DECORRELATION APPROACHES FOR DISTRIBUTED FUSION WITH TARGET TRACKING ALGORITHMS

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

BY

DUYGU ACAR

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

JANUARY 2021

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: Duygu Acar

Signature :

ABSTRACT

DECORRELATION APPROACHES FOR DISTRIBUTED FUSION WITH TARGET TRACKING ALGORITHMS

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

January 2021, 182 pages

The sensor data fusion can be described as merging complementary and/or overlapping information provided by sensors to produce superior information while ensuring consistency in every step of the process. The optimal fusion (centralized fusion) can be done when all measurements from the sensors are made available to fusion center. With modern real-world systems, due to communication constraints, the sensor data are usually first processed in local sensor/fusion nodes and the processed data is sent to other fusion nodes to be fused to produce better quality information. This approach is called track/estimate fusion.

A fusion system where track/estimate fusion is made is called as a decentralised fusion system. The main problem of track fusion is the correlated estimation errors of the local agents. These errors originate from common process noise or common information due to previous communication which is not be easily identifiable. In addition, in noisy (i.e. false alarms, clutter) environment, decorrelated information can not be obtained precisely. Neglecting the correlations between the estimates to be fused has consequences ranging from low fusion performance to filter divergence due to overconfident estimates. In addition, the incompatible cases to be handled can

occur during common information removal.

This thesis proposes decorrelation algorithms to be employed in sensor fusion nodes in a decentralised fusion system using local IMM (Interacting Multiple Model), PDA (Probabilistic Data Association) or combination of these, IMM-PDA filters. Being instances of information decorrelation approaches, the proposed methods attack only the correlations caused by the previous communication. The basic idea is to recognize the common information and then remove it. The application of decorrelation has a significant effect on the filter performance in the fusion node. Information decorrelation is applied either on the Gaussian mixtures or the merged Gaussians produced by the IMM, PDA or IMM-PDA filters.

The decorrelation approaches for the IMM, PDA and IMM-PDA filters are derived and analyzed in different chapters in the thesis along with the corresponding fusion operations. The implementation issues that might arise while applying the decorrelation approach are addressed in detail. The investigated approaches are compared with alternatives on simple distributed single (maneuvering) target tracking examples with various communication rates.

Keywords: Information decorrelation, tracklet, distributed estimation, maneuvering target tracking, track fusion, interacting multiple model, IMM, probabilistic data association, PDA, IMM-PDA

ÖZ

HEDEF İZLEME ALGORİTMALARIYLA DAĞITIK FÜZYON İÇİN İLİNTİSİZLEŞTİRME YAKLAŞIMLARI

Acar, Duygu
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Umut Orguner

Ocak 2021 , 182 sayfa

Sensör bilgi füzyonu, sensörlerden elde edilen çakışan veya tamamlayıcı bilginin, tutarlılığın her işlem adımında sağlandığı bir ortamda, daha üstün bilginin oluşturulması kapsamında birleştirilmesi olarak tarif edilebilir. Optimum füzyon (merkezi füzyon), sensörlerden elde edilen tüm ölçümlerin füzyon merkezine sağlandığı zaman yapılabilir. Modern gerçek sistemlerde, haberleşme kısıtlarından dolayı, sensör bilgisi lokal olarak işlenmekte ve bu işlenmiş veri daha üstün bilgi oluşturmak için diğer füzyon merkezine iletilmektedir. Bu yaklaşım, iz/kestirim füzyonu olarak ifade edilir.

Bu işlemin yapıldığı füzyon sistemi ise merkezi olmayan füzyon sistemi olarak tanımlanır. İz füzyonunun en büyük problemi lokal etmenlerin ilintili kestirim hatalarıdır. Bu hatalar, ortak süreç gürültüleri veya önceki haberleşmeden dolayı oluşan ortak bilgiden kaynaklanır. İlave olarak, gürültülü ortamlarda ilintisiz bilgi kesin olarak elde edilemez. Füzyonu yapılacak verilerin arasındaki ilintilerin göz ardı edilmesi durumunda filtre uyuşmazlığından dolayı kaynaklanan düşük füzyon performansından kendine aşırı güvenen kestirimlerin elde edilmesine kadar birçok olumsuz sonuç oluşabilecektir. İlave olarak, ortak bilginin kaldırılması sırasında oluşabilecek uygun

olmayan durumların ise ortadan kaldırılması gerekmektedir.

Bu tez, lokal olarak IMM (Interacting Multiple Model), PDA (Olasılıksal Veri İlişkilendirme) veya bunların kombinasyonu, IMM-PDA filtreleri kullanan merkezi olmayan bir füzyon sisteminde sensör füzyon düğümlerinde kullanılacak ilintisizleştirme algoritmalarını önermektedir. Bilgi ilintisizleştirme yaklaşımlarının örnekleri olarak, önerilen yöntemler yalnızca önceki iletişimin neden olduğu korelasyonlara işlem yapar. Temel fikir, diğer füzyon merkezinden alınan yerel bilgideki ortak bilginin tespit edilmesi ve hesaplamalardan çıkarılmasıdır. Bilgi İlintisizleştirme uygulaması füzyon merkezindeki filtre performansı üzerinde önemli bir etkiye sahiptir. Bilgi ilintisizleştirme ya Gauss karışımı ya da IMM, PDA veya IMM-PDA filtre tarafından üretilen birleştirilmiş Gauss dağılımlarına uygulanır.

IMM, PDA ve IMM-PDA filtreleri için ilintisizleştirme yaklaşımları, ilgili füzyon işlemleriyle birlikte tezin farklı bölümlerinde türetilmiş ve analiz edilmiştir. İlişkilendirme yaklaşımı uygulanırken ortaya çıkabilecek uygulama sorunları ayrıntılı olarak ele alınmıştır. İncelenen yaklaşımlar, çeşitli haberleşme hızlarına sahip basit dağıtık tek (manevralı) hedef izleme örneklerindeki alternatiflerle karşılaştırılır.

Anahtar Kelimeler: Bilgi dekorelasyonu, tracklet, dağıtık kestirim, manevralı hedef takibi, iz füzyonu, etkileşimli çoklu model, IMM, olasılıksal veri ilişkilendirmesi, PDA, IMM-PDA

To all the people who have taken me forward and whom I believe are by my side,

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my thesis supervisor, Prof. Dr. Umut Orguner, for his support, trust, friendship and encouragement at every stage of this thesis study. It was a great honor for me to work with a supervisor whose academic competence has always impressed me.

I would also like to thank the members of my thesis committee, Prof. Dr. Çağatay Candan and Prof. Dr. Sinan Gezici for their useful perspective, encouragement and suggestions which improved the quality of the thesis study.

I would like to express my gratitude to ASELSAN Inc. for supporting this thesis. I am grateful to my managers at ASELSAN Inc. for the opportunity to improve myself and my colleagues for their support and encouragement.

I would like to thank the members of my thesis defense committee Prof. Dr. Murat Efe and Assoc. Prof. Dr. Emre Özkan for reviewing my thesis and providing useful views.

I wish to express my thanks to Kenan Ahıska who have contributed to the process of review of this dissertation.

I would like to thank Prof. Dr. Mübeccel Demirekler, who improved my academic perspective and played a role in advancing my academic career.

I express my sincerest thanks to my mother Şuran Acar for her endless love, my sister Ebru Acar for her endless patience and support, my father Devlet Acar for his encouragement, Hüseyin İçer for being with me on my sleepless days and my best friend Sema Akbulut Koç for her moral support.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xvi

LIST OF FIGURES xvii

LIST OF ABBREVIATIONS xx

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 1

1.1.1 Motivation 1

1.1.2 Problem Definition 4

1.2 Proposed Methods and Models 5

1.3 Contributions and Novelties 6

1.4 The Outline of the Thesis 7

2 BACKGROUND 11

2.1 Track Fusion 11

2.1.1 Optimal Track Fusion 12

2.1.2 Track Fusion Algorithms 14
2.1.2.1 Linear Combination Methods 15
Naive Fusion (Basic Convex Combination): 15
Linear Combination with Cross Covariance: 16
2.1.2.2 Decorrelation Based Methods 17
Channel Filter (Information Matrix Filter): 17
Restarted Local/Global Filters: 19
Equivalent Measurement: 19
2.1.2.3 Conservative Fusion Methods 19
Chernoff Fusion: 19
Ellipsoidal Intersection (EI): 20
Inverse Covariance Intersection (ICI): 21
2.2 Target Tracking 21
2.2.1 State Estimation 21
2.2.2 Data Association 22
2.2.3 Tracking Algorithms 23
2.2.3.1 Interacting Multiple Model (IMM) Filter 23
2.2.3.2 Probabilistic Data Association (PDA) Filter 25
2.2.3.3 IMM-PDA Filter 30
2.3 Gaussian Algebra 37
2.3.1 Gaussian Multiplication 37
2.3.2 Gaussian Division 39
3 DECORRELATION FOR IMM FILTER 41

3.1 Introduction 41
3.2 Problem Definition 43
3.3 Decorrelation Approach 46
3.3.1 Output Level Decorrelation (OLD) 47
3.3.2 Mode Level Decorrelation (MLD) 49
3.4 Implementation Issues 53
3.4.1 Incompatible Covariance Problem 53
3.4.2 Outlier Decorrelation Problem 57
3.4.3 Solution for Incompatible Covariance and Outlier Decorrelation Problems 59
3.4.3.1 κ for the incompatible covariance problem 60
3.4.3.2 κ for the outlier decorrelation problem 61
3.4.3.3 Use of covariance inflation in OLD & MLD for the IMM filter 63
3.5 Fusion 66
3.6 Simulation Results 68
3.6.1 Model-Match Case 68
3.6.1.1 Comparison of Different OLD & MLD Approaches . . 69
3.6.1.2 Comparison of OLD & MLD with Alternative Approaches 70
3.6.2 Model-Mismatch Case 76
3.7 Conclusion 78
4 DECORRELATION FOR PDA FILTER 85
4.1 Introduction 85
4.2 Problem Definition 86

4.3 Decorrelation Approach 88
4.3.1 Output Level Decorrelation (OLD) 89
4.3.2 Hypothesis Level Decorrelation (HLD) 91
4.4 Implementation Issues 96
4.4.1 Use of covariance inflation in OLD & HLD for the PDA filter . 96
4.5 Fusion 99
4.6 Simulation Results 101
4.6.1 Model-Match Case 101
4.6.2 Model-Mismatch Case 105
4.7 Conclusion 106
5 DECORRELATION FOR IMM-PDA FILTER 113
5.1 Introduction 113
5.2 Problem Definition 115
5.3 Decorrelation Approach 117
5.3.1 Output Level Decorrelation (OLD) 118
5.3.2 Mode Level Decorrelation (MLD) 120
5.3.3 Hypothesis Level Decorrelation (HLD) 123
5.4 Implementation Issues 129
5.4.1 Use of covariance inflation in OLD & MLD & HLD for the IMM-PDA filter 129
5.5 Fusion 134
5.6 Simulation Results 137
5.6.1 Model-Match Case 137

5.6.2 Model-Mismatch Case 141
5.7 Conclusion 142
6 CONCLUSIONS 149
6.1 Future Work 150
REFERENCES 153
A DECORRELATION THEORY 161
A.1 Decorrelation for a Linear Gaussian System 161
A.2 Decorrelation for a Jump Markov Linear System 169
B COMPUTATIONAL COMPLEXITY 175
B.1 IMM Filter - Computational Load 175
B.2 PDA Filter - Computational Load 176
B.3 IMM-PDA Filter - Computational Load 177
B.4 Summary of the Computational Complexity Results 178
CURRICULUM VITAE 181

LIST OF TABLES

TABLES

Table 2.1 A classification of the track fusion methods 15

Table B.1 Summary of the computational complexity results 179

LIST OF FIGURES

FIGURES

Figure 1.1 Centralized fusion. 1

Figure 1.2 Decentralized fusion. 2

Figure 1.3 (a) Fusion architecture. (b) Information graph. 5

Figure 2.1 Single step of the IMM filter. 26

Figure 2.2 Gating illustration. 27

Figure 2.3 Single step of the PDA filter. 29

Figure 2.4 Single step of the IMM-PDA filter. 31

Figure 2.5 Detailed IMM-PDA filter block diagram. 36

Figure 3.1 OLD decorrelation & transmission operations for the IMM filter. 48

Figure 3.2 MLD decorrelation & transmission operations for the IMM filter. 52

Figure 3.3 Illustration of the incompatible covariance problem on the JMLS in (3.19). 56

Figure 3.4 Means and the 99% confidence ellipsoids of the current, predicted and decorrelated quantities in the example for outlier decorrelation problem. The region into which the decorrelated mean has to fall for the outlier decorrelation problem not to happen is shown with the legend “gate”. 58

Figure 3.5 Relationship between the lower bound κ_{OD} and the threshold γ_{OD} for Example 2. 62

Figure 3.6 Current, predicted and decorrelated quantities for different values of the inflation constant κ for Example 2. The solid, dashed and dotted lines denote the predicted covariance, decorrelated covariance and the gate, respectively. 63

Figure 3.7 Fusion operation for the IMM filter. 67

Figure 3.8 Comparison of performances of different OLD and MLD approaches for $N = 3$. 73

Figure 3.9 Restarted LF for the IMM filter. 74

Figure 3.10 IMM Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 3$. 80

Figure 3.11 IMM Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$. 81

Figure 3.12 True trajectory used for the model-mismatch case. The red dot denotes the start of the target trajectory. 82

Figure 3.13 IMM Filter - Model Mismatch Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 3$. 83

Figure 3.14 IMM Filter - Model Mismatch Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$. . 84

Figure 4.1 OLD decorrelation & transmission operations for the PDA filter. 90

Figure 4.2 HLD Decorrelation & transmission operations for the PDA filter. 94

Figure 4.3 Fusion operations for the PDA filter. 100

Figure 4.4 PDA Filter - Model Match Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 3$. 108

Figure 4.5 PDA Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$. 109

Figure 4.6 PDA Filter - Model Mismatch Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 3$. 110

Figure 4.7 PDA Filter - Model Mismatch Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$. . 111

Figure 5.1 OLD decorrelation and transmission operations for the IMM-PDA filter. 119

Figure 5.2 MLD decorrelation and transmission operations for the IMM-PDA filter. 122

Figure 5.3 HLD decorrelation and transmission operations for the IMM-PDA filter. 127

Figure 5.4 Fusion operations for the IMM-PDA filter. 136

Figure 5.5 IMM-PDA Filter - Model Match Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 3$. . . 144

Figure 5.6 IMM-PDA Filter - Model Match Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$. 145

Figure 5.7 IMM-PDA Filter - Model Mismatch Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 3$. . 146

Figure 5.8 IMM-PDA Filter - Model Mismatch Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$. 147

LIST OF ABBREVIATIONS

ABBREVIATIONS

2D	2 Dimensional
CF	Channel Filter
CI	Covariance Intersection
EI	Ellipsoidal Intersection
EKF	Extended Kalman Filter
EMD	Exponential Mixture Density
FPFD	Forward-Prediction Fusion and De-correlation
GCI	Generalized Covariance Intersection
GMD	Geometric Mean Density
HLD	Hypothesis Level Decorrelation
IC	Incompatible Covariance
ICI	Inverse Covariance Intersection
IMF	Information Matrix Filter
IMM	Interacting Multiple Model
JMLS	Jump Markov Linear System
JPDA	Joint Probabilistic Data Association
KF	Kalman Filter
LF	Local Filter
MLD	Mode Level Decorrelation
MHT	Multiple Hypothesis Tracker
NEES	Normalized Estimation Error Squared
NN	Nearest Neighbors

OD	Outlier Decorrelation
OLD	Output Level Decorrelation
OOST	Out-of-Sequence Track
PCCI	Pairwise Component Covariance Intersection
PCF	Pseudo-Chernoff Fusion
PDA	Probabilistic Data Association
PHD	Probability Hypothesis Density
RMS	Root Mean Square
RLF	Restarting Local Filter
WCF	Weighted Covariance Fusion
WLS	Weighted Least Squares

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

1.1.1 Motivation

"*Everything we hear is an opinion, not a fact. Everything we see is a perspective, not the truth.*"

— Marcus Aurelius

The sensor data fusion can be described as merging complementary and/or overlapping information provided by sensors to produce superior information. The optimal fusion can be done when all measurements from the sensors are made available to a fusion center as in Figure 1.1. In modern real-world systems, multiple sensors and

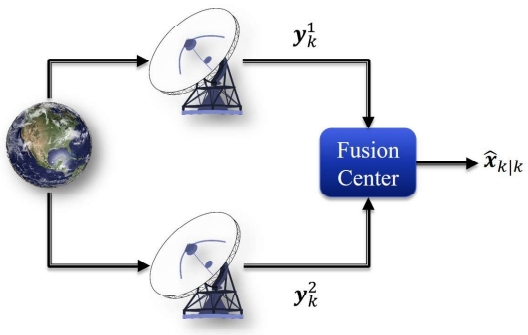


Figure 1.1: Centralized fusion.

fusion centers are physically distributed with perpective of systems of systems. In a distributed target tracking framework, the sensor data are usually first processed in

local sensor/fusion nodes and the processed data are sent to other fusion nodes to be fused due to communication constraints as in Figure 1.2. Therefore, correlations between the estimates/tracks (or estimation errors) to be fused, which are either caused by common process noise [1] or previous communication, are inevitable. When the correlations between the estimates to be fused are neglected, the consequences ranging from low fusion performance to filter divergence occur due to overconfident estimates and/or rumor propagation. The correlations between sensor/fusion nodes in a

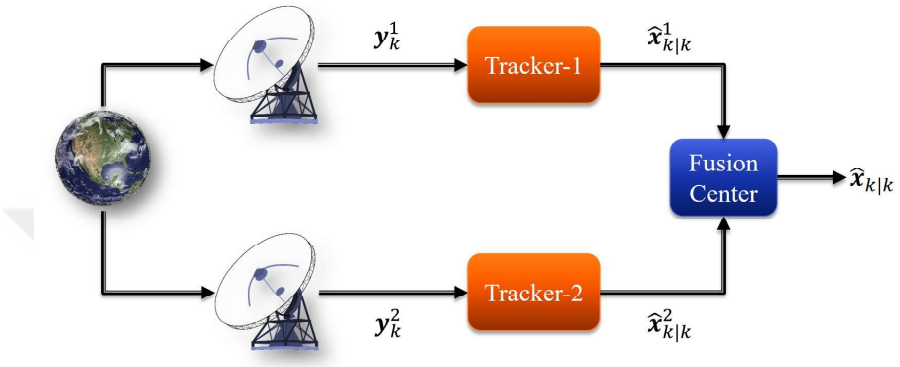


Figure 1.2: Decentralized fusion.

distributed system are handled in three categories in the literature as follows.

- **Linear Combination Methods**

  These methods identify and estimate the common information. For compensation of the effect of common process noise [2, 3], it is required that the approach has extra communication (e.g., Kalman gains). For compensation of the effects of previous communication, a global knowledge of the overall network topology, extra computation or storage resources are required [4, 5]. It should be clear that this type of data is not available on many practical designs.

- **Decorrelation Based Methods**

  "*If you have a past with which you feel dissatisfied, then forget it, now. Imagine a new story for your life and believe in it. Focus only on the moments when you achieved what you desired, and that strength will help you to get what you want.*"

  — Paulo Coelho

The decorrelation methods are applied in some form to either transmitted or received information. These approaches usually attack the correlations caused by previous communication by applying information decorrelation to remove the common information at the local node before transmission. In this case, the transmitted decorrelated quantities are called equivalent-measurements [6] or tracklets [7]. Unlike these methods, for the same purpose, a standard method, which is running a parallel auxiliary local filter which is restarted after every information transmission in addition to a primary/principal local filter, is used to obtain decorrelated estimates. As another point of view of the same category, there are also methods where the decorrelation is applied at receiver center, e.g., information matrix fusion [8, 9], channel filter [10, 11].

- **Conservative Fusion Methods**
  These methods sacrifice independent information conservatively to avoid double counting. The main focus is to preserve (approximate) consistency.[1] The well-known instances of such methods are covariance intersection [12, 13], ellipsoidal intersection [14] and the recent approach called inverse covariance intersection [15]. In these methods, over-conservative results might occur since the correlation amount is not estimated. In addition, these methods have reduced performance compared to the previous two approaches when the correlations can be estimated and compensated [11].

With this motivation in mind, in this thesis, the problem of information decorrelation for compensating for the correlations caused by previous communication between two local agents in a distributed estimation framework where multiple model/hypothesis filters are used in the local and fusion nodes is investigated. While looking for a solution to our problem, the following questions are discussed:

- How is the information decorrelation applied in the local node?
- What information is transmitted to the fusion node?
- How is the transmitted information fused in the fusion node?

[1] See Section 2.1.2 for a mathematical definition of consistency.

- What is the performance of the proposed methods with various transmission periods?

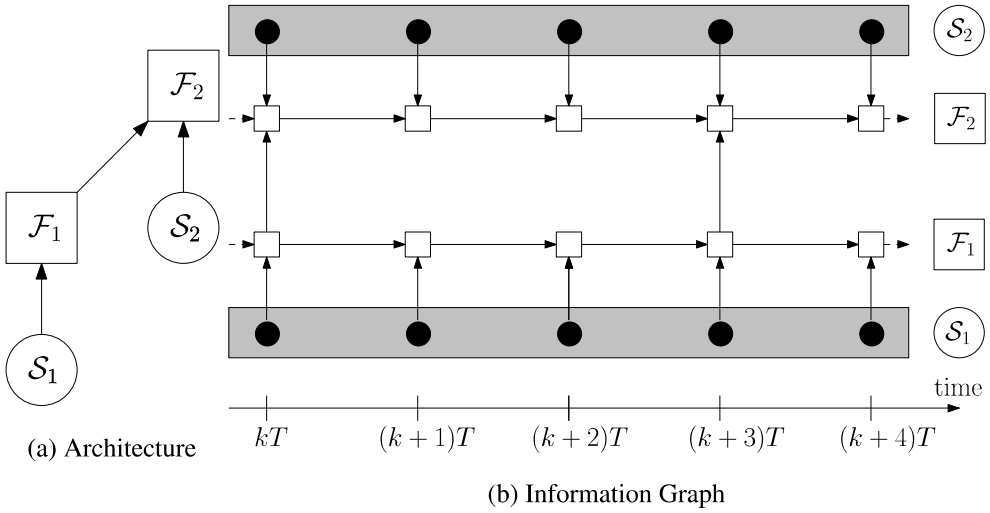
- Is the consistency ensured in all cases with the methods applied?

1.1.2 Problem Definition

The main problem of track fusion is the correlated estimation errors of the local agents. These correlations originate from common process noise or common information due to previous communication, which is not easily identifiable. In addition, in a noisy (i.e. false alarms, clutter) environment, decorrelated information can not be obtained precisely. Neglecting the correlations between the estimates to be fused has consequences ranging from low fusion performance to filter divergence due to overconfident estimates. The idea of this thesis study is to design a decentralized fusion system having multiple sensor fusion nodes using local IMM (Interacting Multiple Model), PDA (Probabilistic Data Association) or a combination of these, IMM-PDA filters with decorrelation approaches.

Two local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ in a sensor network are considered. Local agents achieve their own state estimation with the measurements provided by their local sensors called $\mathcal{S}_1$ and $\mathcal{S}_2$. The fusion architecture and the corresponding information graph taken as base in this thesis are shown in Fig. 1.3a and Fig. 1.3b, respectively. The local agent $\mathcal{F}_1$ communicates its recent information to $\mathcal{F}_2$ with a constant period N. In general, in such fusion architectures, the state estimates of the local agents are correlated due to the common process noise. After $\mathcal{F}_1$ transmits its information to $\mathcal{F}_2$ for the first time, the estimates of the local agents become correlated with previously transmitted estimates to $\mathcal{F}_2$. It is important to note that in this thesis the effects of the process noise will be neglected. In other words, the system models having no or little process noise are considered. Therefore, the correlations between the quantities to be fused are dominated by the previous communication.

The uncompensated correlations due to previous communication might result in inconsistent fused quantities. This situation causes poor performance and/or the divergence of the state estimation algorithm. The process of compensating for the effects



(a) Architecture

(b) Information Graph

Figure 1.3: (a) Fusion architecture. (b) Information graph.

Figure is taken from: M.E. Liggins and Kuo-Chu Chang, "Distributed Fusion Architectures, Algorithms and Performance within a Network-Centric Architecture," 2009. [16]

of previous communication is called as *information decorrelation* or just *decorrelation*.

1.2 Proposed Methods and Models

In this thesis, a novel information decorrelation approach is proposed for multiple model/hypothesis filters. The proposed method is applied on IMM filter, PDA filter and combination of IMM and PDA filters (i.e. IMM-PDA filter). The proposed method is applied at more than one level, starting from the output level and going to the innermost level possible. Applying the proposed method at an internal level reduces the effects of the approximations the filters make and results in improved performance.

The filters and the decorrelation levels considered are;

- Decorrelation for IMM Filter
  - Output level decorrelation (OLD),
  - Mode level decorrelation (MLD).

- Decorrelation for PDA Filter
  - Output level decorrelation (OLD),
  - Hypothesis level decorrelation (HLD).
- Decorrelation for IMM-PDA Filter
  - Output level decorrelation (OLD),
  - Mode level decorrelation (MLD),
  - Hypothesis level decorrelation (HLD).

It has been observed during this thesis study that some implementation issues arise during the implementation of the proposed method. A new method has been proposed for solving these implementation issues: Covariance Inflation (CI). The proposed main method is then accompanied with the covariance inflation method and the results are evaluated as a whole.

1.3 Contributions and Novelties

Our contributions are as follows:

- The information decorrelation approach investigated for the IMM filter was first proposed in,
  D. Acar and U. Orguner, "Information Decorrelation for an Interacting Multiple Model Filter", in *Proceedings of the 21st International Conference on Information Fusion*, Jul. 2018, see [17].
- The implementation issues that might be encountered while applying this decorrelation approach are investigated and the covariance inflation method was first proposed in,
  D. Acar and U. Orguner, "Decorrelation of Previously Communicated Information for an Interacting Multiple Model Filter," *IEEE Transactions on Aerospace and Electronic Systems*, Aug. 2020, see [18].
- A new theory called "*Decorrelation Theory*" was first proposed in [18].

- A comprehensive comparison of the proposed solutions for the implementation issues with alternative ideas that might be used to solve similar problems is presented. It is shown that this novel approach is consistent and can be an alternative of the well-known methods in the literature.

- An overview of the computation load of the investigated decorrelation approaches is provided. It is shown that the proposed methods are quite advantageous in terms of computational complexity.

- The information decorrelation with covariance inflation approach which is proposed in this thesis study is applied to PDA and IMM-PDA filters separately. It is shown that the proposed approach beats the well-known approaches in the literature in terms of all performance criteria.

- The information decorrelation approach applied for the PDA filter will be submitted to International Conference on Information Fusion:
  D. Acar and U. Orguner, "Information Decorrelation for a Probabilistic Data Association Filter".

- The information decorrelation approach applied for the IMM-PDA filter will be submitted to IEEE Transactions on Aerospace and Electronic Systems:
  D. Acar and U. Orguner, "Decorrelation Approaches for Distributed Fusion in a Dense Cluttered Environment".

1.4 The Outline of the Thesis

The rest of the thesis is organized as follows. In Chapter 2, the background information that the reader should possess is given. In this chapter, track fusion and optimum track fusion are introduced. It was emphasized that optimum track fusion cannot be applied in practice and therefore sub-optimal methods should be used. Subsequently, the tracking algorithms are detailed and single steps of IMM, PDA and IMM-PDA algorithms used in this thesis are given in detail. The fusion architecture used in this thesis is also given in this chapter. Furthermore, the reasons for the correlated errors problem that will arise during the application of track fusion algorithms are mentioned and all track fusion algorithms that are used to solve these problems and that

have been introduced in the literature are mentioned.

In Chapter 3, the information decorrelation approach for the IMM filter is introduced. Two decorrelation methodologies (i.e. output level decorrelation (OLD) and mode level decorrelation (MLD)) are proposed. In OLD, the merged output estimates and covariances of the local IMM filter are decorrelated and transmitted to the fusion node. In MLD, unmerged mode-conditioned estimates and covariances of the local IMM filter are decorrelated, merged and transmitted. This chapter, also, investigates and proposes solutions for the implementation issues (i.e. incompatible covariance (IC) problem and outlier decorrelation (OD) problem) that might be encountered while applying the chosen decorrelation approach. These problems are illustrated with an example on a toy problem and with reference to Appendix A Section A.2. In addition, in this part of the thesis, the fusion operations to be applied in the fusion node are described. Furthermore, The numerical results comparing the investigated decorrelation approach with the alternative algorithms on simple distributed maneuvering target tracking examples are presented. Brief remarks on the computational load of the decorrelation methods are given in Appendix B Section B.1. At last, the conclusions are drawn.

In Chapter 4, the information decorrelation approach for the PDA filter is introduced. Two decorrelation methodologies (i.e. output level decorrelation (OLD) and hypothesis level decorrelation (HLD)) are proposed. In OLD, the merged output estimates and covariances of the local PDA filter are decorrelated and transmitted to the fusion node. In HLD, unmerged hypothesis-conditioned estimates and covariances of the local PDA filter are decorrelated, merged and transmitted. The decorrelation theory for the PDA filter with a linear Gaussian system is developed in Appendix A Section A.1. Furthermore, the proposed covariance inflation solution for the IMM filter is applied within OLD and HLD approaches for the PDA filter. In addition, in this chapter, the fusion operations to be applied in the fusion node are described. Furthermore, the numerical results are presented comparing the investigated decorrelation approach with the alternative algorithms on simple distributed target tracking within cluttered area examples. Brief remarks on the computational load of the decorrelation methods are given in Appendix B Section B.2. At last, the conclusions are drawn.

In Chapter 5, the information decorrelation approach for the IMM-PDA filter is introduced. Three decorrelation methodologies (i.e. output level decorrelation (OLD), mode level decorrelation (MLD) and hypothesis level decorrelation (HLD)) are proposed. In OLD, the merged output estimates and covariances of the local IMM-PDA filter are decorrelated and transmitted to the fusion node. In MLD, mode-conditioned estimates and covariances of the local IMM-PDA filter are decorrelated, merged and transmitted. In HLD, mode-hypothesis-conditioned estimates and covariances of the local IMM-PDA filter are decorrelated, merged and transmitted. Furthermore, the proposed covariance inflation solution for the IMM filter is applied within OLD, MLD and HLD approaches for the IMM-PDA filter. In addition, in this chapter, the fusion operations to be applied at the hypothesis level in the fusion node are described. Furthermore, the numerical results comparing the investigated decorrelation approach with the alternative algorithms on simple distributed maneuvering target tracking examples in a cluttered area are presented. Brief remarks on the computational load of the decorrelation methods are given in Appendix B Section B.3. At last, the conclusions are drawn.

In Chapter 6, overall conclusions are drawn for the decorrelation approaches developed for the IMM, PDA and IMM-PDA filters. The importance of information decorrelation is emphasized and it is highlighted that similar methods can be put forward for different types of filters using the ideas presented in the thesis, e.g. JPDA, IMM-JPDA, etc.

CHAPTER 2

BACKGROUND

In this chapter background information that the reader should possess for the remaining parts of the thesis is presented. In Section 2.1, track fusion and optimum track fusion are introduced and track fusion algorithms proposed in the literature are briefly described. In Section 2.2, brief background information about target tracking and single steps of well-known target tracking algorithms (i.e. IMM, PDA and IMM-PDA filters) are given. In Section 2.3, Gaussian algebra used in some of the derivations of the thesis is given.

2.1 Track Fusion

The sensor data fusion can be described as merging complementary and/or overlapping information provided by sensors to produce superior information while ensuring consistency (see Section 2.1.2 for a mathematical definition of consistency) in every step of the process. The optimal fusion can be done when all measurements from the sensors are made available to a fusion center. This is called centralized fusion or measurement fusion.

Modern real-world systems, in which multiple sensors and fusion centers are distributed physically, are designed as systems of systems considering modality, modularity and reliability requirements. With such realistic systems, due to communication constraints, the sensor data are usually first processed in local sensor/fusion nodes and the processed data are sent to other fusion nodes to be fused to produce better quality information. This approach, i.e., communicating/fusing the processed data, is called track/estimate fusion. A fusion system where track/estimate fusion is made is called

as a decentralized fusion system. Similarly, fusion systems using this methodology are said to have a decentralized/distributed fusion architecture.

As stated above, the measurement fusion is the optimal approach. However, due to less strict communication requirements, track/estimate fusion is usually preferred. In a measurement fusion system (i.e., centralized fusion system), the acquired measurements need to be communicated between the nodes at every time instant a measurement is collected. In a decentralized fusion system, on the other hand, the estimates/-tracks might be communicated much less frequently since the sufficient/summary statistics of them are already held in the local nodes. The topic of track fusion is investigated in [5, 4, 19, 20, 10].

2.1.1 Optimal Track Fusion

We consider a simple Bayesian fusion scenario and illustrate the optimal track/estimate fusion technique. Suppose that we have an unknown quantity $\mathbf{x} \in \mathbb{R}^n$. Suppose that two sets of measurements, $\mathbf{Z}_1$ and $\mathbf{Z}_2$, exist providing information about the unknown quantity $\mathbf{x}$. These sets of measurements might be considered to be obtained from different sensors. It is assumed that there might be some common measurements which are in both $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Hence $\mathbf{Z}_1 \cap \mathbf{Z}_2$ is, in general, not an empty set. The aim of track/estimate fusion is to obtain the fused posterior information $p(\mathbf{x}|\mathbf{Z}_1 \cup \mathbf{Z}_2)$ from the locally processed posterior information $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$.

The main assumption in this section is that the collected measurements are conditionally independent given $\mathbf{x}$. This assumption is described mathematically as [4],

$$p(\mathbf{Z}_1 \cup \mathbf{Z}_2|\mathbf{x}) = \frac{p(\mathbf{Z}_1|\mathbf{x})p(\mathbf{Z}_2|\mathbf{x})}{p(\mathbf{Z}_1 \cap \mathbf{Z}_2|\mathbf{x})}. \tag{2.1}$$

The optimal fusion formula expressing the fused posterior information $p(\mathbf{x}|\mathbf{Z}_1 \cup \mathbf{Z}_2)$ in terms of locally processed posterior information $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$ is given as [4],

$$p(\mathbf{x}|\mathbf{Z}_1 \cup \mathbf{Z}_2) = \frac{p(\mathbf{Z}_1 \cup \mathbf{Z}_2|\mathbf{x})p(\mathbf{x})}{p(\mathbf{Z}_1 \cup \mathbf{Z}_2)}, \tag{2.2a}$$

$$\propto \frac{p(\mathbf{x}|\mathbf{Z}_1)p(\mathbf{x}|\mathbf{Z}_2)}{p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2)}. \tag{2.2b}$$

The fused posterior information is obtained by multiplying the locally processed posterior information terms $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$ (just as in the case of Bayes rule) along with a division by the density $p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2)$ which represents the common information in the locally processed posterior information terms $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$. The division by the common information in (2.2) is made to avoid double counting this common information which exist both in $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$.

In practice, even though the locally processed information terms $p(\mathbf{x}|\mathbf{Z}_1)$ and $p(\mathbf{x}|\mathbf{Z}_2)$ are available to a fusion center, the common information term $p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2)$ would not be easily identifiable in a sensor network with a complicated topology. This is the main difficulty for the application of the optimal fusion approach proposed here in real world applications.

When all of the assumed information terms are Gaussian densities, i.e., $p(\mathbf{x}|\mathbf{Z}_1)$, $p(\mathbf{x}|\mathbf{Z}_2)$ and $p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2)$ are Gaussian densities given below,

$$p(\mathbf{x}|\mathbf{Z}_1) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1), \tag{2.3a}$$

$$p(\mathbf{x}|\mathbf{Z}_2) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2), \tag{2.3b}$$

$$p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{1\cap 2}, \mathbf{P}_{1\cap 2}), \tag{2.3c}$$

the optimally fused information $p(\mathbf{x}|\mathbf{Z}_1 \cup \mathbf{Z}_2)$ given by (2.2) has the following analytical solution.

$$p(\mathbf{x}|\mathbf{Z}_1 \cup \mathbf{Z}_2) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{1\cup 2}, \mathbf{P}_{1\cup 2}), \tag{2.4}$$

where

$$(\mathbf{P}_{1\cup 2})^{-1} = (\mathbf{P}_1)^{-1} + (\mathbf{P}_2)^{-1} - (\mathbf{P}_{1\cap 2})^{-1}, \tag{2.5a}$$

$$(\mathbf{P}_{1\cup 2})^{-1}\hat{\mathbf{x}}_{1\cup 2} = (\mathbf{P}_1)^{-1}\hat{\mathbf{x}}_1 + (\mathbf{P}_2)^{-1}\hat{\mathbf{x}}_2 - (\mathbf{P}_{1\cap 2})^{-1}\hat{\mathbf{x}}_{1\cap 2}. \tag{2.5b}$$

Note that the subtractions in the equations above, represent the discount of the common information. The main focus of this thesis study will be obtaining fusion rules when the densities $p(\mathbf{x}|\mathbf{Z}_1)$, $p(\mathbf{x}|\mathbf{Z}_2)$ and $p(\mathbf{x}|\mathbf{Z}_1 \cap \mathbf{Z}_2)$ are Gaussian mixtures estimated in the IMM, PDA and IMM-PDA filters. Note that in the case of Gaussian mixtures, the analytical division in the optimal approach cannot be made. Therefore, sub-optimal approaches have to be pursued.

2.1.2 Track Fusion Algorithms

The main problem of track fusion is the correlated estimation errors of the local agents. There are two types of correlation existent in the sensor fusion systems:

Common Process Noise: While the measurement noise can be assumed to be conditionally independent among different sensors, process noise cannot due to the targets maneuvering behavior which would lead to correlated errors in the different sensors. This type of correlation is named as track spatial dependence in [3]. This topic is investigated also in [2, 5, 1, 9].

Common Prior Estimates: In distributed fusion architectures, there exists communication between the local agents which would lead to correlations between the estimates of communicating agents. This type correlation is named as track temporal dependence in [3]. This topic is investigated also in [5].

The significant requirement of a track fusion algorithm is that it is able to combine information while satisfying a consistency requirement in every step of the fusion process. A more rigorous definition of consistency can be made as follows.

We consider a filter which calculates the estimate $\hat{\mathbf{x}}_k$ of a random vector $\mathbf{x}_k \in \mathbb{R}^{n_x}$ at time k from a set of random measurements $\mathbf{y}_{0:k} \triangleq \{\mathbf{y}_0, \ldots, \mathbf{y}_k\}$. Suppose that the filter also calculates the error covariance $\mathbf{P}_k$ for the estimate $\hat{\mathbf{x}}_k$ at time k. We can now make the following definition of consistency which was taken from [21, Ch. 5.4].

Definition 1 *The filter is called consistent at time k if the estimate $\hat{\mathbf{x}}_k$ and the covariance $\mathbf{P}_k$ satisfies the following conditions.*

$$E\left[\hat{\mathbf{x}}_k - \mathbf{x}_k\right] = 0, \tag{2.6a}$$

$$E\left[(\hat{\mathbf{x}}_k - \mathbf{x}_k)(\hat{\mathbf{x}}_k - \mathbf{x}_k)^{\mathrm{T}}\right] = \mathbf{P}_k. \tag{2.6b}$$

The conditions (2.6a) and (2.6b) are called *unbiasedness* and *covariance consistency*. In this thesis when we refer to the term consistency, we mean the covariance consistency, which is simply the equality of the filter-calculated covariance $\mathbf{P}_k$ to the actual mean square error (MSE) of the estimate $\hat{\mathbf{x}}_k$.

If the fused covariance is larger than the true covariance of the fused estimate, a target tracking/fusion algorithm is called underconfident. If the fused covariance is smaller than the true covariance of the fused estimate, a target tracking/fusion algorithm is called overconfident. It is usually better for a target tracking/fusion algorithm to be underconfident than to be overconfident so we usually would like to have larger fused covariance than the true covariance.

Basically, with Gaussian densities, for a track fusion algorithm to be consistent, the fused covariance should be larger than or equal to the true covariance of the fused estimate. Otherwise, the fusion algorithm can overestimate the information or underestimate the uncertainty.

Track fusion algorithms have been investigated over the past two decades within especially decentralized or distributed fusion architecture. In the literature, there are three approaches to track fusion which are also listed in tabular format in Table 2.1.

Table 2.1: A classification of the track fusion methods

Linear Combination Methods	**Decorrelation Based Methods**	**Conservative Fusion Methods**
Naive Fusion [5, 22, 11] (Basic Convex Combination)	Channel Filter [22, 11] (Information Matrix Filter)	Chernoff Fusion [23]
Linear Combination with Cross Covariance [1]	Restarted Local or Global Filters [5]	Ellipsoidal Intersection [19, 24]
	Equivalent Measurements [5, 25, 6, 26]	Inverse Covariance Intersection [15]

2.1.2.1 Linear Combination Methods

Naive Fusion (Basic Convex Combination): Naive fusion is the simplest fusion approach but sub-optimal because it assumes that the correlation between fused quantities is negligible. This fusion approach is investigated in [5, 22, 11]. The aim of Naive fusion is to obtain the fused posterior information $p(\mathbf{x})$ from the locally processed posterior information $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$. The Naive fusion formula can be

written for density functions as follows,

$$p(\mathbf{x}) = \frac{p_1(\mathbf{x})p_2(\mathbf{x})}{\int p_1(\mathbf{x})p_2(\mathbf{x})d\mathbf{x}}. \tag{2.7}$$

As seen above, the fused information is obtained by multiplying the local information terms $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ as in the case of Bayes rule. The integral $\int p_1(\mathbf{x})p_2(\mathbf{x})d\mathbf{x}$ in (2.7) is a normalizing constant. When the densities, i.e., $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, are Gaussian densities given below,

$$p_1(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1), \tag{2.8a}$$

$$p_2(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2), \tag{2.8b}$$

the fused information $p(\mathbf{x})$ given by (2.7) has the following form

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, P), \tag{2.9}$$

where

$$\mathbf{P}^{-1} = \mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}, \tag{2.10a}$$

$$\mathbf{P}^{-1}\hat{\mathbf{x}} = \mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2. \tag{2.10b}$$

The fusion formulae given above have a simple interpretation. The inverses of the covariances are called as information matrices in the literature. Since the naive fusion formula assumes that the two pieces of information to be fused are independent, it directly adds them. Hence the information matrices are simply added.

As mentioned above, in track fusion systems, as the processed data is fused due to the previous communication between the nodes, the pieces of information to be fused are never independent. Therefore, the use of naive fusion formula given above leads to inconsistency in the estimate covariances, which leads to bad performance or even instability in the state estimators in target trackers. The inconsistency in this case appears as a too small error covariance in the filter compared to true errors of the estimates. In order to obtain better performance the existence of common/correlated information has to be taken into account.

Linear Combination with Cross Covariance: This fusion algorithm accounts for correlations between the local estimates due to common process noise but not correlations caused by common prior estimates or communication. This algorithm explicitly

calculates cross-correlations between the local estimates and for this purpose it needs some additional information (Kalman filter gains etc.) which sometimes cannot be exchanged between the nodes due to the communication constraints. This method is investigated in [1].

2.1.2.2 Decorrelation Based Methods

The algorithms in this category handle double counting of information by calculating and subtracting the common information from local information or sending uncorrelated information between fusion nodes. The information decorrelation algorithms can be used for non-Gaussian probability distributions and they are optimal under no or little process noise assumption. The information decorrelation algorithms account for the correlations caused by previous communications between the nodes. The algorithms given in this section are investigated in [5].

Channel Filter (Information Matrix Filter): The Channel Filter (CF)/Information Matrix Filter (IMF) is one of the information decorrelation approaches. The basic idea is to recognize the common information in the received local information and remove it. This algorithm is also called as the channel filter approach [22, 11]. The algorithm removes only the common prior information due to previous communication. The channel filter fusion equation for density functions is given as,

$$p(\mathbf{x}) = \frac{p_1(\mathbf{x})p_2(\mathbf{x})/\bar{p}(\mathbf{x})}{\int p_1(\mathbf{x})p_2(\mathbf{x})/\bar{p}(\mathbf{x})d\mathbf{x}}, \tag{2.11}$$

where $\bar{p}(\mathbf{x})$ is the density corresponding to the common information. As seen above, the fused posterior information is obtained by multiplying the information terms $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ along with a division by the density $\bar{p}(\mathbf{x})$ which represents the common information (i.e., the common prior information due to previous communication). Suppose now that $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ represent the information in local agent 1 and 2 respectively. When the local agent 1 receives the information $p_2(\mathbf{x})$, it checks whether it received any previous information from local agent 2 in the past. If it did, the density $\bar{p}(\mathbf{x})$ is selected as the predicted probability density function for the previously received information from the local agent 2. The division by this common

information is made to avoid double counting the common information which exist both in $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$. Since, the common information terms $\bar{p}(\mathbf{x})$ are selected to be the (predicted form of) previous information in the same communication channel (from sensor 2 to sensor 1 in the example above), the fusion method described here is called as the channel filter. When the densities are Gaussian, i.e., $p_1(\mathbf{x})$, $p_2(\mathbf{x})$ and $\bar{p}(\mathbf{x})$ are Gaussian densities given below,

$$p_1(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1), \tag{2.12a}$$

$$p_2(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2), \tag{2.12b}$$

$$\bar{p}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{P}}). \tag{2.12c}$$

The fused information $p(\mathbf{x})$ given by (2.11) has the following analytical solution

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P}), \tag{2.13}$$

where

$$\mathbf{P}^{-1} = \mathbf{P}_1^{-1} + \mathbf{P}_2^{-1} - \bar{\mathbf{P}}^{-1}, \tag{2.14a}$$

$$\mathbf{P}^{-1}\hat{\mathbf{x}} = \mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2 - \bar{\mathbf{P}}^{-1}\bar{\mathbf{x}}. \tag{2.14b}$$

The subtraction made above represents the division by the common information. In the thesis, we will focus on channel filter and try to generalize it for Gaussian mixture densities provided by the IMM, PDA and IMM-PDA filter.

In the literature, CF was implemented with another fusion algorithm (covariance intersection filter) in [27] both to overcome spatial and temporal correlations. This algorithm is proposed as split covariance intersection filter-information matrix filter (SCIF-IMF). The graphical model formulation of CF was proposed in [28]. CF was compared with Weighted Covariance Fusion (WCF) in [29] and efficiency and robustness of CF are shown by simulation results. CF was implemented within a hybrid architecture and compared with Weighted Least Square (WLS) estimator in [30]. Simulation results showed that both algorithms have the same computational cost. CF was described with two different levels of common process noise and a nonlinear classifier was used to make self-adjustment in [31] where the method is named as Dual-Band IMF (DBIMF). CF was implemented within distributed fusion architecture with partial feedback and compared with both other memoryless-optimal and

approximate track-to-track fusion algorithms in [32]. Simulation results showed that CF/IMF is consistent and simple to implement. CF was implemented with Forward-Prediction Fusion and De-correlation (FPFD) method to solve Out-of-Sequence Track (OOST) integration problem in [33]. It was shown in [33] that the proposed algorithm is efficient in terms of the computational and memory requirements.

Restarted Local/Global Filters: Restarting the global filter after each fusion process is another solution to the correlation problem due to common prior estimates. In this case, the information decorrelation is not applied; instead the case which causes correlation is tried to be not generated. These approaches are investigated in [5]. It is important to note that these methods do not solve the correlation problem due to common process noise.

Restarting the local filter (Restarted LF) after fusing the last information coming from the same local filter is one of the fusion techniques. Since the local filter is restarted after every fusion step, only low quality estimates are available to the local agent if there is no feedback from the fusion center to the local agent, which is the main disadvantage of this method.

Equivalent Measurement: In this method, local filter calculates and sends the equivalent measurement which is not correlated with the global track. This concept is also named as "tracklets". This method is investigated and related equations are given in [5, 25, 6, 26]. This method is the same as channel filter. Only the place where the decorrelation is applied changes. In the channel filter, the decorrelation is made in the receiving node while in the case of tracklets, the decorrelation is made in the transmitting node. The decorrelation is applied in the transmitting node in this thesis.

2.1.2.3 Conservative Fusion Methods

Chernoff Fusion: If the correlation between two sensor data is totally unknown or we do not have available information to decorrelate the data, Chernoff fusion [23] can

be used [22, 11] for track fusion purposes. The Chernoff fusion formula is given as,

$$p(\mathbf{x}) = \frac{p_1^w(\mathbf{x}) p_2^{1-w}(\mathbf{x})}{\int p_1^w(\mathbf{x}) p_2^{1-w}(\mathbf{x}) d\mathbf{x}}, \tag{2.15}$$

where $w \in [0, 1]$ is a parameter which is selected based on an optimization process with a pre-defined cost function. When the cost function is chosen as Chernoff information which is defined as the denominator of the above equation, this method is named as Chernoff fusion. Although Chernoff fusion can be applied to non-Gaussian/nonlinear density functions [34], the fused density function cannot be described as a known density or with finite number of parameters due to exponential weights. In addition, optimization problem to be solved for finding the w parameter can be too computational costly for the fusion center.

In the literature, the Chernoff fusion algorithm is also known as Generalized Covariance Intersection (GCI) [35, 36, 37], Exponential Mixture Densities (EMD) [38, 39, 40], Geometric Mean Densities (GMDs) [41]. There are some special cases of Chernoff Fusion. In Shannon fusion Chernoff fusion formula [42] is applied while the parameter w is selected to minimize the Shannon entropy of the resulting fused density $p(\mathbf{x})$. Bhattacharyya fusion algorithm [22, 11] is another special case of Chernoff fusion. In this fusion technique, the parameter w is selected to be constant and equal to 0.5. This methodology is analyzed in [11]; and it is seen that the results seem similar to Naive fusion but with larger fused covariance. Covariance Intersection (CI) is the special case of Chernoff fusion when the fused densities are Gaussian densities [28, 36, 43]. Some researchers generalized the CI algorithm to non-Gaussian densities. CI algorithm is extended in [44] for Gaussian mixtures. The proposed solution in [44] is implemented with the idea of the application CI to each cross pair of Gaussian densities in the density multiplication in (2.15). This solution was called as pairwise component CI (PCCI) in [45]. In [45], a superior approach to PCCI was proposed and it was called as pseudo-Chernoff fusion (PCF). In [46, 36], the idea of sigma-points is used to make approximate Chernoff fusion for Gaussian mixtures.

Ellipsoidal Intersection (EI): The covariance intersection method is consistent; however it can overestimate the information in some scenarios which results in performance degradation. Therefore, to handle this disadvantage of the covariance inter-

section method, a fusion algorithm whose name is Largest Ellipsoid [19, 24] is proposed and investigated. The proposed filter underestimates the uncertainty without decrease in track fusion performance. This algorithm is investigated also in [47, 14] with the name ellipsoidal intersection.

Inverse Covariance Intersection (ICI): The well-known instance of conservative fusion methods is the recent approach called inverse covariance intersection [15]. In this proposed approach, ellipsoidal intersection method is extended so that the fused quantities are less conservative by using specific correlation model. However, consistency is not guaranteed when correlation structure varies which is examined in [48].

2.2 Target Tracking

This section gives brief background information about target tracking. Target tracking is the general name given to the state estimation of multiple entities, called as targets, in noisy environments, under measurement origin uncertainty and model mismatch (called as maneuvers) with possibly multiple sensors. A computer program handling the operations in target tracking is called as a target tracker. A target tracker has to deal with two important tasks, namely, state estimation and data association which are investigated in the following subsections.

2.2.1 State Estimation

State estimation is the estimation of kinematic properties (position, velocity etc.) and attributes (type, size etc.) of a target out of noise corrupted measurements. In state estimation for target tracking, state space models for the target motion and the measurement models are utilized. For linear Gaussian estimation problems, Kalman filter (KF) is the optimal algorithm. For nonlinear systems, extended Kalman filter (EKF) is utilized which uses a linearization of the nonlinear state and measurement models. KF and EKF are investigated in [49]. Due to the limitations of KF and EKF, unscented transform is proposed and investigated in [50] which lead to unscented Kalman fil-

ter. With increase in computational resources, more complicated state estimators like multiple model filters [51, 21] and particle filters [49, 52] were proposed and used.

2.2.2 Data Association

A data association algorithm is the part of a tracker which deals with the measurement origin uncertainty. In multiple target tracking applications it is difficult to find the correspondences between the collected measurements from the sensor and the tracks of targets which are estimated. This is called the measurement origin uncertainty in the literature. When there are missing and false reports from the sensor(s), this problem becomes even more difficult to solve. Data association algorithms are approximate solutions for this difficult problem. The single/composite hypothesis data association approaches (like global nearest neighbors (NN) [53, 3] and joint probabilistic data association (PDA) [54, 55]) are the early approaches proposed in the literature for this problem. The NN algorithm selects a single measurement closest to the predicted target from the measurements falling within the validation region. However, in dense cluttered environments, selecting a single measurement may result in discarding the measurement coming from the actual target. In order to prevent this, PDA / JPDA algorithms are generally used in cluttered environments. Hypothesis-based algorithms calculate probabilities of whether all measurements falling within the validation region will come from the real target and find the combined measurement as the weighted average of the measurements falling within the validation region. Since hypothesis-based algorithms contain approximations, the consistency and robustness are investigated in [56]. IMM and (J)PDA algorithms are combined (i.e. IMM-PDA [57]) to ensure continuous target tracking both in dense cluttered environments and in environments with maneuvering targets. It is shown in [58, 59, 60] that the combination of IMM and JPDA algorithms prevents sensitivity to track coalescence problem even with partially measured targets. In [61] a novel IMM/JPDA is proposed for tracking of highly maneuvering targets. To solve the maneuvering target tracking problem within a cluttered environment, joint IMM-PDA particle filter (JIMM-PDAP) [62] is proposed and compared with IMMJPDA based algorithms. With the increase in the computational power, the implementation of multiple hypothesis trackers (MHT) [4, 42], which are more sophisticated approaches to data

association, is now possible. Another fold of developments in multiple target tracking involve the use of the probability hypothesis density (PHD) filters [63] in multiple target tracking which are based on random finite sets.

2.2.3 Tracking Algorithms

This section gives single steps of well-known target tracking algorithms (i.e. IMM, PDA and IMM-PDA filters) used in this thesis.

2.2.3.1 Interacting Multiple Model (IMM) Filter

We consider the following jump Markov linear system (JMLS).

$$\mathbf{x}_k = \mathbf{A}_{r_k}\mathbf{x}_{k-1} + \mathbf{B}_{r_k}\mathbf{w}_{r_k,k}, \tag{2.16a}$$

$$\mathbf{y}_k = \mathbf{C}_{r_k}\mathbf{x}_k + \mathbf{D}_{r_k}\mathbf{v}_{r_k,k}, \tag{2.16b}$$

where $\mathbf{x}_k$ denotes the base state, $\mathbf{y}_k$ denotes the measurement and $r_k \in \{1, \ldots, N_r\}$ denotes the mode index. The mode sequence r_k is assumed to evolve according to a homogeneous Markov chain with transition probability matrix $\mathbf{\Pi} = [\pi_{ji} \triangleq P(r_k = i|r_{k-1} = j)]$. The terms $\mathbf{w}_{r_k,k} \sim \mathcal{N}(\mathbf{w}_{r_k,k}; \mathbf{0}, \mathbf{Q}_{r_k})$ and $\mathbf{v}_{r_k,k} \sim \mathcal{N}(\mathbf{v}_{r_k,k}; \mathbf{0}, \mathbf{R}_{r_k})$ represent process and measurement noise respectively. The models of the form (2.16) are commonly used in maneuvering target tracking, which is a sub-area of target tracking, where the aim is to adaptively estimate the state of the target when the target state evolves with time-varying dynamics.

The state estimation algorithms using jump Markov models are called *multiple model approaches* [21, Section 11.6],[64], whose most well-known instance is the so-called interacting multiple model filter (IMM) [51].
IMM filter [51] is a sub-optimal solution to the state estimation problem for the JMLS in (2.16). At each time step IMM filter approximates the posterior state density $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ as a Gaussian mixture with N_r components.

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}) = \sum_{i=1}^{N_r} \mu_k^i \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i), \tag{2.17}$$

where

- $\mu_k^i \triangleq P(r_k = i|\mathbf{y}_{0:k})$, $i = 1, \ldots, N_r$ are the posterior mode probabilities;
- $\hat{\mathbf{x}}_k^i \triangleq E\left[\mathbf{x}_k|r_k = i, \mathbf{y}_{0:k}\right]$, $i = 1, \ldots, N_r$ are the mode-conditioned state estimates;
- $\mathbf{P}_k^i \triangleq E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\cdot)^{\mathrm{T}}|r_k = i, \mathbf{y}_{0:k}\right]$, $i = 1, \ldots, N_r$ are the mode-conditioned state covariances;

where we use the notation $(\mathbf{x})(\cdot)^{\mathrm{T}}$ denoting the outer product $\mathbf{x}\mathbf{x}^{\mathrm{T}}$ for the sake of brevity. The block diagram related to the IMM filter is illustrated in Figure (2.1) and a single step of the IMM filter is given in Algorithm 1.

Algorithm 1 (A Single Step of the IMM filter) *Suppose that we have the previous summary statistics* $\{\mu_{k-1}^i, \hat{\mathbf{x}}_{k-1}^i, \mathbf{P}_{k-1}^i\}_{i=1}^{N_r}$. *Then, a single step of the IMM filter obtains the current summary statistics* $\{\mu_k^i, \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i\}_{i=1}^{N_r}$ *as follows.*

- ***IMM Prediction Update***
  - ***Mode Probability Prediction:*** *Calculate the predicted mode probabilities* $\{\mu_{k|k-1}^i\}_{i=1}^{N_r}$ *as*

$$\mu_{k|k-1}^i = \sum_{j=1}^{N_r} \pi_{ji} \mu_{k-1}^j. \tag{2.18}$$

  - ***Mixing:***
    - *Calculate the mixing probabilities* $\{\mu_{k-1}^{ji}\}_{i,j=1}^{N_r}$ *as*

$$\mu_{k-1}^{ji} = \frac{\pi_{ji}\mu_{k-1}^j}{\mu_{k|k-1}^i}. \tag{2.19}$$

    - *Calculate the mixed estimates* $\{\hat{\mathbf{x}}_{k-1}^{0i}\}_{i=1}^{N_r}$ *and covariances* $\{\mathbf{P}_{k-1}^{0i}\}_{i=1}^{N_r}$ *as*

$$\hat{\mathbf{x}}_{k-1}^{0i} = \sum_{j=1}^{N_r} \mu_{k-1}^{ji} \hat{\mathbf{x}}_{k-1}^j,$$

$$\mathbf{P}_{k-1}^{0i} = \sum_{j=1}^{N_r} \mu_{k-1}^{ji} \left[\mathbf{P}_{k-1}^j + (\hat{\mathbf{x}}_{k-1}^j - \hat{\mathbf{x}}_{k-1}^{0i})(\cdot)^{\mathrm{T}}\right].$$

- ***Mode Matched Prediction Update:*** *For the ith model, $i = 1, \ldots, N_r$, calculate the predicted estimate $\hat{\mathbf{x}}^i_{k|k-1}$ and covariance $\mathbf{P}^i_{k|k-1}$ from the mixed estimate $\hat{\mathbf{x}}^{0i}_{k-1}$ and covariance $\mathbf{P}^{0i}_{k-1}$ as*

$$\hat{\mathbf{x}}^i_{k|k-1} = \mathbf{A}_i \hat{\mathbf{x}}^{0i}_{k-1}, \tag{2.20a}$$

$$\mathbf{P}^i_{k|k-1} = \mathbf{A}_i \mathbf{P}^{0i}_{k-1} \mathbf{A}_i^{\mathrm{T}} + \mathbf{B}_i Q_i \mathbf{B}_i^{\mathrm{T}}. \tag{2.20b}$$

- ***IMM Measurement Update***
  - ***Mode-Matched Measurement Update:*** *For the ith model, $i = 1, \ldots, N_r$, calculate the updated estimate $\hat{\mathbf{x}}^i_k$ and covariance $\mathbf{P}^i_k$ from the predicted estimate $\hat{\mathbf{x}}^i_{k|k-1}$ and covariance $\mathbf{P}^i_{k|k-1}$ as*

$$\hat{\mathbf{x}}^i_k = \hat{\mathbf{x}}^i_{k|k-1} + \mathbf{K}^i_k (\mathbf{y}_k - \hat{\mathbf{y}}^i_{k|k-1}), \tag{2.21a}$$

$$\mathbf{P}^i_k = \mathbf{P}^i_{k|k-1} - \mathbf{K}^i_k \mathbf{S}^i_k \mathbf{K}^{iT}_k, \tag{2.21b}$$

$$\hat{\mathbf{y}}^i_{k|k-1} = \mathbf{C}_i \hat{\mathbf{x}}^i_{k|k-1}, \tag{2.21c}$$

$$\mathbf{S}^i_k = \mathbf{C}_i \mathbf{P}^i_{k|k-1} \mathbf{C}_i^{\mathrm{T}} + \mathbf{D}_i \mathbf{R}_i \mathbf{D}_i^{\mathrm{T}}, \tag{2.21d}$$

$$\mathbf{K}^i_k = \mathbf{P}^i_{k|k-1} \mathbf{C}_i^{\mathrm{T}} (\mathbf{S}^i_k)^{-1}. \tag{2.21e}$$

  - ***Mode Probability Update:*** *Calculate the updated mode probabilities $\{\mu^i_k\}_{i=1}^{N_r}$ as*

$$\mu^i_k = \frac{\mu^i_{k|k-1} \mathcal{N}(\mathbf{y}_k; \mathbf{y}^i_{k|k-1}, \mathbf{S}^i_k)}{\sum_{\ell=1}^{N_r} \mu^\ell_{k|k-1} \mathcal{N}(\mathbf{y}_k; \mathbf{y}^\ell_{k|k-1}, \mathbf{S}^\ell_k)}. \tag{2.22}$$

- ***Output Estimate Calculation:*** *Calculate the overall estimate $\hat{\mathbf{x}}_k$ and covariance $\mathbf{P}_k$ as*

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_r} \mu^i_k \hat{\mathbf{x}}^i_k, \tag{2.23a}$$

$$\mathbf{P}_k = \sum_{i=1}^{N_r} \mu^i_k \left[\mathbf{P}^i_k + (\hat{\mathbf{x}}^i_k - \hat{\mathbf{x}}_k)(\cdot)^{\mathrm{T}} \right]. \tag{2.23b}$$

2.2.3.2 Probabilistic Data Association (PDA) Filter

We consider the following linear Gaussian system,

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_k, \tag{2.24a}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{v}_k. \tag{2.24b}$$

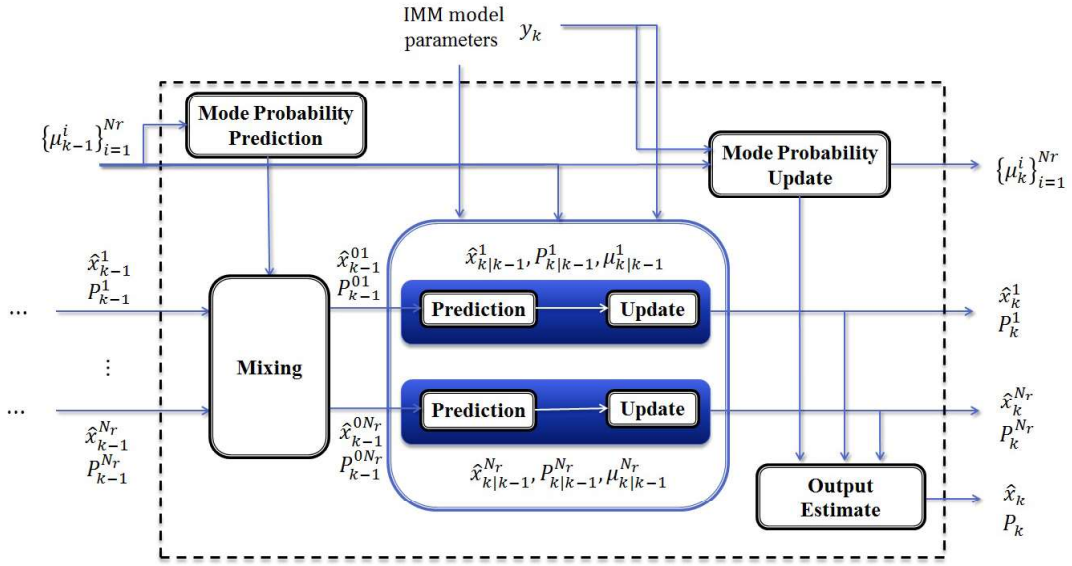


Figure 2.1: Single step of the IMM filter.

where

- $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the base state,
- $\mathbf{y}_k \in \mathbb{R}^{n_y}$ is the measurement,
- $\mathbf{w}_k \in \mathbb{R}^{n_w}$ is the process noise distributed with $\mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q})$,
- $\mathbf{v}_k \in \mathbb{R}^{n_v}$ is the measurement noise distributed with $\mathbf{v}_k \sim \mathcal{N}(\mathbf{v}_k; \mathbf{0}, \mathbf{R})$,

Measurements used in tracking systems not only may have originated from target, but also from false alarms, clutters. Therefore, the validation procedure which limits the region in the measurement space is needed. Measurement space is a space that we desire to find the measurement from the target of interest. In the validation procedure, the validated region is determined. This region can also be defined as gate or association region. Measurements outside the validation region which are unlikely to have originated from the target of interest can be ignored. This holds if the gate probability is close to unity and the model used to obtain the gate is correct. However, in the validation region, more than one detection i.e. several measurements may exist. The problem of tracking a single target in clutter considers the situation where there are possibly several measurements in the validation region (gate) of a target. It is assumed that the measurement contains all the information that could be used

to discard the undesirable measurements. Therefore, any measurement that has been validated could have originated from the target of interest.

The validation region is a two dimensional ellipsoid region. It is centered at the predicted measurement. Example validation regions can be seen in Figure 2.2 where $\hat{\mathbf{y}}^i_{k|k-1}$ is the predicted position for the ith target and $\mathbf{y}^j_k$ is the jth measurement. The

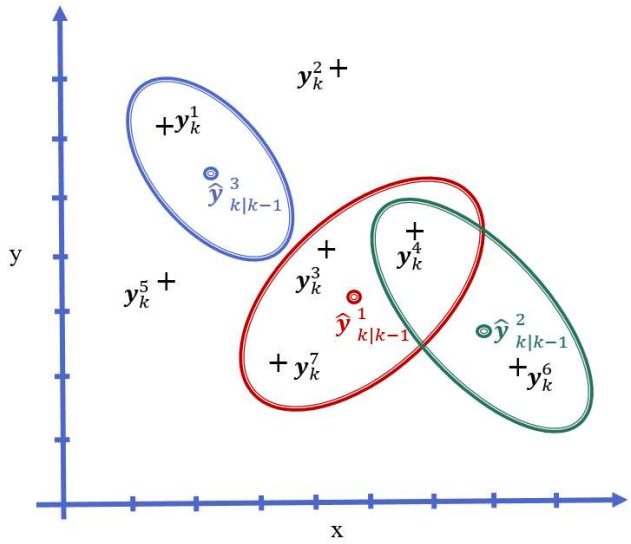


Figure 2.2: Gating illustration.

validation region is basically a form of distance measuring between the predicted measurement and the measurements depending on the target uncertainty.

$$(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^T \mathbf{S}^{-1}_{k|k-1} (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \leq \gamma_G, \tag{2.25}$$

where γ_G is the gate threshold. When the target state and measurement models are correct,

$$\tilde{\mathbf{y}}_k \triangleq \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1} \sim \mathcal{N}(\mathbf{0}_{n_y}, \mathbf{S}_{k|k-1}). \tag{2.26}$$

Hence $(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \mathbf{S}^{-1}_{k|k-1} (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})$ is χ^2 distributed with n_y degrees of freedom.

By using χ^2 distribution, the gate threshold can easily be found. For example, when degrees of freedom is 2 and gate probability is 0.9, then the gate threshold would be approximately 4.7.

Measurements in the gate at time k are shown as $\mathbf{Y}_k = \{\mathbf{y}^i_k\}_{i=1}^{m_k}$. There are the following hypotheses about these measurements,

- $(\theta_k = 0) \triangleq$ All of $\mathbf{Y}_k$ is false alarm i.e. no target originated measurement in the gate.

- $(\theta_k = i) \triangleq$ Measurement $\mathbf{y}_k^i$ belongs to target, all the rest are false alarms.

for $i = 1, ..., m_k$. The problem is to estimate the posterior distribution $p(\mathbf{x}_k|\mathbf{Y}_{0:k})$.

A Bayesian methodology is used in the PDA filter [54], which accounts for the measurement origin uncertainty. This association method is the soft version of Nearest Neighbor association method [53]. The PDA filter does not make a hard decision. It uses all of the measurements in the gate to the extent that they suit the prediction. The PDA filter calculates the association probabilities to the target being tracked for each validated measurement at the current time.
At each time step PDA filter approximates the posterior state density $p(\mathbf{x}_k|\mathbf{Y}_{0:k})$ as a Gaussian mixture with $m_k + 1$ components.

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}) = \sum_{i=0}^{m_k} \mu_k^i \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i), \tag{2.27}$$

where

- $\mu_k^i \triangleq P(\theta_k = i|\mathbf{Y}_{0:k})$, $i = 0, \ldots, m_k$ are the posterior hypothesis probabilities;

- $\hat{\mathbf{x}}_k^i \triangleq E[\mathbf{x}_k|\theta_k = i, \mathbf{Y}_{0:k}]$, $i = 0, \ldots, m_k$ are the hypothesis-conditioned state estimates;

- $\mathbf{P}_k^i \triangleq E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\cdot)^{\mathrm{T}}|\theta_k = i, \mathbf{Y}_{0:k}\right]$, $i = 0, \ldots, m_k$ are the hypothesis-conditioned state covariances;

where we use the notation $(\mathbf{x})(\cdot)^{\mathrm{T}}$ denoting the outer product $\mathbf{x}\mathbf{x}^{\mathrm{T}}$ for the sake of brevity. The block diagram related to the PDA filter is illustrated in Figure (2.3) and a single step of the PDA filter is given in Algorithm 2.

Algorithm 2 (A Single Step of the PDA filter) *Suppose that we have the previous summary statistics* $\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1}$. *Then, a single step of the PDA filter obtains the current summary statistics* $\hat{\mathbf{x}}_k, \mathbf{P}_k$ *as follows.*

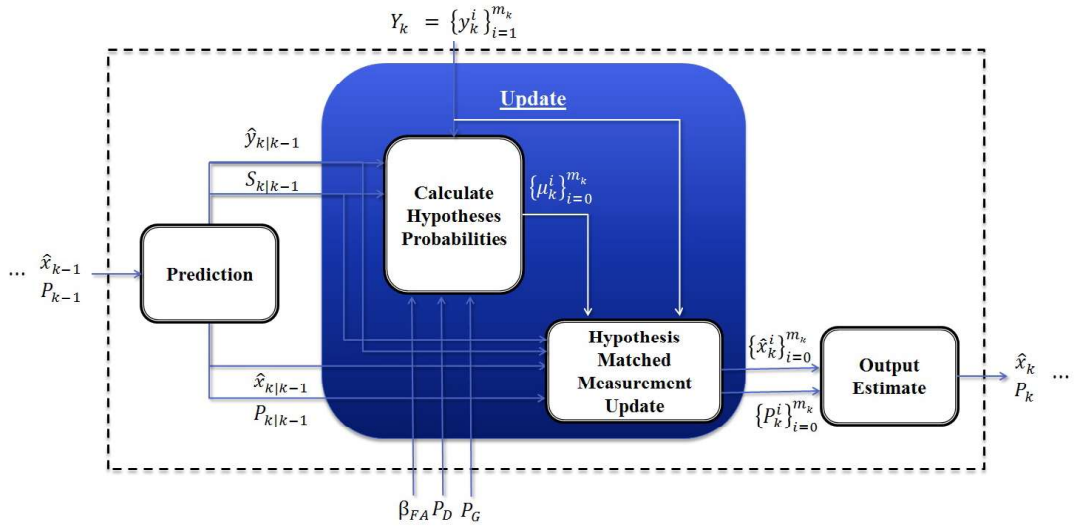


Figure 2.3: Single step of the PDA filter.

- ***PDA Prediction Update:*** *Calculate the predicted estimate* $\hat{\mathbf{x}}_{k|k-1}$ *and covariance* $\mathbf{P}_{k|k-1}$ *from the previous estimate* $\hat{\mathbf{x}}_{k-1}$ *and covariance* $\mathbf{P}_{k-1}$ *as*

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}\hat{\mathbf{x}}_{k-1}, \tag{2.28}$$

$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{Q}\mathbf{B}^{\mathrm{T}}. \tag{2.29}$$

- ***PDA Measurement Update***

  - ***Hypothesis Probability Calculation:*** *Calculate the hypotheses probabilities* $\{\mu_k^i\}_{i=0}^{m_k}$ *as*

$$\mu_k^i \propto \begin{cases} (1 - P_D P_G)\beta_{FA}, & i = 0 \quad \text{(2.30a)} \\ P_D p_{k|k-1}(\mathbf{y}_k^i), & \textit{otherwise} \quad \text{(2.30b)} \end{cases}$$

  where $p_{k|k-1}(\mathbf{y}_k) = \mathcal{N}(\mathbf{y}_k; \hat{\mathbf{y}}_{k|k-1}, \mathbf{S}_{k|k-1})$ *is the innovation (measurement prediction) likelihood* β_{FA} *is the false alarm rate,* P_G *is the gate probability which is defined as the probability that the target-originated measurement falls in the validation region and* P_D *is the probability of detection.*

  - ***Hypothesis-Matched Measurement Update:*** *For the* i*th hypothesis,* $i = 1, ..., m_k$*, calculate the updated estimate* $\hat{\mathbf{x}}_{k|k}^i$ *and covariance* $\mathbf{P}_{k|k}^i$ *from the predicted estimate* $\hat{\mathbf{x}}_{k|k-1}$ *and covariance* $\mathbf{P}_{k|k-1}$ *as follows. In the*

case of a linear Gaussian system,

$$\hat{\mathbf{x}}^i_{k|k} = \begin{cases} \hat{\mathbf{x}}_{k|k-1}, & i = 0 \quad (2.31a) \\ \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}^i_k - \hat{\mathbf{y}}_{k|k-1}), & \textit{otherwise} \quad (2.31b) \end{cases}$$

$$\mathbf{P}^i_{k|k} = \begin{cases} \mathbf{P}_{k|k-1}, & i = 0 \quad (2.32a) \\ \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_{k|k-1}\mathbf{K}^T_k, & \textit{otherwise} \quad (2.32b) \end{cases}$$

Note that the quantities $\mathbf{P}^i_{k|k}$ and $\mathbf{K}_k$, are the same for $i = 1, ..., m_k$.

- ***Output Estimate Calculation:*** *Calculate the overall estimate $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ as below. In a linear Gaussian system case, the overall state estimate $\hat{\mathbf{x}}_{k|k}$ can be calculated as*

$$\hat{\mathbf{x}}_{k|k} = \sum_{i=0}^{m_k} \mu^i_k \hat{\mathbf{x}}^i_{k|k} = \hat{\mathbf{x}}^i_{k|k-1} + \mathbf{K}_k(\mathbf{y}^{eq}_k - \hat{\mathbf{y}}_{k|k-1}), \quad (2.33)$$

where

$$\mathbf{y}^{eq}_k = \mu^0_k \hat{\mathbf{y}}_{k|k-1} + \sum_{i=1}^{m_k} \mu^i_k \mathbf{y}^i_k, \quad (2.34)$$

*is the **equivalent** measurement.*

The covariance $\mathbf{P}_{k|k}$ corresponding to $\hat{\mathbf{x}}_{k|k}$ is given by

$$\mathbf{P}_{k|k} = \sum_{i=0}^{m_k} \mu^i_k [\mathbf{P}^i_{k|k} + (\hat{\mathbf{x}}^i_{k|k} - \hat{\mathbf{x}}_{k|k})(\hat{\mathbf{x}}^i_{k|k} - \hat{\mathbf{x}}_{k|k})^T]. \quad (2.35)$$

2.2.3.3 IMM-PDA Filter

We consider the following jump Markov linear system (JMLS),

$$\mathbf{x}_k = \mathbf{A}_{r_k}\mathbf{x}_{k-1} + \mathbf{B}_{r_k}\mathbf{w}_{r_k,k}, \quad (2.36a)$$

$$\mathbf{y}_k = \mathbf{C}_{r_k}\mathbf{x}_k + \mathbf{D}_{r_k}\mathbf{v}_{r_k,k}, \quad (2.36b)$$

where $\mathbf{x}_k$ denotes the base state, $\mathbf{y}_k$ denotes the measurement and $r_k \in \{1, \ldots, N_r\}$ denotes the mode index. The mode sequence r_k is assumed to evolve according to a homogeneous Markov chain with transition probability matrix $\Pi = [\pi_{ji} \triangleq P(r_k = i | r_{k-1} = j)]$. The terms $\mathbf{w}_{r_k,k} \sim \mathcal{N}(\mathbf{w}_{r_k,k}; \mathbf{0}, \mathbf{Q}_{r_k})$ and $\mathbf{v}_{r_k,k} \sim \mathcal{N}(\mathbf{v}_{r_k,k}; \mathbf{0}, \mathbf{R}_{r_k})$

represent process and measurement noise respectively. Measurements in the gate are shown as $\mathbf{Y}_k = \{\mathbf{y}_k^i\}_{i=1}^{m_k}$. There are the following hypotheses about these measurements,

- $(\theta_k = 0) \triangleq$ All of $\mathbf{Y}_k$ is false alarm i.e. no target originated measurement in the gate.
- $(\theta_k = i) \triangleq$ Measurement $\mathbf{y}_k^i$ belongs to target, all the rest are false alarms.

for $i = 1, ..., m_k$.

The main idea of IMM-PDA is illustrated in Figure (2.4). The number of hypotheses

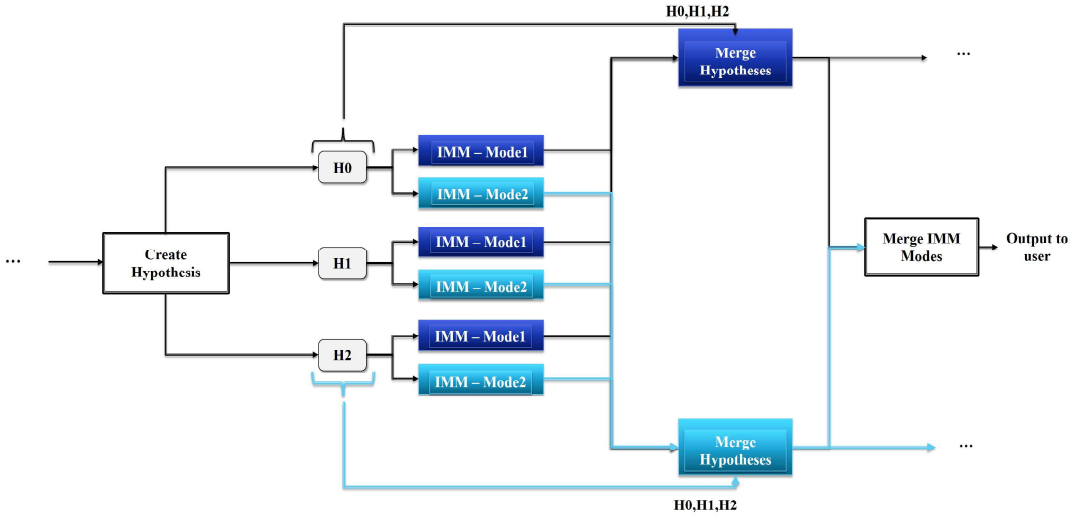


Figure 2.4: Single step of the IMM-PDA filter.

changes at every step. IMM mode statistics (i.e. state estimates, covariances and mode probabilities) are adaptively updated with each hypothesis related quantities at every step. PDA hypothesis probabilities are used to merge the sufficient statistics for each IMM mode.

The main problem is to estimate the posterior distribution $p(\mathbf{x}_k|\mathbf{Y}_{0:k})$. The optimal $p(\mathbf{x}_k|\mathbf{Y}_{0:k})$ is a mixture of Gaussians and an approximation is required. IMM filter makes the approximation,

$$p(\mathbf{x}_k|\mathbf{Y}_{0:k}) = \sum_{i=1}^{N_r} \mu_k^i \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^i, \mathbf{P}_{k|k}^i), \tag{2.37}$$

where

$$\mu_k^i = P(r_k = i|\mathbf{Y}_{0:k}), \tag{2.38}$$

is the posterior mode probability of the ith IMM mode. The mode-conditioned means $\{\hat{\mathbf{x}}_{k|k}^i\}_{i=1}^{N_r}$, covariances $\{\mathbf{P}_{k|k}^i\}_{i=1}^{N_r}$ and the mode probabilities $\{\mu_k^i\}_{i=1}^{N_r}$ must be calculated recursively from their previous quantities $\{\hat{\mathbf{x}}_{k-1|k-1}^i\}_{i=1}^{N_r}$, $\{\mathbf{P}_{k-1|k-1}^i\}_{i=1}^{N_r}$ and $\{\mu_{k-1}^i\}_{i=1}^{N_r}$.

We write the posterior density as follows,

$$p(\mathbf{x}_k|\mathbf{Y}_{0:k}) = \sum_{i=1}^{N_r}\sum_{j=0}^{m_k} p(\mathbf{x}_k|\theta_k = j, r_k = i, \mathbf{Y}_{0:k})P(\theta_k = j, r_k = i|\mathbf{Y}_{0:k}), \tag{2.39a}$$

$$= \sum_{i=1}^{N_r}\sum_{j=0}^{m_k} \rho_k^{ij}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{ij}, \mathbf{P}_{k|k}^{ij}), \tag{2.39b}$$

where

$$\rho_k^{ij} = P(\theta_k = j, r_k = i|\mathbf{Y}_{0:k}), \tag{2.40}$$

is the posterior mode-hypothesis probability of the ith IMM mode and jth hypothesis. A single step of the IMM-PDA filters is given in Algorithm 3 and the detailed block diagram is illustrated in Figure (2.5).

Algorithm 3 ***A Single Step of IMM-PDA filter:*** *Suppose we have the previous sufficient statistics* $\{\mu_{k-1}^i, \hat{\mathbf{x}}_{k-1}^i, \mathbf{P}_{k-1}^i\}_{j=1}^{N_r}$. *Then, a single step of the IMM-PDA algorithm to obtain current sufficient statistics* $\{\mu_k^i, \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i\}_{i=1}^{N_r}$ *is given as follows.*

- ***IMM Prediction Update***
  - ***Mode Probability Prediction:*** *Calculate the predicted mode probability* $\mu_{k|k-1}^i$ *as*

$$\mu_{k|k-1}^i = \sum_{\ell=1}^{N_r} \pi_{\ell i}\mu_{k-1}^{\ell}. \tag{2.41}$$

  - ***Mixing:***
    - *Calculate the mixing probabilities* $\{\mu_{k-1}^{\ell i}\}_{i,\ell=1}^{N_r}$ *as*

$$\mu_{k-1}^{\ell i} = \frac{\pi_{\ell i}\mu_{k-1}^{\ell}}{\sum_{m=1}^{N_r}\pi_{mi}\mu_{k-1}^{m}}. \tag{2.42}$$

* *Calculate the mixed estimates* $\{\hat{\mathbf{x}}_{k-1}^{0i}\}_{i=1}^{N_r}$ *and covariances* $\{\mathbf{P}_{k-1}^{0i}\}_{i=1}^{N_r}$ *as*

$$\hat{\mathbf{x}}_{k-1}^{0i} = \sum_{\ell=1}^{N_r} \mu_{k-1}^{\ell i} \hat{\mathbf{x}}_{k-1}^{l}, \tag{2.43a}$$

$$\mathbf{P}_{k-1}^{0i} = \sum_{\ell=1}^{N_r} \mu_{k-1}^{\ell i} \left[\mathbf{P}_{k-1}^{l} + (\hat{\mathbf{x}}_{k-1}^{l} - \hat{\mathbf{x}}_{k-1}^{0i})(\hat{\mathbf{x}}_{k-1}^{l} - \hat{\mathbf{x}}_{k-1}^{0i})^{\mathrm{T}}\right]. \tag{2.43b}$$

- ***Mode-Matched Prediction Update:*** *For ith model,* $i = 1, \ldots, N_r$*, calculate the predicted estimate* $\hat{\mathbf{x}}_{k|k-1}^{i}$ *and covariance* $\mathbf{P}_{k|k-1}^{i}$ *from the mixed estimate* $\hat{\mathbf{x}}_{k-1}^{0i}$ *and covariance* $\mathbf{P}_{k-1}^{0i}$ *as*

$$\hat{\mathbf{x}}_{k|k-1}^{i} = \mathbf{A}_i \hat{\mathbf{x}}_{k-1}^{0i}, \tag{2.44a}$$

$$\mathbf{P}_{k|k-1}^{i} = \mathbf{A}_i \mathbf{P}_{k-1}^{0i} A_i^{\mathrm{T}} + \mathbf{B}_i \mathbf{Q}_i \mathbf{B}_i^{\mathrm{T}}. \tag{2.44b}$$

- ***Gating***
  - ***Merge Predicted Statistics:*** *Calculate the merged predicted estimate* $\bar{\mathbf{x}}_{k|k-1}$ *and covariance* $\overline{\mathbf{P}}_{k|k-1}$ *from the mode-matched predicted estimate* $\hat{\mathbf{x}}_{k|k-1}^{i}$ *and covariance* $\mathbf{P}_{k|k-1}^{i}$ *as*

$$\bar{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{N_r} \mu_{k|k-1}^{i} \hat{\mathbf{x}}_{k|k-1}^{i}, \tag{2.45a}$$

$$\overline{\mathbf{P}}_{k|k-1} = \sum_{i=1}^{N_r} \mu_{k|k-1}^{i} \left[\mathbf{P}_{k|k-1}^{i} + (\hat{\mathbf{x}}_{k|k-1}^{i} - \bar{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k-1}^{i} - \bar{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\right]. \tag{2.45b}$$

  - ***Gated Measurement Determination:***
    * *Find predicted measurement* $\hat{\mathbf{y}}_{k|k-1}$ *and innovation* $\mathbf{S}_{k|k-1}$ *from the merged predicted estimate* $\bar{\mathbf{x}}_{k|k-1}$ *and covariance* $\overline{\mathbf{P}}_{k|k-1}$ *as*

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{C}\bar{\mathbf{x}}_{k|k-1}, \tag{2.46a}$$

$$\mathbf{S}_{k|k-1} = \mathbf{C}\overline{\mathbf{P}}_{k|k-1}\mathbf{C}^{T} + \mathbf{D}\mathbf{R}\mathbf{D}^{T}. \tag{2.46b}$$

    * *Apply gating and find gated measurements* $\mathbf{Y}_k = \{\mathbf{y}_k^j\}_{j=1}^{m_k}$ *from predicted measurement* $\hat{\mathbf{y}}_{k|k-1}$ *and innovation* $\mathbf{S}_{k|k-1}$ *as*

$$(\mathbf{y}_k^j - \hat{\mathbf{y}}_{k|k-1})^T \mathbf{S}_{k|k-1}^{-1} (\mathbf{y}_k^j - \hat{\mathbf{y}}_{k|k-1}) \leq \gamma_G, \tag{2.47}$$

where γ_G is the gate threshold.

- ***Mode-Hypothesis-Matched Statistics Calculation:*** *For jth hypothesis, $j = 0, \ldots, m_k$ and for ith model, $i = 1, \ldots, N_r$,*

  - ***Mode-Hypothesis Probability Calculation:*** *Calculate mode-hypothesis probabilities ρ_k^{ij} from predicted measurement $\mathbf{y}_{k|k-1}^i$ and innovation $\mathbf{S}_{k|k-1}^i$ quantities for each IMM mode as*

$$\rho_k^{ij} = \frac{\ell_k^{ij} \mu_{k|k-1}^i}{\sum_{m=0}^{m_k} \sum_{l=1}^{N_r} \ell_k^{ml} \mu_{k|k-1}^l}, \tag{2.48}$$

where

$$\ell_k^{ij} = \begin{cases} \beta_{FA}(1 - P_D P_G), & j = 0, \quad \text{(2.49a)} \\ P_D \mathcal{N}(\mathbf{y}_k^j; \hat{\mathbf{y}}_{k|k-1}^i, \mathbf{S}_{k|k-1}^i), & \textit{otherwise} \quad \text{(2.49b)} \end{cases}$$

$p_{k|k-1}^i(\mathbf{y}_k^j) = \mathcal{N}(\mathbf{y}_k^j; \hat{\mathbf{y}}_{k|k-1}^i, \mathbf{S}_{k|k-1}^i)$ is the innovation (measurement prediction) likelihood and

$$\hat{\mathbf{y}}_{k|k-1}^i = \mathbf{C}_i \mathbf{x}_{k|k-1}^i, \tag{2.50a}$$

$$\mathbf{S}_{k|k-1}^i = \mathbf{C}_i \mathbf{P}_{k|k-1}^i \mathbf{C}_i^{\mathrm{T}} + \mathbf{D}_i \mathbf{R}_i \mathbf{D}_i^{\mathrm{T}}. \tag{2.50b}$$

  - ***Mode-Hypothesis-Matched Measurement Update:*** *Calculate the updated estimate $\hat{\mathbf{x}}_k^{ij}$ and covariance $\mathbf{P}_k^{ij}$ from the predicted estimate $\hat{\mathbf{x}}_{k|k-1}^i$ and covariance $\mathbf{P}_{k|k-1}^i$ as*

$$\hat{\mathbf{x}}_k^{ij} = \begin{cases} \hat{\mathbf{x}}_{k|k-1}^i & j = 0, \quad \text{(2.51a)} \\ \hat{\mathbf{x}}_{k|k-1}^i + \mathbf{K}_k^i(\mathbf{y}_k^j - \hat{\mathbf{y}}_{k|k-1}^i) & \textit{otherwise} \quad \text{(2.51b)} \end{cases}$$

$$\mathbf{P}_k^{ij} = \begin{cases} \mathbf{P}_{k|k-1}^i & j = 0, \quad \text{(2.51c)} \\ \mathbf{P}_{k|k-1}^i - \mathbf{K}_k^i \mathbf{S}_k^i (\mathbf{K}_k^i)^{\mathrm{T}} & \textit{otherwise} \quad \text{(2.51d)} \end{cases}$$

where

$$\mathbf{K}_k^i = \mathbf{P}_{k|k-1}^i \mathbf{C}_i^{\mathrm{T}} (\mathbf{S}_k^i)^{-1}. \tag{2.52}$$

- ***Mode-Matched Estimate Calculation (Merge Hypotheses):*** *Calculate the mode-*

matched estimate $\hat{\mathbf{x}}_k^i$*, covariance* $\mathbf{P}_k^i$ *and mode probability* μ_k^i *as*

$$\mu_{k|k}^i = \sum_{j=0}^{m_k} \rho_k^{ij}, \tag{2.53a}$$

$$\hat{\rho}_{k|k}^{ij} = \frac{\rho_k^{ij}}{\sum_{j=0}^{m_k} \rho_k^{ij}}, \tag{2.53b}$$

$$\hat{\mathbf{x}}_k^i = \sum_{j=0}^{m_k} \hat{\rho}_k^{ij} \hat{\mathbf{x}}_k^{ij}, \tag{2.53c}$$

$$\mathbf{P}_k^i = \sum_{j=0}^{m_k} \hat{\rho}_k^{ij} \left[\mathbf{P}_k^{ij} + (\hat{\mathbf{x}}_k^{ij} - \hat{\mathbf{x}}_k^i)(\hat{\mathbf{x}}_k^{ij} - \hat{\mathbf{x}}_k^i)^{\mathrm{T}} \right]. \tag{2.53d}$$

- ***Output Estimate Calculation (Merge IMM Modes):*** *Calculate the overall estimate* $\hat{\mathbf{x}}_k$ *and covariance* $\mathbf{P}_k$ *as*

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_r} \mu_k^i \hat{\mathbf{x}}_k^i, \tag{2.54a}$$

$$\mathbf{P}_k = \sum_{i=1}^{N_r} \mu_k^i \left[\mathbf{P}_k^i + (\hat{\mathbf{x}}_k^i - \hat{\mathbf{x}}_k)(\hat{\mathbf{x}}_k^i - \hat{\mathbf{x}}_k)^{\mathrm{T}} \right]. \tag{2.54b}$$

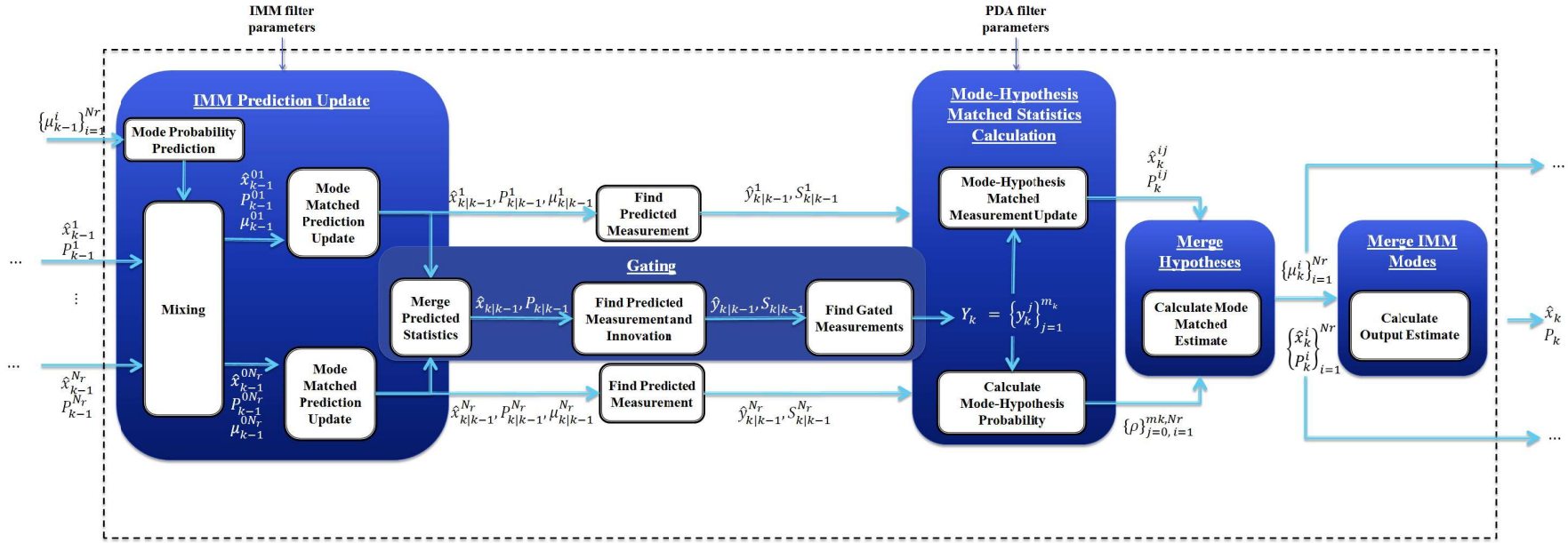


Figure 2.5: Detailed IMM-PDA filter block diagram.

2.3 Gaussian Algebra

This section gives brief results on multiplication and division of Gaussian densities which are used in some of the derivations of the thesis.

2.3.1 Gaussian Multiplication

Consider two vectors $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \mathbb{R}^{n_x}$ and two invertible matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{n_x \times n_x}$. $\mathbf{P}_1$ and $\mathbf{P}_2$ are also symmetric positive definite. With this assumption we can define two Gaussian distributions, $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1)$ and $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2)$.[1] The multiplication of these Gaussian distributions is given as

$$\begin{aligned}\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1)\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2) =& \frac{1}{\sqrt{|2\pi\mathbf{P}_1|}} \exp\big(-0.5\|\mathbf{x}-\hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}}\big) \\ &\times \frac{1}{\sqrt{|2\pi\mathbf{P}_2|}} \exp\big(-0.5\|\mathbf{x}-\hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}}\big), \qquad (2.55a)\\ =& \frac{1}{2\pi\sqrt{|\mathbf{P}_1\mathbf{P}_2|}} \exp\big(-0.5\big(\|\mathbf{x}-\hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}} + \|\mathbf{x}-\hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}}\big)\big), \qquad (2.55b)\end{aligned}$$

where $\|\mathbf{x}\|^2_{\mathbf{P}} \triangleq \mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x}$. It is a fact that the matrix $\mathbf{P}_1 + \mathbf{P}_2$ is invertible [2] if and only if the matrix $\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}$ is invertible

$$(\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1})^{-1} = \mathbf{P}_1(\mathbf{P}_1 + \mathbf{P}_2)^{-1}\mathbf{P}_2. \qquad (2.56)$$

Hence, if the matrix $\mathbf{P}_1 + \mathbf{P}_2$ is invertible, we can write the following expressions

$$\|\mathbf{x}-\hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}} + \|\mathbf{x}-\hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}} = \|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1+\mathbf{P}_2)^{-1}} + \|\mathbf{x}-\overline{\mathbf{x}}\|^2_{\overline{\mathbf{P}}^{-1}}, \qquad (2.57a)$$

$$|\mathbf{P}_1| \cdot |\mathbf{P}_2| = |\mathbf{P}_1 + \mathbf{P}_2| \cdot |\overline{\mathbf{P}}|, \qquad (2.57b)$$

where

$$\overline{\mathbf{P}} = (\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1})^{-1}, \qquad (2.58a)$$

$$\overline{\mathbf{x}} = \overline{\mathbf{P}}(\mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2). \qquad (2.58b)$$

[1] Note that without the property of positive definiteness imposed on the matrix $\mathbf{P}$, the Gaussian distribution $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P})$ might not be well-defined. This is due to the factor $\frac{1}{\sqrt{|\mathbf{P}|}}$, which would be undefined if the determinant $|\mathbf{P}|$ is zero or complex if the determinant $|\mathbf{P}|$ is negative.

[2] According to the Matrix Inversion Lemma in [65], if $\mathbf{A}$ and $\mathbf{C}$ are invertible matrices, $(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$.

We can use the expression (2.57a) on the argument of the exponential function above to write

$$\begin{aligned}\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_2,\mathbf{P}_2) =& \frac{1}{2\pi\sqrt{|\mathbf{P}_1\mathbf{P}_2|}}\exp\big(-0.5\big(\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1+\mathbf{P}_2)^{-1}}\big)\\ &\times\exp\big(-0.5\big(\|\mathbf{x}-\overline{\mathbf{x}}\|^2_{\overline{\mathbf{P}}^{-1}}\big).\end{aligned} \tag{2.59}$$

Using the expression (2.57b) on the argument of the exponential function above to write

$$\begin{aligned}\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_2,\mathbf{P}_2) =& \frac{1}{\sqrt{2\pi|\mathbf{P}_1+\mathbf{P}_2|}}\exp\big(-0.5\big(\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1+\mathbf{P}_2)^{-1}}\big)\\ &\times\frac{1}{\sqrt{2\pi|\overline{\mathbf{P}}|}}\exp\big(-0.5\big(\|\mathbf{x}-\overline{\mathbf{x}}\|^2_{\overline{\mathbf{P}}^{-1}}\big).\end{aligned} \tag{2.60}$$

Then the multiplication of these Gaussian distributions results in

$$\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_2,\mathbf{P}_2) = \mathcal{N}(\hat{\mathbf{x}}_2;\hat{\mathbf{x}}_1,\mathbf{P}_1+\mathbf{P}_2)\mathcal{N}(\mathbf{x};\overline{\mathbf{x}},\overline{\mathbf{P}}). \tag{2.61}$$

Hence, the multiplication of two Gaussian distributions results in a Gaussian distribution and the result can be expressed as a multiplication of another two Gaussian distributions.

The equations given in (2.58) can be rewritten by using matrix inversion lemma as,

$$\overline{\mathbf{x}} = \hat{\mathbf{x}}_1 + \mathbf{P}_1(\mathbf{P}_1+\mathbf{P}_2)^{-1}(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1), \tag{2.62a}$$

$$\overline{\mathbf{P}} = \mathbf{P}_1 - \mathbf{P}_1(\mathbf{P}_1+\mathbf{P}_2)^{-1}\mathbf{P}_1. \tag{2.62b}$$

We can derive the Kalman filter measurement update equations by using the expression (2.62). Using this analogy we can calculate the updated estimate $\overline{\mathbf{x}}$ and covariance $\overline{\mathbf{P}}$ from the predicted estimate $\hat{\mathbf{x}}_1$ and covariance $\mathbf{P}_1$ by considering $\hat{\mathbf{x}}_2$ as the measurement and $\mathbf{P}_2$ as its error covariance as

$$\overline{\mathbf{x}} = \hat{\mathbf{x}}_1 + \mathbf{K}(\hat{\mathbf{x}}_2 - \mathbf{C}\hat{\mathbf{x}}_1), \tag{2.63a}$$

$$\overline{\mathbf{P}} = \mathbf{P}_1 - \mathbf{K}\mathbf{S}\mathbf{K}^T, \tag{2.63b}$$

$$\mathbf{S} = \mathbf{C}\mathbf{P}_1\mathbf{C}^{\mathrm{T}} + \mathbf{P}_2, \tag{2.63c}$$

$$\mathbf{K} = \mathbf{P}_1\mathbf{C}^{\mathrm{T}}\mathbf{S}^{-1}. \tag{2.63d}$$

The Kalman filter measurement update operation can be described as the multiplication of the predicted statistics with the measurement statistics as

$$\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)\mathcal{N}(\hat{\mathbf{x}}_2;\mathbf{C}\mathbf{x},\mathbf{P}_2) = \mathcal{N}(\mathbf{x};\overline{\mathbf{x}},\overline{\mathbf{P}})\mathcal{N}(\hat{\mathbf{x}}_2;\mathbf{C}\hat{\mathbf{x}}_1,\mathbf{C}\mathbf{P}_1\mathbf{C}^{\mathrm{T}}+\mathbf{P}_2). \tag{2.64}$$

2.3.2 Gaussian Division

Consider two vectors $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \mathbb{R}^{n_x}$ and two invertible matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{n_x \times n_x}$. Remember the fact that the matrix $\mathbf{P}_1 + \mathbf{P}_2$ is invertible if and only if the matrix $\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}$ is invertible thanks to the fact that

$$(\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1})^{-1} = \mathbf{P}_1(\mathbf{P}_1 + \mathbf{P}_2)^{-1}\mathbf{P}_2. \tag{2.65}$$

Hence, if the matrix $\mathbf{P}_1 + \mathbf{P}_2$ is invertible, we can write the following expressions

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}} + \|\mathbf{x} - \hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}} &= \|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1+\mathbf{P}_2)^{-1}} \\ &\quad + \|\mathbf{x} - \bar{\mathbf{x}}\|^2_{\overline{\mathbf{P}}^{-1}}, && \text{(2.66a)} \\ |\mathbf{P}_1| \cdot |\mathbf{P}_2| &= |\mathbf{P}_1 + \mathbf{P}_2| \cdot |\overline{\mathbf{P}}|, && \text{(2.66b)} \end{aligned}$$

where $\|\mathbf{x}\|^2_{\mathbf{P}} \triangleq \mathbf{x}^{\mathrm{T}}\mathbf{P}\mathbf{x}$ and

$$\bar{\mathbf{x}} = \overline{\mathbf{P}}(\mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2), \qquad \overline{\mathbf{P}} = \left(\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}\right)^{-1}. \tag{2.67a}$$

Suppose now that the matrix $\mathbf{P}_1 - \mathbf{P}_2$ is invertible. By substituting $\mathbf{P}_2 \leftarrow -\mathbf{P}_2$ in (2.66), we get

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}} - \|\mathbf{x} - \hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}} &= \|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1-\mathbf{P}_2)^{-1}} \\ &\quad + \|\mathbf{x} - \tilde{\mathbf{x}}\|^2_{\widetilde{\mathbf{P}}^{-1}}, && \text{(2.68a)} \\ |\mathbf{P}_1| \cdot |-\mathbf{P}_2| &= |\mathbf{P}_1 - \mathbf{P}_2| \cdot |\widetilde{\mathbf{P}}|, && \text{(2.68b)} \end{aligned}$$

where

$$\tilde{\mathbf{x}} = \widetilde{\mathbf{P}}(\mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 - \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2), \qquad \widetilde{\mathbf{P}} = \left(\mathbf{P}_1^{-1} - \mathbf{P}_2^{-1}\right)^{-1}. \tag{2.69}$$

From this point on in this derivation, we are going to assume that the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ in (2.68) are not only invertible but also symmetric positive definite. With this assumption we can define two Gaussian distributions, $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1)$ and $\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2)$. The ratio of these Gaussian distributions is given as

$$\begin{aligned} \frac{\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_1, \mathbf{P}_1)}{\mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_2, \mathbf{P}_2)} &= \frac{\sqrt{|\mathbf{P}_2|}}{\sqrt{|\mathbf{P}_1|}} \frac{\exp\left(-0.5\|\mathbf{x} - \hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}}\right)}{\exp\left(-0.5\|\mathbf{x} - \hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}}\right)}, && \text{(2.70a)} \\ &= \frac{\sqrt{|\mathbf{P}_2|}}{\sqrt{|\mathbf{P}_1|}} \exp\left(-0.5\big(\|\mathbf{x} - \hat{\mathbf{x}}_1\|^2_{\mathbf{P}_1^{-1}} - \|\mathbf{x} - \hat{\mathbf{x}}_2\|^2_{\mathbf{P}_2^{-1}}\big)\right). && \text{(2.70b)} \end{aligned}$$

Assuming that $\mathbf{P}_2 > \mathbf{P}_1$ (i.e., $\mathbf{P}_1^{-1} > \mathbf{P}_2^{-1}$), we can use the expression (2.68a) on the argument of the exponential function above to write

$$\begin{aligned}\frac{\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)}{\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_2,\mathbf{P}_2)} \propto & \frac{\sqrt{|\mathbf{P}_2|}}{\sqrt{|\mathbf{P}_1|}} \exp\big(-0.5\|\mathbf{x}-\tilde{\mathbf{x}}\|^2_{\widetilde{\mathbf{P}}^{-1}}\big) \\ & \times \exp\big(-0.5\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_1-\mathbf{P}_2)^{-1}}\big), \qquad (2.71a) \\ = & \frac{\sqrt{|\mathbf{P}_2|}}{\sqrt{|\mathbf{P}_1|}} \frac{\exp\big(-0.5\|\mathbf{x}-\tilde{\mathbf{x}}\|^2_{\widetilde{\mathbf{P}}^{-1}}\big)}{\exp\big(-0.5\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\|^2_{(\mathbf{P}_2-\mathbf{P}_1)^{-1}}\big)}, \qquad (2.71b) \\ = & \frac{\sqrt{|\mathbf{P}_2|}}{\sqrt{|\mathbf{P}_1|}} \frac{\sqrt{|\widetilde{\mathbf{P}}|}}{\sqrt{|\mathbf{P}_2-\mathbf{P}_1|}} \frac{\mathcal{N}(\mathbf{x};\tilde{\mathbf{x}},\widetilde{\mathbf{P}})}{\mathcal{N}(\hat{\mathbf{x}}_2;\hat{\mathbf{x}}_1,\mathbf{P}_2-\mathbf{P}_1)}. \qquad (2.71c)\end{aligned}$$

Note that both of the matrices $\mathbf{P}_2 - \mathbf{P}_1$ and $\widetilde{\mathbf{P}}$ are positive definite (and hence invertible) under the assumption $\mathbf{P}_2 > \mathbf{P}_1$ (i.e., $\mathbf{P}_1^{-1} > \mathbf{P}_2^{-1}$). Therefore, the square-roots in the term $\frac{\sqrt{|\widetilde{\mathbf{P}}|}}{\sqrt{|\mathbf{P}_2-\mathbf{P}_1|}}$ in (2.71c) do not give zero or complex values. Noting that the expression (2.68b) is equivalent to

$$|\mathbf{P}_1| \cdot |\mathbf{P}_2| = |\mathbf{P}_2 - \mathbf{P}_1| \cdot |\widetilde{\mathbf{P}}|, \qquad (2.72)$$

we can write

$$\frac{|\widetilde{\mathbf{P}}|}{|\mathbf{P}_1|} = \frac{|\mathbf{P}_2|}{|\mathbf{P}_2 - \mathbf{P}_1|}. \qquad (2.73)$$

Using this equality in (2.71c) results in

$$\frac{\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_1,\mathbf{P}_1)}{\mathcal{N}(\mathbf{x};\hat{\mathbf{x}}_2,\mathbf{P}_2)} = \frac{|\widetilde{\mathbf{P}}|}{|\mathbf{P}_1|} \frac{\mathcal{N}(\mathbf{x};\tilde{\mathbf{x}},\widetilde{\mathbf{P}})}{\mathcal{N}(\hat{\mathbf{x}}_2;\hat{\mathbf{x}}_1,\mathbf{P}_2-\mathbf{P}_1)}, \qquad (2.74)$$

under the assumption that $\mathbf{P}_2 > \mathbf{P}_1$ (i.e., $\mathbf{P}_1^{-1} > \mathbf{P}_2^{-1}$).

CHAPTER 3

DECORRELATION FOR IMM FILTER

In this chapter we investigate the problem of information decorrelation for compensating for the correlations caused by previous communication between two local agents in a distributed estimation framework where IMM filters are used in the local and fusion nodes. In Section 3.1 it is given how the information decorrelation task is related to the IMM filter and the contributions with same purpose in the literature are examined. We give the problem formulation in Section 3.2 where we also review the task of information decorrelation. The information decorrelation approach for the IMM filter, which was taken from [17] is introduced in Section 3.3. Section 3.4 introduces solutions for the implementation issues that might be encountered while applying the chosen decorrelation approach. Section 3.5 describes the fusion operations to be applied in the fusion node. Brief remarks on the computational load of the decorrelation method are made in Appendix B Section B.1. In Section 3.6 we present the numerical results comparing the investigated decorrelation approach with the alternative algorithms on simple distributed maneuvering target tracking examples. The conclusions are drawn in Section 3.7.

3.1 Introduction

The interacting multiple model (IMM) filter [51] is the most well-known and common instance of multiple model filters. The IMM filter is a suboptimal state estimator which approximates the posterior distribution of the base-state (i.e., the continuous-valued state) as a Gaussian mixture. This makes the information decorrelation task for an IMM filter directly related to the division of Gaussian mixtures, which, in general,

does not yield easily to analytical approaches [66]. In [66], it is cleverly noticed that the ratio of two Gaussian mixtures is again a mixture but with non-Gaussian components. The components of the ratio (mixture) are then approximated with Gaussians using importance sampling [66]. In this chapter we avoid the problem of ratio of mixtures by applying the decorrelation on the joint posterior (of the base-state and mode-state) provided by the IMM filter.

In this chapter, we first give the information decorrelation approach for the IMM filter which was first proposed in [17]. We introduce the implementation issues that might be encountered while applying this decorrelation approach and give the proposed solutions to these issues which was first proposed in [18]. We also present a comparison of the proposed solutions for the implementation issues with alternative ideas that might be used to solve similar problems. Furthermore, we provide a brief overview of the computation load of the investigated decorrelation approach and give simulation results.

Since the work [17], there has been recent interest in using all the information in the internal structure of the IMM filters in track fusion. As a significant analytical achievement, it is shown that approximate cross-correlation information between the mode-conditioned estimates of identical local IMM filters can be obtained in [67]. In [67], there is a fusion architecture without memory (i.e., on demand fusion which does not involve correlations due to previous communication) and the posterior mixtures of local IMM filters are fused by using the cross-correlation information. The results of [67] show that the use of the posterior mixtures of the IMM filters in fusion beats the use only the output posteriors (moment-matched Gaussian distributions) of the IMM filters in terms of overall fusion performance. In this chapter, unlike [67], we assume a fusion architecture where correlations are dominated by previous communication between the nodes (rather than the process noise), which necessitates the decorrelation/subtraction of previously communicated information before fusion.

3.2 Problem Definition

We consider two local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ in a sensor network equipped with their own sensors denoted as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ run local Bayesian filters for the following jump Markov linear system (JMLS).

$$\mathbf{x}_k = \mathbf{A}_{r_k}\mathbf{x}_{k-1} + \mathbf{B}_{r_k}\mathbf{w}_{r_k,k}, \tag{3.1a}$$

$$\mathbf{y}_k^s = \mathbf{C}_{r_k}\mathbf{x}_k + \mathbf{D}_{r_k}\mathbf{v}_{r_k,k}^s, \tag{3.1b}$$

where the superscript $s \in \{1, 2\}$ denotes the local agent index; $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denotes the base-state; $r_k \in \{1, \ldots, N_r\}$ denotes the mode-state; $\mathbf{y}_k^s \in \mathbb{R}^{n_y}$ denotes the local measurements of $\mathcal{F}_s$. $\mathbf{w}_{r_k,k} \sim \mathcal{N}(\mathbf{w}_{r_k,k}; \mathbf{0}, \mathbf{Q}_{r_k})$ denotes the i.i.d. process noise. $\mathbf{v}_{r_k,k}^s \sim \mathcal{N}(\mathbf{v}_{r_k,k}^s; \mathbf{0}, \mathbf{R}_{r_k}^s)$ denotes the i.i.d. measurement noise for $\mathcal{F}_s$. The measurement noise sequences $\{\mathbf{v}_{r_k,k}^1\}_{k=0}^{\infty}$ and $\{\mathbf{v}_{r_k,k}^2\}_{k=0}^{\infty}$ belonging to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, are assumed to be independent. The mode sequence r_k is assumed to evolve according to a homogeneous Markov chain with transition probability matrix $\mathbf{\Pi} = [\pi_{ji} \triangleq P(r_k = i | r_{k-1} = j)]$. The initial joint distribution of $\{\mathbf{x}_0, r_0\}$ is given as

$$p(\mathbf{x}_0, r_0) = \pi_0^{r_0}\mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0^{r_0}, \mathbf{P}_0^{r_0}). \tag{3.2}$$

The models of the form (3.1) are commonly used in maneuvering target tracking, which is a sub-area of target tracking, where the aim is to adaptively estimate the state of the target when the target state and/or measurements evolve with time-varying dynamics. The state estimation algorithms using jump Markov models are called *multiple model approaches* [21, Section 11.6], [64], whose most well-known instance is the so-called interacting multiple model (IMM) filter [51], which is described in Section 2.2.3.1.

We assume that the local Bayesian filter of the local agent $\mathcal{F}_s$ provides the (approximate) joint posterior distribution $p^s(\mathbf{x}_k, r_k | \mathbf{y}_{0:k}^s)$ of $\{\mathbf{x}_k, r_k\}$ given $\mathbf{y}_{0:k}^s \triangleq \{\mathbf{y}_0^s, \mathbf{y}_1^s, \ldots, \mathbf{y}_k^s\}$. The simple fusion architecture we consider and the corresponding information graph are shown in Figure 1.3a and Figure 1.3b, respectively. The local agent $\mathcal{F}_1$ communicates its recent information to $\mathcal{F}_2$ with a constant period N. In Figure 1.3b the case with $N = 3$ is shown. We here constrain ourselves to the case where $\mathcal{F}_1$ transmits only the information pertaining to the base-state $\mathbf{x}_k$ to $\mathcal{F}_2$,

i.e., the information related to the marginal posterior distribution $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})$ of the base-state $\mathbf{x}_k$ is to be communicated to $\mathcal{F}_2$.

In general, the state estimation errors of the local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ are correlated due to the common process noise in (3.1). After $\mathcal{F}_1$ transmits its local information to $\mathcal{F}_2$ for the first time and $\mathcal{F}_2$ fuses this information with the statistics of its own Bayesian filter, the estimates of the local agents become correlated due to previous communication as well. Since IMM filter is the state-of-the-art adaptive estimation technique in maneuvering target tracking, the process noise would affect the estimation errors to much smaller an extent than it would to those of Kalman filter and its variants. Therefore, it is assumed that the correlations caused by the common process noise will be negligible with an IMM filter. Here, we restrict ourselves to the case where the correlations between the quantities to be fused are dominated by the previous communication, i.e., we will neglect the effects of the common process noise. Even when the effects of the common process noise are negligible, when the correlations due to previous communication are not compensated, the fused quantities might be inconsistent (see Section 2.1.2 for a mathematical definition of consistency), which might lead to poor performance and/or the divergence of the state estimator. We call the process of compensating for the effects of previous communication *information decorrelation* or just *decorrelation*. In general, decorrelation has to be carried out either at the transmitting local agent $\mathcal{F}_1$ or at the receiving local agent $\mathcal{F}_2$, which is the fusion center. In this chapter we assume that the decorrelation operation is conducted at the transmitting local agent $\mathcal{F}_1$. Hence, the quantities that the local agent $\mathcal{F}_1$ transmits can be called *(quasi-)tracklets* [7].

Suppose now that we are at an intermediate stage of estimation at time k and $\mathcal{F}_1$ transmitted its last information at time $k-N$. Let the marginal posterior distribution $p^1(\mathbf{x}_{k-N}|\mathbf{y}^1_{0:k-N})$ represent the transmitted information from the local node $\mathcal{F}_1$ to the fusion node $\mathcal{F}_2$ up to (and including) time $k-N$. Since this information has already been used and fused by the fusion center $\mathcal{F}_2$, when the local node $\mathcal{F}_1$ wants to transmit its updated information at time k represented by the current marginal posterior $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})$, it first has to decorrelate the previously communicated information $p^1(\mathbf{x}_{k-N}|\mathbf{y}^1_{0:k-N})$ from the marginal posterior information $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})$ and then transmit the decorrelated information. We here make the following assumption about the

JMLS in (3.1).

Assumption 1 *The state transition matrix* $\mathbf{A}_{r_k}$ *of* (3.1) *is independent of the mode sequence* r_k *and is invertible, i.e.,* $\mathbf{A}_{r_k} = \mathbf{A}$ *for all* $r_k \in \{1, \ldots, N_r\}$*, where* $\mathbf{A}$ *is invertible.*

When we neglect the effects of the process noise and make Assumption 1, it is shown in Appendix A that the decorrelation operation amounts to the division of the updated marginal distribution $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})$ to the predicted marginal distribution $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k-N})$ i.e., the information $\tilde{p}^1(\mathbf{x}_k)$ that has to be sent to $\mathcal{F}_2$ is given as

$$\tilde{p}^1(\mathbf{x}_k) \propto \frac{p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k-N})}. \tag{3.3}$$

Note here that in general the function $\tilde{p}^1_k(\mathbf{x}_k)$, which we call *decorrelated information*, might not be normalizable with respect to $\mathbf{x}_k$ and hence cannot be interpreted as a valid probability density function. When $\tilde{p}^1_k(\mathbf{x}_k)$ is normalizable, we are going to normalize it and call the result the *decorrelated distribution*, and its statistics the *decorrelated statistics*, e.g., decorrelated mean and covariance. Note that since there is actually process noise in the JMLS in (3.1), in our application, the decorrelation operation in (3.3) will only be an approximation of the case considered in Appendix A.

When the JMLS contains only a single model, i.e., $N_r = 1$, the estimation problem can be solved using a Kalman filter, which is the optimal estimator in minimum mean square error sense. In this case, the posterior $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k})$ and the predicted distribution $p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k-N})$ would be Gaussian distributions given as

$$p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^1_k, \mathbf{P}^1_k), \tag{3.4a}$$

$$p^1(\mathbf{x}_k|\mathbf{y}^1_{0:k-N}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^1_{k|k-N}, \mathbf{P}^1_{k|k-N}), \tag{3.4b}$$

where $\hat{\mathbf{x}}^1_k$ and $\hat{\mathbf{x}}^1_{k|k-N}$ are the current and predicted state estimates, respectively, while $\mathbf{P}^1_k$ and $\mathbf{P}^1_{k|k-N}$ are the corresponding covariances. The decorrelated information $\tilde{p}^1(\mathbf{x}_k)$ in (3.3) can then be parameterized by the information vector $\tilde{\boldsymbol{\jmath}}^1_k$ and the information matrix $\tilde{\mathcal{J}}^1_k$ given as

$$\tilde{\boldsymbol{\jmath}}^1_k \triangleq (\mathbf{P}^1_k)^{-1}\hat{\mathbf{x}}^1_k - (\mathbf{P}^1_{k|k-N})^{-1}\hat{\mathbf{x}}^1_{k|k-N}, \tag{3.5a}$$

$$\tilde{\mathcal{J}}^1_k \triangleq (\mathbf{P}^1_k)^{-1} - (\mathbf{P}^1_{k|k-N})^{-1}. \tag{3.5b}$$

These quantities can be transmitted to $\mathcal{F}_2$ and then fused with the local state estimate and covariance of $\mathcal{F}_2$ using the information form of the Kalman filter measurement update equations. Note that the information matrix $\widetilde{\boldsymbol{\mathcal{J}}}_k$ in (3.5b) is guaranteed to be only positive semi-definite and hence it might not be invertible. When it is invertible, the decorrelated information $\tilde{p}^1(\mathbf{x}_k)$ in (3.3) corresponding to quantities in (3.4) would be normalizable and when normalized, the result can be shown to be the following Gaussian distribution.

$$\tilde{p}^1(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1), \tag{3.6}$$

where the decorrelated mean $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ are given by the expressions below.

$$\tilde{\mathbf{x}}_k^1 \triangleq (\widetilde{\boldsymbol{\mathcal{J}}}_k^1)^{-1}\tilde{\boldsymbol{\jmath}}_k^1 = \widetilde{\mathbf{P}}_k^1\big((\mathbf{P}_k^1)^{-1}\hat{\mathbf{x}}_k^1 - (\mathbf{P}_{k|k-N}^1)^{-1}\hat{\mathbf{x}}_{k|k-N}^1\big), \tag{3.7a}$$

$$\widetilde{\mathbf{P}}_k^1 \triangleq (\widetilde{\boldsymbol{\mathcal{J}}}_k^1)^{-1} = \big((\mathbf{P}_k^1)^{-1} - (\mathbf{P}_{k|k-N}^1)^{-1}\big)^{-1}. \tag{3.7b}$$

When $\widetilde{\boldsymbol{\mathcal{J}}}_k^1$ is invertible, it would be equivalent to transmit the decorrelated mean $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to $\mathcal{F}_2$ instead of $\{\tilde{\boldsymbol{\jmath}}_k^1, \widetilde{\boldsymbol{\mathcal{J}}}_k^1\}$. The quantities $\{\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1\}$ can be fused with the local filter statistics of $\mathcal{F}_2$ by considering the decorrelated mean $\tilde{\mathbf{x}}_k^1$ as a conditionally independent measurement of the true base-state $\mathbf{x}_k$ with the measurement error covariance $\widetilde{\mathbf{P}}_k^1$.

When $N_r > 1$, the marginal posterior distribution $p^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1)$ and predicted marginal distribution $p^1(\mathbf{x}_k|\mathbf{y}_{0:k-N}^1)$ become Gaussian mixtures. Hence making the decorrelation in (3.3) is not straightforward since a division of Gaussian mixtures, which does not have an analytical expression [66], is required. In [18], the approximate solution of this decorrelation problem first proposed in [17] is used when the posterior densities are obtained using an IMM filter.

3.3 Decorrelation Approach

In this section we describe the information decorrelation methodology for the IMM filter of the local agent $\mathcal{F}_1$, which was first proposed in [17]. We assume that we are at an intermediate stage of estimation at time k. We assume that the last decorrelated information transmission from $\mathcal{F}_1$ to $\mathcal{F}_2$ happened at time $k-N$. In [17], two decorrelation methodologies were proposed:

- Output level decorrelation (OLD),
- Mode level decorrelation (MLD).

In OLD, the merged output estimates and covariances of the local IMM filter (i.e., $\{\hat{\mathbf{x}}_k^1, \mathbf{P}_k^1\}$) are decorrelated and transmitted. In MLD, unmerged mode-conditioned estimates and covariances of the local IMM filter, i.e., $\{\mu_k^{1,i}, \hat{\mathbf{x}}_k^{1,i}, \mathbf{P}_k^{1,i}\}_{i=1}^{N_r}$, are decorrelated, merged and transmitted. In the following parts of this chapter, we are going to use the parts of the IMM filter as building blocks for the information decorrelation approach we describe. For this purpose, we define the following functions.

- $\big[\{\bar{\mu}^i, \bar{\mathbf{x}}^i, \overline{\mathbf{P}}^i\}_{i=1}^{N_r}\big]$**=IMMPrediction**$\big(\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=1}^{N_r}\big)$**:** This is the functional representation of the IMM filter prediction update in Algorithm 1 which takes the mode probabilities, estimates and covariances $\{\mu_{k-1}^{s,i}, \hat{\mathbf{x}}_{k-1}^{s,i}, \mathbf{P}_{k-1}^{s,i}\}_{i=1}^{N_r}$ and returns the predicted mode probabilities, estimates and covariances $\{\mu_{k|k-1}^{s,i}, \hat{\mathbf{x}}_{k|k-1}^{s,i}, \mathbf{P}_{k|k-1}^{s,i}\}_{i=1}^{N_r}$.
- $\big[\hat{\mathbf{x}}, \mathbf{P}\big]$**=Merging**$\big(\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=1}^{N_r}\big)$**:** This is the functional representation of the output estimate calculation part in Algorithm 1 which takes the mode probabilities, estimates and covariances $\{\mu_k^{s,i}, \hat{\mathbf{x}}_k^{s,i}, \mathbf{P}_k^{s,i}\}_{i=1}^{N_r}$ and returns the merged estimate $\hat{\mathbf{x}}_k^s$ and covariance $\mathbf{P}_k^s$.

3.3.1 Output Level Decorrelation (OLD)

The OLD decorrelates and transmits the output estimate and the covariance of the local IMM filter (i.e., $\hat{\mathbf{x}}_k^1$ and $\mathbf{P}_k^1$). The output estimate and its covariance at the last transmission time $k-N$ are $\hat{\mathbf{x}}_{k-N}^1$ and $\mathbf{P}_{k-N}^1$, respectively. The transmitted information up to (and including) time $k-N$ is hence represented by the Gaussian distribution $N(\mathbf{x}_{k-N}; \hat{\mathbf{x}}_{k-N}^1, \mathbf{P}_{k-N}^1)$. The decorrelation operation requires the predicted version of this distribution to time k. Since there are N_r state transition models unlike the case of Kalman filter, we need to use an IMM filter for this prediction operation. For this purpose, we initialize an N_r model IMM predictor with mode probabilities at time $k-N$, equal mode-conditioned estimates and covariances, i.e., $\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}$, $\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^1$ and $\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^1$ for $i = 1, \ldots, N_r$, where we use overlined quantities to separate them from the respective quantities of the primary IMM filter at

the local node. We then use the IMM prediction operation N-times to obtain the predicted mode probabilities $\bar{\mu}^{1,i}_{k|k-N}$, state estimates $\bar{\mathbf{x}}^{1,i}_{k|k-N}$ and covariances $\overline{\mathbf{P}}^{1,i}_{k|k-N}$ for $i = 1, \ldots, N_r$. Finally these quantities are merged with the merging operation to obtain the predicted decorrelated estimate $\bar{\mathbf{x}}^{1}_{k|k-N}$ and covariance $\bar{\mathbf{P}}^{1}_{k|k-N}$. Note that the predicted information obtained in this manner will also be represented by a single Gaussian density, i.e., $\bar{p}^1(\mathbf{x}_k|\mathbf{y}^1_{0:k-N}) = \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})$. Hence the new decorrelated information will be represented by a Gaussian $\tilde{p}^1(\mathbf{x}_k|\mathbf{y}^1_{0:k}) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)$, whose parameters can be calculated using the information filter expressions given in (3.7).

In OLD, the decorrelated merged estimate $\tilde{\mathbf{x}}^1_k$ and covariance $\widetilde{\mathbf{P}}^1_k$ are transmitted to the second node. The weight $\tilde{\mu}^1_k$, which is due to the division operation, is ignored, where $\tilde{\mu}^1_k \propto \frac{|\widetilde{\mathbf{P}}^1_k|}{|\mathbf{P}^1_k|} \times \left(\mathcal{N}(\bar{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}^1_k, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)\right)^{-1}$. This corresponds to the transmission of the density $\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)$ to the fusion node. The OLD is summarized in Algorithm 4. Section 3.5 gives the fusion formula to be applied into the receiving node. The decorrelation and transmission operations made in the first node in OLD are illustrated in Figure 3.1.

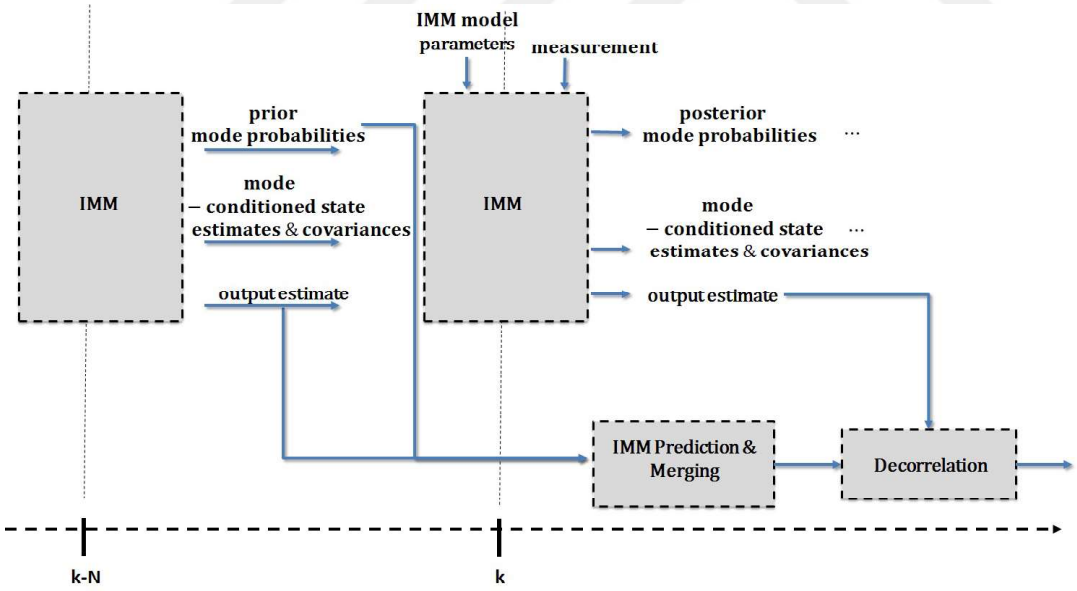


Figure 3.1: OLD decorrelation & transmission operations for the IMM filter.

Algorithm 4 (OLD)

- *Given:* $\{\mu^{1,i}_{k-N}\}_{i=1}^{N_r}, \hat{\mathbf{x}}^1_{k-N}, \mathbf{P}^1_{k-N}$.
- *Initialization:*

$$\bar{\mu}^{1,i}_{k-N} = \mu^{1,i}_{k-N}, \quad i = 1, \dots, N_r, \tag{3.8a}$$

$$\bar{\mathbf{x}}^{1,i}_{k-N} = \hat{\mathbf{x}}^1_{k-N}, \quad i = 1, \dots, N_r, \tag{3.8b}$$

$$\overline{\mathbf{P}}^{1,i}_{k-N} = \mathbf{P}^1_{k-N}, \quad i = 1, \dots, N_r. \tag{3.8c}$$

- *IMM predictor:*

  for $j = k-N+1, \dots, k-N$

  $\big[\{\bar{\mu}^{1,i}_{j|k-N}, \bar{\mathbf{x}}^{1,i}_{j|k-N}, \overline{\mathbf{P}}^{1,i}_{j|k-N}\}_{i=1}^{N_r}\big]$ = *IMMPrediction* $\big(\{\bar{\mu}^{1,i}_{j-1|k-N}, \bar{\mathbf{x}}^{1,i}_{j-1|k-N}, \overline{\mathbf{P}}^{1,i}_{j-1|k-N}\}_{i=1}^{N_r}\big)$.

  end

- *Merging:*

$$\big[\bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N}\big] = \textit{Merging}\big(\{\bar{\mu}^{1,i}_{k|k-N}, \bar{\mathbf{x}}^{1,i}_{k|k-N}, \overline{\mathbf{P}}^{1,i}_{k|k-N}\}_{i=1}^{N_r}\big).$$

- *Decorrelation:*

$$\tilde{\mathbf{x}}^1_k = \widetilde{\mathbf{P}}^1_k\big((\mathbf{P}^1_k)^{-1}\hat{\mathbf{x}}^1_k - (\overline{\mathbf{P}}^1_{k|k-N})^{-1}\bar{\mathbf{x}}^1_{k|k-N}\big), \tag{3.9a}$$

$$\widetilde{\mathbf{P}}^1_k = \big((\mathbf{P}^1_k)^{-1} - (\overline{\mathbf{P}}^1_{k|k-N})^{-1}\big)^{-1}. \tag{3.9b}$$

- *Transmission:* $\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k$.

An important point to note here is that, unlike the case in a single Kalman filter, the decorrelation operation in OLD does not always result in positive semi-definite decorrelated covariance which was examined at the end of Section 3.2. In such a case the transmission of the decorrelated information to the fusion center at the end of OLD cannot be done. We give details about such issues and propose methods to deal with them in Section 3.4.

3.3.2 Mode Level Decorrelation (MLD)

The transmitted information about the base-state up to (and including) time $k-N$ is represented by a Gaussian mixture, i.e., $p^1(\mathbf{x}_{k-N}|\mathbf{y}^1_{0:k-N}) =$

$\sum_{i=1}^{N_r} \mu_{k-N}^{1,i} \mathcal{N}(\mathbf{x}_{k-N}; \hat{\mathbf{x}}_{k-N}^{1,i}, \mathbf{P}_{k-N}^{1,i})$. Using this information, we can initialize an IMM predictor by setting $\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}$, $\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^{1,i}$ and $\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^{1,i}$ for $i = 1, \ldots, N_r$, where we use overlined quantities to separate them from the respective quantities of the primary IMM filter at the local agent $\mathcal{F}_1$. If we predict this information to time k using an IMM predictor, we can obtain the predicted mode probabilities $\bar{\mu}_{k|k-N}^{1,i}$, mode-conditioned estimates $\bar{\mathbf{x}}_{k|k-N}^{1,i}$ and covariances $\overline{\mathbf{P}}_{k|k-N}^{1,i}$ of the predicted marginal distribution
$p^1(\mathbf{x}_k|\mathbf{y}_{0:k-N}^1) = \sum_{i=1}^{N_r} \bar{\mu}_{k|k-N}^{1,i} \mathcal{N}(\mathbf{x}_{k-N}; \bar{\mathbf{x}}_{k|k-N}^1, \overline{\mathbf{P}}_{k|k-N}^1)$. Since the current marginal posterior is also a Gaussian mixture, i.e., $p^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1) = \sum_{i=1}^{N_r} \mu_k^{1,i} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,i}, \mathbf{P}_k^{1,i})$, the decorrelation in (3.3) involves the division of two Gaussian mixtures, $p^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1)$ and $p^1(\mathbf{x}_k|\mathbf{y}_{0:k-N}^1)$, which cannot be carried out analytically [66]. In MLD, this problem is overcome by applying the decorrelation over the joint densities $p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^1)$ and $p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k-N}^1)$ provided by the IMM filter and, in effect, by making the following approximation.

$$\frac{p^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1)}{p^1(\mathbf{x}_k|\mathbf{y}_{0:k-N}^1)} \approx \sum_{r_k} \frac{p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^1)}{p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k-N}^1)}, \tag{3.10}$$

up to a constant multiplier with respect to $\mathbf{x}_k$. The joint densities $p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^1)$ and $p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k-N}^1)$ are the scaled Gaussian distributions provided by the IMM filter and predictor as follows.

$$p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^1) = \mu_k^{1,r_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k}, \mathbf{P}_k^{1,r_k}), \tag{3.11a}$$

$$p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k-N}^1) = \bar{\mu}_{k|k-N}^{1,r_k} \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k}). \tag{3.11b}$$

Division of these two joint distributions gives the following joint decorrelated information.

$$\tilde{p}^1(\mathbf{x}_k, r_k) \propto \frac{p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^1)}{p^1(\mathbf{x}_k, r_k|\mathbf{y}_{0:k-N}^1)}, \tag{3.12a}$$

$$= \frac{\mu_k^{1,r_k}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k}, \mathbf{P}_k^{1,r_k})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k})}. \tag{3.12b}$$

Under some assumptions of normalizability with respect to $\mathbf{x}_k$, the division of two Gaussian distributions above can be written as a scaled Gaussian (see Section 2.3.2 for details), which results in the following joint decorrelated distribution.

$$\tilde{p}^1(\mathbf{x}_k, r_k) = \tilde{\mu}_k^{1,r_k} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_k}, \widetilde{\mathbf{P}}_k^{1,r_k}), \tag{3.13}$$

where

$$\tilde{\mu}_k^{1,r_k} \propto \frac{\mu_k^{1,r_k}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{|\widetilde{\mathbf{P}}_k^{1,r_k}|}{|\mathbf{P}_k^{1,r_k}|} \times \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,r_k}; \hat{\mathbf{x}}_k^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k} - \mathbf{P}_k^{1,r_k})\right)^{-1},$$

$$\tilde{\mathbf{x}}_k^{1,r_k} = \widetilde{\mathbf{P}}_k^{1,r_k}\left((\mathbf{P}_k^{1,r_k})^{-1}\hat{\mathbf{x}}_k^{1,r_k} - (\overline{\mathbf{P}}_{k|k-N}^{1,r_k})^{-1}\bar{\mathbf{x}}_{k|k-N}^{1,r_k}\right), \tag{3.14a}$$

$$\widetilde{\mathbf{P}}_k^{1,r_k} = \left((\mathbf{P}_k^{1,r_k})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1,r_k})^{-1}\right)^{-1}. \tag{3.14b}$$

Marginalizing out the mode variable r_k from the joint decorrelated distribution $\tilde{p}^1(\mathbf{x}_k, r_k)$, we get the approximate marginal decorrelated distribution below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r} \tilde{\mu}_k^{1,i} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}). \tag{3.15}$$

It is important to note that the decorelated information in MLD is a Gaussian mixture. Therefore, the decorrelated Gaussian mixture is merged before transmission to the fusion node since a Gaussian mixture transmission is computationally more demanding than transmission of a single Gaussian distribution. We apply a merging operation to the decorrelated mixture $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$.

The MLD is summarized in Algorithm 5. Note that the decorrelation operations in MLD might occasionally result in indefinite or even negative definite decorrelated covariances unlike the case in a single Kalman filter, which was examined at the end of Section 3.2. In such a case the merging operation at the end of MLD cannot be applied. We give details about such issues and propose methods to deal with them in Section 3.4. The decorrelation and transmission operations made in the first node in MLD are illustrated in Figure 3.2. Section 3.5 gives the fusion formula to be applied into the receiving node.

Algorithm 5 (MLD)

- *Given:* $\{\mu_{k-N}^{1,i}, \hat{\mathbf{x}}_{k-N}^{1,i}, \mathbf{P}_{k-N}^{1,i}\}_{i=1}^{N_r}$.
- *Initialization:*

$$\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}, \quad i = 1, \ldots, N_r, \tag{3.16a}$$

$$\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^{1,i}, \quad i = 1, \ldots, N_r, \tag{3.16b}$$

$$\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^{1,i}, \quad i = 1, \ldots, N_r. \tag{3.16c}$$

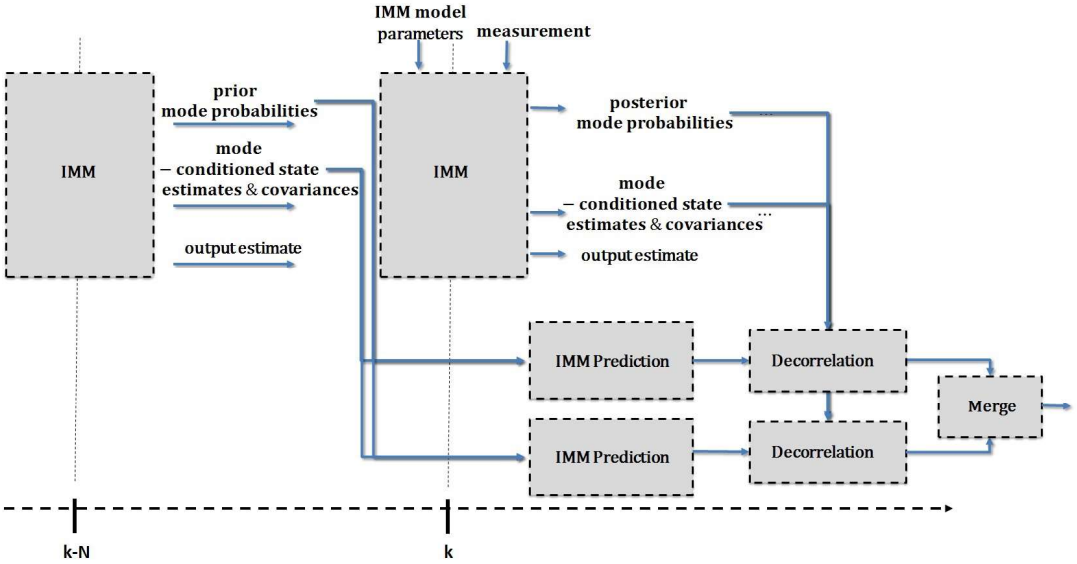


Figure 3.2: MLD decorrelation & transmission operations for the IMM filter.

- *IMM predictor:*

  For $j = k - N + 1, \ldots, k$,

  $$\left[\{\bar{\mu}^{1,i}_{j|k-N}, \bar{\mathbf{x}}^{1,i}_{j|k-N}, \overline{\mathbf{P}}^{1,i}_{j|k-N}\}_{i=1}^{N_r}\right] = \textit{IMMPrediction}\left(\{\bar{\mu}^{1,i}_{j-1|k-N}, \bar{\mathbf{x}}^{1,i}_{j-1|k-N}, \overline{\mathbf{P}}^{1,i}_{j-1|k-N}\}_{i=1}^{N_r}\right).$$

  End

- *Decorrelation:*

  For $i = 1, \ldots, N_r$,

  $$\tilde{\mu}^{1,i}_k \propto \frac{\mu^{1,i}_k}{\bar{\mu}^{1,i}_{k|k-N}} \frac{|\widetilde{\mathbf{P}}^{1,i}_k|}{|\mathbf{P}^{1,i}_k|} \times \left(\mathcal{N}(\bar{\mathbf{x}}^{1,i}_{k|k-N}; \hat{\mathbf{x}}^{1,i}_k, \overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,i}_k)\right)^{-1}, \tag{3.17a}$$

  $$\tilde{\mathbf{x}}^{1,i}_k = \widetilde{\mathbf{P}}^{1,i}_k\left((\mathbf{P}^{1,i}_k)^{-1}\hat{\mathbf{x}}^{1,i}_k - (\overline{\mathbf{P}}^{1,i}_{k|k-N})^{-1}\bar{\mathbf{x}}^{1,i}_{k|k-N}\right), \tag{3.17b}$$

  $$\widetilde{\mathbf{P}}^{1,i}_k = \left((\mathbf{P}^{1,i}_k)^{-1} - (\overline{\mathbf{P}}^{1,i}_{k|k-N})^{-1}\right)^{-1}. \tag{3.17c}$$

  End

- *Merging*

  $$\left[\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k\right] = \textit{Merging}\left(\{\tilde{\mu}^{1,i}_k, \tilde{\mathbf{x}}^{1,i}_k, \widetilde{\mathbf{P}}^{1,i}_k\}_{i=1}^{N_r}\right).$$

- *Transmission:* $\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k$.

3.4 Implementation Issues

Both of the proposed decorrelation approaches in the previous section involve the information subtraction operations given below.

$$\tilde{\mathbf{x}} = \widetilde{\mathbf{P}}(\mathbf{P}^{-1}\hat{\mathbf{x}} - \overline{\mathbf{P}}^{-1}\overline{\mathbf{x}}), \tag{3.18a}$$

$$\widetilde{\mathbf{P}} = \left(\mathbf{P}^{-1} - \overline{\mathbf{P}}^{-1}\right)^{-1}, \tag{3.18b}$$

where we removed the subscripts related to time and superscripts related to the local agent index and mode-state for the sake of simplicity. In (3.18), the quantities $\{\hat{\mathbf{x}}, \mathbf{P}\}$, $\{\overline{\mathbf{x}}, \overline{\mathbf{P}}\}$, $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ represent the mean and covariance (of a component) of the current distribution, predicted distribution and decorrelated distribution, respectively. In our studies we have encountered two important issues related to the information matrix difference in (3.18b) while applying decorrelation process for the IMM filter. These issues are called as *incompatible covariance problem* and *outlier decorrelation problem*. These problems, which are related to each other to some degree, are introduced in two separate subsections below. The solution to these problems is presented in Section 3.4.3.

3.4.1 Incompatible Covariance Problem

Even with a single Kalman filter the information matrix subtraction operation in (3.18b) might result in a singular difference matrix [see the discussion below (3.5)], which would make the matrix inverse operation in (3.18b) meaningless. Having a complicated prediction and update structure with approximations, the situation is worse with an IMM filter, in that the difference matrix in (3.18b) might occasionally be indefinite or even negative-(semi-)definite. We call these problems, which would make applying OLD and MLD impossible, the incompatible covariance problem. In mathematical terms we can say that the impompatible covariance problem appears when the predicted covariance $\overline{\mathbf{P}}$ is not larger (in a positive-definite sense) than the current covariance $\mathbf{P}$, i.e., $\overline{\mathbf{P}} \not\succ \mathbf{P}$.

Note that OLD applies the information matrix substraction operation in (3.18b) to the (predicted and current) output covariances of the IMM filter. When we obtain

improper distribution as a result of the subtraction operation, the transmission operation at the end of OLD in Algorithm 4 cannot be done. On the other hand, MLD applies the information matrix substraction operation in (3.18b) to the (predicted and current) covariances of all modes of the IMM filter. When we encounter incompatible covariance problem for one or more of the modes, the merging operation at the end of MLD in Algorithm 5 cannot be used. Merging operation is a form of Kullback-Leibler (or information) projection which cannot be applied to improper distributions to the author's best knowledge. Furthermore, sending indefinite or negative definite (decorrelated) covariances which would be obtained after an incompatible covariance problem to the fusion node $\mathcal{F}_2$ might also harm the covariances of $\mathcal{F}_2$.

We observed that the incompatible covariance problem is likely to happen especially when there is a mode switch of the IMM filter from a non-maneuvering mode to a maneuvering mode. In such a case, the predicted mode remains at the non-maneuvering mode since the IMM predictor is unaware of the recent measurements. Hence the predicted covariances do not grow sufficiently fast compared to the current (updated) mode covariance which significantly grows due to the switching to the maneuvering mode. The following example illustrates this phenomenon on a toy problem.

Example 1 *Consider the following JMLS with the scalar base-state $x_k \in \mathbb{R}$ and the mode-state $r_k \in \{1, 2\}$.*

$$x_k = x_{k-1} + w_{r_k,k}, \tag{3.19a}$$

$$y_k = x_k + v_k, \tag{3.19b}$$

which has the following parameters in the JMLS model (2.16).

$$A_1 = A_2 = 1, \qquad B_1 = B_2 = 1, \qquad C_1 = C_2 = 1, \tag{3.20a}$$

$$D_1 = D_2 = 1, \qquad R_1 = R_2 = 10, \qquad Q_1 = 1,\ Q_2 = 10^2. \tag{3.20b}$$

The diagonal elements of the transition probability matrix $\mathbf{\Pi} \in \mathbb{R}^{2\times 2}$ are given as $\pi_{11} = \pi_{22} = 0.95$. Note that the different models in the JMLS (3.19) *differ from each other only in their process noise variances, Q_1, Q_2. Since $Q_2 \gg Q_1$, the first ($r_k = 1$) and the second ($r_k = 2$) modes of the JMLS* (3.19) *correspond to non-maneuvering and maneuvering modes, respectively.*

Suppose that we are running an IMM filter on this JMLS and at time $k-2$, we have the following posterior IMM statistics,

$$\mu^1_{k-2} = 0.95, \qquad \hat{x}^1_{k-2} = 0, \qquad P^1_{k-2} = 1, \tag{3.21a}$$

$$\mu^2_{k-2} = 0.05, \qquad \hat{x}^2_{k-2} = 0, \qquad P^2_{k-2} = 1. \tag{3.21b}$$

Note that these IMM statistics means that the filter thinks that the active mode of the JMLS is the non-maneuvering mode. Applying two consecutive IMM prediction steps to the statistics above, we get

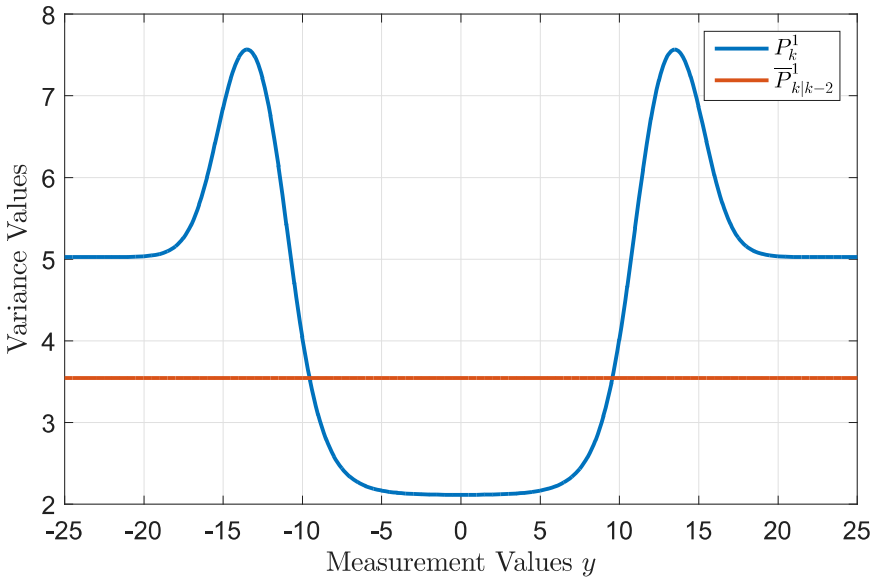
$$\bar{\mu}^1_{k|k-2} = 0.8645, \qquad \bar{x}^1_{k|k-2} = 0, \qquad \overline{P}^1_{k|k-2} = 3.5440, \tag{3.22a}$$

$$\bar{\mu}^2_{k|k-2} = 0.1355, \qquad \bar{x}^2_{k|k-2} = 0, \qquad \overline{P}^2_{k|k-2} = 167.9391. \tag{3.22b}$$

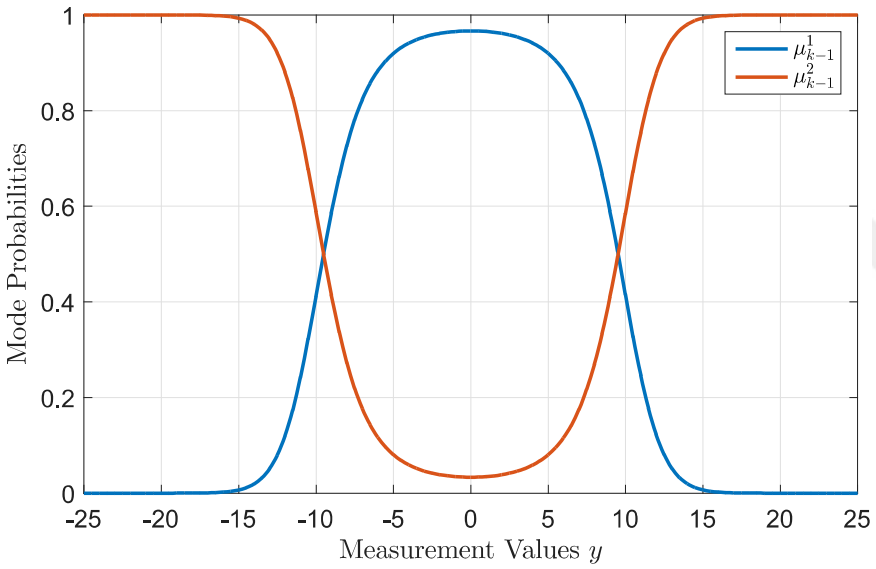
Suppose that at time $k-1$ and k, we got equal measurements $y_{k-1} = y_k = y$. We apply IMM prediction and measurement updates on the statistics in (3.21) *twice with the measurements $y_{k-1} = y_k = y$ by changing y in the interval $[-25, 25]$. The resulting mode-conditioned updated variance quantities P^1_k for $r_k = 1$ are shown in Figure 3.3a along with the mode-conditioned predicted variance $\overline{P}^1_{k|k-2}$ given in* (3.22)*.*

For small absolute values of the measurement y, the updated variance P^1_k is strictly lower than the predicted variance $\overline{P}^1_{k|k-2}$, which means that there would be no incompatible covariance problem in the non-maneuvering mode $r_k = 1$. On the other hand, as the absolute value of the measurement value y gets larger, the updated variance P^1_k counter-intuitively becomes higher than the predicted variance $\overline{P}^1_{k|k-2}$, which would cause incompatible covariance problem in the non-maneuvering mode $r_k = 1$. This phenomenon is observed because of the mode switch of the IMM filter to the maneuvering mode for large absolute values of the measurement y. The evidence of this fact can be seen in the mode probabilities μ^1_{k-1}, μ^2_{k-1} shown in Figure 3.3b with respect to the value of the measurement y.

The mechanism of the incompatible covariance problem for large absolute values of y is the following. Since the filter initially thinks that the non-maneuvering mode is active with high probability, the predicted mode probabilities of the non-maneuvering mode would dominate those of the maneuvering mode (with the prediction operation being unaware of the recent measurements). This means that only little amount of variance mixing would take place in the prediction operation resulting in the variances of the non-maneuvering mode remaining small. On the other hand, in the filter,



(a) Variances P_k^1 and $\overline{P}_{k|k-2}^1$



(b) Mode Probabilities μ_{k-1}^1 and μ_{k-1}^2

Figure 3.3: Illustration of the incompatible covariance problem on the JMLS in (3.19).

since a mode switch happens due to the measurements, updated variance of the non-maneuvering mode gets large values due to the maneuver and the resulting aggressive variance mixing operation.

3.4.2 Outlier Decorrelation Problem

Even when the covariances are compatible (i.e., no indefinite or negative-(semi-)definite covariance differences), at some rare time instances, the decorrelated quantities can be significant outliers compared to the predicted and current mean and covariance, which we call the *outlier decorrelation problem*. We illustrate the problem using the following example.

Example 2 *Let us consider a two-dimensional decorrelation problem where we try to apply decorrelation for Gaussian distributions over* xy*-plane. We choose the current mean and covariance* $\{\hat{\mathbf{x}}, \mathbf{P}\}$ *and the predicted mean and covariance* $\{\bar{\mathbf{x}}, \overline{\mathbf{P}}\}$ *as follows.*

$$\hat{\mathbf{x}} \triangleq \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathbf{P} \triangleq \begin{bmatrix} 12 & 2 \\ 2 & 5 \end{bmatrix}, \tag{3.23a}$$

$$\bar{\mathbf{x}} \triangleq \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \qquad \overline{\mathbf{P}} \triangleq \begin{bmatrix} 34 & 16 \\ 16 & 14 \end{bmatrix}. \tag{3.23b}$$

The corresponding decorrelated mean and covariance $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ *we obtain using* (3.18) *are then given as*

$$\tilde{\mathbf{x}} \triangleq \begin{bmatrix} 102 \\ -67 \end{bmatrix}, \qquad \widetilde{\mathbf{P}} \triangleq \begin{bmatrix} 368 & -228 \\ -228 & 158 \end{bmatrix}. \tag{3.24}$$

Note that all three covariances given above are positive definite. We illustrate the means and the 99% confidence ellipsoids of the current, predicted and decorrelated distributions in Figure 3.4. It is clearly seen in Figure 3.4 that the decorrelated quantities are significant outliers to the current and predicted quantities.

In our simulations we have observed that the outlier decorrelation problem appears especially when the predicted and the current covariances are very similar in size along one or more directions of the state-space, which usually leads to a decorrelated covariance with a high condition number. As an example notice that the semi-minor axes of the confidence ellipsoids for the current and predicted covariances in Figure 3.4 are similar to each other and the confidence ellipsoid of the decorrelated covariance is quite elongated which shows that the decorrelated covariance has a high condition

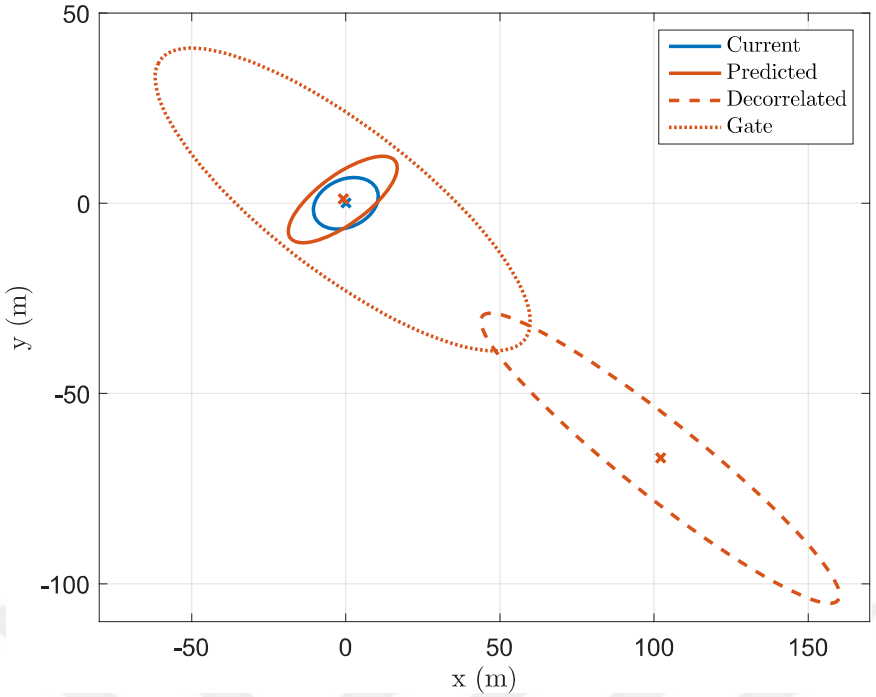


Figure 3.4: Means and the 99% confidence ellipsoids of the current, predicted and decorrelated quantities in the example for outlier decorrelation problem. The region into which the decorrelated mean has to fall for the outlier decorrelation problem not to happen is shown with the legend "gate".

number. With this observation, we can say that the outlier decorrelation problem is actually a weaker form of the incompatible covariance problem, which happens when the predicted and current covariances get close to each other along some directions in the state-space, i.e., when the covariances get close to incompatibility.

The outlier decorrelation problem can be detected during the decorrelation process by using a χ^2-test as follows. It should be clear that when we update the predicted quantities $\{\bar{\mathbf{x}}, \overline{\mathbf{P}}\}$ with the decorrelated quantities $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ using the Kalman filter measurement update (by considering $\tilde{\mathbf{x}}$ as the measurement and $\widetilde{\mathbf{P}}$ as its error covariance), we would get the current mean and covariance $\{\hat{\mathbf{x}}, \mathbf{P}\}$. Using this analogy we can detect the outlier decorrelation problem using the following χ^2-test which is usually used in gating in target tracking.

$$g_{\text{OD}} \triangleq (\tilde{\mathbf{x}} - \bar{\mathbf{x}})^{\text{T}} (\overline{\mathbf{P}} + \widetilde{\mathbf{P}})^{-1} (\tilde{\mathbf{x}} - \bar{\mathbf{x}}) \lessgtr \gamma_{\text{OD}}, \tag{3.25}$$

where the test statistic g_{OD} is χ^2 distributed with degrees of freedom n_x when there is

no outlier decorrelation problem, i.e., when the decorrelated quantities are not outliers to the predicted quantities. The constant γ_{OD} is an appropriate test threshold that can be set such that the test statistic g_{OD} is less than the threshold γ_{OD} with probability $P_{\mathrm{OD}} \approx 1$. In Figure 3.4 we show the region into which the decorrelated mean has to fall (with probability $P_{\mathrm{OD}} = 0.99$) for the outlier decorrelation problem not to happen with the legend "Gate".

3.4.3 Solution for Incompatible Covariance and Outlier Decorrelation Problems

In this section we present a heuristic solution for both of the problems discussed above. Two extreme solution ideas are the following.

- To abort information transmission when such problems appear;
- To abort only decorrelation and transmit the current information (without any decorrelation) when such problems appear.

The former solution might cause important loss of information or latency in the sensor network since no information is transmitted to other nodes. On the other hand, the latter solution might cause overconfidence in the fusion nodes since no decorrelation is applied. The solution that we present below will try to find the middle ground between these extreme ideas.

Note that two pieces of information enter into the decorrelation process, namely the current mean and covariance $\{\hat{\mathbf{x}}, \mathbf{P}\}$ and the predicted mean and covariance $\{\overline{\mathbf{x}}, \overline{\mathbf{P}}\}$. If the predicted covariance $\overline{\mathbf{P}}$ is not larger than the current covariance $\mathbf{P}$ in the positive definite sense, i.e., if $\overline{\mathbf{P}} \not> \mathbf{P}$, we would encounter the incompatible covariance problem. When $\overline{\mathbf{P}} > \mathbf{P}$, as the predicted covariance $\overline{\mathbf{P}}$ approaches the current covariance $\mathbf{P}$, i.e., as $\overline{\mathbf{P}} \to \mathbf{P}$, the decorrelated covariance $\widetilde{\mathbf{P}}$ would go to infinity, which means that effectively no information would be transmitted. Hence the case $\overline{\mathbf{P}} \to \mathbf{P}$ corresponds to aborting the transmission. On the other hand, as the predicted covariance $\overline{\mathbf{P}}$ gets larger and larger, the decorrelated quantities $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ would converge to the current quantities $\{\hat{\mathbf{x}}, \mathbf{P}\}$. This convergence would be equivalent to

aborting the decorrelation. With this reasoning, a middle ground between aborting transmission and aborting decorrelation can be found by simply sufficiently enlarging the predicted covariance with a scalar constant $\kappa > 1$, i.e., by setting $\overline{\mathbf{P}} \leftarrow \kappa\overline{\mathbf{P}}$, so that no incompatible covariance or outlier decorrelation problems are encountered. We will call the process of enlarging the predicted covariance $\overline{\mathbf{P}}$ *covariance inflation*. The enlargement of the predicted covariance $\overline{\mathbf{P}}$ would reduce the effect of the predicted quantities on the decorrelated quantities and the inflation constant κ should be selected

- Sufficiently large so that we avoid the incompatible covariance or outlier decorrelation problems;
- Sufficiently small so that decorrelation is still effective to the largest possible extent.

For incompatible covariance and outlier decorrelation problems we might need different values of the inflation constant κ as investigated below.

3.4.3.1 κ for the incompatible covariance problem

When incompatible covariance problem appears, we know that the difference matrix $\mathbf{P}^{-1} - \overline{\mathbf{P}}^{-1}$ becomes indefinite or negative-(semi-)definite. By replacing the predicted covariance $\overline{\mathbf{P}}$ with its inflated version $\kappa\overline{\mathbf{P}}$ we can find a constraint on the inflation constant κ to avoid the incompatible covariance problem as follows.

$$\mathbf{P}^{-1} - (\kappa\overline{\mathbf{P}})^{-1} > 0. \tag{3.26}$$

Using some straightforward matrix algebra we can see that a sufficient condition for the matrix $\mathbf{P}^{-1} - (\kappa\overline{\mathbf{P}})^{-1}$ to be positive definite is that

$$\kappa > \kappa_{\mathrm{IC}} \triangleq \lambda_{\max}(\mathbf{P}\overline{\mathbf{P}}^{-1}), \tag{3.27}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the argument matrix.

3.4.3.2 κ for the outlier decorrelation problem

When outlier decorrelation problem appears, we know that the test statistic g_{OD} in (3.25) becomes greater than the threshold γ_{OD}. By replacing the predicted covariance $\overline{\mathbf{P}}$ with its inflated version $\kappa\overline{\mathbf{P}}$ we can find a constraint on the inflation constant κ to avoid the outlier decorrelation problem as follows.

$$g_{\mathrm{OD}}(\kappa) \triangleq (\tilde{\mathbf{x}} - \overline{\mathbf{x}})^{\mathrm{T}}(\kappa\overline{\mathbf{P}} + \widetilde{\mathbf{P}})^{-1}(\tilde{\mathbf{x}} - \overline{\mathbf{x}}) < \gamma_{\mathrm{OD}}. \tag{3.28}$$

Note that the decorrelated quantities $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ in the test statistic $g_{\mathrm{OD}}(\kappa)$ above also depend on κ through the equations (3.18) with the substitution $\overline{\mathbf{P}} \leftarrow \kappa\overline{\mathbf{P}}$. Using the expressions

$$\hat{\mathbf{x}} = \overline{\mathbf{x}} + \kappa\overline{\mathbf{P}}(\kappa\overline{\mathbf{P}} + \widetilde{\mathbf{P}})^{-1}(\tilde{\mathbf{x}} - \overline{\mathbf{x}}), \tag{3.29a}$$

$$\mathbf{P} = \kappa\overline{\mathbf{P}} - \kappa\overline{\mathbf{P}}(\kappa\overline{\mathbf{P}} + \widetilde{\mathbf{P}})^{-1}\kappa\overline{\mathbf{P}}, \tag{3.29b}$$

which are the standard Kalman filter update formulae corresponding to the information form expressions in (3.18) (with the substitution $\overline{\mathbf{P}} \leftarrow \kappa\overline{\mathbf{P}}$), we can write

$$g_{\mathrm{OD}}(\kappa) = (\hat{\mathbf{x}} - \overline{\mathbf{x}})^{\mathrm{T}}(\kappa\overline{\mathbf{P}} - \mathbf{P})^{-1}(\hat{\mathbf{x}} - \overline{\mathbf{x}}). \tag{3.30}$$

The test statistic $g_{\mathrm{OD}}(\kappa)$ in (3.30) can be shown to be a convex and decreasing function of the inflation constant κ. As a result, we can easily find the solution of the equation $g_{\mathrm{OD}}(\kappa) = \gamma_{\mathrm{OD}}$ using Newton's method with the convergent iteration

$$\kappa^{(0)} = 1, \tag{3.31a}$$

$$\kappa^{(m+1)} = \kappa^{(m)} + \frac{(\hat{\mathbf{x}} - \overline{\mathbf{x}})^{\mathrm{T}}\mathbf{\Lambda}_m(\hat{\mathbf{x}} - \overline{\mathbf{x}}) - \gamma_{\mathrm{OD}}}{(\hat{\mathbf{x}} - \overline{\mathbf{x}})^{\mathrm{T}}\mathbf{\Lambda}_m\overline{\mathbf{P}}\mathbf{\Lambda}_m(\hat{\mathbf{x}} - \overline{\mathbf{x}})}, \tag{3.31b}$$

for $m \geq 0$ where $\mathbf{\Lambda}_m \triangleq \left(\kappa^{(m)}\overline{\mathbf{P}} - \mathbf{P}\right)^{-1}$. The solution $\kappa_{\mathrm{OD}} \triangleq \kappa^{(\infty)}$ of the equation $g_{\mathrm{OD}}(\kappa) = \gamma_{\mathrm{OD}}$ can be considered as a lower bound for the inflation constant κ to avoid the outlier decorrelation problem, i.e.,

$$\kappa > \kappa_{\mathrm{OD}} \triangleq \kappa^{(\infty)}. \tag{3.32}$$

Note that there is only a single lower bound κ_{IC} of the inflation constant κ in order to avoid the incompatible covariance problem. On the other hand, different lower bounds κ_{OD} for κ can be found for avoiding the outlier decorrelation problem depending on the selected test threshold γ_{OD}. As the test threshold γ_{OD} gets larger, the

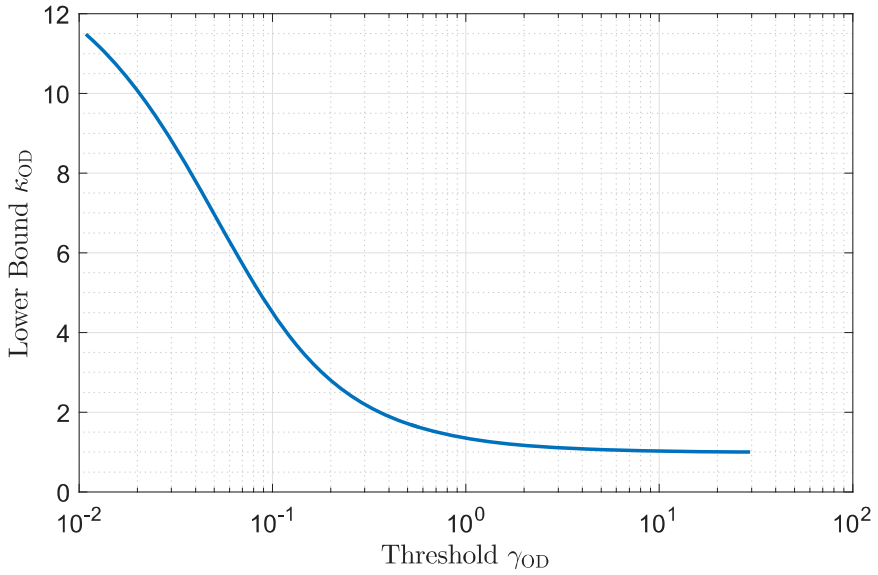


Figure 3.5: Relationship between the lower bound κ_{OD} and the threshold γ_{OD} for Example 2.

lower bound κ_{OD} for the inflation constant κ gets smaller, i.e., there is less need for inflating the predicted covariance $\overline{\mathbf{P}}$. As γ_{OD} goes to infinity, there remains no need for covariance inflation and as γ_{OD} goes to zero, the lower bound κ_{OD} goes to infinity, i.e., the predicted covariance should be inflated by an infinite factor. The relationship between the lower bound κ_{OD} and the threshold γ_{OD} is shown in Figure 3.5 for the quantities in Example 2. For Example 2, the gate statistics $g_{\mathrm{OD}}(\kappa)$ for $\kappa = 1$, i.e., the gate statistics with no covariance inflation, turns out to be 29.5. We do not plot the values of κ_{OD} for $\gamma_{\mathrm{OD}} > 29.5$ in Figure 3.5 since we do not need covariance inflation for these values of the threshold γ_{OD}.

We illustrate the effect of covariance inflation with different values of the inflation constant $\kappa \in \{1, 1.05, 1.5\}$ on the predicted and decorrelated quantities of Example 2 in Figure 3.6. A surprising observation in Figure 3.6 is that as the inflation constant κ gets larger, the gates get smaller. This is due to the fact that the decorrelated covariance $\widetilde{\mathbf{P}}$, which is a part of the gate covariance $\kappa\overline{\mathbf{P}} + \widetilde{\mathbf{P}}$, gets much more smaller than the predicted covariance $\kappa\overline{\mathbf{P}}$ gets larger as κ increases. As a result, the gate covariance can actually get smaller as κ increases. Note also that as κ gets larger, the decorrelated quantities approach the current quantities.

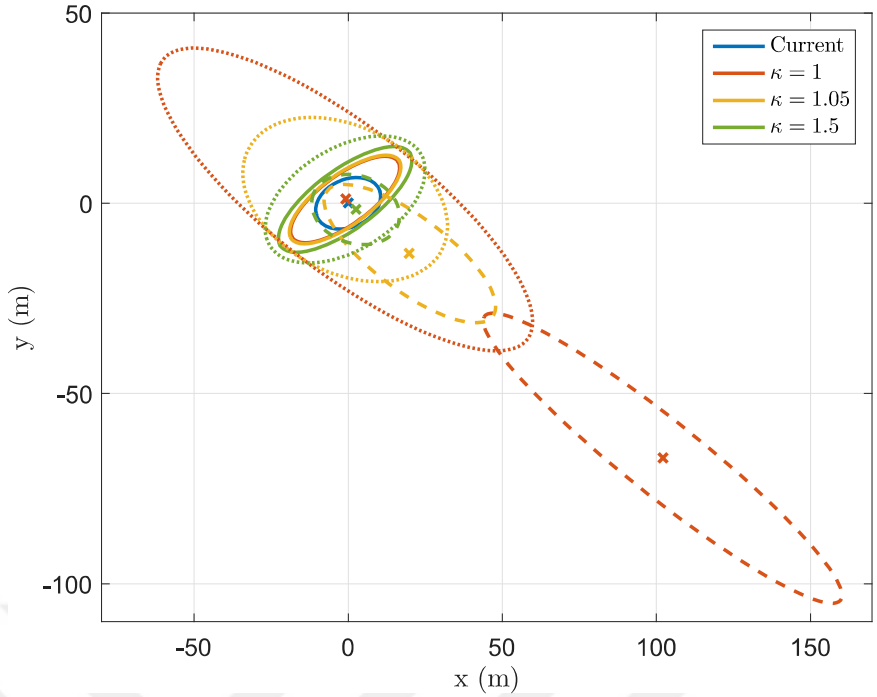


Figure 3.6: Current, predicted and decorrelated quantities for different values of the inflation constant κ for Example 2. The solid, dashed and dotted lines denote the predicted covariance, decorrelated covariance and the gate, respectively.

3.4.3.3 Use of covariance inflation in OLD & MLD for the IMM filter

In the implementation of OLD and MLD, at each decorrelation step we check the existence of the incompatible covariance and outlier decorrelation problems. When one or more of these problems are detected, a compensation in the form of covariance inflation is applied. The related steps to be added into the OLD and MLD are given in Algorithms 6 and 7, respectively.

In both MLD and OLD we first check for the existence of the incompatible covariance problem. If there is one, we calculate the lower bound κ_{IC} for the inflation constant κ in OLD and inflate the predicted covariance with the constant $c_{\text{IC}}\kappa_{\text{IC}}$ where $c_{\text{IC}} > 1$ is a safety factor. In MLD we first check for the existence of the incompatible covariance problem for each mode ($i = 1, \ldots, N_r$). If we find such a problem in a mode, we calculate the lower bound κ^i_{IC}, $i = 1, \ldots, N_r$ for the inflation constant κ for that mode. If there exists incompatible covariance problem in at least one mode, we inflate the predicted covariances of all modes with the maximum of the values $c_{\text{IC}}\kappa^i_{\text{IC}}$,

$i = 1, \ldots, N_r$, where $c_{\mathrm{IC}} > 1$ is a safety factor. In MLD, in the case of incompatible covariance problem in a mode, the predicted covariances of all modes should be inflated with the same inflation constant in order not to create artificial irregularities in the decorrelated weights $\tilde{\mu}_k^{1,i}$, $i = 1, \ldots, N_r$. Inflating only the predicted covariance of the mode(s) in which there is an incompatible covariance problem would cause unintended irregularities in the decorrelated weights $\tilde{\mu}_k^{1,i}$, $i = 1, \ldots, N_r$ which are highly dependent on the predicted covariances.

After handling the incompatible covariance problems, we then check for the existence of the outlier decorrelation problems in both OLD and MLD. If there is one, we calculate the lower bound κ_{OD} for the inflation constant κ using the Newton method with a fixed number of iterations M_{OD} in OLD and inflate the predicted covariance with the constant $c_{\mathrm{OD}}\kappa_{\mathrm{OD}}$ where $c_{\mathrm{OD}} > 1$ is a safety factor. In MLD we check for the existence of the outlier decorrelation problem for each mode ($i = 1, \ldots, N_r$). If we find such a problem in a mode, we calculate the lower bound κ_{OD}^i, $i = 1, \ldots, N_r$ for the inflation constant κ for that mode similarly to what we do in OLD. In MLD if there exists outlier decorrelation problem in at least one mode, we inflate the predicted covariances of all modes with the maximum of the values $c_{\mathrm{OD}}\kappa_{\mathrm{OD}}^i$, $i = 1, \ldots, N_r$ to avoid artificial irregularities in the decorrelated weights $\tilde{\mu}_k^i$, $i = 1, \ldots, N_r$.

Algorithm 6 (OLD Covariance Inflation) *To be inserted before the decorrelation step of OLD in Algorithm 4*

- *Given:* $\hat{\mathbf{x}}_k^1$, $\mathbf{P}_k^1$, $\bar{\mathbf{x}}_{k|k-N}^1$, $\overline{\mathbf{P}}_{k|k-N}^1$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Check and compensate for the incompatible covariance problem*

  $\kappa_{\mathrm{IC}} = \lambda_{\max}\big(\mathbf{P}_k^1(\overline{\mathbf{P}}_{k|k-N}^1)^{-1}\big)$.
  If $\kappa_{\mathrm{IC}} \geq 1$,
  Set $\overline{\mathbf{P}}_{k|k-N}^1 \leftarrow c_{\mathrm{IC}}\kappa_{\mathrm{IC}}\overline{\mathbf{P}}_{k|k-N}^1$.
  End

- *Check and compensate for the outlier decorrelation problem*

  $g_{\mathrm{OD}} = (\hat{\mathbf{x}}_k^1 - \bar{\mathbf{x}}_{k|k-N}^1)^{\mathrm{T}}(\overline{\mathbf{P}}_{k|k-N}^1 - \mathbf{P}_k^1)^{-1}(\hat{\mathbf{x}}_k^1 - \bar{\mathbf{x}}_{k|k-N}^1)$.

If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,

$\kappa^{(0)} = 1$.

For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

$\mathbf{\Lambda}_m = (\kappa^{(m)} \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)^{-1}$.

$\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}} \mathbf{\Lambda}_m (\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})}{(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}} \mathbf{\Lambda}_m \overline{\mathbf{P}}^1_{k|k-N} \mathbf{\Lambda}_m (\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})}$.

End

$\kappa_{\mathrm{OD}} = \kappa^{(M_{\mathrm{OD}})}$.

Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{OD}} \kappa_{\mathrm{OD}} \overline{\mathbf{P}}^1_{k|k-N}$.

End

Algorithm 7 (MLD Covariance Inflation) *To be inserted before the decorrelation step of MLD in Algorithm 5.*

- *Given:* $\{\hat{\mathbf{x}}^{1,i}_k, \mathbf{P}^{1,i}_k\}_{i=1}^{N_r}$, $\{\bar{\mathbf{x}}^{1,i}_{k|k-N}, \overline{\mathbf{P}}^{1,i}_{k|k-N}\}_{i=1}^{N_r}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Check and compensate for the incompatible covariance problem:*

For $i = 1, \ldots, N_r$,

$\kappa^i_{\mathrm{IC}} = \lambda_{\max}\big(\mathbf{P}^{1,i}_k (\overline{\mathbf{P}}^{1,i}_{k|k-N})^{-1}\big)$.

End

$\kappa_{\mathrm{IC}} = \max\{\kappa^1_{\mathrm{IC}}, \ldots, \kappa^{N_r}_{\mathrm{IC}}\}$.

If $\kappa_{\mathrm{IC}} \geq 1$,

For $i = 1, \ldots, N_r$,

Set $\overline{\mathbf{P}}^{1,i}_{k|k-N} \leftarrow c_{\mathrm{IC}} \kappa_{\mathrm{IC}} \overline{\mathbf{P}}^{1,i}_{k|k-N}$.

End

End

- *Check and compensate for the outlier decorrelation problem:*

For $i = 1, \ldots, N_r$,

$g_{\mathrm{OD}} = (\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} (\overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1} (\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})$.

If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,

$\kappa^{(0)} = 1$.

For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

$\mathbf{\Lambda}_m = (\kappa^{(m)}\overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1}$.

$\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} \mathbf{\Lambda}_m (\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})}{(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} \mathbf{\Lambda}_m \overline{\mathbf{P}}^{1,i}_{k|k-N} \mathbf{\Lambda}_m (\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})}$.

End

$\kappa^i_{\mathrm{OD}} = \kappa^{(M_{\mathrm{OD}})}$.

Else

$\kappa^i_{\mathrm{OD}} = 0$.

End

End

$\kappa_{\mathrm{OD}} = \max\{\kappa^1_{\mathrm{OD}}, \ldots, \kappa^{N_r}_{\mathrm{OD}}\}$.

If $\kappa_{\mathrm{OD}} \geq 1$,

For $i = 1, \ldots, N_r$,

Set $\overline{\mathbf{P}}^{1,i}_{k|k-N} \leftarrow c_{\mathrm{OD}} \kappa_{\mathrm{OD}} \overline{\mathbf{P}}^{1,i}_{k|k-N}$.

End

End

3.5 Fusion

In this section we describe the fusion operation to be applied by the local agent $\mathcal{F}_2$ upon receiving the decorrelated information from $\mathcal{F}_1$. After receiving the decorrelated distribution $\tilde{p}^1(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)$, i.e., its mean and covariance $\{\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k\}$, from $\mathcal{F}_1$ at time k, $\mathcal{F}_2$ has to incorporate this information into the summary statistics of its own IMM filter.

The current joint posterior distribution of the IMM filter of $\mathcal{F}_2$ before fusion is given as

$$p^2(\mathbf{x}_k, r_k | \mathcal{I}^1_{0:k-N}, \mathbf{y}^2_{0:k}) = \mu^{2,r_k}_{k|k-N,k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,r_k}_{k|k-N,k}, \mathbf{P}^{2,r_k}_{k|k-N,k}), \tag{3.33}$$

where $\mathcal{I}^1_{0:k-N}$ denotes the information that $\mathcal{F}_2$ has got from $\mathcal{F}_1$ so far (see Appendix A for the reasons for the use of this notation). Note that the summary statistics of the IMM filter of $\mathcal{F}_2$ before fusion are given as $\{\mu^{2,i}_{k|k-N,k}, \hat{\mathbf{x}}^{2,i}_{k|k-N,k}, \mathbf{P}^{2,i}_{k|k-N,k}\}^{N_r}_{i=1}$. $\mathcal{F}_2$ applies naive fusion to fuse the pieces of information described by the distributions

$\tilde{p}^1(\mathbf{x}_k)$ and $p^2(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k-N}, \mathbf{y}^2_{0:k})$.

$$p^2(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k}, \mathbf{y}^2_{0:k}) \propto \tilde{p}^1(\mathbf{x}_k)p^2(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k-N}, \mathbf{y}^2_{0:k}), \tag{3.34a}$$

$$= \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)\mu^{2,r_k}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,r_k}_{k|k-N,k}, \mathbf{P}^{2,r_k}_{k|k-N,k}), \tag{3.34b}$$

$$= \mu^{2,r_k}_{k|k,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,r_k}_{k|k,k}, \mathbf{P}^{2,r_k}_{k|k,k}), \tag{3.34c}$$

where

$$\mu^{2,r_k}_{k|k,k} \propto \mu^{2,r_k}_{k|k-N,k}\mathcal{N}(\tilde{\mathbf{x}}^1_k; \hat{\mathbf{x}}^{2,r_k}_{k|k-N,k}, \mathbf{P}^{2,r_k}_{k|k-N,k} + \widetilde{\mathbf{P}}^1_k), \tag{3.35a}$$

$$\hat{\mathbf{x}}^{2,r_k}_{k|k,k} \triangleq \mathbf{P}^{2,r_k}_{k|k,k}\big((\mathbf{P}^{2,r_k}_{k|k-N,k})^{-1}\hat{\mathbf{x}}^{2,r_k}_{k|k-N,k} + (\widetilde{\mathbf{P}}^1_k)^{-1}\tilde{\mathbf{x}}_k\big), \tag{3.35b}$$

$$\mathbf{P}^{2,r_k}_{k|k,k} \triangleq \big((\mathbf{P}^{2,r_k}_{k|k-N,k})^{-1} + (\widetilde{\mathbf{P}}^1_k)^{-1}\big)^{-1}. \tag{3.35c}$$

The summary statistics of the IMM filter of $\mathcal{F}_2$ after fusion are then given as $\{\mu^{2,i}_{k|k,k}, \hat{\mathbf{x}}^{2,i}_{k|k,k}, \mathbf{P}^{2,i}_{k|k,k}\}_{i=1}^{N_r}$. The fusion operations made in the second node are illustrated in Figure 3.7.

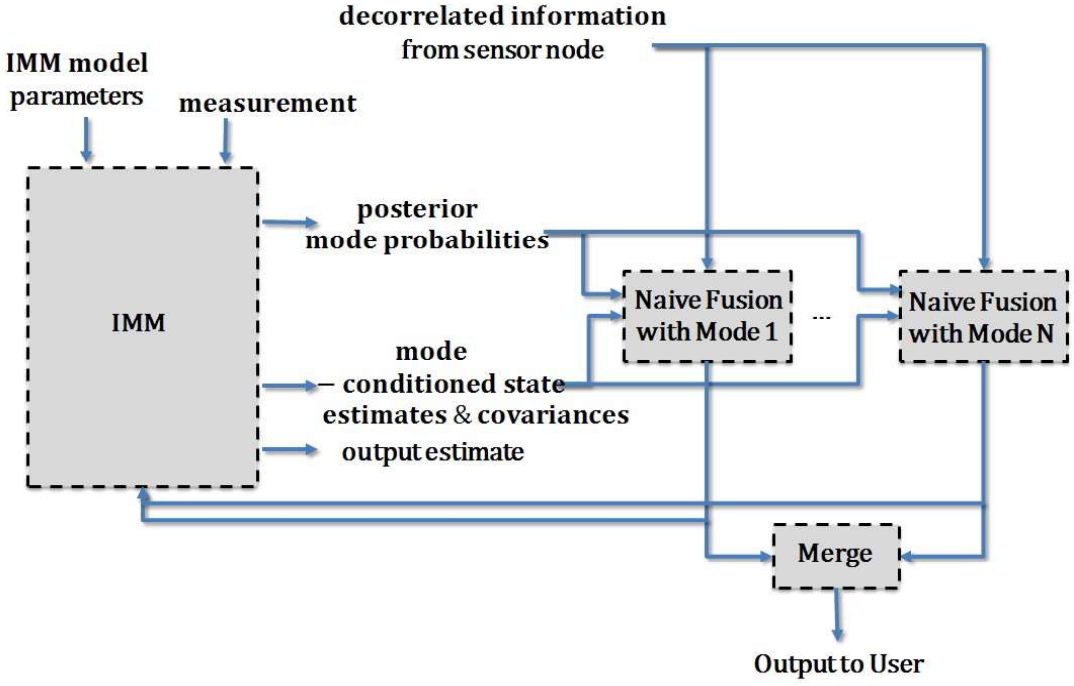


Figure 3.7: Fusion operation for the IMM filter.

3.6 Simulation Results

In order to test the performance of the investigated decorrelation strategy, we consider two 2D maneuvering target tracking scenarios, which are given in two different subsections below.

3.6.1 Model-Match Case

For this case, we generate the true target data using a JMLS in which the base-state $\mathbf{x}_k \in \mathbb{R}^4$ is composed of Cartesian xy positions and velocities and the mode-state $r_k \in \{1, 2\}$ takes two values. The JMLS has two nearly constant velocity models (directly written in discrete-time) which only differ in their process noise covariance values.

$$\mathbf{A}_1 = \mathbf{A}_2 = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{bmatrix}, \qquad \mathbf{B}_1 = \mathbf{B}_2 = \begin{bmatrix} \frac{T^2}{2}\mathbf{I}_2 \\ T\mathbf{I}_2 \end{bmatrix},$$

$$\mathbf{C}_1 = \mathbf{C}_2 = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}, \qquad \mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}_2, \tag{3.36a}$$

$$\mathbf{Q}_i = \sigma_i^2 \mathbf{I}_2, \quad i = 1, 2, \qquad \mathbf{R}_1 = \mathbf{R}_2 = \sigma_y^2 \mathbf{I}_2, \tag{3.36b}$$

where $\sigma_1 = 1\,\mathrm{m/s^2}$ and $\sigma_2 = 10\,\mathrm{m/s^2}$, respectively and $\sigma_y = 20\,\mathrm{m}$. The diagonal elements of the transition probability matrix $\mathbf{\Pi} \in \mathbb{R}^{2\times 2}$ for the JMLS are given as $\pi_{11} = \pi_{22} = 0.95$. We initiate the base and mode-states at the values given below.

$$\mathbf{x}_0 = \begin{bmatrix} 0\,\mathrm{m} & 0\,\mathrm{m} & 25\,\mathrm{m/s} & 25\,\mathrm{m/s} \end{bmatrix}^{\mathrm{T}}, \qquad r_0 = 1.$$

We consider the two local agents shown as $\mathcal{F}_1$ and $\mathcal{F}_2$ in Figure 1.3 running their own local IMM filters on the measurements collected by the local sensors shown as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The IMM filters of both nodes have identical parameters which are matched to the true target data. In the simulations the sampling period of $T = 1\,\mathrm{s}$ is used and each node collects a total of 100 measurements. The mode-conditioned estimates and covariances of the local IMM filters are initialized using standard two-point initiation from the first two local measurements. The initial mode probabilities are selected to be equal in both IMM filters. Throughout the simulations we give detailed results for the communication period $N = 3$. We are going to illustrate the summary of the results for the cases $N = 1, 2, 3, 4, 5$.

3.6.1.1 Comparison of Different OLD & MLD Approaches

We first compare the performances of different OLD & MLD approaches. For this purpose we present the results obtained by OLD/MLD with the following alternative solutions for the incompatible covariance and outlier decorrelation problems.

- Aborted Transmission: The information is not transmitted to $\mathcal{F}_2$ when the incompatible covariance or outlier decorrelation problem is encountered in $\mathcal{F}_1$ at the communication times.
- Aborted Decorrelation: Decorrelation is not applied and the un-decorrelated (i.e., correlated) information is transmitted to $\mathcal{F}_2$ when the incompatible covariance or outlier decorrelation problem is encountered.
- Covariance Inflation IC: The solution proposed in Section 3.4 is used for only the incompatible covariance problem. The parts related to the outlier decorrelation problem in Algorithms 6 and 7 are not applied. Note that covariance inflation should always be applied for avoiding the incompatible covariance problem, but the parts related to the outlier decorrelation problem are optional.
- Covariance Inflation IC+OD: The solution proposed in Section 3.4 is used for both incompatible covariance and outlier decorrelation problems. Hence all parts of Algorithms 6 and 7 are applied. The gate probability for the outlier decorrelation problem is selected as $P_{\mathrm{OD}} = 0.99$, which results in the gate/test threshold $\gamma_{\mathrm{OD}} \approx 13.28$.

A total of 10000 Monte Carlo (MC) runs are made by changing the target trajectory and measurement noise realizations in each run. The root mean square (RMS) position errors, RMS velocity errors and the mean normalized estimation error squared (NEES) [21] divided by the state dimension $n_x = 4$ for the fusion node $\mathcal{F}_2$ are shown in Figure 3.8a, Figure 3.8b and Figure 3.8c, respectively.[1] Note that we present only

[1] The NEES of the local agent $\mathcal{F}_s$ at time k (divided by the state dimension n_x) is defined as

$$\mathrm{NEES}_k^s \triangleq \frac{1}{n_x}(\hat{\mathbf{x}}_k^s - \mathbf{x}_k)^{\mathrm{T}}(\mathbf{P}_k^s)^{-1}(\hat{\mathbf{x}}_k^s - \mathbf{x}_k). \tag{3.37}$$

The mean NEES at time k is the average of NEES_k^s values over the MC ensemble. The closer the value of the mean NEES to unity, the better the consistency properties of the filter are, i.e., the filter-calculated (error) covariance matrix $\mathbf{P}_k^s$ represents the true estimation error $\hat{\mathbf{x}}_k^s - \mathbf{x}_k$ in a better way. The larger/smaller values of mean NEES than unity indicate an overconfident/underconfident filter.

the results at the communication/fusion times to avoid cluttered plots.

It is seen in the Figure 3.8 that, in terms of the RMS errors, the best OLD/MLD approach is the one using covariance inflation for both incompatible covariance and outlier decorrelation problems, i.e., OLD/MLD: Covariance Inflation IC+OD. Aborting transmission has bad RMS error results because it simply prevents the recent information from reaching the fusion node. Aborting decorrelation, on the other hand, causes overconfidence in the fusion node which prevents consistent processing of the future data. The overconfidence caused by the aborted decorrelation is confirmed by the high mean NEES values of OLD/MLD: Aborted Transmission in Figure 3.8c. It is observed that in OLD applying covariance inflation for only incompatible covariance problem has nearly the same performance with applying it for both incompatible covariance and outlier decorrelation problems. It is expected result because the outlier problem does not occur frequenctly as in MLD. In MLD, it is seen that applying covariance inflation for only incompatible covariance problem gives significantly bad RMS error results (especially for velocity estimation) compared to applying it for both incompatible covariance and outlier decorrelation problems. It is also seen in Figure 3.8c that MLD: Covariance Inflation IC is more overconfident than MLD: Covariance Inflation IC+OD. This shows that outlier decorrelation problem, which we interpret as a weaker form of incompatible covariance problem, should also be compensated in decorrelation.

In the subsequent parts of this section, we are only going to consider the OLD approach called OLD: Covariance Inflation IC+OD, which we will call simply as the OLD and the MLD approach called MLD: Covariance Inflation IC+OD, which we will call simply as the MLD, since these have been seen to be the best OLD/MLD approach.

3.6.1.2 Comparison of OLD & MLD with Alternative Approaches

In the following we compare the performances of the following algorithms on the same scenario.

- Baseline Methods:

– Local Agent: This is the local IMM filter of the agent $\mathcal{F}_1$.

– Centralized IMM Filter (Centralized): This is a centralized IMM filter using the measurements of both $\mathcal{F}_1$ and $\mathcal{F}_2$.

– Restarted Local Filter (Restarted LF): In this method, the decorrelated quantities in $\mathcal{F}_1$ are calculated by running an auxiliary (shadow) IMM filter which is restarted after every information transmission (in addition to the primary/principal IMM filter which is not restarted). The auxiliary IMM filter has the same parameters as the primary IMM filter and it is initiated after each data transmission using equal initial mode probabilities. The initial mode-conditioned state estimates of the auxiliary IMM filter are set to those of the primary IMM filter at the same time instant (the time instant of the data transmission). The initial mode-conditioned covariances of the auxiliary IMM are set to $10^{14}\mathbf{I}_4$, where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$. This initial mode-conditioned covariance selection makes the values of initial mode-conditioned state estimates unimportant. The auxiliary IMM filter is run with the local measurements of $\mathcal{F}_1$ until the next transmission time and the output estimate and covariance at the next transmission time is transmitted to $\mathcal{F}_2$. The main steps of Restarted LF for the IMM filter, whose block diagram for $N = 3$ is given in Figure 3.9, are given below;

  Suppose that the last information transmission happened at time $k - N$ and the next information transmission will be at time k. Then we apply the following steps in the auxiliary IMM filter.

  * Initialization: Initialize the auxiliary IMM filter at time $k - N + 1$ with independent information.

$$\mu^{r_k}_{k-N+1|k-N+1} = 1/N_r, \tag{3.38a}$$

$$\hat{\mathbf{x}}^{r_k}_{k-N+1|k-N+1} = \hat{\mathbf{x}}^{r_k}_{k-N|k-N}, \tag{3.38b}$$

$$\mathbf{P}^{r_k}_{k-N+1|k-N+1} = 10^{14}\mathbf{I}_n, \tag{3.38c}$$

    where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$.

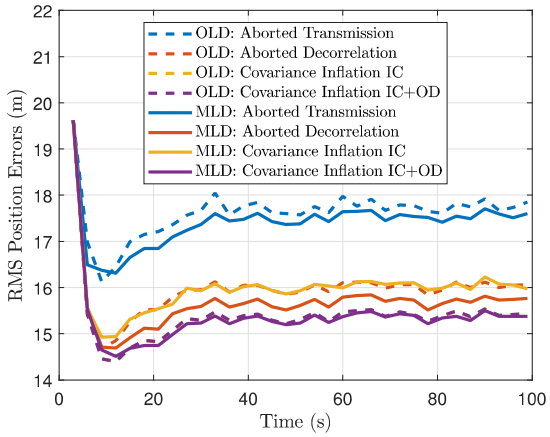
  * Recursion: Update the IMM filter recursively in time using the measurements $\mathbf{y}_{k-N+1:k}$ and obtain the weights $\{\mu^i_{k|k}\}^{N_r}_{i=1}$, estimates $\{\hat{\mathbf{x}}^i_{k|k}\}^{N_r}_{i=1}$ and covariances $\{\mathbf{P}^i_{k|k}\}^{N_r}_{i=1}$.

* Merging: Merge the mode-conditioned statistics.

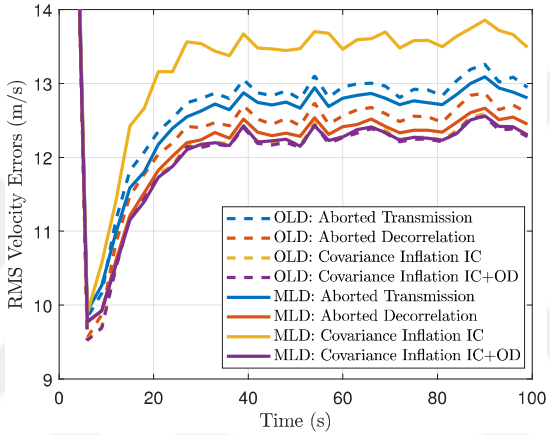
$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_r} \mu_k^i \hat{\mathbf{x}}_k^i, \tag{3.39a}$$

$$\mathbf{P}_k = \sum_{i=1}^{N_r} \mu_k^i \left[\mathbf{P}_k^i + (\hat{\mathbf{x}}_k^i - \hat{\mathbf{x}}_k)(\cdot)^{\mathrm{T}} \right]. \tag{3.39b}$$

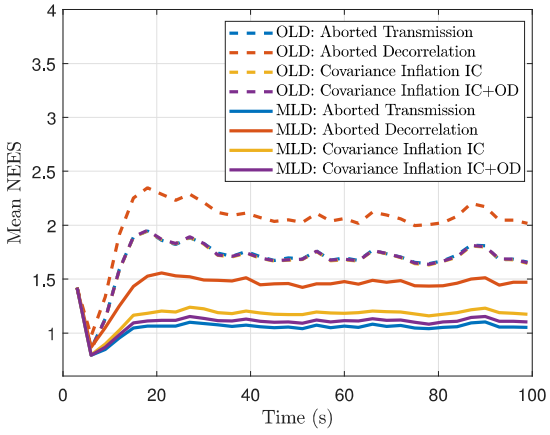
* Transmission: Transmit merged estimate $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ to the fusion node.



(a) RMS Position Errors



(b) RMS Velocity Errors



(c) Mean NEES

Figure 3.8: Comparison of performances of different OLD and MLD approaches for $N = 3$.

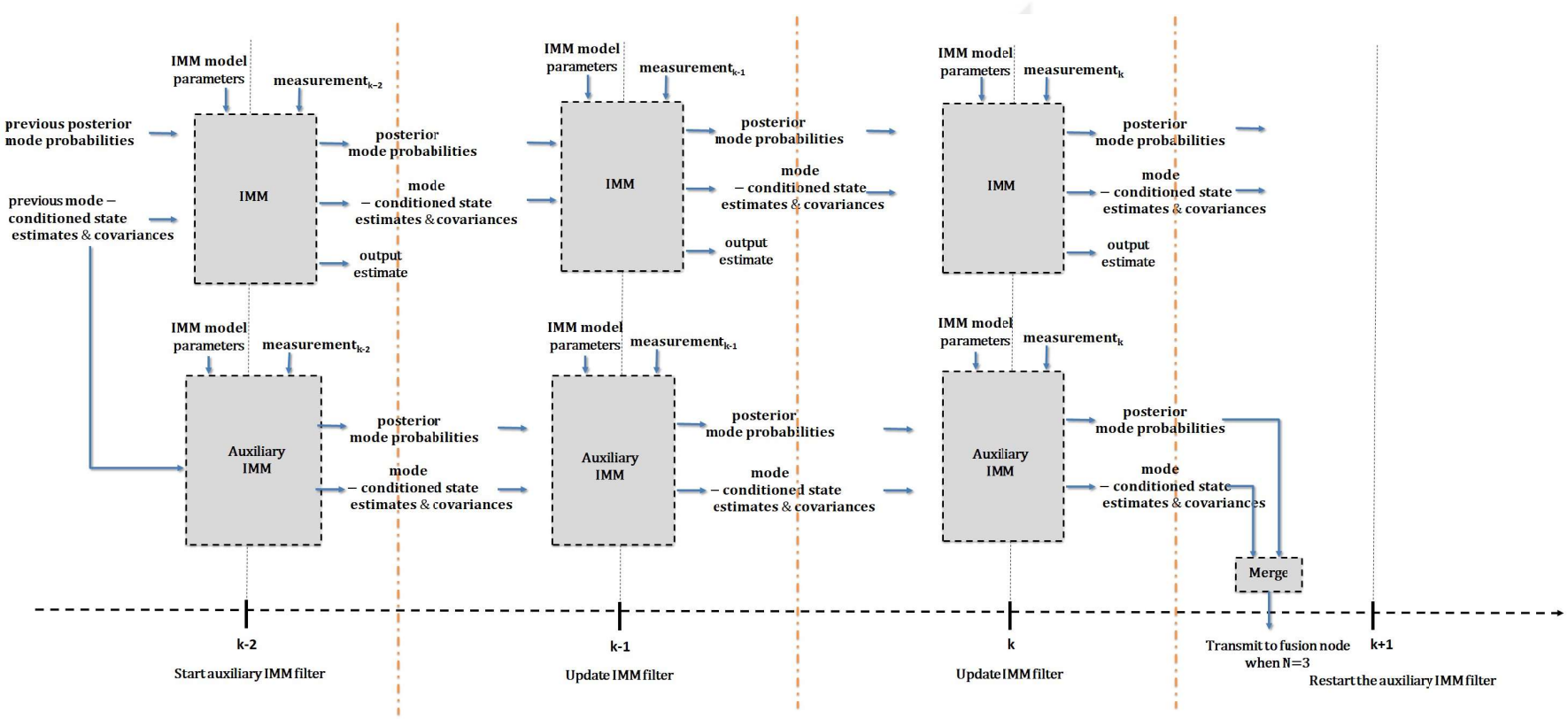


Figure 3.9: Restarted LF for the IMM filter.

- Proposed Method:
  - OLD: This is the OLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.
  - MLD: This is the MLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.

RMS position errors, RMS velocity errors and mean NEES values obtained from 10000 MC runs are shown in Figure 3.10a, Figure 3.10b and Figure 3.10c, respectively. As expected we see that the centralized approach yields the best RMS error performance. In terms of RMS position and velocity errors, both MLD and Restarted LF have very similar performances. The results for OLD are slightly worse than the other fusion methods which have very similar performance.
We see, on the other hand, that OLD has the worst NEES performance. This confirms that the effects of the approximations are more significant in OLD. We see, on the other hand, that the Restarted LF has higher mean NEES values than MLD. This has been observed to happen due to the Restarted LF being unable to estimate the maneuvers, i.e., Restarted LF being unable to compensate for the process noise, sufficiently well. This is because the (Restarted) IMM filter takes time before compensating for the maneuver with sufficient accuracy after a restart.

We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 10000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 3.11a, Figure 3.11b, and Figure 3.11c, respectively. Here, the averages of the RMS errors and mean NEES values are taken over the time instants in the interval $[20\,\mathrm{s},\ 100\,\mathrm{s}]$ to avoid the initial transient period. It is seen that, for $N = 1$ there are no approximations in MLD other than merging the mode-conditioned decorrelated estimates. As a result MLD gives almost the same results as the centralized filter for $N = 1$, similar to the Restarted LF. Since there is no effect of process noise when $N = 1$, MLD is also perfectly consistent for $N = 1$. OLD, on the other hand, suffers from high RMS errors and very high NEES values even for $N = 1$. For $N \geq 2$ approximations involved in MLD makes its results differ from those of the centralized filter as expected. For all N, MLD and Restarted LF show similar performances in RMS errors. For $N = 1; 2; 3$, OLD is worse than MLD in both RMS errors

and average NEES. On the other hand for $N \geq 4$ RMS errors of OLD are slightly better than those of MLD. However in terms of consistency OLD still remains worse than MLD. As N increases, the effects of previous communication get less important and effects of process noise start to dominate. OLD seems to be more effective for large N (less dominant effects of previously communicated information) in calculating the decorrelated estimates (better RMS errors), on the other hand, its inaccurate covariance calculations result in inconsistency problems. Although MLD seems to be slightly worse in decorrelated estimate calculation, its uncertainty estimates, i.e., decorrelated covariances, reflect its own errors in a better way than OLD. As N gets higher we see that Restarted LF becomes significantly inconsistent due to the inability of the Restarted IMM filter in estimating and compensating for the maneuvers (i.e., the process noise) sufficiently well in a short-time period after the restart. MLD consistency is also affected by the process noise but to a much less extent than the Restarted LF.

3.6.2 Model-Mismatch Case

In this case we use the fixed true 2D target trajectory shown in Figure 3.12 where the target makes a coordinated-turn maneuver in the middle of the scenario. The fusion architecture is kept the same as the model-match case and the sensors collect position measurements with measurement noise standard deviation $\sigma_y = 20\,m$. The IMM filters of the nodes use the parameters of the JMLS given in the model-match case. Hence, in this scenario, we have a mismatch between the true target motion and the filter parameters. We run the same algorithms as in Section 3.6.1.2 and make 1000 MC runs by changing the measurement noise realizations in each run. The reason why the number of MC runs in the model-mismatch case is less than the number of MC cases in the model-match case is that the MC results in the case of model-mismatch of different approaches are smooth enough to be easily compared among each other. We show the resulting RMS position errors, RMS velocity errors and mean NEES values in Figure 3.13a, Figure 3.13b and Figure 3.13c, respectively.

Note that in the model-mismatch case the closeness of the performance measures to those of centralized IMM filter is important. A striking observation in Figure 3.13a

and 3.13b is that OLD beats the centralized filter in the non-maneuvering parts of the scenario. This can be explained as follows. When the target moves with constant velocity, there is no process noise. On the other hand, IMM filters use non-zero process noise standard deviations making them sub-optimal. Since OLD cannot compensate correlations caused by previous communications sufficiently well, IMM filter in the fusion node becomes over-confident. This is, in effect, equivalent to the IMM filter in the fusion node using a lower process noise covariance which is closer to the true target behavior making it closer to optimality. Note that this non-intuitive behavior of OLD ends as soon as maneuver starts when OLD performance gets worse than the other fusion methods due to over-confidence, which is also visible in Figure 3.13c.

The most important deviations of the performance measures of MLD and Restarted LF from those of the centralized IMM filter are in the RMS velocity error performance in Figure 3.13b. These are detailed below.

- Performance of MLD is slightly better than that of the centralized IMM filter in the non-maneuvering parts of the scenario. This surprising phenomenon can actually be explained as follows. When the target moves with constant velocity, there is no process noise. On the other hand, IMM filters use non-zero process noise standard deviations making them suboptimal. MLD can compensate for most of the effects of the correlations caused by previous communications but it cannot do this perfectly. As a result, the corresponding IMM filter in the fusion node becomes slightly overconfident. This is, in effect, equivalent to the IMM filter in the fusion node using a slightly lower process noise covariance, which is closer to the true target behavior and to optimality. The evidence of the slight overconfidence of MLD cannot be detected in Figure 3.13c since the (squared) estimation errors are reduced by approximately the same amount as the calculated covariances get smaller, which results in the NEES values virtually remaining the same.

- Performance of Restarted LF is slightly worse than that of the centralized IMM filter in the maneuvering parts of the scenario. This is again due to the inability of Restarted LF to compensate for the process noise in a short time period after the restart. As a result, more process noise enters into the estimation

errors. This makes the Restarted LF overconfident compared to the centralized IMM filter in the maneuvering parts of the scenario, which is also visible in Figure 3.13c.

We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 1000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 3.14a, Figure 3.14b, and Figure 3.14c, respectively. Here, the averages of the RMS errors and mean NEES values are taken again over the time instants in the interval $[20\,\text{s},\ 150\,\text{s}]$ to avoid the initial transient period. It is seen that the results for the Restarted LF, MLD and centralized filter are almost the same for $N = 1$. OLD has better RMS results in the non-maneuvering parts of the scenario due to its inability to decorrelate the previously communicated information. It is seen that Restarted LF has a similar behavior to OLD (as evidenced by the RMS velocity errors) as N gets larger. This shows that the Restarted LF has smaller decorrelated covariance than necessary as N gets larger. The Restarted LF is also unable to estimate and compensate for the maneuver in the short period after the restart, which results in inconsistency due to process noise during the maneuver for $N \geq 3$. Note that we do not claim that MLD is immune to maneuver, in fact, it becomes slightly inconsistent during the maneuver as well when N is increased and the process noise effects become dominant. However it is more consistent than Restarted LF (i.e., it has closer results to the centralized filter) since the Restarted LF is over-sensitive to the process noise. Overall MLD is the decorrelation method yielding the closest results to the centralized filter in both RMS errors and average NEES for the scenario with the model-mismatch.

3.7 Conclusion

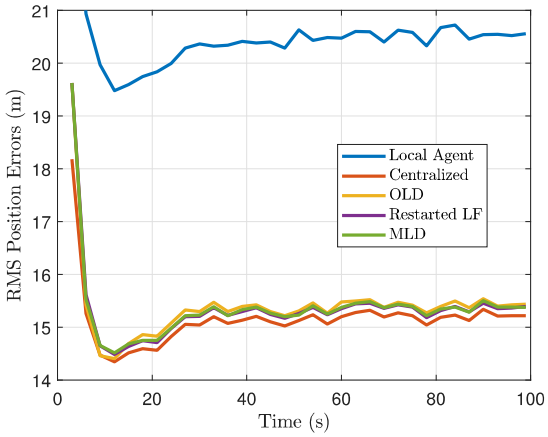
We have investigated a decorrelation approach for the IMM filter which can be alternative to the standard local auxiliary/shadow filter restart approach. In our investigations we examined the issues that might be encountered in applying decorrelation in an IMM filter in detail and proposed a method, called Covariance Inflation, for alleviating the effects of such problems. The proposed solution Covariance Inflation is used within OLD and MLD approaches for the IMM filter and both MLD and OLD

algorithm steps are updated to include the covariance inflation solution.

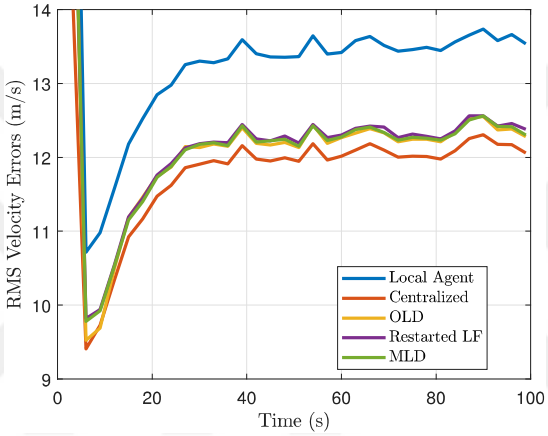
In simulation results, OLD and MLD are compared with both alternative solutions without covariance inflation (i.e. aborted transmission and aborted decorrelation) and alternative approaches previously proposed in literature (i.e. Centralized fusion, Restarted local filter).

We have seen that the Covariance Inflation method, having smaller RMS errors and being more consistent than its alternatives (i.e. aborted transmission and aborted decorrelation), seems to be a reasonable middle ground between the extreme solutions of aborting transmission and aborting decorrelation. We, on the other hand, have seen that the OLD, which makes the decorrelation on the IMM filter's output estimates, has undesirable consistency characteristics. The MLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems has the closest RMS error and NEES performance results to the centralized filter and actually has better consistency properties than the restarted local filter.

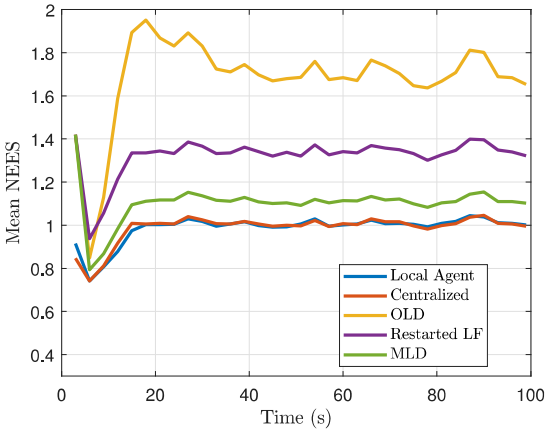
As a result, the MLD seems to be a promising candidate for decorrelation for an IMM filter for low or moderate process noise levels and/or short communication periods. Although, in this chapter, we concentrated only on IMM filters, covariance inflation can also be of value while applying decorrelation for other multiple model filters.



(a) RMS Position Errors

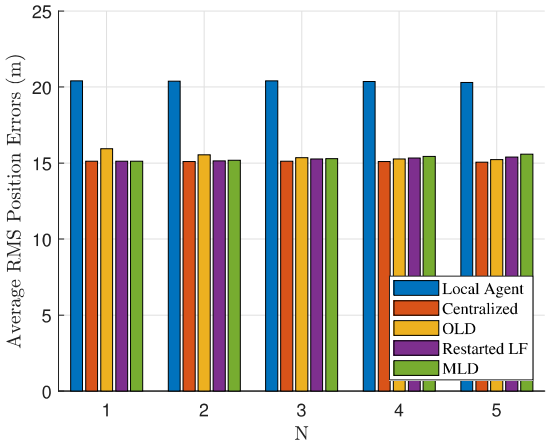


(b) RMS Velocity Errors

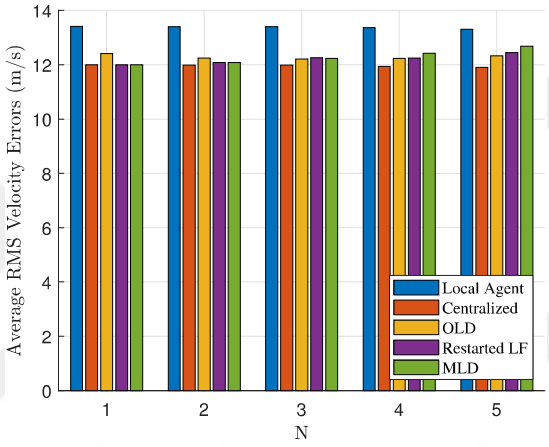


(c) Mean NEES

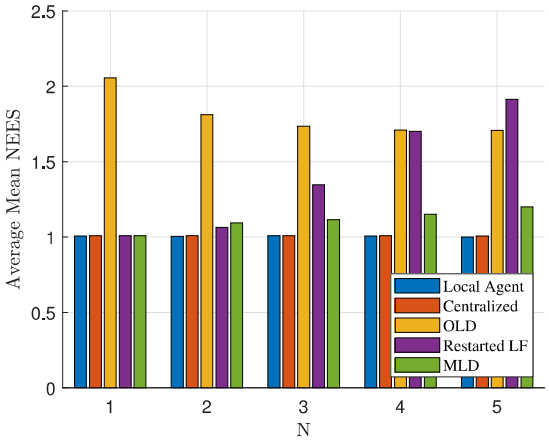
Figure 3.10: IMM Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors



(b) Average RMS Velocity Errors



(c) Average Mean NEES

Figure 3.11: IMM Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$.

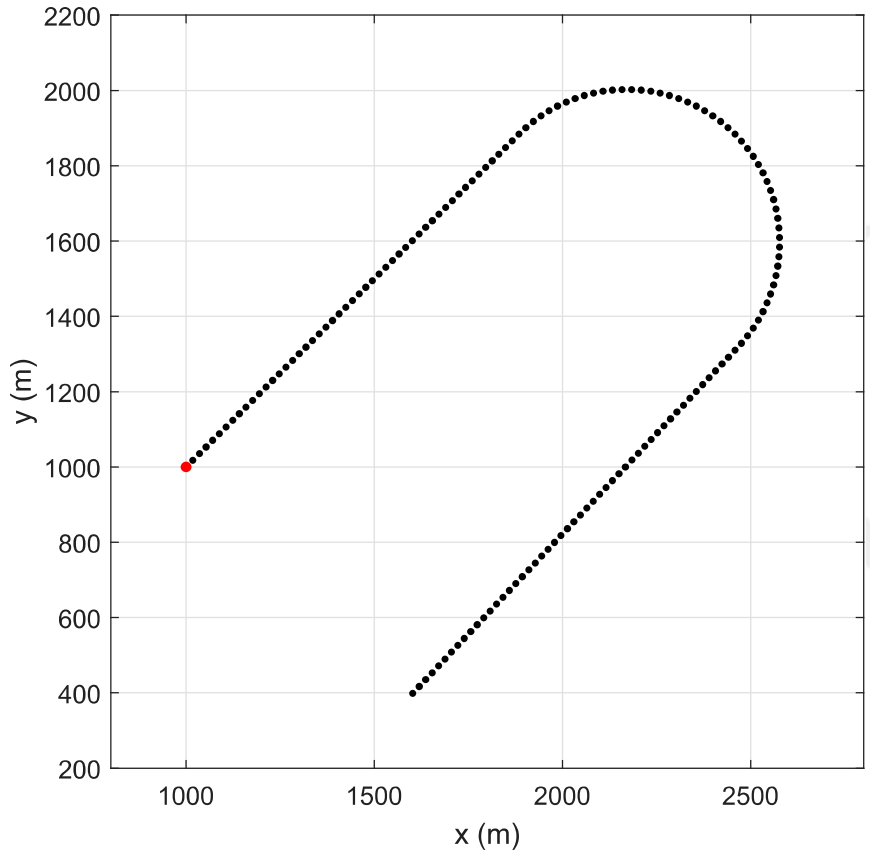
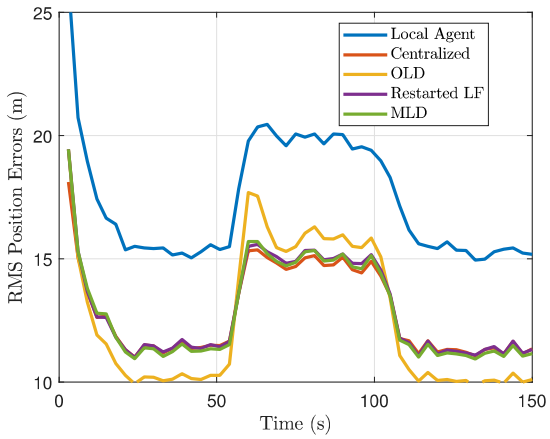
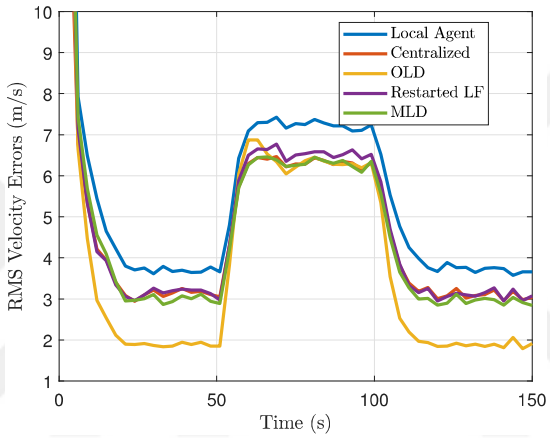


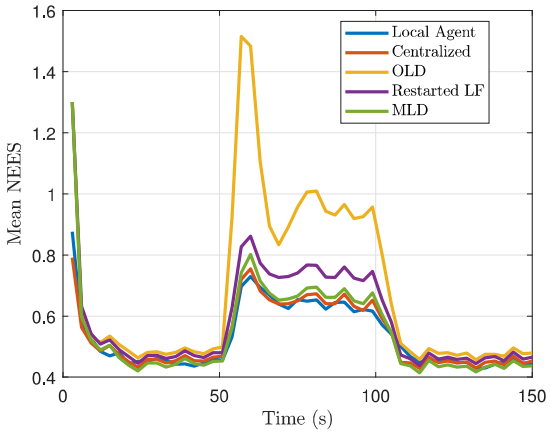
Figure 3.12: True trajectory used for the model-mismatch case. The red dot denotes the start of the target trajectory.



(a) RMS Position Errors

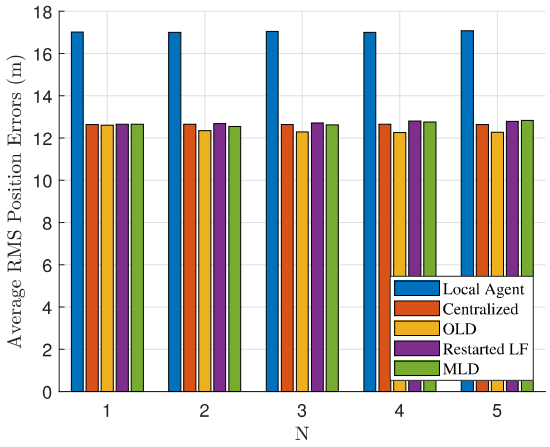


(b) RMS Velocity Errors

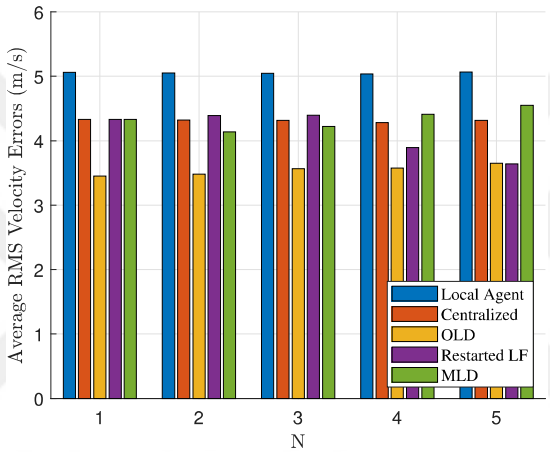


(c) Mean NEES

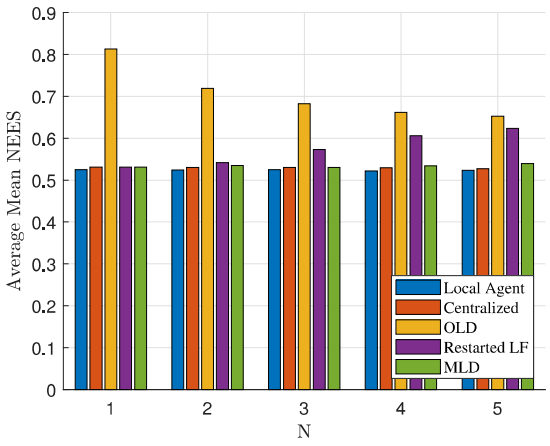
Figure 3.13: IMM Filter - Model Mismatch Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors



(b) Average RMS Velocity Errors



(c) Average Mean NEES

Figure 3.14: IMM Filter - Model Mismatch Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$.

CHAPTER 4

DECORRELATION FOR PDA FILTER

In this chapter we investigate the problem of information decorrelation for compensating for the correlations caused by previous communication between two local agents in a distributed estimation framework where PDA filters are used in the local and fusion nodes. In Section 4.1 it is given how the information decorrelation task is related to the PDA filter. We give a problem formulation for the task of information decorrelation in fusion architectures using PDA filter as a local filter in Section 4.2. The information decorrelation approach for the PDA filter is introduced in Section 4.3. Section 4.4 investigates and proposes solutions for the implementation issues that might be encountered while applying the chosen decorrelation approach for the PDA filter. Section 4.5 describes the fusion operations to be applied in the fusion node. Brief remarks on the computational load of the decorrelation method are made in Appendix B Section B.2. In Section 4.6 we present the numerical results comparing the investigated decorrelation approach with the alternative algorithms on simple distributed maneuvering target tracking examples. The conclusions are drawn in Section 4.7.

4.1 Introduction

PDA filter uses all of the measurements in the validation region and approximates the posterior state density as a Gaussian mixture. In addition to the hypothesis that none of the measurements falling within the validation region will come from the real target, the hypotheses that each measurement within the validation region may belong to the real target are evaluated together with their hypothesis probabilities.

The state estimate and covariance of each hypothesis are calculated. These statistics are merged based on hypothesis weights. Information decorrelation can be done over merged statistics, as well as hypothesis-conditioned statistics. As in the IMM filter, the information decorrelation task for a PDA filter is directly related to the division of Gaussian mixtures. In this chapter we avoid the problem of ratio of mixtures by applying the decorrelation on the joint posterior (of the base-state and hypothesis-state) provided by the PDA filter.

In this chapter, we first give the information decorrelation approach for the PDA filter. We investigate and propose solutions for the implementation issues that might be encountered while applying this decorrelation approach, which were proposed in [18]. Furthermore, we provide a brief overview of the computation load of the investigated decorrelation approach and give simulation results.

In this chapter, we assume a fusion architecture where correlations are dominated by previous communication between the nodes (rather than the process noise), which necessitates the decorrelation/subtraction of previously communicated information before fusion.

4.2 Problem Definition

We consider two local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ in a sensor network equipped with their own sensors denoted as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ run local Bayesian filters for the following linear Gaussian system.

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_k, \tag{4.1a}$$

$$\mathbf{y}_k^s = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{v}_k^s, \tag{4.1b}$$

where the superscript $s \in \{1, 2\}$ denotes the local agent index; $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denotes the base-state; $\mathbf{y}_k^s \in \mathbb{R}^{n_y}$ denotes the local measurements of $\mathcal{F}_s$. $\mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}_{r_k})$ denotes the i.i.d. process noise. $\mathbf{v}_k^s \sim \mathcal{N}(\mathbf{v}_k^s; \mathbf{0}, \mathbf{R}_{r_k}^s)$ denotes the i.i.d. measurement noise for $\mathcal{F}_s$. The measurement noise sequences $\{\mathbf{v}_k^1\}_{k=0}^{\infty}$ and $\{\mathbf{v}_k^2\}_{k=0}^{\infty}$ belonging to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, are assumed to be independent. The initial posterior distri-

bution of $\{\mathbf{x}_0\}$ is given as

$$p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0, \mathbf{P}_0). \tag{4.2}$$

Measurements in the gate are shown as $\mathbf{Y}_k^s = \{\mathbf{y}_k^{i,s}\}_{i=1}^{m_k}$. There are the following hypotheses about these measurements,

- $\theta_k = 0$: All of $\mathbf{Y}_k^s$ is false alarm i.e. no target originated measurement in the gate,
- $\theta_k = i$: Measurement $\mathbf{y}_k^{i,s}$ belongs to target, all the rest are false alarms,

for $i = 1, ..., m_k$. The problem is to estimate the posterior distribution $p(\mathbf{x}_k|\mathbf{Y}_{0:k})$.

The PDA filter [54] calculates the association probabilities to the target being tracked for each validated measurement at the current time and uses all of the measurements in the gate to the extent that they suit the prediction.

We assume that the local Bayesian filter of the local agent $\mathcal{F}_s$ provides the joint posterior distribution $p^s(\mathbf{x}_k, \theta_k|\mathbf{Y}_{0:k}^s)$ of $\{\mathbf{x}_k, \theta_k\}$ given $\mathbf{Y}_{0:k}^s \triangleq \{\mathbf{Y}_0^s, \mathbf{Y}_1^s, \ldots, \mathbf{Y}_k^s\}$. The simple fusion architecture we consider and the corresponding information graph are shown in Figure 1.3a and Figure 1.3b, respectively. The local agent $\mathcal{F}_1$ communicates its recent information to $\mathcal{F}_2$ with a constant period $N = 3$. $\mathcal{F}_1$ transmits only the information related to the marginal posterior distribution $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1)$ of the base-state $\mathbf{x}_k$ to $\mathcal{F}_2$. The estimates of the local agents are correlated due to previous communication since $\mathcal{F}_2$ fuses local information which is transmitted from $\mathcal{F}_1$ with the statistics of its own Bayesian filter. It is also known that the state estimation errors of the local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ are correlated due to the common process noise in (4.1). Although the effects of the common process noise are negligible, the correlations due to previous communication has to be compensated for the fused quantities to be consistent. For satisfactory performance of the state estimator, the decorrelation has to be carried out at the transmitting local agent $\mathcal{F}_1$ before transmission to $\mathcal{F}_2$, which is the fusion center.

Suppose now that we are at an intermediate stage of estimation at time k and $\mathcal{F}_1$ transmitted its last information at time $k - N$. Let the marginal posterior distribution $p^1(\mathbf{x}_{k-N}|\mathbf{Y}_{0:k-N}^1)$ represent the transmitted information from the local node $\mathcal{F}_1$

to the fusion node $\mathcal{F}_2$ up to (and including) time $k-N$. Since this information has already been used and fused by the fusion center $\mathcal{F}_2$, when the local node $\mathcal{F}_1$ wants to transmit its updated information at time k represented by the current marginal posterior $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$, it first has to decorrelate the previously communicated information $p^1(\mathbf{x}_{k-N}|\mathbf{Y}^1_{0:k-N})$ from the marginal posterior information $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ and then transmit the decorrelated information. We here make the following assumption about the linear Gaussian system in (4.1).

Assumption 2 *The state transition matrix* $\mathbf{A}$ *of* (4.1) *is invertible.*

When we neglect the effects of the process noise and make Assumption 2, it is shown in Appendix A that the decorrelation operation amounts to the division of the updated marginal distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ to the predicted distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$ i.e., the information $\tilde{p}^1(\mathbf{x}_k)$ that has to be sent to $\mathcal{F}_2$ is given as

$$\tilde{p}^1(\mathbf{x}_k) \propto \frac{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})}. \tag{4.3}$$

For the function $\tilde{p}^1_k(\mathbf{x}_k)$ be interpreted as a valid probability density function, it has to be normalizable with respect to $\mathbf{x}_k$. When $\tilde{p}^1_k(\mathbf{x}_k)$ is normalizable, the normalized function is interpreted as the *decorrelated distribution*, and its statistics the *decorrelated statistics*, e.g., decorrelated mean and covariance. The linear Gaussian system in (4.1) includes process noise. Therefore, the decorrelation operation in (4.3) will only be an approximation of the case considered in Appendix A.

4.3 Decorrelation Approach

In this section we describe the information decorrelation methodology for the PDA filter of the local agent $\mathcal{F}_1$. We assume that we are at an intermediate stage of estimation at time k. We assume that the last decorrelated information transmission from $\mathcal{F}_1$ to $\mathcal{F}_2$ happened at time $k-N$. For the PDA filter, two decorrelation methodologies were proposed:

- Output level decorrelation (OLD),
- Hypothesis level decorrelation (HLD).

In OLD, the merged output estimates and covariances of the local PDA filter (i.e., $\{\hat{\mathbf{x}}_k^1, \mathbf{P}_k^1\}$) are decorrelated and transmitted. In HLD, unmerged hypothesis-conditioned estimates and covariances of the local PDA filter, i.e., $\{\mu_k^{1,i}, \hat{\mathbf{x}}_k^{1,i}, \mathbf{P}_k^{1,i}\}_{i=0}^{m_k}$, are decorrelated, merged and transmitted. In the following parts of this chapter, we are going to use the parts of the PDA filter as building blocks for the information decorrelation approach we describe. For this purpose, we define the following functions.

- $\left[\bar{\mathbf{x}}, \overline{\mathbf{P}}\right]$**=PDAPrediction**$\left(\hat{\mathbf{x}}, \mathbf{P}\right)$**:** This is the functional representation of the PDA filter prediction update in Algorithm 2 which takes the state estimate and covariance $\hat{\mathbf{x}}_{k-1}^s, \mathbf{P}_{k-1}^s$ and returns the predicted state estimate and covariance $\hat{\mathbf{x}}_{k|k-1}^s, \mathbf{P}_{k|k-1}^s$. This prediction operation is basically same with the Kalman filter prediction update operation and equal to each other for all hypotheses.

$$\hat{\mathbf{x}}_{k|k-1}^s = \mathbf{A}\hat{\mathbf{x}}_{k-1}^s, \tag{4.4a}$$

$$\mathbf{P}_{k|k-1}^s = \mathbf{A}\mathbf{P}_{k-1}^s\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{Q}\mathbf{B}^{\mathrm{T}}. \tag{4.4b}$$

- $\left[\hat{\mathbf{x}}, \mathbf{P}\right]$**=Merging**$\left(\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=0}^{m_k}\right)$**:** This is the functional representation of the output estimate calculation part in Algorithm 2 which takes the hypothesis probabilities, estimates and covariances $\{\mu_k^{s,i}, \hat{\mathbf{x}}_k^{s,i}, \mathbf{P}_k^{s,i}\}_{i=0}^{m_k}$ and returns the merged estimate $\hat{\mathbf{x}}_k^s$ and covariance $\mathbf{P}_k^s$.

4.3.1 Output Level Decorrelation (OLD)

The OLD decorrelates and transmits the output estimate and the covariance of the local PDA filter (i.e., $\hat{\mathbf{x}}_k^1$ and $\mathbf{P}_k^1$). The output estimate and its covariance at the last transmission time $k-N$ are $\hat{\mathbf{x}}_{k-N}^1$ and $\mathbf{P}_{k-N}^1$, respectively. The transmitted information up to (and including) time $k-N$ is hence represented by the Gaussian distribution $\mathcal{N}(\mathbf{x}_{k-N}; \hat{\mathbf{x}}_{k-N}^1, \mathbf{P}_{k-N}^1)$. The decorrelation operation requires the predicted version of this distribution to time k. Since linear Gaussian system contains a single target model like the case of Kalman filter, we need to use a Kalman filter for this prediction operation. For this purpose, we initialize a single model PDA predictor with the state estimate and covariance at the last transmission time $k-N$, i.e., $\bar{\mathbf{x}}_{k-N}^1 = \hat{\mathbf{x}}_{k-N}^1$ and $\overline{\mathbf{P}}_{k-N}^1 = \mathbf{P}_{k-N}^1$, where we use overlined quantities to separate them from the respective quantities of the primary PDA filter at

the local node. We then use the PDA prediction operation N-times to obtain state estimate $\bar{\mathbf{x}}^1_{k|k-N}$ and covariance $\overline{\mathbf{P}}^1_{k|k-N}$. Note that the predicted information obtained in this manner will also be represented by a single Gaussian density, i.e., $\bar{p}^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N}) = \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})$. Hence the new decorrelated information will be represented by a Gaussian $\tilde{p}^1(\mathbf{x}_k|\mathbf{Y}_{0:k}) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)$, whose parameters can be calculated using the information filter expressions given in (3.7).

In OLD, the decorrelated estimate $\tilde{\mathbf{x}}^1_k$ and covariance $\widetilde{\mathbf{P}}^1_k$ are transmitted to the second node. The weight $\tilde{\mu}^1_k$, which is due to the division operation, is ignored, where $\tilde{\mu}^1_k \propto \frac{|\widetilde{\mathbf{P}}^1_k|}{|\overline{\mathbf{P}}^1_k|} \times \left(\mathcal{N}(\bar{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}^1_k, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)\right)^{-1}$. This corresponds to the transmission of the density $\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)$ to the fusion node. The OLD is summarized in Algorithm 8. Section 4.5 gives the fusion formula to be applied into the receiving node. The general decorrelation and transmission operations made in the first node in OLD are illustrated in Figure 4.1.

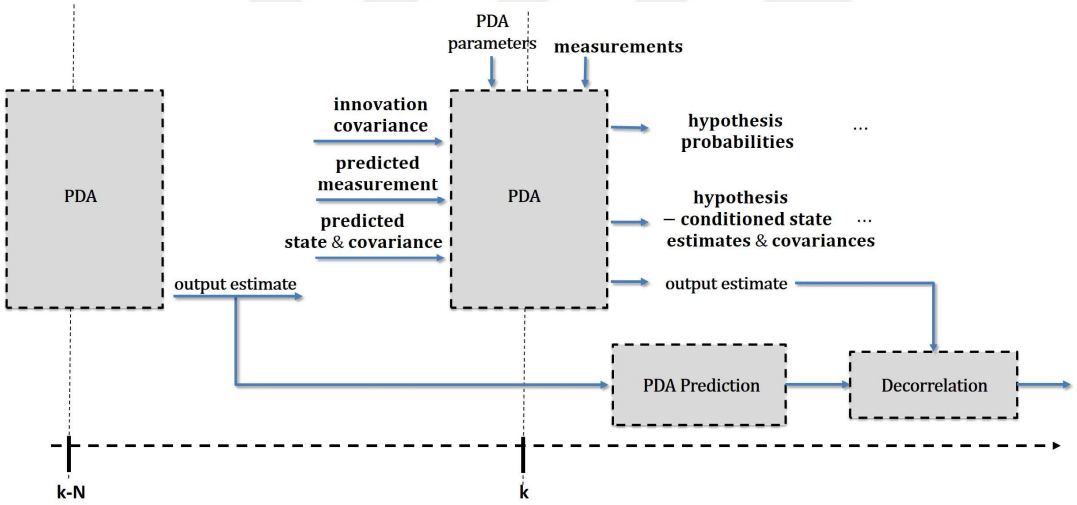


Figure 4.1: OLD decorrelation & transmission operations for the PDA filter.

Algorithm 8 (OLD)

- *Given:* $\hat{\mathbf{x}}^1_{k-N}, \mathbf{P}^1_{k-N}$.

- *Initialization:*

$$\bar{\mathbf{x}}^1_{k-N} = \hat{\mathbf{x}}^1_{k-N}, \tag{4.5a}$$

$$\overline{\mathbf{P}}^1_{k-N} = \mathbf{P}^1_{k-N}, \tag{4.5b}$$

- *PDA predictor:*

for $j = k-N+1, \ldots, k-N$

$$\left[\bar{\mathbf{x}}^1_{j|k-N}, \overline{\mathbf{P}}^1_{j|k-N}\right] = \textit{PDAPrediction}\left(\bar{\mathbf{x}}^1_{j-1|k-N}, \overline{\mathbf{P}}^1_{j-1|k-N}\right).$$

end

- *Decorrelation:*

$$\tilde{\mathbf{x}}^1_k = \widetilde{\mathbf{P}}^1_k \left(\left(\mathbf{P}^1_k\right)^{-1} \hat{\mathbf{x}}^1_k - \left(\overline{\mathbf{P}}^1_{k|k-N}\right)^{-1} \bar{\mathbf{x}}^1_{k|k-N} \right), \tag{4.6a}$$

$$\widetilde{\mathbf{P}}^1_k = \left(\left(\mathbf{P}^1_k\right)^{-1} - \left(\overline{\mathbf{P}}^1_{k|k-N}\right)^{-1} \right)^{-1}. \tag{4.6b}$$

- *Transmission:* $\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k$.

The decorrelation operation in OLD, like the case in the IMM filter, does not always result in positive semi-definite decorrelated covariance which was examined at the end of Section 4.2. In such a case the transmission of the decorrelated information to the fusion center can be done after applying the proposed methods in Section 4.4.

4.3.2 Hypothesis Level Decorrelation (HLD)

The main idea of HLD is to apply decorrelation to each hypothesis of the PDA filter. The transmitted information is represented by a Gaussian distribution, i.e., $p^1(\mathbf{x}_{k-N}|\mathbf{Y}^1_{0:k-N}) = \mathcal{N}(\mathbf{x}_{k-N}; \hat{\mathbf{x}}^1_{k-N}, \mathbf{P}^1_{k-N})$ which is about the base-state up to (and including) time $k-N$. Using this information, we can initialize a PDA predictor by setting $\bar{\mathbf{x}}^1_{k-N} = \hat{\mathbf{x}}^1_{k-N}$ and $\overline{\mathbf{P}}^1_{k-N} = \mathbf{P}^1_{k-N}$, where we use overlined quantities to separate them from the respective quantities of the primary PDA filter at the local agent $\mathcal{F}_1$. If we predict this information to time k using an PDA predictor, we can obtain the state estimate $\bar{\mathbf{x}}^1_{k|k-N}$ and covariance $\overline{\mathbf{P}}^1_{k|k-N}$ of the predicted distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$. Since the current marginal posterior is a Gaussian mixture, i.e., $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k}) = \sum_{i=0}^{m_k} \mu^{1,i}_k \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{1,i}_k, \mathbf{P}^{1,i}_k)$, the decorrelation in (4.3) involves the division of a Gaussian mixture and a Gaussian density, $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ and $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$.

In HLD, the decorrelation is applied over the densities $p^1(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:k})$ and $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$ provided by the PDA filter.

$$\frac{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})} = \sum_{\theta_k} \frac{p^1(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})}, \tag{4.7}$$

up to a constant multiplier with respect to $\mathbf{x}_k$. The joint density $p^1(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:k})$ and the predicted density $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$ are the scaled Gaussian distributions provided by the PDA filter and predictor as follows.

$$p^1(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:k}) = \mu_k^{1,\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,\theta_k}, \mathbf{P}_k^{1,\theta_k}), \tag{4.8a}$$

$$p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N}) = \mathcal{N}(\mathbf{x}_k; \overline{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N}). \tag{4.8b}$$

Division of these two distributions gives the following joint decorrelated information.

$$\tilde{p}^1(\mathbf{x}_k, \theta_k) \propto \frac{p^1(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})}, \tag{4.9a}$$

$$= \mu_k^{1,\theta_k} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,\theta_k}, \mathbf{P}_k^{1,\theta_k})}{\mathcal{N}(\mathbf{x}_k; \overline{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})}. \tag{4.9b}$$

Under some assumptions of normalizability with respect to $\mathbf{x}_k$, the division of two Gaussian distributions above can be written as a scaled Gaussian (see Section A for details), which results in the following joint decorrelated distribution.

$$\tilde{p}^1(\mathbf{x}_k, \theta_k) = \tilde{\mu}_k^{1,\theta_k} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,\theta_k}, \widetilde{\mathbf{P}}_k^{1,\theta_k}), \tag{4.10}$$

where

$$\tilde{\mu}_k^{1,\theta_k} \propto \mu_k^{1,\theta_k} \frac{|\widetilde{\mathbf{P}}_k^{1,\theta_k}|}{|\mathbf{P}_k^{1,\theta_k}|} \left(\mathcal{N}(\overline{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}_k^{1,\theta_k}, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}_k^{1,\theta_k})\right)^{-1}, \tag{4.11a}$$

$$\tilde{\mathbf{x}}_k^{1,\theta_k} = \widetilde{\mathbf{P}}_k^{1,\theta_k} \left((\mathbf{P}_k^{1,\theta_k})^{-1} \hat{\mathbf{x}}_k^{1,\theta_k} - (\overline{\mathbf{P}}^1_{k|k-N})^{-1} \overline{\mathbf{x}}^1_{k|k-N}\right), \tag{4.11b}$$

$$\widetilde{\mathbf{P}}_k^{1,\theta_k} = \left((\mathbf{P}_k^{1,\theta_k})^{-1} - (\overline{\mathbf{P}}^1_{k|k-N})^{-1}\right)^{-1}. \tag{4.11c}$$

Marginalizing out the hypothesis variable θ_k from the joint decorrelated distribution $\tilde{p}^1(\mathbf{x}_k, \theta_k)$, we get the approximate marginal decorrelated distribution below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=0}^{m_k} \tilde{\mu}_k^{1,i} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}). \tag{4.12}$$

Finally we apply a merging operation to the decorrelated mixture $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$.

When the transmission period N is equal to 1, the joint decorrelated information related to the zero-hypothesis (i.e, no measurement belonging to the true target in the validation region) can be written as.

$$\tilde{p}^1(\mathbf{x}_k, \theta_k = 0) \propto \frac{p^1(\mathbf{x}_k, \theta_k = 0|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})}, \tag{4.13a}$$

$$= \mu_k^{1,0} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,0}, \mathbf{P}_k^{1,0})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})}, \tag{4.13b}$$

$$= \mu_k^{1,0} \frac{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N})}, \tag{4.13c}$$

$$= \mu_k^{1,0}. \tag{4.13d}$$

Marginalizing out the hypothesis variable θ_k from the joint decorrelated information $\tilde{p}^1(\mathbf{x}_k, \theta_k)$, we get the approximate decorrelated information below.

$$\tilde{p}^1(\mathbf{x}_k) = \tilde{\mu}_k^{1,0} + (1 - \tilde{\mu}_k^{1,0}) \sum_{j=1}^{m_k} \tilde{\mu}_k^{1,j} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,j}, \widetilde{\mathbf{P}}_k^{1,j}), \tag{4.14}$$

where

$$\tilde{\mu}_k^{1,0} = \frac{\mu_k^{1,0}}{\mu_k^{1,0} + \sum_{i=1}^{m_k} \mu_k^{1,i} \frac{|\widetilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|} \left(\mathcal{N}(\bar{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}_k^{1,i}, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}_k^{1,i})\right)^{-1}}, \tag{4.15a}$$

$$\tilde{\mu}_k^{1,i} = \frac{\mu_k^{1,i} \frac{|\widetilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|} \left(\mathcal{N}(\bar{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}_k^{1,i}, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}_k^{1,i})\right)^{-1}}{\sum_{i=1}^{m_k} \mu_k^{1,i} \frac{|\widetilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|} \left(\mathcal{N}(\bar{\mathbf{x}}^1_{k|k-N}; \hat{\mathbf{x}}_k^{1,i}, \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}_k^{1,i})\right)^{-1}}. \tag{4.15b}$$

Finally we apply a merging operation only to the decorrelated mixture part of $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$ with a constant $\tilde{\mu}_k^{1,0}$.

$$\tilde{p}^1(\mathbf{x}_k) = \tilde{\mu}_k^{1,0} + (1 - \tilde{\mu}_k^{1,0}) \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1), \tag{4.16}$$

where $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ are given as

$$\tilde{\mathbf{x}}_k^1 = \sum_{j=1}^{m_k} \tilde{\mu}_k^{1,j} \tilde{\mathbf{x}}_k^{1,j}, \tag{4.17a}$$

$$\widetilde{\mathbf{P}}_k^1 = \sum_{j=1}^{m_k} \tilde{\mu}_k^{1,j} \left[\widetilde{\mathbf{P}}_k^{1,j} + (\tilde{\mathbf{x}}_k^{1,j} - \tilde{\mathbf{x}}_k^1)(\cdot)^{\mathrm{T}}\right]. \tag{4.17b}$$

The HLD is summarized in Algorithm 9. Note that the decorrelation operations in HLD might not always result in positive semi-definite decorrelated covariances unlike

the case in a single Kalman filter, which was examined at the end of Section 4.2. In such a case the decorrelated quantities cannot be merged at the end of HLD. We give details about such issues and propose methods to deal with them in Section 4.4. The decorrelation and transmission operations made in the first node in HLD are illustrated in Figure 4.2. Section 4.5 gives the fusion formula to be applied into the receiving node.

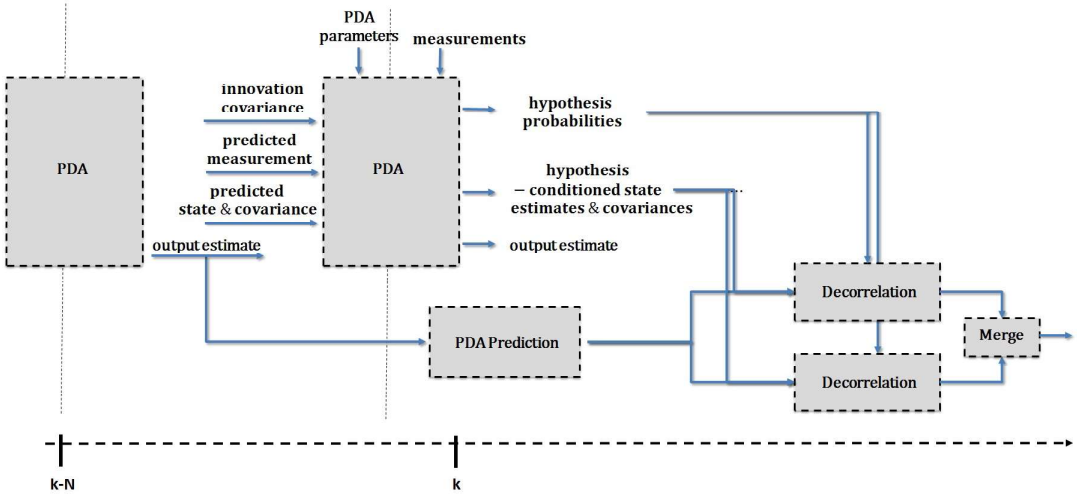


Figure 4.2: HLD Decorrelation & transmission operations for the PDA filter.

Algorithm 9 (HLD)

- *Given:* $\hat{\mathbf{x}}^1_{k-N}, \mathbf{P}^1_{k-N}$.
- *Initialization:*

$$\bar{\mathbf{x}}^1_{k-N} = \hat{\mathbf{x}}^1_{k-N}, \tag{4.18a}$$

$$\overline{\mathbf{P}}^1_{k-N} = \mathbf{P}^1_{k-N}. \tag{4.18b}$$

- *PDA predictor:*

  For $j = k-N+1, \ldots, k$,

  $\left[\bar{\mathbf{x}}^1_{j|k-N}, \overline{\mathbf{P}}^1_{j|k-N}\right]$ = *PDAPrediction*$\left(\bar{\mathbf{x}}^1_{j-1|k-N}, \overline{\mathbf{P}}^1_{j-1|k-N}\right)$.

  End

- *Decorrelation:*

If $N == 1$

For $i = 1, \ldots, m_k$,

$$\tilde{\mathbf{x}}_k^{1,i} = \widetilde{\mathbf{P}}_k^{1,i}\big((\mathbf{P}_k^{1,i})^{-1}\hat{\mathbf{x}}_k^{1,i} - (\overline{\mathbf{P}}_{k|k-N}^{1})^{-1}\bar{\mathbf{x}}_{k|k-N}^{1}\big),$$

$$\widetilde{\mathbf{P}}_k^{1,i} = \big((\mathbf{P}_k^{1,i})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1})^{-1}\big)^{-1}.$$

$$\tilde{\mu}_k^{1,i} = \frac{\mu_k^{1,i}\frac{|\tilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|}\Big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1};\hat{\mathbf{x}}_k^{1,i},\overline{\mathbf{P}}_{k|k-N}^{1}-\mathbf{P}_k^{1,i})\Big)^{-1}}{\sum_{i=1}^{m_k}\mu_k^{1,i}\frac{|\tilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|}\Big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1};\hat{\mathbf{x}}_k^{1,i},\overline{\mathbf{P}}_{k|k-N}^{1}-\mathbf{P}_k^{1,i})\Big)^{-1}}.$$

End

$$\tilde{\mu}_k^{1,0} = \frac{\mu_k^{1,0}}{\mu_k^{1,0}+\sum_{i=1}^{m_k}\mu_k^{1,i}\frac{|\tilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|}\times\Big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1};\hat{\mathbf{x}}_k^{1,i},\overline{\mathbf{P}}_{k|k-N}^{1}-\mathbf{P}_k^{1,i})\Big)^{-1}}.$$

else

For $i = 0, \ldots, m_k$,

$$\tilde{\mu}_k^{1,i} = \frac{\mu_k^{1,i}\frac{|\tilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|}\Big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1};\hat{\mathbf{x}}_k^{1,i},\overline{\mathbf{P}}_{k|k-N}^{1}-\mathbf{P}_k^{1,i})\Big)^{-1}}{\sum_{i=0}^{m_k}\mu_k^{1,i}\frac{|\tilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|}\Big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1};\hat{\mathbf{x}}_k^{1,i},\overline{\mathbf{P}}_{k|k-N}^{1}-\mathbf{P}_k^{1,i})\Big)^{-1}},$$

$$\tilde{\mathbf{x}}_k^{1,i} = \widetilde{\mathbf{P}}_k^{1,i}\big((\mathbf{P}_k^{1,i})^{-1}\hat{\mathbf{x}}_k^{1,i} - (\overline{\mathbf{P}}_{k|k-N}^{1})^{-1}\bar{\mathbf{x}}_{k|k-N}^{1}\big),$$

$$\widetilde{\mathbf{P}}_k^{1,i} = \big((\mathbf{P}_k^{1,i})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1})^{-1}\big)^{-1}.$$

End

End

- *Merging*

If $N == 1$

$$\big[\tilde{\mathbf{x}}_k^{1}, \widetilde{\mathbf{P}}_k^{1}\big] = \textit{Merging}\big(\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=1}^{m_k}\big).$$

else

$$\big[\tilde{\mathbf{x}}_k^{1}, \widetilde{\mathbf{P}}_k^{1}\big] = \textit{Merging}\big(\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=0}^{m_k}\big).$$

end

- *Transmission:*

If $N == 1$

$$\tilde{\mu}_k^{1,0}, \tilde{\mathbf{x}}_k^{1}, \widetilde{\mathbf{P}}_k^{1}.$$

else

$$\tilde{\mathbf{x}}_k^{1}, \widetilde{\mathbf{P}}_k^{1}.$$

end

4.4 Implementation Issues

Both of the proposed decorrelation approaches for the PDA filter involve the information subtraction operations given in (3.18), the quantities $\{\hat{\mathbf{x}}, \mathbf{P}\}$, $\{\bar{\mathbf{x}}, \overline{\mathbf{P}}\}$, $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ represent the mean and covariance (of a component) of the current distribution, predicted distribution and decorrelated distribution, respectively. As in the IMM filter case, we have encountered both incompatible covariance problem and outlier decorrelation problem related to the information matrix difference in (3.18b) while applying decorrelation process for the PDA filter.

4.4.1 Use of covariance inflation in OLD & HLD for the PDA filter

In the implementation of OLD and HLD, at each decorrelation step we check the existence of the incompatible covariance and outlier decorrelation problems. When one or more of these problems are detected, a compensation in the form of covariance inflation is applied. The related steps to be added into the OLD and HLD are given in Algorithms 10 and 11, respectively.

In both HLD and OLD we first check for the existence of the incompatible covariance problem. If there is one, we calculate the lower bound κ_{IC} for the inflation constant κ in OLD and inflate the predicted covariance with the constant $c_{\mathrm{IC}}\kappa_{\mathrm{IC}}$ where $c_{\mathrm{IC}} > 1$ is a safety factor. In HLD we first check for the existence of the incompatible covariance problem for each hypothesis ($i = 0, \ldots, m_k$). If we find such a problem in a hypothesis, we calculate the lower bound κ^i_{IC}, $i = 0, \ldots, m_k$ for the inflation constant κ for that hypothesis. If there exists incompatible covariance problem in at least one hypotheses, we inflate the predicted covariance for all hypotheses with the maximum of the values $c_{\mathrm{IC}}\kappa^i_{\mathrm{IC}}$, $i = 0, \ldots, m_k$, where $c_{\mathrm{IC}} > 1$ is a safety factor. In HLD, in the case of incompatible covariance problem in a hypothesis, the predicted covariance of all hypotheses should be inflated with the same inflation constant in order not to create artificial irregularities in the decorrelated weights $\tilde{\mu}^{1,i}_k$, $i = 0, \ldots, m_k$. Inflating only the predicted covariance of the hypotheses in which there is an incompatible covariance problem would cause unintended irregularities in the decorrelated weights $\tilde{\mu}^{1,i}_k$, $i = 0, \ldots, m_k$ which are highly dependent on the predicted covariance.

After handling the incompatible covariance problems, we then check for the existence of the outlier decorrelation problems in both OLD and HLD. If there is one, we calculate the lower bound κ_{OD} for the inflation constant κ using the Newton method with a fixed number of iterations M_{OD} in OLD and inflate the predicted covariance with the constant $c_{\mathrm{OD}}\kappa_{\mathrm{OD}}$ where $c_{\mathrm{OD}} > 1$ is a safety factor. In HLD we check for the existence of the outlier decorrelation problem for each hypothesis ($i = 0, \ldots, m_k$). If we find such a problem in a hypothesis, we calculate the lower bound κ^i_{OD}, $i = 0, \ldots, m_k$ for the inflation constant κ for that hypothesis similarly to what we do in OLD. In HLD if there exists outlier decorrelation problem in at least one hypothesis, we inflate the predicted covariance of all hypotheses with the maximum of the values $c_{\mathrm{OD}}\kappa^i_{\mathrm{OD}}$, $i = 0, \ldots, m_k$ to avoid artificial irregularities in the decorrelated weights $\tilde{\mu}^i_k$, $i = 0, \ldots, m_k$.

Algorithm 10 (OLD Covariance Inflation) *To be inserted before the decorrelation step of OLD in Algorithm 8*

- *Given:* $\hat{\mathbf{x}}^1_k$, $\mathbf{P}^1_k$, $\bar{\mathbf{x}}^1_{k|k-N}$, $\overline{\mathbf{P}}^1_{k|k-N}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Check and compensate for the incompatible covariance problem*

  $\kappa_{\mathrm{IC}} = \lambda_{\max}\big(\mathbf{P}^1_k\big(\overline{\mathbf{P}}^1_{k|k-N}\big)^{-1}\big)$.
  If $\kappa_{\mathrm{IC}} \geq 1$,
    Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{IC}}\kappa_{\mathrm{IC}}\overline{\mathbf{P}}^{\mathbf{1}}{}_{k|k-N}$.
  End

- *Check and compensate for the outlier decorrelation problem*

  $g_{\mathrm{OD}} = (\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}(\overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)^{-1}(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})$.
  If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,
    $\kappa^{(0)} = 1$.
    For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,
      $\mathbf{\Lambda}_m = (\kappa^{(m)}\overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)^{-1}$.
      $\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})}{(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m\overline{\mathbf{P}}^1_{k|k-N}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})}$.
    End
    $\kappa_{\mathrm{OD}} = \kappa^{(M_{\mathrm{OD}})}$.

Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{OD}}\kappa_{\mathrm{OD}}\overline{\mathbf{P}}^1_{k|k-N}$.

End

Algorithm 11 (HLD Covariance Inflation) *To be inserted before the decorrelation step of HLD in Algorithm 9.*

- *Given:* $\{\hat{\mathbf{x}}^{1,i}_k, \mathbf{P}^{1,i}_k\}_{i=0}^{m_k}$, $\bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Find the initial hypothesis number for algorithm to be applied:*

  If $N == 1$

  $l = 1$.

  else

  $l = 0$.

  End

- *Check and compensate for the incompatible covariance problem:*

  For $i = l, \ldots, m_k$,

  $\kappa^i_{\mathrm{IC}} = \lambda_{\max}(\mathbf{P}^{1,i}_k(\overline{\mathbf{P}}^1_{k|k-N})^{-1})$.

  End

  $\kappa_{\mathrm{IC}} = \max\{\kappa^l_{\mathrm{IC}}, \ldots, \kappa^{m_k}_{\mathrm{IC}}\}$.

  If $\kappa_{\mathrm{IC}} \geq 1$,

  Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{IC}}\kappa_{\mathrm{IC}}\overline{\mathbf{P}}^1_{k|k-N}$.

  End

- *Check and compensate for the outlier decorrelation problem:*

  For $i = l, \ldots, m_k$,

  $g_{\mathrm{OD}} = (\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}(\overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1}(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})$.

  If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,

  $\kappa^{(0)} = 1$.

  For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

  $\mathbf{\Lambda}_m = (\kappa^{(m)}\overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1}$.

  $\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})}{(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m\overline{\mathbf{P}}^1_{k|k-N}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^{1,i}_k - \bar{\mathbf{x}}^1_{k|k-N})}$.

End

$\kappa_{\mathrm{OD}}^{i} = \kappa^{(M_{\mathrm{OD}})}$.

Else

$\kappa_{\mathrm{OD}}^{i} = 0$.

End

End

$\kappa_{\mathrm{OD}} = \max\{\kappa_{\mathrm{OD}}^{l}, \ldots, \kappa_{\mathrm{OD}}^{N_r}\}$.

If $\kappa_{\mathrm{OD}} \geq 1$,

Set $\overline{\mathbf{P}}_{k|k-N}^{1} \leftarrow c_{\mathrm{OD}}\kappa_{\mathrm{OD}}\overline{\mathbf{P}}_{k|k-N}^{1}$.

End

4.5 Fusion

In this section we describe the fusion operation to be applied by the local agent $\mathcal{F}_2$ upon receiving the decorrelated information from $\mathcal{F}_1$. After receiving the decorrelated distribution $\tilde{p}^1(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1)$, i.e., its mean and covariance $\{\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1\}$, from $\mathcal{F}_1$ at time k, $\mathcal{F}_2$ has to incorporate this information into the summary statistics of its own PDA filter.

The current joint posterior distribution of the PDA filter of $\mathcal{F}_2$ before fusion is given as

$$p^2(\mathbf{x}_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2) = \mu_{k|k-N,k}^{2,\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,\theta_k}, \mathbf{P}_{k|k-N,k}^{2,\theta_k}), \tag{4.19}$$

where $\mathcal{I}_{0:k-N}^1$ denotes the information that $\mathcal{F}_2$ has got from $\mathcal{F}_1$ so far (see Appendix A for the reasons for the use of this notation). Note that the summary statistics of the PDA filter of $\mathcal{F}_2$ before fusion are given as $\{\mu_{k|k-N,k}^{2,i}, \hat{\mathbf{x}}_{k|k-N,k}^{2,i}, \mathbf{P}_{k|k-N,k}^{2,i}\}_{i=0}^{m_k}$. $\mathcal{F}_2$ applies naive fusion to fuse the pieces of information described by the distributions $\tilde{p}^1(\mathbf{x}_k)$ and $p^2(\mathbf{x}_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2)$.

$$p^2(\mathbf{x}_k, \theta_k | \mathcal{I}_{0:k}^1, \mathbf{Y}_{0:k}^2) \propto \tilde{p}^1(\mathbf{x}_k) p^2(\mathbf{x}_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2), \tag{4.20a}$$

$$= \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1) \mu_{k|k-N,k}^{2,\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,\theta_k}, \mathbf{P}_{k|k-N,k}^{2,\theta_k}), \tag{4.20b}$$

$$= \mu_{k|k,k}^{2,\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k,k}^{2,\theta_k}, \mathbf{P}_{k|k,k}^{2,\theta_k}), \tag{4.20c}$$

where

$$\mu^{2,\theta_k}_{k|k,k} \propto \mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\tilde{\mathbf{x}}^1_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k} + \widetilde{\mathbf{P}}^1_k), \tag{4.21a}$$

$$\hat{\mathbf{x}}^{2,\theta_k}_{k|k,k} \triangleq \mathbf{P}^{2,\theta_k}_{k|k,k}\big((\mathbf{P}^{2,\theta_k}_{k|k-N,k})^{-1}\hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k} + (\widetilde{\mathbf{P}}^1_k)^{-1}\tilde{\mathbf{x}}_k\big), \tag{4.21b}$$

$$\mathbf{P}^{2,\theta_k}_{k|k,k} \triangleq \big((\mathbf{P}^{2,\theta_k}_{k|k-N,k})^{-1} + (\widetilde{\mathbf{P}}^1_k)^{-1}\big)^{-1}. \tag{4.21c}$$

The summary statistics of the PDA filter of $\mathcal{F}_2$ after fusion are then given as $\{\mu^{2,i}_{k|k,k}, \hat{\mathbf{x}}^{2,i}_{k|k,k}, \mathbf{P}^{2,i}_{k|k,k}\}^{m_k}_{i=0}$. The fusion operations made in the second node are illustrated in Figure (4.3).

As given in Algorithm 9, as a result of decorrelation when $N = 1$, the transmission

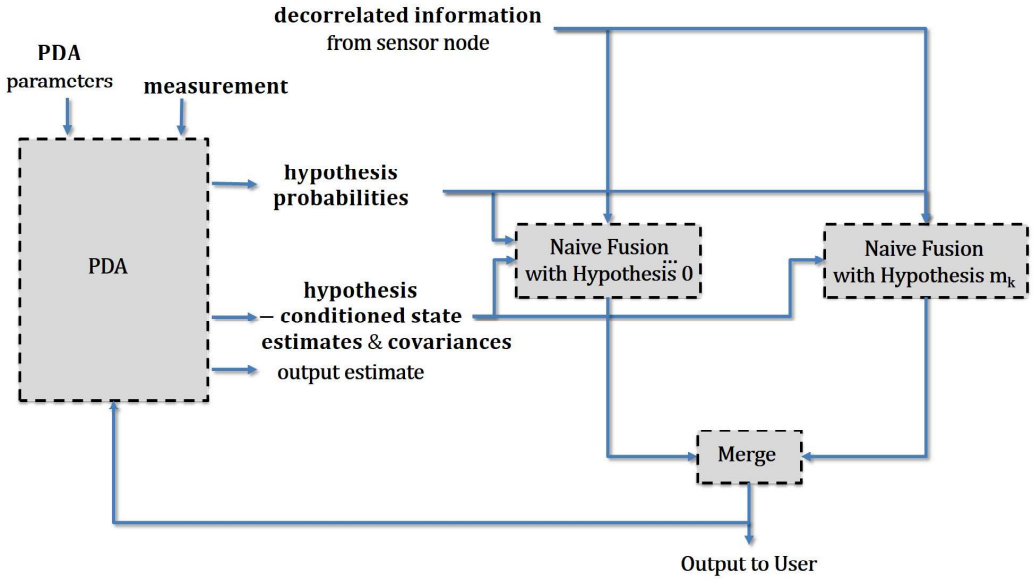


Figure 4.3: Fusion operations for the PDA filter.

of the weight $\tilde{\mu}^{1,0}_k$, mean $\tilde{\mathbf{x}}^1_k$ and covariance $\widetilde{\mathbf{P}}^1_k$ from $\mathcal{F}_1$ at time k to the fusion center is required.

$$\tilde{p}^1(\mathbf{x}_k) = \tilde{\mu}^{1,0}_k + (1 - \tilde{\mu}^{1,0}_k)\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k). \tag{4.22}$$

$\mathcal{F}_2$ applies naive fusion to fuse the pieces of information described by the distributions

$\tilde{p}^1(\mathbf{x}_k)$ and $p^2(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k-N}, \mathbf{Y}^2_{0:k})$.

$$p^2(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k}) \propto \tilde{p}^1(\mathbf{x}_k) p^2(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k-N}, \mathbf{Y}^2_{0:k}), \tag{4.23a}$$

$$= \big(\tilde{\mu}^{1,0}_k + (1-\tilde{\mu}^{1,0}_k)\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)\big) \times \mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k}), \tag{4.23b}$$

$$= \tilde{\mu}^{1,0}_k \mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k}) + (1-\tilde{\mu}^{1,0}_k)\mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k)\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k}), \tag{4.23c}$$

$$= \tilde{\mu}^{1,0}_k \mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k}) + (1-\tilde{\mu}^{1,0}_k)\mu^{2,\theta_k}_{k|k,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k,k}, \mathbf{P}^{2,\theta_k}_{k|k,k}), \tag{4.23d}$$

where

$$\mu^{2,\theta_k}_{k|k,k} \propto \mu^{2,\theta_k}_{k|k-N,k}\mathcal{N}(\tilde{\mathbf{x}}^1_k; \hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k}, \mathbf{P}^{2,\theta_k}_{k|k-N,k} + \widetilde{\mathbf{P}}^1_k), \tag{4.24a}$$

$$\hat{\mathbf{x}}^{2,\theta_k}_{k|k,k} \triangleq \mathbf{P}^{2,\theta_k}_{k|k,k}\big((\mathbf{P}^{2,\theta_k}_{k|k-N,k})^{-1}\hat{\mathbf{x}}^{2,\theta_k}_{k|k-N,k} + (\widetilde{\mathbf{P}}^1_k)^{-1}\tilde{\mathbf{x}}_k\big), \tag{4.24b}$$

$$\mathbf{P}^{2,\theta_k}_{k|k,k} \triangleq \big((\mathbf{P}^{2,\theta_k}_{k|k-N,k})^{-1} + (\widetilde{\mathbf{P}}^1_k)^{-1}\big)^{-1}. \tag{4.24c}$$

Marginalizing out the hypothesis variable θ_k from the joint fused distribution $p^2(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k})$, we get the approximate marginal fused distribution below.

$$p^2(\mathbf{x}_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k}) = \sum_{i=0}^{m_k} \big[\tilde{\mu}^{1,0}_k \mu^{2,i}_{k|k-N,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,i}_{k|k-N,k}, \mathbf{P}^{2,i}_{k|k-N,k}) + (1-\tilde{\mu}^{1,0}_k)\mu^{2,i}_{k|k,k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{2,i}_{k|k,k}, \mathbf{P}^{2,i}_{k|k,k})\big]. \tag{4.25}$$

4.6 Simulation Results

In order to test the performance of the investigated decorrelation strategy, we consider two 2D maneuvering target tracking scenarios, which are given in two different subsections below.

4.6.1 Model-Match Case

For this case, we generate the true target data using a linear Gaussian system in which the base-state $\mathbf{x}_k \in \mathbb{R}^4$ is composed of Cartesian xy positions and velocities. The tar-

get moves according to a nearly constant velocity model which is written in discrete-time directly.

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \frac{T^2}{2}\mathbf{I}_2 \\ T\mathbf{I}_2 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}, \qquad \mathbf{D} = \mathbf{I}_2, \tag{4.26a}$$

$$\mathbf{Q} = \sigma^2 \mathbf{I}_2, \qquad \mathbf{R} = \sigma_y^2 \mathbf{I}_2, \tag{4.26b}$$

where $\sigma = 5\,\mathrm{m/s^2}$ and $\sigma_y = 20\,\mathrm{m}$. We initiate the base state at the values given below.

$$\mathbf{x}_0 = \begin{bmatrix} 0\,\mathrm{m} & 0\,\mathrm{m} & 25\,\mathrm{m/s} & 25\,\mathrm{m/s} \end{bmatrix}^{\mathrm{T}}.$$

An ellipsoidal gate is used with gate probability 0.99. The gate threshold γ_G is calculated as 9.2103 by using χ^2-distribution with degrees of freedom 2. The false alarm rate β_{FA} and the probability of detection P_D which are used in hypothesis probability calculation are taken as 10^{-6} and 0.9, respectively.

We consider the two local agents shown as $\mathcal{F}_1$ and $\mathcal{F}_2$ in Figure 1.3 running their own local PDA filters on the measurements collected by the local sensors shown as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The PDA filters of both nodes have identical parameters which are matched to the true target data. In the simulations the sampling period of $T = 1\,\mathrm{s}$ is used and each node collects a total of 100 measurements. The initial state estimates and covariances of the local PDA filters are initialized using standard two-point initiation from the first two local measurements. Throughout the simulations we give detailed results for the communication period $N = 3$. We are going to illustrate the summary of the results for the cases $N = 1, 2, 3, 4, 5$. In the following we compare the performances of the following algorithms on the same scenario.

- Baseline Methods:
  - Local Agent: This is the local PDA filter of the agent $\mathcal{F}_1$.
  - Centralized PDA Filter (Centralized): This is a centralized PDA filter using the measurements of both $\mathcal{F}_1$ and $\mathcal{F}_2$.
  - Restarted Local Filter (Restarted LF): In this method, the decorrelated quantities in $\mathcal{F}_1$ are calculated by running an auxiliary (shadow) PDA filter which is restarted after every information transmission (in addition to

the primary/principal PDA filter which is not restarted). The auxiliary PDA filter has the same parameters as the primary PDA filter and it is initiated after each data transmission using posterior state estimate of the primary PDA filter at the same time instant (the time instant of the data transmission). The initial covariance of the auxiliary PDA is set to $10^{14}\mathbf{I}_4$, where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$. This initial covariance selection makes the values of initial state estimate unimportant. The auxiliary PDA filter is run with the local gated measurements of $\mathcal{F}_1$ until the next transmission time and the output estimate and covariance at the next transmission time is transmitted to $\mathcal{F}_2$. The main steps of Restarted LF for the PDA filter are given below;

Suppose that the last information transmission happened at time $k-N$ and the next information transmission will be at time k. Then we apply the following steps in the auxiliary PDA filter.

* Initialization: Initialize the auxiliary PDA filter at time $k-N+1$ with independent information.

$$\hat{\mathbf{x}}_{k-N+1|k-N+1} = \hat{\mathbf{x}}_{k-N|k-N}, \tag{4.27a}$$

$$\mathbf{P}_{k-N+1|k-N+1} = 10^{14}\mathbf{I}_n, \tag{4.27b}$$

  where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$.

* Recursion: Update the PDA filter recursively in time using the measurements $\mathbf{Y}_{k-N+1:k}$ and obtain the estimate $\hat{\mathbf{x}}^i_{k|k}$ and covariance $\mathbf{P}^i_{k|k}$.
* Transmission: Transmit the estimate $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ to the fusion node.

- Proposed Method:
  - OLD: This is the OLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.
  - HLD: This is the HLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.

RMS position errors, RMS velocity errors and mean NEES values obtained from 10000 MC runs are shown in Figure 4.4a, Figure 4.4b and Figure 4.4c, respectively.

It is seen from the Figure 4.4c that the centralized approach yields the best RMS error performance as expected. In terms of RMS position and velocity errors, both OLD and HLD have very similar performances. In terms of RMS position the results for OLD are slightly better than HLD. However, the RMS velocity error results for OLD is slightly over-confident. As can be seen in Figure 4.4b OLD beats the centralized filter. This can be explained as follows. When the target moves with constant velocity, there is no process noise. On the other hand, PDA filters use non-zero process noise standard deviations making them sub-optimal. Since OLD cannot compensate correlations caused by previous communications sufficiently well, PDA filter in the fusion node becomes over-confident. This is, in effect, equivalent to the PDA filter in the fusion node using a lower process noise covariance which is closer to the true target behavior making it closer to optimality. On the other hand, RMS error performance of Restarted LF is worse than the other fusion methods and similar to the Local Agent performance. It is observed that Restarted LF is not preferable for the PDA filter.

It is observed that OLD has better NEES performance unlike what we observed for OLD with the IMM filter. This confirms that the effects of the approximations are not very significant in OLD. The Restarted LF and local agent have slightly lower mean NEES values than OLD/HLD. This is due to the fact that the effects of the hypotheses that measurements do not belong to the real target within the validation region are more dominant in Restarted LF/local agent filter compared to OLD/HLD. RMS erros of Restarted LF is higher than the other filters (i.e. centralized filter and all proposed decorrelation approaches) so the number of false alarms falling inside the validatin region might be higher in Restarted LF. Since in the PDA filter, as all hypotheses are merged according to their hypothesis weights to calculate the posterior, the covariance estimates of the Restarted LF will result in lower NEES values (i.e. underconfidence). We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 10000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 4.5a, Figure 4.5b, and Figure 4.5c, respectively. Here, the averages of the RMS errors and mean NEES values are taken over the time instants in the interval $[20\,\mathrm{s},\ 100\,\mathrm{s}]$ to avoid the initial transient period. It is seen that, OLD and HLD give almost the same results as the centralized filter for $N = 1$ since there is no effect of process noise when $N = 1$. Restarted LF, on the other hand, suffers from

high RMS errors even for $N = 1$. However, Restarted LF has lower NEES values than the other approaches same as local agent. For $N \geq 2$ approximations involved in both OLD & HLD make their results differ from those of the centralized filter as expected. For all N, OLD and HLD show similar performances in RMS errors. For $N = 1; 2; 3$, OLD is slightly better than HLD in RMS errors. On the other hand for $N \geq 4$ RMS errors of OLD are similar to those of HLD. As N increases, the effects of previous communication get less important and effects of process noise start to dominate. For all N, HLD shows better performance in average NEES. As N gets higher we see that Restarted LF has still the worst RMS error performance, but better average NEES values. Overall HLD seems to be more effective for a PDA filter with model-match scenario.

4.6.2 Model-Mismatch Case

In this case we use the fixed true 2D target trajectory shown in Figure 3.12 where the target makes a coordinated-turn maneuver in the middle of the scenario. The fusion architecture is kept the same as the model-match case and the sensors collect position measurements with measurement noise standard deviation $\sigma_y = 20\,m$. The PDA filters of the nodes use the parameters of the linear system given in the model-match case. Hence, in this scenario, we have a mismatch between the true target motion and the filter parameters. We run the same algorithms as in Section 4.6.1 and make 1000 MC runs by changing the measurement noise realizations in each run. The reason why the number of MC runs in the model-mismatch case is less than the number of MC cases in the model-match case is that the MC results in the case of model-mismatch of different approaches are smooth enough to be easily compared among each other. We show the resulting RMS position errors, RMS velocity errors and mean NEES values in Figure 4.6a, Figure 4.6b and Figure 4.6c, respectively. Note that in the model-mismatch case the closeness of the performance measures to those of centralized PDA filter is important. A striking observation in Figure 4.6a and 4.6b is that OLD gives similar results with the centralized filter in all (both maneuvering and non-maneuvering) parts of the scenario. In addition, OLD beats HLD in maneuvering parts of the scenario. However OLD performance gets worse than HLD due to over-confidence, which is also visible in Figure 4.6c. When the

target moves with constant velocity, there is no process noise. On the other hand, PDA filters use non-zero process noise standard deviations making them suboptimal. OLD can compensate for most of the effects of the correlations caused by previous communications but it cannot do this perfectly. As a result, the corresponding PDA filter in the fusion node becomes slightly overconfident. Performance of Restarted LF is worse than that of the centralized PDA filter in all parts of the scenario. This is again due to the inability of Restarted LF to compensate for the process noise in a short time period after the restart. As a result, more process noise enters into the estimation errors. This makes the Restarted LF overconfident compared to the centralized PDA filter, which is also visible in Figure 4.6c.

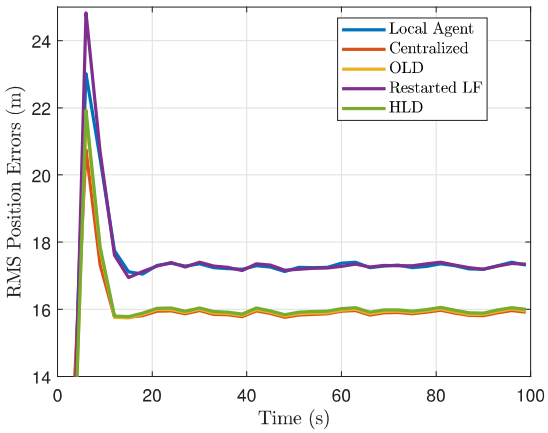
We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 1000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 4.7a, Figure 4.7b, and Figure 4.7c, respectively. Here, the averages of the RMS errors and mean NEES values are taken again over the time instants in the interval $[20\,\mathrm{s},\ 150\,\mathrm{s}]$ to avoid the initial transient period. It is seen that the results for OLD and centralized filter are almost the same for $N = 1$. OLD and HLD have better RMS results than the other fusion methods. However OLD is unable to compensate the correlations caused by previous communications, which results in inconsistency due to process noise for $N \geq 3$. This shows that OLD has smaller decorrelated covariance than necessary as N gets larger. HLD is more consistent compared to OLD as N gets larger. It is also seen that Restarted LF has a similar behavior to the local agent for all N. Overall HLD is the decorrelation method yielding the closest results to the centralized filter in both RMS errors and average NEES values for large N with the model-mismatch scenario. OLD is also the decorrelation method yielding the closest results to the centralized filter in both RMS errors and average NEES values for short N with the model-mismatch scenario.

4.7 Conclusion

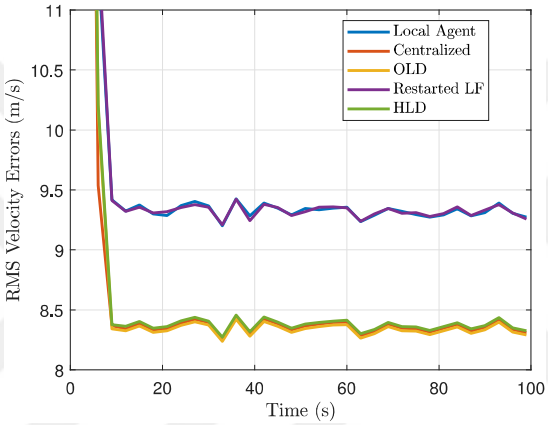
We have investigated a decorrelation approach for the PDA filter. In our investigations we examined the issues that might be encountered in applying decorrelation in a PDA filter in detail and applied covariance inflation method which is proposed for the IMM

filter.

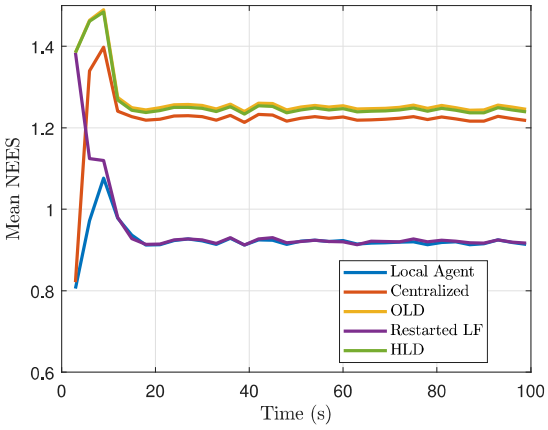
We have seen that the OLD and HLD approaches with covariance inflation applied for both incompatible covariance and outlier decorrelation problems have the closest RMS error to the centralized filter and actually HLD has better consistency properties than the other fusion methods. We have observed also that Restarted LF gives similar results with the local agent filter, i.e., the filter in which the fusion process is not done. As a result, OLD can be used for a decorrelation method for a PDA filter for short communication periods. In addition, HLD seems to be a promising candidate for decorrelation for a PDA filter for low or moderate process noise levels and/or high communication periods.



(a) RMS Position Errors

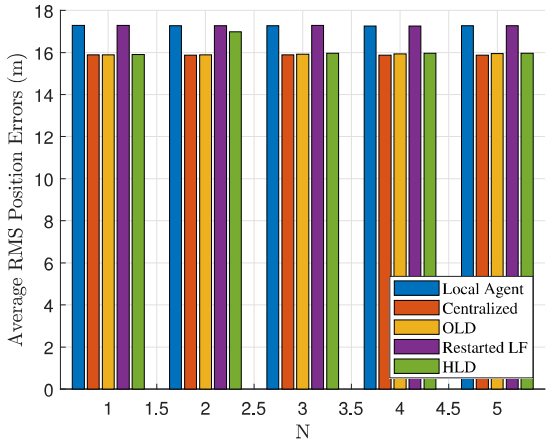


(b) RMS Velocity Errors

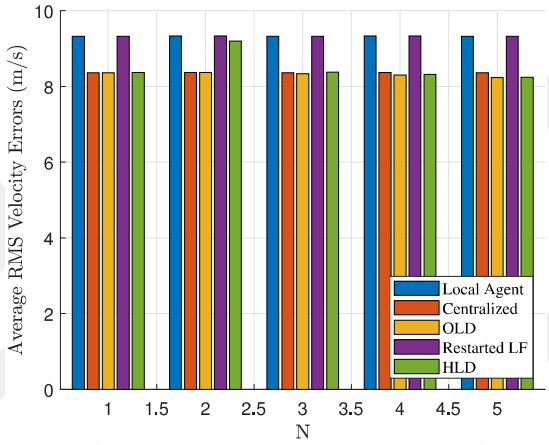


(c) Mean NEES

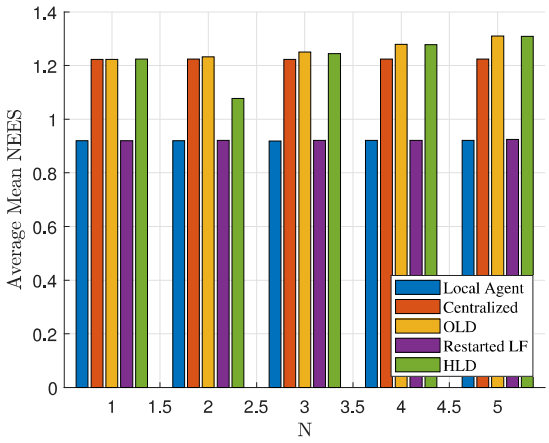
Figure 4.4: PDA Filter - Model Match Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors

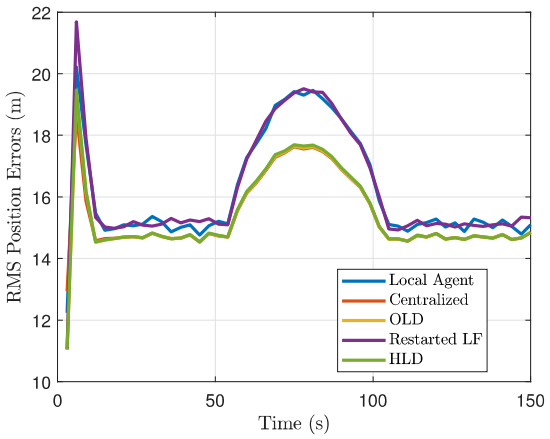


(b) Average RMS Velocity Errors

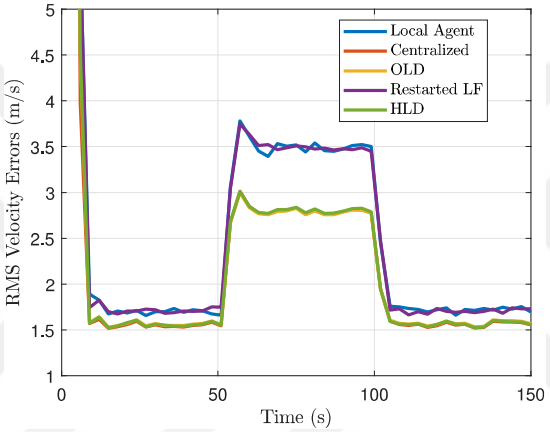


(c) Average Mean NEES

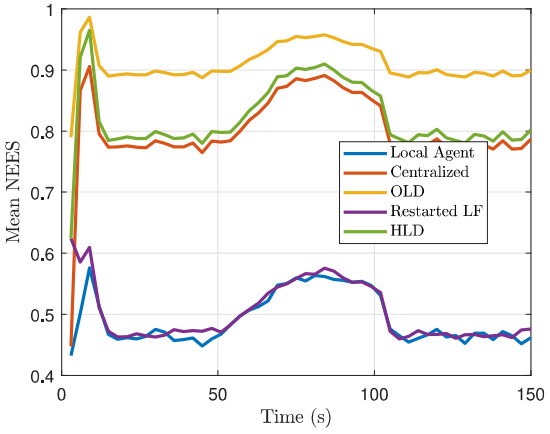
Figure 4.5: PDA Filter - Model Match Case - Comparison of performances of OLD, MLD and alternative approaches for $N = 1, 2, 3, 4, 5$.



(a) RMS Position Errors

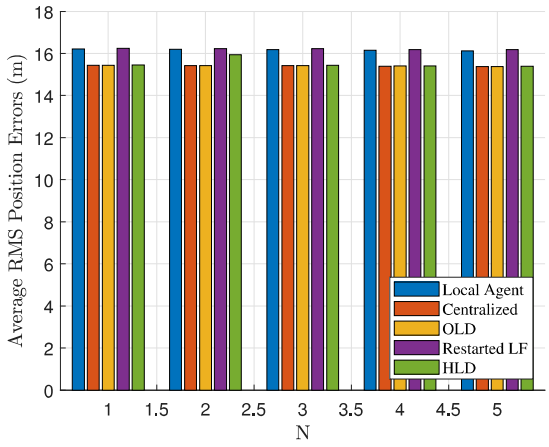


(b) RMS Velocity Errors

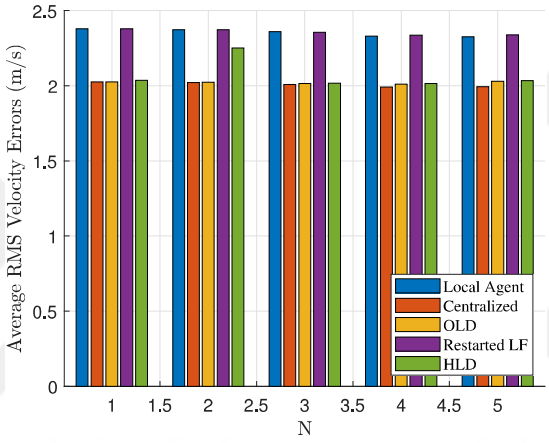


(c) Mean NEES

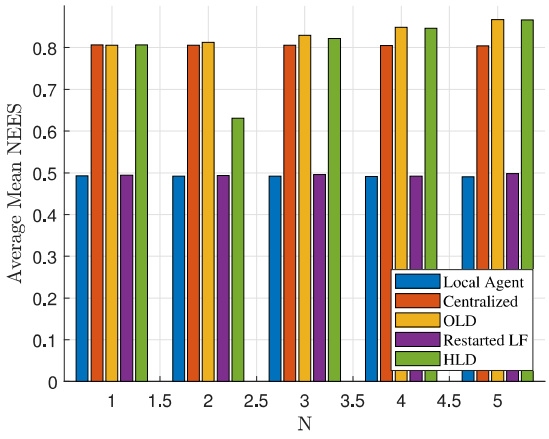
Figure 4.6: PDA Filter - Model Mismatch Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors



(b) Average RMS Velocity Errors



(c) Average Mean NEES

Figure 4.7: PDA Filter - Model Mismatch Case - Comparison of performances of OLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$.

CHAPTER 5

DECORRELATION FOR IMM-PDA FILTER

In this chapter we investigate the problem of information decorrelation for compensating for the correlations caused by previous communication between two local agents in a distributed estimation framework where IMM-PDA filters are used in the local and fusion nodes. In Section 5.1 it is given how the information decorrelation task is related to the IMM-PDA filter. We present a problem formulation in Section 5.2 where we also review the task of information decorrelation. The information decorrelation approach for the IMM-PDA filter is introduced in Section 5.3. Section 5.4 investigates and proposes solutions for the implementation issues that might be encountered while applying the chosen decorrelation approach for the IMM-PDA filter. Section 5.5 describes the fusion operations to be applied in the fusion node. Brief remarks on the computational load of the decorrelation method are made in Appendix B Section B.3. In Section 5.6 we present the numerical results comparing the investigated decorrelation approach with the alternative algorithms on simple distributed maneuvering target tracking examples. The conclusions are drawn in Section 5.7.

5.1 Introduction

IMM and PDA algorithms are combined (i.e. IMM-PDA [57]) to ensure continuous maneuvering target tracking in cluttered environments. While the IMM filter successfully tracks maneuvering targets, the PDA filter ensures that the system is resistant to clutter. The PDA filter evaluates not only the hypotheses that each measurement within the validation region may belong to the real target but also the hypothesis that

none of the measurements falling within the validation region.

In this chapter, the IMM and PDA algorithms are combined in a way that the IMM filter is applied on each hypothesis of the PDA filter and the PDA filter uses all of the validated measurements which are obtained by using merged statistics of the predicted mode matched statistics of the IMM filter. Since the IMM filter is run on each hypothesis, the mode-hypothesis matched state estimates and covariances are calculated. These statistics are merged based on mode-hypothesis weights same as in the IMM filter to obtain mode matched statistics. The mode-hypothesis matched statistics are determined by using each predicted mode matched statistics of the IMM filter. The updated IMM mode probabilities are calculated by using mode-hypothesis probabilities. These mode matched statistics are merged based on IMM mode probabilities with the merging operation to obtain the output estimate. Information decorrelation proposed in this chapter can be done over merged statistics, as well as mode-conditioned statistics or mode-hypothesis-conditioned statistics. As in the IMM/PDA filter, the information decorrelation task for an IMM-PDA filter is directly related to the division of Gaussian mixtures. In this chapter we avoid the problem of ratio of mixtures by applying the decorrelation on the joint posterior (of the base-state, mode-state and hypothesis-state) provided by the IMM-PDA filter.

In this chapter, we first give the information decorrelation approach for the IMM-PDA filter. We investigate and propose solutions for the implementation issues that might be encountered while applying this decorrelation approach which were proposed in [18]. Furthermore, we provide a brief overview of the computation load of the investigated decorrelation approach and give simulation results.

In this chapter, we assume a fusion architecture where correlations are dominated by previous communication between the nodes (rather than the process noise), which necessitates the decorrelation/subtraction of previously communicated information before fusion.

5.2 Problem Definition

We consider two local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ in a sensor network equipped with their own sensors denoted as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ run local Bayesian filters for the following jump Markov linear system (JMLS).

$$\mathbf{x}_k = \mathbf{A}_{r_k}\mathbf{x}_{k-1} + \mathbf{B}_{r_k}\mathbf{w}_{r_k,k}, \tag{5.1a}$$

$$\mathbf{y}_k^s = \mathbf{C}_{r_k}\mathbf{x}_k + \mathbf{D}_{r_k}\mathbf{v}_{r_k,k}^s, \tag{5.1b}$$

where the superscript $s \in \{1, 2\}$ denotes the local agent index; $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denotes the base-state; $r_k \in \{1, \ldots, N_r\}$ denotes the mode-state. Measurements in the gate are shown as $\mathbf{Y}_k = \{\mathbf{y}_k^i\}_{i=1}^{m_k}$. There are the following hypotheses about these measurements,

- $(\theta_k = 0)$: All of $\mathbf{Y}_k$ is false alarm i.e. no target originated measurement in the gate,
- $(\theta_k = i)$: Measurement $\mathbf{y}_k^i$ belongs to target, all the rest are false alarms,

for $i = 1, ..., m_k$. $\mathbf{w}_{r_k,k} \sim \mathcal{N}(\mathbf{w}_{r_k,k}; \mathbf{0}, \mathbf{Q}_{r_k})$ denotes the i.i.d. process noise. $\mathbf{v}_{r_k,k}^s \sim \mathcal{N}(\mathbf{v}_{r_k,k}^s; \mathbf{0}, \mathbf{R}_{r_k}^s)$ denotes the i.i.d. measurement noise for $\mathcal{F}_s$. The measurement noise sequences $\{\mathbf{v}_{r_k,k}^1\}_{k=0}^{\infty}$ and $\{\mathbf{v}_{r_k,k}^2\}_{k=0}^{\infty}$ belonging to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, are assumed to be independent. The mode sequence r_k is assumed to evolve according to a homogeneous Markov chain with transition probability matrix $\mathbf{\Pi} = [\pi_{ji} \triangleq P(r_k = i|r_{k-1} = j)]$. The initial joint distribution of $\{\mathbf{x}_0, r_0\}$ is given as

$$p(\mathbf{x}_0, r_0) = \pi_0^{r_0}\mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0^{r_0}, \mathbf{P}_0^{r_0}). \tag{5.2}$$

We assume that the local Bayesian filter of the local agent $\mathcal{F}_s$ provides the (approximate) joint posterior distribution $p^s(\mathbf{x}_k, r_k|\mathbf{Y}_{0:k}^s)$ of $\{\mathbf{x}_k, r_k\}$ given $\mathbf{Y}_{0:k}^s \triangleq \{\mathbf{Y}_0^s, \mathbf{Y}_1^s, \ldots, \mathbf{Y}_k^s\}$. The simple fusion architecture we consider and the corresponding information graph are shown in Figure 1.3a and Figure 1.3b, respectively. The local agent $\mathcal{F}_1$ communicates its recent information to $\mathcal{F}_2$ with a constant period N. In Figure 1.3b the case with $N = 3$ is shown. We here constrain ourselves to the case where $\mathcal{F}_1$ transmits only the information pertaining to the base-state $\mathbf{x}_k$ to $\mathcal{F}_2$, i.e., the information related to the marginal posterior distribution $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1)$ of

the base-state $\mathbf{x}_k$ is to be communicated to $\mathcal{F}_2$. In general, the state estimation errors of the local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ are correlated due to the common process noise in (5.1a). After $\mathcal{F}_1$ transmits its local information to $\mathcal{F}_2$ for the first time and $\mathcal{F}_2$ fuses this information with the statistics of its own Bayesian filter, the estimates of the local agents become correlated due to previous communication as well. Here, we assume that the correlations between the quantities to be fused are dominated by the previous communication. Even when the effects of the common process noise are negligible, when the correlations due to previous communication are not compensated, the fused quantities might be inconsistent, which might lead to poor performance and/or the divergence of the state estimator.

Suppose now that we are at an intermediate stage of estimation at time k and $\mathcal{F}_1$ transmitted its last information at time $k-N$. Let the marginal posterior distribution $p^1(\mathbf{x}_{k-N}|\mathbf{Y}^1_{0:k-N})$ represent the transmitted information from the local node $\mathcal{F}_1$ to the fusion node $\mathcal{F}_2$ up to (and including) time $k-N$. Since this information has already been used and fused by the fusion center $\mathcal{F}_2$, when the local node $\mathcal{F}_1$ wants to transmit its updated information at time k represented by the current marginal posterior $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$, it first has to decorrelate the previously communicated information $p^1(\mathbf{x}_{k-N}|\mathbf{Y}^1_{0:k-N})$ from the marginal posterior information $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ and then transmit the decorrelated information. We make the following assumption about the JMLS in (5.1a).

Assumption 3 *The state transition matrix $\mathbf{A}_{r_k}$ of (3.1) is independent of the mode sequence r_k and is invertible, i.e., $\mathbf{A}_{r_k} = \mathbf{A}$ for all $r_k \in \{1, \ldots, N_r\}$, where $\mathbf{A}$ is invertible.*

When we neglect the effects of the process noise and make Assumption 3, the decorrelation operation amounts to the division of the updated marginal distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ to the predicted marginal distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$ i.e., the information $\tilde{p}^1(\mathbf{x}_k)$ that has to be sent to $\mathcal{F}_2$ is given as

$$\tilde{p}^1(\mathbf{x}_k) \propto \frac{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})}. \tag{5.3}$$

5.3 Decorrelation Approach

In this section we describe the information decorrelation methodology for the IMM-PDA filter of the local agent $\mathcal{F}_1$. We assume that we are at an intermediate stage of estimation at time k. We assume that the last decorrelated information transmission from $\mathcal{F}_1$ to $\mathcal{F}_2$ happened at time $k-N$. For the IMM-PDA filter, three decorrelation methodologies were proposed:

- Output level decorrelation (OLD),
- Mode level decorrelation (MLD),
- Hypothesis level decorrelation (HLD).

In OLD, the merged output estimates and covariances of the local IMM-PDA filter (i.e., $\{\hat{\mathbf{x}}_k^1, \mathbf{P}_k^1\}$) are decorrelated and transmitted. In MLD, mode-conditioned estimates and covariances of the local IMM-PDA filter, i.e., $\{\mu_k^{1,i}, \hat{\mathbf{x}}_k^{1,i}, \mathbf{P}_k^{1,i}\}_{i=1}^{N_r}$, are decorrelated, merged and transmitted. In HLD, mode-hypothesis matched estimates and covariances of the local IMM-PDA filter, i.e., $\{\rho_k^{1,ij}, \hat{\mathbf{x}}_k^{1,ij}, \mathbf{P}_k^{1,ij}\}_{i=1,j=0}^{N_r,m_k}$, are decorrelated, merged and transmitted. In the following parts of this chapter, we are going to use the parts of the IMM-PDA filter as building blocks for the information decorrelation approach we describe. For this purpose, we define the following functions.

- $\big[\{\bar{\mu}^i, \bar{\mathbf{x}}^i, \overline{\mathbf{P}}^i\}_{i=1}^{N_r}\big]$=**IMMPrediction**$\big(\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=1}^{N_r}\big)$**:** This is the functional representation of the IMM filter prediction update in Algorithm 3 which takes the mode probabilities, estimates and covariances $\{\mu_{k-1}^{s,i}, \hat{\mathbf{x}}_{k-1}^{s,i}, \mathbf{P}_{k-1}^{s,i}\}_{i=1}^{N_r}$ and returns the predicted mode probabilities, estimates and covariances $\{\mu_{k|k-1}^{s,i}, \hat{\mathbf{x}}_{k|k-1}^{s,i}, \mathbf{P}_{k|k-1}^{s,i}\}_{i=1}^{N_r}$.
- $\big[\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=1}^{N_r}\big]$=**MergeHypotheses**$\big(\{\rho^{ij}, \hat{\mathbf{x}}^{ij}, \mathbf{P}^{ij}\}_{i=1,j=0}^{N_r,m_k}\big)$**:** This is the functional representation of the merged estimate calculation part in Algorithm 3 which takes the mode-hypothesis probabilities, state estimates and covariances $\{\rho_k^{s,ij}, \hat{\mathbf{x}}_k^{s,ij}, \mathbf{P}_k^{s,ij}\}_{i=1,j=0}^{N_r,m_k}$ and returns the mode prbabilities, mode-conditioned state estimates and covariances $\{\mu_k^{s,i}, \hat{\mathbf{x}}_k^{s,i}, \mathbf{P}_k^{s,i}\}_{i=1}^{N_r}$.

- $\big[\hat{\mathbf{x}}, \mathbf{P}\big]$=**Merging**$\big(\{\mu^i, \hat{\mathbf{x}}^i, \mathbf{P}^i\}_{i=1}^{N_r}\big)$**:** This is the functional representation of the output estimate calculation part in Algorithm 3 which takes the mode probabilities, estimates and covariances $\{\mu_k^{s,i}, \hat{\mathbf{x}}_k^{s,i}, \mathbf{P}_k^{s,i}\}_{i=1}^{N_r}$ and returns the merged estimate $\hat{\mathbf{x}}_k^s$ and covariance $\mathbf{P}_k^s$.

5.3.1 Output Level Decorrelation (OLD)

The OLD decorrelates and transmits the output estimate and the covariance of the local IMM-PDA filter (i.e., $\hat{\mathbf{x}}_k^1$ and $\mathbf{P}_k^1$). The output estimate and its covariance at the last transmission time $k-N$ are $\hat{\mathbf{x}}_{k-N}^1$ and $\mathbf{P}_{k-N}^1$, respectively. The transmitted information up to (and including) time $k-N$ is hence represented by the Gaussian distribution $N(\mathbf{x}_{k-N}; \hat{\mathbf{x}}_{k-N}^1, \mathbf{P}_{k-N}^1)$. The decorrelation operation requires the predicted version of this distribution to time k. Since no measurement updates will be performed during the prediction steps of the IMM-PDA filter, the PDA filter will not be run and only the IMM prediction process will be applied. Since there are N_r state transition models unlike the case of Kalman filter, we need to use an IMM filter for this prediction operation. For this purpose, we initialize an N_r model IMM predictor with mode probabilities at time $k-N$, equal mode-conditioned estimates and covariances, i.e., $\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}$, $\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^1$ and $\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^1$ for $i = 1, \ldots, N_r$, where we use overlined quantities to separate them from the respective quantities of the primary IMM-PDA filter at the local node. We then use the IMM prediction operation N-times to obtain the predicted mode probabilities $\bar{\mu}_{k|k-N}^{1,i}$, state estimates $\bar{\mathbf{x}}_{k|k-N}^{1,i}$ and covariances $\overline{\mathbf{P}}_{k|k-N}^{1,i}$ for $i = 1, \ldots, N_r$. Finally these quantities are merged with the merging operation to obtain the predicted estimate $\bar{\mathbf{x}}_{k|k-N}^1$ and covariance $\bar{P}_{k|k-N}^1$. Note that the predicted information obtained in this manner will also be represented by a single Gaussian density, i.e., $\bar{p}^1(\mathbf{x}_k|\mathbf{Y}_{0:k-N}^1) = \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^1, \overline{\mathbf{P}}_{k|k-N}^1)$. Hence the new decorrelated information will be represented by a Gaussian $\tilde{p}^1(\mathbf{x}_k|\mathbf{Y}_{0:k}) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1)$, whose parameters can be calculated using the information filter expressions given in (3.7).

In OLD, the decorrelated merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ are transmitted to the second node. The weight $\tilde{\mu}_k^1$, which is due to the division operation, is ignored, where $\tilde{\mu}_k^1 \propto \frac{|\widetilde{\mathbf{P}}_k^1|}{|\mathbf{P}_k^1|} \times \big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^1; \hat{\mathbf{x}}_k^1, \overline{\mathbf{P}}_{k|k-N}^1 - \mathbf{P}_k^1)\big)^{-1}$. This corresponds to the trans-

mission of the density $\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1)$ to the fusion node. The OLD is summarized in Algorithm 12. Section 5.5 gives the fusion formula to be applied into the receiving node. The general decorrelation and transmission operations made in the first node in OLD are illustrated in Figure (5.1).

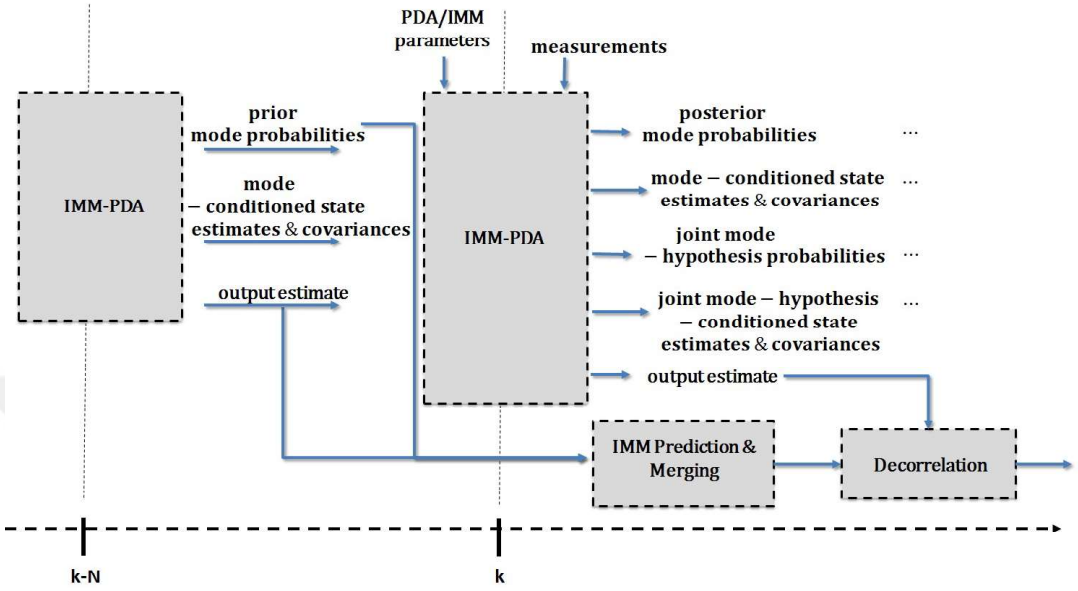


Figure 5.1: OLD decorrelation and transmission operations for the IMM-PDA filter.

Algorithm 12 (OLD)

- *Given:* $\{\mu_{k-N}^{1,i}\}_{i=1}^{N_r}, \hat{\mathbf{x}}_{k-N}^1, \mathbf{P}_{k-N}^1$.
- *Initialization:*

$$\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}, \quad i = 1, \ldots, N_r, \tag{5.4a}$$

$$\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^1, \quad i = 1, \ldots, N_r, \tag{5.4b}$$

$$\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^1, \quad i = 1, \ldots, N_r. \tag{5.4c}$$

- *IMM predictor:*

for $j = k - N + 1, \ldots, k - N$

$$\left[\{\bar{\mu}_{j|k-N}^{1,i}, \bar{\mathbf{x}}_{j|k-N}^{1,i}, \overline{\mathbf{P}}_{j|k-N}^{1,i}\}_{i=1}^{N_r}\right] = \textit{IMMPrediction} \left(\{\bar{\mu}_{j-1|k-N}^{1,i}, \bar{\mathbf{x}}_{j-1|k-N}^{1,i}, \overline{\mathbf{P}}_{j-1|k-N}^{1,i}\}_{i=1}^{N_r}\right).$$

end

- *Merging:*

$$\big[\bar{\mathbf{x}}^1_{k|k-N}, \overline{\mathbf{P}}^1_{k|k-N}\big] = \textit{Merging}\big(\{\bar{\mu}^{1,i}_{k|k-N}, \bar{\mathbf{x}}^{1,i}_{k|k-N}, \overline{\mathbf{P}}^{1,i}_{k|k-N}\}_{i=1}^{N_r}\big).$$

- *Decorrelation:*

$$\tilde{\mathbf{x}}^1_k = \widetilde{\mathbf{P}}^1_k\big(\big(\mathbf{P}^1_k\big)^{-1}\hat{\mathbf{x}}^1_k - \big(\overline{\mathbf{P}}^1_{k|k-N}\big)^{-1}\bar{\mathbf{x}}^1_{k|k-N}\big), \tag{5.5a}$$

$$\widetilde{\mathbf{P}}^1_k = \big(\big(\mathbf{P}^1_k\big)^{-1} - \big(\overline{\mathbf{P}}^1_{k|k-N}\big)^{-1}\big)^{-1}. \tag{5.5b}$$

- *Transmission:* $\tilde{\mathbf{x}}^1_k, \widetilde{\mathbf{P}}^1_k$.

It is important that the decorrelation operation in OLD does not always result in positive semi-definite decorrelated covariance which was examined at the end of Section 5.2. In such a case the transmission of the decorrelated information to the fusion center at the end of OLD cannot be done. We give details about such issues and propose methods to deal with them in Section 5.4.

5.3.2 Mode Level Decorrelation (MLD)

The transmitted information about the base-state up to (and including) time $k-N$ is represented by a Gaussian mixture, i.e., $p^1(\mathbf{x}_{k-N}|\mathbf{Y}^1_{0:k-N}) = \sum_{i=1}^{N_r} \mu^{1,i}_{k-N}\mathcal{N}(\mathbf{x}_{k-N}; \hat{\mathbf{x}}^{1,i}_{k-N}, \mathbf{P}^{1,i}_{k-N})$. Using this information, we can initialize an IMM predictor by setting $\bar{\mu}^{1,i}_{k-N} = \mu^{1,i}_{k-N}$, $\bar{\mathbf{x}}^{1,i}_{k-N} = \hat{\mathbf{x}}^{1,i}_{k-N}$ and $\overline{\mathbf{P}}^{1,i}_{k-N} = \mathbf{P}^{1,i}_{k-N}$ for $i = 1, \ldots, N_r$, where we use overlined quantities to separate them from the respective quantities of the primary IMM-PDA filter at the local agent $\mathcal{F}_1$. If we predict this information to time k using an IMM predictor, we can obtain the predicted mode probabilities $\bar{\mu}^{1,i}_{k|k-N}$, mode-conditioned estimates $\bar{\mathbf{x}}^{1,i}_{k|k-N}$ and covariances $\overline{\mathbf{P}}^{1,i}_{k|k-N}$ of the predicted marginal distribution $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N}) = \sum_{i=1}^{N_r} \bar{\mu}^{1,i}_{k|k-N}\mathcal{N}(\mathbf{x}_{k-N}; \bar{\mathbf{x}}^{1,i}_{k|k-N}, \overline{\mathbf{P}}^{1,i}_{k|k-N})$. Since the current marginal posterior is also a Gaussian mixture, i.e., $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k}) = \sum_{i=1}^{N_r} \mu^{1,i}_k\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^{1,i}_k, \mathbf{P}^{1,i}_k)$, the decorrelation in (5.2) involves the division of two Gaussian mixtures, $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})$ and $p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})$. In MLD, the decorrelation is applied over the joint densities $p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k})$ and $p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N})$ provided by the IMM-PDA filter by making

the following approximation.

$$\frac{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k|\mathbf{Y}^1_{0:k-N})} \approx \sum_{r_k} \frac{p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N})}, \tag{5.6}$$

up to a constant multiplier with respect to $\mathbf{x}_k$. The joint densities $p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k})$ and $p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N})$ are the scaled Gaussian distributions provided by the IMM-PDA filter and IMM predictor as follows.

$$p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k}) = \mu_k^{1,r_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k}, \mathbf{P}_k^{1,r_k}), \tag{5.7a}$$

$$p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N}) = \bar{\mu}_{k|k-N}^{1,r_k} \mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k}). \tag{5.7b}$$

Division of these two joint distributions gives the following joint decorrelated information.

$$\tilde{p}^1(\mathbf{x}_k, r_k) \propto \frac{\mu_k^{1,r_k}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k}, \mathbf{P}_k^{1,r_k})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k})}. \tag{5.8a}$$

Under some assumptions of normalizability with respect to $\mathbf{x}_k$ and marginalizing out the mode variable r_k from the joint decorrelated distribution $\tilde{p}^1(\mathbf{x}_k, r_k)$, we get the approximate marginal decorrelated distribution below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r} \tilde{\mu}_k^{1,i} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}), \tag{5.9}$$

where

$$\tilde{\mu}_k^{1,r_k} \propto \frac{\mu_k^{1,r_k}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{|\widetilde{\mathbf{P}}_k^{1,r_k}|}{|\mathbf{P}_k^{1,r_k}|} \big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,r_k}; \hat{\mathbf{x}}_k^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k} - \mathbf{P}_k^{1,r_k})\big)^{-1}, \tag{5.10a}$$

$$\tilde{\mathbf{x}}_k^{1,r_k} = \widetilde{\mathbf{P}}_k^{1,r_k} \big((\mathbf{P}_k^{1,r_k})^{-1} \hat{\mathbf{x}}_k^{1,r_k} - (\overline{\mathbf{P}}_{k|k-N}^{1,r_k})^{-1} \bar{\mathbf{x}}_{k|k-N}^{1,r_k}\big), \tag{5.10b}$$

$$\widetilde{\mathbf{P}}_k^{1,r_k} = \big((\mathbf{P}_k^{1,r_k})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1,r_k})^{-1}\big)^{-1}. \tag{5.10c}$$

Finally we apply a merging operation to the decorrelated mixture $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$.

The MLD is summarized in Algorithm 13. It is important that the decorrelation operations in MLD might occasionally result in indefinite or even negative definite decorrelated covariances, which was examined at the end of Section 5.2. In such a case the merging operation at the end of MLD cannot be applied. We give details about such issues and propose methods to deal with them in Section 5.4. The general decorrelation and transmission operations made in the first node in MLD are illustrated in Figure (5.2).

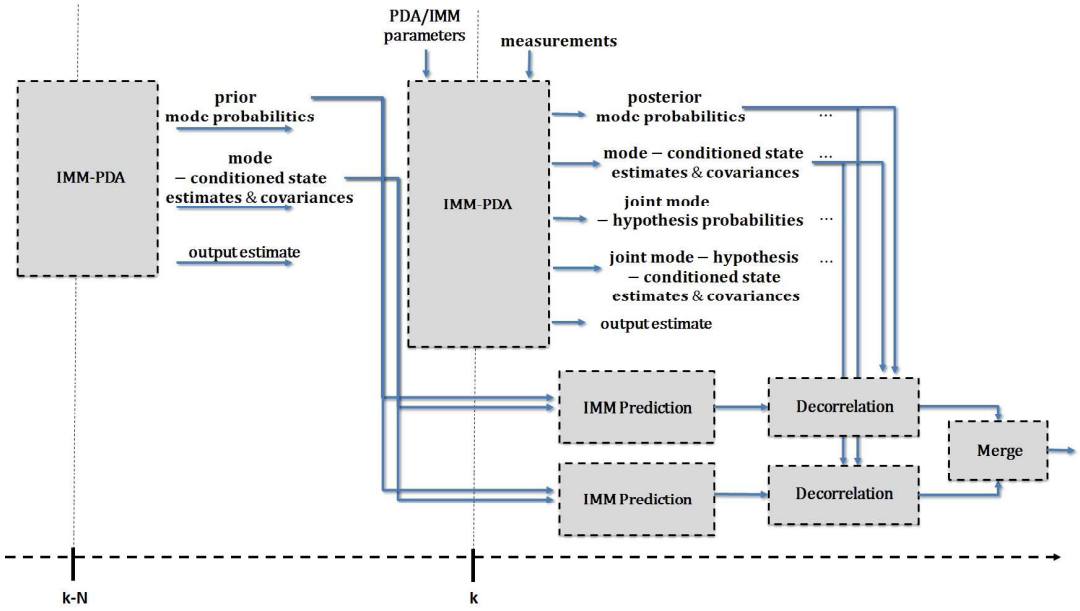


Figure 5.2: MLD decorrelation and transmission operations for the IMM-PDA filter.

Algorithm 13 (MLD)

- *Given:* $\{\mu^{1,i}_{k-N}, \hat{\mathbf{x}}^{1,i}_{k-N}, \mathbf{P}^{1,i}_{k-N}\}_{i=1}^{N_r}$.

- *Initialization:*

$$\bar{\mu}^{1,i}_{k-N} = \mu^{1,i}_{k-N}, \quad i = 1, \ldots, N_r, \tag{5.11a}$$

$$\bar{\mathbf{x}}^{1,i}_{k-N} = \hat{\mathbf{x}}^{1,i}_{k-N}, \quad i = 1, \ldots, N_r, \tag{5.11b}$$

$$\overline{\mathbf{P}}^{1,i}_{k-N} = \mathbf{P}^{1,i}_{k-N}, \quad i = 1, \ldots, N_r. \tag{5.11c}$$

- *IMM predictor:*

For $j = k - N + 1, \ldots, k$,

$$\left[\{\bar{\mu}^{1,i}_{j|k-N}, \bar{\mathbf{x}}^{1,i}_{j|k-N}, \overline{\mathbf{P}}^{1,i}_{j|k-N}\}_{i=1}^{N_r}\right] = \textit{IMMPrediction} \left(\{\bar{\mu}^{1,i}_{j-1|k-N}, \bar{\mathbf{x}}^{1,i}_{j-1|k-N}, \overline{\mathbf{P}}^{1,i}_{j-1|k-N}\}_{i=1}^{N_r}\right).$$

End

- *Decorrelation:*

For $i = 1, \ldots, N_r$,

$$\tilde{\mu}_k^{1,i} \propto \frac{\mu_k^{1,i}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\widetilde{\mathbf{P}}_k^{1,i}|}{|\mathbf{P}_k^{1,i}|} \big(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,i}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,i})\big)^{-1}, \tag{5.12a}$$

$$\tilde{\mathbf{x}}_k^{1,i} = \widetilde{\mathbf{P}}_k^{1,i} \big((\mathbf{P}_k^{1,i})^{-1} \hat{\mathbf{x}}_k^{1,i} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1} \bar{\mathbf{x}}_{k|k-N}^{1,i}\big), \tag{5.12b}$$

$$\widetilde{\mathbf{P}}_k^{1,i} = \big((\mathbf{P}_k^{1,i})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1}\big)^{-1}. \tag{5.12c}$$

End

- *Merging*

$$\big[\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1\big] = \textit{Merging}\big(\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=1}^{N_r}\big).$$

- *Transmission:* $\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1$.

5.3.3 Hypothesis Level Decorrelation (HLD)

The transmitted information about the base-state up to (and including) time $k-N$ is represented by a Gaussian mixture, i.e., $p^1(\mathbf{x}_{k-N}|\mathbf{Y}_{0:k-N}^1) = \sum_{i=1}^{N_r} \mu_{k-N}^{1,i} \mathcal{N}(\mathbf{x}_{k-N}; \hat{\mathbf{x}}_{k-N}^{1,i}, \mathbf{P}_{k-N}^{1,i})$. Using this information, we can initialize an IMM predictor by setting $\bar{\mu}_{k-N}^{1,i} = \mu_{k-N}^{1,i}$, $\bar{\mathbf{x}}_{k-N}^{1,i} = \hat{\mathbf{x}}_{k-N}^{1,i}$ and $\overline{\mathbf{P}}_{k-N}^{1,i} = \mathbf{P}_{k-N}^{1,i}$ for $i = 1, \ldots, N_r$ at the local agent $\mathcal{F}_1$. If we predict this information to time k using an IMM predictor, we can obtain the predicted mode probabilities $\bar{\mu}_{k|k-N}^{1,i}$, mode-conditioned estimates $\bar{\mathbf{x}}_{k|k-N}^{1,i}$ and covariances $\overline{\mathbf{P}}_{k|k-N}^{1,i}$ of the predicted marginal distribution $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k-N}^1) = \sum_{i=1}^{N_r} \bar{\mu}_{k|k-N}^{1,i} \mathcal{N}(\mathbf{x}_{k-N}; \bar{\mathbf{x}}_{k|k-N}^{1,i}, \overline{\mathbf{P}}_{k|k-N}^{1,i})$. Since the current marginal posterior is also a Gaussian mixture, i.e., $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1) = \sum_{i=1}^{N_r} \sum_{j=0}^{m_k} \rho_k^{1,ij} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,ij}, \mathbf{P}_k^{1,ij})$, the decorrelation in (5.2) involves the division of two Gaussian mixtures, $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1)$ and $p^1(\mathbf{x}_k|\mathbf{Y}_{0:k-N}^1)$. In HLD, this problem is overcome by applying the decorrelation over the joint densities $p^1(\mathbf{x}_k, r_k, \theta_k|\mathbf{Y}_{0:k}^1)$ and $p^1(\mathbf{x}_k, r_k|\mathbf{Y}_{0:k-N}^1)$ provided by the IMM-PDA filter and, in effect, by making the following approximation.

$$\frac{p^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1)}{p^1(\mathbf{x}_k|\mathbf{Y}_{0:k-N}^1)} \approx \sum_{r_k} \sum_{\theta_k} \frac{p^1(\mathbf{x}_k, r_k, \theta_k|\mathbf{Y}_{0:k}^1)}{p^1(\mathbf{x}_k, r_k|\mathbf{Y}_{0:k-N}^1)}, \tag{5.13}$$

up to a constant multiplier with respect to $\mathbf{x}_k$. The joint densities $p^1(\mathbf{x}_k, r_k, \theta_k|\mathbf{Y}_{0:k}^1)$ and $p^1(\mathbf{x}_k, r_k|\mathbf{Y}_{0:k-N}^1)$ are the scaled Gaussian distributions provided by the IMM-

PDA filter and predictor as follows.

$$p^1(\mathbf{x}_k, r_k, \theta_k|\mathbf{Y}^1_{0:k}) = \rho_k^{1,r_k\theta_k}\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k\theta_k}, \mathbf{P}_k^{1,r_k\theta_k}), \tag{5.14a}$$

$$p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N}) = \bar{\mu}^{1,r_k}_{k|k-N}\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^{1,r_k}_{k|k-N}, \overline{\mathbf{P}}^{1,r_k}_{k|k-N}). \tag{5.14b}$$

Division of these two joint distributions gives the following joint decorrelated information.

$$\tilde{p}^1(\mathbf{x}_k, r_k, \theta_k) \propto \frac{p^1(\mathbf{x}_k, r_k, \theta_k|\mathbf{Y}^1_{0:k})}{p^1(\mathbf{x}_k, r_k|\mathbf{Y}^1_{0:k-N})}, \tag{5.15a}$$

$$= \frac{\rho_k^{1,r_k\theta_k}}{\bar{\mu}^{1,r_k}_{k|k-N}} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k\theta_k}, \mathbf{P}_k^{1,r_k\theta_k})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}^{1,r_k}_{k|k-N}, \overline{\mathbf{P}}^{1,r_k}_{k|k-N})}. \tag{5.15b}$$

Under some assumptions of normalizability with respect to $\mathbf{x}_k$, the division of two Gaussian distributions above can be written as a scaled Gaussian, which results in the following joint decorrelated distribution.

$$\tilde{p}^1(\mathbf{x}_k, r_k, \theta_k) = \tilde{\rho}_k^{1,r_k\theta_k}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_k\theta_k}, \widetilde{\mathbf{P}}_k^{1,r_k\theta_k}), \tag{5.16}$$

where

$$\tilde{\rho}_k^{1,r_k\theta_k} \propto \frac{\rho_k^{1,r_k\theta_k}}{\bar{\mu}^{1,r_k}_{k|k-N}} \frac{|\widetilde{\mathbf{P}}_k^{1,r_k\theta_k}|}{|\mathbf{P}_k^{1,r_k\theta_k}|} \big(\mathcal{N}(\bar{\mathbf{x}}^{1,r_k}_{k|k-N}; \hat{\mathbf{x}}_k^{1,r_k\theta_k}, \overline{\mathbf{P}}^{1,r_k}_{k|k-N} - \mathbf{P}_k^{1,r_k\theta_k})\big)^{-1}, \tag{5.17a}$$

$$\tilde{\mathbf{x}}_k^{1,r_k\theta_k} = \widetilde{\mathbf{P}}_k^{1,r_k\theta_k}\big((\mathbf{P}_k^{1,r_k\theta_k})^{-1}\hat{\mathbf{x}}_k^{1,r_k\theta_k} - (\overline{\mathbf{P}}^{1,r_k}_{k|k-N})^{-1}\bar{\mathbf{x}}^{1,r_k}_{k|k-N}\big), \tag{5.17b}$$

$$\widetilde{\mathbf{P}}_k^{1,r_k\theta_k} = \big((\mathbf{P}_k^{1,r_k\theta_k})^{-1} - (\overline{\mathbf{P}}^{1,r_k}_{k|k-N})^{-1}\big)^{-1}. \tag{5.17c}$$

Marginalizing out the hypothesis variable θ_k from the joint decorrelated distribution $\tilde{p}^1(\mathbf{x}_k, r_k, \theta_k)$, we get the approximate distribution below.

$$\tilde{p}^1(\mathbf{x}_k, r_k) = \sum_{j=0}^{m_k} \tilde{\rho}_k^{1,r_kj}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_kj}, \widetilde{\mathbf{P}}_k^{1,r_kj}). \tag{5.18}$$

We apply a merging operation to the mixture $\tilde{p}^1(\mathbf{x}_k, r_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^{1,r_k}$ and covariance $\widetilde{\mathbf{P}}_k^{1,r_k}$.

$$\tilde{p}^1(\mathbf{x}_k, r_k) = \tilde{\mu}_k^{1,r_k}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_k}, \widetilde{\mathbf{P}}_k^{1,r_k}), \tag{5.19}$$

where

$$\tilde{\mu}_k^{1,r_k} = \sum_{j=0}^{m_k} \tilde{\rho}_k^{1,r_kj}. \tag{5.20}$$

Marginalizing out the mode variable r_k from the joint decorrelated distribution $\tilde{p}^1(\mathbf{x}_k, r_k)$, we get the approximate marginal decorrelated distribution below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r} \tilde{\mu}_k^{1,i} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}). \tag{5.21}$$

Finally we apply a merging operation again to the decorrelated mixture $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$.

When the transmission period N is equal to 1, the joint decorrelated information related to the zero-hypothesis (i.e, no measurement belonging to the true target in the validation region) can be written as.

$$\begin{aligned}
\tilde{p}^1(\mathbf{x}_k, r_k, \theta_k = 0) \propto & \frac{p^1(\mathbf{x}_k, r_k, \theta_k = 0 | \mathbf{Y}_{0:k}^1)}{p^1(\mathbf{x}_k, r_k | \mathbf{Y}_{0:k-N}^1)}, && \text{(5.22a)} \\
= & \frac{\rho_k^{1,r_k 0}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^{1,r_k 0}, \mathbf{P}_k^{1,r_k 0})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k})}, && \text{(5.22b)} \\
= & \frac{\rho_k^{1,r_k 0}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k})}{\mathcal{N}(\mathbf{x}_k; \bar{\mathbf{x}}_{k|k-N}^{1,r_k}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k})}, && \text{(5.22c)} \\
= & \frac{\rho_k^{1,r_k 0}}{\bar{\mu}_{k|k-N}^{1,r_k}}, && \text{(5.22d)} \\
= & \tilde{\rho}_k^{1,r_k 0}. && \text{(5.22e)}
\end{aligned}$$

Marginalizing out the hypothesis variable θ_k from the joint decorrelated information $\tilde{p}^1(\mathbf{x}_k, r_k, \theta_k)$, we get the approximate decorrelated information below.

$$\tilde{p}^1(\mathbf{x}_k, r_k) = \tilde{\rho}_k^{1,r_k 0} + (1 - \tilde{\rho}_k^{1,r_k 0}) \sum_{j=1}^{m_k} \tilde{\rho}_k^{1,r_k j} \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_k j}, \widetilde{\mathbf{P}}_k^{1,r_k j}), \tag{5.23}$$

where

$$\tilde{\rho}_k^{1,r_k 0} = \frac{\tilde{\rho}_k^{1,r_k 0}}{\sum_{i=1}^{N_r} \rho_k^{1,i0} + \sum_{i=1}^{N_r} \sum_{j=1}^{m_k} \frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\widetilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij}) \right)^{-1}}, \tag{5.24a}$$

$$\tilde{\rho}_k^{1,r_k j} = \frac{\frac{\rho_k^{1,r_k j}}{\bar{\mu}_{k|k-N}^{1,r_k}} \frac{|\widetilde{\mathbf{P}}_k^{1,r_k j}|}{|\mathbf{P}_k^{1,r_k j}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,r_k}; \hat{\mathbf{x}}_k^{1,r_k j}, \overline{\mathbf{P}}_{k|k-N}^{1,r_k} - \mathbf{P}_k^{1,r_k j}) \right)^{-1}}{\sum_{i=1}^{N_r} \sum_{j=1}^{m_k} \frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\widetilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij}) \right)^{-1}}. \tag{5.24b}$$

We apply a merging operation to the mixture $\tilde{p}^1(\mathbf{x}_k, r_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^{1,r_k}$ and covariance $\widetilde{\mathbf{P}}_k^{1,r_k}$.

$$\tilde{p}^1(\mathbf{x}_k, r_k) = \tilde{\rho}_k^{1,r_k 0} + (1 - \tilde{\rho}_k^{1,r_k 0})\tilde{\mu}_k^{1,r_k}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,r_k}, \widetilde{\mathbf{P}}_k^{1,r_k}), \tag{5.25}$$

where

$$\tilde{\mu}_k^{1,r_k} = \sum_{j=1}^{m_k} \tilde{\rho}_k^{1,r_k j}, \tag{5.26a}$$

$$\hat{\rho}_{k|k}^{1,r_k j} = \frac{\tilde{\rho}_k^{1,r_k j}}{\sum_{j=0}^{m_k} \tilde{\rho}_k^{1,r_k j}}, \tag{5.26b}$$

$$\tilde{\mathbf{x}}_k^{1,r_k} = \sum_{j=1}^{m_k} \hat{\rho}_k^{1,r_k j}\tilde{\mathbf{x}}_k^{1,r_k j}, \tag{5.26c}$$

$$\widetilde{\mathbf{P}}_k^{1,r_k} = \sum_{j=1}^{m_k} \hat{\rho}_k^{1,r_k j}\left[\widetilde{\mathbf{P}}_k^{1,r_k j} + (\tilde{\mathbf{x}}_k^{1,r_k j} - \tilde{\mathbf{x}}_k^{1,r_k j})(\cdot)^{\mathrm{T}}\right]. \tag{5.26d}$$

Marginalizing out the mode variable r_k from the mixture part of the joint decorrelated information $\tilde{p}^1(\mathbf{x}_k, r_k)$, we get the approximate marginal decorrelated information below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r} \tilde{\rho}_k^{1,i0} + \sum_{i=1}^{N_r} \left(1 - \tilde{\rho}_k^{1,i0}\right) \tilde{\mu}_k^{1,i}\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}), \tag{5.27}$$

Finally we apply a merging operation only to the decorrelated mixture part of $\tilde{p}^1(\mathbf{x}_k)$ to obtain the merged estimate $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ to be transmitted to $\mathcal{F}_2$ with a constant $\sum_{i=1}^{N_r} \tilde{\rho}_k^{1,i0}$.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r} \tilde{\rho}_k^{1,i0} + \left(1 - \sum_{i=1}^{N_r} \tilde{\rho}_k^{1,i0}\right)\mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1) \tag{5.28}$$

where $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ are given as

$$\tilde{\mathbf{x}}_k^1 = \sum_{i=1}^{N_r} \tilde{\mu}_k^{1,i}\tilde{\mathbf{x}}_k^{1,i}, \tag{5.29a}$$

$$\widetilde{\mathbf{P}}_k^1 = \sum_{i=1}^{N_r} \tilde{\mu}_k^{1,i}\left[\widetilde{\mathbf{P}}_k^{1,i} + (\tilde{\mathbf{x}}_k^{1,i} - \tilde{\mathbf{x}}_k^1)(\cdot)^{\mathrm{T}}\right]. \tag{5.29b}$$

The HLD is summarized in Algorithm 14. Note that the decorrelation operations in HLD might occasionally result in indefinite or even negative definite decorrelated covariances, which was examined at the end of Section 5.2. In such a case the merging

operation at the end of HLD cannot be applied. We give details about such issues and propose methods to deal with them in Section 5.4. The general decorrelation and transmission operations made in the first node in HLD are illustrated in Figure (5.3).

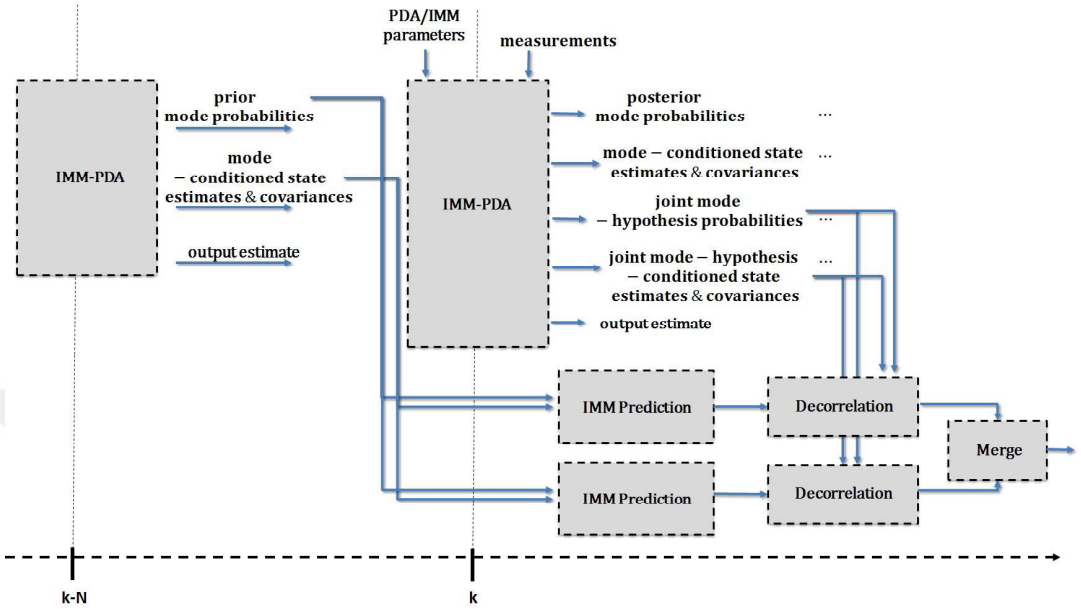


Figure 5.3: HLD decorrelation and transmission operations for the IMM-PDA filter.

Algorithm 14 (HLD)

- *Given:* $\{\mu^{1,i}_{k-N}, \hat{\mathbf{x}}^{1,i}_{k-N}, \mathbf{P}^{1,i}_{k-N}\}_{i=1}^{N_r}$.
- *Initialization:*

$$\bar{\mu}^{1,i}_{k-N} = \mu^{1,i}_{k-N}, \quad i = 1, \ldots, N_r, \tag{5.30a}$$

$$\bar{\mathbf{x}}^{1,i}_{k-N} = \hat{\mathbf{x}}^{1,i}_{k-N}, \quad i = 1, \ldots, N_r, \tag{5.30b}$$

$$\overline{\mathbf{P}}^{1,i}_{k-N} = \mathbf{P}^{1,i}_{k-N}, \quad i = 1, \ldots, N_r. \tag{5.30c}$$

- *IMM predictor:*

For $j = k - N + 1, \ldots, k$,

$$\left[\{\bar{\mu}^{1,i}_{j|k-N}, \bar{\mathbf{x}}^{1,i}_{j|k-N}, \overline{\mathbf{P}}^{1,i}_{j|k-N}\}_{i=1}^{N_r}\right] = \textit{IMMPrediction} \left(\{\bar{\mu}^{1,i}_{j-1|k-N}, \bar{\mathbf{x}}^{1,i}_{j-1|k-N}, \overline{\mathbf{P}}^{1,i}_{j-1|k-N}\}_{i=1}^{N_r}\right).$$

End

- *Decorrelation:*

For $i = 1, \ldots, N_r$,

$if N == 1$,

For $j = 1, \ldots, m_k$,

$$\tilde{\rho}_k^{1,ij} = \frac{\frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\tilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij})\right)^{-1}}{\sum_{i=1}^{N_r} \sum_{j=1}^{m_k} \frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\tilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij})\right)^{-1}},$$

$$\tilde{\mathbf{x}}_k^{1,ij} = \widetilde{\mathbf{P}}_k^{1,ij} \left((\mathbf{P}_k^{1,ij})^{-1} \hat{\mathbf{x}}_k^{1,ij} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1} \bar{\mathbf{x}}_{k|k-N}^{1,i}\right),$$

$$\widetilde{\mathbf{P}}_k^{1,ij} = \left((\mathbf{P}_k^{1,ij})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1}\right)^{-1}.$$

End

$$\tilde{\rho}_k^{1,i0} = \frac{\check{\rho}_k^{1,i0}}{\sum_{i=1}^{N_r} \rho_k^{1,i0} + \sum_{i=1}^{N_r} \sum_{j=1}^{m_k} \frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\tilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij})\right)^{-1}}.$$

else

For $j = 0, \ldots, m_k$,

$$\tilde{\rho}_k^{1,ij} = \frac{\frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\tilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij})\right)^{-1}}{\sum_{i=1}^{N_r} \sum_{j=0}^{m_k} \frac{\rho_k^{1,ij}}{\bar{\mu}_{k|k-N}^{1,i}} \frac{|\tilde{\mathbf{P}}_k^{1,ij}|}{|\mathbf{P}_k^{1,ij}|} \left(\mathcal{N}(\bar{\mathbf{x}}_{k|k-N}^{1,i}; \hat{\mathbf{x}}_k^{1,ij}, \overline{\mathbf{P}}_{k|k-N}^{1,i} - \mathbf{P}_k^{1,ij})\right)^{-1}},$$

$$\tilde{\mathbf{x}}_k^{1,ij} = \widetilde{\mathbf{P}}_k^{1,ij} \left((\mathbf{P}_k^{1,ij})^{-1} \hat{\mathbf{x}}_k^{1,ij} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1} \bar{\mathbf{x}}_{k|k-N}^{1,i}\right),$$

$$\widetilde{\mathbf{P}}_k^{1,ij} = \left((\mathbf{P}_k^{1,ij})^{-1} - (\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1}\right)^{-1}.$$

End

End

End

- *Hypothesis Merging:*

$If N == 1$

$$\left[\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=1}^{N_r}\right] = \textit{MergeHypotheses}\left(\{\tilde{\rho}_k^{1,ij}, \tilde{\mathbf{x}}_k^{1,ij}, \widetilde{\mathbf{P}}_k^{1,ij}\}_{i=1,j=1}^{N_r,m_k}\right).$$

else

$$\left[\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=1}^{N_r}\right] = \textit{MergeHypotheses}\left(\{\tilde{\rho}_k^{1,ij}, \tilde{\mathbf{x}}_k^{1,ij}, \widetilde{\mathbf{P}}_k^{1,ij}\}_{i=1,j=0}^{N_r,m_k}\right).$$

End

- *Mode Merging:*

$$\left[\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1\right] = \textit{Merging}\left(\{\tilde{\mu}_k^{1,i}, \tilde{\mathbf{x}}_k^{1,i}, \widetilde{\mathbf{P}}_k^{1,i}\}_{i=1}^{N_r}\right).$$

- *Transmission:*

If $N == 1$,

$\sum_{i=1}^{N_r} \tilde{\rho}_k^{1,0i}, \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1$.

else

$\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1$.

End

5.4 Implementation Issues

All of the proposed decorrelation approaches for the IMM-PDA filter involve the information subtraction operations given in (3.18), the quantities $\{\hat{\mathbf{x}}, \mathbf{P}\}$, $\{\overline{\mathbf{x}}, \overline{\mathbf{P}}\}$, $\{\tilde{\mathbf{x}}, \widetilde{\mathbf{P}}\}$ represent the mean and covariance (of a component) of the current distribution, predicted distribution and decorrelated distribution, respectively. As in the IMM/PDA filter case, we have encountered both incompatible covariance and outlier decorrelation problem related to the information matrix difference in (3.18b) while applying decorrelation process to the merged, mode-conditioned or mode-hypothesis matched statistics for the IMM-PDA filter.

5.4.1 Use of covariance inflation in OLD & MLD & HLD for the IMM-PDA filter

In the implementation of OLD, MLD and HLD, at each decorrelation step we check the existence of the incompatible covariance and outlier decorrelation problems. When one or more of these problems are detected, a compensation in the form of covariance inflation is applied. The related steps to be added into the OLD, MLD and HLD are given in Algorithms 15, 16 and 17, respectively.

In all decorrelation algorithms, we first check for the existence of the incompatible covariance problem, then we handle the outlier decorrelation problems.

- In OLD, we first check for the existence of the incompatible covariance problem. In the case of incompatible covariance problem, we calculate the lower bound κ_{IC} for the inflation constant κ and inflate the predicted covariance with the constant $c_{\mathrm{IC}}\kappa_{\mathrm{IC}}$ where $c_{\mathrm{IC}} > 1$ is a safety factor.

- In MLD we first check for the existence of the incompatible covariance problem for each mode ($i = 1, \ldots, N_r$). If we find such a problem in a mode, we calculate the lower bound κ_{IC}^{i}, $i = 1, \ldots, N_r$ for the inflation constant κ for that mode. If there exists incompatible covariance problem in at least one mode, we inflate the predicted covariances of all modes with the maximum of the values $c_{\mathrm{IC}}\kappa_{\mathrm{IC}}^{i}$, $i = 1, \ldots, N_r$, where $c_{\mathrm{IC}} > 1$ is a safety factor. In MLD, in the case of incompatible covariance problem in a mode, the predicted covariances of all modes should be inflated with the same inflation constant in order not to create artificial irregularities in the decorrelated weights $\tilde{\mu}_k^{1,i}$, $i = 1, \ldots, N_r$. Inflating only the predicted covariance of the modes in which there is an incompatible covariance problem would cause unintended irregularities in the decorrelated weights $\tilde{\mu}_k^{1,i}$, $i = 1, \ldots, N_r$ which are highly dependent on the predicted covariances.

- In HLD we first check for the existence of the incompatible covariance problem for each mode-hypothesis ($i = 1, \ldots, N_r, j = 0, \ldots, m_k$). If we find such a problem in a mode,hypothesis, we calculate the lower bound $\kappa_{\mathrm{IC}}^{ij}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$ for the inflation constant κ for that mode-hypothesis. If there exists incompatible covariance problem in at least one mode-hypothesis, we inflate the predicted covariances of all modes-hypotheses with the maximum of the values $c_{\mathrm{IC}}\kappa_{\mathrm{IC}}^{ij}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$, where $c_{\mathrm{IC}} > 1$ is a safety factor. In HLD, in the case of incompatible covariance problem in a mode-hypothesis, the predicted covariances of all modes-hypotheses should be inflated with the same inflation constant in order not to create artificial irregularities in the decorrelated weights $\tilde{\rho}_k^{1,ij}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$. Inflating only the predicted covariance of the modes-hypotheses in which there is an incompatible covariance problem would cause unintended irregularities in the decorrelated weights $\tilde{\rho}_k^{1,ij}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$.

After handling the incompatible covariance problems, we then check for the existence of the outlier decorrelation problems in OLD, MLD and HLD.

- In OLD, if there is outlier decorrelation problem, we calculate the lower bound κ_{OD} for the inflation constant κ using the Newton method with a fixed num-

ber of iterations M_{OD} and inflate the predicted covariance with the constant $c_{\mathrm{OD}}\kappa_{\mathrm{OD}}$ where $c_{\mathrm{OD}} > 1$ is a safety factor.

- In MLD we check for the existence of the outlier decorrelation problem for each mode ($i = 1, \ldots, N_r$). If we find such a problem in a mode, we calculate the lower bound κ^i_{OD}, $i = 1, \ldots, N_r$ for the inflation constant κ for that mode. If there exists outlier decorrelation problem in at least one mode, we inflate the predicted covariances of all modes with the maximum of the values $c_{\mathrm{OD}}\kappa^i_{\mathrm{OD}}$, $i = 1, \ldots, N_r$ to avoid artificial irregularities in the decorrelated weights $\tilde{\mu}^i_k$, $i = 1, \ldots, N_r$.

- In HLD we check for the existence of the outlier decorrelation problem for each mode-hypothesis ($i = 1, \ldots, N_r, j = 0, \ldots, m_k$). If we find such a problem in a mode-hypothesis, we calculate the lower bound $\kappa^{ij}_{\mathrm{OD}}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$ for the inflation constant κ for that mode-hypothesis. In HLD if there exists outlier decorrelation problem in at least one mode-hypothesis, we inflate the predicted covariances of all modes-hypotheses with the maximum of the values $c_{\mathrm{OD}}\kappa^{ij}_{\mathrm{OD}}$, $i = 1, \ldots, N_r, j = 0, \ldots, m_k$.

Algorithm 15 (OLD Covariance Inflation) *To be inserted before the decorrelation step of OLD in Algorithm 12*

- *Given:* $\hat{\mathbf{x}}^1_k$, $\mathbf{P}^1_k$, $\bar{\mathbf{x}}^1_{k|k-N}$, $\overline{\mathbf{P}}^1_{k|k-N}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.

- *Check and compensate for the incompatible covariance problem*

  $\kappa_{\mathrm{IC}} = \lambda_{\max}\big(\mathbf{P}^1_k(\overline{\mathbf{P}}^1_{k|k-N})^{-1}\big)$.
  If $\kappa_{\mathrm{IC}} \geq 1$,
  Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{IC}}\kappa_{\mathrm{IC}}\overline{\mathbf{P}}^1_{k|k-N}$.
  End

- *Check and compensate for the outlier decorrelation problem*

  $g_{\mathrm{OD}} = (\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}}(\overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)^{-1}(\hat{\mathbf{x}}^1_k - \bar{\mathbf{x}}^1_{k|k-N})$.
  If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,
  $\kappa^{(0)} = 1$.

For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

$\boldsymbol{\Lambda}_m = (\kappa^{(m)} \overline{\mathbf{P}}^1_{k|k-N} - \mathbf{P}^1_k)^{-1}$.

$\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^1_k - \overline{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}} \boldsymbol{\Lambda}_m (\hat{\mathbf{x}}^1_k - \overline{\mathbf{x}}^1_{k|k-N})}{(\hat{\mathbf{x}}^1_k - \overline{\mathbf{x}}^1_{k|k-N})^{\mathrm{T}} \boldsymbol{\Lambda}_m \overline{\mathbf{P}}^1_{k|k-N} \boldsymbol{\Lambda}_m (\hat{\mathbf{x}}^1_k - \overline{\mathbf{x}}^1_{k|k-N})}$.

End

$\kappa_{\mathrm{OD}} = \kappa^{(M_{\mathrm{OD}})}$.

Set $\overline{\mathbf{P}}^1_{k|k-N} \leftarrow c_{\mathrm{OD}} \kappa_{\mathrm{OD}} \overline{\mathbf{P}}^1_{k|k-N}$.

End

Algorithm 16 (MLD Covariance Inflation) *To be inserted before the decorrelation step of MLD in Algorithm 13.*

- *Given:* $\{\hat{\mathbf{x}}^{1,i}_k, \mathbf{P}^{1,i}_k\}_{i=1}^{N_r}$, $\{\overline{\mathbf{x}}^{1,i}_{k|k-N}, \overline{\mathbf{P}}^{1,i}_{k|k-N}\}_{i=1}^{N_r}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Check and compensate for the incompatible covariance problem:*

  For $i = 1, \ldots, N_r$,

  $\kappa^i_{\mathrm{IC}} = \lambda_{\max}\big(\mathbf{P}^{1,i}_k (\overline{\mathbf{P}}^{1,i}_{k|k-N})^{-1}\big)$.

  End

  $\kappa_{\mathrm{IC}} = \max\{\kappa^1_{\mathrm{IC}}, \ldots, \kappa^{N_r}_{\mathrm{IC}}\}$.

  If $\kappa_{\mathrm{IC}} \geq 1$,

  For $i = 1, \ldots, N_r$,

  Set $\overline{\mathbf{P}}^{1,i}_{k|k-N} \leftarrow c_{\mathrm{IC}} \kappa_{\mathrm{IC}} \overline{\mathbf{P}}^{1,i}_{k|k-N}$.

  End

  End

- *Check and compensate for the outlier decorrelation problem:*

  For $i = 1, \ldots, N_r$,

  $g_{\mathrm{OD}} = (\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} (\overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1} (\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})$.

  If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,

  $\kappa^{(0)} = 1$.

  For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

  $\boldsymbol{\Lambda}_m = (\kappa^{(m)} \overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,i}_k)^{-1}$.

  $\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} \boldsymbol{\Lambda}_m (\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})}{(\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}} \boldsymbol{\Lambda}_m \overline{\mathbf{P}}^{1,i}_{k|k-N} \boldsymbol{\Lambda}_m (\hat{\mathbf{x}}^{1,i}_k - \overline{\mathbf{x}}^{1,i}_{k|k-N})}$.

End

$\kappa_{\mathrm{OD}}^{i} = \kappa^{(M_{\mathrm{OD}})}$.

Else

$\kappa_{\mathrm{OD}}^{i} = 0$.

End

End

$\kappa_{\mathrm{OD}} = \max\{\kappa_{\mathrm{OD}}^{1}, \ldots, \kappa_{\mathrm{OD}}^{N_r}\}$.

If $\kappa_{\mathrm{OD}} \geq 1$,

For $i = 1, \ldots, N_r$,

Set $\overline{\mathbf{P}}_{k|k-N}^{1,i} \leftarrow c_{\mathrm{OD}} \kappa_{\mathrm{OD}} \overline{\mathbf{P}}_{k|k-N}^{1,i}$.

End

End

Algorithm 17 (HLD Covariance Inflation) *To be inserted before the decorrelation step of HLD in Algorithm 14.*

- *Given:* $\{\hat{\mathbf{x}}_k^{1,ij}, \mathbf{P}_k^{1,ij}\}_{i=1,j=0}^{N_r,m_k}$, $\{\bar{\mathbf{x}}_{k|k-N}^{1,i}, \overline{\mathbf{P}}_{k|k-N}^{1,i}\}_{i=1}^{N_r}$, γ_{OD}, M_{OD}, c_{IC}, c_{OD}.
- *Find the initial hypothesis number for algorithm to be applied:*

If $N == 1$

$l = 1$,

else

$l = 0$.

End

- *Check and compensate for the incompatible covariance problem:*

For $i = 1, \ldots, N_r$,

For $j = l, \ldots, m_k$,

$\kappa_{\mathrm{IC}}^{ij} = \lambda_{\max}\big(\mathbf{P}_k^{1,ij}(\overline{\mathbf{P}}_{k|k-N}^{1,i})^{-1}\big)$.

End

End

$\kappa_{\mathrm{IC}} = \max\{\kappa_{\mathrm{IC}}^{1l}, \ldots, \kappa_{\mathrm{IC}}^{N_r m_k}\}$.

If $\kappa_{\mathrm{IC}} \geq 1$,

For $i = 1, \ldots, N_r$,

Set $\overline{\mathbf{P}}^{1,i}_{k|k-N} \leftarrow c_{\mathrm{IC}}\kappa_{\mathrm{IC}}\overline{\mathbf{P}}^{1,i}_{k|k-N}$.

End

End

- *Check and compensate for the outlier decorrelation problem:*

For $i = 1, \ldots, N_r$,

For $j = l, \ldots, m_k$,

$g_{\mathrm{OD}} = (\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}}(\overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,ij}_k)^{-1}(\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})$.

If $g_{\mathrm{OD}} \geq \gamma_{\mathrm{OD}}$,

$\kappa^{(0)} = 1$.

For $m = 0, \ldots, M_{\mathrm{OD}} - 1$,

$\mathbf{\Lambda}_m = (\kappa^{(m)}\overline{\mathbf{P}}^{1,i}_{k|k-N} - \mathbf{P}^{1,ij}_k)^{-1}$.

$\kappa^{(m+1)} = \kappa^m + \frac{(\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})}{(\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})^{\mathrm{T}}\mathbf{\Lambda}_m\overline{\mathbf{P}}^{1,i}_{k|k-N}\mathbf{\Lambda}_m(\hat{\mathbf{x}}^{1,ij}_k - \bar{\mathbf{x}}^{1,i}_{k|k-N})}$.

End

$\kappa^{ij}_{\mathrm{OD}} = \kappa^{(M_{\mathrm{OD}})}$.

Else

$\kappa^{ij}_{\mathrm{OD}} = 0$.

End

End

End

$\kappa_{\mathrm{OD}} = \max\{\kappa^{1l}_{\mathrm{OD}}, \ldots, \kappa^{N_r m_k}_{\mathrm{OD}}\}$.

If $\kappa_{\mathrm{OD}} \geq 1$,

For $i = 1, \ldots, N_r$,

Set $\overline{\mathbf{P}}^{1,i}_{k|k-N} \leftarrow c_{\mathrm{OD}}\kappa_{\mathrm{OD}}\overline{\mathbf{P}}^{1,i}_{k|k-N}$.

End

End

5.5 Fusion

In this section we describe the fusion operation to be applied by the local agent $\mathcal{F}_2$ upon receiving the decorrelated information from $\mathcal{F}_1$. After receiving the decorrelated

distribution $\tilde{p}^1(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1)$, i.e., its mean and covariance $\{\tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1\}$, from $\mathcal{F}_1$ at time k, $\mathcal{F}_2$ has to incorporate this information into the summary statistics of its own IMM-PDA filter.

The current joint posterior distribution of the IMM-PDA filter of $\mathcal{F}_2$ before fusion is given as

$$p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2) = \rho_{k|k-N,k}^{2,r_k\theta_k} \times \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k}), \tag{5.31}$$

where $\mathcal{I}_{0:k-N}^1$ denotes the information that $\mathcal{F}_2$ has got from $\mathcal{F}_1$ so far. Note that the summary statistics of the IMM-PDA filter of $\mathcal{F}_2$ before fusion are given as $\{\rho_{k|k-N,k}^{2,ij}, \hat{\mathbf{x}}_{k|k-N,k}^{2,ij}, \mathbf{P}_{k|k-N,k}^{2,ij}\}_{i=1,j=0}^{N_r,m_k}$. $\mathcal{F}_2$ applies naive fusion to fuse the pieces of information described by the distributions $\tilde{p}^1(\mathbf{x}_k)$ and $p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2)$.

$$p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k}^1, \mathbf{Y}_{0:k}^2) \propto \tilde{p}^1(\mathbf{x}_k) p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2), \tag{5.32a}$$

$$= \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1) \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k}), \tag{5.32b}$$

$$= \rho_{k|k,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k,k}^{2,r_k\theta_k}), \tag{5.32c}$$

where

$$\rho_{k|k,k}^{2,r_k\theta_k} \propto \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\tilde{\mathbf{x}}_k^1; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k} + \widetilde{\mathbf{P}}_k^1), \tag{5.33a}$$

$$\hat{\mathbf{x}}_{k|k,k}^{2,r_k\theta_k} \triangleq \mathbf{P}_{k|k,k}^{2,r_k\theta_k} \left((\mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})^{-1} \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k} + (\widetilde{\mathbf{P}}_k^1)^{-1} \tilde{\mathbf{x}}_k^1 \right), \tag{5.33b}$$

$$\mathbf{P}_{k|k,k}^{2,r_k\theta_k} \triangleq \left((\mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})^{-1} + (\widetilde{\mathbf{P}}_k^1)^{-1} \right)^{-1}. \tag{5.33c}$$

The summary statistics of the IMM-PDA filter of $\mathcal{F}_2$ after fusion are then given as $\{\rho_{k|k,k}^{2,ij}, \hat{\mathbf{x}}_{k|k,k}^{2,ij}, \mathbf{P}_{k|k,k}^{2,ij}\}_{i=1,j=0}^{N_k,m_k}$. As given in Algorithm 14, The general fusion operations made in the fusion node in are illustrated in Figure (5.4).

As given in the merged estimation calculation part of the Algorithm 3, the merged mode probability $\mu_{k|k,k}^{2,r_k}$, estimate $\hat{\mathbf{x}}_{k|k,k}^{2,r_k}$ and covariance $\mathbf{P}_{k|k,k}^{2,r_k}$ are calculated as,

$$\mu_{k|k,k}^{2,r_k} = \sum_{j=0}^{m_k} \rho_{k|k,k}^{2,jr_k}, \tag{5.34a}$$

$$\hat{\mathbf{x}}_{k|k,k}^{2,r_k} = \sum_{j=0}^{m_k} \rho_{k|k,k}^{2,jr_k} \hat{\mathbf{x}}_{k|k,k}^{2,r_kj}, \tag{5.34b}$$

$$\mathbf{P}_{k|k,k}^{2,r_k} = \sum_{j=0}^{m_k} \rho_{k|k,k}^{2,jr_k} \left[\mathbf{P}_{k|k,k}^{2,r_kj} + (\hat{\mathbf{x}}_{k|k,k}^{2,r_kj} - \hat{\mathbf{x}}_{k|k,k}^{2,r_k})(.)^{\mathrm{T}} \right]. \tag{5.34c}$$

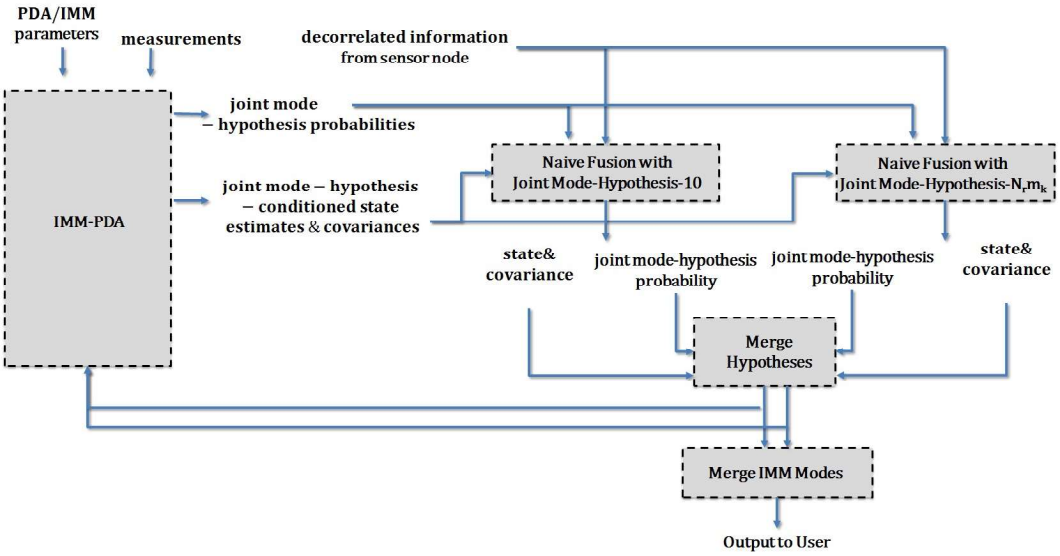


Figure 5.4: Fusion operations for the IMM-PDA filter.

As stated in Section 5.3.3, as a result of $N = 1$ decorrelation, the weight $\sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i}$, mean $\tilde{\mathbf{x}}_k^1$ and covariance $\widetilde{\mathbf{P}}_k^1$ are transmitted to the fusion center, $\mathcal{F}_2$ as below.

$$\tilde{p}^1(\mathbf{x}_k) = \sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i} + \left(1 - \sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i}\right) \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1), \tag{5.35}$$

where N_r^1 is the number of modes of IMM filter in$\mathcal{F}_2$.

$\mathcal{F}_2$ applies naive fusion to fuse the pieces of information described by the distributions $\tilde{p}^1(\mathbf{x}_k)$ and $p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2)$.

$$p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k}^1, \mathbf{Y}_{0:k}^2) \propto \tilde{p}^1(\mathbf{x}_k) p^2(\mathbf{x}_k, r_k, \theta_k | \mathcal{I}_{0:k-N}^1, \mathbf{Y}_{0:k}^2), \tag{5.36a}$$

$$= \sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i} \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})$$

$$+ \left(1 - \sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i}\right) \mathcal{N}(\mathbf{x}_k; \tilde{\mathbf{x}}_k^1, \widetilde{\mathbf{P}}_k^1) \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k}), \tag{5.36b}$$

$$= \left(\sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i}\right) \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})$$

$$+ \left(1 - \sum_{i=1}^{N_r^1} \tilde{\rho}_k^{1,0i}\right) \rho_{k|k,k}^{2,r_k\theta_k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k,k}^{2,r_k\theta_k}), \tag{5.36c}$$

where

$$\rho_{k|k,k}^{2,r_k\theta_k} \propto \rho_{k|k-N,k}^{2,r_k\theta_k} \mathcal{N}(\tilde{\mathbf{x}}_k^1; \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k}, \mathbf{P}_{k|k-N,k}^{2,r_k\theta_k} + \widetilde{\mathbf{P}}_k^1), \tag{5.37a}$$

$$\hat{\mathbf{x}}_{k|k,k}^{2,r_k\theta_k} \triangleq \mathbf{P}_{k|k,k}^{2,r_k\theta_k} \big((\mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})^{-1} \hat{\mathbf{x}}_{k|k-N,k}^{2,r_k\theta_k} + (\widetilde{\mathbf{P}}_k^1)^{-1} \tilde{\mathbf{x}}_k^1 \big), \tag{5.37b}$$

$$\mathbf{P}_{k|k,k}^{2,r_k\theta_k} \triangleq \big((\mathbf{P}_{k|k-N,k}^{2,r_k\theta_k})^{-1} + (\widetilde{\mathbf{P}}_k^1)^{-1} \big)^{-1}. \tag{5.37c}$$

5.6 Simulation Results

In order to test the performance of the investigated decorrelation strategy for the IMM-PDA filter, we consider two 2D maneuvering target tracking scenarios, which are given in two different subsections below.

5.6.1 Model-Match Case

For this case, we generate the true target data using a JMLS in which the base-state $\mathbf{x}_k \in \mathbb{R}^4$ is composed of Cartesian xy positions and velocities and the mode-state $r_k \in \{1, 2\}$ takes two values. The JMLS has two nearly constant velocity models (directly written in discrete-time) which only differ in their process noise covariance values.

$$\mathbf{A}_1 = \mathbf{A}_2 = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{bmatrix}, \qquad \mathbf{B}_1 = \mathbf{B}_2 = \begin{bmatrix} \frac{T^2}{2}\mathbf{I}_2 \\ T\mathbf{I}_2 \end{bmatrix},$$

$$\mathbf{C}_1 = \mathbf{C}_2 = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}, \qquad \mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}_2, \tag{5.38a}$$

$$\mathbf{Q}_i = \sigma_i^2 \mathbf{I}_2, \quad i = 1, 2, \qquad \mathbf{R}_1 = \mathbf{R}_2 = \sigma_y^2 \mathbf{I}_2, \tag{5.38b}$$

where $\sigma_1 = 1\,\mathrm{m/s^2}$ and $\sigma_2 = 5\,\mathrm{m/s^2}$, respectively and $\sigma_y = 20\,\mathrm{m}$. The diagonal elements of the transition probability matrix $\mathbf{\Pi} \in \mathbb{R}^{2\times 2}$ for the JMLS are given as $\pi_{11} = \pi_{22} = 0.95$. We initiate the base and mode-states at the values given below.

$$\mathbf{x}_0 = \begin{bmatrix} 0\,\mathrm{m} & 0\,\mathrm{m} & 25\,\mathrm{m/s} & 25\,\mathrm{m/s} \end{bmatrix}^{\mathrm{T}}, \qquad r_0 = 1.$$

An ellipsoidal gate is used with gate probability 0.99. The gate threshold γ_G is calculated as 9.2103 by using χ^2-distribution with degrees of freedom 2. The false alarm

rate β_{FA} and the probability of detection P_D which are used in hypothesis probability calculation are taken as 10^{-6} and 0.9, respectively.

We consider the two local agents shown as $\mathcal{F}_1$ and $\mathcal{F}_2$ in Figure 1.3 running their own local IMM-PDA filters on the measurements collected by the local sensors shown as $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. The IMM-PDA filters of both nodes have identical parameters which are matched to the true target data. In the simulations the sampling period of $T = 1\,\text{s}$ is used and each node collects a total of 100 measurements. The mode-conditioned estimates and covariances of the local IMM-PDA filters are initialized using standard two-point initiation from the first two local measurements. The initial mode probabilities are selected to be equal in both IMM-PDA filters. Throughout the simulations we give detailed results for the communication period $N = 3$. We are going to illustrate the summary of the results for the cases $N = 1, 2, 3, 4, 5$.

We compare the performances of the following algorithms on the same scenario.

- Baseline Methods:
  - Local Agent: This is the local IMM-PDA filter of the agent $\mathcal{F}_1$.
  - Centralized IMM-PDA Filter (Centralized): This is a centralized IMM-PDA filter using the measurements of both $\mathcal{F}_1$ and $\mathcal{F}_2$.
  - Restarted Local Filter (Restarted LF): In this method, the decorrelated quantities in $\mathcal{F}_1$ are calculated by running an auxiliary (shadow) IMM-PDA filter which is restarted after every information transmission (in addition to the primary/principal IMM-PDA filter which is not restarted). The auxiliary IMM-PDA filter has the same parameters as the primary IMM-PDA filter and it is initiated after each data transmission using equal initial mode probabilities. The initial mode-conditioned state estimates of the auxiliary IMM-PDA filter are set to those of the primary IMM-PDA filter at the same time instant (the time instant of the data transmission). The initial mode-conditioned covariances of the auxiliary IMM-PDA are set to $10^{14}\mathbf{I}_4$. This initial mode-conditioned covariance selection makes the values of initial mode-conditioned state estimates unimportant. The auxiliary IMM-PDA filter is run with the local gated measurements of

$\mathcal{F}_1$ until the next transmission time and the output estimate and covariance at the next transmission time is transmitted to $\mathcal{F}_2$. The main steps of Restarted LF for the IMM-PDA filter are given below;

Suppose that the last information transmission happened at time $k-N$ and the next information transmission will be at time k. Then we apply the following steps in the auxiliary IMM-PDA filter.

* Initialization: Initialize the auxiliary IMM filter at time $k-N+1$ with independent information.

$$\mu^{r_k}_{k-N+1|k-N+1} = 1/N_r, \tag{5.39a}$$

$$\hat{\mathbf{x}}^{r_k}_{k-N+1|k-N+1} = \hat{\mathbf{x}}^{r_k}_{k-N|k-N}, \tag{5.39b}$$

$$\mathbf{P}^{r_k}_{k-N+1|k-N+1} = 10^{14}\mathbf{I}_n, \tag{5.39c}$$

where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$.

* Recursion: Update the IMM-PDA filter recursively in time using the measurements $\mathbf{Y}_{k-N+1:k}$ and obtain the weights $\{\mu^i_{k|k}\}_{i=1}^{N_r}$, estimates $\{\hat{\mathbf{x}}^i_{k|k}\}_{i=1}^{N_r}$ and covariances $\{\mathbf{P}^i_{k|k}\}_{i=1}^{N_r}$.
* Merging: Merge the mode-conditioned statistics.

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_r} \mu^i_k \hat{\mathbf{x}}^i_k, \tag{5.40a}$$

$$\mathbf{P}_k = \sum_{i=1}^{N_r} \mu^i_k \left[\mathbf{P}^i_k + (\hat{\mathbf{x}}^i_k - \hat{\mathbf{x}}_k)(\cdot)^{\mathrm{T}}\right]. \tag{5.40b}$$

* Transmission: Transmit merged estimate $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ to the fusion node.

- Proposed Method:
  - OLD: This is the OLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.
  - MLD: This is the MLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.
  - HLD: This is the HLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems.

RMS position errors, RMS velocity errors and mean NEES values obtained from 10000 MC runs are shown in Figure 5.5a, Figure 5.5b and Figure 5.5c, respectively. As expected we see that the centralized approach yields the best RMS error performance. In terms of RMS position and velocity errors, both OLD, MLD and HLD have very similar performances. The RMS error results for Restarted LF are slightly worse than the other fusion methods which have very similar performance. It is striking result that Restarted LF has similar performance with the local agent.

OLD has slightly worse NEES performance compared with the other decorrelation approaches MLD and HLD. This confirms that the effects of the approximations are more significant in OLD. In terms of RMS position and velocity errors, MLD has slightly better performance compared to HLD. However, HLD is more consistent than OLD as seen in Figure 5.5c. We see, on the other hand, that the Restarted LF has lower NEES values than all decorrelation approaches and it behaves same with local agent. In addition, it is striking result that Restarted LF seems to have lower NEES values than the centralized filter. This is due to the fact that the effects of the hypotheses that measurements do not belong to the real target within the validation region are more dominant in Restarted LF/local agent filter compared to OLD/HLD. RMS erros of Restarted LF is higher than the other filters (i.e. centralized filter and all proposed decorrelation approaches) so the number of false alarms falling inside the validatin region might be higher in Restarted LF. Since in the IMM-PDA filter, as all hypotheses are merged according to their hypothesis weights to calculate the posterior, the covariance estimates of the Restarted LF will result in lower NEES values (i.e. underconfidence).

We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 10000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 5.6a, Figure 5.6b, and Figure 5.6c, respectively. Here, the averages of the RMS errors and mean NEES values are taken over the time instants in the interval $[20\,\mathrm{s},\ 100\,\mathrm{s}]$ to avoid the initial transient period. It is seen that, for $N = 1$ there are no approximations in MLD other than merging the mode-conditioned decorrelated estimates and in HLD other than merging the mode-hypothesis-conditioned decorrelated estimates. As a result MLD and HLD give almost the same results as the centralized filter for $N = 1$. Restarted LF, on the other hand, suffers from high

RMS errors even for $N = 1$. For $N \geq 2$ approximations involved in MLD and HLD make their results differ from those of the centralized filter as expected. For $N = 1$, OLD is slightly worse than MLD and HLD in RMS errors. On the other hand for $N \geq 2$, RMS errors of MLD are slightly better than those of OLD and HLD. For all N, the local agent and Restarted LF show similar performances in both RMS errors and NEES values. In terms of consistency OLD is getting worse than MLD and HLD when the decorrelated data transmission period increases. For all N the closest results to the centralized fusion results belong to the HLD method in terms of NEES values. For all N we see that Restarted LF has the worst RMS error performance despite it has low NEES values and the transmission period does not affect its performance significantly.

5.6.2 Model-Mismatch Case

In this case we use the fixed true 2D target trajectory shown in Figure 3.12 where the target makes a coordinated-turn maneuver in the middle of the scenario. The fusion architecture is kept the same as the model-match case and the sensors collect position measurements with measurement noise standard deviation $\sigma_y = 20\,m$. The IMM-PDA filters of the nodes use the parameters of the JMLS given in the model-match case. Hence, in this scenario, we have a mismatch between the true target motion and the filter parameters. We make 1000 MC runs by changing the measurement noise realizations in each run. The reason why the number of MC runs in the model-mismatch case is less than the number of MC cases in the model-match case is that the MC results in the case of model-mismatch of different approaches are smooth enough to be easily compared among each other. We show the resulting RMS position errors, RMS velocity errors and mean NEES values in Figure 5.7a, Figure 5.7b and Figure 5.7c, respectively.

Note that in the model-mismatch case the closeness of the performance measures to those of centralized IMM-PDA filter is important. The most important deviations of the performance measures of MLD and HLD from those of the centralized IMM-PDA filter are in the mean NEES performance in Figure 5.7b. Performance of MLD/HLD is better than that of OLD in the maneuvering parts of the scenario. This surpris-

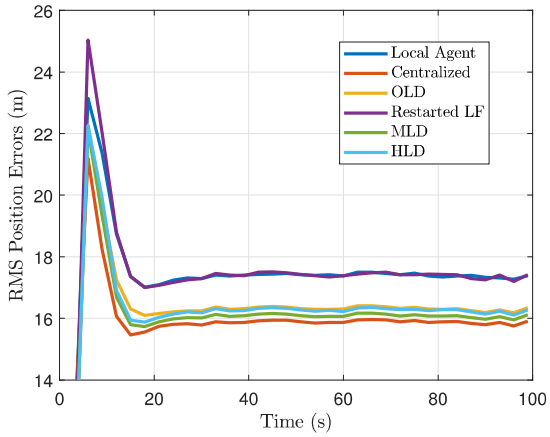
ing phenomenon can actually be explained as follows. When the target moves with constant velocity, there is no process noise. On the other hand, IMM-PDA filters use non-zero process noise standard deviations making them suboptimal. OLD can compensate for most of the effects of the correlations caused by previous communications but it cannot do this perfectly as can be seen in Figure 5.7c. The RMS errors performance of Restarted LF is worse than that of the centralized IMM-PDA filter in the maneuvering parts of the scenario. However, in the maneuvering parts of the scenario, Restarted LF has good mean NEES values compared to other decorrelation approaches (i.e. OLD). In the non-maneuvering parts of the scenario, Restarted LF shows under-confident behaviour compared to the centralized IMM-PDA filter but in these parts of the scenario it has satisfactory RMS errors performance similar to the other decorrelation approaches.

We show the average RMS position errors, average RMS velocity errors and average mean NEES values obtained from 1000 MC runs for the communication periods $N = 1, 2, 3, 4, 5$ in Figure 5.8a, Figure 5.8b, and Figure 5.8c, respectively. Here, the averages of the RMS errors and mean NEES values are taken again over the time instants in the interval $[20\,\mathrm{s},\ 150\,\mathrm{s}]$ to avoid the initial transient period. It is seen that all results (RMS error and average NEES results) for MLD and HLD are the nearest to the centralized filter results for all N. It is seen that Restarted LF has a similar behavior to local agent for all N. This shows that the Restarted LF is not preferable for the IMM-PDA filter. It is significant result that MLD and HLD are immune to the process noise effects for all N. OLD has better results compared to Restarted LF. However, average NEES values of OLD show that OLD is not a consistent decorrelation approach for IMM-PDA filter. HLD is the decorrelation method yielding the closest results to the centralized filter for high communication periods in both RMS errors and average NEES for the scenario with the model-mismatch.

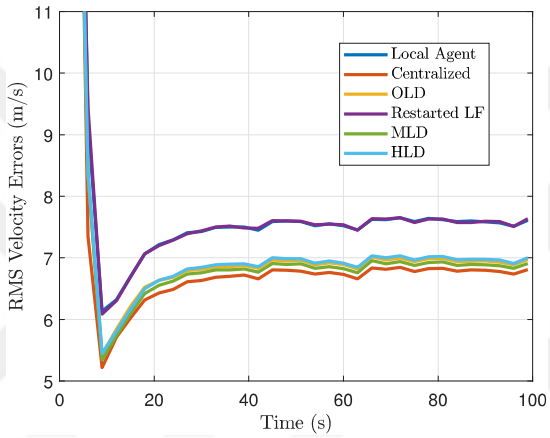
5.7 Conclusion

We have investigated a decorrelation approach for the IMM-PDA filter. In our investigations we examined the issues that might be encountered in applying decorrelation in an IMM-PDA filter in detail and applied covariance inflation method.

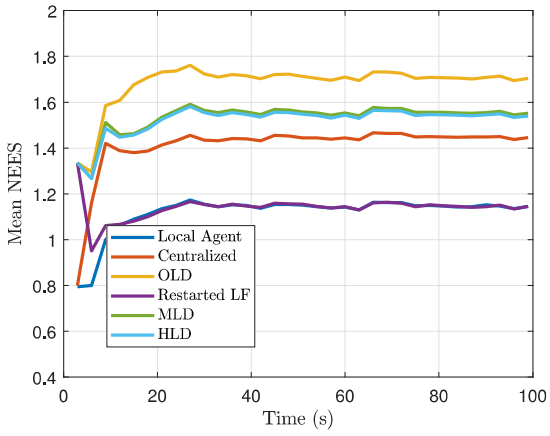
We have seen that the OLD, MLD and HLD approach with covariance inflation applied for both incompatible covariance and outlier decorrelation problems have the closest RMS error to the centralized filter. We have observed also that Restarted LF does not work for the IMM-PDA filter since it gives similar results with the local agent filter. As a result, the MLD can be used for low or moderate process noise levels and/or low communication periods. Furthermore, the HLD can be used for low or moderate process noise levels and/or high communication periods.



(a) RMS Position Errors

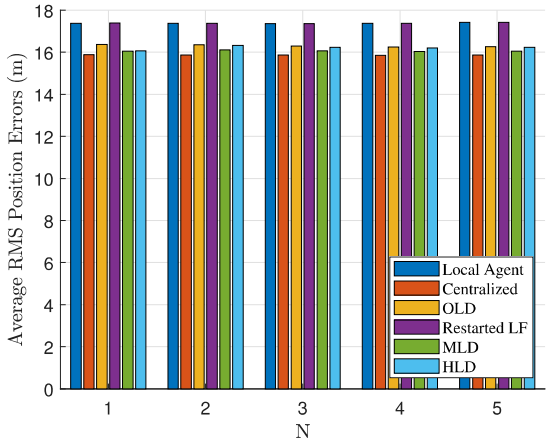


(b) RMS Velocity Errors

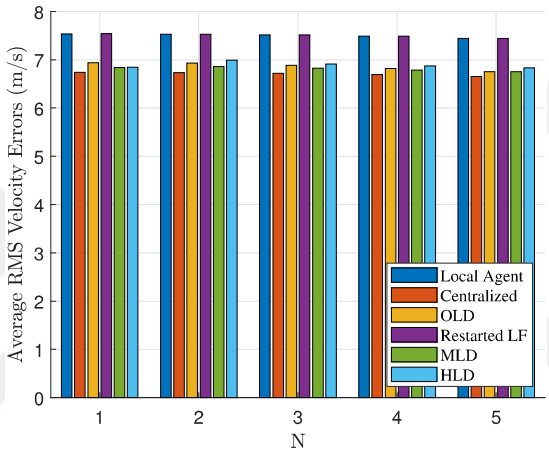


(c) Mean NEES

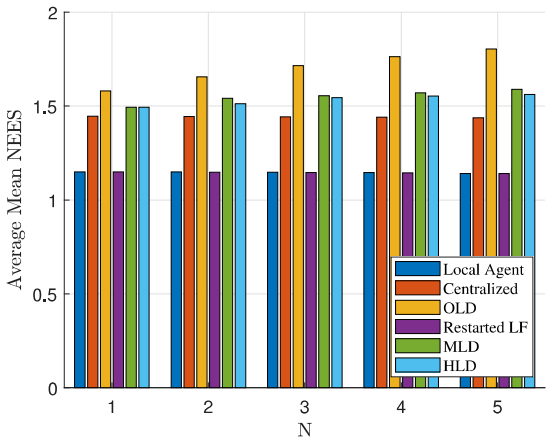
Figure 5.5: IMM-PDA Filter - Model Match Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors

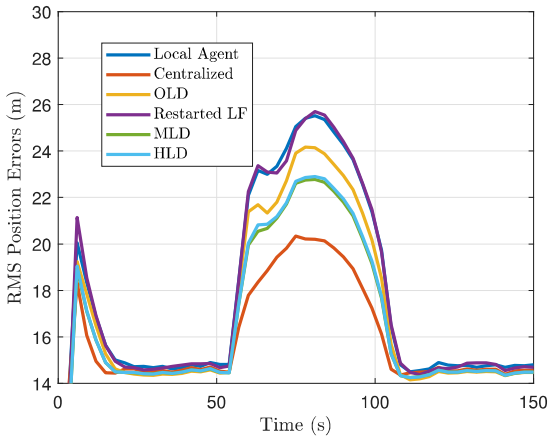


(b) Average RMS Velocity Errors

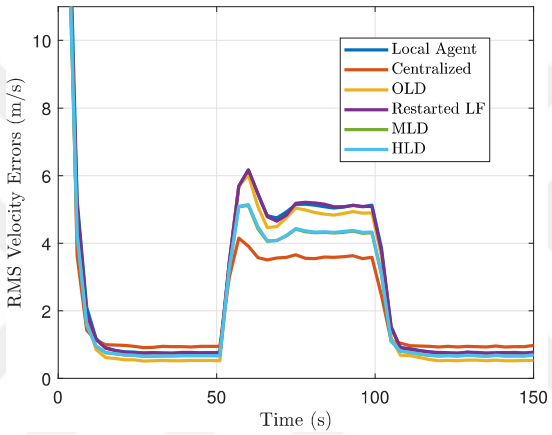


(c) Average Mean NEES

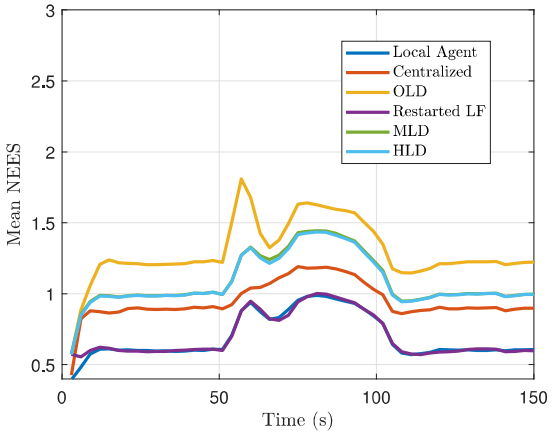
Figure 5.6: IMM-PDA Filter - Model Match Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$.



(a) RMS Position Errors

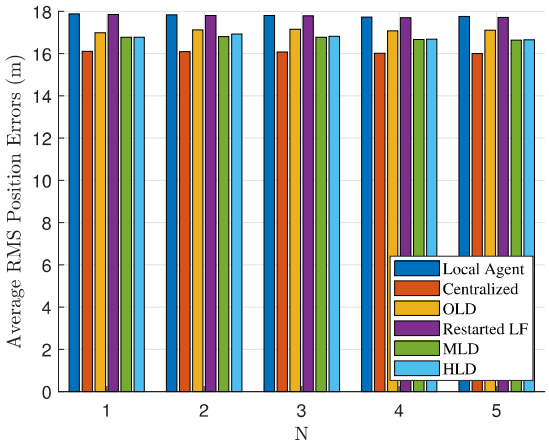


(b) RMS Velocity Errors

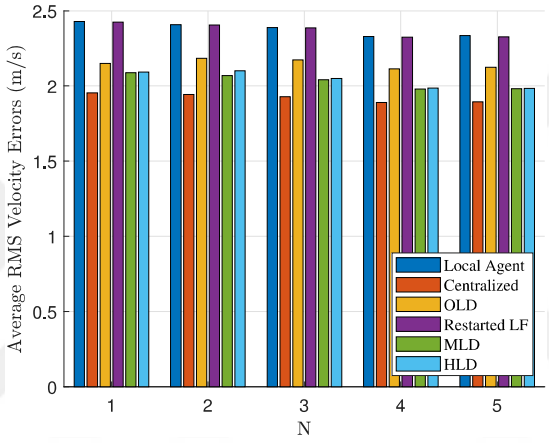


(c) Mean NEES

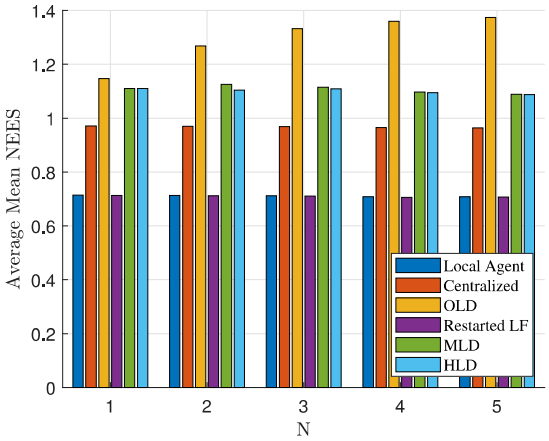
Figure 5.7: IMM-PDA Filter - Model Mismatch Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 3$.



(a) Average RMS Position Errors



(b) Average RMS Velocity Errors



(c) Average Mean NEES

Figure 5.8: IMM-PDA Filter - Model Mismatch Case - Comparison of performances of OLD, MLD, HLD and alternative approaches for $N = 1, 2, 3, 4, 5$.

CHAPTER 6

CONCLUSIONS

In this thesis, the problem of information decorrelation for compensating for the correlations caused by previous communication between two local agents in a distributed estimation framework where multiple model filters (i.e. IMM filter, PDA filter and IMM-PDA filter) are used in the local and fusion nodes was investigated.

First, a decorrelation approach for the IMM filter has been investigated. In the investigations, the issues that might be encountered in applying decorrelation in an IMM filter are examined in detail and a method is proposed, called covariance inflation, for alleviating the effects of both incompatible covariance and outlier decorrelation problems. According to the simulation results, it is seen that the MLD approach with covariance inflation has the closest RMS error and NEES performance results to the centralized filter. In addition, it is a significant observation that MLD has better consistency properties than the restarted local filter. As a result, the MLD seems to be a promising candidate for decorrelation in an IMM filter for low or moderate process noise levels and/or short communication periods.

Second, the proposed information decorrelation methods have been applied on the PDA filter. It is seen that the OLD and HLD approaches with covariance inflation applied for both incompatible covariance and outlier decorrelation problems have the closest RMS errors to the centralized filter. Moreover, HLD has better consistency properties than the other fusion methods. It is a striking result that Restarted LF gives similar results with the local agent, i.e., the filter in which the fusion process is not done. As a result, HLD seems to be a promising candidate for decorrelation for a PDA filter for low or moderate process noise levels and/or high communication periods.

At last, the proposed information decorrelation methods have been applied to the IMM-PDA filter. Similar to the earlier cases, the OLD, MLD and HLD approaches with covariance inflation applied for both incompatible covariance and outlier decorrelation problems have the closest RMS errors to the centralized filter. It has been furthermore observed that Restarted LF does not work well as an information decorrelation approach for the IMM-PDA filter since it gives similar results with the local agent filter as in the case of PDA filter. As a result, the MLD seems to be a promising candidate for decorrelation for an IMM-PDA filter for low or moderate process noise levels and/or short communication periods and HLD can be used for low or moderate process noise levels and/or high communication periods.

6.1 Future Work

This thesis study will shed light on the realization of the future works listed below.

- Investigations were carried out only on simple single target tracking scenarios. The approaches proposed in this thesis can be generalized and applied to multiple target tracking scenarios.

- The problem that motivated this thesis study has been solved for cases where multiple model filters IMM, PDA and IMM-PDA are used. In addition to these filters, proposed approaches can be applied on other multiple model filters in the literature like JPDA filter, GM-PHD filter etc. The basic idea is an application of the proposed information decorrelation approaches at more than one level, starting from the output level and going to the innermost level possible with different communication periods with the aim of recognition the common information and removing it. Applying the proposed methods at inner levels avoids the merging approximations that the multiple model filters inherently make.

- In case where PDA filter is used, a generic false alarm rate is used. By varying the amount of clutter (for example, low-level or dense) in cluttered environment, the performance of methods proposed in this thesis can be measured.

- Only constant velocity models are used as a target motion model due to Assumptions 1 and 2. As a further approximation the methods proposed in this thesis can be applied by using additional target motion models (e.g. acceleration model, coordinated-turn model) in case the multiple model filter used is IMM or IMM-PDA filters, and their performance can be examined.

- A decentralized fusion system having multiple sensor fusion nodes using local IMM, PDA or a combination of these, IMM-PDA filters is designed. By adding a feedback mechanism to the fusion architecture used, it can be ensured that the local node receives the information obtained as a result of fusion from the fusion node. Thus, the sensor node will have acquired additional information about the state of the target and new information decorrelation methods can be developed for the fusion architectures with feedback.

REFERENCES

[1] Y. Bar-Shalom, "On the track-to-track correlation problem," *IEEE Trans. Autom. Control*, vol. 26, no. 2, pp. 571–572, Apr. 1981.

[2] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 803–805, Nov. 1986.

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

[4] M. E. Liggins, C.-Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proc. IEEE*, vol. 85, no. 1, pp. 95–107, Jan. 1997.

[5] W. H. B. Chee-Yee Chong, S. Mori and K.-C. Chang, "Architectures and algorithms for track association and fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 15, no. 1, pp. 5–13, Jan. 2000.

[6] O. E. Drummond, "Feedback in track fusion without process noise," in *Signal and Data Processing of Small Targets 1995*, O. E. Drummond, Ed., vol. 2561, International Society for Optics and Photonics. SPIE, 1995, pp. 369–383.

[7] O. E. Drummond, "On track and tracklet fusion filtering," in *Signal and Data Processing of Small Targets 2002*, O. E. Drummond, Ed., vol. 4728, International Society for Optics and Photonics. SPIE, 2002, pp. 176–195.

[8] C.-Y. Chong, "Hierarchical estimation," in *Proceedings of 2nd MIT/ONR Workshop on Distributed Information and Decision Systems Motivated by Naval Command Control and Communication (C3) Problems*, Jul. 1979, pp. 205–220.

[9] K. C. Chang, T. Zhi, and R. K. Saha, "Performance evaluation of track fusion with information matrix filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 455–466, Apr. 2002.

[10] S. Grime and H. Durrant-Whyte, "Data fusion in decentralized sensor networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[11] K. Chang, C.-Y. Chong, and S. Mori, "Analytical and computational evaluation of scalable distributed fusion algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 4, pp. 2022–2034, Oct. 2010.

[12] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the American Control Conference*, vol. 4, Jun. 1997, pp. 2369–2373.

[13] S. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection (CI)," in *Handbook of Multisensor Data Fusion*, D. L. Hall and J. Llinas, Eds. Boca Raton, FL: CRC Press LLC, 2001, ch. 12.

[14] J. Sijs and M. Lazar, "State fusion with unknown correlation: Ellipsoidal intersection," *Automatica*, vol. 48, no. 8, pp. 1874–1878, 2012.

[15] B. Noack, J. Sijs, M. Reinhardt, and U. D. Hanebeck, "Decentralized data fusion with inverse covariance intersection," *Automatica*, vol. 79, pp. 35–41, 2017.

[16] M. Liggins and K. Chang, *Distributed Fusion Architectures, Algorithms, and Performance within a Network-Centric Architecture*, Sep. 2008, pp. 411–435.

[17] D. Acar and U. Orguner, "Information decorrelation for an interacting multiple model filter," in *Proceedings of the 21st International Conference on Information Fusion*, Jul. 2018.

[18] D. Acar and U. Orguner, "Decorrelation of previously communicated information for an interacting multiple model filter," *IEEE Transactions on Aerospace and Electronic Systems*, Aug. 2020.

[19] A. Benaskeur, "Consistent fusion of correlated data sources," in *Proceedings of the 28th Annual Conference of the Industrial Electronics Society (IECON 02)*, vol. 4, Nov. 2002, pp. 2652–2656.

[20] A. K. Raz and D. A. DeLaurentis, "Performance evaluation of distributed track-to-track fusion systems," in *Sys., Man and Cybernetics, 2014 IEEE International Conference on*. IEEE, Oct. 2014.

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

[22] K. Chang, C.-Y. Chong, and S. Mori, "On scalable distributed sensor fusion," in *Proceedings of the 11th International Conference on Information Fusion*, Jul. 2008.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.

[24] Y. Zhou and J. Li, "Data fusion of unknown correlations using internal ellipsoidal approximation," in *Proceedings of the 17th IFAC World Congress*, Jul. 2008.

[25] O. E. Drummond, "Hybrid sensor fusion algorithm architecture and tracklets," in *Signal and Data Processing of Small Targets 1997*, O. E. Drummond, Ed., vol. 3163, International Society for Optics and Photonics. SPIE, 1997, pp. 485–502.

[26] O. E. Drummond, "Tracklets and a hybrid fusion with process noise," in *Signal and Data Processing of Small Targets 1997*, O. E. Drummond, Ed., vol. 3163, International Society for Optics and Photonics. SPIE, 1997, pp. 512–524.

[27] H. Li, F. Nashashibi, B. Lefaudeux, and E. Pollard, "Track-to-track fusion using split covariance intersection filter-information matrix filter (scif-imf) for vehicle surrounding environment perception," in *Proceedings of the 16th international IEEE annual conference on Intelligent Trans. Sys.*, Oct. 2013, pp. 1430–1435.

[28] A. Makarenko, A. Brooks, T. Kaupp, and H. Durrant-Whyte, "Decentralized data fusion: A graphical model approach," in *Proceedings of the 12th International Conference on Information Fusion*, Jul. 2009, pp. 545–554.

[29] L. Fong, "Distributed data fusion algorithms for tracking a maneuvering target," in *Proceedings of the 10th International Conference on Information Fusion*, 2007, pp. 1–8.

[30] L. Fong, "Distributed data fusion via hybrid approach," in *IECON 2007 - 33rd Annual Conference of the IEEE Industrial Electonics Society*, 2007, pp. 604–609.

[31] L. Fong, "Adaptive information matrix filtering fusion with nonlinear classifier," in *Proceedings of SCIE Annual Conference 2010*, 2010, pp. 2214–2219.

[32] X. Tian and Y. Bar-Shalom, "On algorithms for asynchronous tract-to-track fusion," in *Proceedings of the 13th International Conference on Information Fusion*, Aug. 2010, pp. 1–8.

[33] B. Duraisamy and T. Schwarz, "Track to track fusion incorporating out of sequence track based on information matrix fusion," in *Proceedings of the 18th International Conference on Intelligent Transportation Sys.*, 2015, pp. 2651–2658.

[34] M. Gunay, U. Orguner, and M. Demirekler, "Chernoff fusion of gaussian mixtures based on sigma-point approximation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2732–2746, Sep. 2016.

[35] R. Mahler, "Optimal/robust distributed data fusion: a unified approach," in *Proceedings of SPIE 4052, Signal Processing, Sensor Fusion, and Target Recognition IX*, 2000, pp. 128–138.

[36] M. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proceedings of the International Conference on Information Fusion*, vol. 1, Jul. 2002, pp. 505–511.

[37] D. Clark, S. Julier, R. Mahler, and B. Ristic, "Robust multi-object sensor fusion with unknown correlations," in *Proceedings of Sensor Signal Processing for Defence (SSPD 2010)*, Sep. 2010.

[38] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using exponential mixture models for suboptimal distributed data fusion," in *Proceedings of IEEE Nonlinear Statistical Signal Processing Workshop*, Sep. 2006, pp. 160–163.

[39] B. Noack, M. Reinhardt, and U. D. Hanebeck, "On nonlinear track-to-track fusion with gaussian mixtures," in *Proceedings of the 17th International Conference on Information Fusion*, Jul. 2014.

[40] M. Uney, D. Clark, and S. Julier, "Information measures in distributed multitarget tracking," in *Proceedings of the 14th International Conference on Information Fusion*, Jul. 2011.

[41] T. Bailey, S. Julier, and G. Agamennoni, "On conservative fusion of information with unknown non-gaussian dependence," in *Proceedings of the International Conference on Information Fusion*, Jul. 2012, pp. 1876–1883.

[42] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.

[43] S. Julier and J. Uhlmann, *General decentralized data fusion with covariance intersection (CI)*, Jun. 2001.

[44] B. Upcroft, L. Ong, S. Kumar, M. Ridley, T. Bailey, S. Sukkarieh, and H. Durrant-Whyte, "Rich probabilistic representations for bearing only decentralised data fusion," in *Proceedings of the 8th International Conference on Information Fusion*, vol. 2, Jul. 2005.

[45] S. Julier, "An empirical study into the use of Chernoff information for robust, distributed fusion of gaussian mixture models," in *Proceedings of the 9th International Conference on Information Fusion*, Jul. 2006, pp. 1–8.

[46] M. Gunay, U. Orguner, and M. Demirekler, "Approximate Chernoff fusion of gaussian mixtures using sigmapoints," in *Proceedings of the 17th International Conference on Information Fusion*, Jul. 2014.

[47] D. Smith and S. Singh, "Approaches to multisensory data fusion in target tracking: A survey," vol. 18, no. 12, Dec. 2006.

[48] B. Noack, J. Sijs, and U. Hanebeck, "Inverse covariance intersection: New insights and properties," Jul. 2017.

[49] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[50] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[51] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.

[52] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings on Radar and Signal Processing*, vol. 140, Apr. 1993, pp. 107–113.

[53] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. ArtechHouse, Norwood, 2011.

[54] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and Data Fusion: Handbook of Algorithms*. YBS Publishing, 2011.

[55] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, Sep. 1975.

[56] Y. Bar-Shalom and K. Birmiwal, "Consistency and robustness of PDAF for target tracking in cluttered environments."

[57] H. A. and B.-S. Y., "Multisensor tracking of a maneuvering target in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 2, pp. 176–189, 1989.

[58] H. A. P. 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. FUSION 2002.*, vol. 1, Jul. 2002, pp. 705–712 vol.1.

[59] H. A. P. Blom and E. A. Bloem, "Interacting multiple model joint probabilistic data association avoiding track coalescence," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 3, Dec. 2002, pp. 3408–3415.

[60] H. Blom and E. A. Bloem, "Joint particle filtering of multiple maneuvering targets from unassociated measurements," *J. Adv. Inf. Fusion*, vol. 1, pp. 15–34, 2006.

[61] B. Chen and J. K. Tugnait, "Tracking of multiple maneuvering targets in clutter using IMM/JPDA filtering and fixed-lag smoothing," *Automatica*, vol. 37, no. 2, pp. 239 – 249, 2001.

[62] H. A. P. Blom and E. A. Bloem, "Joint IMMPDA particle filter," in *Proceedings*

of the 6th International Conference on Information Fusion. FUSION 2003., Jul. 2003, pp. 785– 792.

[63] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, Norwood, 2007.

[64] X. R. Li and V. P. 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, Oct. 2005.

[65] C. Simpkins, "System identification: Theory for the user, 2nd edition (ljung, l.; 1999) [on the shelf]," *Robotics and Automation Magazine, IEEE*, vol. 19, pp. 95–96, Jun. 2012.

[66] N. R. Ahmed, "What's one mixture divided by another?: A unified approach to high-fidelity distributed data fusion with mixture models," in *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sep. 2015, pp. 289–296.

[67] R. Visina, Y. Bar-Shalom, P. Willett, and D. Dey, "On-demand track-to-track fusion using local IMM inside information," in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII*, vol. 11018, International Society for Optics and Photonics. SPIE, 2019, pp. 22–32.

[68] M. Mallick and B. L. Scala, "Comparison of single-point and two-point difference track initiation algorithms using position measurements," *Acta Automatica Sinica*, vol. 34, no. 3, pp. 1874–1029, Mar. 2008.

Appendix A

DECORRELATION THEORY

We first consider the case of a linear Gaussian system with no process noise. We believe that the result for linear Gaussian system we present is known in the literature to our best knowledge but we still give it here for pedagogical reasons. Then we obtain a similar result for a family of jump Markov linear systems with mode independent state transition matrix.

A.1 Decorrelation for a Linear Gaussian System

We consider two local agents $\mathcal{F}_1$ and $\mathcal{F}_2$ in a sensor network which achieves the state estimation of the following linear Gaussian system.

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1}, \tag{A.1a}$$

$$\mathbf{y}_k^s = \mathbf{C}^s\mathbf{x}_k + \mathbf{v}_k^s, \tag{A.1b}$$

where

- the superscript $s \in \{1, 2\}$ denotes the local agent index;
- $\mathbf{x}_k \in \mathbb{R}^{n_x}$, denotes the state at time k with the initial distribution $\mathbf{x}_0 \sim p(\mathbf{x}_0) \triangleq \mathcal{N}(\mathbf{x}_0, \bar{\mathbf{x}}_0, \mathbf{P}_0)$;
- $\mathbf{y}_k^s \in \mathbb{R}^{n_y}$ denotes the noisy measurement of $\mathcal{F}_s$, $s = 1, 2$, at time k;
- $\mathbf{v}_k^s \sim \mathcal{N}(\mathbf{v}_k^s; \mathbf{0}, \mathbf{R}^s)$ denotes the i.i.d. measurement noise of $\mathcal{F}_s$, $s = 1, 2$. The measurement noise sequences $\{\mathbf{v}_k^1\}_{k=0}^{\infty}$ and $\{\mathbf{v}_k^2\}_{k=0}^{\infty}$ of $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, are assumed to be independent.

- The state transition matrix $\mathbf{A} \in \mathbb{R}^{n_x \times n_x}$ is invertible.

With this information, we have the following state transition density and measurement likelihoods for this system.

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}), \tag{A.2a}$$

$$p(\mathbf{y}_k^s|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k^s; \mathbf{C}^s\mathbf{x}_k, \mathbf{R}^s), \tag{A.2b}$$

for $s = 1, 2$, where $\delta(\cdot)$ denotes Dirac delta function over the state space $\mathbb{R}^{n_x}$.

Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ run their own local Kalman filters to calculate the local posterior distributions denoted as $p^1_{k|k}(\mathbf{x}_k|\mathbf{y}^1_{0:k})$ and $p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k})$, respectively. We assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ do not know the true distribution $p(\mathbf{x}_0)$ of $\mathbf{x}_0$ and have no prior information about $\mathbf{x}_0$ whatsoever. As a result, $\mathcal{F}_1$ and $\mathcal{F}_2$ initiate their Kalman filter statistics from a number of local measurements and measurement covariances. Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ uses the measurements $\mathbf{y}^1_{0:k_0}$ and $\mathbf{y}^2_{0:k_0}$, respectively, to initialize their Kalman filter estimates.[1] This procedure is equivalent to $\mathcal{F}_1$ and $\mathcal{F}_2$ assuming a diffuse (non-informative) prior $q(\mathbf{x}_{k_0})$ to calculate their initial posteriors $p^s_{k_0|k_0}(\mathbf{x}_{k_0}|\mathbf{y}^s_{0:k_0})$ as follows.[2]

$$p^s_{k_0|k_0}(\mathbf{x}_{k_0}|\mathbf{y}^s_{0:k_0}) = \mathcal{N}(\mathbf{x}_{k_0}; \hat{\mathbf{x}}_{k_0|k_0}, \mathbf{P}_{k_0|k_0}) \propto p(\mathbf{y}^s_{0:k_0}|\mathbf{x}_{k_0})q(\mathbf{x}_{k_0}) \propto p(\mathbf{y}^s_{0:k_0}|\mathbf{x}_{k_0}), \tag{A.3}$$

where we assume that k_0 is sufficiently high so that the likelihood $p(\mathbf{y}^s_{0:k_0}|\mathbf{x}_{k_0})$ can be normalized with respect to $\mathbf{x}_{k_0}$. A very common instance of such a filter initialization is the so-called *two-point-initialization* in target tracking where the state $\mathbf{x}_k$ is composed of position and velocity components and position measurements are collected [68].[3]

[1] Generalization to the case of $\mathcal{F}_1$ and $\mathcal{F}_2$ using different numbers of local measurements to initialize their local Kalman filters is straightforward.

[2] One can intuitively think of a non-informative prior as a Gaussian distribution over the state-space with infinite covariance.

[3] In the case of two-point-initialization, we have $k_0 = 1$, i.e., we have $k_0 + 1 = 2$ measurements, namely, $\mathbf{y}^s_0$ and $\mathbf{y}^s_1$. The likelihood $p(\mathbf{y}^s_{0:1}|\mathbf{x}_1)$ is then normalizable with respect to $\mathbf{x}_1$ and the normalized likelihood gives the following initial posterior state distribution $p^s_{1|1}(\mathbf{x}_1|\mathbf{y}^s_{0:1})$.

$$p^s_{1|1}(\mathbf{x}_1|\mathbf{y}^s_{0:1}) = \mathcal{N}\left(\mathbf{x}_1; \begin{bmatrix} \mathbf{y}^s_1 \\ \frac{\mathbf{y}^s_1 - \mathbf{y}^s_0}{T} \end{bmatrix}, \begin{bmatrix} 1 & \frac{1}{T} \\ \frac{1}{T} & \frac{1}{T^2} \end{bmatrix} \otimes \mathbf{R}^s\right) \tag{A.4}$$

where T denotes the sampling time and $\otimes$ denotes the Kronecker product.

After initialization, $\mathcal{F}_1$ and $\mathcal{F}_2$ continue to run their Kalman filters. At each time step the time update is carried out with Chapman-Kolmogorov equation as follows.

$$p^s_{k|k-1}(\mathbf{x}_k|\mathbf{y}^s_{0:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{y}^s_{0:k-1}) \, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.5a}$$

$$= \int \delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{y}^s_{0:k-1}) \, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.5b}$$

$$= \frac{1}{|\mathbf{A}|} \int \delta(\mathbf{A}^{-1}\mathbf{x}_k - \mathbf{x}_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{y}^s_{0:k-1}) \, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.5c}$$

$$= \frac{1}{|\mathbf{A}|} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k|\mathbf{y}^s_{0:k-1}), \tag{A.5d}$$

where we used the property of the Dirac delta function given as

$$\delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}) = \frac{1}{|\mathbf{A}|} \delta(\mathbf{A}^{-1}\mathbf{x}_k - \mathbf{x}_{k-1}), \tag{A.6}$$

for writing the equality in (A.5c). Since we have $p^s_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{y}^s_{0:k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^s_{k-1|k-1}, \mathbf{P}^s_{k-1|k-1})$, the predicted distribution $p^s_{k|k-1}(\mathbf{x}_k|\mathbf{y}^s_{0:k-1})$ in (A.5d) is given as

$$p^s_{k|k-1}(\mathbf{x}_k|\mathbf{y}^s_{0:k-1}) = \frac{1}{|\mathbf{A}|} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k|\mathbf{y}^s_{0:k-1}), \tag{A.7a}$$

$$= \mathcal{N}(\mathbf{x}_k; \mathbf{A}\hat{\mathbf{x}}^s_{k-1|k-1}, \mathbf{A}\mathbf{P}^s_{k-1|k-1}\mathbf{A}^{\mathrm{T}}), \tag{A.7b}$$

$$\triangleq \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}^s_{k|k-1}, \mathbf{P}^s_{k|k-1}), \tag{A.7c}$$

where $\hat{\mathbf{x}}^s_{k|k-1} \triangleq \mathbf{A}\hat{\mathbf{x}}^s_{k-1|k-1}$ and $\mathbf{P}^s_{k|k-1} \triangleq \mathbf{A}\mathbf{P}^s_{k-1|k-1}\mathbf{A}^{\mathrm{T}}$. The measurement update is made using the Bayes rule as follows.

$$p^s_{k|k}(\mathbf{x}_k|\mathbf{y}^s_{0:k}) \propto p(\mathbf{y}^s_k|\mathbf{x}_k) p^s_{k|k-1}(\mathbf{x}_k|\mathbf{y}^s_{0:k-1}), \tag{A.8}$$

which gives a Gaussian distribution whose mean $\hat{\mathbf{x}}^s_{k|k}$ and covariance $\mathbf{P}^s_{k|k}$ are given by the Kalman filter update equations for the predicted mean $\hat{\mathbf{x}}^s_{k|k-1}$ and covariance $\mathbf{P}^s_{k|k-1}$ with the measurement $\mathbf{y}^s_k$.

Suppose now that $\mathcal{F}_1$ sends its posterior distribution to $\mathcal{F}_2$ for the first time at time $\ell \geq k_0$, i.e., $p^1(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell})$ is transmitted to $\mathcal{F}_2$. Since

- there is no communication between the local agents before time ℓ;
- the system (A.1) does not have any process noise;
- the filters are not initialized with common priors but with conditionally independent measurements;

the fused distribution, which we denote as $p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})$, at $\mathcal{F}_2$ can be obtained using the so-called naive fusion.

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell}) \propto p^1_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell})p^2_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{y}^2_{0:\ell}), \tag{A.9}$$

which can be calculated to be a Gaussian distribution. From time ℓ and on, $\mathcal{F}_2$ updates the fused distribution $p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})$ using its local Kalman filter. The first time update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$\begin{aligned}
p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell}) &= \int p(\mathbf{x}_{\ell+1}|\mathbf{x}_\ell)p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})\,\mathrm{d}\mathbf{x}_\ell, && \text{(A.10a)}\\
&= \int \delta(\mathbf{x}_{\ell+1}-\mathbf{A}\mathbf{x}_\ell)p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})\,\mathrm{d}\mathbf{x}_\ell, && \text{(A.10b)}\\
&= \frac{1}{|\mathbf{A}|}\int \delta(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}-\mathbf{x}_\ell)p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})\,\mathrm{d}\mathbf{x}_\ell, && \text{(A.10c)}\\
&= \frac{1}{|\mathbf{A}|}p^{f,2}_{\ell|\ell,\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell}), && \text{(A.10d)}\\
&\propto \frac{1}{|\mathbf{A}|}p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell}), && \text{(A.10e)}\\
&\propto |\mathbf{A}|\underbrace{\frac{1}{|\mathbf{A}|}p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})}_{=p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})}\underbrace{\frac{1}{|\mathbf{A}|}p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell})}_{=p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell})}, && \text{(A.10f)}\\
&\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell}). && \text{(A.10g)}
\end{aligned}$$

Hence the locally predicted fused distribution $p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell})$ can be written as the naive fusion of the individual locally predicted distributions $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})$ and $p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell})$. The first measurement update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$\begin{aligned}
p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell+1}) &\propto p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1})p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell}), && \text{(A.11a)}\\
&\propto p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1})p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell}), && \text{(A.11b)}\\
&= p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})\underbrace{p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1})p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell})}_{\propto p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell+1})}, && \text{(A.11c)}\\
&\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell+1}). && \text{(A.11d)}
\end{aligned}$$

Hence the locally updated fused distribution $p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:\ell+1})$ is the naive fusion of the locally predicted distribution $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})$ and the locally updated distribution $p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}|\mathbf{y}^2_{0:\ell+1})$.

Repeating the time and measurement updates given above until time $k \triangleq \ell + N$, one can show that the locally updated fused distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:k})$ is given as

$$p^{f,2}_{k|\ell,k}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:k}) = p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell})p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k}). \tag{A.12}$$

Suppose now that, at time $k \triangleq \ell+N$, $\mathcal{F}_1$ sends its local information, which we denote as $\tilde{p}^1_k(\mathbf{x}_k)$, to $\mathcal{F}_2$ for the second time. We assume here that $\mathcal{F}_2$ would like to fuse the $\tilde{p}^1_k(\mathbf{x}_k)$ with its current fused distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:k})$ using naive fusion to obtain the updated fused distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k|\mathbf{y}^1_{0:k},\mathbf{y}^2_{0:k})$ as follows.

$$p^{f,2}_{k|k,k}(\mathbf{x}_k|\mathbf{y}^1_{0:k},\mathbf{y}^2_{0:k}) \propto \tilde{p}^1_k(\mathbf{x}_k)p^{f,2}_{k|\ell,k}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell},\mathbf{y}^2_{0:k}), \tag{A.13a}$$

$$\propto \tilde{p}^1_k(\mathbf{x}_k)p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell})p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k}). \tag{A.13b}$$

The question at this stage would be what should be the information $\tilde{p}^1_k(\mathbf{x}_k)$ that is transmitted by $\mathcal{F}_1$ to satisfy (A.13). We are going to answer this question by thinking about what would happen if the first communication time between $\mathcal{F}_1$ and $\mathcal{F}_2$ was k instead of ℓ. In that case, similar to (A.9), the fused distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k|\mathbf{y}^1_{0:k},\mathbf{y}^2_{0:k})$ would be given as

$$p^{f,2}_{k|k,k}(\mathbf{x}_k|\mathbf{y}^1_{0:k},\mathbf{y}^2_{0:k}) \propto p^1_{k|k}(\mathbf{x}_k|\mathbf{y}^1_{0:k})p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k}). \tag{A.14}$$

Since the left hand sides of (A.13) and (A.14) are equal, the right hand sides should be equal up to a multiplicative constant, i.e.,

$$\tilde{p}^1_k(\mathbf{x}_k)p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell})p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k}) \propto p^1_{k|k}(\mathbf{x}_k|\mathbf{y}^1_{0:k})p^2_{k|k}(\mathbf{x}_k|\mathbf{y}^2_{0:k}), \tag{A.15}$$

which gives

$$\tilde{p}^1_k(\mathbf{x}_k) \propto \frac{p^1_{k|k}(\mathbf{x}_k|\mathbf{y}^1_{0:k})}{p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell})}. \tag{A.16}$$

Note that in general the function $\tilde{p}^1_k(\mathbf{x}_k)$ might not be normalizable with respect to $\mathbf{x}_k$ and hence cannot be interpreted as a valid probability density function.

Measurement space is a space that we desire to find the measurement from the target of interest. Measurements used in tracking systems not only may have originated

from target, but also from false alarms, clutters. Therefore, the validation region which limits the region in the measurement space is determined. Any measurement that has been validated could have originated from the target of interest. Therefore, association probabilities are calculated for each validated measurement at the current time and this Bayesian information can be used in PDA filter algorithm.

We consider measurements in the gate are shown as $\mathbf{Y}_k^s = \{\mathbf{y}_k^{i,s}\}_{i=1}^{m_k}$. There are the following hypotheses about these measurements,

- $\theta_k^0 =$ All of $\mathbf{Y}_k^s$ is due to false alarm i.e. no target originated measurement in the gate.
- $\theta_k^i =$ Measurement $\mathbf{y}_k^{i,s}$ belongs to target, all the rest are false alarms.

for $i = 1, ..., m_k$. With this information, we have the following measurement likelihoods for this system,

$$p(\mathbf{Y}_k^s|\mathbf{x}_k) = \mathcal{N}(\mathbf{Y}_k^s; \mathbf{C}^s\mathbf{x}_k, \mathbf{R}^s). \tag{A.17}$$

Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ run their own local PDA filters to calculate the local posterior distributions denoted as $p_{k|k}^1(\mathbf{x}_k|\mathbf{Y}_{0:k}^1)$ and $p_{k|k}^2(\mathbf{x}_k|\mathbf{Y}_{0:k}^2)$, respectively. It is assumed that $\mathcal{F}_1$ and $\mathcal{F}_2$ uses the measurements $\mathbf{Y}_{0:k_0}^1$ and $\mathbf{Y}_{0:k_0}^2$, respectively, to initialize their PDA filter estimates. The initial posterior distribution of $\{\mathbf{x}_0\}$ is given as

$$p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0, \mathbf{P}_0). \tag{A.18}$$

After initialization, $\mathcal{F}_1$ and $\mathcal{F}_2$ continue to run their PDA filters. The time update is carried out to obtain the predicted marginal base-state distribution by using Chapman-Kolmogorov equation as follows.

$$p_{k|k-1}^s(\mathbf{x}_k|\mathbf{Y}_{0:k-1}^s) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p_{k-1|k-1}^s(\mathbf{x}_{k-1}|\mathbf{Y}_{0:k-1}^s)\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.19a}$$

$$= \int \delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}) p_{k-1|k-1}^s(\mathbf{x}_{k-1}|\mathbf{Y}_{0:k-1}^s)\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.19b}$$

$$= \frac{1}{|\mathbf{A}|} \int \delta(\mathbf{A}^{-1}\mathbf{x}_k - \mathbf{x}_{k-1}) p_{k-1|k-1}^s(\mathbf{x}_{k-1}|\mathbf{Y}_{0:k-1}^s)\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.19c}$$

$$= \frac{1}{|\mathbf{A}|} p_{k-1|k-1}^s(\mathbf{A}^{-1}\mathbf{x}_k|\mathbf{Y}_{0:k-1}^s). \tag{A.19d}$$

The measurement update is made using the Bayes rule as follows.

$$p^s_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^s_{0:k}) \propto p(\mathbf{Y}^s_k|\mathbf{x}_k, \theta_k) p^s_{k|k-1}(\mathbf{x}_k|\mathbf{Y}^s_{0:k-1}). \tag{A.20}$$

Suppose now that $\mathcal{F}_1$ sends its marginal base-state posterior $p^1_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{Y}^1_{0:\ell})$ to $\mathcal{F}_2$ for the first time at time ℓ. We will assume here that $\mathcal{F}_2$ fuses this information with its joint posterior $p^2_{\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathbf{Y}^2_{0:\ell})$ using the naive fusion as follows.

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}) \propto p^1_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{Y}^1_{0:\ell}) p^2_{\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathbf{Y}^2_{0:\ell}). \tag{A.21}$$

Notice that we call the fused joint posterior as $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$ instead of $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathbf{Y}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$ since we can write that

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}) \neq p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, \theta_\ell|\mathbf{Y}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}). \tag{A.22}$$

$\mathcal{F}_1$ sends only the information related to the base-state $\mathbf{x}_k$ to $\mathcal{F}_2$ and it does not send any information about the hypothesis-state θ_k. As a result, $\mathcal{F}_1$ sends only partial information about $\mathbf{Y}^1_{0:\ell}$, which we show as $\mathcal{I}^1_{0:\ell}$, to $\mathcal{F}_2$.

Therefore the fused joint posterior $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$ can be described as an approximation of the ideally fused joint posterior $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, \theta_\ell|\mathbf{Y}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$.

From time ℓ and on, $\mathcal{F}_2$ merged the fused distribution $p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$ as follows.

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}) = \sum_{\theta_\ell} p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, \theta_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}). \tag{A.23}$$

Then $\mathcal{F}_2$ updates the merged fused distribution using its own local Bayesian filter.

The first time update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}) = \int p(\mathbf{x}_{\ell+1}|\mathbf{x}_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.24a}$$

$$= \int \delta(\mathbf{x}_{\ell+1} - \mathbf{A}\mathbf{x}_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.24b}$$

$$= \frac{1}{|\mathbf{A}|} \int \delta(\mathbf{A}^{-1}\mathbf{x}_{\ell+1} - \mathbf{x}_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.24c}$$

$$= \frac{1}{|\mathbf{A}|} p^{f,2}_{\ell|\ell,\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}), \tag{A.24d}$$

$$\propto \frac{1}{|\mathbf{A}|} p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell}) p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell}), \tag{A.24e}$$

$$\propto |\mathbf{A}| \underbrace{\frac{1}{|\mathbf{A}|} p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell})}_{=p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell})} \underbrace{\frac{1}{|\mathbf{A}|} p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell})}_{=p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell})}, \tag{A.24f}$$

$$\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell}). \tag{A.24g}$$

Hence the locally predicted fused joint distribution $p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell})$ can be described as the naive fusion of the predicted marginal distributions of $\mathcal{F}_1$ $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell})$ and predicted distribution $p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell})$ of $\mathcal{F}_2$. The first measurement update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}, \theta_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell+1}) \propto p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, \theta_{\ell+1}) p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell}),$$

(A.25a)

$$\propto p(\mathbf{Y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, \theta_{\ell+1}) p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell}), \tag{A.25b}$$

$$= p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell}) \underbrace{p(\mathbf{Y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, \theta_{\ell+1}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^2_{0:\ell})}_{\propto p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, \theta_{\ell+1}|\mathbf{Y}^2_{0:\ell+1})}, \tag{A.25c}$$

$$\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell}) p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, \theta_{\ell+1}|\mathbf{Y}^2_{0:\ell+1}). \tag{A.25d}$$

Therefore the locally updated fused distribution $p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}, \theta_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:\ell+1})$ can be described as the naive fusion of the locally predicted distribution $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{Y}^1_{0:\ell})$ and the locally updated joint distribution $p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, \theta_{\ell+1}|\mathbf{Y}^2_{0:\ell+1})$.

Repeating both prediction and measurement updates given above until time $k = \ell + N$, the locally updated fused joint distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:k})$ is given as

$$p^{f,2}_{k|\ell,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:k}) = p^1_{k|\ell}(\mathbf{x}_k|\mathbf{Y}^1_{0:\ell}) p^2_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^2_{0:k}). \tag{A.26}$$

Suppose now that, at time $k = \ell + N$, $\mathcal{F}_1$ sends its local marginal information, as $\tilde{p}^1_k(\mathbf{x}_k)$, to $\mathcal{F}_2$ for the second time. It is assumed that $\mathcal{F}_2$ would like to fuse $\tilde{p}^1_k(\mathbf{x}_k)$ with its current fused joint distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:\ell}, \mathbf{Y}^2_{0:k})$ using naive fusion to get the updated fused joint distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k})$ as follows.

$$p^{f,2}_{k|k,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k}) \propto \tilde{p}_k(\mathbf{x}_k) p^{f,2}_{k|\ell,k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^1_{0:\ell}, \mathbf{Y}^2_{0:k}), \tag{A.27a}$$

$$\propto \tilde{p}^1_k(\mathbf{x}_k) p^1_{k|\ell}(\mathbf{x}_k|\mathbf{Y}^1_{0:\ell}) p^2_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^2_{0:k}). \tag{A.27b}$$

If the first communication time between $\mathcal{F}_1$ and $\mathcal{F}_2$ was k instead of ℓ, similar to (A.21), we would like the fused joint distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k})$ to be given as

$$p^{f,2}_{k|k,k}(\mathbf{x}_k, \theta_k|\mathcal{I}^1_{0:k}, \mathbf{Y}^2_{0:k}) \propto p^1_{k|k}(\mathbf{x}_k|\mathbf{Y}^1_{0:k}) p^2_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^2_{0:k}). \tag{A.28}$$

Since the left hand sides of (A.27) and (A.28) are equal, the right hand sides should be equal up to a multiplicative constant, i.e.,

$$\tilde{p}^1_k(\mathbf{x}_k) p^1_{k|\ell}(\mathbf{x}_k|\mathbf{Y}^1_{0:\ell}) p^2_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^2_{0:k}) \propto p^1_{k|k}(\mathbf{x}_k|\mathbf{Y}^1_{0:k}) p^2_{k|k}(\mathbf{x}_k, \theta_k|\mathbf{Y}^2_{0:k}), \tag{A.29}$$

which gives

$$\tilde{p}^1_k(\mathbf{x}_k) \propto \frac{p^1_{k|k}(\mathbf{x}_k|\mathbf{Y}^1_{0:k})}{p^1_{k|\ell}(\mathbf{x}_k|\mathbf{Y}^1_{0:\ell})}. \tag{A.30}$$

A.2 Decorrelation for a Jump Markov Linear System

We now consider the following jump Markov linear systems.

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1}, \tag{A.31a}$$

$$\mathbf{y}^s_k = \mathbf{C}^s_{r_k}\mathbf{x}_k + \mathbf{D}^s_{r_k}\mathbf{v}^s_{r_k,k}, \tag{A.31b}$$

where

- the superscript $s \in \{1, 2\}$ denotes the local agent index;
- $\mathbf{x}_k \in \mathbb{R}^{n_x}$, denotes the base-state at time k;
- $r_k \in \{1, \ldots, N_r\}$, denotes the mode-state at time k;
- r_k evolves according to a homogeneous Markov chain with transition probability matrix $\mathbf{\Pi} \triangleq [\pi_{ji} \triangleq P(r_k = i|r_{k-1} = j)]$;

- the initial joint distribution of $\mathbf{x}_0$ and r_0 is given as $p(\mathbf{x}_0, r_0) = \pi_0^{r_0} \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0^{r_0}, \mathbf{P}_0^{r_0})$;

- $\mathbf{y}_k^s \in \mathbb{R}^{n_y}$ denotes the noisy measurement of $\mathcal{F}_s$, $s = 1, 2$, at time k;

- $\mathbf{v}_{r_k,k}^s \sim \mathcal{N}(\mathbf{v}_{r_k,k}^s; \mathbf{0}, \mathbf{R}_{r_k}^s)$ denotes the i.i.d. measurement noise sequence of $\mathcal{F}_s$, $s = 1, 2$. The measurement noise sequences $\{\mathbf{v}_{r_k,k}^1\}_{k=0}^{\infty}$ and $\{\mathbf{v}_{r_k,k}^2\}_{k=0}^{\infty}$ of $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, are assumed to be independent;

- The mode-independent state transition matrix $\mathbf{A} \in \mathbb{R}^{n_x \times n_x}$ is an invertible matrix;

Note that the overall state of the JMLS in (2.16) is $\{\mathbf{x}_k, r_k\}$, i.e., the overall state is composed of both the base-state $\mathbf{x}_k$ and the mode-state r_k. With this information, we have the following state transition density and measurement likelihoods for the JMLS in (2.16).

$$p(\mathbf{x}_k, r_k | \mathbf{x}_{k-1}, r_{k-1}) = \pi_{r_{k-1} r_k} \delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}), \tag{A.32a}$$

$$p(\mathbf{y}_k^s | \mathbf{x}_k, r_k) = \mathcal{N}(\mathbf{y}_k^s; \mathbf{C}_{r_k}^s \mathbf{x}_k, \mathbf{D}_{r_k}^s \mathbf{R}_{r_k}^s \mathbf{D}_{r_k}^{s,\mathrm{T}}), \tag{A.32b}$$

for $s = 1, 2$, where $\delta(\cdot)$ denotes Dirac delta function over $\mathbb{R}^{n_x}$. Suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ run their own local Bayesian filters to calculate the local posterior distributions denoted as $p_{k|k}^1(\mathbf{x}_k, r_k | \mathbf{y}_{0:k}^1)$ and $p_{k|k}^2(\mathbf{x}_k, r_k | \mathbf{y}_{0:k}^2)$. Note that these exact posteriors would be computationally intractable to calculate in general as they would be Gaussian mixtures with ever increasing number of components. However, in this part we would not care about this fact and will assume that the local Bayesian filters can calculate these exact posteriors.

We assume that $\mathcal{F}_1$ and $\mathcal{F}_2$ do not know the true distribution $p(\mathbf{x}_0, r_0)$ of $\mathbf{x}_0, r_0$ and have no prior information about $\mathbf{x}_0, r_0$ whatsoever. As a result, $\mathcal{F}_1$ and $\mathcal{F}_2$ initiate their Bayesian filter statistics from a number of local measurements and measurement covariances by using the non-informative prior $q(\mathbf{x}_0, r_0)$. Suppose that $\mathcal{F}_s$ uses its local measurements $\mathbf{y}_{0:k_0}^s$ to initialize its own local filter.

After initialization, $\mathcal{F}_1$ and $\mathcal{F}_2$ continue to run their Bayesian filters. The time update

is carried out with Chapman-Kolmogorov equation as follows.

$$p^s_{k|k-1}(\mathbf{x}_k, r_k|\mathbf{y}^s_{0:k-1}) = \int \sum_{r_{k-1}} p(\mathbf{x}_k, r_k|\mathbf{x}_{k-1}, r_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}, r_{k-1}|\mathbf{y}^s_{0:k-1})\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.33a}$$

$$= \int \sum_{r_{k-1}} \pi_{r_{k-1}r_k} \delta(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}, r_{k-1}|\mathbf{y}^s_{0:k-1})\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.33b}$$

$$= \frac{1}{|\mathbf{A}|} \sum_{r_{k-1}} \pi_{r_{k-1}r_k} \int \delta(\mathbf{A}^{-1}\mathbf{x}_k - \mathbf{x}_{k-1}) p^s_{k-1|k-1}(\mathbf{x}_{k-1}, r_{k-1}|\mathbf{y}^s_{0:k-1})\, \mathrm{d}\mathbf{x}_{k-1}, \tag{A.33c}$$

$$= \frac{1}{|\mathbf{A}|} \sum_{r_{k-1}} \pi_{r_{k-1}r_k} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k, r_{k-1}|\mathbf{y}^s_{0:k-1}). \tag{A.33d}$$

Note that the predicted marginal base-state distribution is now given as

$$p^s_{k|k-1}(\mathbf{x}_k|\mathbf{y}^s_{0:k-1}) = \sum_{r_k} p^s_{k|k-1}(\mathbf{x}_k, r_k|\mathbf{y}^s_{0:k-1}), \tag{A.34a}$$

$$= \sum_{r_k} \frac{1}{|\mathbf{A}|} \sum_{r_{k-1}} \pi_{r_{k-1}r_k} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k, r_{k-1}|\mathbf{y}^s_{0:k-1}), \tag{A.34b}$$

$$= \frac{1}{|\mathbf{A}|} \sum_{r_{k-1}} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k, r_{k-1}|\mathbf{y}^s_{0:k-1}), \tag{A.34c}$$

$$= \frac{1}{|\mathbf{A}|} p^s_{k-1|k-1}(\mathbf{A}^{-1}\mathbf{x}_k|\mathbf{y}^s_{0:k-1}), \tag{A.34d}$$

which is the same result as the one for the linear Gaussian system. This is natural since the base-state is not dependent on the mode-state in the JMLS we consider. The measurement update is made using the Bayes rule as follows.

$$p^s_{k|k}(\mathbf{x}_k, r_k|\mathbf{y}^s_{0:k}) \propto p(\mathbf{y}^s_k|\mathbf{x}_k, r_k) p^s_{k|k-1}(\mathbf{x}_k, r_k|\mathbf{y}^s_{0:k-1}). \tag{A.35}$$

Suppose now that $\mathcal{F}_1$ sends its marginal base-state posterior $p^1_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell})$ to $\mathcal{F}_2$ for the first time at time ℓ. We will assume here that $\mathcal{F}_2$ fuses this information with its joint posterior $p^2_{\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathbf{y}^2_{0:\ell})$ using the naive fusion as follows.

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}) \propto p^1_{\ell|\ell}(\mathbf{x}_\ell|\mathbf{y}^1_{0:\ell}) p^2_{\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathbf{y}^2_{0:\ell}). \tag{A.36}$$

Notice that we call the fused joint posterior as $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$ and not as $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathbf{y}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$ since here we can clearly say that

$$p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}) \neq p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathbf{y}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}). \tag{A.37}$$

The reasons are two fold.

- Due to the stochastic nature of the mode-state, there is still some (implicit) process noise of the system causing correlated errors in the local nodes. Hence the fused posterior cannot be obtained by naive fusion, i.e., by the multiplication of the local posteriors;

- $\mathcal{F}_1$ sends only the information related to the base-state $\mathbf{x}_k$ to $\mathcal{F}_2$ and it does not send any information about the mode-state r_k. As a result, $\mathcal{F}_1$ sends only partial information about $\mathbf{y}^1_{0:\ell}$, which we show as $\mathcal{I}^1_{0:\ell}$, to $\mathcal{F}_2$;

Hence the fused joint posterior $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$ is only an approximation of the ideally fused joint posterior $p^{f,2}_{\ell,\ell|\ell}(\mathbf{x}_\ell, r_\ell|\mathbf{y}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$.

From time ℓ and on, $\mathcal{F}_2$ updates the fused distribution $p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$ using its own local Bayesian filter. The first time update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}) = \int \sum_{r_\ell} p(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{x}_\ell, r_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.38a}$$

$$= \int \sum_{r_\ell} \pi_{r_\ell r_{\ell+1}} \delta(\mathbf{x}_{\ell+1} - \mathbf{A}\mathbf{x}_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.38b}$$

$$= \frac{1}{|\mathbf{A}|} \sum_{r_\ell} \pi_{r_\ell r_{\ell+1}} \int \delta(\mathbf{A}^{-1}\mathbf{x}_{\ell+1} - \mathbf{x}_\ell) p^{f,2}_{\ell|\ell,\ell}(\mathbf{x}_\ell, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})\, \mathrm{d}\mathbf{x}_\ell, \tag{A.38c}$$

$$= \frac{1}{|\mathbf{A}|} \sum_{r_\ell} \pi_{r_\ell r_{\ell+1}} p^{f,2}_{\ell|\ell,\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}, r_\ell|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}), \tag{A.38d}$$

$$\propto \frac{1}{|\mathbf{A}|} \sum_{r_\ell} \pi_{r_\ell r_{\ell+1}} p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell}) p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}, r_\ell|\mathbf{y}^2_{0:\ell}), \tag{A.38e}$$

$$\propto |\mathbf{A}| \underbrace{\frac{1}{|\mathbf{A}|} p^1_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})}_{=p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})} \underbrace{\frac{1}{|\mathbf{A}|} \sum_{r_\ell} \pi_{r_\ell r_{\ell+1}} p^2_{\ell|\ell}(\mathbf{A}^{-1}\mathbf{x}_{\ell+1}, r_\ell|\mathbf{y}^2_{0:\ell})}_{=p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell})}, \tag{A.38f}$$

$$\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell}). \tag{A.38g}$$

Hence the locally predicted fused joint distribution $p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell})$ can be written as the naive fusion of the individual locally predicted marginal distributions $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})$ and locally predicted joint $p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell})$. The first measure-

ment update of $\mathcal{F}_2$ after time ℓ is given as follows.

$$p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell+1}) \propto p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, r_{\ell+1}) p^{f,2}_{\ell+1|\ell,\ell}(\mathbf{x}_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell}),$$

(A.39a)

$$\propto p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, r_{\ell+1}) p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell}), \tag{A.39b}$$

$$= p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell}) \underbrace{p(\mathbf{y}^2_{\ell+1}|\mathbf{x}_{\ell+1}, r_{\ell+1}) p^2_{\ell+1|\ell}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell})}_{\propto p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell+1})}, \tag{A.39c}$$

$$\propto p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell}) p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell+1}). \tag{A.39d}$$

Hence the locally updated fused distribution $p^{f,2}_{\ell+1|\ell,\ell+1}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:\ell+1})$ is the naive fusion of the locally predicted marginal distribution $p^1_{\ell+1|\ell}(\mathbf{x}_{\ell+1}|\mathbf{y}^1_{0:\ell})$ and the locally updated joint distribution $p^2_{\ell+1|\ell+1}(\mathbf{x}_{\ell+1}, r_{\ell+1}|\mathbf{y}^2_{0:\ell+1})$.

Repeating the time and measurement updates given above until time $k = \ell + N$, one can show that the locally updated fused joint distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:k})$ is given as

$$p^{f,2}_{k|\ell,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:k}) = p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell}) p^2_{k|k}(\mathbf{x}_k, r_k|\mathbf{y}^2_{0:k}). \tag{A.40}$$

Suppose now that, at time $k = \ell + N$, $\mathcal{F}_1$ sends its local marginal information, which we denote as $\tilde{p}^1_k(\mathbf{x}_k)$, to $\mathcal{F}_2$ for the second time. We assume here that $\mathcal{F}_2$ would like to fuse $\tilde{p}^1_k(\mathbf{x}_k)$ with its current fused joint distribution $p^{f,2}_{k|\ell,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:\ell}, \mathbf{y}^2_{0:k})$ again using naive fusion to obtain the updated fused joint distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k}, \mathbf{y}^2_{0:k})$ as follows.

$$p^{f,2}_{k|k,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k}, \mathbf{y}^2_{0:k}) \propto \tilde{p}_k(\mathbf{x}_k) p^{f,2}_{k|\ell,k}(\mathbf{x}_k, r_k|\mathbf{y}^1_{0:\ell}, \mathbf{y}^2_{0:k}), \tag{A.41a}$$

$$\propto \tilde{p}^1_k(\mathbf{x}_k) p^1_{k|\ell}(\mathbf{x}_k|\mathbf{y}^1_{0:\ell}) p^2_{k|k}(\mathbf{x}_k, r_k|\mathbf{y}^2_{0:k}). \tag{A.41b}$$

We here again ask what $\tilde{p}^1_k(\mathbf{x}_k)$ that is transmitted by $\mathcal{F}_1$ should be to satisfy (A.41). We again consider what would happen if the first communication time between $\mathcal{F}_1$ and $\mathcal{F}_2$ was k instead of ℓ. In that case, similar to (A.36), we would like the fused joint distribution $p^{f,2}_{k|k,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k}, \mathbf{y}^2_{0:k})$ to be given as

$$p^{f,2}_{k|k,k}(\mathbf{x}_k, r_k|\mathcal{I}^1_{0:k}, \mathbf{y}^2_{0:k}) \propto p^1_{k|k}(\mathbf{x}_k|\mathbf{y}^1_{0:k}) p^2_{k|k}(\mathbf{x}_k, r_k|\mathbf{y}^2_{0:k}). \tag{A.42}$$

Since the left hand sides of (A.41) and (A.42) are equal, the right hand sides should be equal up to a multiplicative constant, i.e.,

$$\tilde{p}_k^1(\mathbf{x}_k)p_{k|\ell}^1(\mathbf{x}_k|\mathbf{y}_{0:\ell}^1)p_{k|k}^2(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^2) \propto p_{k|k}^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1)p_{k|k}^2(\mathbf{x}_k, r_k|\mathbf{y}_{0:k}^2), \tag{A.43}$$

which gives

$$\tilde{p}_k^1(\mathbf{x}_k) \propto \frac{p_{k|k}^1(\mathbf{x}_k|\mathbf{y}_{0:k}^1)}{p_{k|\ell}^1(\mathbf{x}_k|\mathbf{y}_{0:\ell}^1)}. \tag{A.44}$$

Appendix B

COMPUTATIONAL COMPLEXITY

B.1 IMM Filter - Computational Load

In this section we investigate OLD, MLD and covariance inflation in terms of their computational complexity and compare them to a restarted (shadow/secondary) local IMM filter. A restarted local filter (Restarted LF) is an IMM filter with the same parameters as the principal/primary IMM filter which is restarted after each data transmission. Restarted LF makes N IMM prediction updates (with computational load $\mathcal{O}(N_r n_x^3 + N_r^2 n_x^2)$) and N IMM measurement updates (with computational load $\mathcal{O}(N_r(n_x^3 + n_y^3))$ in information form) with a measurement of size $n_y \times 1$. As a result the total computational complexity of Restarted LF for each data transmission is $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2 + N_r(n_x^3 + n_y^3)))$.

OLD and MLD both make N IMM prediction updates similar to Restarted LF. Each decorrelation operation applied after the prediction updates in OLD and MLD can be assumed to have the same computational complexity of a single Kalman filter measurement update with a measurement of dimension $n_x \times 1$ (with computational load $\mathcal{O}(n_x^3)$). After the prediction updates, OLD and MLD apply decorrelation on a single estimate and on N_r estimates, respectively. Hence the total computational complexities of OLD and MLD for each data transmission is $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + n_x^3)$ and $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r n_x^3)$ respectively. As a result OLD and MLD would be advantageous in terms of computational load thanks to less number of measurement updates (i.e., decorrelations) when the data transmission period N is large.

We have to note here that checking the existence of incompatible covariance and outlier decorrelation problems in OLD and MLD has computational complexities

$\mathcal{O}(n_x^3)$ and $\mathcal{O}(N_r n_x^3)$, respectively. Once we have these problems, applying covariance inflation in OLD and MLD has computational complexities $\mathcal{O}(M_{\mathrm{IC/OD}} n_x^3)$ and $\mathcal{O}(N_r M_{\mathrm{IC/OD}} n_x^3)$, respectively.

B.2 PDA Filter - Computational Load

In this section we investigate OLD, HLD and covariance inflation in terms of their computational complexity and compare them to a restarted shadow local PDA filter. Restarted LF is a PDA filter with the same parameters as the principal/primary PDA filter which is restarted after each data transmission. The only difference of shadow PDA filter from the primary PDA filter is that the gating is not applied in the shadow one. The shadow PDA filter uses the same gated measurements with the ones validated in primary PDA filter. Restarted LF makes N PDA prediction updates (with computational load $\mathcal{O}(n_x^3)$) and N PDA measurement updates (with computational load $\mathcal{O}(m_k(n_x^3 + n_y^3))$ in information form) m_k times with a measurement of size $n_y \times 1$ where m_k is the number of hypothesis generated for the time k. It is important to note that m_k changes every time step. Especially in dense cluttered environment, the number of measurements falling within the validation region will be numerous. Therefore, the number of hypotheses produced for that step will also be high causing computational load to increase. As a result the total computational complexity of Restarted LF for each data transmission is $\mathcal{O}(N(n_x^3 + m_k(n_x^3 + n_y^3)))$.

OLD and HLD both make N PDA prediction updates similar to Restarted LF. Each decorrelation operation applied after the prediction updates in OLD and HLD can be assumed to have the same computational complexity of a single Kalman filter measurement update with a measurement of dimension $n_x \times 1$ (with computational load $\mathcal{O}(n_x^3)$). After the prediction updates, OLD and HLD apply decorrelation on a single estimate and on $m_k + 1$ estimates, respectively. In HLD, the zero-hypothesis, which is the hypothesis that none of the measurements falling within the validation region will come from the real target, is also taken into account for decorrelation operation. Hence the total computational complexities of OLD and HLD for each data transmission is $\mathcal{O}(N(n_x^3)+n_x^3)$ and $\mathcal{O}(N(n_x^3)+(m_k+1)n_x^3)$, respectively. Therefore, OLD and HLD would be advantageous in terms of computational load due to the

less number of measurement updates (i.e., decorrelations) when the data transmission period N is large.

In OLD and HLD, checking the existence of incompatible covariance and outlier decorrelation problems has computational complexities $\mathcal{O}(n_x^3)$ and $\mathcal{O}((m_k+1)n_x^3)$, respectively. When these problems occur, applying covariance inflation in OLD and HLD has computational complexities $\mathcal{O}(M_{\mathrm{IC/OD}}n_x^3)$ and $\mathcal{O}((m_k+1)M_{\mathrm{IC/OD}}n_x^3)$, respectively.

B.3 IMM-PDA Filter - Computational Load

In this section we investigate OLD, MLD, HLD and covariance inflation in terms of their computational complexity and compare them to a restarted secondary local IMM-PDA filter. A restarted local filter (Restarted LF) is an IMM-PDA filter with the same parameters as the principal IMM-PDA filter which is restarted after each data transmission. The only difference of secondary IMM-PDA filter from the principal IMM-PDA filter is that the gating is not applied in the secondary one. The secondary IMM-PDA filter uses the same gated measurements with the ones validated in principal IMM-PDA filter. Restarted LF first makes N IMM prediction updates (with computational load $\mathcal{O}(N_r n_x^3 + N_r^2 n_x^2)$) and N IMM $\times$ PDA measurement updates (with computational load $\mathcal{O}(N_r m_k(n_x^3 + n_y^3))$ in information form) with a measurement of size $n_y \times 1$. As a result the total computational complexity of Restarted LF for each data transmission is $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2 + N_r m_k(n_x^3 + n_y^3)))$.

OLD, MLD and HLD both make N IMM prediction updates similar to Restarted LF. Each decorrelation operation applied after the prediction updates in OLD, MLD and HLD can be assumed to have the same computational complexity of a single Kalman filter measurement update with a measurement of dimension $n_x \times 1$ (with computational load $\mathcal{O}(n_x^3)$). After the prediction updates, OLD, MLD and MHD apply decorrelation on a single estimate, on N_r estimates and $N_r \times (m_k+1)$ estimates, respectively. Hence the total computational complexities of OLD, MLD and HLD for each data transmission is $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + n_x^3)$, $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r n_x^3)$ and $\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r(m_k+1)n_x^3)$, respectively. OLD, MLD and HLD

are advantageous in terms of computational load similar with other filter types (IMM filter and PDA filter) when the data transmission period N is large.

The incompatible covariance and outlier decorrelation problems in OLD, MLD and HLD have computational complexities $\mathcal{O}(n_x^3)$, $\mathcal{O}(N_r n_x^3)$ and $\mathcal{O}(N_r(m_k+1)n_x^3)$, respectively. Once we have these problems, applying covariance inflation in OLD, MLD and HLD has computational complexities $\mathcal{O}(M_{\text{IC/OD}} n_x^3)$, $\mathcal{O}(N_r M_{\text{IC/OD}} n_x^3)$ and $\mathcal{O}(N_r(m_k+1)M_{\text{IC/OD}} n_x^3)$, respectively.

B.4 Summary of the Computational Complexity Results

In this section we summarize the computational complexity results for all proposed decorrelation approaches, i.e., OLD, MLD, HLD with covariance inflation and compare them to a restarted secondary local filter. All of the results are presented in Table B.1, where we see that OLD, MLD and HLD are all advantageous when the data transmission period N is large thanks to less number of measurement updates.

Table B.1: Summary of the computational complexity results

	Restarted LF	**OLD**	**MLD**	**HLD**
IMM	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2 + N_r(n_x^3 + n_y^3)))$	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + n_x^3)$	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r n_x^3)$	-
IMM - Check CI	-	$\mathcal{O}(n_x^3)$	$\mathcal{O}(N_r n_x^3)$	-
IMM - Apply CI	-	$\mathcal{O}(M_{\text{IC/OD}} n_x^3)$	$\mathcal{O}(N_r M_{\text{IC/OD}} n_x^3)$	-
PDA	$\mathcal{O}(N(n_x^3 + m_k(n_x^3 + n_y^3)))$	$\mathcal{O}(N(n_x^3) + n_x^3)$	-	$\mathcal{O}(N(n_x^3) + (m_k + 1)n_x^3)$
PDA - Check CI	-	$\mathcal{O}(n_x^3)$	-	$\mathcal{O}((m_k + 1)n_x^3)$
PDA - Apply CI	-	$\mathcal{O}(M_{\text{IC/OD}} n_x^3)$	-	$\mathcal{O}((m_k + 1)M_{\text{IC/OD}} n_x^3)$
IMM-PDA	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2 + N_r m_k(n_x^3 + n_y^3)))$	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + n_x^3)$	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r n_x^3)$	$\mathcal{O}(N(N_r n_x^3 + N_r^2 n_x^2) + N_r(m_k + 1)n_x^3)$
IMM-PDA - Check CI	-	$\mathcal{O}(n_x^3)$	$\mathcal{O}(N_r n_x^3)$	$\mathcal{O}(N_r(m_k + 1)n_x^3)$
IMM-PDA - Apply CI	-	$\mathcal{O}(M_{\text{IC/OD}} n_x^3)$	$\mathcal{O}(N_r M_{\text{IC/OD}} n_x^3)$	$\mathcal{O}(N_r(m_k + 1)M_{\text{IC/OD}} n_x^3)$