum_{\mathbf{r}_{n-1}} \bar{\beta}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} && \text{(C.5a)} \\ \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq & \bar{\beta}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} / \bar{\beta}_n^{f,\mathbf{r}_n}, && \text{(C.5b)} \\ \mathbf{S}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq & \overline{\mathbf{C}}_n^{\mathbf{r}_n} \mathbf{P}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} (\overline{\mathbf{C}}_n^{\mathbf{r}_n})^{\mathrm{T}} + \overline{\mathbf{R}}_n^{\mathbf{r}_n}, && \text{(C.5c)} \\ \bar{\beta}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq & |\mathbf{R}_n^b|^{\frac{1-\gamma_n^f}{2}} \alpha_{n-1}^{f,\mathbf{r}_{n-1}} p(\mathbf{r}_n) \mathcal{N}\left(\bar{\mathbf{y}}_n^{\mathbf{r}_n}; \overline{\mathbf{C}}_n^{\mathbf{r}_n}\mathbf{m}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, \mathbf{S}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}\right), && \text{(C.5d)} \\ \mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq & \left((\mathbf{P}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n})^{-1} + (\overline{\mathbf{C}}_n^{\mathbf{r}_n})^{\mathrm{T}} (\overline{\mathbf{R}}_n^{\mathbf{r}_n})^{-1} \overline{\mathbf{C}}_n^{\mathbf{r}_n}\right)^{-1}, && \text{(C.5e)} \\ \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq & \mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \left((\mathbf{P}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n})^{-1} \mathbf{m}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} + (\overline{\mathbf{C}}_n^{\mathbf{r}_n})^{\mathrm{T}} (\overline{\mathbf{R}}_n^{\mathbf{r}_n})^{-1} \bar{\mathbf{y}}_n^{\mathbf{r}_n}\right) && \text{(C.5f)} \\ \bar{\mathbf{v}}_n^{f,\mathbf{r}_n} \triangleq & \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}, && \text{(C.5g)} \\ \overline{\mathbf{V}}_n^{f,\mathbf{r}_n} \triangleq & \sum_{\mathbf{r}_{n-1}} \beta_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \left[\mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} + (\mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} - \bar{\mathbf{v}}_n^{f,\mathbf{r}_n})(\cdot)^{\mathrm{T}}\right]. && \text{(C.5h)}\end{aligned}$$

The density, $\psi_n^f(\mathbf{X}_n, \mathbf{r}_n)$ in (4.19b) is written for $\gamma_n^f > 0$ as

$$\begin{aligned}\psi_n^f(\mathbf{X}_n, \mathbf{r}_n) \propto & |\mathbf{R}_n^b|^{\frac{1-\gamma_n^f}{2}} \alpha_n^{f,\mathbf{r}_n} \mathcal{N}\left(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{x}_n, \frac{\mathbf{R}_n^b}{\gamma_n^f}\right) \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}), && \text{(C.6a)} \\ = & \beta_n^{f,\mathbf{r}_n} \alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n, \mathbf{v}_n^{f,\mathbf{r}_n}, \mathbf{V}_n^{f,\mathbf{r}_n}), && \text{(C.6b)}\end{aligned}$$

where

$$\beta_n^{f,\mathbf{r}_n} \triangleq |\mathbf{R}_n^b|^{\frac{1-\gamma_n^f}{2}} p(\mathbf{r}_n) \mathcal{N}\left(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{S}_n^{f,\mathbf{r}_n}\right), \tag{C.7a}$$

$$\mathbf{S}_n^{f,\mathbf{r}_n} \triangleq \mathbf{C}_n^b \mathbf{P}_n^{f,\mathbf{r}_n} (\mathbf{C}_n^b)^{\mathrm{T}} + \frac{\mathbf{R}_n^b}{\gamma_n^f}, \tag{C.7b}$$

$$\mathbf{V}_n^{f,\mathbf{r}_n} \triangleq \left((\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} + \gamma_n^f (\mathbf{C}_n^b)^{\mathrm{T}} (\mathbf{R}_n^b)^{-1} \mathbf{C}_n^b \right)^{-1}, \tag{C.7c}$$

$$\mathbf{v}_n^{f,\mathbf{r}_n} \triangleq \mathbf{V}_n^{f,\mathbf{r}_n} \left((\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} \mathbf{m}_n^{f,\mathbf{r}_n} + \gamma_n^f (\mathbf{C}_n^b)^{\mathrm{T}} (\mathbf{R}_n^b)^{-1} \mathbf{y}_n^b \right). \tag{C.7d}$$

When $\gamma_n^f = 0$ and $\mathbf{r}_n^k \neq 0 \; \exists k$, the derivation procedure for the tilted density $\psi_n^f(\cdot, \cdot)$ remains the same except that the parameters of the backward factor vanish from the augmented quantities in (C.3) as follows.

$$\overline{\mathbf{y}}_n^{\mathbf{r}_n} \triangleq \left[\begin{array}{ccc} \left(\overline{\mathbf{y}}_n^{\mathbf{r}_n,1}\right)^{\mathrm{T}} & \cdots & \left(\overline{\mathbf{y}}_n^{\mathbf{r}_n,K}\right)^{\mathrm{T}} \end{array} \right]^{\mathrm{T}}, \tag{C.8a}$$

$$\overline{\mathbf{C}}_n^{\mathbf{r}_n} \triangleq \left[\begin{array}{ccc} \left(\overline{\mathbf{C}}_n^{\mathbf{r}_n,1}\right)^{\mathrm{T}} & \cdots & \left(\overline{\mathbf{C}}_n^{\mathbf{r}_n,K}\right)^{\mathrm{T}} \end{array} \right]^{\mathrm{T}}, \tag{C.8b}$$

$$\overline{\mathbf{R}}_n^{\mathbf{r}_n} \triangleq \text{blkdiag} \left[\begin{array}{ccc} \overline{\mathbf{R}}_n^{\mathbf{r}_n,1}, & \cdots, & \overline{\mathbf{R}}_n^{\mathbf{r}_n,K} \end{array} \right]. \tag{C.8c}$$

If $\gamma_n^f = 0$ and $\mathbf{r}_n^k = 0 \; \forall k$, then we replace the parameters in between (C.5d)–(C.5f) as

$$\bar{\beta}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq \alpha_{n-1}^{f,\mathbf{r}_{n-1}} (1 - P_D)^K {\beta_{FA}}^K \tag{C.9a}$$

$$\mathbf{V}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq \mathbf{P}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n})^{-1}, \tag{C.9b}$$

$$\mathbf{v}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n} \triangleq \mathbf{m}_{n,\mathbf{r}_{n-1}}^{f,\mathbf{r}_n}. \tag{C.9c}$$

When $\gamma_n^f = 0$, the density $\psi_n^f(\mathbf{X}_n, \mathbf{r}_n)$ turns out to be simply

$$\psi_n^f(\mathbf{X}_n, \mathbf{r}_n) = \rho_n^f(\mathbf{X}_n, \mathbf{r}_n) = \alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}). \tag{C.10}$$

Due to the absence of a backward factor in this case, the solution to the M-projection problem becomes

$$\mathbf{P}_n^{f,\mathbf{r}_n} \triangleq \overline{\mathbf{V}}_n^{f,\mathbf{r}_n}, \tag{C.11a}$$

$$\mathbf{m}_n^{f,\mathbf{r}_n} \triangleq \overline{\mathbf{v}}_n^{f,\mathbf{r}_n}, \tag{C.11b}$$

$$\alpha_n^{f,\mathbf{r}_n} \triangleq \overline{\beta}_n^{f,\mathbf{r}_n}. \tag{C.11c}$$

C.2 Update of the Backward Factor $q_n^b(\mathbf{X}_n)$

We write the tilted density, $\bar{\psi}_n^b(\mathbf{X}_n)$ for $\gamma_n^b > 0$ and $\mathbf{r}_n^k \neq 0$, $\exists k$ as

$$\bar{\psi}_n^b(\mathbf{X}_n) \propto \sum_{\mathbf{r}_n} \left[\alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}) \right]^{\gamma_n^b} \sum_{\mathbf{r}_{n+1}} \int_{\mathbf{X}_{n+1}} \left[\mathcal{N}(\mathbf{X}_{n+1}; \overline{\mathbf{A}} \mathbf{X}_n, \overline{\mathbf{Q}}) \right.$$

$$\times \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{x}_{n+1}, \overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}) p(\mathbf{r}_{n+1}) \Big] \tag{C.12a}$$

$$\propto \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \Big[(\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} \left| \mathbf{P}_n^{f,\mathbf{r}_n} \right|^{\frac{1-\gamma_n^b}{2}} p(\mathbf{r}_{n+1}) \mathcal{N}\left(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} \right)$$

$$\times \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \overline{\mathbf{A}} \mathbf{X}_n, \mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}}) \Big] \tag{C.12b}$$

$$= \bar{\beta}_n^b \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \beta_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \mathcal{N}(\mathbf{X}_n; \mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}, \mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}) \tag{C.12c}$$

$$\propto \mathcal{N}(\mathbf{X}_n; \bar{\mathbf{v}}_n^b, \overline{\mathbf{V}}_n^b) \tag{C.12d}$$

where the augmented quantities $\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}$, $\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}}$, $\overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}$ are as in (C.3) and

$$\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}} \triangleq \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{Q} (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}})^{\mathrm{T}} + \overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}, \tag{C.13a}$$

$$\bar{\beta}_n^b \triangleq \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \tag{C.13b}$$

$$\beta_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} / \bar{\beta}_n^b \tag{C.13c}$$

$$\bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq (\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} \left| \mathbf{P}_n^{f,\mathbf{r}_n} \right|^{\frac{1-\gamma_n^b}{2}} \prod_{\substack{0 \le k \le K \\ 0 < r_n^k}} \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \mathbf{m}_n^{f,\mathbf{r}_n}, \overline{\mathbf{S}}_{n+1}^{\mathbf{r}_{n+1}})$$

$$\times \prod_{\substack{1 \le j \le m_n \\ 0 \le k \le K \\ r_n^k \neq j}} \frac{1}{V}, \tag{C.13d}$$

$$\overline{\mathbf{S}}_{n+1}^{b,\mathbf{r}_{n+1}} \triangleq \mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}} + \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}}, \tag{C.13e}$$

$$\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \big(\gamma_n^b (\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} + (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}} (\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}})^{-1} \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \big)^{-1}, \tag{C.13f}$$

$$\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \big(\gamma_n^b (\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} \mathbf{m}_n^{f,\mathbf{r}_n} + (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}} (\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}})^{-1} \bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}} \big)^{-1}, \tag{C.13g}$$

$$\overline{\mathbf{v}}_n^b \triangleq \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \overline{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}, \tag{C.13h}$$

$$\overline{\mathbf{V}}_n^b \triangleq \sum_{\mathbf{r}_{n+1}} \sum_{\mathbf{r}_{n+1}} \overline{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \big(\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} + (\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} - \overline{\mathbf{v}}_n^b)(\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} - \overline{\mathbf{v}}_n^b)^{\mathrm{T}} \big). \tag{C.13i}$$

Note that when $\bar{\psi}_n^b(\mathbf{X}_n) = 1$ and $\mathbf{r}_n^k = 0 \; \forall k$, (C.16f) and (C.16g) changes to

$$\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} = \frac{\mathbf{P}_n^{b,\mathbf{r}_n}}{\gamma_n^f}, \tag{C.14a}$$

$$\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} = \mathbf{m}_n^{b,\mathbf{r}_n}. \tag{C.14b}$$

We write the tilted density, $\bar{\psi}_n^b(\mathbf{X}_n)$ for $\gamma_n^b > 0$and $\mathbf{r}_n^k \neq 0$, $\exists k$ as

$$\bar{\psi}_n^b(\mathbf{X}_n) \propto \sum_{\mathbf{r}_n} \left[\alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}) \right]^{\gamma_n^b} \sum_{\mathbf{r}_{n+1}} \int_{\mathbf{X}_{n+1}} \Big[\mathcal{N}(\mathbf{X}_{n+1}; \overline{\mathbf{A}} \mathbf{X}_n, \overline{\mathbf{Q}})$$

$$\times \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{x}_{n+1}, \overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}) p(\mathbf{r}_{n+1}) \Big] \tag{C.15a}$$

$$\propto \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \Big[(\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} \left| \mathbf{P}_n^{f,\mathbf{r}_n} \right|^{\frac{1-\gamma_n^b}{2}} p(\mathbf{r}_{n+1}) \mathcal{N}\left(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} \right)$$

$$\times \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \overline{\mathbf{A}} \mathbf{X}_n, \mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}}) \Big] \tag{C.15b}$$

$$= \bar{\beta}_n^b \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \beta_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \mathcal{N}(\mathbf{X}_n; \mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}, \mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}) \tag{C.15c}$$

$$\propto \mathcal{N}(\mathbf{X}_n; \bar{\mathbf{v}}_n, \overline{\mathbf{V}}_n) \tag{C.15d}$$

where the augmented quantities $\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}$, $\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}}$, $\overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}$ are as in (C.3) and

$$\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}} \triangleq \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{Q} (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}})^{\mathrm{T}} + \overline{\mathbf{R}}_{n+1}^{\mathbf{r}_{n+1}}, \tag{C.16a}$$

$$\bar{\beta}_n^b \triangleq \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \tag{C.16b}$$

$$\beta_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} / \bar{\beta}_n^b \tag{C.16c}$$

$$\bar{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq (\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} \left| \mathbf{P}_n^{f,\mathbf{r}_n} \right|^{\frac{1-\gamma_n^b}{2}} \prod_{\substack{0 \le k \le K \\ 0 < r_n^k}} \mathcal{N}(\bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}}; \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \mathbf{m}_n^{f,\mathbf{r}_n}, \overline{\mathbf{S}}_{n+1}^{\mathbf{r}_{n+1}})$$

$$\times \prod_{\substack{1 \le j \le m_n \\ 0 \le k \le K \\ r_n^k \ne j}} \frac{1}{V}, \tag{C.16d}$$

$$\overline{\mathbf{S}}_{n+1}^{b,\mathbf{r}_{n+1}} \triangleq \mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}} + \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}}, \tag{C.16e}$$

$$\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \big(\gamma_n^b (\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} + (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}} (\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}})^{-1} \overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A} \big)^{-1}, \tag{C.16f}$$

$$\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \triangleq \mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \big(\gamma_n^b (\mathbf{P}_n^{f,\mathbf{r}_n})^{-1} \mathbf{m}_n^{f,\mathbf{r}_n} + (\overline{\mathbf{C}}_{n+1}^{\mathbf{r}_{n+1}} \mathbf{A})^{\mathrm{T}} (\mathbf{S}_{n+1}^{b,\mathbf{r}_{n+1}})^{-1} \bar{\mathbf{y}}_{n+1}^{\mathbf{r}_{n+1}} \big)^{-1}, \tag{C.16g}$$

$$\overline{\mathbf{v}}_n^b \triangleq \sum_{\mathbf{r}_n} \sum_{\mathbf{r}_{n+1}} \overline{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}}, \tag{C.16h}$$

$$\overline{\mathbf{V}}_n^b \triangleq \sum_{\mathbf{r}_{n+1}} \sum_{\mathbf{r}_{n+1}} \overline{\beta}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} \big(\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} + (\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} - \overline{\mathbf{v}}_n^b)(\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} - \overline{\mathbf{v}}_n^b)^{\mathrm{T}} \big). \tag{C.16i}$$

Note that when $\bar{\psi}_n^b(\mathbf{X}_n) = 1$ and $\mathbf{r}_n^k = 0 \; \forall k$, (C.16f) and (C.16g) changes to

$$\mathbf{V}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} = \frac{\mathbf{P}_n^{b,\mathbf{r}_n}}{\gamma_n^f}, \tag{C.17a}$$

$$\mathbf{v}_{n,\mathbf{r}_n}^{b,\mathbf{r}_{n+1}} = \mathbf{m}_n^{b,\mathbf{r}_n}. \tag{C.17b}$$

The density $\psi_n^b(\mathbf{X}_n)$ is written as

$$\psi_n^b(\mathbf{X}_n) \propto \mathcal{N}(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \mathbf{R}_n^b) \sum_{\mathbf{r}_n} \left[\alpha_n^{f,\mathbf{r}_n} \mathcal{N}(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \mathbf{P}_n^{f,\mathbf{r}_n}) \right]^{\gamma_n^b} \tag{C.18a}$$

$$\propto \mathcal{N}(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \mathbf{R}_n^b) \sum_{\mathbf{r}_n} \left[(\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} |\mathbf{P}_n^{f,\mathbf{r}_n}|^{\frac{1-\gamma_n^b}{2}} \right.$$

$$\left. \times \mathcal{N}\left(\mathbf{X}_n; \mathbf{m}_n^{f,\mathbf{r}_n}, \frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} \right) \right] \tag{C.18b}$$

$$\propto \mathcal{N}(\mathbf{y}_n^b; \mathbf{C}_n^b \mathbf{X}_n, \mathbf{R}_n^b) \mathcal{N}(\mathbf{X}_n; \mathbf{v}_n^b, \mathbf{V}_n^b), \tag{C.18c}$$

where

$$\beta_n^{b,\mathbf{r}_n} = \frac{(\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} |\mathbf{P}_n^{f,\mathbf{r}_n}|^{\frac{1-\gamma_n^b}{2}}}{\sum_{\mathbf{r}_n} (\alpha_n^{f,\mathbf{r}_n})^{\gamma_n^b} |\mathbf{P}_n^{f,\mathbf{r}_n}|^{\frac{1-\gamma_n^b}{2}}}, \tag{C.19a}$$

$$\mathbf{v}_n^b \triangleq \sum_{\mathbf{r}_n} \beta_n^{b,\mathbf{r}_n} \mathbf{m}_n^{f,\mathbf{r}_n}, \tag{C.19b}$$

$$\mathbf{V}_n^b \triangleq \sum_{\mathbf{r}_n} \beta_n^{b,\mathbf{r}_n} \left(\frac{\mathbf{P}_n^{f,\mathbf{r}_n}}{\gamma_n^b} + (\mathbf{m}_n^{f,\mathbf{r}_n} - \mathbf{v}_n^b)(\mathbf{m}_n^{f,\mathbf{r}_n} - \mathbf{v}_n^b)^{\mathrm{T}} \right). \tag{C.19c}$$

APPENDIX D

GAUSSIAN IDENTITIES

D.1 Gaussian Algebra

Lemma 1 *Given two Gaussian densities* $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ *and* $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$*, their product is*

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)\,\mathcal{N}(\mathbf{x};, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) = c\,\mathcal{N}(\mathbf{x}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \tag{D.1}$$

where

$$c = \mathcal{N}(\boldsymbol{\mu}_a; \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_b), \tag{D.2a}$$

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b)^{-1}, \tag{D.2b}$$

$$\hat{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}_a^{-1} + \boldsymbol{\Sigma}_b^{-1})^{-1}. \tag{D.2c}$$

Lemma 2 *If random variables* $\mathbf{x} \in \mathbb{R}^{d_x}$ *and* $\mathbf{y} \in \mathbb{R}^{d_y}$ *have the Gaussian probability distributions*

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}), \tag{D.3}$$

$$\mathbf{y} \mid \mathbf{x} \sim \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{x}, \mathbf{R}), \tag{D.4}$$

then the joint distribution of $\mathbf{x}$*,* $\mathbf{y}$ *and the marginal distribution of* $\mathbf{y}$ *are given as*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}; \begin{pmatrix} \mathbf{m} \\ \mathbf{C}\mathbf{m} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{P}\mathbf{C}^{\mathrm{T}} \\ \mathbf{C}\mathbf{P} & \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R} \end{pmatrix}\right), \tag{D.5}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{m}, \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R}). \tag{D.6}$$

D.2 Moment Matching between a Mixture of Scaled Gaussians and a Scaled Gaussian

Lemma 3 *Consider the joint density* $p(\mathbf{x}, r)$ *defined as*

$$p(\mathbf{x}, r) \triangleq \sum_i \bar{\beta}_i^r \, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_i^r, \boldsymbol{\Sigma}_i^r) \tag{D.7}$$

for $\mathbf{x} \in \mathbb{R}^d$ *and* $r \in \{1, \ldots, R\}$ *where* $\bar{\beta}_i^r \geq 0$ *and* $\sum_i \sum_r \bar{\beta}_i^r = 1$. *The solution of the optimization problem*

$$\{\pi^{r*}, \boldsymbol{\mu}^{r*}, \boldsymbol{\Sigma}^{r*}\}_{r=1}^R = \arg \min_{\{\pi^r, \boldsymbol{\mu}^r, \boldsymbol{\Sigma}^r\}_{r=1}^R} \mathrm{KL}(p(\cdot, \cdot) || q(\cdot, \cdot)), \tag{D.8}$$

where

$$q(\mathbf{x}, r) \triangleq \pi^r \, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}^r, \boldsymbol{\Sigma}^r), \tag{D.9}$$

is given as

$$\pi^{r*} = \bar{\beta}^r, \tag{D.10a}$$

$$\boldsymbol{\mu}^{r*} = \sum_i \beta_i^r \boldsymbol{\mu}_i^r, \tag{D.10b}$$

$$\boldsymbol{\Sigma}^{r*} = \sum_i \beta_i^r (\boldsymbol{\Sigma}_i^r + (\boldsymbol{\mu}_i^r - \boldsymbol{\mu}^{r*})(\boldsymbol{\mu}_i^r - \boldsymbol{\mu}^{r*})), \tag{D.10c}$$

where

$$\bar{\beta}^r \triangleq \sum_i \bar{\beta}_i^r, \qquad \beta_i^r \triangleq \bar{\beta}_i^r / \bar{\beta}^r. \tag{D.11}$$

APPENDIX E

ABEL'S THEOREM [1]

Theorem 4 (Abel's Theorem [1, p. 53]) *Let* $\mathbf{C} \in \mathbb{R}^{\ell \times d}$, $\mathbf{U} \in \mathbb{R}^{d \times (d-\ell)}$ *be two matrices satisfying* $\mathbf{CU} = 0$ *and let* $\mathbf{P} \in \mathbb{R}^{d \times d}$ *be a positive definite matrix. We also assume that* $\mathbf{C}$ *is full row-rank and the augmented square matrix* $\begin{bmatrix} \mathbf{C}^{\mathrm{T}} & \mathbf{P}^{-1}\mathbf{U} \end{bmatrix} \in \mathbb{R}^{d \times d}$ *is full rank and hence invertible. Then,*

$$\mathbf{C}^{\mathrm{T}}(\mathbf{CPC}^{\mathrm{T}})^{-1}\mathbf{C} = \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{P}^{-1}. \tag{E.1}$$

APPENDIX F

PROOF OF (3.9) AND (3.10)

The factor $\rho_n^f(\mathbf{x}_n, r_n)$ defined in (3.7) can be written as follows.

$$\rho_n^f(\mathbf{x}_n, r_n) \propto q_n^{f,r_n}(\mathbf{x}_n, r_n) p(\mathbf{y}_n \mid \mathbf{x}_n, r_n), \tag{F.1a}$$

$$= \pi_n^{f,r_n} \exp\Big(-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n}(\cdot) \Big) \mathcal{N}(\mathbf{y}_n; \mathbf{C}_n^{r_n}\mathbf{x}_n, \mathbf{R}_n^{r_n}), \tag{F.1b}$$

$$= \pi_n^{f,r_n} \exp\Big(-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n}(\cdot) \Big)$$
$$\times \frac{1}{\sqrt{|2\pi\mathbf{R}_n^{r_n}|}} \exp\Big(-\frac{1}{2}(\mathbf{y}_n - \mathbf{C}_n^{r_n}\mathbf{x}_n)^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1}(\cdot) \Big), \tag{F.1c}$$

$$= \frac{\pi_n^{f,r_n}}{\sqrt{|2\pi\mathbf{R}_n^{r_n}|}} \exp\Big(-\frac{1}{2}\Big(\mathbf{y}_n^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n + (\boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n} \Big) \Big)$$
$$\times \exp\Big(-\frac{1}{2}\Big(\mathbf{x}_n^{\mathrm{T}} \big(\boldsymbol{\Phi}_n^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{C}_n^{r_n} \big) \mathbf{x}_n$$
$$- 2\mathbf{x}_n^{\mathrm{T}} \big(\boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n} + (\mathbf{C}_n^{r_n})^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n \big) \Big) \Big), \tag{F.1d}$$

$$= \frac{\pi_n^{f,r_n}}{\sqrt{|2\pi\mathbf{R}_n^{r_n}|}} \exp\Big(-\frac{1}{2}\Big(\mathbf{y}_n^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n + (\boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n} \Big) \Big)$$
$$\times \exp\Big(-\frac{1}{2}\Big(\mathbf{x}_n^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1} \mathbf{x}_n - 2\mathbf{x}_n^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1} \mathbf{m}_n^{f,r_n} \Big) \Big), \tag{F.1e}$$

$$= \frac{\pi_n^{f,r_n}}{\sqrt{|2\pi\mathbf{R}_n^{r_n}|}} \exp\Big(-\frac{1}{2}\Big(\mathbf{y}_n^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n + (\boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n}$$
$$- (\mathbf{m}_n^{f,r_n})^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1} \mathbf{m}_n^{f,r_n} \Big) \Big)$$
$$\times \exp\Big(-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n^{f,r_n})^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1}(\cdot) \Big), \tag{F.1f}$$

$$= \frac{\pi_n^{f,r_n} \sqrt{|2\pi\mathbf{P}_n^{f,r_n}|}}{\sqrt{|2\pi\mathbf{R}_n^{r_n}|}} \exp\Big(-\frac{1}{2}\Big(\mathbf{y}_n^{\mathrm{T}} (\mathbf{R}_n^{r_n})^{-1} \mathbf{y}_n + (\boldsymbol{\mu}_n^{f,r_n})^{\mathrm{T}} \boldsymbol{\Phi}_n^{f,r_n} \boldsymbol{\mu}_n^{f,r_n}$$

$$- (\mathbf{m}_n^{f,r_n})^{\mathrm{T}} (\mathbf{P}_n^{f,r_n})^{-1} \mathbf{m}_n^{f,r_n} \Big) \Big) \mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}), \tag{F.1g}$$

$$\propto \alpha_n^{f,r_n} \mathcal{N}(\mathbf{x}_n; \mathbf{m}_n^{f,r_n}, \mathbf{P}_n^{f,r_n}), \tag{F.1h}$$

which is (3.9), where α_n^{f,r_n}, $\mathbf{m}_n^{f,r_n}$ and $\mathbf{P}_n^{f,r_n}$ are given in (3.10).

APPENDIX G

PROOF OF (A.17)

The expression for $\mathbf{y}^*$ in (A.8) can be written as

$$\mathbf{y}^* = \mathbf{R}\big(\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\big)^{-1}\big(\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\widetilde{\mathbf{P}}\mathbf{P}^{-1}\hat{\mathbf{x}}\big). \tag{G.1}$$

We can rewrite $\widetilde{\mathbf{P}}$ in (A.4a) in covariance form as

$$\widetilde{\mathbf{P}} = \mathbf{P} - \mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}. \tag{G.2}$$

Based on this, we can find the terms $\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}}$ and $\mathbf{C}\widetilde{\mathbf{P}}\mathbf{P}^{-1}$ as follows.

$$\begin{aligned}
\mathbf{C}\widetilde{\mathbf{P}}\mathbf{C}^{\mathrm{T}} &= \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} - \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}, && \text{(G.3a)}\\
&= \big((\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1} + \mathbf{R}^{-1}\big)^{-1}, && \text{(G.3b)}\\
&= \big((\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1}\big)^{-1}, && \text{(G.3c)}\\
&= \mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}, && \text{(G.3d)}\\
\mathbf{C}\widetilde{\mathbf{P}}\mathbf{P}^{-1} &= \mathbf{C}\big(\mathbf{P} - \mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}\big)\mathbf{P}^{-1}, && \text{(G.3e)}\\
&= \big(\mathbf{I} - \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\big)\mathbf{C}, && \text{(G.3f)}\\
&= \mathbf{R}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}, && \text{(G.3g)}\\
&= (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1} + \mathbf{I})^{-1}\mathbf{C}, && \text{(G.3h)}\\
&= (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})^{-1} - \mathbf{I} + \mathbf{I})^{-1}\mathbf{C}, && \text{(G.3i)}\\
&= \mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}, && \text{(G.3j)}
\end{aligned}$$

where we substituted $(\mathbf{R}^*)^{-1}$ in (A.15) for $\mathbf{R}^{-1}$ twice. Substituting the results in (G.3) back into (G.1), we get

$$\begin{aligned}
\mathbf{y}^* &= \mathbf{R}^*\big(\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\big)^{-1}\big(\mathbf{C}\bar{\mathbf{x}} - (\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}})(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\hat{\mathbf{x}}\big), && \text{(G.4a)}\\
&= \mathbf{R}^*\big(\big(\mathbf{C}\overline{\mathbf{P}}\mathbf{C}^{\mathrm{T}}\big)^{-1}\mathbf{C}\bar{\mathbf{x}} - (\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\hat{\mathbf{x}}\big), && \text{(G.4b)}
\end{aligned}$$

which is the same as (A.17).

APPENDIX H

PROOF OF (A.19)

We obtain this determinant identity from the following well-known identity about the Gaussian distributions which is, in a sense, the basis of Kalman filter update expressions.

$$\mathcal{N}(\mathbf{y}; \mathbf{C}\hat{\mathbf{x}}, \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})\,\mathcal{N}(\mathbf{x}; \tilde{\mathbf{x}}, \widetilde{\mathbf{P}}) = \mathcal{N}(\mathbf{y}; \mathbf{C}\mathbf{x}, \mathbf{R})\,\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P}). \tag{H.1}$$

Gaussian distributions above have the scaling terms $|2\pi(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}+\mathbf{R})|^{-1/2}$, $|2\pi\widetilde{\mathbf{P}}|^{-1/2}$, $|2\pi\mathbf{R}|^{-1/2}$ and $|2\pi\mathbf{P}|^{-1/2}$. Equating these scaling terms on both sides would give

$$|2\pi\widetilde{\mathbf{P}}|^{-1/2}|2\pi(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R})|^{-1/2} = |2\pi\mathbf{P}|^{-1/2}|2\pi\mathbf{R}|^{-1/2}, \tag{H.2}$$

which is equivalent to

$$|\widetilde{\mathbf{P}}||\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R}| = |\mathbf{P}||\mathbf{R}|. \tag{H.3}$$

The reviewer is right that the RHS should be the determinant of the matrix $\begin{bmatrix} \mathbf{P} & \mathbf{P}\mathbf{C}^{\mathrm{T}} \\ \mathbf{C}\mathbf{P} & \mathbf{R} + \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} \end{bmatrix}$ but the determinant of this matrix is exactly equal to $|\mathbf{P}||\mathbf{R}|$. We can here use the following identity for the determinant of a block partitioned matrix.

$$\left|\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right| = |\mathbf{A}||\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| \tag{H.4}$$

whenever $\mathbf{A}$ is invertible. If we apply this to the aforementioned matrix, we get

$$\left|\begin{bmatrix} \mathbf{P} & \mathbf{P}\mathbf{C}^{\mathrm{T}} \\ \mathbf{C}\mathbf{P} & \mathbf{R} + \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} \end{bmatrix}\right| = |\mathbf{P}||\mathbf{R} + \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} - \mathbf{C}\mathbf{P}\mathbf{P}^{-1}\mathbf{P}\mathbf{C}^{\mathrm{T}}| \tag{H.5a}$$

$$= |\mathbf{P}||\mathbf{R} + \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} - \mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}}| \tag{H.5b}$$

$$= |\mathbf{P}||\mathbf{R}|. \tag{H.5c}$$

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Sarıtaş, Elif

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Middle East Technical University	2015
B.S.	Middle East Technical University	2012

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2021-Present	ROMER - METU	Lecturer
2012-2021	Middle East Technical University	Research/ Teaching Assistant

PUBLICATIONS

1. E. Sarıtaş, U. Orguner, "Expectation Propagation with Context Adjustment for Smoothing of Jump Markov Linear Systems," IEEE Transactions on Aerospace and Electronic Systems (submitted on May 29, 2022, revised on January 03, 2023.)

International Conference Publications

1. E. Sarıtaş, U. Orguner, “A Random Matrix Measurement Update Using Taylor-Series Approximations", International Conference on Information Fusion (FU-

SION), Cambridge, 2018.

2. E. Sarıtaş, U. Orguner, “Posterior Cramér-Rao Lower Bounds for Extended Target Tracking in a Random Matrix Framework”, International Conference on Information Fusion (FUSION), Heidelberg, 2016.

3. S. Özgen, ¨ E. Sarıtaş, U. Orguner, and D. Acar, “Delay estimation based on kinematic track information without time stamps”, Signal Processing and Communications Applications Conference (SIU), Trabzon, 2014.