

A Particle Markov Chain Monte Carlo Algorithm for Random Finite Set based Multi-Target Tracking

Tuyet Thi Anh Vu

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

National ICT Australia, NICTA
Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

August 2011

Produced on Archival Quality Paper

Copyright © 2011 Tuyet Thi Anh Vu

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the author.

Abstract

The multi target tracking (MTT) problem is essentially that of estimating the presence and associated time trajectories of moving objects based on measurements from a variety of sensors. Tracking a large number of unknown targets which move close and cross each other such as biological cells becomes difficult. The targets being tracked may randomly appear and disappear from the field of view, they may be temporarily obscured by other objects, may merge and split, may spawn other targets, and may cross or travel very close to each other for extended periods of time. Sensor measurements also present a number of challenging characteristics, such as noise which introduces location errors and may cause missed detection of targets, false measurements which do not belong to a valid target of interest, ghosting, misidentification etc.

A new approach to this problem is proposed by first formulating the problem in a random set finite framework and then using the Particle Markov Chain Monte Carlo (PMCMC) method for solving the problem. Under the random finite set (RFS) framework originally proposed by Mahler, a multi-target posterior distribution is propagated recursively via a Bayesian framework. The intractability of the posterior distribution is computed by using the PMCMC method that uses the sequential Monte Carlo outputs for the Markov Chain Monte Carlo (MCMC) method.

A RFS is a finite-set valued random variable. Alternatively, RFS can be interpreted as a random variable that is random in number of elements and in the values of these elements themselves and that the order of its elements is irrelevant. As a result, the RFS framework is a mathematically rigorous tool for capturing all uncertainties of its elements and its cardinality. With the uncertain properties of the MTT problem, the RFS framework is naturally used to formulate the MTT problem to capture the essence of MTT problem and then allows the multi-target posterior distribution to be propagated via a Bayesian framework. The first contribution of this dissertation is to derive the posterior distribution for the trajectories of the targets that is the special case for the multi-target posterior distribution. The multi-target posterior distribution is intractable so an approximation method such as PMCMC is required. PMCMC methods proposed by [4] use the Sequential Monte Carlo (SMC) algorithm to design an efficient high dimensional proposal distribution for the Markov Chain Monte Carlo (MCMC) method. The premise of this method is to sample from any distribution which has no closed form solution and which applying the traditional MCMC method or SMC method fails to give a reliable solution or is unfeasible on its own. The second contribution is to derive a RFS based PMCMC algorithm and implement this algorithm for the multi-target tracking problem when targets move close and/or cross each other in a dense environment and the number of targets is unknown.

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Tuyet Thi Anh Vu, August 2011

Acknowledgements

I wish to express my sincerest gratitude to all those who have helped and assisted me during my Ph.D candidature at the University of Melbourne.

First of all I would like to thank my supervisors. I would like to thank Professor Rob Evans for his support, guidance and help. His broad knowledge, deep insight, advice and helpfulness have been greatly appreciated. I would like to thank Professor Ba-Ngu Vo for his guidance, honest critique, advice and help. I would especially like to thank him for introducing me to random finite set theory and Particle Markov Chain Monte Carlo methods which I have found very interesting and enjoyed working on. I would also like to thank Dr Thomas Hanselmann for his helpful and friendly advice and in particular for encouraging me and helping me to study for a PhD.

I would like to thank National Information and Communications Technology Australia (NICTA) for their financial support during the entire course of my PhD, and for providing me with the opportunity to attend workshops and conferences. I would also like to thank the Department of Electrical and Electronic Engineering and the University of Melbourne for providing me with an inspiring environment and the necessary resources to complete this thesis.

I would like to thank Dr Mark Morelande for his advice, valuable recommendations and helpful discussions. In particular I would like to thank him for all the time he spent on giving very helpful comments on my work and especially on my conference presentations. I would like to thank Dr Rajib Chakravorty for his explanations and discussion on conventional target tracking techniques. I also want to thank Dr Ba-Tuong Vo for the short but valuable discussion and advice at Fusion 2011 and Dr Sanjeev Arulampalam for his advice and discussions. I would like to thank Dr Branko Ristic for the discussions about "tracking and data association" after attending his course. I also thank him for being so friendly during the course and at Fusion 2011 and for introducing me to people in my field. I would like to thank Professor Arnaud Doucet for the short but valuable discussion about Particle Markov Chain Monte Carlo and the discussion on the topic "the life of a PhD student". I would like to thank Haseeb Malik and Andrey Kan for proofreading chapters of my thesis.

I would like to thank Dr Robert Schmid and Dr Marcus Brazil for giving me the opportunity to tutor the subject Engineering Analysis A which wonderful opportunity for me to gain experience in undergraduate teaching at the University of Melbourne.

I would also like to thank Natasha Baxter, Domenic Santilli-Centofanti, Tracy Painter and all staff members, past and present, of NICTA and the Department of Electrical & Electronic Engineering for all their help.

My friends and fellow students from the Department of Electrical and Electronic Engineering were the spice of my PhD life. I am especially grateful to Dr An Tran, Dr. Ta Minh Chien, Dr Christian Vecchiola, Dr Marco A. S. Netto, Dr Bahman Tahayori, Dr. Matthieu Gilson, Elma O'Sullivan-Greene, Huynh Trong Anh, Tran Nhan, Mathias Foo, Kelvin Layton, Dean Freestone, Andre Peterson, Andreas Schutt, Sei Zhen Khong, Mohd Asyraf Zulkifley, Michelle Chong, Adel Ali, Sajeeb Saha, Haseeb Malik, Andrey Kan and all of my other friends.

I also thank to my niece Trần Tiên, my sister Vũ Thị Ánh Hồng and all the rest of my family for their help and advice in all aspects of life. Their smiles and greeting always make my life enjoyable and happy. I also thank to my friend Nguyễn Sơn in Vietnam for his help and encouragement.

I would also like to thank my former supervisor Dr Trần Minh Thuyết for his dedication and help during my studies in Vietnam which inspired me to do research. His deep insight and enthusiasm in mathematics encouraged me to study further to understand the beauty of mathematics and its applications.

Finally, I wish to give a whole-hearted thank to the most important person in my life - my partner Erik for his amazing encouragement, love and support, and for his dedication and enthusiasm in understanding my research work which lead to many valuable discussion and for proofreading my thesis and advising me on academic writing. There are no words to express my love for him and to thank him for his care, sense of humor and especially for his unconditional love and great support. His inspiration and smile have been a great encouragement for me during my PhD candidature. He has always been a best friend and a soul-mate with whom I could share not only all matters in daily life but also everything related to my studies. I also thank his family for their thoughts and support.

List of Abbreviations

SNR	:	Signal to noise ratio
KF	:	Kalman filter
EKF	:	Extended Kalman filter
UKF	:	Unscented Kalman filter
SIS	:	Sequential importance resampling
SMC	:	Sequential Monte Carlo
MC	:	Markov Chain
MCMC	:	Markov Chain Monte Carlo
MH	:	Metropolis-Hastings
MTT	:	Multi-target tracking
NNSF	:	Nearest neighbor standard filter
GNNF	:	Global nearest neighbor filter
PDA	:	Probabilistic data association
JPDA	:	Joint probabilistic data association
JIPDA	:	Joint Integrated probabilistic data association
MHT	:	Multiple hypothesis tracking
MCMCDA	:	Markov Chain Monte Carlo data association
RJMCMC	:	Reversible Jump Markov Chain Monte Carlo
PMCMC	:	Particle Markov Chain Monte Carlo
PMMH	:	Particle Marginal Metropolis-Hastings Algorithm
RFS	:	Random finite set
FISST	:	Finite set statistics
PHD	:	Probability hypothesis density
GM-PHD,GMPHD	:	Gaussian mixture implementation of PHD
SMC-PHD	:	SMC implementation of the PHD filter
CPHD	:	Cardinality probability hypothesis density
GM-CPHD	:	Gaussian mixture implementation of CPHD
MeMBeR	:	Multi-target multi-Bernoulli
CBMeMBeR	:	Cardinality balanced MeMBeR
GM-CBMeMBeR	:	Gaussian mixture implementation of CBMeMBeR
SMC-CBMeMBeR	:	SMC implementation of the CBMeMBeR
GP-CBMeMBeR	:	Gaussian particle MeMBeR

List of Figures

1.1	Measurements collected from the sensors	2
1.2	Underlying trajectory of target	2
1.3	Underlying trajectories of targets	2
2.1	Relationship between the state space and the observation space	8
2.2	Propagation of probability density function	13
2.3	Bayes filter and its implementations	14
2.4	One cycle in the Kalman filter equation for a linear system	15
4.1	Properties of states	59
4.2	Property of a state of MC	60
4.3	Intercommunication property	61
4.4	Recurrent and transient states of a MC	61
4.5	Transient set and irreducible sets of MC	62
4.6	Overview of Markov chains	66
5.1	Multi-target Bayes Filter and its first-order multi-target moment D_t	94
5.2	Overview of the approximation developments of the multi-target Bayes filter	101
6.1	Trajectories of targets	109
6.2	Proposal moves of the MC	120
6.3	Neighborhood of measurement	122
6.4	The Birth move and the Death move	124
6.5	Description of the Split and Merge move	124
6.6	Relationship between the extension and reduction move	125
6.7	The Switch move	126
6.8	Extension Merge move and its reverse move Extension Merge move	127
6.9	Extension Merge move and Birth Merge move	127
6.10	Proposal of an Birth Merge move and its reverse move Birth Merge move	128
6.11	Proposal of an Extension Merge move and Split Delete move	128
6.12	Extension Backward and Reduction Backward move	129
6.13	Update moves	129
6.14	Point Update moves at middle position	130
6.15	Point Update moves	130
6.16	Point Update moves	131
6.17	Point Update moves for two tracks	131
6.18	Point Update moves for two tracks at first time	132
6.19	Point Update moves for two tracks at first time	132
6.20	Example of aperiodic state	147
7.1	Location of the appearance of targets with mean $m_\gamma^{(i)}$ and $P_\gamma^{(i)}$, $i = 1, \dots, J$	153

7.2	Scenario of dense targets in 2D	154
7.3	Ground truth is plotted against to PMMH and GMPHD	155
7.4	OSPA errors between GM-PHD and PMMH	156
7.5	Cardinality and localization errors	156
7.6	Cardinality of GMPHD, PMMH and ground truth	157
7.7	Ground truth and its estimates versus time	157

List of Symbols

x_t	Single-target state at time t	7
\mathcal{X}	Single-target state space	7
z_t	Single-target measurement at time t	7
n_x	Dimensionality of a single-target state	7
\mathcal{Z}	Single-target measurement space	7
n_z	Dimensionality of a single-target measurement	7
$\mathcal{N}(\cdot; m, P)$	Gaussian with mean m and covariance P	8
T	The duration of surveillance	8
\mathcal{T}	The set of time indices	8
$\bar{f}_{t t-1}$	The single target transition density from time $t - 1$ to time t	9
\bar{g}_t	The single target measurement likelihood at time t	9
$z_{1:t}$	A sequence of single-target measurements up to time t	10
$p_{1:t}$	The posterior distribution up to time t	10
p_t	The posterior distribution at time t	10
$p_{t t-1}$	The predicted distribution at time t	10
$\delta(\cdot)$	Dirac delta function	20
$\mathcal{F}(A)$	The collection of finite subset of A	28
S^n	The Cartesian product of S taken n times	34
β_Σ	Belief functional of an RFS Σ	35
$\langle u, 1 \rangle$	$\int u(x)dx$	39
G_Σ	The probability-generating functional (p.g.fl.) of an RFS Σ	41
$\mathbf{1}_S$	The indicator function of S	42
X_t	Multi-target state at time t	43
K_x	The unit of volume on state space \mathcal{X}	44
γ_t	The intensity of new born target at time t	47
$\beta_{t t-1}(\cdot x)$	The intensity of spawnings at time t from target x	47
$f_{t t-1}$	The multi-target transition density from time $t - 1$ to time t	47
D_t	The RFS of target-generated measurement at time t	48
Z_t	Multi-target measurement at time t	48
K_z	The unit of volume on measurement space \mathcal{Z}	48
Λ_t	The RFS of clutter at time t	49
g_t	The multi-target likelihood function at time t	49
κ_t	The intensity of clutter at time t	49
$Z_{1:t}$	A sequence of multi-target measurements up to time t	50
\mathbb{N}^*	The set of natural number starting from 0	87
K	The maximum number of target in the region of interest	107
\mathcal{K}	The list of target labels	107
T	The number of measurement scans	107
\mathcal{T}	The set of time indices	107

\tilde{x}	The augmented single-target state	107
τ	A track (k, t, x_0, \dots, x_m) with label k , initial time t and target states x_0, \dots, x_m	108
m^*	A track gate	108
$\mathfrak{T}_0(\tau), \mathfrak{T}_0(\tilde{\theta}_\tau)$	The initial time of τ	108
$\mathfrak{T}_f(\tau), \mathfrak{T}_f(\tilde{\theta}_\tau)$	The last existing time of τ	108
$\mathcal{L}(\tau), \mathcal{L}(\tilde{\theta}_\tau)$	The track label of τ	108
$\mathbf{x}_t(\tau)$	The single-target state of track τ at time t'	108
ω	A track hypothesis	108
$\mathbf{X}_t(\omega)$	The multi-target state of track hypothesis ω at time t'	108
\tilde{X}_t	The augmented multi-target state at time t	109
$f_{t t-1}(\tilde{X}_t \tilde{X}_{t-1})$	The augmented multi-target transition density	110
$g_t(Z_t \tilde{X}_t)$	The augmented multi-target likelihood	111
θ_t	Auxiliary variable at time t	111
$\tilde{\theta}_t$	The augmented auxiliary variable at time t	112
$g_t(\tilde{Z}_t \tilde{X}_t, \theta_t)$	The density of \tilde{Z}_t conditional on \tilde{X}_t and θ_t	112
\tilde{Z}_t	An ordered multi-target measurement at time t	112
$\tilde{\theta}_{1:t}$	A sequence of augmented auxiliary variables up to time t	112
$\tilde{X}, \tilde{X}_{1:T}$	A sequence of augmented multi-target states up to time T	113
$\tilde{Z}, \tilde{Z}_{1:T}$	A sequence of ordered multi-target measurement up to time T	113
$\tilde{\theta}, \tilde{\theta}_{1:T}$	A sequence of augmented auxiliary variable up to time T	113
$\Lambda_t(\omega)$	The clutter associated with track hypothesis ω	118

List of Algorithms

1	Sequential Importance Sampling	23
2	Metropolis-Hastings Algorithm	67
3	Gibbs Algorithm	68
4	SMC Algorithm	71
5	Particle Independent Metropolis-Hastings Sampler	74
6	Particle Marginal Metropolis-Hastings Sampler	75
7	Conditional SMC algorithm	76
8	Sub-block Update using Conditional SMC algorithm	76
9	Particle Gibbs Algorithm	77
10	Multi-scan MCMC DA Algorithm	89
11	SMC Algorithm	115
12	PMMH Algorithm	117
13	Conditional SMC Algorithm	148
14	PMMH Algorithm for MTT	148

Contents

1	Introduction	1
1.1	Motivation and scope	1
1.2	Organization	3
1.3	Contributions	5
2	Bayesian Filtering	7
2.1	Single Target System Model	7
2.2	Bayes Approach	9
2.2.1	Bayes Theorem	9
2.2.2	Bayes Estimators	9
2.3	Bayes Filter and Its Implementations	11
2.3.1	Bayes Filter	11
2.3.2	The Kalman Filter and Its Variants	14
2.3.3	Particle Filter	19
3	Random Finite Set (RFS) for Filtering	27
3.1	Background on Random Finite Set	27
3.1.1	Mathematical Preliminaries	27
3.1.2	Measure and Integral of RFSs	30
3.1.3	Finite Set Statistic (FISST)	31
3.1.4	Probability Theory and FISST	38
3.1.5	Some Important Multi-target Probability Distribution	38
3.2	RFS Model for Multi-target Tracking	43
3.2.1	Multi-target Dynamical Model	43
3.2.2	Multi-target Measurement Model	48
3.2.3	Multi-target Bayes Filter	49
3.2.4	Multi-target Moment Densities	53
3.3	Conclusion	54
4	Particle Markov Chain Monte Carlo (PMCMC) Methods	55
4.1	Markov Chain Monte Carlo	55
4.1.1	Markov chains	56
4.1.2	Metropolis-Hastings Algorithm	67
4.1.3	Gibbs Sampler	68
4.2	Particle Markov Chain Monte Carlo methods	69
4.2.1	Sequential Monte Carlo Algorithm	70
4.2.2	Particle Independent Metropolis Hastings Sampler	73
4.2.3	Particle Marginal Metropolis-Hastings (PMMH) Sampler	74
4.2.4	Modified Particle Gibbs Sampler	75

4.3	Conclusion	77
5	Literature Review in Target Tracking	79
5.1	Conventional Target Tracking Techniques	79
5.1.1	Single-target Tracking in Clutter	80
5.1.2	Multiple Target Tracking (MTT) Techniques	83
5.2	RFS-based Target Tracking Techniques	89
5.2.1	Single-target Tracking	89
5.2.2	Multi-target Tracking	93
5.3	Conclusion and Discussion	102
6	PMCMC Method for RFS based Multi-target Tracking	103
6.1	Introduction	103
6.2	Formulation of the MTT problem in an RFS framework	104
6.2.1	Multi-target System Model in Random Finite Set Framework	104
6.2.2	Track Hypothesis in RFS Framework	107
6.2.3	Posterior Distribution	110
6.3	PMMH Algorithm for RFS-based Multi-target Tracking	114
6.3.1	PMMH Algorithm	114
6.3.2	Design and Construction of Proposal Distribution	117
6.3.3	PMMH Algorithm for RFS based Multi-target tracking	148
6.4	Summary and Discussion	149
7	Simulation and Performance	151
7.1	Multi-object Miss-distance	151
7.2	Simulation and Performance	152
7.3	Numerical Result and Discussion	154
7.4	Conclusion	157
8	Conclusion	159
8.1	Summary and Conclusion	159
8.2	Future Research	160
Appendices		
A	Background Mathematics	161
A.1	Probability and measures	161
A.2	Topology	162
A.3	Integration of measurable function	163
A.4	Markov Chains	164
B	Mathematical Proofs	167
Bibliography		171

Chapter 1

Introduction

Target tracking is the process of extracting information about one or more targets of interest based on the data or measurements collected from one or more sensors. Target tracking is challenging for one or more of the following reasons: 1) the origin of the measurements is unknown; 2) a measurement generated from a target is corrupted by noise; 3) the sensor(s) may not detect the target(s); and 4) the number of targets is unknown. The difficulty increases in direct proportion to the number of these conditions which apply. The successive estimates of target states conditional on all available measurements give a trajectory of the target which is called a track. Some typical applications of target tracking are military applications such as surveillance, air-to-air defence, battle field intelligence and defence; and non-military application such as robotics, image processing, automatic control and medicine [8, 17, 22, 65].

1.1 Motivation and scope

Assume that sensors have collected a large number of measurements at time steps 1, 2 and 3 which are represented as planes Z_1 , Z_2 and Z_3 respectively in Figure 1.1. We consider the two following scenarios

- (P.1) Only one target moves in the region of interest in a noisy and cluttered environment and the sensor(s) may not reliably detect the target in Figure 1.2 . This problem is called single target tracking in clutter.
- (P.2) More than one target move in the region of interest in a noisy and highly dense cluttered environment. This problem is called multi target tracking (MTT) and is considered more challenging than single target tracking. Consider Figure 1.3, where the tracks in the region are represented by different colors. The problem becomes even more difficult when a large unknown number of targets move close to each other, may also cross each other. In addition, they may spawn other targets or die unpredictable which increases the difficulty of the problem. Such problems occur in medicine when tracking the biological cell movement which plays an important role for understanding the cell development and helps in detecting cancer cells.

There are many existing techniques in the literature for handling problems (P.1) which take the uncertainty of measurement origin and missed detections into account including the nearest neigh-

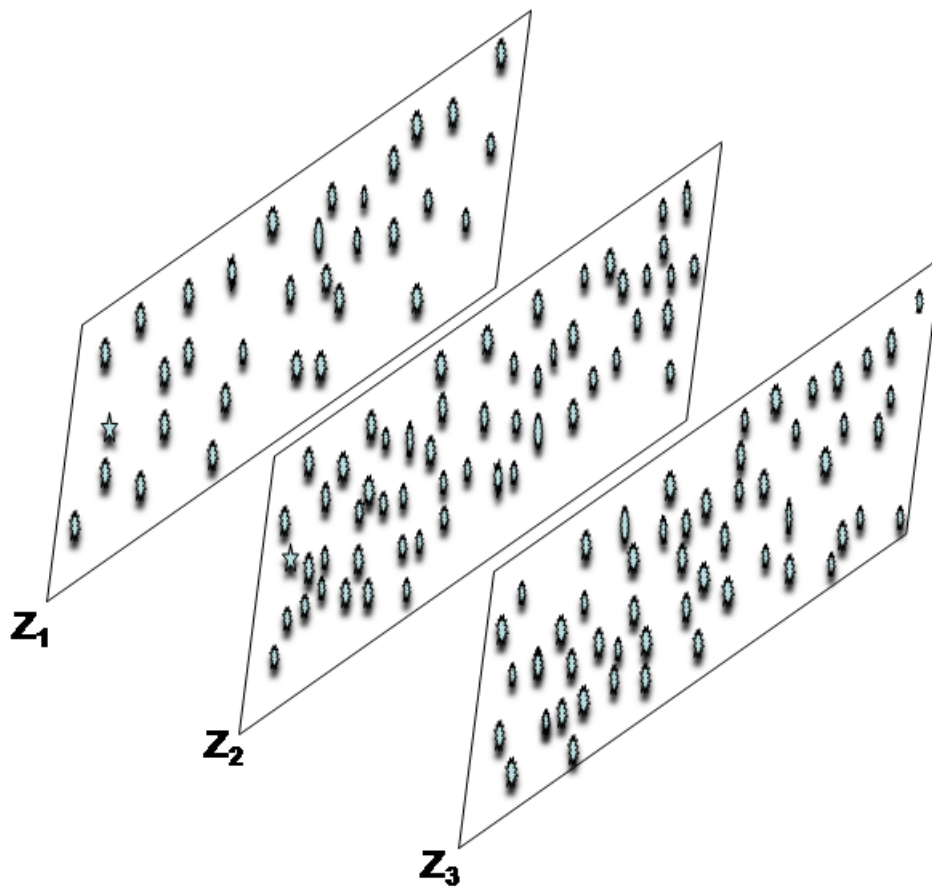


Figure 1.1: Set of measurements collected from the sensors at time step 1, 2, and 3.

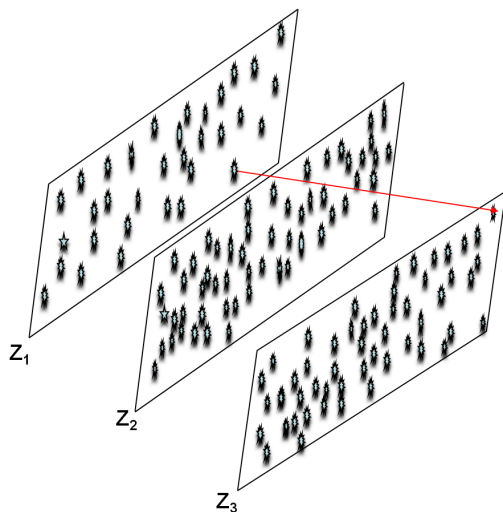


Figure 1.2: A possible underlying trajectory of target where the target at time 2 is not detected by the sensors.

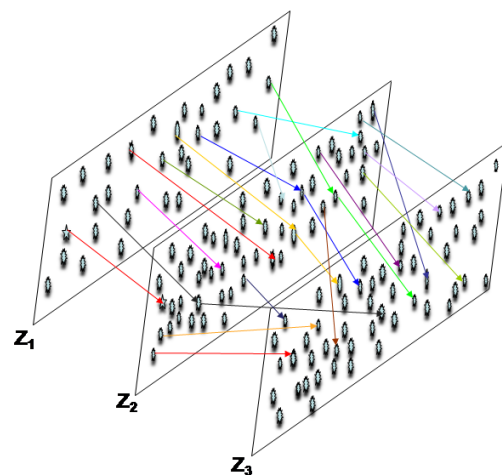


Figure 1.3: Possible underlying trajectories of the targets where tracks are represented by lines of different colors.

bor standard filter [8], probabilistic density association filter (PDAF) [8, 11] and their variants; and RFS-based technique [172, 182].

For problem (P.2), many conventional techniques, which combine data association methods and Bayesian filtering, were derived to track multiple targets provided that the number of targets is moderate and targets do not move too close to each other. If the number of targets is known and moderate, the global nearest neighbor filter [8, 14, 17], Joint PDAF [8, 17] or their variants can be applied. If the number of targets is unknown and moderate, the multiple hypothesis tracking can be used [16, 145, 160]. In the last decade, new techniques were derived to deal with MTT problems based on random finite set (RFS) theory which deals with finite-set-valued random variable with the properties that its number and values are random and the order of its elements is not important [60, 101]. Modeling the MTT problem in the RFS framework not only captures the uncertainty caused by the four above-mentioned difficulties but also allows the full multi-target Bayesian filter to be propagated in a similar way as the single-target Bayesian filter. The advantage of these techniques compared to conventional techniques is that the number of targets can be estimated in an optimal manner along with their states. For the problem described in (P.2), the existing techniques break down when there is a large unknown number of targets and when tracks are closely-spaced and crossing each other. Even the current RFS-based techniques break down under these conditions.

This thesis addresses and proposes a solution for the tracking problem under such conditions. The proposed method is based on the use of the batch processing to estimate a set of tracks (the trajectories of targets) from the multi-target posterior distribution obtained from a Bayesian recursive framework. The complicated Bayesian recursion, which results from multiple integrals of a sequence of sets, can be solved by sampling methods in order to find a sample which maximizes the multi-target posterior distribution. This method involves three issues: 1) Formulate the posterior distribution of trajectories of targets conditional on all measurements available which captures all information about target states and their labels. This distribution is also the distribution of a sequence of multi-target states; 2) Find independent samples from this posterior distribution where each sample is a sequence of multi-target states over all time scans; and 3) Find the optimal estimator which can deal with a set of tracks where the number of track is random and the number of states in each track is also random.

In this thesis, the first two issues are addressed while the last issue is briefly considered as a question for further research. The first two issues are solved by the development of a Bayesian multi-target batch processing algorithm based on RFS modeling and a Particle Markov Chain Monte Carlo (PMCMC) numerical approximation with a Gaussian Mixture Probability Hypothesis Density (GM-PHD) initialization. This algorithm is capable of tracking a large number of unknown targets in very high density situations and in a highly dense cluttered environment. This contribution has been published in Fusion 2011 [185].

1.2 Organization

This dissertation is organized as follows

Chapter 2 presents single-target Bayesian filtering. Bayesian filtering is the foundation for most of the approaches for single target tracking. At each time when measurements are received, the new estimate of the target state is obtained by combining the new information from the measurements with the current estimate. The two most popular approaches to estimate the target states, Minimum Mean Square Error Estimation (MMSE) and Maximum A Posterior (MAP) Estimation, are also introduced. Some special cases and approximations of the Bayes filter such as the Kalman filter and the particle filter are presented.

Chapter 3 summarizes the random finite set (RFS) theory. Concepts like the transition density and likelihood functions for multi-target tracking are introduced leading to the formulation of Bayesian multi-target filtering in the RFS framework [3].

Chapter 4 describes Particle Markov Chain Monte Carlo (PMCMC) methods [4, 70], a numerical approximation which combines the Markov Chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC) methods by utilizing the strength of each of these methods. The approach of PMCMC is to use the SMC algorithm to design efficient high dimensional proposal distributions for MCMC algorithms when these high dimensional proposal distributions cannot be satisfactorily sampled using either SMC or MCMC on its own.

Chapter 5 reviews the target tracking in clutter. It discusses the traditional techniques which deal with single-target tracking [8, 10, 14, 17, 116, 121] and multi-target tracking [8, 118, 123] in clutter by applying data association techniques and filtering algorithms. Data association problem in multi-target tracking problem assigns each measurement to a target and then this measurement is used to update the target state through filtering technique so that the trajectories of each target can be estimated recursively. A new approach for target tracking based on random finite set (RFS) framework is introduced Section 5.2. Random finite set based single target tracking filtering [182] is introduced in Subsection 5.2.1. This Subsection also describe the mathematically rigorous Bayes' recursion for tracking a target that generates multiple measurements in the presence of clutter. Subsection 5.2.2 presents the multi-target tracking algorithms based on RFS. One of the most popular approach, e.g. the PHD filter derived by Mahler [96] which is an approximation of the full multi-target Bayesian filtering, is presented. A closed form solution, the Gaussian mixture PHD filter (GM-PHD) recursion is also presented in this Subsection.

Chapter 6, which contains the main contribution of this thesis, proposes a new technique for multi-target tracking under high target density and clutter. Section 6.2 formulates the problem in a RFS framework, and derives the Bayesian recursion for propagating the posterior distribution of the target trajectories. This posterior distribution is computationally demanding as all possible pairings of measurement and targets must be considered. The complexity of the problem is reduced by the introduction of an auxiliary variable which is expressed as a relationship between target labels and measurement indices at a time instance. Section 6.3 proposes a viable solution for estimating this posterior distribution which has no closed form expression by using Particle Marginal Metropolis-Hastings Algorithm (PMMH) which is an PMCMC method.

Chapter 7 illustrates the PMMH algorithm for RFS based Multi-target tracking described in Chapter 6 on a simulation example and evaluates its performance. Some discussion and performance evaluation are presented in this Chapter.

Chapter 8 summarizes the dissertation. Future research direction for tracking closely spaced and crossing targets at low computational cost are outlined.

1.3 Contributions

This thesis presents a number of contributions to the area of multi-target tracking. Four minor but important contributions are presented in Chapters 2-5. These contributions serve as a foundation to the development of the three major contributions of this thesis, related to Bayesian multi-target batch processing. The proposed method is based on RFS modeling and PMCMC numerical approximation with the Gaussian Mixture Probability Hypothesis Density (GM-PHD) initialization and it is capable of tracking a large number of unknown targets in very high density situations and in a highly dense clutter environment. Chapters 6 and 7 contain the three major contributions. The contributions are summarized as follows

1. The first minor contribution is a comprehensive overview of Bayesian filtering for single target tracking and estimation presented in Chapter 2. This representation of Bayesian filtering is the foundation for most of the tracking techniques such as the conventional target tracking techniques found in Chapter 5 and the derivation of Bayesian filtering for multi-target tracking found in Chapter 3.2.
2. The second minor contribution of this thesis is an overview of random finite set theory presented in Chapter 3.1. This overview and the Bayesian filtering in Chapter 2 lead to the modeling of multi-target tracking problems and the derivation of the multi-target Bayesian recursions found in Chapter 3.2. The multi-target Bayesian recursion is a fundamental tool for deriving all the RFS-based techniques presented in Chapter 5.2.
3. The third minor contribution is a focussed overview of the various simulations and sampling methods presented in Chapter 2.3.3 and Chapter 4. In this overview, new sampling methods, particle Markov Chain Monte Carlo (PMCMC) methods which use the output of the Sequential Monte Carlo (SMC) method as the Markov Chain Monte Carlo (MCMC) update, are presented. They are important techniques which are able to sample from a complicated distributions and the main contributions of this thesis are based upon these techniques.
4. The final minor contribution of this thesis is a concise summary of target tracking techniques found in the literature presented in Chapter 5. This summary shows the development of existing techniques as they attempted to address increasingly complicated problems arising over time. This includes conventional techniques existing for the past 50 years and the RFS-based techniques existing for a decade or so. These two kinds of techniques are still under development (especially the RFS-based techniques) to give better solutions to the multi-target tracking problem. As a result, this summary has shown that there is no existing technique which can handle the problem where a large number of dense targets move and cross each other in noisy and cluttered environment.

5. The first major contribution of the thesis is the RFS-based formulation of the MTT problem where a large number of dense targets move close and cross each other and can be found in Chapter 6.2. In this chapter, the posterior distribution of a track set (trajectories of the targets) is derived based on the Bayesian recursion. By formulating an augmented multi-target state as an extension of the multi-target state, conditional on all available measurements the posterior distribution of a set track is a posterior distribution of a sequence of the augmented multi-target states. The posterior of the track set is computationally intractable when there exists a large number of dense and crossing targets in densely cluttered environment and in highly dense clutter and the introduction of augmented auxiliary variable is needed. Conditional on all available measurements, a posterior distribution of track set and a sequence of the augmented auxiliary variables is derived.
6. The second major contribution is the derivation of an Algorithm using PMCMC method for sampling from the posterior distribution of a track set and a sequence of augmented auxiliary variables. This algorithm can be found in Chapter 6.3. In this chapter, the discussion of the disadvantages of using two powerful sampling techniques such as the MCMC and SMC on their own leads to the idea of choosing the approximation methods which combine the strength of these techniques to generate sample from this distribution such as the Particle MCMC (PMCMC) methods derived in [4]. A well known property of MCMC as well as PMCMC is that the rate of convergence depends on the initial distribution. Thus an estimate from a popular filtering technique such as GM-PHD filter is used as the initial state of a Markov chain in order to reduce the computational cost of PMCMC.
7. The last major contribution of this thesis is the simulation and associated discussion found in Chapter 7. The simulation illustrates the performance of the algorithm which show that the algorithm is capable of tracking a large unknown number of dense targets in a highly dense cluttered environment.

The publications based on this thesis are

Conference:

- A.-T. Vu, B.-N. Vo, and R. Evans, "Particle Markov Chain Monte Carlo for Bayesian Multi-target Tracking," Proc. 14th Annual Conf. Information Fusion, Chicago, USA, 2011. (Best Student Paper Award Finalist).

Journal:

- A.-T. Vu, B.-N. Vo, and R. Evans, "Particle Marginal Metropolis-Hastings Algorithm for Bayesian Multi-target Tracking". In preparation

Chapter 2

Bayesian Filtering

The purpose of tracking is to extract information about the targets from the available measurements. The target tracking is usually deemed successful when the useful properties of the targets are efficiently obtained from the observations. In practice, tracking aims to estimate the trajectories of the targets observed in the area of interest. This chapter provides an overview of Bayesian filtering for single target, which is based on general Bayesian filtering [3, 21, 44] or [9, 22] (Bayesian filtering for target tracking).

The outline of the chapter is as follows. Section 2.1 introduces a common model for single target tracking. Section 2.2 describes the Bayes approach which is the central foundation for most target tracking techniques. Section 2.2.2 introduces the two most common estimators for target tracking. Section 2.3 is devoted to presenting the Bayes filter and its application to target tracking.

2.1 Single Target System Model

The target which is tracked can be an air-craft, a person, a weather balloon, a biological cell etc. The target states and target behavior are normally unknown. Depending on the type of the target and the (noisy) environment, the target behavior can be modeled systematically with or without the presence of the noise. The measurements obtained from the target can be e.g. radar measurements or video images. Based on the type of the measurements, the measurements can also be modeled systematically in order to establish the relationship between target states and the measurements. In practice, the target states are hidden and only partially observed in the observation space or can only be measured with error. In general, the available measurements are noisy and are not the same as target states (see Figure 2.1). Furthermore, the measurements are received at regular time interval therefore the target dynamic system or measurement system can be modeled as the discrete time system as follows.

At each time t , the target state is represented by the vector x_t taking values in a *state space* $\mathcal{X} \subset \mathbb{R}^{n_x}$, and is indirectly observed via a noisy measurement vector z_t taking value in a *measurement*

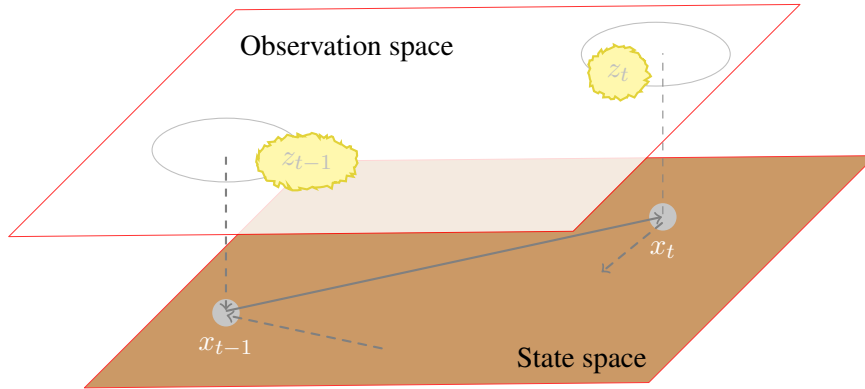


Figure 2.1: Based on [97]. When a target moves, it generates target states x_{t-1} and x_t on the state space. The target motion and the target states x_{t-1} , x_t are only known through the measurements z_{t-1} , z_t generated from the target states respectively on the measurement space.

space $\mathcal{Z} \subset \mathbb{R}^{n_z}$. The time evolution of target state is described by

$$x_t = \begin{cases} F_{t-1}x_{t-1} + B_t u_t + v_t, & \text{for linear system;} \\ \mathbf{f}_{t-1}(x_{t-1}, u_t) + v_t, & \text{for non-linear system.} \end{cases} \quad (2.1)$$

where

- F_{t-1} is the state transition matrix of the linear system model at time $t - 1$,
- $\mathbf{f}_{t-1}(\cdot)$ is the transition function for non-linear system at time $t - 1$,
- B_t is the control-input matrix which is pre-multiplied with the control vector u_t ,
- v_t is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_t , $v_t \sim \mathcal{N}(v_t; 0, Q_t)$.

(2.1) specifies the transformation of any given target state x_{t-1} at time $t - 1$ to a new state x_t , taking vector noise v_t into account.

The target state x_t is observed by the noisy measurement

$$z_t = \begin{cases} H_t x_t + w_t, & \text{for linear system;} \\ \mathbf{h}_t(x_t) + w_t, & \text{for non-linear system.} \end{cases} \quad (2.2)$$

where

- H_t is the observation matrix which maps the true state vector into the measurement space at time t ,
- $\mathbf{h}_t(\cdot)$ is the known observation function at time t ,
- w_t is the observation noise which is assumed to be drawn from a zero mean multivariate Gaussian white noise with covariance R_t , $w_t \sim \mathcal{N}(w_t; 0, R_t)$.

Let T be the duration of surveillance and let $\mathcal{T} = \{1, \dots, T\}$ be the set of time indices. The initial state x_1 , and the noise vectors at each time step $v_2, \dots, v_T, w_1, \dots, w_T$ are all assumed to be mutually independent. By this assumption and the form of (2.1), the sequence of target state $\{x_t : t \in \mathcal{T}\}$ follows a first order Markov process¹. Given the probability distribution $p_0(x_1)$

¹see definition of the first order Markov process in A.17

of the initial state x_1 , the time evolution of the target state is alternatively described by a *Markov transition density* $\bar{f}_{t|t-1}(\cdot|\cdot)$ ($t > 1$) where

$$\bar{f}_{t|t-1}(x_t|x_{t-1}) \quad (2.3)$$

is the probability density of the target state x_t at time t given the target state x_{t-1} at time $t - 1$ i.e. it describes how the target state at time $t - 1$ moves to a new target state at time t .

Similarly, the measurement vector at time t is alternatively modeled by the likelihood function $\bar{g}_t(\cdot|\cdot)$ where

$$\bar{g}_t(z_t|x_t) \quad (2.4)$$

describes at time t how likely it is that the target state x_t generates the measurement z_t .

2.2 Bayes Approach

The Bayesian approach is widely used in statistical inference, and in many areas of science and engineering. In target tracking, it is the standard approach to modeling and the development of target tracking algorithms. When new measurements are collected from the sensor(s), the current estimate of the target state is updated by combining the new information in the new measurements with the previous estimate of the target state. This update process can be implemented recursively in time, and it is formalized using the Bayes's theorem which was first developed by Thomas Bayes [12]. The material can be found in many mathematical books such as [3, 21] or in target tracking literature e.g.[22].

2.2.1 Bayes Theorem

Bayesian estimation consider the problem of estimating a random variable x based on measurements of another random variable z . In such estimation problems the conditional density $p(x|z)$ plays an important role. It is also called the posterior distribution since it describes the distribution of x after having obtained the measurement z .

Bayes theorem relates $p(x|z)$ to $p(z|x)$ and $p(x)$ and states that

$$p(x|z) = \frac{p(z|x)p(x)}{\int p(z|x)p(x)dx}.$$

In target tracking x is usually a target state at a specific time or a sequence of the target states. Similarly, z is a measurement at a specific time or a sequence of the measurements.

2.2.2 Bayes Estimators

Let $\hat{x}(z)$ be an estimator of x given measurement z and let $L(x, \hat{x}(z))$ be the loss function or cost function e.g. squared error. The Bayes risk of $\hat{x}(z)$ is defined as $E_{x,z}[L(x, \hat{x}(z))]$ where

the expectation is taken over the joint distribution of x and z . This defines the risk function as a function of $\hat{x}(\cdot)$. An estimator $\hat{x}(z)$ is said to be a Bayes estimator if it minimizes the Bayes risk. The estimator which minimizes the posterior expected loss $E_{x|z}[L(x, \hat{x}(z))|z]$ for each z where the expectation is with respect to the conditional distribution of x given z also minimizes the Bayes risk and therefore is a Bayes estimator. The most frequent Bayes estimators are the *Minimum Mean Square error (MMSE)* and the *maximum a posteriori probability (MAP)* estimators.

2.2.2.1 Minimum Mean Square Error (MMSE) Estimator

MMSE uses the mean square error (MSE) as the risk function. Thus the Bayes risk is called the squared error risk and defined as

$$MSE(\hat{x}(z)) = E_{x,z}[(\hat{x}(z) - x)^2] \quad (2.5)$$

where the expectation is taken with respect to the joint distribution of x and z . This can be also written as

$$MSE(\hat{x}(z)) = E_{x,z}[(\hat{x}(z) - x)^2] = E_z(E_{x|z}[(\hat{x}(z) - x)^2|z]) \quad (2.6)$$

where the expectation $E_{x|z}$ is with respect to the conditional distribution of x given z , and E_z is the expectation with respect to the distribution of z . Hence the Bayes estimator

$$\begin{aligned} \hat{\mathbf{x}}(z) &= \underset{\hat{x}(z)}{\operatorname{argmin}} E_{x,z}[(\hat{x}(z) - x)^2] = \underset{\hat{x}(z)}{\operatorname{argmin}} E_z(E_{x|z}[(\hat{x}(z) - x)^2|z]) \\ &= \underset{\hat{x}(z)}{\operatorname{argmin}} E_{x|z}[(\hat{x}(z) - x)^2|z] \end{aligned} \quad (2.7)$$

That is the MMSE estimator is the $\hat{\mathbf{x}}(z)$ such that $E_{x|z}[(\hat{x}(z) - x)^2|z]$ is minimum. Equivalently, setting the derivative of $E_{x|z}[(\hat{x}(z) - x)^2|z]$ to zero, we have

$$\frac{d}{d\hat{x}(z)} E_{x|z}[(\hat{x}(z) - x)^2|z] = E_{x|z}[2(\hat{x}(z) - x)|z] = 2(\hat{x}(z) - E(x|z)) = 0. \quad (2.8)$$

Therefore, the MMSE estimate $\hat{\mathbf{x}}(z)$ of the x is simply the mean of a posterior distribution

$$\hat{\mathbf{x}}(z) = E[x|z] = \int xp(x|z)dx \quad (2.9)$$

and it is also called the *Expected A Posteriori (EAP) estimator* [101, p.63]

2.2.2.2 Maximum A Posteriori (MAP) Estimator

Maximum a posteriori (MAP) estimator maximizes the posterior probability distribution

$$\hat{\mathbf{x}}(z) = \underset{x}{\operatorname{argmax}} p(x|z) = \underset{x}{\operatorname{argmax}} p(z|x)p(x). \quad (2.10)$$

2.3 Bayes Filter and Its Implementations

Estimating the target states from noisy observations is an important problem in engineering and it can be found in many text books [3, 9, 75]. In the context of target tracking, the Bayes filter is the standard approach to recursive state estimation. The main algorithm development in this thesis Chapter 6 is utilizing the posterior distribution as derived in the Bayes filter. Given the measurement history $z_{1:l} = (z_1, \dots, z_l)$, the (discrete) estimation problem involves computing an estimate of the state x_t . The problem is called (discrete) smoothing if $t < l$, (discrete) filtering if $t = l$ and (discrete) prediction if $t > l$. Filtering and prediction are used in real time operation to estimate the current and future states given the measurement or data up to time l . The accuracy of the estimate can be improved by smoothing when more observations are accumulated. Since the probability density of the state x_t given the state history $x_{1:t-1} = (x_1, \dots, x_{t-1})$ is modeled as $p_{t|t-1}(x_t|x_{1:t-1})$ ($p_{t|t-1}(x_t|x_{1:t-1}) = \bar{f}_{t|t-1}(x_t|x_{t-1})$ for the model given in (2.3)) with initial density $p_0(x_1)$ and that the probability density of the observation $p_{1:t}(z_{1:t}|x_{1:t})$ is given, Bayes rule allows one to calculate the posterior probability density $p_{1:t}(x_{1:t}|z_{1:t})$. The posterior density $p_{1:t}(x_{1:t}|z_{1:t})$ is of importance for estimation problem because it encapsulate all the information about the state trajectories available from the measurements and prior information. In the following, a summary of the Bayes filter is given as well as some of its applications to linear and non-linear system models. The reader is referred to [45, 69, 147] for further details.

2.3.1 Bayes Filter

The Bayes filter for target tracking computes the posterior density of the target state given the history of measurements and the initial density. The posterior distribution encapsulate all the information about the target states and depends on the likelihood function and the prior distribution. Here the prior distribution is determined by the dynamic system model for the target and the prior distribution; and the likelihood function can be found from the measurement model.

Given an initial distribution p_0 . Let $x_{1:t} = (x_1, \dots, x_t)$ and $z_{1:t} = (z_1, \dots, z_t)$. Applying Bayes rule, the posterior distribution of the target state up to time t is given by

$$p_{1:t}(x_{1:t}|z_{1:t}) = \frac{p(z_{1:t}|x_{1:t})p(x_{1:t})}{p(z_{1:t})} \quad (2.11)$$

where $p(z_{1:t}) = \int p(z_{1:t}|x_{1:t})p(x_{1:t})dx_{1:t}$ is a normalizing constant, $p(x_{1:t})$ is the prior density, and $p(z_{1:t}|x_{1:t})$ is the likelihood. The likelihood is calculated as follows

$$\begin{aligned} p(z_{1:t}|x_{1:t}) &= p(z_t|x_{1:t}, z_{1:t-1})p(z_{t-1}|x_{1:t}, z_{1:t-2}) \dots p(z_1|x_{1:t}) \\ &\stackrel{(a)}{=} \bar{g}_t(z_t|x_t)\bar{g}_{t-1}(z_{t-1}|x_{t-1}) \dots \bar{g}_1(z_1|x_1) \end{aligned}$$

where (a) holds because of the measurement model in (2.2). We have

$$\begin{aligned}
p(z_{1:t+1}) &= \int p(z_{1:t+1}|x_{1:t+1})p(x_{1:t+1})dx_{1:t+1} \\
&= \int p(z_{t+1}|x_{1:t+1}, z_{1:t})p(z_{1:t}|x_{1:t+1})p(x_{1:t+1})dx_{1:t+1} \\
&= \int p(z_{t+1}|x_{1:t+1}, z_{1:t})p(z_{1:t}, x_{1:t+1})dx_{1:t+1} \\
&= \int p(z_{t+1}|x_{1:t+1}, z_{1:t})p(x_{t+1}|z_{1:t}, x_{1:t})p(z_{1:t}, x_{1:t})dx_{1:t+1} \\
&= \int \bar{g}(z_{t+1}|x_{t+1})p_{t+1|t}(x_{t+1}|z_{1:t})dx_{t+1} \int p(z_{1:t}, x_{1:t})dx_{1:t} \\
&= p(z_{t+1}|z_{1:t})p(z_{1:t})
\end{aligned} \tag{2.12}$$

Using Bayes recursion, the posterior distribution at time $t + 1$ is

$$\begin{aligned}
p_{1:t+1}(x_{1:t+1}|z_{1:t+1}) &= \frac{p(z_{1:t+1}|x_{1:t+1})p(x_{1:t+1})}{p(z_{1:t+1})} \\
&= \frac{p(z_{t+1}|x_{1:t+1}, z_{1:t})p(z_{1:t}|x_{1:t+1})p(x_{t+1}|x_{1:t})p(x_{1:t})}{p(z_{t+1}|z_{1:t})p(z_{1:t})} \\
&\stackrel{(a)}{=} \frac{p(z_{1:t}|x_{1:t})p(x_{1:t})}{p(z_{1:t})} \frac{\bar{g}_{t+1}(z_{t+1}|x_{t+1})p(x_{t+1}|x_{1:t})}{p(z_{t+1}|z_{1:t})} \\
&\stackrel{(b)}{=} p_{1:t}(x_{1:t}|z_{1:t}) \frac{\bar{g}_{t+1}(z_{t+1}|x_{t+1})\bar{f}_{t+1|t}(x_{t+1}|x_t)}{p(z_{t+1}|z_{1:t})}
\end{aligned} \tag{2.13}$$

where $p(z_{t+1}|z_{1:t}) = \int \bar{g}_{t+1}(z_{t+1}|x_{t+1})p_{t+1|t}(x_{t+1}|z_{1:t})dx_{t+1}$ by (2.12); (a) holds because of the measurement model in (2.2); and (b) holds because the dynamic system is a hidden Markov process i.e. $\bar{f}_{t+1|t}(x_{t+1}|x_t) = p(x_{t+1}|x_t)$. By (2.13), filtering recursion is also calculated

$$\begin{aligned}
p_{t+1}(x_{t+1}|z_{1:t+1}) &= \int p_{1:t+1}(x_{1:t+1}|z_{1:t+1})dx_{1:t} \\
&= \frac{\bar{g}_{t+1}(z_{t+1}|x_{t+1})}{p(z_{t+1}|z_{1:t})} \int \bar{f}_{t+1|t}(x_{t+1}|x_t)p_{1:t}(x_{1:t}|z_{1:t})dx_{1:t} \\
&\stackrel{(a)}{=} \frac{\bar{g}_{t+1}(z_{t+1}|x_{t+1})}{p(z_{t+1}|z_{1:t})} \int \bar{f}_{t+1|t}(x_{t+1}|x_t)p_t(x_t|z_{1:t})dx_t
\end{aligned} \tag{2.14}$$

where (a) holds because $p_t(x_t|z_{1:t}) = \int p_{1:t}(x_{1:t}|z_{1:t})dx_{1:t-1}$. Thus given initial distribution p_0 and the dynamic model and the measurement model, the Bayes filter can be computed in three steps

Initialization: Given an initial distribution p_0 which is usually to be an independently and identically distributed distribution or a Poisson distribution. The calculation of (2.14) can be divided in 2-step procedure: predict and update step

Predict:

$$p_{t+1|t}(x_{t+1}|z_{1:t}) = \int \bar{f}_{t+1|t}(x_{t+1}|x_t)p_t(x_t|z_{1:t})dx_t. \tag{2.15}$$

Update:

$$p_{t+1}(x_{t+1}|z_{1:t+1}) = \frac{\bar{g}_{t+1}(z_{t+1}|x_{t+1})p_{t+1|t}(x_{t+1}|z_{1:t})}{p(z_{t+1}|z_{1:t})}. \quad (2.16)$$

The propagation of the posterior distribution p_{t+1} , $t \geq 0$ over time is illustrated in Figure 2.2².

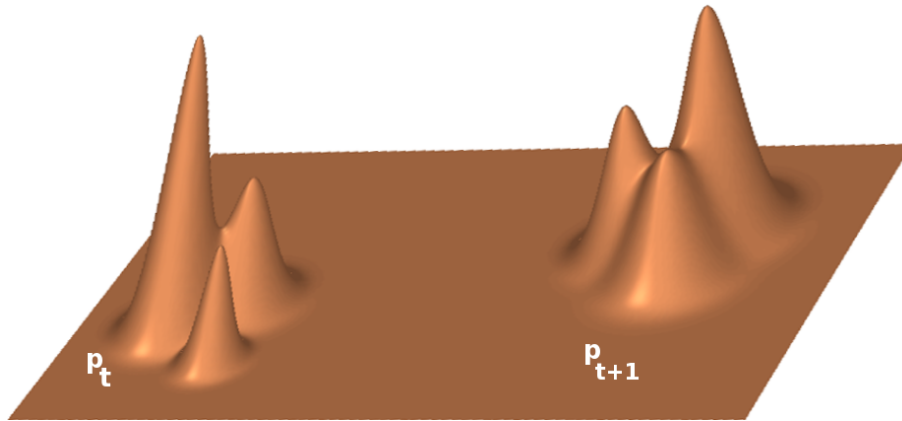


Figure 2.2: Illustration of the propagation of probability density function over time

Due to the multiple integral on the right hand side in the posterior density (2.16), the full implementation of the Bayes filter is generally intractable in practice.

Given the posterior distribution $p_t(x_t|z_{1:t})$ at time t , an optimal estimate of the state vector at time t given the history of measurements $z_{1:t}$ can be obtained using the estimators described in (2.9) and (2.10) as follows

$$\begin{aligned} \hat{\mathbf{x}}_t^{EAP}(z_{1:t}) &= \hat{\mathbf{x}}^{MMSE}(z_{1:t}) = E[x_t|z_{1:t}] = \int x_t p_t(x_t|z_{1:t}) dx_t \\ \hat{\mathbf{x}}_t^{MAP}(z_{1:t}) &= \arg \sup_{x_t} p_t(x_t|z_{1:t}) \end{aligned}$$

The following sections will discuss some approximations of the Bayes filter for single target tracking. In a linear system with Gaussian noise and Gaussian initial distribution, all posterior distributions in the Bayes recursions are Gaussian and hence completely determined by their mean values and covariance matrices. The Kalman filter (KF) propagates their means and covariance matrices and is described in Section 2.3.2.1. For Gaussian linear system with Gaussian noise, Bayes filter is KF. Approximations based on the KF for non-linear systems are described in Sections 2.3.2.2 and 2.3.2.3. Another approximation of Bayes filter is the Particle filter (PF) which is presented in Subsection 2.3.3. Figure 2.3³ shows the relationship between KF and PF.

²This figure is based on [175].

³This graph is based on the graph [96].

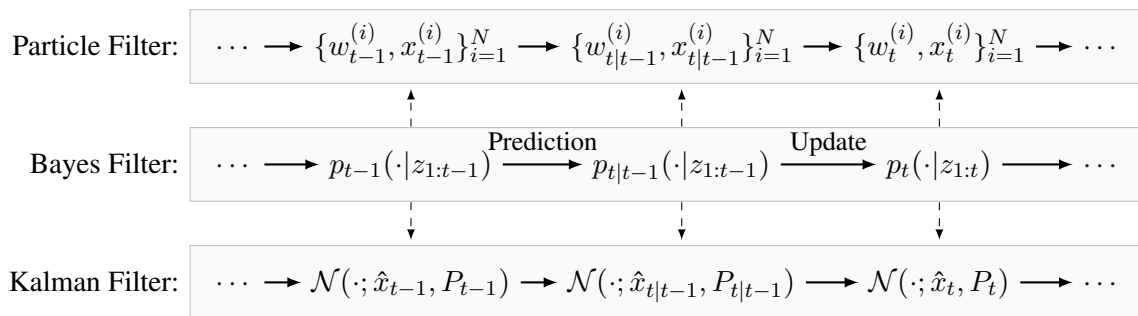


Figure 2.3: Single-target Bayes Filter and two of its implementations

2.3.2 The Kalman Filter and Its Variants

This section presents the Kalman filter (KF) which is the optimal Bayes filter for linear Gaussian systems and is described in Subsection 2.3.2.1. An approximation for non-linear system is to linearize the non-linear systems along the state trajectories and applying the KF to the linearized systems. This approach is called the extended Kalman filter (EKF), and it is described in Subsection 2.3.2.2. When the system is too skewed⁴, the Unscented Kalman filter (UKF) was derived to improve the performance of the EKF and is presented in Subsection 2.3.2.3. The material in this section can be found in books on tracking e.g. [101, 147] or in books in general filtering e.g. [3]

2.3.2.1 The Kalman Filter

Kalman filter (KF) was first developed by Kalman [81] and applied ubiquitously in many areas such as control system, tracking etc. KF is popular because it is easy to implement and it provides the closed form solution of the Bayes filter for Gaussian linear system. This section sketches its derivation.

The KF assumes a linear system given by

$$x_t = F_{t-1}x_{t-1} + B_t u_t + v_t \quad (2.17)$$

$$z_t = H_t x_t + w_t \quad (2.18)$$

The initial state, and the noise vectors at each step $x_1, v_2, \dots, v_T, w_1, \dots, w_T$ are all assumed to be mutually independent where v_t, w_t are zero mean vector valued Gaussian random variable with covariance matrices Q_t and R_t respectively. The initial state is a Gaussian vector with mean $E[x_1]$ and covariance $cov(x_1)$. The KF is represented by two variables for $t > 1$:

- \hat{x}_t , state estimate at time t given observations up to time t ($t > 1$),
- \mathbf{P}_t , the error covariance matrix where the error is defined as $x_t - \hat{x}_t$.

The state vector \hat{x}_t contains information about the target at time t based on the measurement collected up to time t and the error covariance matrix \mathbf{P}_t describes the uncertainty in \hat{x}_t .

⁴Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable, e.g when the distribution is symmetric then there is zero skewness.

Denote $\hat{x}_1 = E[x_1]$ and $\mathbf{P}_1 = cov(x_1)$. At time $t - 1, t > 1$, the state estimate \hat{x}_{t-1} and the associated covariance \mathbf{P}_{t-1} are given. Then at time t , the state estimate \hat{x}_t and covariance estimate \mathbf{P}_t are constructed in two steps: Prediction and Update; as summarized in Figure 2.4.

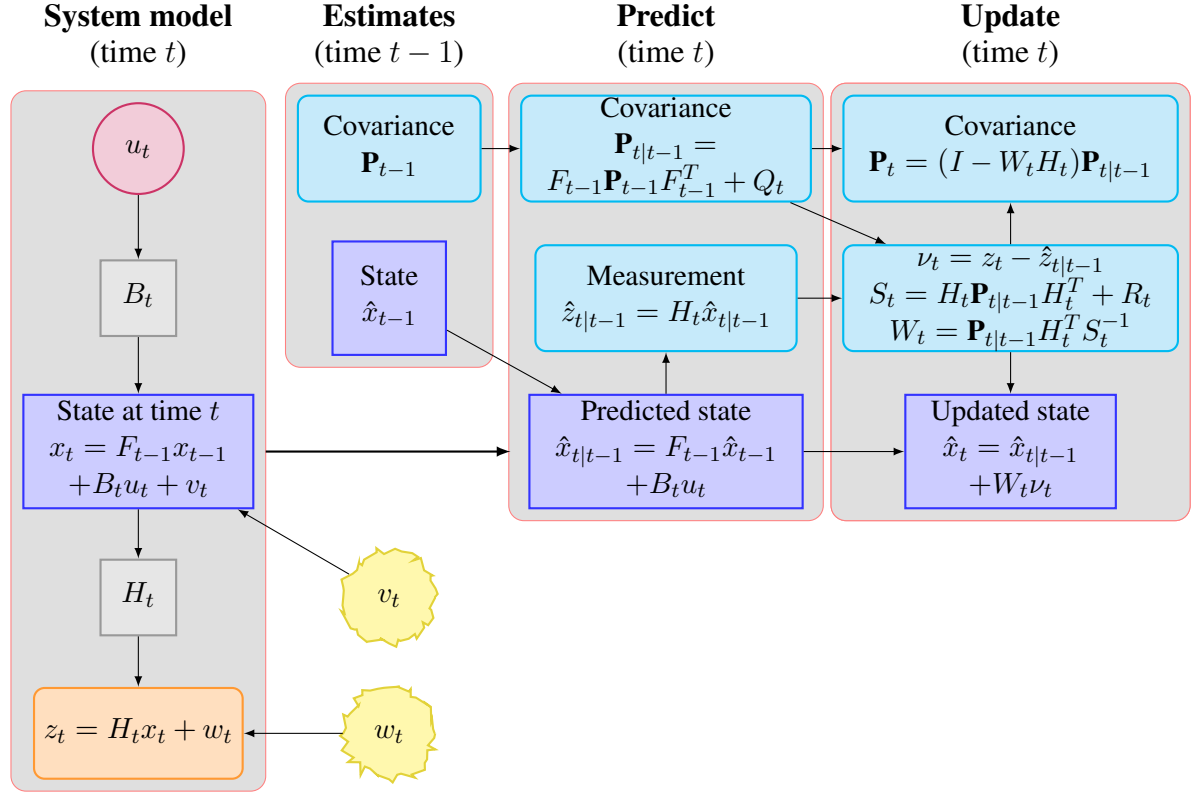


Figure 2.4: One cycle in the Kalman filter equation for a linear system

Prediction:

By the assumption of Gaussian process noise with $E[v_t] = 0$ for $t > 0$; the independence of the noise vector v_t and the state x_{t-1} ; and x_1 is Gaussian distribution, the predicted density $p_{t|t-1}(x_t | z_{1:t-1})$ in (2.15) is the following Gaussian density

$$p_{t|t-1}(x_t | z_{1:t-1}) = \mathcal{N}(x_t; \hat{x}_{t|t-1}, \mathbf{P}_{t|t-1})$$

where $\hat{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ are given by

$$\begin{aligned} \hat{x}_{t|t-1} &= E[x_{t|t-1}] = E[x_t | z_{1:t-1}] = E[F_{t-1}x_{t-1} + B_t u_t + v_t | z_{1:t-1}] \\ &\stackrel{(a)}{=} F_{t-1}E[x_{t-1}] + E[B_t u_t] + E[v_t] = F_{t-1}\hat{x}_{t-1} + B_t u_t \end{aligned} \quad (2.19)$$

$$\begin{aligned} \mathbf{P}_{t|t-1} &= E[(x_{t|t-1} - \hat{x}_{t|t-1})(x_{t|t-1} - \hat{x}_{t|t-1})^{Tr}] = E[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^{Tr} | z_{1:t-1}] \\ &\stackrel{(b)}{=} E[(F_{t-1}x_{t-1} - F_{t-1}\hat{x}_{t-1} + v_t)(F_{t-1}x_{t-1} - F_{t-1}\hat{x}_{t-1} + v_t)^{Tr} | z_{1:t-1}] \\ &\stackrel{(c)}{=} F_{t-1}\mathbf{P}_{t-1}F_{t-1}^{Tr} + Q_t \end{aligned} \quad (2.20)$$

where F^{Tr} denotes the transpose of F ; (a) holds since u_t , and v_t are statistically independent of $z_{1:t-1}$ and $E(v_t) = 0$; and (b) holds because of system model in (2.1) and (2.19) and (c) because v_t is statistically independent of x_{t-1} .

Update:

Having extrapolated the target state $\hat{x}_{t|t-1}$ at time t from \hat{x}_{t-1} , the KF updates the predicted state $\hat{x}_{t|t-1}$ using z_t , the measurement collected at time t , theorem 3.2 and theorem 3.3 in [6, 219-220] as follows

$$\hat{x}_t = E[x_t|z_{1:t}] = \hat{x}_{t|t-1} + W_t \nu_t \quad (2.21)$$

where the matrix W_t is the Kalman gain and the difference $\nu_t = z_t - H_t \hat{x}_{t|t-1}$ between the actual measurement z_t and the predicted measurement $H_t \hat{x}_{t|t-1}$ in (2.21) is called the measurement innovation or the residual. The matrix W_t in (2.21) is the gain that minimizes the a posteriori error covariance

$$\mathbf{P}_t = \text{cov}[x_t - \hat{x}_t]. \quad (2.22)$$

By the assumption of Gaussian measurement noise with $E[w_t] = 0$ and covariance matrix $\text{cov}[w_t] = R_t$; the independence of noise vector v_t and the state x_{t-1} ; and the linear system model in (2.17) and (2.18), the covariance \mathbf{P}_t in (2.22) becomes

$$\begin{aligned} \mathbf{P}_t &= \text{cov}[x_t - \hat{x}_t] = \text{cov}[x_t - (\hat{x}_{t|t-1} + W_t \nu_t)] = \text{cov}[x_t - (\hat{x}_{t|t-1} + W_t(z_t - H_t \hat{x}_{t|t-1}))] \\ &= \text{cov}[x_t - \hat{x}_{t|t-1} - W_t(H_t x_t + w_t - H_t \hat{x}_{t|t-1})] \\ &= (I - W_t H_t) \text{cov}[x_t - \hat{x}_{t|t-1}] (I - W_t H_t)^{Tr} + W_t \text{cov}[w_t] W_t^{Tr} \\ &= (I - W_t H_t) \mathbf{P}_{t|t-1} (I - W_t H_t)^{Tr} + W_t R_t W_t^{Tr}. \end{aligned} \quad (2.23)$$

Since $\frac{d^2}{dW_t^2} \mathbf{P}_t = R_t \geq 0$, \mathbf{P}_t in (2.23) is minimum when $\frac{d}{dW_t} \mathbf{P}_t = 0$. It follows that

$$W_t = \mathbf{P}_{t|t-1} H_t^{Tr} (H_t \mathbf{P}_{t|t-1} H_t^{Tr} + R_t)^{-1}. \quad (2.24)$$

The covariance of the measurement innovation $z_t - H_t \hat{x}_{t|t-1}$ can be computed as follows

$$S_t = H_t \mathbf{P}_{t|t-1} H_t^{Tr} + R_t. \quad (2.25)$$

Substitute (2.24) and (2.25) into (2.23), the covariance matrix \mathbf{P}_t can be rewritten as follows

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - W_t S_t^{-1} W_t^{Tr}. \quad (2.26)$$

Thus the posterior density at time t is a Gaussian of the form [69, p.335-336]

$$p_{t|1:t}(x_t|z_{1:t}) = \mathcal{N}(x_t; \hat{x}_t, \mathbf{P}_t)$$

where \hat{x}_t is given in (2.21) and \mathbf{P}_t is given in (2.26).

It can also be shown that when the Gaussian assumption of $x_1, v_m, m \geq 2$ and $w_n, n \geq 1$ are dropped, the KF is the best linear estimator which produces an estimate minimizing a mean square error [3, p.46]. For further details, the reader is referred to [3, p.46-49].

As shown when the system is a Gaussian linear system, the KF is the optimal Bayes filter. However, when the process is non-linear, other methods must be sought. In the next section, the Extended KF is presented as an approximation method.

2.3.2.2 The extended Kalman Filter (EKF)

In the EKF, the non-linear state transition and measurement models are linearized and the KF is applied to the linearized equations [3, p.195-205], [147, p.19-22]. The state transition and measurement models are given by

$$x_t = f_{t-1}(x_{t-1}, u_t) + v_t \quad (2.27)$$

$$z_t = h_t(x_t) + w_t \quad (2.28)$$

where v_t, w_t are process noise and measurement noise with covariance Q_t and R_t respectively.

The EKF assumes that initial distribution is Gaussian with mean \hat{x}_1 and covariance \mathbf{P}_1 and is constructed in two steps: Prediction and Update

Prediction:

At each time step t , the process model (2.27) is linearized around the previous estimate \hat{x}_{t-1} using a first order Taylor series expansion,

$$x_t \approx f_{t-1}(\hat{x}_{t-1}, u_t) + F_{t-1}(x_{t-1} - \hat{x}_{t-1}) + v_t \quad (2.29)$$

where $F_{t-1} = \frac{\partial f_{t-1}}{\partial x}(\hat{x}_{t-1}, u_t)$. with the gradient derivative $\frac{\partial f_{t-1}}{\partial x}(\hat{x}_{t-1}, u_t)$. The expansion (2.29) ignores the high order terms because x_{t-1} is assumed to be close to the \hat{x}_{t-1} . Applying the linear KF prediction formula, the predicted state $\hat{x}_{t|t-1}$ and predicted estimate covariance $\mathbf{P}_{t|t-1}$ are

$$\hat{x}_{t|t-1} = F_{t-1}\hat{x}_{t-1} + f_{t-1}(\hat{x}_{t-1}, u_t) - F_{t-1}\hat{x}_{t-1} = f_{t-1}(\hat{x}_{t-1}, u_t) \quad (2.30)$$

$$\mathbf{P}_{t|t-1} = F_{t-1}\mathbf{P}_{t-1}F_{t-1}^\top + Q_{t-1}. \quad (2.31)$$

Hence the predicted distribution $p_{t|t-1}(x_t|z_{1:t})$ is approximated by a Gaussian with mean $\hat{x}_{t|t-1}$ and covariance $\mathbf{P}_{t|t-1}$.

Update:

At each time step t , the measurement model (2.28) is linearized around the estimate $\hat{x}_{t|t-1}$ using a first order Taylor series expansion,

$$z_t \approx h_t(\hat{x}_{t|t-1}) + H_t(x_t - \hat{x}_{t|t-1}) + w_t \quad (2.32)$$

where $H_t = \frac{\partial h_t}{\partial x}(\hat{x}_{t|t-1})$. The high order terms are ignored in the expansion (2.32) because x_t is assumed to be close to the $\hat{x}_{t|t-1}$. Applying the linear KF update formula, the predicted

measurement, the update state estimate \hat{x}_t and update estimate covariance \mathbf{P}_t are

$$\begin{aligned} z_{t|t-1} &= H_t \hat{x}_{t|t-1}^T + \mathfrak{h}_t(\hat{x}_{t|t-1}) - H_t \hat{x}_{t|t-1}^T = \mathfrak{h}_t(\hat{x}_{t|t-1}) \\ \hat{x}_t &= \hat{x}_{t|t-1} + W_t \nu_t = \hat{x}_{t|t-1} + W_t (z_t - z_{t|t-1}) \end{aligned} \quad (2.33)$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - W_t S_t^{-1} W_t^T \quad (2.34)$$

where W_t and S_t are given as for the KF in (2.24) and (2.26) respectively.

At each time $t \geq 1$, the EKF approximates the posterior distribution $p_t(x_t|z_{1:t})$ with a Gaussian with mean \hat{x}_t and covariance \mathbf{P}_t by linearizing $\mathfrak{f}_t(\cdot)$ and $\mathfrak{h}_t(\cdot)$ in the non-linear system models (2.27) and (2.28) respectively. If the system models are severely non-linear, the Gaussian property of the posterior distribution is violated. In these cases, the performance of EKF will be unreliable (see example in [147, p. 21-22]). Unlike its linear counterpart, Julier [80] shows that the extended Kalman filter in general is not an optimal estimator because it linearizes all non-linear system so that the KF can be applied. This linearization causes two well-know drawbacks:

1. The state estimate is unreliable if the assumption of local linearity is incorrect.
2. The derivation of the gradient derivatives $\frac{\partial \mathfrak{h}_t}{\partial x}(\hat{x}_{t|t-1})$ and/or $\frac{\partial \mathfrak{f}_t}{\partial x}(\hat{x}_{t-1}, u_t)$ is complicated and difficult in most application.

In addition, if the initial estimate of the state is wrong, or if the process is modeled incorrectly, the filter may quickly diverge. Another problem with the EKF is that the estimated covariance matrix tends to underestimate the true covariance matrix and therefore risks becoming inconsistent in the statistical sense without the addition of "stabilizing noise". Attempts to improve the extended Kalman filter led to the development of the Unscented Kalman filter (UKF) which was derived by Julier [79, 80] and detailed in the next subsection.

2.3.2.3 Unscented Kalman filter (UKF)

The assumed initial distribution is the same as for the EKF, that is an n_x dimensional Gaussian with mean \hat{x}_1 and covariance \mathbf{P}_1 . At each time $t - 1$, $t > 2$ instead of linearizing the function \mathfrak{f}_t and \mathfrak{h}_t , the UKF assumes that the posterior distribution $p_{t-1}(x_{t-1}|z_{1:t-1})$ is approximately by Gaussian with mean \hat{x}_{t-1} and covariance \mathbf{P}_{t-1} i.e. $p_{t-1}(x_{t-1}|z_{1:t-1}) \approx \mathcal{N}(x_{t-1}; \hat{x}_{t-1}, \mathbf{P}_{t-1})$. The UKF is constructed to propagate the mean and covariance to time t through the nonlinear system (2.27) and (2.28) as follows

1. $2n_x + 1$ points are chosen to capture the true mean \hat{x}_{t-1} and covariance \mathbf{P}_{t-1} to represent $\mathcal{N}(x_{t-1}; \hat{x}_{t-1}, \mathbf{P}_{t-1})$ by $\bar{x}_{t-1}^i = \hat{x}_{t-1} + \sigma^i$ for $i = 0, \dots, 2n_x$ where
 - $W_0 = \kappa_u / (n_x + \kappa_u)$, $W_i = 1/2(n_x + \kappa_u)$ are the weight associated with the i th point,
 - $\sigma^0 = 0$, $\sigma^j = \pm \left(\sqrt{(n_x + \kappa_u) \mathbf{P}_{t-1}} \right)_j$ where $\left(\sqrt{(n_x + \kappa_u) \mathbf{P}_{t-1}} \right)_j$ is the j th row of the matrix square root of $(n_x + \kappa_u) \mathbf{P}_{t-1}$, $j = 1, \dots, n_x$.
2. Each point is predicted as $\hat{x}_{t|t-1}^i = \mathfrak{f}_{t-1}(\bar{x}_{t-1}^i, u_t)$,

3. The predicted state $\hat{x}_{t|t-1}$ and its covariance $\mathbf{P}_{t|t-1}$ are calculated

$$\hat{x}_{t|t-1} = \sum_{i=0}^{2n_x} W_i \bar{x}_{t|t-1}^i \quad (2.35)$$

$$\mathbf{P}_{t|t-1} = \sum_{i=0}^{2n_x} W_i (\hat{x}_{t|t-1}^i - \hat{x}_{t|t-1})(\hat{x}_{t|t-1}^i - \hat{x}_{t|t-1}, u_t)^{Tr} + Q_t \quad (2.36)$$

4. The update state \hat{x}_t and its covariance \mathbf{P}_t are calculated as

$$\hat{x}_t = \hat{x}_{t|t-1} + W_t \nu_t, \quad (2.37)$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - W_t \mathbf{P}_{t|t-1}^{zz} W_t^{Tr} \quad (2.38)$$

where the innovation ν_t , the innovation covariance $\mathbf{P}_{t|t-1}^{zz}$ and the cross-correlation matrix $\mathbf{P}_{t|t-1}^{xz}$ and Unscented Kalman gain W_t are calculated as

$$\nu_t = z_t - z_{t|t-1}, \quad z_{t|t-1} = \sum_{i=1}^{2n_x} W_i \bar{z}_{t|t-1}^i, \quad \bar{z}_{t|t-1}^i = \mathfrak{h}_t(\hat{x}_{t|t-1}^i),$$

$$\mathbf{P}_{t|t-1}^{zz} = \sum_{i=0}^{2n_x} W_i (\bar{z}_{t|t-1}^i - z_{t|t-1})(\bar{z}_{t|t-1}^i - z_{t|t-1})^{Tr} + R_t$$

$$\mathbf{P}_{t|t-1}^{xz} = \sum_{i=0}^{2n_x} W_i (\bar{z}_{t|t-1}^i - z_{t|t-1})(\hat{x}_{t|t-1}^i - \hat{x}_{t|t-1})^{Tr}$$

$$W_t = \mathbf{P}_{t|t-1}^{xz} (\mathbf{P}_{t|t-1}^{zz})^{-1}$$

The UKF outperforms the EKF for nonlinear system when the probability distributions are monomodal or not heavily skewed. A more general filter, Particle filter, is derived to deal with systems with heavily skewed (asymmetric) and/or multimodal probability density function. The general idea behind the particle filter is to represent the posterior distribution by a set of random samples with associated weights. The advantage of this filter is that when the number of samples is very large, these samples represent the posterior distribution very well. Estimates of mean and covariance of the states or any function of the states are easily compute from these samples, and the accuracy of the estimates will improve as the number of samples increases.

2.3.3 Particle Filter

The particle filter is an approach derived to deal with non-linear non-Gaussian system where the conventional techniques fail. Since its first appearance in 1993 [61], the particle filter has become a useful approach for finding numerical solutions to estimation problem when there is no general analytic (closed form) expression for the probability distribution in (2.16). This section is based on the tutorial and overview articles in [5, 20, 46]. The particle filter is a suboptimal filter and it performs sequential Monte Carlo (SMC) estimation based on random samples or point mass

approximations to the posterior distributions in the Bayes filter [147]. A summary of the basic idea is presented in the following.

Let $l_t : \mathcal{X}^t = \mathcal{X} \times \dots \times \mathcal{X} \mapsto \mathbb{R}^{n_t}$ be a function of the parameters to be estimated. l_t is assumed to be integrable with respect to $p_{1:t}(x_{1:t}|z_{1:t})$ given in (2.13). We would like to compute

$$I(l_t) = E(l_t(x_{1:t})) = \int l_t(x_{1:t})p_{1:t}(x_{1:t}|z_{1:t})dx_{1:t}. \quad (2.39)$$

For example, if we want to estimate the mean of target state $x_{1:t}$ up to time t , then we use the function $l_t(x_{1:t}) = x_{1:t}$. The computation of (2.39) is infeasible when there is no analytic solution to integration. One can resort to Monte carlo methods by sampling the target probability distribution $p_{1:t}(x_{1:t}|z_{1:t})$. Denote

$$p_{1:t}(x_{1:t}|z_{1:t}) = \frac{\pi_t(x_{1:t}, z_{1:t})}{Z_t} \quad (2.40)$$

where the normalizing constant

$$Z_t = \int \pi_t(x_{1:t}, z_{1:t})dx_{1:t} \quad (2.41)$$

may be unknown.

The perfect Monte Carlo sampling which is used to approximate any probability distribution is summarized in Subsection 2.3.3.1. When it is difficult to sample from the probability distribution or it is only known up to a normalizing constant, then important sampling is often used and it is described in Subsection 2.3.3.2. Sequential Monte Carlo is used to reduce the computations by applying Monte Carlo sampling to the posterior distribution $p_{1:t}(x_{1:t}|z_{1:t})$ utilizing the previously drawn samples $\{x_{1:t-1}^n, n = 1, \dots, N\}$. It is presented in Subsection 2.3.3.3 .

2.3.3.1 Perfect Monte Carlo Sampling

If $p_{1:t}(x_{1:t}|z_{1:t})$ can be sampled, then Monte Carlo sampling is carried out by drawing N independent and identically distributed samples $\{x_{1:t}^n\}_{n=1}^N$ from the distribution $p_{1:t}(x_{1:t}|z_{1:t})$. Based on the samples the distribution $p_{1:t}(\cdot|z_{1:t})$ can be approximated by

$$\frac{1}{N} \sum_{n=1}^N \delta(x_{1:t}^n - x_{1:t})$$

where $\delta(\cdot)$ is the Dirac delta function. The expectation of a function $l_t : \mathcal{X}^t = \mathcal{X} \times \dots \times \mathcal{X} \mapsto \mathbb{R}^{n_t}$ given by

$$I = \int l_t(x_{1:t})p_{1:t}(x_{1:t}|z_{1:t})dx_{1:t} \quad (2.42)$$

can be estimated by $\frac{1}{N} \sum_{n=1}^N l_t(x_{1:t}^n)$. By the law of large numbers, $\frac{1}{N} \sum_{n=1}^N l_t(x_{1:t}^n)$ will under natural conditions almost surely converge to I [43], i.e.

$$\frac{1}{N} \sum_{n=1}^N l_t(x_{1:t}^n) \xrightarrow[n \rightarrow \infty]{a.s.} I.$$

Moreover, the rate of convergence is only dependent on the number of samples N but it is not dependent on the dimension of $x_{1:t}$.

Often the distribution is only known up to a proportionally constant. This is for example the case for the posterior distribution $p_{1:t}(x_{1:t}|z_{1:t})$ in Bayes filter. A possible solution for this type of distributions is to use importance sampling which is described next.

2.3.3.2 Importance sampling

Consider now the case when it is difficult to sample from the distribution $p_{1:t}(x_{1:t}|z_{1:t})$ in (2.40). The distribution $p_{1:t}(x_{1:t}|z_{1:t})$ is general of the non-standard form such as a non-linear non-Gaussian model, a mixture model or the product of non-linear non-Gaussian models/mixture models. Computing this distribution in closed-form may be intractable so resorting to the numerical methods to sample from this distribution is an option. The idea behind importance sampling (IS) methods is to sample from a distribution $q_{1:t}(x_{1:t}|z_{1:t})$ instead of $p_{1:t}(x_{1:t}|z_{1:t})$ where the support⁵ of $q_{1:t}(x_{1:t}|z_{1:t})$ contains the support of $p_{1:t}(x_{1:t}|z_{1:t})$ [43]. $q_{1:t}(x_{1:t}|z_{1:t})$ is called an importance distribution. We have that

$$\begin{aligned} I(l_t) &= E(l_t(x_{1:t})) = \int l(x_{1:t}) p_{1:t}(x_{1:t}|z_{1:t}) dx_{1:t} = \frac{\int l(x_{1:t}) \pi_t(x_{1:t}, z_{1:t}) dx_{1:t}}{\int \pi_t(x_{1:t}, z_{1:t}) dx_{1:t}} \\ &= \frac{\int l(x_{1:t}) w_t(x_{1:t}) q_{1:t}(x_{1:t}|z_{1:t}) dx_{1:t}}{\int w_t(x_{1:t}) q_{1:t}(x_{1:t}|z_{1:t}) dx_{1:t}} \end{aligned} \quad (2.43)$$

where $w_t(x_{1:t})$ is known as the importance weight and is given by

$$w_t(x_{1:t}) = \frac{\pi_t(x_{1:t}, z_{1:t})}{q_{1:t}(x_{1:t}|z_{1:t})}, \quad (2.44)$$

and by (2.41),

$$E_{q_{1:t}}[w_t(x_{1:t})] = \int w_t(x_{1:t}) q_{1:t}(x_{1:t}|z_{1:t}) dx_{1:t} = Z_t. \quad (2.45)$$

An approximation of Z_t is

$$\hat{Z}_t(N) = \frac{1}{N} \sum_{n=1}^N w_t(x_{1:t}^n). \quad (2.46)$$

⁵the support of a function is the set of domain where the function is not zero.

Therefore, if N weighted particles $\{x_{1:t}^n, w_t(x_{1:t}^n)\}_{n=1}^N$ are i.i.d. samples from the importance distribution $q_{1:t}(x_{1:t}|z_{1:t})$, then the distribution $\pi_t(x_{1:t}, z_{1:t})$ can be approximated by

$$\frac{1}{N} \sum_{n=1}^N w_t(x_{1:t}^n) \delta(x_{1:t} - x_{1:t}^n).$$

By a weighted particle $\{x_{1:t}^n, w_t(x_{1:t}^n)\}_{n=1}^N$ we understand that $x_{1:t}^n$ are sampled from the importance distribution $q_{1:t}(x_{1:t}|z_{1:t})$ and $w_t(x_{1:t}^n)$ are computed according to (2.44). Since by (2.40) the posterior distribution $p_{1:t}(x_{1:t}|z_{1:t})$ is proportional to $\pi_t(x_{1:t}, z_{1:t})$, the distribution $p_{1:t}(x_{1:t}|z_{1:t})$ can be approximated by

$$P_N(x_{1:t}|z_{1:t}) = \sum_{n=1}^N \bar{w}_t(x_{1:t}^n) \delta(x_{1:t} - x_{1:t}^n) \quad (2.47)$$

where the normalized importance weights $\bar{w}_t(x_{1:t}^n)$ are

$$\bar{w}_t(x_{1:t}^n) = \frac{w_t(x_{1:t}^n)}{\sum_{i=1}^N w_t(x_{1:t}^i)} \quad (2.48)$$

Thus, an estimate of $I(l_t)$ is

$$I_N(l_t) = \int l_t(x_{1:t}) P_N(x_{1:t}|z_{1:t}) dx_{1:t} = \sum_{n=1}^N l_t(x_{1:t}^n) \bar{w}_t(x_{1:t}^n) \quad (2.49)$$

If the mean and variance of $l_t(x_{1:t})$ satisfy $I(l_t) < \infty, \sigma_{l_t}^2 = \text{var}[l_t(x_{1:t})] < \infty$ then from the strong law of large numbers, $I_N(l_t)$ converges almost surely to $I(l_t)$, that is [57, theorem 1]

$$I_N(l_t) \xrightarrow[a.s.]{N \rightarrow \infty} I(l_t). \quad (2.50)$$

Moreover, under these conditions the central limit theorem says that $\sqrt{N}[I_N(l_t) - I(l_t)]$ will converge in distribution to the standard normal distribution $\mathcal{N}(0, \sigma_{l_t}^2)$ as n approaches infinity [57, theorem 2], i.e.

$$\sqrt{N}[I_N(l_t) - I(l_t)] \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, \sigma_{l_t}^2). \quad (2.51)$$

where $\sigma_{l_t}^2$ is given by [57]

$$\sigma_{l_t}^2 = \int [l_t(x_{1:t}) - I(l_t)]^2 \frac{p^2(x_{1:t}|z_{1:t})}{q_t(x_{1:t}|z_{1:t})} dx_{1:t} \quad (2.52)$$

From (2.51), the convergence rate of this estimate does not depend on the dimension of the integrand. It only depends on the sample size N . In order to reduce the number of computations, the IS is modified to compute an estimate $P_N(x_{1:t}|z_{1:t})$ of $p_{1:t}(x_{1:t}|z_{1:t})$ by reusing the past samples $\{x_{1:t-1}^n, n = 1, \dots, N\}$. This modification called Sequential Importance Sampling (SIS) uses an

importance distribution of the following form

$$q_{1:t}(x_{1:t}|z_{1:t}) = q_{1:t-1}(x_{1:t-1}|z_{1:t-1})q_t(x_t|x_{1:t-1}, z_{1:t}) \quad (2.53)$$

$$= q_1(x_1) \prod_{i=2}^t q_i(x_i|x_{1:i-1}, z_i) \quad (2.54)$$

since conditionally on $x_{1:t-1}$, x_t is statistically independent of $z_{1:t-1}$.

2.3.3.3 Sequential importance sampling

A sample $x_{1:t}$ can be generated recursively in time by a sample $x_t \sim q_t(x_t|x_{1:t-1}, z_t)$ at present time t and then use (2.53). The importance weight $w_t(x_{1:t})$ is calculated recursively

$$\begin{aligned} w_t(x_{1:t}) &= \frac{\pi_t(x_{1:t}, z_{1:t})}{q_{1:t}(x_{1:t}|z_{1:t})} = \frac{\pi_{t-1}(x_{1:t-1}, z_{1:t-1}) \pi_t(x_{1:t}, z_{1:t}) q_{1:t-1}(x_{1:t-1}|z_{1:t-1})}{q_{1:t-1}(x_{1:t-1}|z_{1:t-1}) \pi_{t-1}(x_{1:t-1}, z_{1:t-1}) q_{1:t}(x_{1:t}|z_{1:t})} \\ &\stackrel{(a)}{=} w_{t-1}(x_{1:t-1}) \frac{p_t(x_t, z_t|x_{1:t-1}, z_{1:t-1})}{q_t(x_t|x_{1:t-1}, z_t)} \\ &= w_1(x_1) \prod_{i=2}^t \frac{p_i(x_i, z_i|x_{1:i-1}, z_{1:i-1})}{q_i(x_i|x_{1:i-1}, z_i)} \end{aligned} \quad (2.55)$$

where (a) holds because of (2.53) and

$$\pi_t(x_{1:t}, z_{1:t}) = p_t(z_t, x_t|x_{1:t-1}, z_{1:t-1})\pi_{t-1}(x_{1:t-1}, z_{1:t-1})$$

The SIS algorithm is summarized below where each time t , x_t^n , $n = 1, \dots, N$ samples are generated.

Algorithm 1 : Sequential Importance Sampling

- At time $t = 1$: draw the samples $x_1^n \sim q_1(x_1)$, $n = 1, \dots, N$ and compute the weight

$$w_1(x_1^n) = \frac{p(x_1^n)}{q_1(x_1)} \quad (2.56)$$

and the normalized weights $\bar{w}_1^n = \frac{w_1(x_1^n)}{\sum_{n=1}^N w_1(x_1^n)}$.

- At time $t = 2, \dots, T$: draw the samples $x_t^n \sim q_t(x_t|x_{1:t-1}^n)$, $n = 1, \dots, N$ and the weights are calculated as in (2.55).

$$w_t(x_{1:t}^n) = w_{t-1}(x_{1:t-1}^n) \frac{p_t(x_t, z_t|x_{1:t-1}^n, z_{1:t-1})}{q_t(x_t^n|x_{1:t-1}^n, z_{1:t})} \quad (2.57)$$

and the normalized weights

$$\bar{w}_t^n = \frac{w_t(x_{1:t}^n)}{\sum_{n=1}^N w_t(x_{1:t}^n)} \quad (2.58)$$

Degeneracy problem

Let $\alpha_t(x_{1:t}) = \frac{p_t(x_t, z_t | x_{1:t-1}, z_{1:t-1})}{q_t(x_t | x_{1:t-1}, z_t)}$, (2.55) can be written as follows

$$w_t(x_{1:t}) = w_1(x_1) \prod_{i=2}^t \alpha_i(x_{1:i}). \quad (2.59)$$

Then the variance of $\widehat{Z}_t(N)$ is

$$\begin{aligned} \text{Var}[\widehat{Z}_t(N)] &= E[(\widehat{Z}_t(N) - E[\widehat{Z}_t(N)])^2] \\ &= E\left[\left(\frac{1}{N} \sum_{n=1}^N w_t(x_{1:t}^n) - \frac{1}{N} Z_t\right)^2\right] = \frac{1}{N^2} E\left[\left(\sum_{n=1}^N (w_t(x_{1:t}^n) - Z_t)\right)^2\right] \\ &= \frac{1}{N^2} \sum_{n=1}^N \text{Var}[w_t(x_{1:t}^n)] = \frac{1}{N} \text{Var}[(w_t(x_{1:t}))] \\ &= \frac{1}{N} \left(E[w_t^2(x_{1:t})] - Z_t^2\right) \end{aligned} \quad (2.60)$$

Then the ratio between the variance of $\widehat{Z}_t(N)$ and Z_t^2 is

$$\frac{\text{Var}[\widehat{Z}_t(N)]}{Z_t^2} = \frac{1}{N} \left(\frac{1}{Z_t^2} E[w_t^2(x_{1:t})] - 1\right). \quad (2.61)$$

Hence, we have

$$\begin{aligned} \frac{\text{Var}[\widehat{Z}_{t+1}(N)]}{Z_{t+1}^2} &= \frac{1}{N} \left(\frac{1}{Z_{t+1}^2} E[w_{t+1}^2(x_{1:t+1})] - 1\right) = \frac{1}{N} \left(\frac{E[w_t^2(x_{1:t}) \alpha_{t+1}^2(x_{1:t+1})]}{Z_t^2 \widehat{\alpha}_{t+1}^2(x_{1:t+1})} - 1\right) \\ &= \frac{1}{N} \left(\frac{E[w_t^2(x_{1:t})]}{Z_t^2} \frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} - 1\right) \end{aligned} \quad (2.62)$$

where $\widehat{\alpha}_{t+1}(x_{1:t+1}) = E_q[\alpha_{t+1}(x_{1:t+1})] = \int \alpha_{t+1}(x_{1:t+1}) q_{t+1}(x_{t+1} | x_{1:t}, z_{t+1}) dx_{t+1}$.

By $\text{Var}[\alpha_{t+1}(x_{1:t+1})] = E[\alpha_{t+1}^2(x_{1:t+1})] - \widehat{\alpha}_{t+1}^2(x_{1:t+1}) > 0$,

and $E[\alpha_{t+1}^2(x_{1:t+1})], \widehat{\alpha}_{t+1}^2(x_{1:t+1}) > 0$, the following is true

$$\frac{E[\alpha_{t+1}(x_{1:t+1})]}{\widehat{\alpha}_{t+1}(x_{1:t+1})} > 1. \quad (2.63)$$

Therefore,

$$\begin{aligned} \frac{\text{Var}[\widehat{Z}_{t+1}(N)]}{Z_{t+1}^2} &= \frac{1}{N} \left(\frac{E[w_t^2(x_{1:t})]}{Z_t^2} \frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} - 1\right) \\ &= \frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} \frac{1}{N} \left(\frac{E[w_t^2(x_{1:t})]}{Z_t^2} - 1\right) + \frac{1}{N} \left(\frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} - 1\right) \\ &= \frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} \frac{\text{Var}[\widehat{Z}_t(N)]}{Z_t^2} + \frac{1}{N} \left(\frac{E[\alpha_{t+1}^2(x_{1:t+1})]}{\widehat{\alpha}_{t+1}^2(x_{1:t+1})} - 1\right) \\ &> \frac{\text{Var}[\widehat{Z}_t(N)]}{Z_t^2}. \end{aligned} \quad (2.64)$$

From (2.64), it follows that the variance of resulting estimates increases exponentially over time t . As a result, after some iterations only one particle survives. Thus resampling technique addresses this problem and is introduced in the next section.

2.3.3.4 Resampling

Resampling is a technique which helps to solve the degeneracy problem by eliminating the low weighted samples and multiplying the high weighted samples. Consider $P_N(x_{1:t}|z_{1:t})$ given in (2.47) which is the IS approximation of the distribution $p_{1:t}(x_{1:t}|z_{1:t})$. This IS approximation is based on the weighted samples which are drawn from the important sampling $q_t(x_{1:t}|z_{1:t})$ and may not be a good approximation of $p_{1:t}(x_{1:t}|z_{1:t})$. As a result, the degeneracy problem can arise as shown earlier. To obtain samples which are approximately distributed according to the distribution $p_{1:t}(x_{1:t}|z_{1:t})$, resampling method is applied by simply drawing new samples based on the IS approximation $P_N(x_{1:t}|z_{1:t})$ given in (2.47) and use these resampled values in the approximation of the distribution $p_{1:t}(x_{1:t}|z_{1:t})$. There are many resampling methods in statistics. Four popular approaches in the literature are

- Stratified resampling [52]
- Systematic resampling [83]
- Residual resampling [87]
- Multinomial resampling [77, p.31-92], [13, p.441-448]

The comparison between the resampling techniques is discussed in [42]. Some version of particle filters are derived from SIS by choosing an appropriate choice of sampling distribution and/or modifying the resampling step in order to overcome the degeneracy problem. Examples are Bootstrap filter [61], Regularized Particle filter [124], auxiliary resampling importance resampling [139], and local linearization particle filter [45, 167].

Chapter 3

Random Finite Set (RFS) for Filtering

Random finite sets play a crucial role in the generalization of the single target Bayes filter to the multi-target Bayes filter which is used in multi-target tracking. A Random finite set (RFS) is a finite-set-valued random variable. An RFS is different from a random vector in two important aspects: 1) the number of elements in an RFS is random while the number of entries in a vector is fixed, 2) the order of the elements are irrelevant for an RFS, but the order of the entries in a vector is important. Another property which distinguishes an RFS from a random vector is that the elements constituting the random vector may be the same but all the elements in an RFS are different.

The purpose of this section is twofold. Firstly, some key concepts in the theory of RFSs are introduced in Subsection 3.1. Then Subsection 3.2 models the multi-target tracking problem in an RFS framework, and it also introduces multi-target Bayes filtering and estimation in an RFS framework. Definitions of the mathematical concepts used in this Chapter are given in Appendix A.1. Most of the material in this Chapter is based on [60], [101], and [40, p.111-156]

3.1 Background on Random Finite Set

This section briefly provides the basic concepts of the RFS which are used in the formulation of the multi-target tracking problem. A random set is introduced by generalizing the notion of a random vector to a random set.

3.1.1 Mathematical Preliminaries

Let $(\Omega, \sigma(\Omega), \mathbb{P})$ be a probability space and (U, \mathcal{U}) be an abstract measurable space. Typically \mathcal{U} will be the Borel σ -algebra generated by the open sets of U , i.e. the smallest σ -algebra containing all open sets of U . A random element is defined as a measurable mapping $\Sigma : \Omega \rightarrow U$. Since the mapping is measurable, $\Sigma^{-1}(A) \in \sigma(\Omega)$ for any $A \in \mathcal{U}$. The probability measure on U induced by Σ is $\mathbf{P}_\Sigma = \mathbb{P} \circ \Sigma^{-1} : \mathcal{U} \rightarrow [0, 1]$ where $\mathbf{P}_\Sigma(A) = \mathbb{P} \circ \Sigma^{-1}(A) = \mathbb{P}(\Sigma^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : \Sigma(\omega) \in A\})$ for $A \in \mathcal{U}$. It is common to use $\mathbb{P}(\Sigma \in A) = \mathbb{P}(\{\omega \in \Omega : \Sigma(\omega) \in A\})$. Different spaces U give different random elements.

- When $U = \mathbb{R}$ (resp. \mathbb{R}^d), the random element is called a random variable (resp. a random vector)
- When U is a class of subsets of some space \mathcal{X} (e.g. $\mathcal{X} \subset \mathbb{R}^d$, $d \in \{1, 2, \dots\}$), the random element Σ is called a random set. When U is a collection of finite sets of a topological space $(\mathcal{X}, \mathbb{T}(\mathcal{X}))$, Σ is called an RFS.

An RFS can be defined mathematically as follows

Definition 3.1: A *random finite set* Σ is a measurable mapping from Ω to $\mathcal{F}(\mathcal{X})$

$$\Sigma : \Omega \rightarrow \mathcal{F}(\mathcal{X})$$

where $\mathcal{F}(\mathcal{X})$ is the space of all finite subsets of \mathcal{X} equipped with the myope topology¹ [109, p.3-5].

In multi-target tracking problem, we often deal with a space which is the product of continuous space (e.g. the state space in tracking) and a discrete space (e.g. the space of target labels in tracking). Such a space is called a hybrid space. In the single target tracking problem, the discrete space reduces to a singleton set and hence the hybrid space reduces to the state space. The following section deals with hybrid space in general.

3.1.1.1 Hybrid Space

A hybrid space is the Cartesian product of \mathcal{X} and a finite discrete space \mathcal{K} i.e. $\mathbb{X} = \mathcal{X} \times \mathcal{K}$. An element $s = (x, k) \in \mathbb{X}$ consists of an Euclidean part $x \in \mathcal{X} \subseteq \mathbb{R}^d$ and a discrete part $k \in \mathcal{K}$. Let $S \subseteq \mathbb{X}$. Then for any $k \in \mathcal{K}$, we define $S(k) = \{x \in \mathcal{X} : (x, k) \in S\}$ so $S = \bigcup_{k \in \mathcal{K}} S(k) \times \{k\}$. Here $S(k_1) \times \{k_1\}$ and $S(k_2) \times \{k_2\}$ are disjoint if $k_1 \neq k_2$. In general, for any $S \subseteq \mathbb{X}^n = \mathbb{X} \times \dots \times \mathbb{X}$, $S(k_1, \dots, k_n)$ is defined as $S(k_1, \dots, k_n) = \{(x_1, \dots, x_n) \in \mathcal{X}^n : ((x_1, k_1), \dots, (x_n, k_n)) \in S\}$ so $S = \bigcup_{(k_1, \dots, k_n) \in \mathcal{K}^n} S(k_1, \dots, k_n) \times \{(k_1, \dots, k_n)\}$.

The RFS on the hybrid space is defined in the obvious fashion by extending Definition 3.1 from the state space to the hybrid space as follows.

Definition 3.2: A *random finite set* Σ on a hybrid space is a measurable mapping from Ω to $\mathcal{F}(\mathbb{X})$

$$\Sigma : \Omega \rightarrow \mathcal{F}(\mathbb{X})$$

where $\mathcal{F}(\mathbb{X})$ is the space of all finite subsets of \mathbb{X} equipped with the myope topology on the hybrid space \mathbb{X} [60, p.137].

Denote by $C^{\mathbb{X}}$ the collection of all closed subsets of \mathbb{X} , by $K^{\mathbb{X}}$ the collection of all compact subsets of \mathbb{X} and by $G^{\mathbb{X}}$ the collection of all open subsets of \mathbb{X} . The myope topology is defined as follows. For any open subset $G \subseteq \mathbb{X}$ and any compact subset $K \subseteq \mathbb{X}$, define the collections of

¹Intuitive explanation of this topology can be found in [60, p.94], [101, p.712] or [168, p.47].

closed subsets *hitting* G and *missing* K as

$$\mathcal{A}_G = \{S \in C^{\mathbb{X}} : S \cap G \neq \emptyset\} \quad (3.1)$$

$$\mathcal{A}^K = \{S \in C^{\mathbb{X}} : S \cap K = \emptyset\} \quad (3.2)$$

respectively. The myope topology has the base²

$$\mathbf{B}_G^K = \{\mathcal{A}^K \cap \mathcal{A}_{G_1} \cap \dots \cap \mathcal{A}_{G_n} : n \geq 0, K \in K^{\mathbb{X}}, G_i \in G^{\mathbb{X}}\}. \quad (3.3)$$

This is called the hit-or-miss topology.

Denote by $\mathcal{F}_n(\mathbb{X})$ the collection of all finite subsets of \mathbb{X} which contain exactly n elements. If $n = 0$, then $\mathcal{F}_n(\mathbb{X}) = \{\emptyset\}$. It can be shown [60, p.132] that $\mathcal{F}_n(\mathbb{X}) \in \sigma(C^{\mathbb{X}})$ and hence $\mathcal{F}_n(\mathbb{X})$ is measurable with respect to $\sigma(C^{\mathbb{X}})$.

The closed (resp. open, compact) subsets of \mathbb{X}^n are those $S \subseteq \mathbb{X}^n$ such that $S(k_1, \dots, k_n)$ are closed (resp. open, compact) for all $(k_1, \dots, k_n) \in \mathcal{K}^n$ [60, p.135]. The hybrid space has a topology which is the product topology of the Euclidean on \mathcal{X} and the discrete topology on \mathcal{K} . This means that for any open subset $S \subseteq \mathbb{X}$, $S(k)$ is open for all $k \in \mathcal{K}$.

Definition 3.3: *The product measure $\bar{\lambda} = \lambda \times \bar{c}$ on the space \mathbb{X} is referred to as the (unit) hybrid Lebesgue measure where λ is the (unit) Lebesgue measure on \mathcal{X} and \bar{c} is the counting measure on \mathcal{K} . We say that a set $S \subseteq \mathbb{X}$ is measurable if $S(k)$ is Lebesgue-measurable for every $k \in \mathcal{K}$. Then the hybrid measure is defined by*

$$\bar{\lambda}(S) = \sum_{k \in \mathcal{K}} \lambda(S(k)) \quad (3.4)$$

Generally, $S \subseteq \mathbb{X}^n$ is measurable if $S(k_1, \dots, k_n)$ is measurable and

$$\bar{\lambda}^n(S) = \sum_{(k_1, \dots, k_n) \in \mathcal{K}^n} \lambda^n(S(k_1, \dots, k_n)) \quad (3.5)$$

where $\bar{\lambda}^1(S) = \bar{\lambda}(S)$ for $S \subseteq \mathbb{X}$.

Suppose that volume in the space \mathcal{X} is measured in units of K_x , then $\lambda(\Delta_x)$ is the volume Lebesgue measure of a neighborhood of Δ_x of x in units of K_x , and $\bar{\lambda}(\Delta_x \times \{k\})$ is the volume hybrid Lebesgue measure of a neighborhood of $\Delta_x \times \{k\}$ of (x, k) in units of K_x .

In an obvious way, the concept of Lebesgue integral is extended to the hybrid space [60, p.136],[168, p.50]

Definition 3.4: a) $f : \mathbb{X}^n \rightarrow \mathbb{R}^m$ is an integrable function if and only if the functions $f_{k_1, \dots, k_n} : \mathcal{X}^n \rightarrow \mathbb{R}^m$ defined by $f_{k_1, \dots, k_n}(x_1, \dots, x_n) = f((x_1, k_1), \dots, (x_n, k_n))$ are Lebesgue-integrable for every $(k_1, \dots, k_n) \in \mathcal{K}^n$

²The definition of a base in Appendix A.2.

- b) Let $S \subset \mathbb{X}^n$ be measurable and $\xi_i = (x_i, k_i)$ for $i = 1, \dots, n$. Then for each integrable f , the hybrid integral of f on S is

$$\begin{aligned} & \int_S f(\xi_1, \dots, \xi_n) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \\ &= \sum_{(k_1, \dots, k_n) \in \mathcal{K}^n} \int_{S(k_1, \dots, k_n)} f_{k_1, \dots, k_n}(x_1, \dots, x_n) \lambda(dx_1) \dots \lambda(dx_n) \end{aligned}$$

The concept of a probability measure is also extended to the hybrid space [60, Definition 6, p. 136]

Definition 3.5: A set function \mathbf{P} defined on the measurable subsets S of \mathbb{X} is a probability measure if it has the form $\mathbf{P}(S) = \mathbb{P}(\{\omega \in \Omega : \xi(\omega) \in S\})$ where $\xi = (x, k)$ is a random variable on \mathbb{X} . Then the set functions $\mathbf{P}_k(U) = \mathbf{P}(U \times \{k\}) = \mathbf{P}(x \in U, k \in \{k\})$ are measures³ on \mathcal{X} for any measurable subset U of \mathcal{X} and for any $k \in \mathcal{K}$. Since $S = \uplus_{k \in \mathcal{K}} S(k) \times \{k\}$, we have $\mathbf{P}(S) = \sum_{k \in \mathcal{K}} \mathbf{P}_k(S(k))$ where \uplus denotes the disjoint union operator.

This definition allows us to transform between the probability measure on the hybrid space and the measure on \mathcal{X} . A set derivative also exists on the hybrid space if \mathbf{P} is absolutely continuous with respect to the hybrid measure

Proposition 3.1: Let \mathbf{P} be a probability measure as defined in Definition 3.5. If \mathbf{P} is absolutely continuous with respect to the hybrid measure $\bar{\lambda}$, then there exists an almost everywhere integrable unique function f on \mathbb{X} such that for any measurable subset $S \subseteq \mathbb{X}$

$$\mathbf{P}(S) = \int_S f(\xi) \bar{\lambda}(d\xi) \quad (3.6)$$

3.1.2 Measure and Integral of RFSs

This section aims to outline the construction of the measure and the integral of RFSs which are based on the conventional probability. This section is based on [171] and [101, p.711-716].

Let Σ be a measurable mapping from Ω to $\mathcal{F}(\mathbb{X})$ as in Definition 3.2. Σ induces a probability measure \mathbf{P}_Σ on \mathbb{X} which is defined for any Borel subset \mathcal{O} of $\mathcal{F}(\mathbb{X})$

$$\mathbf{P}_\Sigma(\mathcal{O}) = \mathbb{P}(\Sigma^{-1}(\mathcal{O})) = \mathbb{P}(\{\omega \in \Omega : \Sigma(\omega) \in \mathcal{O}\}). \quad (3.7)$$

Denote $\chi : \uplus_{n=0}^{\infty} \mathbb{X}^n \rightarrow \mathcal{F}(\mathbb{X})$ as the mapping of vectors to finite sets defined for each n by $\chi(\xi_1, \dots, \xi_n) = \{\xi_1, \dots, \xi_n\}$ where $\mathbb{X}^0 = \{\emptyset\}$. Then for any Borel set $\mathcal{O} \subseteq \mathcal{F}(\mathbb{X})$, the measure μ is defined as

$$\mu(\mathcal{O}) = \sum_{n=0}^{\infty} \frac{\bar{\lambda}^n(\chi^{-1}(\mathcal{O}) \cap \mathbb{X}^n)}{n! K_x^n} \quad (3.8)$$

³This is not a probability measure as mentioned in [60, Definition 6, p. 136] otherwise \mathbf{P} cannot be a probability measure

Note that the term K_x^n in each sum will cancel out with K_x^n of the hybrid measure

$$\bar{\lambda}^n \left(\chi^{-1}(\mathcal{O}) \cap \mathbb{X}^n \right).$$

Assume that the probability measure \mathbf{P}_Σ induced by the RFS Σ is absolutely continuous with respect to the measure μ . By the Radon Nikodým theorem there exist an almost everywhere unique integrable function $g_\Sigma : \mathcal{F}(\mathbb{X}) \rightarrow [0, \infty)$ such that

$$\mathbf{P}_\Sigma(\mathcal{O}) = \int_{\mathcal{O}} g_\Sigma(Z) \mu(dZ) \quad (3.9)$$

By the definition of the measure μ in (3.8), (3.9) can be rewritten as

$$\int_{\mathcal{O}} g_\Sigma(Z) \mu(dZ) = \sum_{n=0}^{\infty} \int_{\mathcal{O} \cap \mathcal{F}_n(\mathbb{X})} g_\Sigma(Z) \mu(dZ) \quad (3.10)$$

$$= \frac{1}{n!} \sum_{n=0}^{\infty} \int_{\chi^{-1}(\mathcal{O}) \cap \mathbb{X}^n} g_\Sigma(\{\xi_1, \dots, \xi_n\}) K_x^{-n} \bar{\lambda}^n(d\xi_1 \dots \xi_n) \quad (3.11)$$

for any Borel set $\mathcal{O} \subseteq \mathcal{F}(\mathbb{X})$. Note that the sum of the right hand side in (3.11) holds because each term of the sum is unitless. This is because g_Σ is unitless and $\bar{\lambda}^n$ has unit of K_x^n . In the sequel a particular kind of integrable finite set function is our interest and it is defined in the following Definition

Definition 3.6 (Global density): A global density (function) is a non-negative, integrable finite set function whose total set integral is unity.

By this Definition, non-negative, integrable finite set function g_Σ is a global density because

$$\int_{\mathcal{F}(\mathbb{X})} g_\Sigma(Z) \mu(Z) = \mathbf{P}_\Sigma(\mathcal{F}(\mathbb{X})) = 1.$$

The non-negative, integrable finite set function g_Σ in (3.9) is a global density because

$$\int_{\mathcal{F}(\mathbb{X})} g_\Sigma(Z) \mu(Z) = \mathbf{P}_\Sigma(\mathcal{F}(\mathbb{X})) = 1.$$

3.1.3 Finite Set Statistic (FISST)

This section summarizes the construction of key concepts such as set derivative, set integral in the finite set statistics (FISST) formulation of the multi-target tracking problem. The global density in Definition 3.6 is a particular set derivative which is of main interest for multi-target tracking. The concepts set derivative and set integral are not normal as the normal concepts in ordinary calculus. Their definitions requires a suitable transformation between the product space \mathbb{X}^n , $n = 1, 2, \dots$ and the space $\mathcal{F}(\mathbb{X})$ (i.e. the collection of finite subsets of \mathbb{X}).

Denote by $\mathcal{F}_{\leq n}(\mathbb{X})$ the collection of all finite subsets of \mathbb{X} which contain no more than n elements. Define a mapping which transforms a set of elements in the space \mathbb{X}^n , $n > 0$ to a finite set of \mathbb{X} in the space $\mathcal{F}_{\leq n}(\mathbb{X})$ as follows

$$\begin{aligned} \chi_n : \mathbb{X}^n &\rightarrow \mathcal{F}_{\leq n}(\mathbb{X}) \\ \xi = (\xi_1, \dots, \xi_n) &\mapsto \chi(x) = \{\xi_1, \dots, \xi_n\}, \end{aligned} \quad (3.12)$$

This mapping is many-to-one. In order to make the mapping between \mathbb{X}^n and $\mathcal{F}_n(\mathbb{X})$ bijective, we define the lexicographic ordering denoted by \prec between two elements $\xi, \zeta \in \mathbb{X}$ where $\xi = (x, k_1)$, $\zeta = (y, k_2)$, $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ as follows. $\xi \prec \zeta$ if one of the following statements is true

- $k_1 < k_2$
- $k_1 = k_2$ and $x_1 < y_1$
- $k_1 = k_2$, $x_1 = y_1$ and $x_2 < y_2$
- $k_1 = k_2$, $x_i = y_i$, for $i = 1, \dots, k < d$ and $x_{k+1} < y_{k+1}$

Let $[\mathbb{X}]^n = \{(\xi_1, \dots, \xi_n) \in \mathbb{X}^n : \xi_1 \prec \xi_2 \prec \dots \prec \xi_n\}$. Then by [60, Proposition 2, p.133], the mapping $\underline{\chi}_n : [\mathbb{X}]^n \rightarrow \mathcal{F}_n(\mathbb{X})$, which is the restriction of the map χ_n to $[\mathbb{X}]^n$, is a homeomorphism (equivalence of topological spaces) between the two spaces $[\mathbb{X}]^n$ and $\mathcal{F}_n(\mathbb{X})$.

Let $f : \mathbb{X}^n \rightarrow \mathbb{R}^r$ ($r \geq 1$) be completely symmetric function⁴. Define $f^* : \mathcal{F}_n(\mathbb{X}) \rightarrow \mathbb{R}^r$ by $f^*(\{\xi_1, \dots, \xi_n\}) = f(\xi_1, \dots, \xi_n)$. By (3.12), the composite function $f^* \circ \chi_n = f$ almost everywhere where \circ denotes the composite symbol. Inversely let $F : \mathcal{F}_n(\mathbb{X}) \rightarrow \mathbb{R}^r$. Define $F^* : \mathbb{X}^n \rightarrow \mathbb{R}^r$ by $F^*(\xi_1, \dots, \xi_n) = F(\{\xi_1, \dots, \xi_n\})$ for all distinct ξ_1, \dots, ξ_n (note that F^* is undefined on a set of measure zero).

The correspondences of $f \rightarrow f^*$ and $F \rightarrow F^*$ set up a one-to-one correspondence between the measurable (resp. continuous) almost everywhere defined symmetric functions on \mathbb{X}^n and the measurable (resp. continuous) functions on $\mathcal{F}_n(\mathbb{X})$ [60, Proposition 3, p.135].

3.1.3.1 Set Derivative and Its Properties

Like ordinary calculus in which an inverse operation of the Lebesgue integral is the derivative, the set integral also has an inverse operation which is called a set derivative which is defined in Definition 3.7 below. We also summarize some basics properties of the set derivative and set integral which involves the belief measure [60, p.150-170]. These properties is useful for deriving the multi-target system model which is introduced in Section 3.2.

The following set derivative is base on [60, Definition 12, p.145-146] and [171, 4.5].

Definition 3.7 (Set Derivative): Let $\Phi : C^{\mathbb{X}} \rightarrow [0, \infty)$ be a set function⁵ on \mathbb{X} and let $\xi = (x, u) \in \mathbb{X}$. If it exists, for any closed subset S of \mathbb{X} the set derivative of Φ at ξ is the set function

⁴ $f(x_1, x_2, \dots, x_n)$ is called symmetric or totally symmetric if and only if it is invariant under any permutation of variables.

⁵ A set function is a function whose input is a set.

defined by

$$\frac{\delta\Phi}{\delta\xi}(S) = \lim_{j \rightarrow \infty} \lim_{i \rightarrow \infty} \frac{\Phi((S - (\overline{B_{x,j}} \times \{k\})) \cup (B_{x,i} \times k)) - \Phi(S - (\overline{B_{x,j}} \times \{k\}))}{\lambda(B_{x,i})} \quad (3.13)$$

where $\overline{B_{x,j}}$ is a sequence of closed balls converging to $\{x\}$; and where $B_{x,i}$ is a sequence of open balls whose closures converge to $\{x\}$.

If $\xi \notin S$, (3.13) can be written as follows

$$\frac{\delta\Phi}{\delta\xi}(S) = \lim_{j \rightarrow \infty} \lim_{i \rightarrow \infty} \frac{\Phi(S \cup (B_{x,i} \times k)) - \Phi(S)}{\lambda(B_{x,i})} \quad (3.14)$$

Note that $\frac{\delta\Phi}{\delta\xi}(S)$ has unit of K_x^{-1} because the denominator $\lambda(B_{x,i})$ in (3.13) and (3.14) has unit of K_x .

It has been shown in [60, p.147-148] that the set derivative of a set function is also a set function and thus we can iterate the process by the following definition.

Definition 3.8: For any closed subset $S \subseteq \mathbb{X}$. The iterated set derivative of order n is defined as

$$\frac{\delta^{n+1}\Phi}{\delta\xi_{n+1} \dots \delta\xi_1}(S) = \frac{\delta\Phi}{\delta\xi_{n+1}} \frac{\delta^n\Phi}{\delta\xi_n \dots \delta\xi_1}(S) \quad (3.15)$$

where $\xi_1, \dots, \xi_{n+1} \in \mathbb{X}$.

When the set derivative of order n , $n > 0$ is well defined, the order of differentiation does not matter. The concept of iterated set derivative is defined as follows

Definition 3.9: Let Φ be a set function. Let $n \geq 1$ and let $Z = \{\xi_1, \dots, \xi_n\} \subseteq \mathbb{X}$ be a finite subset with n distinct elements. Assume that all iterated set derivative of Φ exist. Then for any closed subset S of \mathbb{X} , the set derivative of Φ is defined as

$$\frac{\delta\Phi}{\delta Z}(S) = \frac{\delta^n\Phi}{\delta\xi_1 \dots \delta\xi_n}(S) \quad (3.16)$$

$$\frac{\delta\Phi}{\delta\emptyset}(S) = \Phi(S) \quad (3.17)$$

$$\frac{\delta^0\Phi}{\delta Z}(S) = \frac{\delta^0\Phi}{\delta\emptyset}(S) = \Phi(S) \quad (3.18)$$

The following proposition generalizes the sum and the product rules for the derivatives to the case of set derivatives

Proposition 3.2: Let Φ and Ψ be the set functions and let $a, b \in \mathbb{R}$ be scalars. Assume that the set derivatives of Φ and Ψ exist. Then the set derivatives of $a\Phi + b\Psi$ and $\Phi\Psi$ also exist and are given by

$$\frac{\delta(a\Phi + b\Psi)}{\delta Z}(S) = a \frac{\delta\Phi}{\delta Z}(S) + b \frac{\delta\Psi}{\delta Z}(S) \quad (3.19)$$

$$\frac{\delta(\Phi\Psi)}{\delta Z}(S) = \sum_{Y \subseteq Z} \frac{\delta\Phi}{\delta Y}(S) \frac{\delta\Psi}{\delta(Z-Y)}(S) \quad (3.20)$$

for all $S \subseteq \mathbb{X}$. If $Y, Z \subseteq \mathbb{X}$ are finite subset with $Y \cap Z = \emptyset$. Then

$$\frac{\delta}{\delta Y} \frac{\delta(\Phi)}{\delta Z}(S) = \frac{\delta \Phi}{\delta(Y \cup Z)}(S)$$

for all $S \subseteq \mathbb{X}$.

The set integral is naturally given by the following definition.

Definition 3.10: Let $\Phi : \mathcal{F}(\mathbb{X}) \rightarrow [0, \infty)$ be a set function. Let \mathbf{f} be the function defined by $\mathbf{f}(Z) = \frac{\delta \Phi}{\delta Z}(\emptyset)$ for any finite subset $Z \in \mathcal{F}(\mathbb{X})$ and let S be closed subset of \mathbb{X} . Then the **set integral** of \mathbf{f} concentrated at S is

$$\int_S \mathbf{f}(Z) \delta(Z) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{S^n} \mathbf{f}(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.21)$$

where $S^n = S \times \dots \times S$ denotes the Cartesian product of S taken n times, and we assume the convention that $S^0 = \{\emptyset\}$ and $\int_{\emptyset} \mathbf{f}(\{\emptyset\}) \bar{\lambda}(d\emptyset) = \mathbf{f}(\{\emptyset\})$. Let \mathcal{O} be any measurable subset of $C^{\mathbb{X}}$ (i.e. $\mathcal{O} \in \sigma(C^{\mathbb{X}})$). The set integral of \mathbf{f} concentrated on \mathcal{O} is defined as

$$\begin{aligned} \int_{\mathcal{O}} \mathbf{f}(Z) \delta(Z) &= \int_{\mathcal{O} \cap \bigcup_{n=0}^{\infty} \mathcal{F}_n(\mathbb{X})} \mathbf{f}(Z) \delta(Z) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\chi_n^{-1}(\mathcal{O} \cap \mathcal{F}_n(\mathbb{X}))} \mathbf{f}(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \end{aligned} \quad (3.22)$$

where $\chi_n : \mathbb{X}^n \rightarrow \mathcal{F}_n(\mathbb{X})$ is the function defined in (3.12) by $\chi_n(\xi_1, \dots, \xi_n) = \{\xi_1, \dots, \xi_n\}$ with the added condition that $\xi_i \neq \xi_j, i \neq j$ so $\chi_n^{-1}(\mathcal{O} \cap \mathcal{F}_n(\mathbb{X}))$ is a measurable subset of \mathbb{X}^n .

It can be shown that (3.21) is the special case of (3.22). Indeed, for any $\mathcal{O} \in \sigma(C^{\mathbb{X}})$, by the myope topology particularly by (3.1) there exists a closed set S such that

$$\begin{aligned} \mathcal{O} &= (\mathcal{A}_{S^c})^c = \{C \in C^{\mathbb{X}} : C \cap S^c \neq \emptyset\}^c = \{C \in C^{\mathbb{X}} : C \cap S^c = \emptyset\} \\ &= \{C \in C^{\mathbb{X}} : C \subseteq S\} \end{aligned} \quad (3.23)$$

where S^c denotes the complement of S so S^c is open. Then for $n = 0, 1, \dots$

$$\chi_n^{-1}(\mathcal{O} \cap \mathcal{F}_n(\mathbb{X})) = S^n. \quad (3.24)$$

By convention, denote $\chi_0^{-1}(\mathcal{O} \cap \mathcal{F}_0(\mathbb{X})) = S^0$.

3.1.3.2 Belief Functional and Global Density

This section introduces a method for construction of the global density function of an absolutely continuous RFS. A direct construction from the probability measure is impractical in most cases. However, it can also be derived from the belief functional defined below. The relationship between

the belief functional and the global density function is derived by applying the set derivative in Definition 3.7.

Definition 3.11: *The belief functional of an RFS Σ is defined as*

$$\beta_{\Sigma}(S) = \mathbb{P}(\Sigma \subseteq S) = \mathbb{P}(\{\omega \in \Omega : \Sigma(\omega) \subseteq S\}) \quad (3.25)$$

for any closed $S \subseteq \mathcal{X}$.

By (3.23),

$$\mathbf{P}_{\Sigma}(\mathcal{O}) = \mathbf{P}_{\Sigma}(\{C \in C^{\mathbb{X}} : C \subseteq S\}) = \mathbb{P}(\Sigma \subseteq S) = \beta_{\Sigma}(S).$$

for some closed $S \subseteq \mathbb{X}$. Since $\beta_{\Sigma}(S)$ is a restriction of $\mathbf{P}_{\Sigma}(\mathcal{O})$ to a smaller class of measurable subsets of $\mathcal{F}(\mathbb{X})$, by [109, p.30] and [101, p.713] it has been shown that $\mathbf{P}_{\Sigma}(\mathcal{O})$ is equivalent to $\beta_{\Sigma}(S)$.

Applying the properties of the RFS Σ , the belief functional $\beta_{\Sigma}(S)$ can be expanded as shown in the following theorem for any closed subset $S \subseteq \mathbb{X}$.

Theorem 3.1: *Let Σ be an RFS with belief functional*

$$\beta_{\Sigma}(S) = \mathbb{P}(\Sigma \subseteq S) = \mathbb{P}(\{\omega \in \Omega : \Sigma(\omega) \subseteq S\}) \quad (3.26)$$

for any closed subset $S \subseteq \mathbb{X}$. Then $\beta_{\Sigma}(S)$ can be factorized as follows

$$\beta_{\Sigma}(S) = \sum_{i=0}^{\infty} p_{\Sigma}(i) q_{\Sigma,i}(S^i) \quad (3.27)$$

where

$$p_{\Sigma}(i) = \mathbb{P}(|\Sigma| = i) \geq 0 \quad (3.28)$$

as a discrete probability distribution which gives the number of elements in Σ and $q_{\Sigma,i}(\cdot)$ is the probability measure on \mathbb{X}^i

$$q_{\Sigma,i}(S^i) = \mathbb{P}(\Sigma(\omega) \subseteq S \mid |\Sigma| = i) \quad (3.29)$$

with $q_{\Sigma,i}(S^0) = 1$.

The following definition shows the relationship between the absolute continuity of $q_{\Sigma,i}$ and the absolute continuity of an RFS Σ .

Definition 3.12: *Let Σ be an RFS. Then Σ is absolutely continuous if the probability measures $q_{\Sigma,i}$ are absolutely continuous with respect to the product hybrid measure on \mathbb{X}^i for every $i = 1, 2, \dots$*

If $q_{\Sigma,i}(S^i)$ is absolutely continuous with respect to hybrid measure, then by Proposition 3.1 it follows that

$$q_{\Sigma,i}(S^i) = \int_{S^i} f^{(i)}(\xi_1, \dots, \xi_i) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_i)$$

for some density function $f^{(i)}$ which, without loss of generality, assumed to be completely symmetric. Then by Definition 3.12, Σ is absolutely continuous. Then applying Definition 3.8, the following holds for $j, n > 0$ [60, p.159-160]

$$\frac{\delta^n q_{\Sigma,j}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) = \begin{cases} j! f^{(j)}(\xi_1 \dots \xi_j), & \text{if } j = n; \\ 0, & \text{if } j \neq n. \end{cases} \quad (3.30)$$

Then from Theorem 3.1, we find that

$$\frac{\delta^n \beta_{\Sigma}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) = p_{\Sigma}(n) \frac{\delta^n q_{\Sigma,n}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \quad (3.31)$$

$$= n! p_{\Sigma}(n) f^n(\xi_1, \dots, \xi_n) \quad (3.32)$$

Therefore the belief functional of RFS Σ for any closed set $S \subseteq \mathbb{X}$ is

$$\beta_{\Sigma}(S) = \frac{\delta^0 \beta_{\Sigma}}{\delta \emptyset}(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{S^n} \frac{\delta^n \beta_{\Sigma}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.33)$$

$$\stackrel{(1a)}{=} \int_S \frac{\delta \beta_{\Sigma}}{\delta Z}(\emptyset) \delta(dZ) \quad (3.34)$$

where $\frac{\delta^0 \beta_{\Sigma}}{\delta \emptyset}(\emptyset) = \beta_{\Sigma}(\emptyset) = p_{\Sigma}(0)$ and (1a) holds by (3.21).

By (3.23) for any $\mathcal{O} \in \sigma(C^{\mathbb{X}})$ and absolutely continuous Σ , we have

$$\mathbb{P}(\Sigma \in \mathcal{O}) = \mathbb{P}(\Sigma \subseteq S) = \beta_{\Sigma}(S) \quad (3.35)$$

for some closed subset S of \mathbb{X} . Thus, by (3.24) we also have

$$\mathbb{P}(\Sigma \in \mathcal{O}) = \frac{\delta^0 \beta_{\Sigma}}{\delta \emptyset}(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\chi_n^{-1}(\mathcal{O} \cap \mathcal{F}_n(\mathbb{X}))} \frac{\delta^n \beta_{\Sigma}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.36)$$

$$\stackrel{(1b)}{=} \int_{\mathcal{O}} \frac{\delta \beta_{\Sigma}}{\delta Z}(\emptyset) \delta(dZ) \quad (3.37)$$

where (1b) holds by (3.22).

It is easy to see that $\frac{\delta \beta_{\Sigma}}{\delta Z}(\emptyset)$ is also a global density function since

$$\int_{\mathbb{X}} \frac{\delta \beta_{\Sigma}}{\delta Z}(\emptyset) \delta(dZ) \stackrel{(1a)}{=} \beta_{\Sigma}(\mathbb{X}) = \mathbb{P}(\Sigma \subseteq \mathbb{X}) = 1$$

where (1a) holds by (3.33) and (3.34).

Denote by

$$\mathbf{f}_\Sigma(Z) = \frac{\delta \beta_\Sigma}{\delta Z}(\emptyset) \quad (3.38)$$

for all finite subset $Z \subseteq \mathcal{F}(\mathbb{X})$, then \mathbf{f}_Σ is a global density. In particular,

$$\mathbf{f}_\Sigma(\{\xi_1, \dots, \xi_n\}) = \frac{\delta^n \beta_\Sigma}{\delta \xi_1 \dots \delta \xi_n}(\emptyset).$$

Then (3.34) and (3.37) can be rewritten compactly and respectively as follows

$$\beta_\Sigma(S) = \int_S \mathbf{f}_\Sigma(Z) \delta(dZ) \quad (3.39)$$

$$\mathbb{P}(\Sigma \in \mathcal{O}) = \int_{\mathcal{O}} \mathbf{f}_\Sigma(Z) \delta(dZ). \quad (3.40)$$

Denote the cardinality distribution of RFS Σ by $\mathbf{f}_\Sigma(n)$. Then taking integral on \mathbb{X}^n both side of (3.31) with respect to $\bar{\lambda}^n$, we have

$$\mathbf{f}_\Sigma(n) = p(|\Sigma| = n) = \frac{1}{n!} \int_{\mathbb{X}^n} \frac{\delta^n \beta_\Sigma}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.41)$$

$$= \frac{1}{n!} \int_{\mathbb{X}^n} \mathbf{f}_\Sigma(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.42)$$

Then it is easy to see $\sum_{n=0}^{\infty} \mathbf{f}_\Sigma(n) = 1$.

If there are m RFS $\Sigma_1, \dots, \Sigma_m$, (3.38) can be generalized as in the following definition.

Definition 3.13: Let $\Sigma_1, \dots, \Sigma_m$ be absolutely continuous finite random subsets of \mathbb{X} . For closed sets $S_1, \dots, S_m \subseteq \mathbb{X}$, their joint belief functional is

$$\beta_{\Sigma_1, \dots, \Sigma_m}(S_1, \dots, S_m) = \mathbb{P}(\Sigma_1 \subseteq S_1, \dots, \Sigma_m \subseteq S_m) \quad (3.43)$$

The joint global density of these random finite sets is the multi-variable set function

$$\mathbf{f}_{\Sigma_1, \dots, \Sigma_m}(Z_1, \dots, Z_m) = \frac{\delta^m \beta_{\Sigma_1, \dots, \Sigma_m}(S_1, \dots, S_m)}{\delta_1 Z_1 \dots \delta_m Z_m}(\emptyset, \dots, \emptyset).$$

In practice, the joint global densities of RFS is assumed to be statistically independent. The following proposition gives the joint densities of such random sets

Proposition 3.3: Let $\Sigma_1, \dots, \Sigma_m$ be statistically independent, absolutely continuous finite random subsets of \mathbb{X} . Their joint density exists and is given by

$$\mathbf{f}_{\Sigma_1, \dots, \Sigma_m}(Z_1, \dots, Z_m) = \prod_{i=1}^m \mathbf{f}_{\Sigma_i}(Z_i)$$

for all finite sets $Z_1, \dots, Z_m \subseteq \mathbb{X}$.

3.1.4 Probability Theory and FISST

This section shows the relationship between integral defined as finite set statistics (FISST) in Section 3.1.3 and the integral defined as the conventional probability theory in Section 3.1.2. This section is based on [171, p.6] and [101, Appendix F, p.711-716].

As shown in [171, Proposition 1,p.6] that for any close subset S of \mathbb{X} , we have

$$\mathbf{P}_\Sigma(\chi(\bigsqcup_{i=0}^{\infty} S^i)) = \int_{\chi(\bigsqcup_{i=0}^{\infty} S^i)} g_\Sigma(Z) \mu(dZ) \quad (3.44)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{S^n} K_x^{-n} g_\Sigma(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.45)$$

$$= \int_S K_x^{-|Z|} g_\Sigma(Z) \delta(Z) \quad (3.46)$$

Note that $\bar{\lambda}(\xi_i), i = 1, \dots, n$ have unit of K_x . Furthermore, by definition of the belief functional,

$$\beta_\Sigma(S) = \mathbf{P}_\Sigma(\bigsqcup_{i=0}^{\infty} S^i). \quad (3.47)$$

Then by (3.34) and (3.46), we have $\frac{\delta \mathbf{P}_\Sigma}{\delta \mu}(Z) = K^{|Z|} \frac{\delta \beta_\Sigma}{\delta Z}(\emptyset)$. This shows that the density function g_Σ in (3.9) can be calculated through the belief functional by the following proposition.

Proposition 3.4: *Let $g_\Sigma : \mathcal{F}(\mathbb{X}) \rightarrow [0, \infty)$ be an integrable function such that (3.9) holds. For any finite set $Z \in \mathcal{F}(\mathbb{X})$ then*

$$g_\Sigma(Z) = K_x^{|Z|} \frac{\delta \beta_\Sigma}{\delta Z}(\emptyset). \quad (3.48)$$

Note that $g_\Sigma(Z)$ is unitless density because $\frac{\delta \beta_\Sigma}{\delta Z}(\emptyset)$ have units of $K_x^{-|Z|}$. Then $p(|\Sigma| = n)$ the cardinality distribution of RFS Σ in (3.41) is

$$p(|\Sigma| = n) = p_\Sigma(n) = \frac{1}{n!} \int_{S^n} K_x^{-n} g_\Sigma(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.49)$$

3.1.5 Some Important Multi-target Probability Distribution

In this section, some important multi-target probability densities are introduced such as independent identically distributed (i.i.d) cluster processes and multi-target Poisson processes [101, p.343-375].

By (3.48) and (3.32), the global density $g_\Sigma(Z)$ for any finite subset $Z = \{\xi_1, \dots, \xi_n\} \subseteq \mathbb{X}$ is

$$g_\Sigma(Z) = K_x^{|Z|} \frac{\delta \beta_\Sigma}{\delta Z}(\emptyset) = K_x^{|Z|} \frac{\delta \beta_\Sigma}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) = K_x^{|Z|} p_\Sigma(n) n! f^{(n)}(\xi_1, \dots, \xi_n) \quad (3.50)$$

Note that $f^{(n)}(\xi_1, \dots, \xi_n)$ has unit of K_x^{-n} .

3.1.5.1 Independent Identically Distributed Cluster Processes

When $\xi_i, i = 1 \dots, n$ are i.i.d random elements on \mathbb{X} , $f^{(n)}(\xi_1, \dots, \xi_n) = \prod_{i=1}^n f(\xi_i)$. Note that $f^{(n)}(\cdot)$ denote the joint density whereas $f(\cdot)$ denotes the density on \mathcal{X} . Note that $f(\cdot)$ has unit of K_x^{-1} . Hence the global density $g_\Sigma(Z)$ in (3.50) for i.i.d cluster processes is

$$g_\Sigma(Z) = K_x^{|Z|} n! p_\Sigma(n) f(\xi_1) \dots f(\xi_n) = K_x^{|Z|} |Z|! p_\Sigma(|Z|) \prod_{\xi \in Z} f(\xi) \quad (3.51)$$

3.1.5.2 Multi-target Poisson Processes

If the discrete probability distribution $p_\Sigma(n)$ is the Poisson distribution with mean η , and $\xi_i, i = 1 \dots, n$ are i.i.d random elements on \mathbb{X} then (3.51) is written as

$$g_\Sigma(Z) = K_x^{|Z|} |Z|! \frac{e^{-\eta} \eta^{|Z|}}{|Z|!} \prod_{\xi \in Z} f(\xi) = K_x^{|Z|} e^{-\eta} \eta^{|Z|} \prod_{\xi \in Z} f(\xi) \quad (3.52)$$

which is called a multi-dimensional Poisson distribution. For any finite subset $Z \subseteq \mathbb{X}$, any RFS Σ having $g_\Sigma(Z)$ as its distribution is a multi-target Poisson process. The function

$$\gamma(\xi) = \eta f(\xi) \quad (3.53)$$

is called intensity density of the Poisson process and has unit of K_x^{-1} . Thus (3.52) is alternatively written in terms of the intensity $\gamma(\cdot)$ as

$$g_\Sigma(Z) = K_x^{|Z|} e^{-\langle \gamma, 1 \rangle} \prod_{\xi \in Z} \gamma(\xi) \quad (3.54)$$

where $\langle \gamma, 1 \rangle = \int_{\mathbb{X}} \gamma(\xi) \bar{\lambda}(d\xi) = \eta$.

3.1.5.3 Multi-Bernoulli Processes

This section adopts some formulas from [172, p.29-30] and [101, p.368-370]

Bernoulli:

Similar to a Bernoulli trial, a Bernoulli RFS Σ on \mathbb{X} is empty with probability $1 - r$, $r > 0$, singleton with probability r where the element is distributed according to the probability distribution p , and zero otherwise. Thus the Bernoulli RFS Σ is completely determined by $\{r, p\}$ and its probability density $\pi_\Sigma(X)$ is

$$\pi_\Sigma(X) = \begin{cases} 1 - r, & X = \emptyset; \\ K_x r p(\xi), & X = \{\xi\}; \\ 0, & \text{otherwise.} \end{cases} \quad (3.55)$$

Alternatively, (3.55) can be rewritten as

$$\pi_{\Sigma}(X) = K_x^{|X|} (1-r)^{1-|X|} (rp_{\Sigma}(X))^{|X|} \quad (3.56)$$

where $p_{\Sigma}(X) = p(\xi)$ if $X = \{\xi\}$, $p_{\Sigma}(X) = 1$ if $X = \emptyset$ and $p_{\Sigma}(X) = 0$ otherwise.

Multi-Bernoulli:

Assume that Σ is the union of m independent Bernoulli RFS $\Sigma^i, i = 1, \dots, m$ with a probability of existence r^i and probability density p^i respectively, i.e.

$$\Sigma = \bigcup_{i=1}^m \Sigma^i. \quad (3.57)$$

Then Σ is called a multi-Bernoulli RFS and its probability density $\pi_{\Sigma}(X)$ is written as follows

$$\pi_{\Sigma}(X) = \begin{cases} \prod_{i=1}^m (1-r^i), & X = \emptyset; \\ K_x^n \sum_{\substack{\{j_1, \dots, j_n\} \subseteq \{1, \dots, n\} \\ j_i \neq j_r, i \neq r}} \prod_{k=1}^n r^{j_k} p^{j_k}(\xi_k) \prod_{l \in \{1, \dots, m\} - \{j_1, \dots, j_n\}} (1-r^l), & |X| = n \leq m; \\ 0, & |X| = n > m. \end{cases} \quad (3.58)$$

In (3.58), the first product of the second line is the distribution of n independent Bernoulli RFSs $\Sigma_{j_k}, k = 1, \dots, n$ which each of them is a singleton while the second product is the distribution of $m - n$ independent Bernoulli RFSs $\Sigma_l, l \notin \{j_1, \dots, j_n\}$ which each of them is empty.

Alternatively, (3.58) can be rewritten in the compressed form as follows

$$\pi_{\Sigma}(X) = K_x^{|X|} \prod_{i=1}^m (1-r^i) \sum_{\substack{\{j_1, \dots, j_n\} \subseteq \{1, \dots, n\} \\ j_i \neq j_r, i \neq r}} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} p^{j_k}(\xi_k) \quad (3.59)$$

By (3.49), the corresponding cardinality distribution of (3.59) is

$$\begin{aligned} p(n) &= \frac{1}{n!} \int_{|X|=n} K_x^{-n} \pi_{\Sigma}(\{\xi_1 \dots \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \\ &= \frac{1}{n!} \prod_{i=1}^m (1-r^i) \sum_{\substack{\{j_1, \dots, j_n\} \subseteq \{1, \dots, n\} \\ j_i \neq j_r, i \neq r}} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} \int p^{j_k}(\xi_k) \bar{\lambda}(d\xi_k) \\ &= \frac{1}{n!} \prod_{i=1}^m (1-r^i) \sum_{\substack{\{j_1, \dots, j_n\} \subseteq \{1, \dots, n\} \\ j_i \neq j_r, i \neq r}} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} = \frac{1}{n!} \prod_{i=1}^m (1-r^i) n! \sum_{1 \leq j_1 < \dots < j_n \leq n} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} \\ &= \prod_{i=1}^m (1-r^i) \sum_{1 \leq j_1 < \dots < j_n \leq n} \prod_{k=1}^n \frac{r^{j_k}}{1-r^{j_k}} \end{aligned} \quad (3.60)$$

3.1.5.4 Binomial independent and identical distributed (i.i.d.) cluster Processes

If $r^i = r$ and $p^i = p$ for all $i = 1, \dots, n$ then (3.59) reduces to i.i.d. cluster process and the cardinality in (3.59) reduces to

$$p(n) = \binom{m}{n} (1-r)^{m-n} r^n. \quad (3.61)$$

Moreover, the probability density (3.59) reduces to the following simple form

$$\pi_{\Sigma}(\{\xi_1, \dots, \xi_n\}) = K_x^n n! p(n) \prod_{i=1}^n p(\xi_i). \quad (3.62)$$

3.1.5.5 Probability-generating functionals

This subsection is also based on [40, p.111-156] and [96]. The probability-generating functionals can often transform difficult mathematical problems into simpler ones. Let $h(\xi)$ be a non-negative real-valued function of $\xi \in \mathbb{X}$ that has no unit measurement. Let Z be finite subset of \mathbb{X} , i.e. $Z \in \mathcal{F}(\mathbb{X})$, define the power of h with respect to Z to be

$$h^Z = \begin{cases} 1, & \text{if } Z = \emptyset; \\ \prod_{\xi \in Z} h(\xi), & \text{otherwise.} \end{cases} \quad (3.63)$$

In similar to the definition of probability-generating function, the probability generating functional (p.g.fl.) G_{Σ} of an RFS Σ on \mathbb{X}

$$G_{\Sigma}[h] = E[h^{\Sigma}] = E \left[E[h^{\Sigma} | |\Sigma| = n] \right] = \sum_{n=0}^{\infty} p(n) E[h^{\Sigma} | |\Sigma| = n] \quad (3.64)$$

where

$$E[h^{\Sigma} | |\Sigma| = n] = \int_{\mathbb{X}^n} h^{\{\xi_1, \dots, \xi_n\}} P_n(d\xi_1, \dots, d\xi_n) \quad (3.65)$$

where $P_n(\cdot)$ is the joint probability distribution on \mathbb{X}^n . By (3.26), (3.28) and (3.29), we have

$$\mathbb{P}(\{\xi_1, \dots, \xi_n\} \subset \bigcup_{i=1}^n d\xi_i) = p(n) P_n(d\xi_1, \dots, d\xi_n) \quad (3.66)$$

where $P_n(\cdot)$ is given in (3.65). On the other hand, we have

$$\mathbb{P}(\{\xi_1, \dots, \xi_n\} \subset \bigcup_{i=1}^n d\xi_i) \stackrel{(a)}{=} \mathbb{P}(\{\xi_1, \dots, \xi_n\} \in \{\bigcup_{i=1}^n d\xi_i\}) \quad (3.67)$$

$$\stackrel{(b)}{=} \frac{1}{n!} \frac{\delta^n \beta_{\Sigma}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \quad (3.68)$$

where (a) hold by (3.35) and (b) hold by (3.36). Thus, (3.64) becomes

$$G_{\Sigma}[h] = \sum_{n=0}^{\infty} \int_{\mathbb{X}^n} h^{\{\xi_1, \dots, \xi_n\}} p(n) P(d\xi_1, \dots, d\xi_n) \quad (3.69)$$

$$\stackrel{(a)}{=} \sum_{n=0}^{\infty} \int_{\mathbb{X}^n} h^{\{\xi_1, \dots, \xi_n\}} \mathbb{P}(\{\xi_1, \dots, \xi_n\} \in \{\bigcup_{i=1}^n d\xi_i\}) \quad (3.70)$$

$$\stackrel{(b)}{=} \sum_{n=0}^{\infty} \int_{\mathbb{X}^n} h^{\{\xi_1, \dots, \xi_n\}} \frac{1}{n!} \frac{\delta^n \beta_{\Sigma}}{\delta \xi_1 \dots \delta \xi_n}(\emptyset) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \stackrel{(c)}{=} \int_{\mathbb{X}} h^Z \frac{\beta_{\Sigma}}{\delta Z}(\emptyset) \delta Z \quad (3.71)$$

where (a) holds by (3.66) and (3.67), (b) holds by (3.68) and (c) holds by (3.21).

By (3.44), (3.46), (3.48), (3.70) and (3.71), we also have

$$G_{\Sigma}[h] = \int_{\chi(\biguplus_{i=0}^{\infty} S^i)} h^Z g_{\Sigma}(Z) \mu(dZ) = \int_{\mathcal{F}(\mathbb{X})} h^Z g_{\Sigma}(Z) \mu(dZ) \quad (3.72)$$

where μ is given in (3.8). The probability-generating functionals have the following properties

Relation to probability-generating function of the random number

If $h(\xi) = c$ is a constant nonnegative real number for all $\xi \in \mathbb{X}$. Then by (3.72) we have

$$\begin{aligned} G_{\Sigma}[h] &= \int_{\mathcal{F}(\mathbb{X})} h^Z g_{\Sigma}(Z) \mu(dZ) \\ &= p_{\Sigma}(0) + \sum_{n=1}^{\infty} \frac{c^n}{n!} \int_{\mathbb{X}^n} K_x^{-n} g_{\Sigma}(\{\xi_1, \dots, \xi_n\}) \bar{\lambda}(d\xi_1) \dots \bar{\lambda}(d\xi_n) \\ &\stackrel{(a)}{=} p_{\Sigma}(0) + c p_{\Sigma}(1) + c^2 p_{\Sigma}(2) \dots = G_{|\Sigma|}(h) \end{aligned}$$

where (a) holds by (3.49) and $G_{|\Sigma|}(h)$ is the probability-generating functional of the random nonnegative integer $|\Sigma|$.

Relation to Belief functional

The probability-generating functional is related to the Belief function by this relationship

$$\begin{aligned} G_{\Sigma}[\mathbf{1}_S] &= \int_{\mathcal{F}(\mathbb{X})} \mathbf{1}_S^Z g_{\Sigma}(Z) \mu(dZ) \stackrel{(a)}{=} \int_{\mathbb{X}} \mathbf{1}_S^Z K_x^{-|Z|} g_{\Sigma}(Z) \delta(dZ) = \int_S K_x^{-|Z|} g_{\Sigma}(Z) \delta(dZ) \\ &\stackrel{(b)}{=} \int_S \frac{\delta \beta_{\Sigma}}{\delta Z}(\emptyset) \delta(dZ) = \beta_{\Sigma}(S) \end{aligned}$$

where $\mathbf{1}_S$ is the indicator function, (a) holds by (3.44)-(3.46) and (b) holds by (3.48). Thus, it shares the following useful property with the belief functional.

Unions of statistically independent RFSs

Let $\Sigma = \Sigma^1 \cup \dots \cup \Sigma^n$, and let $G_{\Sigma^1}[h], \dots, G_{\Sigma^n}[h]$ be the corresponding probability-generating functionals where $\Sigma^i, i = 1, \dots, \Sigma^n$ are statistically independent. Then for all h , we have

$$G_{\Sigma}[h] = G_{\Sigma^1}[h] \dots G_{\Sigma^n}[h]$$

Examples of some probability generating functionals

Based on the definition of the probability-generating functional, the processes introduced earlier have the following p.g.fl

- p.g.fl of Poisson process is $G_{\Sigma}[h] = e^{\eta \int h(\xi) \bar{f}(\xi) \bar{\lambda}(d\xi) - \eta}$
- p.g.fl of an i.i.d. cluster processes is $G_{\Sigma}[h] = G(\int h(\xi) \bar{f}(\xi) \bar{\lambda}(d\xi))$
- p.g.fl of Bernoulli process is $G_{\Sigma}[h] = (1 - r + r \int h(\xi) \bar{f}(\xi) \bar{\lambda}(d\xi))$
- p.g.fl of multi-Bernoulli process is $G_{\Sigma}[h] = \prod_{i=1}^m (1 - r^i + r^i \int h(\xi) \bar{f}(\xi) \bar{\lambda}(d\xi))$

3.2 RFS Model for Multi-target Tracking

In this section, the multi-target system model for tracking an unknown number of targets is presented in the RFS framework whose theoretical background was covered in the previous section. In the multi-target system, the number of targets is unknown and varies with time due to the appearance and disappearance of the single targets in the surveillance area. Similarly the unknown number of measurement also changes with time due to imperfect sensors and spurious measurements not coming from targets. Furthermore, the origins of the measurements are unknown. The multi-target tracking problem can be naturally modeled in a very flexible manner using random finite sets. Modeling the target states and measurements at each time instant as RFSs captures the unknown and varying number of targets and measurements; and the fact that the order of the target states or measurement is irrelevant. The model of the multi-target tracking problem in RFS framework can be found in many places from the more theoretical and mathematical oriented sources [60, p219-256] to the more engineering oriented sources in [106],[92, Chapter 9,11-12]. The specific application of the underlying RFS model for the multi-object dynamics and multi-object measurements can be found in [96, 171].

This section is organized as follows. First, the underlying multi-target states will be modeled to capture the randomness of general multi-target tracking problems in Subsection 3.2.1. This underlying multi-target states are observed by the measurements which give information about the targets. These measurements are modeled in Subsection 3.2.2. In this section, the single target system model from Chapter 2.1 is used to build the multi-target model.

3.2.1 Multi-target Dynamical Model

In multi-target tracking (MTT), the single targets are usually assumed to move independently in the region of interest and the number of targets changes over time due to the spontaneous birth, death or spawning of the new targets from existing targets (e.g. rocket). This makes the problem more challenging than the single target tracking problem. In the following subsections, some common approaches to constructing the multi-target state model and its Markov transition density are discussed. In practice the common single state space is usually the hybrid space is discussed in the previous section. However the hybrid state space will not be used until Chapter 6.2.

3.2.1.1 Multi-target State

Given a multi-target state $X_{t-1} = \{x'_1, \dots, x'_m\}$ at time $t-1$, each state $x' \in X_{t-1}$ is assumed to follow a Markov process in the following sense. The single target which is given in (2.1) either continues to exist at time $t \in \mathcal{T}, t > 1$ with probability $p_{S_t}(x')$ and moves to the new state x according to the probability density $\bar{f}_{t|t-1}(x|x')$ in (2.3) or dies with probability $1 - p_{S_t}(x')$ and takes on the value \emptyset . Thus, given a single state $x' \in X_{t-1}$ at time $t-1$, its behavior at time t is modeled by the Bernoulli RFS

$$S_{t|t-1}(x')$$

that is either $\{x\}$ when the target survives or \emptyset when the target dies.

Denote by $\beta_{S_{t|t-1}}(\cdot|x')$ the belief functional of an RFS $S_{t|t-1}(x')$. Then for any closed subset S of \mathcal{X} and by Theorem 3.1, we have

$$\begin{aligned} \beta_{S_{t|t-1}}(S|x') &= \mathbb{P}(S_{t|t-1}(x') \subseteq S|x') = p_{\Sigma}(0) + p_{\Sigma}(1)q_{\Sigma,1}(S) \\ &= 1 - p_{S_t}(x') + p_{S_t}(x') \int \bar{f}_{t|t-1}(x|x')\lambda(dx) = \int_S f_{S_{t|t-1}}(Y|x')\delta Y \end{aligned}$$

where μ_s is the dominating measure of the form (3.8) on the Borel subsets of $\mathcal{F}(\mathcal{X})$ where the state space \mathcal{X} is used in place of the hybrid space \mathbb{X} as follows

$$\mu_s(\mathcal{O}) = \sum_{n=0}^{\infty} \frac{\lambda^n(\chi^{-1}(\mathcal{O}) \cap \mathcal{X}^n)}{n!K_x^n} \quad (3.73)$$

for any Borel set $\mathcal{O} \subseteq \mathcal{F}(\mathcal{X})$. Note that K_x is the unit of volume on \mathcal{X} . Thus $\beta_{\Sigma}(S)$ is completely described by the distribution of a target x'

$$f_{S_{t|t-1}}(S_{t|t-1}(x')|x') = \begin{cases} 1 - p_{S_t}(x'), & \text{if } S_{t|t-1}(x') = \emptyset; \\ K_x p_{S_t}(x') \bar{f}_{t|t-1}(x|x'), & \text{if } S_{t|t-1}(x') = \{x\} \\ 0, & \text{otherwise.} \end{cases} \quad (3.74)$$

The survival or death of all existing target from time $t-1$ to time t is hence modeled by the RFS

$$S_{t|t-1}(X_{t-1}) = \bigcup_{x' \in X_{t-1}} S_{t|t-1}(x'). \quad (3.75)$$

Let $X_t = S_{t|t-1}(X_{t-1}) = \{x_1, \dots, x_n\}$ and $|X_{t-1}| = m$. Conditional on X_{t-1} , the RFSs on the right hand side of (3.75) are assumed to be mutual independent. So by (3.43) in Definition 3.13,

the belief functional $\beta_{\mathcal{S}_{t|t-1}}$ of an RFS $\mathcal{S}_{t|t-1}(X_{t-1})$ for the model (3.75) is for any $S \subseteq \mathcal{X}$

$$\begin{aligned} \beta_{\mathcal{S}_{t|t-1}}(S|X_{t-1}) &= \mathbb{P}(\mathcal{S}_{t|t-1}(X_{t-1}) \subseteq S) \\ &= \sum_{\uplus_{x' \in X_{t-1}} \mathcal{S}_{t|t-1}(x')=X_{t-1}} \prod_{x' \in X_{t-1}} \mathbb{P}(\mathcal{S}_{t|t-1}(x') \in S) \\ &= \sum_{\uplus_{x' \in X_{t-1}} \mathcal{S}_{t|t-1}(x')=X_{t-1}} \prod_{x' \in X_{t-1}} p_{x', \mathcal{S}_{t|t-1}(x')}(S). \end{aligned}$$

where

$$p_{x', \mathcal{S}_{t|t-1}(x')}(S) = \begin{cases} 1 - p_{S_t}(x'), & \mathcal{S}_{t|t-1}(x') = \emptyset, \\ p_{S_t}(x') \int_S \bar{f}(x|x') dx, & |\mathcal{S}_{t|t-1}(x')| = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3.76)$$

and by the product rule in (3.20) of Proposition 3.2., the global density $\pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1})$ of the RFS X_t By the product rule in (3.20) of Proposition 3.2 and

By (3.48) the global density $\pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1})$ of the RFS $\mathcal{S}_{t|t-1}(X_{t-1})$ is

$$\pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1}) = K_x^{|X_t|} \frac{\delta \beta_{\mathcal{S}_{t|t-1}}(\emptyset|X_{t-1})}{\delta X_t}. \quad (3.77)$$

By the product rule in (3.20) and (3.19) of Proposition 3.2, (3.77) becomes

$$\pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1}) = K_x^{|X_t|} \sum_{\substack{\uplus_{i=1}^{|X_{t-1}|} u_i = X_t \\ u_i = \emptyset \text{ or } u_i = \{x\} \subseteq X_t}} \prod_{x'} \frac{\delta p_{x', u_i}(\emptyset)}{\delta u_i}$$

where \uplus denotes the disjoint union and K_x is the unit of volume on space \mathcal{X} . Note that $|X_t| \leq |X_{t-1}|$. By the discussion in Section 3.1.4, $\pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1})$ is unitless and each $\frac{\delta \beta_{\mathcal{S}_{t|t-1}}(\emptyset|x')}{\delta x}$, $x' \in X_{t-1}$, $x \in X_t$ has unit of K_x . From equation (3.16), (3.17) and (3.76), we have

$$\frac{\delta p_{x', u_i}(\emptyset)}{\delta u_i} = \begin{cases} p_{x', \emptyset}(\emptyset) = 1 - p_{S_t}(x'), & \text{if } u_i = \emptyset; \\ \frac{\delta p_{x', u_i}(\emptyset)}{\delta x} \stackrel{(a)}{=} p_{S_t}(x') \bar{f}_{t|t-1}(x|x'), & \text{if } u_i = \{x\}; \\ 0, & \text{if } |u_i| > 1. \end{cases}$$

where (a) holds by (3.31) and (3.30). To express the probability density $\pi_{\mathcal{S}_{t|t-1}}(\cdot|X_{t-1})$ of the RFS $\mathcal{S}_{t|t-1}(X_{t-1})$ in a general form, we introduce the following notation.

Let $\mathbb{T}(U, V)$ denote the set of all one-to-one functions taking a finite set U to a finite set V . The set of all 1-1 function $\mathbb{T}(U, V) = \emptyset$ if $|U| > |V|$ and we use the convention that the sum over the empty set is zero. A one-to-one function $\alpha \in \mathbb{T}(X_t, X_{t-1})$ is used to associate the targets at time t with the targets at time $t-1$. Specifically, $x' = \alpha(x)$ means that the target state x' at time $t-1$ has evolved to the state x at time t (i.e. $\alpha(x)$ represents the previous state at time $t-1$ of the target state x). A target state x' at time $t-1$ not associated with any target state state at time

t is dead. With this notation, it follows that the transitional probability density $\pi_{\mathcal{S}_{t|t-1}}(\cdot|X_{t-1})$ of the RFS $\mathcal{S}_{t|t-1}(X_{t-1})$ is

$$\begin{aligned} \pi_{\mathcal{S}_{t|t-1}}(X_t|X_{t-1}) &= K_x^{|X_t|} \sum_{\alpha \in \mathbb{T}(X_t, X_{t-1})} \prod_{x' \in X_{t-1} - \alpha(X_t)} (1 - p_{S_t}(x')) \\ &\quad \times \prod_{x \in X_t} p_{S_t}(\alpha(x)) \bar{f}_{t|t-1}(x|\alpha(x)) \end{aligned} \quad (3.78)$$

where $X_{t-1} - \alpha(X_t)$ means set difference and the sum is $\prod_{x' \in X_{t-1}} (1 - p_{S_t}(x'))$ if $X_t = \emptyset$. The form in (3.78) is originally used in [172, section 2.3.2, p. 33]

A new target at time t may result from either the spontaneous birth (independent of the surviving targets) which is modeled by an RFS of spontaneous births Γ_t or spawning from a target state x' at time $t-1$ which is modeled by an RFS of spawning $B_{t|t-1}(x')$. Thus the multi-target state at time t is the union of the surviving targets, the spawned targets and the spontaneous births

$$\Sigma_t(X_{t-1}) = \mathcal{S}_{t|t-1}(X_{t-1}) \cup \mathcal{B}_{t|t-1}(X_{t-1}) \cup \Gamma_t \quad (3.79)$$

where $\mathcal{B}_{t|t-1}(X_{t-1}) = \bigcup_{x' \in X_{t-1}} B_{t|t-1}(x')$. (3.79) describes how the multi-target state may change from X_{t-1} at time step $t-1$ to $\Sigma_t(X_{t-1})$ at time step t .

3.2.1.2 Markov Transition Density

Assuming the three RFSs on the right hand side of (3.79) are mutually independent conditional on X_{t-1} , the RFS transition density in (3.79) can be described in the form of the *multi-target transition density* $f_{t|t-1}(\cdot|X_{t-1})$ describing the probability of the multi-target state moving from X_{t-1} at time $t-1$ to $\Sigma_t(X_{t-1})$ at time t . Assume that $X_t = \Sigma_t(X_{t-1})$. By (3.43) in Definition 3.13, the belief functional β_{Σ_t} of an RFS $\Sigma_t(X_{t-1})$ for the model (3.79) is for any $S \subseteq \mathcal{X}$

$$\begin{aligned} \beta_{\Sigma_t}(S|X_{t-1}) &= \mathbb{P}(\Sigma_t(X_{t-1}) \subseteq S) \\ &= \mathbb{P}(\mathcal{S}_{t|t-1}(X_{t-1}) \subseteq S) \mathbb{P}(\mathcal{B}_{t|t-1}(X_{t-1}) \subseteq S) \mathbb{P}(\Gamma_t \subseteq S) \\ &= \beta_{\mathcal{S}_{t|t-1}|X_{t-1}}(S|X_{t-1}) \beta_{\mathcal{B}_{t|t-1}}(S|X_{t-1}) \beta_{\Gamma_t}(S) \end{aligned}$$

where by (3.25), $\beta_{\mathcal{B}_{t|t-1}}(\cdot|X_{t-1})$ is the belief functional of an RFS $\mathcal{B}_{t|t-1}(X_{t-1})$; and β_{Γ_t} as the belief functional of the RFS Γ_t . Then by the product rule in (3.20), we have

$$\begin{aligned}
f_{t|t-1}(X_t|X_{t-1}) &= K_x^{|X_t|} \frac{\delta \beta_{\Sigma_t}}{\delta X_t}(\emptyset|X_{t-1}) \\
&= K_x^{|X_t|} \sum_{\uplus_{i=1}^3 U_i=X_t} \frac{\delta \beta_{\mathcal{S}_{t|t-1}}(\emptyset|X_{t-1})}{\delta U_1} \frac{\delta \beta_{\mathcal{B}_{t|t-1}}(\emptyset|X_{t-1})}{\delta U_2} \frac{\delta \beta_{\Gamma_t}(\emptyset)}{\delta U_3} \\
&= \sum_{\uplus_{i=1}^3 U_i=X_t} K_x^{|U_1|} \frac{\delta \beta_{\mathcal{S}_{t|t-1}}(\emptyset|X_{t-1})}{\delta U_1} K_x^{|U_2|} \frac{\delta \beta_{\mathcal{B}_{t|t-1}}(\emptyset|X_{t-1})}{\delta U_2} K_x^{|U_3|} \frac{\delta \beta_{\Gamma_t}(\emptyset)}{\delta U_3} \\
&= \sum_{\uplus_{i=1}^3 U_i=X_t} \pi_{\mathcal{S}_{t|t-1}}(U_1|X_{t-1}) \pi_{\mathcal{B},t|t-1}(U_2|X_{t-1}) \pi_{\Gamma,t}(U_3) \tag{3.80}
\end{aligned}$$

where

- $\pi_{\mathcal{S}_{t|t-1}}(U_1|X_{t-1})$ is given in (3.78).
- $\pi_{\mathcal{B},t|t-1}(U_2|X_{t-1}) = K_x^{|U_2|} \frac{\delta \beta_{\mathcal{B}_{t|t-1}}(\emptyset|X_{t-1})}{\delta U_2}$ is the probability density of the RFS of spawning target from X_{t-1} .
- $\pi_{\Gamma,t}(U_3) = K_x^{|U_3|} \frac{\delta \beta_{\Gamma_t}(\emptyset)}{\delta U_3}$ is the spontaneous birth Γ_t .

Note that $\pi_{\mathcal{B},t|t-1}(\cdot|X_{t-1})$ and $\pi_{\Gamma,t}$ are unitless from the discussion in Section 3.1.4. X_t in (3.80) also considers the new spontaneous birth and spawning target compared to only surviving targets in X_t given in (3.78). (3.79) describes the time evolution of the multi-target state and incorporates the model of target motion, spontaneous birth and spawning which are captured in the multi-target transition density (3.80).

Assuming that Γ_t is a multi-target Poisson process (or Poisson RFS) with intensity function $\gamma_t(\cdot)$ and that $\mathcal{B}_{t|t-1}(x')$ is a Poisson RFS with intensity function $\beta_{t|t-1}(\cdot|x')$ (see multi-target Poisson process in Subsection 3.1.5.2), we have

$$\begin{aligned}
\pi_{\Gamma,t}(X_t) &= K_x^{|X_t|} e^{-\langle \gamma_t, 1 \rangle} \prod_{x \in X_t} \gamma_t(x), \\
\pi_{\mathcal{B},t|t-1}(X_t|X_{t-1}) &= K_x^{|X_t|} e^{-\sum_{x' \in X_{t-1}} \langle \beta_{t|t-1}(\cdot|x'), 1 \rangle} \prod_{x \in X_t} \sum_{x' \in X_{t-1}} \beta_{t|t-1}(x|x')
\end{aligned}$$

where $\langle \gamma_t, 1 \rangle$ is the expected number of spontaneously generated new targets and $\langle \beta_{t|t-1}(\cdot|x), 1 \rangle$ is the expected number of new targets spawned from the target state x .

Then the transition density $f_{t|t-1}(X_t|X_{t-1})$ in (3.80) simplifies to

$$\begin{aligned}
f_{t|t-1}(X_t|X_{t-1}) &= K_x^{|X_t|} \sum_{W \subseteq X_t} \sum_{\alpha \in \mathcal{T}(W, X_{t-1})} e^{-\mu_f(X_{t-1})} \prod_{\bar{x} \in X_t - W} b(\bar{x}|X_{t-1}) \times \\
&\quad \prod_{x' \in X_{t-1} - \alpha(W)} (1 - p_{S_t}(x')) \prod_{x \in W} p_{S_t}(\alpha(x)) \bar{f}_{t|t-1}(x|\alpha(x)) \tag{3.81}
\end{aligned}$$

where

$$\begin{aligned}\mu_f(X_{t-1}) &= \langle \gamma_t, 1 \rangle + \sum_{x' \in X_{t-1}} \langle \beta_{t|t-1}(\cdot|x'), 1 \rangle, \\ b(x|X_{t-1}) &= \gamma_t(x) + \sum_{x' \in X_{t-1}} \beta_{t|t-1}(x|x').\end{aligned}$$

Here $\mu_f(X_{t-1})$ is the expected number of new targets (spontaneous birth and spawning) and $b(\cdot|X_{t-1})$ is the intensity function of a new target state given X_{t-1} . Each $W \subset X_t$ is the set of surviving targets which is evolved from the previous state at time $t-1$ and the second sum is

$$e^{-\mu_f(X_{t-1})} \prod_{\bar{x} \in X_t} b(\bar{x}|X_{t-1}) \prod_{x' \in X_{t-1}} (1 - p_{S_t}(x')) \text{ if } W = \emptyset.$$

3.2.2 Multi-target Measurement Model

In multi-target tracking, the dynamical system is a hidden system so the only known information is the measurements. However, the measurements not only consist of target generated measurements but also include clutter which are measurements generated by other objects which are not the targets of interest. In addition, sensors may not observe the present targets due to sensor imperfection. This subsection will construct the multi-target measurement model.

3.2.2.1 Multi-target Measurement

At time t , each single-target state $x \in X_t$, is either detected with probability $p_{D_t}(x)$ and generates an observation z with likelihood $\bar{g}_t(z|x)$, or it is missed with probability $1 - p_{D_t}(x)$. Thus, at time t , each single-target state $x \in X_t$ generates an RFS $D_t(x)$ that can take either the value $\{z\}$ when the target is observed by a sensor or \emptyset when the target is not detected. The detection and generation of measurements for all targets at time t is hence given by the RFS

$$\mathfrak{D}_t(X_t) = \bigcup_{x \in X_t} D_t(x). \quad (3.82)$$

Assuming that, conditional on the multi-target state X_t , the measurements at time index t are independent of the states at all other time indices and that the RFSs on the right hand side of (3.82) are mutually independent. The independence conditional on target states is a common assumption in tracking algorithms. The probability density of the RFS $Z_t = \mathfrak{D}_t(X_t)$ is calculated similarly to the RFS of the surviving targets which gives

$$\pi_{\mathfrak{D},t}(Z_t|X_t) = K_z^{|Z_t|} \sum_{\alpha \in \mathcal{T}(Z_t, X_t)} \prod_{x \notin \alpha(Z_t)} (1 - p_{D_t}(x)) \prod_{z \in Z_t} p_{D_t}(\alpha(z)) \bar{g}_t(z|\alpha(z)) \quad (3.83)$$

where K_z is the unit of volume on \mathcal{Z} and $\bar{g}_t(z|\alpha(z))$ is given (2.4). The explanation of K_z is similar to that of K_x in the previous section which shows that the density $\pi_{\mathcal{D},t}(Z_t|X_t)$ is unitless. If $Z_t = \emptyset$ the sum is $\prod_{x \in X_t} (1 - p_{D_t}(x))$.

Apart from target-originated measurements, the sensor also receives a set of false/spurious measurements or clutter which is modeled by an RFS Λ_t . Consequently, at time t , the multi-target measurement Z_t is the union of target-generated measurements and clutter,

$$Z_t = \mathcal{D}_t(X_t) \cup \Lambda_t. \quad (3.84)$$

3.2.2.2 Likelihood Function

Assuming that the two RFSs on the right hand side of (3.84) are mutually independent, the RFS multi-target measurement can be expressed in the form of the *multi-target likelihood* $g_{t|t-1}(\cdot|X_t)$. Let $\pi_{\Lambda,t}(\cdot|\cdot)$ be the density of the RFS Λ_t , the multi-target likelihood function $g_t(Z_t|X_t)$ is constructed similarly to the Markov transition density and is given by

$$g_t(Z_t|X_t) = \sum_{\substack{W \subseteq Z_t \\ |W| \leq |X_t|}} \pi_{\mathcal{D},t}(W|X_t) \pi_{\Lambda,t}(Z_t - W|X_t) \quad (3.85)$$

When Λ_t is a Poisson RFS with intensity κ_t ,

$$\pi_{\Lambda,t}(Z) = e^{-\langle \kappa_t, 1 \rangle} K_z^{|Z|} \prod_{z \in Z} \kappa_t(z),$$

the multi-target likelihood function $g_t(Z_t|X_t)$ in (3.85) has the following form

$$g_t(Z_t|X_t) = K_z^{|Z_t|} \sum_{W \subseteq Z_t} \sum_{\alpha \in \mathcal{T}(W, X_t)} e^{-\langle \kappa_t, 1 \rangle} \prod_{z' \in Z_t - W} \kappa_t(z') \prod_{x \in X_t - \alpha(W)} (1 - p_{D_t}(x)) \prod_{z \in W} p_{D_t}(\alpha(z)) \bar{g}(z|\alpha(z)). \quad (3.86)$$

where the second sum is $e^{-\langle \kappa_t, 1 \rangle} \prod_{z' \in Z_t} \kappa_t(z') \prod_{x \in X_t} (1 - p_{D_t}(x))$ if $W = \emptyset$. The formula in (3.86) is originally used in [172, p.35]. The terms in the second sum have the following meanings: the first two terms describe the clutter, the third term (the second product) expresses the missed detections and the last product describe the target-generated measurements.

3.2.3 Multi-target Bayes Filter

Multi-target Bayesian filtering and multi-target estimation in RFS framework is presented in this section. Applying the RFS framework to multi-target tracking was pioneered by Mahler [106] by using random finite sets instead of random vectors. The objective of multi-target Bayes filter is to jointly estimate the number of targets and their states. This filter which generalize the single-target Bayes filter is the theoretical foundation for multi-target fusion, detection, tracking and

identification [101, p.483-537]. This filter can be found in many sources such as [60, 92, 101, 107] upon which this section is based.

3.2.3.1 Multi-target Bayes Filter

This section presents the Multi-target Bayes filter for the multi-target system model described in the previous subsections. Like the single-target Bayes filter, the multi-target Bayes filter also consists of three steps: Initialization, Predictor and Update.

Initialization: The initial step reflect the knowledge of target states before receiving measurements. However, if there is limited information about the target states, the multi-target Poisson process in section 3.1.5.2 are used with a large mean/variance η and a very high-variance spatial distribution $f(x)$ e.g. uniform distribution

$$p_0(X_1) = K_x^{|X_1|} e^{-\eta} \eta^n \prod_{x \in X_1} f(x). \quad (3.87)$$

The density $f(x)$ can be uniform distribution over some known region or the whole region if there is no prior knowledge. Note that density $f(x)$ has unit of K_x^{-1} . Then the posterior distribution $p_1(X_1|Z_1)$ is given by

$$p_1(X_1|Z_1) = \begin{cases} \frac{g_1(Z_1|X_1)p_0(X_1)}{\int_{\mathcal{X}} g_1(Z_1|X_1)p_0(X_1)\mu_s(dX_1)}, & \text{there exist measurements } Z_1; \\ p_0(X_1), & \text{otherwise.} \end{cases} \quad (3.88)$$

where μ_s is the dominating measure given in (3.73) and $g_1(Z_1|X_1)$ is multi-target likelihood function given in (3.86). Note that $g_1(Z_1|X_1)$ and $p_0(X_1)$ are unitless.

Predictor: Given the history of measurements up to time t i.e $Z_{1:t} = (Z_1, \dots, Z_t)$. The predictor for multi-target Bayes filter is the analog of (2.15) with the set integral in (3.10)

$$p_{t+1|t}(X_{t+1}|Z_{1:t}) = \int_{\mathcal{X}} f_{t+1|t}(X_{t+1}|X_t) p_t(X_t|Z_{1:t}) \mu_s(dX_t) \quad (3.89)$$

where $f_{t+1|t}(X_{t+1}|X_t)$ is given in (3.81) and is unitless. Then (3.89) is explicitly written as follows

$$\begin{aligned} & p_{t+1|t}(X_{t+1}|Z_{1:t}) \\ &= \sum_{n=0}^{\infty} \frac{K_x^{-n}}{n!} \int_{\mathcal{X}} f_{t+1|t}(X_t|\{x_1, \dots, x_n\}) p_t(\{x_1, \dots, x_n\}|Z_{1:t}) \lambda^n(dx_1 \dots dx_n) \\ &= \sum_{n=0}^{\infty} \frac{K_x^{-n}}{n!} \int f_{t+1|t}(X_t|\{x_1, \dots, x_n\}) p_t(\{x_1, \dots, x_n\}|Z_{1:t}) \lambda(dx_1) \dots \lambda(dx_n) \end{aligned} \quad (3.90)$$

where λ^n is the n th product (unit) Lebesgue measure on \mathcal{X} and defined in Section (3.1.2). Some examples illustrating this formula can be found in [101, p.487-490].

Update: The update for multi-target Bayes filter is the analog of (2.16)

$$p_{t+1}(X_{t+1}|Z_{1:t+1}) = \frac{g_t(Z_{t+1}|X_{t+1})p_{t+1|t}(X_{t+1}|Z_{1:t})}{p(Z_{t+1}|Z_{1:t})} \quad (3.91)$$

where $p_{t+1|t}(X_{t+1}|Z_{1:t})$ is given in (3.90) and the normalizing factor $p(Z_{t+1}|Z_{1:t})$ is

$$p(Z_{t+1}|Z_{1:t}) = \int_{\mathcal{X}} g_t(Z_{t+1}|X_{t+1})p_{t+1|t}(X_{t+1}|Z_{1:t})\mu_s(X_{t+1}).$$

Note that all the integrals are set integrals as defined in (3.10). Examples of this update are given in [60, p.186-187] and [101, p.491-492].

The multi-target posterior distribution can be alternatively propagated via Bayes recursion as follows. Given the initial distribution p_0 and p_1 is calculated in (3.88). Assume that the posterior distribution $p_{1:t-1}(X_{1:t-1}|Z_{1:t-1})$ up to time $t-1$ is calculated, the posterior distribution $p_{1:t}(X_{1:t}|Z_{1:t})$ at time t can be calculated using the Bayesian recursion

$$p_{1:t}(X_{1:t}|Z_{1:t}) = p_{1:t-1}(X_{1:t-1}|Z_{1:t-1}) \frac{f_{t|t-1}(X_t|X_{t-1})g_t(Z_t|X_t)}{p(Z_t|Z_{1:t-1})}.$$

Denote $f_{1|0}(X_1|X_0) = p_0(X_1)$, the posterior distribution $p_{1:T}(X_{1:T}|Z_{1:T})$ can be written as follows

$$p_{1:T}(X_{1:T}|Z_{1:T}) = \frac{\prod_{t=1}^T f_{t|t-1}(X_t|X_{t-1})g_t(Z_t|X_t)}{p(Z_{1:T})}. \quad (3.92)$$

3.2.3.2 Multi-target Bayes Estimation

Unlike the single-target Bayes estimation, multi-target state estimation poses unexpected difficulties due to the set integral in the posterior distribution. Traditional estimation methods such as EAP and MAP estimators described in Section 2.2.2 do not apply to the multi-target posterior distribution (3.91). This was shown in [101, Section 14.5.1.p.494-497]. Hence Mahler introduced two new Bayes-optimal multi-target state estimators [101, p.494-505]. These estimators are the marginal multitarget (MaM) estimator and the joint multitarget (JoM) estimators.

Failure of Traditional Estimators: Let $p_t(X_t|Z_{1:t})$ denote the posterior distribution at time t . The set integral for EAP estimator is written

$$\hat{\mathbf{X}}^{EAP}(Z_{1:t}) = \int X_t p_t(X_t|Z_{1:t})\mu_s(dX_t) \quad (3.93)$$

The right hand side of the equation is not defined since X_t is a set and the integral of set is not defined.

The MAP estimator is ambiguous as shown next. We consider the example 78 in [101, p.494-495] where the targets move in one dimensional interval $[0, 2]$, the probability of the target being present or absent is equally 0.5 and the distance is measured in meters, the multi-target posterior

distribution is

$$p_t(X_t|Z_{1:t}) = \begin{cases} .5, & \text{if } X_t = \emptyset; \\ .25, & \text{if } X_t = \{x\} \text{ and } 0 \leq x \leq 2; \\ 0, & \text{otherwise.} \end{cases} \quad (3.94)$$

The MAP estimator is given by

$$\hat{\mathbf{X}}^{MAP}(Z_{1:t}) = \underset{X_t}{\operatorname{argmax}} p_t(X_t|Z_{1:t}). \quad (3.95)$$

Therefore $\hat{\mathbf{X}}^{MAP}(Z_{1:t}) = \emptyset$ because $p_t(\emptyset|Z_{1:t}) = 0.5$ is maximum value. However, when we change the unit of measurements from meter to kilometer the multitarget posterior distribution $p_t(X_t|Z_{1:t})$ is

$$p_t(X_t|Z_{1:t}) = \begin{cases} .5, & \text{if } X_t = \emptyset; \\ 250, & \text{if } X_t = \{x\} \text{ and } 0 \leq x \leq 0.002; \\ 0, & \text{otherwise.} \end{cases} \quad (3.96)$$

Note that $p_t(X_t|Z_{1:t})$ is unitless because we multiply the traditional multi-target state and traditional multi-target measurement with $m^{|X_t|}$ in (3.81) and (3.86). Therefore, when we change the unit of measurements, the value of posterior distribution also changes. In this case $\hat{\mathbf{X}}^{MAP}(Z_{1:t}) = \{x\}$ where $x \in [0, 0.002]$.

Marginal Multi-target (MaM) Estimator: The MaM estimator consists of two steps:

- The first step is the computation of the MAP estimate of the target number

$$\hat{n} = \underset{n}{\operatorname{arg sup}} p(n|Z_{1:t})$$

where $p(n|Z_{1:t}) = \frac{1}{n!} \int_{\mathcal{X}^n} p_t(\{x_1, \dots, x_n\}|Z_{1:t}) K_x^{-n} \lambda(dx_1) \dots \lambda(dx_n)$ from (3.49) and $p_t(\cdot|\cdot)$ is given (3.91)

- The second step is computation of a MAP estimate of the states of the individual target given that $n = \hat{n}$

$$\hat{\mathbf{X}}^{MaM}(Z_{1:t}) = \underset{x_1, \dots, x_n}{\operatorname{arg sup}} p_t(\{x_1, \dots, x_n\}|Z_{1:t})$$

The MaM estimator is Bayes-optimal with respect to the risk function defined in [60, 192-194] but it is unknown whether it is statistically consistent [101, p.497-500] (See examples in [101, Section 14.5.3, p.501-503]).

Joint Multi-target (JoM) Estimator: The JoM estimator is defined as

$$\hat{\mathbf{X}}^{JoM}(Z_{1:t}) = \underset{X_t}{\operatorname{arg sup}} p_t(X_t|Z_{1:t}) \frac{c^{|X_t|}}{|X_t|!}$$

where c is a fixed dimensionless constant. Note that c in [101, Section 14.5.2, p.498] is a fixed constant having the same units of measurement as the single target state x . The difference between c in this thesis and the one given in [101, Section 14.5.2, p.498] is due to the fact that the posterior distribution $p_t(X_t|Z_{1:t})$ in this thesis is unitless. The JoM also can be estimated in two steps

1. First, determine the MAP estimate for each $n \geq 0$

$$\hat{X}_n(Z_{1:t}) = \arg \sup_{x_1, \dots, x_n} p_t(\{x_1, \dots, x_n\}|Z_{1:t})$$

2. Then

$$\hat{\mathbf{X}}^{JoM}(Z_{1:t}) = \hat{X}_{\hat{n}}(Z_{1:t}) \quad \text{where} \quad \hat{n} = \arg \sup_n p_t(\hat{X}_n(Z_{1:t})|Z_{1:t}) \frac{c^n}{n!}.$$

The JoM is Bayes optimal and will converge to the true solution provided that there is enough data [101, p. 498]. The value c determines the accuracy of the target state estimate and the rate of convergence. A smaller value of c gives a more accurate target state estimate but a slower rate of convergence [101, Section 14.5.3, p. 498-500].

3.2.4 Multi-target Moment Densities

The multi-target moment density is defined as follows [60, 169] and [104, p.8]

Definition 3.14 (Multi-target Moment Densities): For any finite set $X \in \mathcal{F}(\mathcal{X})$, the multi moment density is

$$\begin{aligned} D_t(X|Z_{1:t}) &= \int K_x^{-|X|} p_t(X \cup W|Z_{1:t}) \mu_s(dW) \\ &= \sum_{n=0}^{\infty} \int K_x^{-|X|+n} p_t(X \cup \{x_1, \dots, x_n\}|Z_{1:t}) \lambda(dx_1) \dots \lambda(dx_n) \end{aligned}$$

where μ_s is given in (3.73) and $p_t(X_t|Z_{1:t})$ given in (3.91).

Notice that $D_t(\emptyset|Z_{1:t}) = 1$ because

$$D_t(\emptyset|Z_{1:t}) = \int p_t(W|Z_{1:t}) \mu_s(dW) = 1. \quad (3.97)$$

If $|X| = n$ then $D_t(X|Z_{1:t})$ is called n th multi-target moment density. Note that $p_t(X \cap W|Z_{1:t}) \mu_s(dW)$ is unitless because μ_s is a unitless measure and $p_t(X \cap W)$ is unitless density by (3.81) so $D_t(X|Z_{1:t})$ has $K_x^{-|X|}$ units. In [104, p.8], author claims that "for any multi-target state $X = \{x_1, \dots, x_n\}$, $D_t(\{x_1, \dots, x_n\}|Z_{1:t})$ is the marginal-posterior likelihood, that is, no matter many targets may be in the multi-target system, exactly n of them have states x_1, \dots, x_n ". When $X = \{x\}$, $D_t(x|Z_{1:t}) = D_t(\{x\}|Z_{1:t})$ and it is called the first-order multi-target moment density, probability hypothesis density (PHD).

3.3 Conclusion

RFS framework has been presented for general space such as a hybrid space which is the product of the state space and a discrete space. The background of RFS and some operations involving RFS such as set integration and set derivative have been introduced. The construction of global densities using two different approaches was presented. The first one used conventional probability and the second approach employed the belief functional. A comparison between the two global densities showed that they were related by the derivative of belief functional at empty set (i.e. \emptyset). Some common probability distributions of RFS were also introduced. The global densities were applied to the multi-target tracking problem in order to formulate the transition densities and likelihood functions. Due to RFS framework, the multi-target Bayes filter was derived and then the multi-target Bayes estimation was derived to accommodate the RFS framework. Finally, the multi-target moment densities were presented.

Chapter 4

Particle Markov Chain Monte Carlo (PMCMC) Methods

Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods are the two main methods for sampling from complicated probability distributions such as the multi-target posterior distribution in (3.91) or (3.92). MCMC and SMC rely on the use of other distributions to explore the state space of interest and if these distributions are poorly chosen or if highly correlated variables of interest are updated independently, the performance of these methods is unreliable. This leads to the derivation of the Particle Markov Chain Monte Carlo (PMCMC) [4] which combines these two methods by taking the advantages of their respective strengths.

Section 4.1 introduces Markov Chain Monte Carlo methods such as the Metropolis-Hastings (MH) algorithm and Gibbs sampler which are used to construct a Markov chain (MC) that converges to the target distribution. Section 4.2 summarizes some PMCMC methods such as the Particle Independent Metropolis-Hastings (PIMH) algorithm, the Particle Marginal Metropolis-Hastings (PMMH) algorithm and the Particle Gibbs algorithm which can be thought of as a natural approximation to the standard MCMC method. They use SMC approach to design efficient high dimensional proposal distributions for MCMC.

4.1 Markov Chain Monte Carlo

Suppose we want to draw samples $\{X^n\}_{n=1}^N$ from a distribution $\pi(X|Z)$. In most cases, it is difficult to sample independently from this distribution because the normalizing constant $p(Z) = \int_{\mathcal{X}} \pi(X, Z) dX$ is general unknown and complicated except in a few cases where the state space \mathcal{X} is linear Gaussian and hidden finite. In our scope, we only consider the space of X is countable¹ so the MCMC which is introduced here is to limit to the countable state space. MCMC method is an algorithm which allows to draw samples from $\pi(\theta|Z)$ which are slightly dependent by using a Markov chain. This section will review general MCMC methods based on [150, p.206-214], [58] and [151].

¹A state space S is countable if S is discrete, with a finite or countable number of elements, and with \mathcal{S} the σ -field of all subsets of S .

MCMC approaches are so-named because they rely on constructing a Markov chain (MC) which has the desired target distribution π as its equilibrium distribution. In MCMC, an MC is constructed from transition kernel $\mathbf{K}(\cdot, \cdot)$ defined on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ such that $\mathbf{K}(x, \cdot)$ is a probability measure and $\mathbf{K}(\cdot, A)$ is measurable for all $A \in \sigma(\mathcal{X})$ where $\sigma(\mathcal{X})$ is a σ -algebra of $\mathcal{X} \subseteq \mathbb{R}^d$ [150]. When \mathcal{X} is discrete, the transition kernel simply is a transition matrix² with elements

$$\mathbb{P}(X_n = z | X_{n-1} = x).$$

In the continuous case, the kernel also denotes the conditional density $\mathbf{K}(x, x')$ of the transition $\mathbf{K}(x, \cdot)$, that is

$$\mathbb{P}(X \in A | x) = \int_A \mathbf{K}(x, x') dx'$$

The following section will describe a Markov chain and its properties.

4.1.1 Markov chains

The purpose of this section is to briefly provide the foundations of a Markov chain (MC) and its properties which are used in MCMC. For further details, the reader is referred to the books which the material in this section are mainly based on [19, 63, 64, 110, 152, 153, 156, 157]. Classification of MC states and their properties are introduced. the MC describes how states evolve over time. In this section we will classify the states and describes their properties. The long term behavior of an MC is characterized by the stationary or equilibrium distribution which is of major importance. Background material for this section is given in Appendix A and Appendix A.4. We start this section by the definition of a Markov chain.

Let (X_0, X_1, \dots) be a sequence of measurable random variables on a space S equipped with σ -algebra. A MC is defined as follows

Definition 4.1: A sequence $\mathbf{X} = (X_0, X_1, \dots)$, $X_0, X_1, \dots \in S$ is called a (discrete-time) **Markov chain** if it satisfies the Markov condition:

$$\mathbb{P}(X_n | X_0, X_1, \dots, X_{n-1}) = \mathbb{P}(X_n | X_{n-1}), \quad \forall n \geq 1, \forall X_0, \dots, X_n \in S. \quad (4.1)$$

Markov chains $\mathbf{X} = (X_0, X_1, \dots)$ on a countable state space³ S are explored in the remaining of this section.

4.1.1.1 Initial distribution and Transition Matrix

Denote by $\Omega = \prod_{n=0}^{\infty} S_n$ the sample space on which the Markov chain \mathbf{X} lives where S_n is a copy of S . Ω is equipped with the σ -algebra $\mathcal{F} = \bigcup_{n=1}^{\infty} \mathcal{S}_n$ where \mathcal{S}_n is a copy of \mathcal{S} [110, p.53].

²A transition matrix is a matrix used to describe the transitions of a Markov chain.

³See in Appendix A.4.

Chains are determined by the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and by the different initial distributions. Let ν be an *initial distribution* on \mathcal{S} , \mathbf{p}_ν is the probability distribution with initial distribution ν as on \mathcal{F} and is constructed as follows for any $A \in \mathcal{F}$

$$\mathbf{p}_\nu(\mathbf{X} \in A) = \sum_{i \in \mathcal{S}} \nu(i) \mathbb{P}(\mathbf{X} \in A | X_0 = i)$$

where

$$\mathbb{P}(\mathbf{X} \in A | X_0 = i) = \sum_{i \in \mathcal{S}} \mathbf{p}_i(\mathbf{X} \in A), \quad (4.2)$$

$\mathbf{p}_i(\mathbf{X} \in A)$ is the probability distribution on \mathcal{F} which is obtained when the initial distribution is the Dirac measure⁴ δ_i at i . From the definition 4.1 of the MC, the MC on the countable state space with initial distribution ν can be defined as follows.

Definition 4.2 (Countable State Space MC): A sequence $\mathbf{X} = (X_0, X_1, X_2, \dots)$ taking values on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a MC if for every n , any initial distribution ν of a chain (i.e. X_0 is distributed from ν) and any sequence of state $\{i_0, i_1, \dots, i_n\}$, then

$$\mathbf{p}_\nu(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \nu(i_0) \prod_{t=1}^n \mathbb{P}(X_t = i_t | X_{t-1} = i_{t-1}). \quad (4.3)$$

In general the transition probability that the MC visits state j from state i after one time step i.e. $\mathbb{P}(X_1 = j | X_0 = i)$ and that the chain visits state j at the $(n+1)$ th step from state i at the n th step, i.e. $\mathbb{P}(X_{n+1} = j | X_n = i)$ are not necessary the same. In order to eliminate this complication of a general MC, the MC here is restricted to the homogeneous case where the transition probabilities only depend on the value of the states in \mathcal{S} but not on the time step n .

Definition 4.3: The chain is called homogeneous if

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) \quad \text{for all } n \geq 1; i, j \in \mathcal{S}.$$

From equation (4.2) and the homogenous property of a MC, equation (4.3) can be written as

$$\mathbf{p}_\nu(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \nu(i_0) \prod_{t=1}^n \mathbb{P}(X_t = i_t | X_{t-1} = i_{t-1}). \quad (4.4)$$

The transition matrix $\mathbf{P} = (p_{ij})$ is the $|\mathcal{S}| \times |\mathcal{S}|$ matrix where transition probabilities are $p_{ij} = \mathbb{P}(X_1 = j | X_0 = i)$. When a MC is homogeneous, from (4.3) and formula of conditional probability, the conditional probabilities of the process \mathbf{X} can be induced as follows

$$\mathbf{p}_\nu(X_{n+1} = j | X_n = i, X_{n-1} = l_{n-1}, \dots, X_0 = l_0) = \mathbb{P}(X_{n+1} = j | X_n = i). \quad (4.5)$$

⁴See definition in Appendix A.1.

The transition matrix \mathbf{P} describes the evolution of X one time step ahead. Similarly the n th step transition matrix describes the evolution of X n step ahead and is defined in the following definition.

Definition 4.4: The n th step transition matrix $\mathbf{P}^n = \{p_{ij}(n) : i, j \in S\}$ is the matrix of n -step transition probabilities $p_{ij}(n) = \mathbb{P}(X_n = j | X_0 = i)$.

\mathbf{P}^n can be expressed using the Chapman-Kolmogorov equation as given in the next theorem.

Theorem 4.1: For any $i, j \in S$,

$$p_{ij}(n) = \sum_{k \in S} p_{ik}(m)p_{kj}(n-m), \quad 0 \leq m \leq n. \quad (4.6)$$

Therefore, $\mathbf{P}^n = \mathbf{P}^m \mathbf{P}^{n-m}$ the n th power of \mathbf{P} . (4.6) is called the Chapman-Kolmogorov.

Let $\mu_i^{(n)} = \mathbb{P}(X_n = i)$ be the mass function of X_n , and let $\boldsymbol{\mu}^{(n)}$ for the row vector with entries $(\mu_i^{(n)} : i \in S)$, $n \geq 0$. The relationship between the mass functions at different time steps is shown in the next lemma.

Lemma 4.1: $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)} \mathbf{P}^n$ and hence $\boldsymbol{\mu}^{(m+n)} = \boldsymbol{\mu}^{(m)} \mathbf{P}^n$.

4.1.1.2 Stopping times and strong Markov property

The evolution of MCs will be explored in this subsection. It describes the strong Markov property that is used for evaluating conditional probabilities given certain 'random times', called **stopping times**. We start with some fundamental material on stopping times.

Definition 4.5: A random variables $\mathfrak{T} : \Omega \rightarrow \mathbb{N}$ is a stopping time for a MC \mathbf{X} (with respect to a filtration⁵ $\mathcal{F}_n, n = 1, 2, \dots$) if for any initial distribution ν the event $\{\omega \in \Omega : \mathfrak{T}(\omega) = n\} \in \mathcal{F}_n$.

The natural filtration $\mathcal{F}_n, n = 0, 1, \dots$ is the σ -algebra generated by (X_0, \dots, X_n) and a stopping time has the property that it can be determined at time n if $\mathfrak{T}(\omega) = n$.

Important examples of stopping times are hitting times. Define the hitting time of a subset $A \subset S$ by $\mathfrak{T} = \min\{n \geq 1 : X_n \in A\}$.

Theorem 4.2 (Strong Markov property): Suppose that \mathfrak{T} is a finite-valued stopping time for a MC \mathbf{X} on S . Then, for any $i \in S$ and $i_1, i_2, \dots, j_1, \dots, j_m \in S$ and $m \geq 1$,

$$\begin{aligned} \mathbb{P}(X_{\mathfrak{T}+1} = j_1, \dots, X_{\mathfrak{T}+m} = j_m | X_0 = i_0, \dots, X_{\mathfrak{T}-1} = i_{\mathfrak{T}-1}, X_{\mathfrak{T}} = i) \\ = \mathbb{P}(X_1 = j_1, \dots, X_m = j_m | X_0 = i). \end{aligned} \quad (4.7)$$

Remark: The strong Markov property in (4.7) is stated m steps to the future, but it also holds for the entire future of the chain. For any $i \in S$ and $B \subset S^\infty$,

$$\mathbb{P}\{(X_{\mathfrak{T}+1}, X_{\mathfrak{T}+2}, \dots) \in B | X_0, \dots, X_{\mathfrak{T}-1}, X_{\mathfrak{T}} = i\} = \mathbb{P}\{(X_1, X_2, \dots) \in B | X_0 = i\}. \quad (4.8)$$

⁵See filtration in Appendix A.1.

Another equivalent statement is that, for any bounded function $f : S^\infty \rightarrow \mathbb{R}_+$,

$$\mathbb{E}(f(X_{\bar{x}+1}, X_{\bar{x}+2}, \dots) | X_0, \dots, X_{\bar{x}-1}, X_{\bar{x}} = i) = \mathbb{E}(f(X_1, X_2, \dots) | X_0 = i). \quad (4.9)$$

Loosely speaking, the strong Markov property means that a MC regenerates or starts anew at the stopping time.

4.1.1.3 Classification of states

Definition 4.6 (Recurrent and Transient state): A state i is called **recurrent** if

$$\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$$

which is to say that the probability of eventual return to i , having started from i , is 1. If this probability is strictly less than 1, the state is called **transient** (see Figure 4.1).

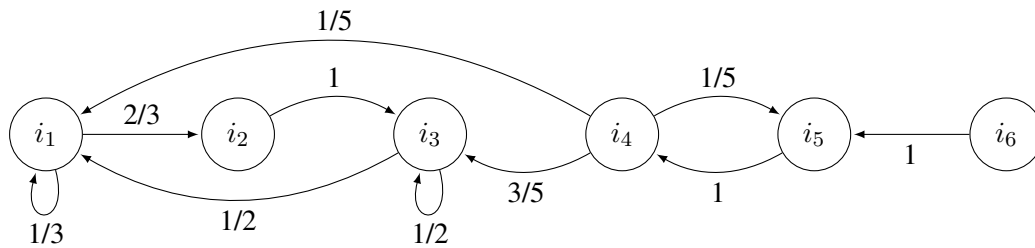


Figure 4.1: States i_4 , i_5 and i_6 are transient whereas states i_1 , i_2 and i_3 are recurrent.

As a result of Definition 4.6, a recurrent state will be revisited several times and a transient state may be revisited several times. One also interested in the probability of the state i ever being visited.

Let $f_{ij}(n) = \mathbb{P}(X_l \neq j, 0 \leq l < n, X_n = j | X_0 = i)$ be the **probability that the first visit to the state j occurs the n th time step** starts from state i at time 0. Define $f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n)$ to be the probability that the chain ever visits to the state j , starting from state i . The state j is recurrent if $f_{jj} = 1$.

We interest in the average time the state i is visited. Let $T_j = \min\{n \geq 1 : X_n = j\}$ be the time the chain first visit state j , with the convention that $T_j = \infty$ if the chain never visits state j . T_j is the hitting time (example of stopping time).

Definition 4.7: The **mean recurrence time** μ_i of a state i is defined as

$$\mu_i = \mathbb{E}(T_i | X_0 = i) = \begin{cases} \sum_n n f_{ii}(n), & \text{if } i \text{ is recurrent,;} \\ \infty, & \text{if } i \text{ is transient.} \end{cases}$$

Definition 4.8: A recurrent state i is called **null** if $\mu_i = \infty$ and **positive (or non-null)** if $\mu_i < \infty$.

The return times of a state of a MC which plays important role for the convergence of a MC is described next

Definition 4.9: Let $d(i) = \gcd\{n : p_{ii}(n) > 0\}$ be the greatest common divisor of the times at which a return to state i is possible. State i is called *periodic* with period $d(i)$ if $d(i) > 1$ and *aperiodic* if $d(i) = 1$.

Definition 4.10: A state i is called *ergodic* if it is positive recurrent and aperiodic.

The properties of a state of the MC is summarized in Figure 4.2 Apart from the property of the

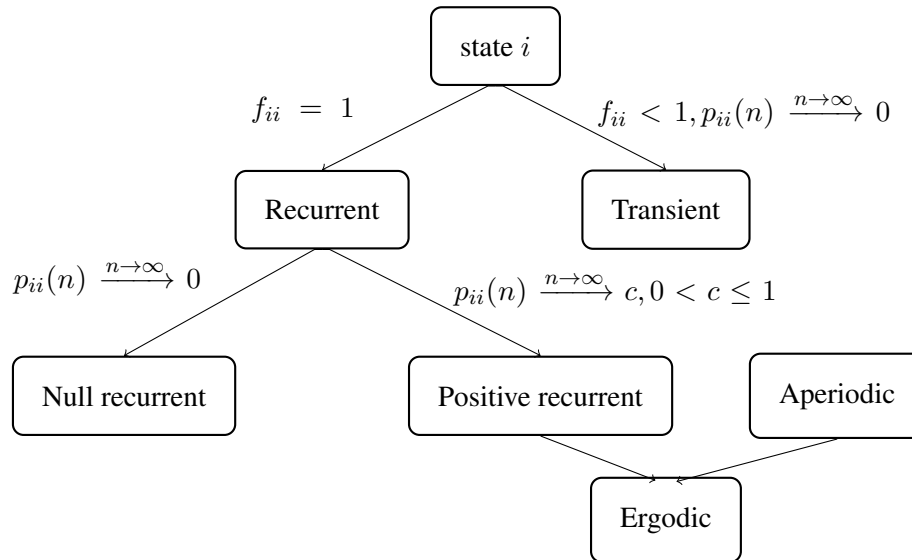


Figure 4.2: Classification of a state of a MC in term of $p_{ii}(n)$.

states of MC, the correlation between the states also plays crucial part to the convergence of the MC. Thus, this relationship is described in the next section.

4.1.1.4 Classification of chains

This section presents the relationship between states of a Markov chain (MC) using the material of the previous subsections.

Definition 4.11: A state i is said to *communicate* with state j , written $i \rightarrow j$, if the chain may ever visit state j with positive probability, having started from i . That is, $i \rightarrow j$ if $p_{ij}(m) > 0$ for some $m \geq 0$. A state i and j are said to *intercommunicate* if $i \rightarrow j$ and $j \rightarrow i$, in which case we write $i \leftrightarrow j$. For completeness, define $p_{ij}(0) = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$

It follows that if $i \neq j$, then $i \rightarrow j$ if and only if $f_{ij} > 0$. In Figure 4.2, state i_6 communicates with state $i_j, j = 1, \dots, 5$; state i_5 communicates with state $i_j, j = 1, \dots, 3$ and intercommunicates with state i_4 ; and state $i_j, j = 1, \dots, 3$ intercommunicates with each other.

The intercommunication property results in the connection between transient, recurrent states which is represented next.

Theorem 4.3: If $i \leftrightarrow j$ then:

1. i and j have the same period,
2. i is transient if and only if j is transient,
3. i is null recurrent if and only if j is null recurrent.

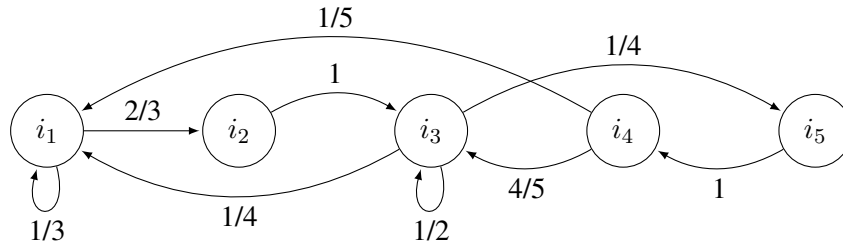


Figure 4.3: All states of a MC are intercommunicate. State i is recurrent so by theorem 4.3, all states are recurrent

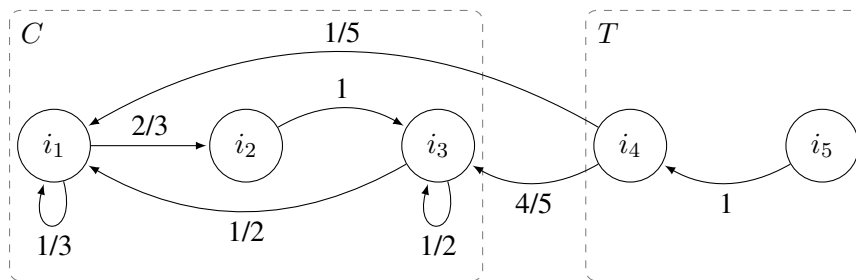


Figure 4.4: The state of the MC in Figure 4.2 are grouped in two sets: a set of recurrent states C and a set of transient states T .

Definition 4.12: A set C of states is called:

1. **closed** if $p_{ij} = 0$ for all $i \in C, j \notin C$,
2. **irreducible** if $i \leftrightarrow j$ for all $i, j \in C$

By this definition, set C in Figure 4.4 is closed. Once a chain takes a value in a closed set C then it never leaves C subsequently. A closed set containing exactly one state is called an absorbing state, then this state is **absorbing**. The equivalence class \leftrightarrow is obviously irreducible. An irreducible set C is said to be aperiodic (or recurrent, null recurrent, positive recurrent, and so on) if all states in C have this property.

Theorem 4.4 (Decomposition theorem): The state space S can be partitioned uniquely as

$$S = T \cup C_1 \cup C_2 \cup \dots$$

where T is the set of transient states, and the C_i are irreducible closed sets of recurrent states. The transient matrix has the form (if $S = T \cup C_1 \cup C_2 \cup \dots \cup C_n$)

$$P = \begin{pmatrix} P_1 & 0 & 0 & \dots & 0 \\ 0 & P_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & P_n & 0 \\ Q_1 & Q_2 & \dots & Q_n & Q \end{pmatrix} \quad (4.10)$$

where $P_k = \{p_{ij} : i, j \in C_k\}$, $Q_k = \{p_{ij} : i \in T, j \in C_k\}$, $Q = \{p_{ij} : i, j \in T\}$, $k \in \{1, 2, \dots, n\}$.

This theorem is illustrated in Figures 4.5. The example in Figure 4.5 has the transition matrix

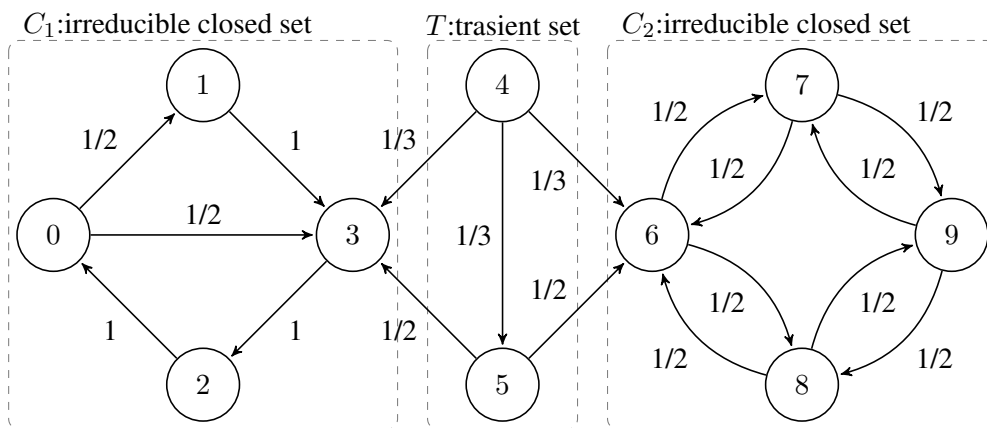


Figure 4.5: Set of states of a Markov chain is categorized into transient set and irreducible sets

4.10 where $n = 2$, and the probabilities are $P_1 = \{p_{ij} : i, j \in \{0, 1, 2, 3\}\}$, $P_2 = \{p_{ij} : i, j \in \{6, 7, 8, 9\}\}$, $Q_1 = \{p_{ij} : i \in \{4, 5\}, j \in \{0, 1, 2, 3\}\}$, $Q_2 = \{p_{ij} : i \in \{4, 5\}, j \in \{6, 7, 8, 9\}\}$, and $Q = \{p_{ij} : i, j \in \{4, 5\}\}$.

The classification of MCs ends this section with few more important and popular terms.

- Two states that intercommunicate each other are said to be in the same class.
- The MC is irreducible if there is only one class - that is, all states intercommunicate each other by Definition 4.12. In this case, by theorems 4.4 and 4.3 all states of the chain are either positive recurrent, null recurrent or transient; and its states have the same period
- The MC is called ergodic if it is irreducible and its states have positive recurrent and aperiodic.

When the state space is finite, the following theorem and lemma are useful in practice [157, Theorem 2, p.581]

Theorem 4.5: *If a finite MC is irreducible and aperiodic then it is positive recurrent.*

Theorem 4.6: *If an irreducible MC is aperiodic and positive recurrent then it is ergodic.*

4.1.1.5 Stationary distribution and the limit theorem

A MC X is more interesting in the long term running than in the short term. The MC may behave randomly in general, but under some condition, the MC converges into some distribution. The distributions in which the MC converges are called stationary distributions. This section will discuss stationary distribution and the limit theorem.

Stationary distribution

Definition 4.13: A Markov chain with transition matrix P has a stationary distribution π if

$$\pi = \pi P. \quad (4.11)$$

If X_0 has distribution π , then from Lemma 4.1 X_n has distribution π for all n .

Theorem 4.7: Suppose a MC on a state space S with transition matrix P satisfies $\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j \geq 0, \forall i, j \in S$. If $\sum_j \pi_j = 1$, then $\pi = (\pi_1, \pi_2, \dots)$ is the unique stationary distribution.

The following theorem shows that under some condition a MC eventually visits its particular state.

Theorem 4.8: For any aperiodic state j of a MC, $p_{jj}(n) \xrightarrow{n \rightarrow \infty} \mu_j^{-1}$. Furthermore, if i is any other state then $p_{ij}(n) \xrightarrow{n \rightarrow \infty} \frac{f_{ij}}{\mu_j}$.

The stationary distribution is an important property of a MC, the following theorem shows under which condition the MC has stationary distribution. and if it has stationary distribution, what its stationary distribution is.

Theorem 4.9: All states of an irreducible aperiodic chain C have the same period and belong to one of the following for all $i, j \in C$:

- (i) Either the states are all transient or all null recurrent. In this case $p_{ij}(n) \xrightarrow{n \rightarrow \infty} 0$ and there exists no stationary distribution
- (ii) Or else, all states are positive recurrent, that is, $p_{ij}(n) \xrightarrow{n \rightarrow \infty} \frac{1}{\mu_j} > 0$. In this case $\{\frac{1}{\mu_j} : \mu_j > 0, j \in C\}$

$$\{\frac{1}{\mu_j} : \mu_j = \lim_{n \rightarrow \infty} p_{ij}(n) > 0, j \in C\} \quad (4.12)$$

is a unique stationary distribution where μ_j is the mean recurrence time of state j .

The proof can be found in [153, p.175-177].

This theorem leads to the following criteria for ergodicity that is very important and useful for applications.

Corollary 4.1: An irreducible aperiodic Markov chain C is ergodic if and only if it has a stationary distribution given in (4.12).

Limiting Distribution

In this subsection, the relationship between the limiting distribution, i.e. $p_{ij}(n)$ as $n \rightarrow \infty$ and the existence of a stationary distribution is further explored.

Theorem 4.10: *Let C be an irreducible aperiodic MC, $p_{ij}(n) \xrightarrow{n \rightarrow \infty} \mu_j^{-1}$ for all $i, j \in C$.*

Theorem 4.11 (Limit distribution): *A MC has a limit distribution if and only if the set S of its states has exactly one aperiodic positive recurrent class C such that $f_{ij} = 1, \forall j \in C$ and $i \in S$.*

Theorem 4.12: *For an irreducible, aperiodic MC, the following statements are equivalent.*

1. *The chain is ergodic.*
2. *The chain has a stationary distribution.*
3. *The chain has a limiting distribution.*

When these statements hold, the limiting distribution and the stationary distribution are the same, and they are positive.

4.1.1.6 Reversibility

Let $\mathbf{X} = \{X_n : 0 \leq n \leq N\}$ be an irreducible positive recurrent MC with transition matrix \mathbf{P} and stationary distribution π . Suppose that X_n has distribution π for every n . Define the 'reverse chain' $\mathbf{Y} = \{Y_n : Y_n = X_{N-n}, 0 \leq n \leq N\}$. \mathbf{Y} is a MC by the following theorem.

Theorem 4.13: *The sequence \mathbf{Y} is a Markov chain with $\mathbb{P}(Y_n = j | Y_n = i) = \frac{\pi_j}{\pi_i} p_{ji}$.*

The chain $\mathbf{Y} = \{Y_n : 0 \leq n \leq N\}$ is called the time-reversal of the chain \mathbf{X} .

Definition 4.14: *Let $\mathbf{X} = \{X_n : 0 \leq n \leq N\}$ be an irreducible MC such that X_n has the stationary distribution π for all n . The chain is called reversible if the transition matrices of \mathbf{X} and its time-reversal are the same, that is*

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j. \quad (4.13)$$

The equations (4.13) is called the detail balance equations and is pivotal to the study of reversible chains. An irreducible MC having a stationary distribution π is called *reversible in equilibrium* if its transition matrix $\mathbf{P} = \{p_{ij}, \text{ for all } i, j\}$ is in detailed balance with π

Theorem 4.14: *Let \mathbf{P} be the transition matrix of an irreducible MC \mathbf{X} and suppose that there exists a distribution π such that $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$. Then π is a stationary distribution of the chain. Furthermore, \mathbf{X} is reversible in equilibrium.*

4.1.1.7 Limit Theorem via coupling

Theorem 4.15: *Suppose \mathbf{X} and \mathbf{Y} are independent, irreducible, aperiodic recurrent MCs on S with arbitrary initial distributions, but with the same transition probabilities. Then*

$$\sup_i [\mathbb{P}(X_n = i) - \mathbb{P}(Y_n = i)] \xrightarrow{n \rightarrow \infty} 0.$$

Theorem 4.16 (Limiting distributions): *If \mathbf{X} is an ergodic MC with stationary distribution π then*

$$\sup_i [\mathbb{P}(X_n = i) - \pi_i] \xrightarrow{n \rightarrow \infty} 0.$$

Hence $\boldsymbol{\pi} = (\pi_i : i \in S)$ is the limiting distribution of \mathbf{X} .

MC is summarized in the Figure 4.6

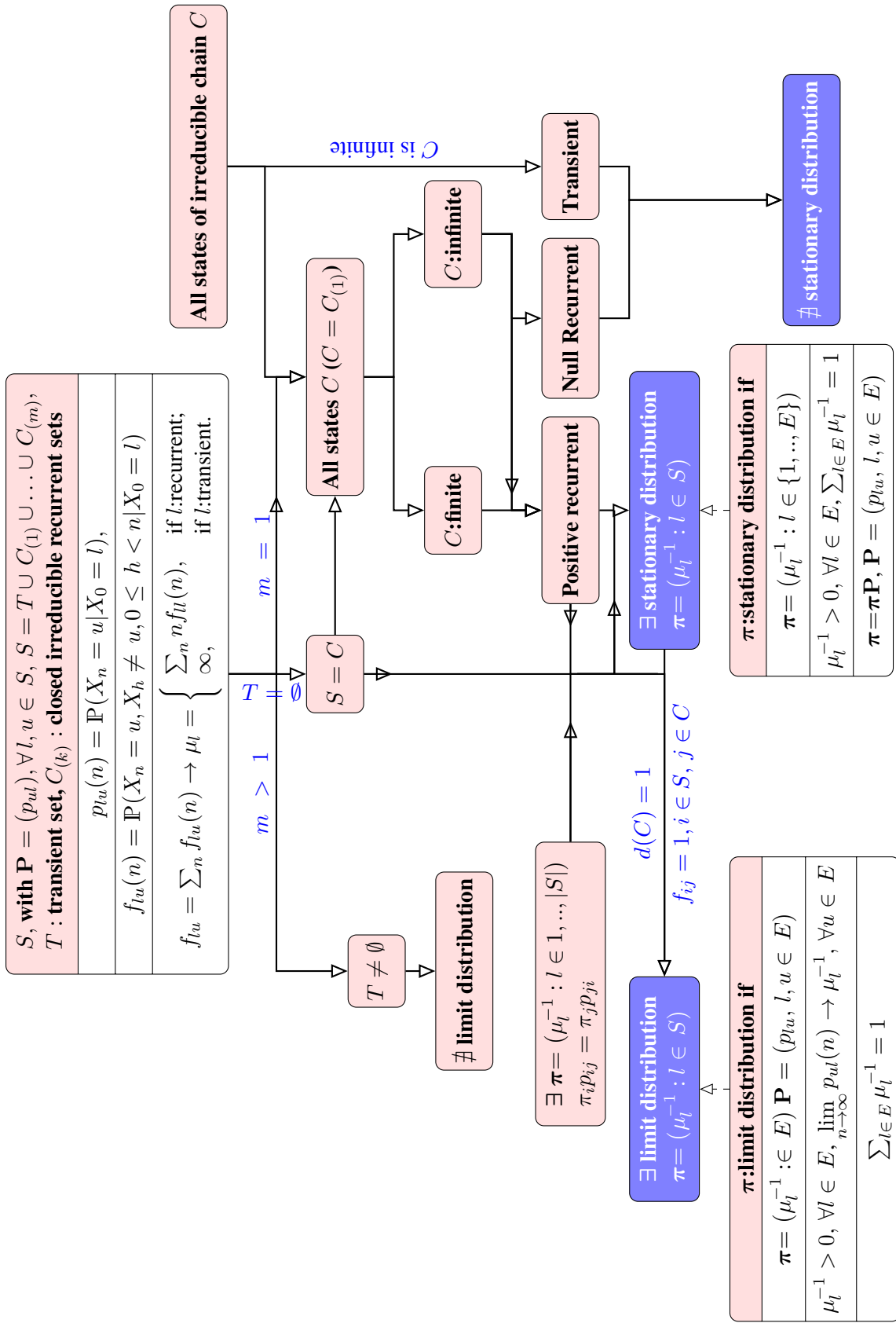


Figure 4.6: Overview of Markov chains

In order to construct a MC which converges to the target distribution π , the following properties must be satisfied

1. The MC satisfies a the detailed balance condition given in Definition 4.13
2. The MC is ergodic.

An irreducible aperiodic MC can be constructed by first defining the starting distribution ν , and then constructing the transition matrix \mathbf{P} given in Definition 4.4 such that the MC finally will reach the target distribution π as the stationary distribution. By Theorem 4.1, it can be written as

$$\pi = \nu_0 \lim_{n \rightarrow \infty} \mathbf{P}^n.$$

At each iteration n , x^n is sampled from the distribution $\nu_0 \mathbf{P}^n$ for $n = 0, 1, \dots$. The time a MC starts from initial distribution ν until it reaches the stationary distributions π is called the burn-in time.

Strictly speaking, it may never reach the stationary distribution exactly with a finite number of steps and the burn-in period is then the period before it is sufficiently close to the stationary distribution. Hence if the initial is close the stationary distribution, the MC converges quickly to the stationary distribution. The two most general and very popular MCMC methods such as Metropolis-Hastings algorithm and Gibbs sampler are designed to construct an ergodic MC which converge to the stationary distribution π .

4.1.2 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm [149] is a Markov Chain Monte Carlo method for obtaining a sequence of random samples from a probability distribution from which direct sampling is difficult. The MH algorithm can draw samples from any distribution $\tilde{\pi}$. This distribution is only known up to a proportionality constant. In Bayesian applications, the normalization factor is often computationally intractable, so the ability to generate a sample without knowing this constant of proportionality is a major virtue of the algorithm. The algorithm uses a proposal density $q(\cdot|x), x \in \mathcal{X}$ to generate a MC. The MH algorithm associated with target density π and the proposal distribution q produces the MC $x(n) \in \mathcal{X}, n = 0, 1, \dots$ given in Algorithm 2.

Algorithm 2 : Metropolis-Hastings (MH) Algorithm

At iteration $n = 0$: initialize arbitrarily $x(0)$

At iteration $n = 1, \dots, N$:

- Sample $y \sim q(\cdot|x(n-1))$
- Compute the acceptance rate

$$\alpha(x(n-1), y) = \min \left\{ 1, \frac{\tilde{\pi}(y)}{\tilde{\pi}(x(n-1))} \frac{q(x(n-1)|y)}{q(y|x(n-1))} \right\} \quad (4.14)$$

- If $\alpha(x(n-1), y) \geq u$ where u is sampled from the uniform distribution on $[0, 1]$, set $x(n) = y$ otherwise set $x(n) = x(n-1)$.
-

Algorithm 2 only depends on the ratios

$$\frac{\tilde{\pi}(y)}{\tilde{\pi}(x(n-1))}, \quad \frac{q(x(n-1)|y)}{q(y|x(n-1))}$$

and it is therefore independent of normalizing constants. Thus $q(\cdot|x(n-1))$ is assumed to be known up to a constant that is independent of $x(n-1)$. The MC starts with $x(0)$ such that $\tilde{\pi}(x(0)) > 0$. As a convention, $\alpha(y, x(n-1))$ is 0 if both $\tilde{\pi}(x(n-1)) = 0$ and $\tilde{\pi}(y) = 0$. According to [149], the MC generated from MH Algorithm 2 converges to its stationary distribution π if the followings hold

- $\text{supp}(\pi) \subset \bigcup_{x \in \text{supp}(\pi)} \text{supp}(q(\cdot|x))$,
- π is bounded and positive on every compact set of its support,
- there exists positive numbers ϵ and δ such that

$$q(y|x) > \epsilon \quad \text{if} \quad |x - y| < \delta$$

This rationale behind this is that the proposal distribution $q(y|x)$ allows moves in a neighborhood of x with diameter δ .

Assume that N_0 is the burn-in time of the MC generated by Algorithm 2, the distribution π can be approximated as follows

$$\pi(x) \approx \sum_{n=N_0+1}^N \delta(x(n) - x)$$

4.1.3 Gibbs Sampler

Gibbs Sampler is a special case of Metropolis-Hastings sampling wherein the random value is always accepted (i.e. $\alpha(x(i-1), y) = 1$ where $\alpha(x(i-1), y)$ is given in (4.14)). This sampling is used to update each components of a state $x(n) = X = (X_1, \dots, X_d) \in \mathbb{R}^d$ where X_i is the i component of X . Denote the full conditional distribution by

$$\pi(X_i | X_j, j \neq i) = \frac{\pi(\{X_j, j \neq i\}, X_i)}{\int \pi(\{X_j, j \neq i\}, X_i) dX_i}$$

which is the distribution of the i th component of X conditional on the other components. The Gibbs sampler is

Algorithm 3 : Gibbs Algorithm

At iteration $n = 0$: initialize arbitrarily $x(0)$

At iteration $n > 0$: for $i = 1, \dots, d$, where $x(n) = (x_1, \dots, x_d) \in \mathcal{X} \subseteq \mathbb{R}^d$

- sample $x_i(n) \sim \pi(\cdot | x_{1:i-1}(n), x_{i+1:d}(n-1))$
-

The Gibbs sampling is useful whenever the conditional distribution of each variable is feasible to sample while the joint distribution is unknown or difficult to sample from. If joint distribution of

all variables and the conditional distribution of any variable are difficult to sample from, the Gibbs sampling can be replaced by the MH algorithm. Similar to the MH algorithm, if N_0 is the burn-in time of the MC generated by Algorithm 3, the distribution π can be approximated as follows

$$\pi(x) \approx \sum_{n=N_0+1}^N \delta(x(n) - x)$$

4.2 Particle Markov Chain Monte Carlo methods

Particle Markov Chain Monte Carlo (PMCMC) methods are algorithms which uses the particles sampling from SMC also known as particle filter described in Chapter 2.3.3 as proposal distribution for MCMC [4]. PMCMC methods actually explore the strengths of SMC and MCMC approaches by combining these algorithms to sample from a high dimension probability distribution that cannot be satisfactorily sampled using either SMC or MCMC on its own. In this section, we will present three PMCMC methods, the Particle Independent Metropolis Hastings Sampler (PIMH), the Particle Marginal Metropolis Hastings (PMMH) sampler and the Particle Gibbs Sampler.

Consider the scenario where we are interested in sampling from the posterior distribution $p(\theta, X_{1:t}, Z_{1:t})$, $t = 1, \dots, T$ where $X_{1:t} = (X_1, \dots, X_t)$, $Z_{1:t} = (Z_1, \dots, Z_t)$ and the random variables $X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ follows the Markov process with initial density $X_1 \sim p_0(X_1)$ and transition density $f(\cdot|X_{t-1}, \theta)$, i.e.

$$X_t \sim f(\cdot|X_{t-1}, \theta)$$

for some static parameter $\theta \in \Theta$ which may be multidimensional. X_t is observed indirectly by the measurement Z_t with the likelihood function $g(Z_t|X_t, \theta)$ i.e.

$$Z_t \sim g(\cdot|X_t, \theta).$$

Given the history of measurements $Z_{1:t} = (Z_1, \dots, Z_t)$, the aim is to perform Bayesian inference. When θ is a known parameter, Bayesian inference relies on the posterior distribution

$$p(X_{1:t}|Z_{1:t}, \theta) \propto p(X_{1:t}, Z_{1:t}|\theta) = \prod_{i=1}^t f(X_i|X_{i-1}, \theta)g(Z_i|X_i, \theta) \quad (4.15)$$

where $f(X_1|X_0, \theta) = p_0(X_1, \theta)$. If θ is unknown, the prior density $p(\theta)$ is ascribed to θ . Then Bayesian inference relies on the posterior distribution

$$p(X_{1:t}, \theta|Z_{1:t}) \propto p(X_{1:t}, Z_{1:t}|\theta)p(\theta) = \prod_{i=1}^t f(X_t|X_{t-1}, \theta)g(Z_t|X_t, \theta)p(\theta) \quad (4.16)$$

When the system is non-linear or non-Gaussian, $p(X_{1:t}, \theta|Z_{1:t})$ and $p(X_{1:t}|Z_{1:t}, \theta)$ do not admit closed form expression. This makes the inference difficult in practice. PMCMC methods are

approximations which provide flexible frameworks to carry out the inference. PMCMC refers to MCMC algorithm target the distribution $p(X_{1:t}, \theta | Z_{1:t})$ or $p(X_{1:t} | Z_{1:t}, \theta)$ which relies on the output of an SMC algorithm targeting $p(X_{1:t} | Z_{1:t}, \theta)$, using $N \gg 1$ particles as a proposal distribution for a Metropolis Hastings (MH) update. Targeting the $p(X_{1:t}, \theta | Z_{1:t})$ or $p(X_{1:t} | Z_{1:t}, \theta)$, PMCMC algorithms are in fact 'exact approximations' to the standard MCMC algorithms in the sense that for any fixed number $N \gg 1$ of particles their transition kernels leave the target density invariant. The next subsection will present the construction of the SMC targeting the distribution $p(X_{1:t} | Z_{1:t}, \theta)$.

4.2.1 Sequential Monte Carlo Algorithm

The general sequential Monte Carlo (SMC) Algorithm can be found in Chapter 2.3.3. In this section, we will presents SMC with a particular proposal distribution to draw samples which are used for the MH update in a MCMC algorithm. In sequential Monte Carlo algorithms, for any given $\theta \in \Theta$ the posterior densities $\{p(X_{1:t} | Z_{1:t}, \theta), t \geq 1\}$ are sequentially approximated by the weighted samples $\{X_{1:t}^n, W_t^n\}_{n=1}^N$

$$\hat{p}(X_{1:t} | Z_{1:t}, \theta) \approx \sum_{n=1}^N W_t^n \delta(X_{1:t}^n - X_{1:t}).$$

Specifically, these methods first approximate $p(X_1 | Z_1, \theta)$ using a proposal density $q(X_1 | Z_1, \theta)$ to generate N particles X_1^n and use the discrepancy between these two densities $q(X_1^n | Z_1, \theta)$ and $p(X_1^n | Z_1, \theta)$ as the normalizing weight W_1^n . To produce $N' \leq N$ particles approximately distributed from $p(X_1 | Z_1, \theta)$, N' samples are drawn from the importance sampling approximation $\hat{p}(X_1 | Z_1, \theta)$ of $p(X_1 | Z_1, \theta)$. For notational simplicity, we denote $p_\theta(A|B) = p(A|B, \theta)$. At time $t > 1$,

$$p_\theta(X_{1:t} | Z_{1:t}) \propto p_\theta(X_{1:t}, Z_{1:t}) = p_\theta(X_{1:t-1} | Z_{1:t-1}) f_\theta(X_t | X_{t-1}) g_\theta(Z_t | X_t). \quad (4.17)$$

This identity allows the use of samples $\{X_{1:t-1}^n, w_{t-1}^n(X_{1:t-1}^n)\}_{n=1}^N$ obtained from the previous time step as a source of samples approximately distributed by $p_\theta(X_{1:t-1} | Z_{1:t-1})$ and extends each such particles through the proposal density $q_\theta(X_t | Z_t, X_{t-1})$ to generate the samples distributed approximately by $p_\theta(X_{1:t-1} | Z_{1:t-1}) f_\theta(X_t | X_{t-1}) g_\theta(Z_t | X_t)$. This means the proposal density is chosen as

$$q_\theta(X_{1:t} | Z_{1:t}) = p_\theta(X_{1:t-1}, Z_{1:t-1}) q_\theta(X_t | X_{t-1}, Z_t).$$

Our aim is to perform Bayesian inference, conditional upon some observation $Z_{1:T}$ for some $T \geq 1$, based on the posterior density $p_\theta(X_{1:T} | Z_{1:T}) \propto p_\theta(X_{1:T}, Z_{1:T})$. From (4.15), we have

$$p_\theta(X_{1:T}, Z_{1:T}) = p_\theta(X_1) \prod_{t=2}^T f_\theta(X_t | X_{t-1}) \prod_{t=1}^T g_\theta(Z_t | X_t). \quad (4.18)$$

If $\theta \in \Theta$ is unknown, one ascribes a prior density $p(\theta)$ to θ and Bayesian reference is based on the joint density $p_\theta(X_{1:T}, Z_{1:T}) = p(X_{1:T}, Z_{1:T}|\theta)$. We use the notation $\mathbf{W}_t := (W_t^1, \dots, W_t^N)$ for the normalized importance weights at time t and $P(\cdot|\mathbf{W}_t)$ for the discrete probability distribution on the set $\{1, \dots, N\}$. That is $P(n|\mathbf{W}_t) = W_t^n$ where the parameter \mathbf{W}_t with $W_t^n \geq 0, n \in \{1, \dots, N\}$ and $\sum_{n=1}^N W_t^n = 1$.

Denote $\mathbf{A}_t = (A_t^1, \dots, A_t^N)$ where the variable A_{t-1}^n represent the index of the 'parent' at time $t-1$ of particle $X_{1:t}^n$ for $t \in \mathcal{T} \setminus \{1\}$. The children particles at time t choose their parent particles at time $t-1$ according to the distribution $r(\mathbf{A}_{t-1}|\mathbf{W}_{t-1}) = \prod_{n=1}^N P(A_{t-1}^n|\mathbf{W}_{t-1})$ using the standard multinomial resampling procedure. The introduction of these variables is useful for finding the *genealogy* of the particles and is necessary for the description of the Particle Gibbs sampler later in Section 4.2.4.

Therefore a pseudo code of the SMC algorithm is provided below (see [4]) to obtain samples from $p_\theta(X_{1:T}|Z_{1:T})$ with initial time t_0 .

Algorithm 4 : SMC Algorithm

Input: $\theta, Z_{1:T}$ and number of samples N , initial time t_0 . In general $t_0 = 1$

Output: $\{X_{1:t}^n, w_t^n(X_{1:t}^n), W_t^n, A_{1:t-1}^n\}_{n=1}^N$

At time $t = t_0$: for $n = 1, \dots, N$,

- sample $X_{t_0}^n \sim q_\theta(\cdot|Z_{t_0})$,
- compute and normalize the weights

$$w_{t_0}(X_{t_0}^n) := \frac{p_\theta(X_{t_0}^n, Z_{t_0})}{q_\theta(X_{t_0}^n|Z_{t_0})}, \quad (4.19)$$

$$W_{t_0}^n := \frac{w_{t_0}(X_{t_0}^n)}{\sum_{m=1}^N w_{t_0}(X_{t_0}^m)}, \quad (4.20)$$

- assign $\mathbf{W}_1 := (W_1^1, \dots, W_1^N)$.

At time $t = t_0 + 1, \dots, T$: for $n = 1, \dots, N$,

- sample $A_{t-1}^n \sim P(\cdot|\mathbf{W}_{t-1})$
- sample $X_t^n \sim q_\theta(\cdot|Z_t, X_{t-1}^{A_{t-1}^n})$, set $X_{1:t}^n = (X_{1:t-1}^{A_{t-1}^n}, X_t^n)$,
- compute and normalize the weights

$$\begin{aligned} w_t(X_{1:t}^n) &:= \frac{p_\theta(X_{1:t}^n, Z_{1:t})}{p_\theta(X_{t-1}^{A_{t-1}^n}, Z_{1:t-1})q_\theta(X_t^n|Z_t, X_{t-1}^{A_{t-1}^n})} \\ &= \frac{f_\theta(X_t^n|X_{t-1}^{A_{t-1}^n})g_\theta(Z_t|X_t^n)}{q_\theta(X_t^n|Z_t, X_{t-1}^{A_{t-1}^n})}, \end{aligned} \quad (4.21)$$

$$W_t^n := \frac{w_t(X_{1:t}^n)}{\sum_{m=1}^N w_t(X_{1:t}^m)},$$

- assign $\mathbf{W}_t := (W_t^1, \dots, W_t^N)$.
-

In this description, for $n = 1, \dots, N$ and $t \in \mathcal{T}$ we introduce B_t^n the index of the ancestor particle of X_t^n at generation t . More formally for $n = 1, \dots, N$ we define $B_T^n = n$ and for $t \in \mathcal{T} \setminus \{T\}$ we have the following backward recursion relation $B_t^n := A_t^{B_{t+1}^n}$. As a result for any $n = 1, \dots, N$ we have $X_{1:T}^n = (X_1^{B_1^n}, X_2^{B_2^n}, \dots, X_{T-1}^{B_{T-1}^n}, X_T^{B_T^n})$ with the ancestral lineage $B_{1:T}^n = (B_1^n, B_2^n, \dots, B_T^n = n)$. Let $O_t^m = \sum_{n=1}^N \mathbb{I}(A_t^n = m)$ be the number of offsprings of the particle m at time t and $s(\cdot | \mathbf{W}_t)$ the corresponding distribution of $\mathbf{O} = (O_t^1, \dots, O_t^N)$.

Once sample particles are collected, an approximation of the target distribution $p_\theta(X_{1:T} | Z_{1:T})$ is given by

$$\hat{p}_\theta(X_{1:T} | Z_{1:T}) = \sum_{n=1}^N W_T^n \delta(X_{1:T} - X_{1:T}^n) \quad (4.22)$$

Furthermore this SMC algorithm provides us with an estimate of the marginal likelihood of $p_\theta(Z_{1:T})$ given by

$$\hat{p}_\theta(Z_{1:T}) := \hat{p}_\theta(Z_1) \prod_{t=2}^T \hat{p}_\theta(Z_t | Z_{1:t-1}) \quad (4.23)$$

where an estimate $\hat{p}_\theta(Z_t | Z_{1:t-1})^6$ of $p_\theta(Z_t | Z_{1:t-1})$ is

$$\hat{p}_\theta(Z_t | Z_{1:t-1}) = \frac{1}{N} \sum_{n=1}^N w_t(X_{t-1:t}^n). \quad (4.24)$$

with convention that $\hat{p}_\theta(Z_1 | Z_0) = \hat{p}_\theta(Z_1)$ and $p_\theta(Z_1 | Z_0) = p_\theta(Z_1)$.

The Algorithm 4, (4.22) and (4.24) will be used in Sections 4.2.2-4.2.4.

In order to derive conditions for the proposal distribution $q_\theta(X_t | X_{1:t-1}) p_\theta(X_{1:t-1} | Z_{1:t-1})$ to approximate $p_\theta(X_{1:t} | Z_{1:t})$, we define, for $t = 1, \dots, T$ [4]

$$\mathcal{S}^\theta = \{X_{1:t} \in \mathcal{X}^t : p_\theta(X_{1:t} | Z_{1:t}) > 0\}, \quad (4.25)$$

$$\mathcal{Q}^\theta = \{X_{1:t} \in \mathcal{X}^t : q_\theta(X_t | X_{1:t-1}, Z_t) p_\theta(X_{1:t-1} | Z_{1:t-1}) > 0\} \quad (4.26)$$

with the convention $p_\theta(X_{1:0} | Z_{1:0}) := 1$ and $q_\theta(X_1 | X_{1:0}, Z_1) = q_\theta(X_1 | Z_1)$. Note that \mathcal{S}^θ and \mathcal{Q}^θ are indeed the supports of $p_\theta(X_{1:t} | Z_{1:t})$ and the proposal distribution

$$q_\theta(X_t | X_{1:t-1}, Z_t) p_\theta(X_{1:t-1} | Z_{1:t-1})$$

respectively. The proposal distribution or importance distribution are used interchangeably throughout the thesis

The following standard minimal assumptions is sufficient to establish the convergence of PMCMC algorithms

(API) For any θ we have $\mathcal{S}^\theta \subseteq \mathcal{Q}^\theta$ for $t = 1, \dots, T$.

⁶See the proof in Appendix B.

(**AP2**) For $n = 1, \dots, N$ and $t = 1, \dots, T$ the re-sampling scheme satisfies

$$\mathbb{E}[O_t^m | \mathbf{W}_t] = NW_t^m, \quad (4.27)$$

$$\text{and } r(A_t^m = n | \mathbf{W}_t) = W_t^n. \quad (4.28)$$

Assumption (**AP1**) shows that sampling from importance density

$$q_\theta(X_t | X_{1:t-1}, Z_t) p_\theta(X_{1:t-1} | Z_{1:t-1})$$

will cover the support of $p_\theta(X_{1:t} | Z_{1:t})$. Assumption (**AP2**) is related to the re-sampling scheme. The unbiased condition in (4.27) ensures that Algorithm 4 will propagate the most promising particles and approximates consistently the distribution $\{p_\theta(X_{1:t} | Z_{1:t}), t = 1, \dots, T\}$ and the normalizing constant $\{p_\theta(Z_{1:t}), t = 1, \dots, T\}$. The condition in (4.28) is not usually satisfied in practice, for computational efficiency, \mathbf{O} is often constructed first according to distribution $s(\cdot | \mathbf{W}_t)$ such that (4.27) holds (i.e. without explicit reference to \mathbf{A}_t). Please see [4] for more details on the construction of \mathbf{O} .

(**AP3**) There exists a sequence of constants $\{c_t : t = 1, \dots, T\}$ such that for any $X_{1:t} \in \tilde{\mathcal{X}}^t$,

$$w(X_{1:t}) \leq c_t \quad (4.29)$$

(**AP4**) There exists a probability $\nu(\cdot)$ on \mathcal{X} and $0 < w_m, w_M, \varepsilon_m, \varepsilon_M < \infty$ such that for any $t = 1, \dots, T$ and any $X_{1:t} \in \mathcal{X}^t$,

$$\begin{aligned} w_m &\leq w_n(X_{1:t}) \leq w_M \text{ and} \\ \varepsilon_m \nu(X_t) &\leq q_\theta(X_t | X_{1:t-1}, Z_t) \leq \varepsilon_M \nu(X_t) \end{aligned} \quad (4.30)$$

As discussed in [4], assumption (**AP3**) aids rapid convergence of SMC and assumption (**AP4**) mitigates the propagation of errors. The impact of these assumptions on the performance of a particular PMCMC algorithms will be addressed below.

4.2.2 Particle Independent Metropolis Hastings Sampler

Assume that θ is known. In the standard independent Metropolis-Hastings (IMH) algorithm, the acceptance rate can be written as follows

$$\alpha = \min \left\{ 1, \frac{p_\theta(X_{1:T}^* | Z_{1:T}) \mathbf{q}_\theta(X_{1:T} | Z_{1:T})}{p_\theta(X_{1:T} | Z_{1:T}) \mathbf{q}_\theta(X_{1:T}^* | Z_{1:T})} \right\} \quad (4.31)$$

The optimal choice for proposal distribution $\mathbf{q}_\theta(X_{1:T} | Z_{1:T})$ is $p_\theta(X_{1:T} | Z_{1:T})$ but in many applications this choice is impossible. The Particle Independent Metropolis Hastings Sampler (PIMH) explores the idea of using SMC approximation of $p_\theta(X_{1:T} | Z_{1:T})$ as a proposal distribution for MH update and is described in Algorithm 5.

Algorithm 5 : Particle Independent Metropolis-Hastings Sampler

Input: $Z_{1:T}$, number of samples L and initial time t_0 . In general $t_0 = 1$.

Output: $X_{1:T}(l)$

At iteration $l = 0$:

- Run a SMC algorithm targeting $p_\theta(X_{1:T}|Z_{1:T})$, sample $X_{1:T}(0) \sim \hat{p}_\theta(\cdot|Z_{1:T})$ and denote by $\hat{p}_\theta^{(0)}(Z_{1:T})$ the marginal likelihood estimate.

At iteration $l = 1, \dots, L$:

- Run a SMC algorithm targeting $p_\theta(X_{1:T}|Z_{1:T})$, sample $X_{1:T}^* \sim \hat{p}_\theta(\cdot|Z_{1:T})$, and

$$\alpha = \min \left\{ 1, \frac{\hat{p}_\theta(Z_{1:T})}{\hat{p}_\theta^{(l-1)}(Z_{1:T})} \right\}. \quad (4.32)$$

- If $\alpha \geq u$ where u is sampled from uniform distribution on $[0, 1]$, set $X_{1:T}(l) = X_{1:T}^*$, $\hat{p}_\theta^{(l)}(Z_{1:T}) = \hat{p}_\theta(Z_{1:T})$ otherwise $X_{1:T}(l) = X_{1:T}(l-1)$, $\hat{p}_\theta^{(l)}(Z_{1:T}) = \hat{p}_\theta^{(l-1)}(Z_{1:T})$

Using the following extremely simple form with $\hat{p}_\theta(Z_{1:T})$ as in (4.23), the acceptance rate $\frac{\hat{p}_\theta^*(Z_{1:T})}{\hat{p}_\theta^{(l-1)}(Z_{1:T})}$ is shown to lead to the target distribution $p_\theta(X_{1:T}|Z_{1:T})$ as the stationary distribution [4] and under weak assumption (AP2) theorem 2 in [4, p.292] showed that the PIMH sampler is ergodic.

When θ is unknown, the PMMH is presented in the following section to deal with this situation.

4.2.3 Particle Marginal Metropolis-Hastings (PMMH) Sampler

When θ is unknown and is part of the estimation problem, which in a Bayesian setting relies on the joint posterior density

$$p(\theta, X_{1:T}|Z_{1:T}) \propto p_\theta(X_{1:T}, Z_{1:T})p(\theta). \quad (4.33)$$

The Particle Marginal Metropolis-Hastings (PMMH) algorithm offers the possibility of designing a good algorithm when θ and $X_{1:T}$ are highly correlated [4]. Assume that sampling from the conditional density $p_\theta(X_{1:T}|Z_{1:T})$ for any $\theta \in \Theta$ is feasible and

$$p(\theta, X_{1:T}|Z_{1:T}) = p(\theta|Z_{1:T})p_\theta(X_{1:T}|Z_{1:T}).$$

Thus the proposal density $q_m(\theta^*, X_{1:T}^*|\theta, X_{1:T})$ from $(X_{1:T}, \theta)$ to $(X_{1:T}^*, \theta^*)$ for an Metropolis-Hastings (MH) update is suggested naturally in the following form

$$q_m(\theta^*, X_{1:T}^*|\theta, X_{1:T}, Z_{1:T}) = \mathbf{q}(\theta^*|\theta)p_{\theta^*}(X_{1:T}^*|Z_{1:T}). \quad (4.34)$$

This form shows that $X_{1:T}^*$ is sampled based on the proposed θ^* and we only need to sample θ^* from $\mathbf{q}(\theta^*|\theta)$. This proposal distribution allows us to sample θ^* on the smaller space Θ (for which

the proposal distribution is easier to design) instead of sampling $\theta^*, X_{1:T}^*$ on the product space $\Theta \times \mathcal{X}^T$.

From (4.34), the MH acceptance ratio is given by

$$\begin{aligned} & \frac{p(\theta^*, X_{1:T}^* | Z_{1:T}) q_m(\theta, X_{1:T} | \theta^*, X_{1:T}^*, Z_{1:T})}{p(\theta, X_{1:T} | Z_{1:T}) q_m(\theta^*, X_{1:T}^* | \theta, X_{1:T}, Z_{1:T})} \\ &= \frac{p_{\theta^*}(Z_{1:T}) \mathbf{q}(\theta | \theta^*) p(\theta^*)}{p_{\theta}(Z_{1:T}) \mathbf{q}(\theta^* | \theta) p(\theta)}. \end{aligned} \quad (4.35)$$

PMMH is proposed naturally whenever samples from $p_{\theta}(X_{1:T} | Z_{1:T})$ and the expression for the marginal likelihood $p_{\theta}(Z_{1:T})$ are needed [4, pp.295] by using $\hat{p}_{\theta}(X_{1:T} | Z_{1:T})$ and $\hat{p}_{\theta}(Z_{1:T})$ in place of $p(X_{1:T} | Z_{1:T}, \theta)$ and $p_{\theta}(Z_{1:T})$ respectively in the MMH update on the right hand side of (4.35). The PMMH sampler is given in algorithm 6 for $l = 1, \dots, L$. The following assumption

Algorithm 6 : Particle Marginal Metropolis-Hastings Sampler

Input: $Z_{1:T}$, number of samples L and initial time t_0 . In general $t_0 = 1$.

Output: $\{X_{1:T}(l), \theta(l)\}_{l=1}^L$

At iteration $l = 0$:

- Set $\theta(0)$ arbitrarily,
- run a SMC algorithm targeting $p_{\theta(0)}(X_{1:T} | Z_{1:T})$, sample $X_{1:T}(0) \sim \hat{p}_{\theta(0)}(\cdot | Z_{1:T})$ and denote by $\hat{p}_{\theta(0)}(Z_{1:T})$ the marginal likelihood estimate.

At iteration $l = 1, \dots, L$:

- Sample $\theta^* \sim \mathbf{q}(\cdot | \theta(l-1))$,
- run a SMC algorithm targeting $p_{\theta^*}(X_{1:T} | Z_{1:T})$, denote by $\hat{p}_{\theta^*}(Z_{1:T})$ the marginal likelihood estimate, and
- calculate the acceptance rate

$$\alpha = \min \left\{ 1, \frac{\hat{p}_{\theta^*}(Z_{1:T}) p(\theta^*) \mathbf{q}(\theta(l-1) | \theta^*)}{\hat{p}_{\theta(l-1)}(Z_{1:T}) p(\theta(l-1)) \mathbf{q}(\theta^* | \theta(l-1))} \right\}. \quad (4.36)$$

If $\alpha \geq u$ where u is sampled from uniform distribution on $[0, 1]$, set $X_{1:T}(l) \sim \hat{p}_{\theta^*}(\cdot | Z_{1:T})$, $\theta(l) = \theta^*$, and $\hat{p}_{\theta(l)}(Z_{1:T}) = \hat{p}_{\theta^*}(Z_{1:T})$ otherwise $X_{1:T}(l) = X_{1:T}(l-1)$, $\theta(l) = \theta(l-1)$ and $\hat{p}_{\theta(l)}(Z_{1:T}) = \hat{p}_{\theta(l-1)}(Z_{1:T})$.

are needed to guarantee the convergence of PMMH [4]

(AP5) The MH sampler of density $p_{\theta}(Z_{1:T})p(\theta)$ and proposal density $\mathbf{q}(\theta^* | \theta)$ is irreducible and aperiodic (and hence converges for $p_{\theta}(\cdot | Z_{1:T})$ almost all starting points).

The assumptions **(AP1)**, **(AP2)** and **(AP5)** ensure that the sequence $\{(\theta(l), X_{1:T}(l))\}$ generated by the PMMH sampler will have $p(\theta, X_{1:T} | Z_{1:T})$ as its limiting distribution (see [4, Theorem 4]).

4.2.4 Modified Particle Gibbs Sampler

An alternative to the MMH algorithm to sample from $p(\theta, X_{1:T} | Z_{1:T})$ consists of using the Gibbs sampler which samples iteratively from $p(\theta | X_{1:T}, Z_{1:T})$ and $p_{\theta}(X_{1:T} | Z_{1:T})$. If the potential tedious design of a proposal density for θ can be bypassed by sampling from $p(\theta | X_{1:T}, Z_{1:T})$,

Particle Gibbs sampler is an option. Moreover sampling from $p_\theta(X_{1:T}|Z_{1:T})$ is typically impossible so the possibility of using a particle approximation to this sampler is suggested. To replace samples from an SMC approximation $\hat{p}_\theta(X_{1:T}|Z_{1:T})$ by samples from $p_\theta(X_{1:T}|Z_{1:T})$ does not admit $p_\theta(X_{1:T}|Z_{1:T})$ as stationary distribution since the prespecified path sample $X_{1:T}$ used as the condition for sampling θ is ignorable. In order to assure that approximation $\hat{p}_\theta(X_{1:T}|Z_{1:T})$ admits $p_\theta(X_{1:T}|Z_{1:T})$ as a stationary distribution, the special type of PMCMC update is proposed and is called Conditional SMC algorithm. This algorithm is similar to SMC but the prespecified path $X_{1:T}$ with its ancestral lineage $B_{1:T}$ is ensured to survive all the resampling steps.

4.2.4.1 Conditional SMC Algorithm

At each time step t , this algorithm generates $N - 1$ particles in the standard way with the remaining particles ascribed to a given particle and guaranteed to survive in the re-sampling step. Given a particle $X_{1:T}^{n*}$, we denote $B_{1:T}^{n*}$ its ancestral lineage. The conditional SMC algorithm proceeds as follows.

Algorithm 7 : Conditional SMC algorithm

Input: $Z_{1:T}$; number of samples N ; initial time t_0 ; and $X_{1:T}^{n*}$ and its ancestral lineage $B_{1:T}^{n*}$. In general $t_0 = 1$

Output: $X_{1:t}^n, w_t^n(X_{1:t}^n), W_t^n, A_{1:t-1}^n$ for $n = 1, \dots, N, n \neq B_t^{n*}$ for $t = 1, \dots, T$

At time $t = t_0$:

- For $n \neq B_t^{n*}$, sample $X_t^n \sim q_{\theta_t}(\cdot|Z_t)$
- Compute $w_t(X_t^n)$ using (4.19) and normalize the weights $W_t^n \propto w_t(X_t^n)$.

At time $t = t_0 + 1, \dots, T$

- For $n \neq B_t^{n*}$, sample $A_{t-1}^n \sim \mathcal{F}(\cdot|W_{t-1})$.
 - For $n \neq B_t^{n*}$, sample $X_t^n \sim q_{\theta_t}(\cdot|Z_t, X_{t-1}^{A_{t-1}^n})$
 - compute $w_t(X_{1:t}^n)$ using Eq.(4.21) and normalize the weights $W_t^n \propto w_t(X_{1:t}^n)$
-

Intuitively, this SMC algorithm is understood as updating $N - 1$ particles while keeping one particle fixed together with its weight. Another advantage of the Conditional SMC algorithm is that updating the sub-blocks $X_{a:b}$ one-at-a time is possible. For any $c, d : 1 \leq c < d \leq T$, a rejection-free-way to update this sub-block proceeds in Algorithm 8

Algorithm 8 : Sub-block Update using Conditional SMC algorithm

- Sample an ancestral lineage $B_{c:d}$ uniformly in $\{1, \dots, N\}^{d-c+1}$
 - Run a conditional SMC algorithm targeting $p_\theta(X_{c:d}|X_{1:c-1}, X_{d+1:T}, Z)$ conditional on $X_{c:d}$ and $B_{c:d}$
 - Sample $X_{c:d} \sim \hat{p}_\theta(X_{c:d}|X_{1:c-1}, X_{d+1:T}, Z)$
-

Thus the following Particle Gibbs Sampler algorithm which always accept a new sample is presented as follows

Algorithm 9 : Particle Gibbs Algorithm

Input: $Z_{1:T}$, number of samples L and initial time t_0 . In general $t_0 = 1$.

Output: $\{X_{1:T}(l), \theta(l)\}_{l=1}^L$.

At iteration $l = 0$: sample $\theta(0), X_{1:T}(0), B_{1:T}(0)$ arbitrarily

At iteration $l = 1, \dots, L$

- Sample $\theta(l) \sim p(\theta|X_{1:T}(l-1), Z_{1:T})$;
 - run a Conditional SMC algorithm targeting $p_{\theta(l)}(X_{1:T}|Z_{1:T})$ conditional on $X_{1:T}(l-1)$ and its ancestral lineage $B_{1:T}(l-1)$; and
 - sample $X_{1:T}(l) \sim \hat{p}_{\theta(l)}(\cdot|Z_{1:T})$ and hence $B_{1:T}(l)$ is also implicitly sampled
-

(**AP6**) The Gibbs sampler that defined by the conditionals $p(\theta|X_{1:T}, Z_{1:T})$ and $p_\theta(X_{1:T}|Z_{1:T})$ is irreducible and aperiodic (and hence converges for $p(\theta, X_{1:T}|z_{1:T})$ almost all starting point).

The theorem 5 in [4] shows that this algorithm admits $p(\theta, X_{1:T}|Z_{1:T})$ as stationary distribution and is ergodic under mild assumptions (**AP1**), (**AP2**), (**AP5**) and (**AP6**).

4.3 Conclusion

PMCMC methods have been presented and can be thought of as natural approximations to MCMC when they can not be implemented in the original form. These methods combine the strengths of SMC and MCMC. This combination is useful for sampling from high dimensional and/or complicated probability distributions that cannot be satisfactorily sampled using either SMC method or MCMC method on its own. PMCMC methods uses the particles from SMC algorithm as the proposal distribution for MCMC method. Different approaches to sample from complicated target distributions by suggesting different proposal distributions have been described leading to different PMCMC methods. The PIMH sampler was first described, and it samples from $p(X_{1:T}|Z_{1:T}, \theta)$ where θ is known. This approach uses a very simple form for MH update by using the SMC approximation of marginal distribution $p_\theta(Z_{1:T})$. When θ is unknown, PMMH sampler and Particle Gibbs sampler were described to deal with distributions $p(\theta, X_{1:T}|Z_{1:T})$ with highly correlated parameter $X_{1:T}$ and θ .

Chapter 5

Literature Review in Target Tracking

A Survey of technical papers in the area of target tracking is presented. Section 5.1 covers the development of conventional target tracking techniques which have been around for the last five decades. The last decades have witnessed development of new target tracking methods which are based on random finite set (RFS) theory, and they are discussed in Section 5.2.

5.1 Conventional Target Tracking Techniques

In this section, a survey of conventional target tracking techniques is presented [7, 8, 10, 14, 17]. These conventional techniques apply data association methods along with the single-target Bayesian filtering to solve the target tracking problem. There are two kinds of target tracking problems, single-target tracking and multiple target tracking problems. The single-target tracking problem requires less effort to find the solution because there is at most only one target in the region of interest. Especially, when there is no clutter, the traditional Bayesian filtering described in Chapter 2.3.1 is employed to estimate the target states from the available measurements collected from sensors. When the motion of a target is governed by a linear system, the Kalman filter (KF), which was first proposed by Kalman [81] in 1960, can be applied to estimate the target states. If the linear system is Gaussian, the KF is an optimal Bayesian filter [3]. When the target motion is governed by a non-linear system, the Extended Kalman filter [3], Unscented Kalman filter [79, 80] or particle filter [5, 20, 46, 61] can be employed. In the case where there are measurements which do not come from the target of interest or the target may generate many measurements, the single target tracking problem is called single target tracking in clutter and more difficult because the origin of the measurements is unknown. Each measurement may be either a target-generated measurement or a false alarm. Hence the single-target Bayesian filtering is not directly applicable and many studies [8, 10, 11, 14, 17] are devoted to the particular problem where a target generates at most one measurement. The conventional solutions [8] such as nearest neighbourhood standard filter (NNSF) and the probability data association filter (PDA) have addressed this problem by cleverly combining the data association problem with conventional Bayesian filtering and are described in Subsection 5.1.1. For the multi-target tracking problem, much more effort is required in order to solve the problem, and it is discussed in Subsection 5.1.2.

5.1.1 Single-target Tracking in Clutter

This section addresses the data association for single target tracking in a cluttered environment with random distributed clutter. The model of the dynamic system is assumed known and the target motion is assumed to follow the hidden Markov system model given in (2.17) in Chapter 2.3. The target state is observed indirectly through the system given in (2.18). Although the linear system model is used here, the techniques to be discussed can be also used for the nonlinear system models by carrying out linearization as in the EKF filter. The simplest approach for tracking a target in clutter is known as the nearest-neighbor standard filter (NNSF) and is described in Subsection 5.1.1.2. Another approach is known as Probability Data Association filter (PDAF) and is described in Subsection 5.1.1.3. Both techniques require the definition of a validation gate described in 5.1.1.1. The objective of a validation gate is to limit the region where a target may generate a measurement. The measurements outside this validation gate are unlikely to originate from the target because they are too far from the expected measurement.

The following notations are used throughout this section. Let η be the expected number of false alarms per unit per volume, and let V be the hypervolume of the surveillance region. Thus ηV is the expected number of false measurements in the surveillance region. The number of false measurements (the measurements not having originated from any targets) follows a Poisson process with parameter ηV

$$\pi_{\Lambda,t}(n) = e^{-\eta V} \frac{(\eta V)^n}{n!}, n = 0, 1, \dots$$

The locations of false measurements are modeled as independently and identical distributed (i.d.d.) random variables with uniform probability density function V^{-1} .

5.1.1.1 Validation of Measurement

In this section which is based on [8] we introduce the validation gate. Assume the linear model (2.17) and (2.18) for the target motion and the target generated measurements. In a clutter environment, the sensors also observe false measurements which are not coming from the targets of interest such as thermal noise, terrain reflections, clouds etc. Assume that the predicted target state at time t is given by $\hat{x}_{t|t-1}$ in (2.19). Then the predicted measurement is $H_t \hat{x}_{t|t-1}$ and the associated measurement covariance S_t is given in (2.25). The target-generated measurements at time t conditional on the history of measurement $Z_{1:t-1}$ is normally distributed

$$p(z_t | Z_{1:t-1}) = \mathcal{N}(z_t; H_t \hat{x}_{t|t-1}, S_t). \quad (5.1)$$

It is impractical to consider all measurements available when updating the state estimate because of the cluttered environment. In order to only consider the measurements z_t at time t which has a high probability (given in (5.1)) of being generated from a target with predicted state $\hat{x}_{t|t-1}$, a

validation gate (region) is defined as follows

$$\bar{V}_t(\gamma) = \{z \in Z_t : [z - H_t \hat{x}_{t|t-1}]^{Tr} S_t^{-1} [z - H_t \hat{x}_{t|t-1}] \leq \gamma\} \quad (5.2)$$

where γ is a parameter obtained from the chi-square distribution (see [8, Appendix C, p. 315-319]) and Z_t is the set of measurements at time t . The volume of the validation gate $\bar{V}_t(\gamma)$ is given by

$$V_t = c_{n_z} \gamma^{n_z/2} |S_t|^{1/2} \quad (5.3)$$

where $|S_t|$ is the determinant of S_t and n_z is the dimension of the measurement vector. The parameter γ is chosen such that the probability

$$P_G = \mathbb{P}(\{z \in \bar{V}_t(\gamma)\}) \quad (5.4)$$

that the true target-generated measurement falls in the validation gate is sufficiently high. The target may not be detected and hence no target-generated measurement may exist in the validation gate. This uncertainty is captured in the detection probability

$$p_{D_t} = \mathbb{P}(\{\text{The true measurement is detected}\}). \quad (5.5)$$

The set of validated measurements at time t which may be originated from target states x_t is denoted by

$$Z_t^\gamma = \bar{V}_t(\gamma) = \{z_1, \dots, z_{|\bar{V}_t(\gamma)|}\} \quad (5.6)$$

where $|\bar{V}_t(\gamma)|$ is the number of elements in $\bar{V}_t(\gamma)$. The set of all validated measurements up to time t is denoted by

$$Z_{1:t}^\gamma = (Z_1^\gamma, \dots, Z_t^\gamma). \quad (5.7)$$

Assigning each measurement with the appropriate target is the crux of the data association technique which is discussed in the next sections.

5.1.1.2 Nearest-Neighbour Standard Filter

The nearest neighbor standard filter (NNSF) in [8] is the simplest technique for solving the single-target tracking in clutter by selecting the measurement in the validated measurement closest to the predicted measurement and using it as the target-generated measurement. The technique is summarized as follows.

At time t , the validated measurement nearest to the predicted measurement is chosen

$$\hat{z}_t = \min_{z \in \bar{V}_t(\gamma)} [z - H_t \hat{x}_{t|t-1}]^{Tr} S_t^{-1} [z - H_t \hat{x}_{t|t-1}].$$

where $\bar{V}_t(\gamma)$ is given (5.2). Then \hat{z}_t is used for updating the state of the target in the same manner as in the Kalman filter.

The problems with this approach is that the closest measurement to the predicted target state may not originate from the target being tracked, and error covariance matrix calculated in the filter equations does not account for the possibility of processing an incorrect measurement. When false measurements occur frequently, NNSF performs poorly because of its high probability of track loss. The PDAF was first proposed in [11] to overcome this limitation of NNSF and is described in the next section.

5.1.1.3 Probability Data Association Filter

The PDAF method considers all measurements in the validation region at current time for update when updating the state estimate. It is a suboptimal Bayesian algorithm and is summarized as follows.

Assume that at time $t - 1$ the mean and covariance of the posterior distribution is \hat{x}_{t-1} and \mathbf{P}_{t-1} . Then PDAF uses the predicted mean $\hat{x}_{t|t-1}$, predicted covariance $\mathbf{P}_{t|t-1}$ and Kalman gain W_t in Kalman filter given in (2.19), (2.20) and (2.24) respectively to predict the state estimate at time t as follows.

Define by $\theta_{t,i}$ the event that the measurement $z_i \in Z_t^\gamma$ is target-generated and $\theta_{t,0}$ the event that none of the measurement in the set of validated measurement is target-generated.

Let $\beta_{t,i}$, $i = 1, \dots, |Z_t^\gamma|$ and $\beta_{t,0}$ be the corresponding probabilities of $\theta_{t,i}$ and $\theta_{t,0}$ respectively where $|Z_t^\gamma|$ is the number of measurements in Z_t^γ . Then

$$\beta_{t,i} = \mathbb{P}(\theta_{t,i}|Z_{1:t}^\gamma) = \frac{\mathcal{N}(z_i; H_t \hat{x}_{t|t-1}, S_t)}{\xi \frac{1-p_{D_t} P_G}{P_G} + \sum_{z \in Z_t^\gamma} \mathcal{N}(z; H_t \hat{x}_{t|t-1}, S_t)} \quad (5.8)$$

$$\beta_{t,0} = \mathbb{P}(\theta_{t,0}|Z_{1:t}^\gamma) = \frac{\xi \frac{1-p_{D_t} P_G}{P_G}}{\xi \frac{1-p_{D_t} P_G}{P_G} + \sum_{z \in Z_t^\gamma} \mathcal{N}(z; H_t \hat{x}_{t|t-1}, S_t)} \quad (5.9)$$

where p_{D_t} is given in (5.5), P_G is given in (5.4), and $\xi = \frac{|Z_t^\gamma|}{V_t}$ with V_t given in (5.3). It follows that

$$\beta_{t,0} + \sum_{z_i \in Z_t^\gamma} \beta_{t,i} = 1.$$

Then the updated state at time t is

$$\hat{x}_t = E[x_t|Z_{1:t}^\gamma] = E[x_t|\theta_{t,0}, Z_{1:t}^\gamma] \mathbb{P}(\theta_{t,0}|Z_{1:t}^\gamma) + \sum_{z_i \in Z_t^\gamma} E[x_t|\theta_{t,i}, Z_{1:t}^\gamma] \mathbb{P}(\theta_{t,i}|Z_{1:t}^\gamma). \quad (5.10)$$

$E[x_t|\theta_{t,i}, Z_{1:t}^\gamma]$ is the expectation of the updated target state at time t given that z_i is the target generated measurements and $E[x_t|\theta_{t,0}, Z_{1:t}^\gamma]$ is the expectation of the predicted target state (=

updated target state when there is no measurement). These expectations are given by the KF

$$E[x_t|\theta_{t,0}, Z_{1:t}^\gamma] = \hat{x}_{t|t-1}, \quad E[x_t|\theta_{t,i}, Z_{1:t}^\gamma] = \hat{x}_{t|t-1} + W_t \nu_t(z_i), i = 1, \dots, |Z_t^\gamma|$$

where $\nu_t(z) = z - H_t \hat{x}_{t|t-1}$. Therefore, \hat{x}_t in (5.10) can be written as follows

$$\begin{aligned} \hat{x}_t &= \hat{x}_{t|t-1} \beta_{t,0} + \sum_{z_i \in Z_t} (\hat{x}_{t|t-1} + W_t \nu_t(z_i)) \beta_{t,z} \\ &= \hat{x}_{t|t-1} + W_t \sum_{z_i \in Z_t} \beta_{t,i} \nu_t(z_i). \end{aligned} \quad (5.11)$$

The covariance associated with the updated target state estimate is [8, Eq. (6.27), p. 165]

$$\mathbf{P}_t = E \left[[x_t - \hat{x}_t][x_t - \hat{x}_t]^T | Z_{1:t}^\gamma \right] = \beta_{t,0} \mathbf{P}_{t|t-1} + (1 - \beta_{t,0}) \hat{\mathbf{P}}_t + \tilde{\mathbf{P}}_t \quad (5.12)$$

where

$$\begin{aligned} \hat{\mathbf{P}}_t &= [I - W_t H_t] \mathbf{P}_{t|t-1} \\ \tilde{\mathbf{P}}_t &= W_t \left[\sum_{z_i \in Z_t^\gamma} \beta_{t,i} \nu_t(z_i) (\nu_t(z_i))^{Tr} - \sum_{z_i \in Z_t^\gamma} \beta_{t,i} \nu_t(z_i) \sum_{z_i \in Z_t^\gamma} \beta_{t,i} (\nu_t(z_i))^{Tr} \right] W_t^{Tr}. \end{aligned}$$

Note that $\left(\sum_{z_i \in Z_t^\gamma} \beta_{t,i} \nu_t(z_i) \right)^{Tr} = \sum_{z_i \in Z_t^\gamma} \beta_{t,i} (\nu_t(z_i))^{Tr}$. PDA neither accounts for the initiation of the track on false sensor returns nor for the termination of the target track [36]. An improvement of PDAF is introduced in [120–122, 142, 154] or combination of nearest neighbours and PDA filter is proposed in [2, 35]. Multiple models used with PDA to track the single maneuvering target in clutter [1, 73].

5.1.2 Multiple Target Tracking (MTT) Techniques

In the multi-target tracking (MTT) problem, the model of the dynamic system is assumed known and the target motions are assumed to follow a hidden Markov system model i.e. each target is considered to follow the single target system model given in (2.17) in Chapter 2.3. Each target is assumed to move independently and generates a measurement which may be observed by sensors. The targets may be not detected by the sensors. These measurements generated by targets are as before called target-generated measurements and are modeled in (2.18) in Chapter 2.3. The sensors may detect false measurements which are not generated by any targets. The false measurements are usually assumed to follow Poisson distribution as for the single target tracking problem.

The multi-target tracking (MTT) problem is significantly more complicated than the single-target tracking problem even in the case where there is no clutter i.e. the sensor do not receive any false measurements. The added complexity comes from the uncertainty of the origins of the measurements which need to be used for the update step. Similarly to the single target case, many conventional techniques address this problem by employing general data association meth-

ods [8, 15, 17] and traditional Bayesian filtering to track multiple targets. For example, the global nearest neighbor filter (GNNF) [8, 14, 17] is a simple approach because it only considers the most likely data association at current time by minimizing the total summed distance function which is the sum of distances between the measurements and the predicted measurements or by maximizing the total summed likelihood. The KF is then employed to update tracks using this data association. The GNNF reduces to the NNSF filter for the single target problem and hence the GNNF also suffers from the same drawback as NNSF. Some researchers have attempted to alleviate this limitation by using data from consecutive scans to defer difficult association decisions [14, 47, 161]. The Joint probability data association filter (JPDAF) was proposed as an improvement to the GNN filter restricted to problems with known and fixed number of targets [8, 17]. This technique is the extension of PDAF to MTT problem and it is used in many tracking applications with imaging sensors see [17] and the reference herein. The JPDAF [8, p.222-228] requires much more computation than PDAF because the association between measurements and multiple targets is more complex and a measurement could have originated from more than one target. The JPDA is presented next.

5.1.2.1 Joint Probability Data Association Filter

This section present briefly the extension of PDA given in Section 5.1.1.3 to the Joint Probability Data Association (JPDA). The JPDA filter propagates individual states the same way as done in the PDA filter except for the computation of the joint probabilities. The probability of a data association event between measurement and target is marginalized out from the joint probability.

Assume that there are a fixed number K of targets to be tracked. The dynamic system model and measurement model for each target are given in (2.17) and (2.18) respectively.

The JPDA filter does not consider different validation regions for each target. The entire surveillance region is the validation region for any target when deriving the joint probability. However, the validation region for each target are used for the selection of "feasible joint events".

Denote by Z_t the set of measurements at time t . For $k = 1, \dots, K$ and $i \in \{1, \dots, |Z_t|\}$, let $\theta_{t,i}^k = (k, i)$ denote the event that the measurement $z_i \in Z_t$ has been generated from target k at time t . Moreover, let $\theta_{t,0}^k = (k, 0)$ denote the event that the target k was undetected at time t , and $\theta_{t,i}^0 = (0, i)$ the event that the measurement z_i was not generated by any target.

For $i_0 \in \{0, 1, \dots, |Z_t|\}$ and $j \in \{0, \dots, K\}$, let $\varrho^\tau(\theta_{t,i}^k) = k$, $\varrho^M(\theta_{t,i}^k) = i$. Let $\theta_t(i_0, j)$ be a set of $\theta_{t,i}^k$, $k \in \{0, \dots, K\}$, $i \in \{0, \dots, |Z_t|\}$ with the following properties

- (1) $\theta_{t,i_0}^j \in \theta_t(i_0, j)$;
- (2) $|\{\theta_{t,i}^k \in \theta_t(i_0, j) : i > 0\}| = |Z_t|$;
- (3) if $k \neq k'$, then $\varrho^M(\theta_{t,i}^k) \neq \varrho^M(\theta_{t,i'}^{k'})$ or $\varrho^M(\theta_{t,i}^k) = \varrho^M(\theta_{t,i'}^{k'}) = 0$;
- (4) if $i \neq i'$, then $\varrho^\tau(\theta_{t,i}^k) \neq \varrho^\tau(\theta_{t,i'}^{k'})$ or $\varrho^\tau(\theta_{t,i}^k) = \varrho^\tau(\theta_{t,i'}^{k'}) = 0$.

Then $\theta_t(i_0, j)$ is one of many possible joint associations between target measurement indices and target indices at time t which associates measurement i_0 with target j (property (1)). Moreover, all measurements at time t are either target generated or clutter (property (2)). No two

targets generate the same measurements (property (3)) and a measurement can not be associated with more than one target (property (4)).

Denote by

$$\Omega_t^{i_0,j} = \{\boldsymbol{\theta}_t(i_0, j) : \boldsymbol{\theta}_t(i_0, j) \text{ satisfies properties (1) - (4)}\}.$$

the set of associations between detected targets and measurements (i.e. excluding clutter and undetected targets).

The key of the JPDA algorithm is the conditional probabilities of the joint association $\boldsymbol{\theta}_t(i_0, j)$ at time t

$$\begin{aligned} \mathbf{P}(\boldsymbol{\theta}_t(i_0, j) | Z_{1:t}) &= \frac{1}{c} \lambda^{|Z_t| - |\varrho^M(\alpha(\boldsymbol{\theta}_t(i_0, j)))|} \prod_{k \in \{1, \dots, K\} - \varrho^\tau(\alpha(\boldsymbol{\theta}_t(i_0, j)))} (1 - p_{D_t}^k) \\ &\times \prod_{u \in \alpha(\boldsymbol{\theta}_t(i_0, j))} \mathcal{N}(z_{\varrho^M(u)}; H_t \hat{x}_{t|t-1}^{\varrho^\tau(u)}, S_t^{\varrho^\tau(u)}) p_{D_t}^{\varrho^\tau(u)}. \end{aligned} \quad (5.13)$$

where $p_{D_t}^k$ is the probability of detecting the target k at time t , $\hat{x}_{t|t-1}^k$ is the predicted target state of target k at time t , c is a normalizing constant and $\boldsymbol{\theta}_t(i_0, j)$ is a set with the properties (1) – (4).

There are many sets which satisfy properties (1) – (4). Let $\Omega_t^{i_0,j}$ denote the collection of all sets which satisfy (1) – (4), i.e.

$$\Omega_t^{i_0,j} = \{\boldsymbol{\theta}_t(i_0, j) : \boldsymbol{\theta}_t(i_0, j) \text{ satisfies properties (1) - (4)}\}.$$

Let $\beta_{t,i}^k$, $i \in \{0, 1, \dots, |Z_t|\}$, $k \in \{1, \dots, K\}$ be the marginal association probability of the event $\theta_{t,i}^k$. It is given by

$$\beta_{t,i}^k = \mathbf{P}(\theta_{t,i}^k | Z_{1:t}) = \sum_{\boldsymbol{\theta}_t(i,k) \in \Omega_t^{i,k}} \mathbf{P}(\boldsymbol{\theta}_t(i, k) | Z_{1:t}) \quad (5.14)$$

where $\mathbf{P}(\boldsymbol{\theta}_t(i, k) | Z_{1:t})$ given in (5.13).

The updated state estimate $\hat{x}_{t|t}^k$ and the covariance $\mathbf{P}_{t|t}^k$ of the target k , $k \in \{1, \dots, K\}$ corresponding to (5.11) and (5.12) are as follows

$$\begin{aligned} \hat{x}_{t|t}^k &= \hat{x}_{t|t-1}^k + W_t^k \sum_{z_i \in Z_t} \beta_{t,i}^k \nu_t^k(z_i) \\ \mathbf{P}_{t|t}^k &= \beta_{t,0}^k \mathbf{P}_{t|t-1}^k + (1 - \beta_{t,0}^k) \hat{\mathbf{P}}_t^k + \tilde{\mathbf{P}}_t^k \end{aligned}$$

where $\nu_t^k(z_i) = z_i - H_t x_{t|t-1}^k$ is the innovation of target k with respect to measurement z_i , W_t^k is Kalman gain of target k and

$$\begin{aligned}\hat{\mathbf{P}}_t^k &= [I - W_t^k H_t] \mathbf{P}_{t|t-1}^k \\ \tilde{\mathbf{P}}_t^k &= W_t^k \left[\sum_{z_i \in Z_t} \beta_{t,i}^k \nu_t^k(z_i) (\nu_t^k(z_i))^{Tr} - \sum_{z_i \in Z_t} \beta_{t,i}^k \nu_t^k(z_i) \sum_{z_i \in Z_t} \beta_{t,i}^k (\nu_t^k(z_i))^{Tr} \right] (W_t^k)^{Tr} \\ W_t^k &= \mathbf{P}_{t|t-1}^k H_t^{Tr} [S_t^k]^{-1}\end{aligned}$$

and where $\mathbf{P}_{t|t-1}^k$ is the predicted covariance of target k and S_t^k is the innovation covariance of target k .

Some extensions and modifications of JPDA are proposed to reduce the computational cost [51, 117], to deal with track coalescence for closely spaced targets [14, 17, 18, 53, 54], to implement the JPDA in a multiprocessor system [201], or to track targets which share the same measurement [119]. For nonlinear systems or a system with non-Gaussian noise, a method based on the JPDA and Monte Carlo methods was used to jointly estimate the target state vectors and association probabilities [55, 74, 82]. Since the JPDA filter and its variants only handle a fixed and known number of targets and lacks track initiation, a more general technique which can handle an unknown number of targets was proposed in [160]. This technique is called Multiple Hypothesis tracking (MHT), and it searches all hypotheses, maintains these hypotheses and defer the decision in order to solve the uncertainty at the current time [16, 145]. Here a hypothesis is a joint data association which assigns measurements to tracks such that each measurement is associated with either a single track or a false alarm. Blackman [16] demonstrated the advantages of MHT over the conventional single hypothesis approaches such as GNNF and JPDA filter. The next section presents the MHT and its variants.

5.1.2.2 Multiple Hypothesis Tracking (MHT)

Multiple Hypothesis tracking (MHT) was first derived in [160] and it is described systematically in [145]. Using Bayes rule the hypotheses are updated by assigning each of the latest measurements to a false alarm, an existing hypothesis or a new track [14, 16, 17]. With this assignment, MHT obviously can initialize a new track or terminate the current track, and hence can deal with an unknown and time-varying number of targets. However it is impractical because the computational complexity grows exponentially with the number of targets and/or the number of measurements. Furthermore, this assignment also causes the number of hypotheses to increase exponentially over time when more measurements are collected. Many studies have been carried out with the aim of reducing the complexity by pruning and gating in order to only keeping the high probability hypotheses for propagation [37, 38, 125]. These hypothesis-oriented MHT method can be improved by track-oriented MHT [23, 24, 41, 84]. The track-oriented MHT method forms track hypotheses which are the collection of hypotheses generating from the same target while the hypothesis-oriented MHT method forms hypotheses from scan to scan. The differences

between the hypothesis-oriented MHT method and the track-oriented MHT method are: 1) the high probability hypotheses remains from scan to scan in hypothesis-oriented MHT method and deletes the low probability hypotheses while the track-oriented MHT method keeps high probability tracks and discards the low probability tracks 2) Current hypotheses spawn new hypotheses for hypothesis-oriented MHT method while hypotheses are formed after tracks formed in the track-oriented MHT method. This pruning and gating may however also eliminate the correct association [56]. The probabilistic MHT (PMHT), a batch algorithm, was proposed to reduce the complexity by assuming the association variables to be statistically independent [56, 163, 164]. For a linear system model, [144] showed that PMHT outperforms JPDA in terms of track loss and mean square estimation error under a two-target scenario. A thorough review on various approaches is discussed in [14, 16, 17] some references herein.

An advanced and completely different batch approach, the Markov Chain Monte Carlo data association (MCMCDA) for multi-target tracking, was recently proposed to handle the problem where a large number of targets moves close to each other with a low probability of detection and high probability of false alarm [126, 127]. This technique employs a Reversible Jump Markov Chain Monte Carlo (RJMCMC) method on the space of data associations to sample from the posterior distribution. Such approximation not only inherit the advantages of RJMCMC method such as sampling from the complicated posterior distribution with high dimensional parameters; but also circumvent the ad hoc of reduction of number of hypotheses by pruning and gating. The following summary of track-oriented MHT in [126, 127] is presented as follows

Markov Chain Monte Carlo Data Association

Let m_t be the number of measurements available at time t i.e. $m_t = |Z_t|$. Define by $\mathcal{I}_t = \{(t, 1), \dots, (t, m_t)\}$, $m_t > 0$ a set of augmented measurement indices at time t where (t, i) denotes an measurement index i at time t and $\mathcal{I}_t = \emptyset$ if $m_t = 0$. Denote by $\mathcal{I}_{1:t}^U = \bigcup_{i=1}^t \mathcal{I}_i$ be a set of augmented measurement indices up to time t . Let $\tau_k, k > 0$ be the set of augmented indices of measurements which are generated from a target k and τ_0 be the set of augmented indices of measurements which are generated from clutter.

Let ω be a partition of $\mathcal{I}_{1:t}^U$, then the following conditions must hold

- $\omega = \{\tau_0, \tau_1, \dots, \tau_K\}$ for some K
- $\bigcup_{k=0}^K \tau_k = \mathcal{I}_{1:t}^U$ and $\tau_k \cap \tau_j = \emptyset$ for $k \neq j$.
- $|\tau_k \cap \mathcal{I}_j| \leq 1$ for any $k = 1, \dots, K$ and $j = 1, \dots, t$.
- $|\tau_k| > 1$ for $k > 0$.

The first condition shows that each partition ω represents a hypothesis whose $\tau_k, k > 0$ represents the list of indices of augmented measurements generated from the target k . The first two requirements imply that all of the measurements must be assigned as a target or clutter and imply that no two tracks can share any measurements at any time. The second last condition implies that each track has at most one measurement at each time. The last condition says that each track has at least one measurement otherwise it is considered as clutter.

With this definition, any partition ω represent a hypothesis up to time t . Denote $m_{1:t}^Z = (m_1, \dots, m_t)$ as a sequence of measurement numbers up to time t and $\Omega_{1:t}^{m_{1:t}^Z}$ be a collection of all partitions ω given $m_{1:t}^Z$. Denote by $\Omega_{1:t} = \{\omega \in \Omega_{1:t}^{m_{1:t}^Z} : m_{1:t}^Z \in \mathbb{N}_*^t\}$ a collection of all partitions

of all possible $m_{1:t}^Z \in \mathbb{N}_*^t$ and \mathbb{N}_*^t is the product of \mathbb{N}_* (note that $\mathbb{N}^* = \{0, 1, \dots\}$). Let T be the duration of the scan time. By Bayes rule the posterior distribution of ω can be computed as follows

$$p(\omega|Z_{1:T}) = \frac{p(\omega, Z_{1:T})}{\int p(\omega, Z_{1:T})d\omega} = \frac{p(Z_{1:T}|\omega)p(\omega)}{p(Z_{1:T})} \quad (5.15)$$

where $p(\omega)$ is the prior distribution and $p(Z_{1:T}|\omega)$ is the likelihood of $Z_{1:T}$ given ω .

We only present the multi-scan Markov Chain Monte Carlo Data Association (MCMCDA) [127] to find a partition ω which maximizes the distribution $p(\omega|Z_{1:T})$, i.e.

$$\hat{\omega} = \arg \max_{\omega \in \Omega_{1:T}} p(\omega|Z_{1:T}). \quad (5.16)$$

The state of target k at time t can be estimated by using MMSE as follows

$$\hat{x}_t^k = \sum_{\omega \in \Omega_{1:T}: \tau_k \in \omega} \int x_t^k p(dx_t^k|\omega, Z_{1:T})p(\omega|Z_{1:T}). \quad (5.17)$$

This considers all ω that contain a target with label k .

By using the property of reversible jump MCMC (RJMCMC), a powerful computational tool for analysis of complex posterior distribution on spaces of varying dimensions to handle the unknown number of targets, the main objective of MCMCDA algorithm is to construct the proposal distribution $q(\omega'|\omega)$ on the space $\Omega_{1:T}$ for MH update. The construction of a Markov chain is proposed given the state of a MC ω . The proposal distribution consists of eight moves grouped in five groups:

- birth/death,
- split/merge,
- extension/reduction,
- update,
- switch.

Each group of the first three groups consists two moves where each move of the group is the reverse of the other. Each group of the last two groups only has one move so a move and its reverse move are the same. With this construction, the MC on space $\Omega_{1:T}$ is reversible. The MCMCDA algorithm uses the acceptance rate

$$\alpha = \frac{p(\omega^*|Z_{1:T})q(\omega|\omega^*)}{p(\omega|Z_{1:T})q(\omega^*|\omega)} = \frac{p(\omega^*|Z_{1:T})}{p(\omega|Z_{1:T})}$$

where $p(\omega|Z_{1:T})$ is given in (5.15). Then the MCMCDA algorithm is described as follows With the construction of the proposal move, the switch move and the death move may make a track labeled k in a partition ω and a track labeled k in the proposed partition ω' totally different tracks but which in the estimation (5.17) is considered as being the same track. Thus it may lead to unreliable estimates of the target states.

Algorithm 10 : Multi-scan MCMC DA Algorithm**Input:** ω_0 , $Z_{1:T}$ and number of samples L **Output:** $\hat{\omega}$ At iteration $l = 0$:

- Set $\omega = \omega_0, \hat{\omega} = \omega_0$

For $l=1, \dots, L$:

- Sample $\omega^* \sim q(\cdot|\omega)$,
- Compute the acceptance rate

$$\alpha = \frac{p(\omega^*|Z_{1:T})}{p(\omega|Z_{1:T})} \quad (5.18)$$

Set $\omega = \omega^*$ if $\alpha > u$ where u is sampled from the uniform distribution on $[0, 1]$,Set $\hat{\omega} = \omega$ if $\frac{p(\omega|Z_{1:T})}{p(\hat{\omega}|Z_{1:T})} > 1$

Remark 1: *In general, the conventional techniques are able to deal with multi-target tracking well in the moderate scenarios. In the severe scenarios where the density of targets is high and the number of false measurements is large, these techniques do not give the reliable solution. Furthermore, these techniques do not estimate the number of existing targets, the probability of detection and the clutter rate at each time step. The new tracking algorithms based on RFS framework are able to estimate these parameters along with the target state [100, 102]. The based RFS techniques are briefly summarized in the following section.*

5.2 RFS-based Target Tracking Techniques

The Random finite set (RFS) approach to data fusion was first pioneered by Mahler [91, 106], and later it was developed as the theory of finite set statistic (FISST) [60]. An RFS is a finite-set valued random variable in the sense that it is random in number of elements and in the values of these elements. Moreover, the order of the elements is irrelevant. As a result, the RFS framework is a mathematically rigorous tool for capturing uncertainties in its elements and its cardinality [62]. In MTT problem, the number of target states and measurements at each time index are random and the order of them is not important. The RFS captures these properties and is a natural way to represent target states and measurements. Modeling the targets and measurements as RFS allows a Bayesian problem formulation by treating the target set and the measurement set as a single meta-target and a single meta-measurement respectively, and hence the multi-target Bayes filter is the analogue of the single-target Bayesian filtering with the provision of the mathematical tools in FISST [92, 93, 95, 171].

5.2.1 Single-target Tracking

Considering the single target Bayesian filtering as a special case of multi-target Bayesian filtering, [172, 182] addresses the more difficult problem of single target filtering with multiple measure-

ments generated by the target, non-uniform sensor field of view, and clutter as discussed in Section 5.1. A Bayes' recursion technique for this problem is proposed using the random finite set (RFS) framework. In this technique, the single target model is in (2.3) and (2.4). A summary of this technique is given in the next section (for details, see [172, 182])

5.2.1.1 RFS measurement model

Given target state x , an RFS of measurements at time t is modeled as the union of $D_t(x)$ the RFS of target-generated measurement at time t , $E_t(x)$ the RFS of extraneous target-generated measurements at time t and Ξ_t the RFS of clutter at time t , i.e.

$$Z_t = D_t(x) \cup E_t(x) \cup \Xi_t. \quad (5.19)$$

The RFS $E_t(x)$ represents that a target may generate more than one measurement. Conditional on x , $D_t(x)$, $E_t(x)$ and Ξ_t are assumed to be independent. The RFS $D_t(x)$ is modeled as a Bernoulli RFS with probability density given by

$$\pi_{D_t}(Z|x) = \begin{cases} 1 - p_{D_t}(x), & \text{if } Z = \emptyset; \\ K_z p_{D_t}(x) \bar{g}_t(z|x), & \text{if } Z = \{z\} \\ 0, & \text{otherwise.} \end{cases} \quad (5.20)$$

where K_z is the unit of volume on space \mathcal{Z} , $p_{D_t}(x)$ is the probability of detection and $\bar{g}_t(z|x)$ is the likelihood function.

The RFSs of extraneous measurements and clutter $E_t(x)$ and Ξ_t in (5.19) are modeled as Poisson RFS with intensities $\varsigma_{E_t}(\cdot|x)$ and ς_{Ξ_t} respectively. For simple notation, these two independent RFSs are grouped together as an RFS Λ_t

$$\Lambda_t(x) = E_t(x) \cup \Xi_t. \quad (5.21)$$

Since two independent RFSs $E_t(x)$ and Ξ_t are Poisson RFSs, their union Λ_t is also Poisson RFS with intensity

$$\kappa_t(z|x) = \varsigma_{E_t}(z|x) + \varsigma_{\Xi_t}(z). \quad (5.22)$$

5.2.1.2 Measurement likelihood

If an RFS Z_t is given in (5.19), the likelihood function $\hat{g}_t(Z_t|x)$ is [172, Proposition 3.1, p.45]

$$\hat{g}_t(Z_t|x) = K_z^{|Z_t|} \frac{1 - p_{D_t}(x)}{e^{\langle \kappa_t(\cdot|x), 1 \rangle}} \prod_{z \in Z_t} \kappa_t(z|x) + K_z^{|Z_t|} \frac{p_{D_t}(x)}{e^{\langle \kappa_t(\cdot|x), 1 \rangle}} \sum_{z^* \in Z_t} \bar{g}_t(z^*|x) \prod_{z \in Z_t - \{z^*\}} \kappa_t(z|x) \quad (5.23)$$

The first term corresponds to a missed detection while the second term refers to a target detection.

The probability of $\kappa_t(x)$ having exactly n_t elements is

$$\rho_t(n_t|x) = \frac{\langle \kappa_t(\cdot|x), 1 \rangle^{n_t} e^{-\langle \kappa_t(\cdot|x), 1 \rangle}}{n_t!}. \quad (5.24)$$

Each measurement $z \in Z_t$ is independent and identically distributed according to the probability density

$$c_t(z|x) = \frac{\kappa(z|x)}{\langle \kappa_t(\cdot|x), 1 \rangle}. \quad (5.25)$$

(5.23) can be alternatively rewritten in the following form

$$\begin{aligned} \hat{g}_t(Z_t|x) = & K_z^{|Z_t|} |Z_t|! (1 - p_{D_t}(x)) \rho_t(|Z_t||x) \prod_{z \in Z_t} c_t(z|x) + \\ & K_z^{|Z_t|} (|Z_t| - 1)! p_{D_t}(x) \rho_t(|Z_t| - 1|x) \sum_{z^* \in Z_t} \bar{g}_t(z^*|x) \prod_{z \in Z_t - \{z^*\}} c_t(z|x). \end{aligned} \quad (5.26)$$

The likelihood function $\hat{g}_t(Z_t|x)$ in (5.26) reduces to the conventional single measurement likelihood $\bar{g}_t(z|x)$ when $|Z_t| = 1$ and $p_{D_t}(x) = 1$.

5.2.1.3 RFS Single-target Bayes recursion

The Bayes recursion in (2.15) and (2.16) can be generalized to accommodate the multiple measurements generated by a target, non-uniform field of view, and clutter with the likelihood function given in (5.23) or (5.26)

$$\begin{aligned} p_{t+1|t}(x_{t+1}|Z_{1:t}) &= \int \bar{f}_{t+1|t}(x_{t+1}|x_t) p_t(x_t|Z_{1:t}) dx_t && \text{(Prediction step)} \\ p_{t+1}(x_{t+1}|Z_{1:t+1}) &= \frac{\hat{g}_{t+1}(Z_{t+1}|x_{t+1}) p_{t+1|t}(x_{t+1}|Z_{1:t})}{\int \hat{g}_{t+1}(Z_{t+1}|x_{t+1}) p_{t+1|t}(x_{t+1}|Z_{1:t})} && \text{(Update step)}. \end{aligned}$$

5.2.1.4 Closed form Solution for Linear Gaussian model

The target is assumed to follow a linear Gaussian transition and the measurements are a linear combination of the states in Gaussian noise

$$\begin{aligned} \bar{f}_{t+1|t}(x|\zeta) &= \mathcal{N}(x; F_{t-1}\zeta, Q_t) \\ \bar{g}_{t+1|t}(z|x) &= \mathcal{N}(z; H_t x, R_t). \end{aligned}$$

The probability of detection is assumed to be constant, i.e. $p_{D_t}(x) = p_{D_t}$ and the intensity of extraneous target-generated measurements is linear Gaussian i.e.

$$\varsigma_{E_t}(z|x) = \eta_t^1 c_t^1(z|x) \quad (5.27)$$

$$c_t^1(z|x) = \mathcal{N}(z; O_t x + o_t, \mathbf{Q}_t) \quad (5.28)$$

where η_t^1 is the expected number of extraneous object-generated measurements, and $c_t^1(\cdot|\cdot)$ is the likelihood of individual extraneous observation at time t , O_t is the extraneous measurement matrix, o_t is a constant vector, and \mathbf{Q}_t is the extraneous covariance at time t . The clutter has intensity

$$\varsigma_{\Xi_t}(z) = \eta_t^0 c_t^0(z)$$

where η_t^0 is the mean clutter, and $c_t^0(\cdot)$ is the density of clutter at time t . Then $\Lambda_t(x)$ in (5.21) is a Poisson RFS with intensity $\kappa_t(z|x)$ given in (5.22) and the cardinality distribution of $\Lambda_t(x)$ is Poisson with rate $\eta_t^0 + \eta_t^1$ and individual elements of $\Lambda_t(x)$ are i.i.d. according to the probability density

$$c_t(z|x) = w_{c_t,t}^0 c_t^0(z) + w_{c_t,t}^1 \mathcal{N}(z; O_t x + o_t, \mathbf{Q}_t) \quad (5.29)$$

where $w_{c_t}^i = \eta_t^i / (\eta_t^0 + \eta_t^1)$ for $i = 0, 1$.

Assume that the posterior distribution p_{t-1} is a Gaussian mixture of the form

$$p_{t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_{t-1}} w_{t-1}^i \mathcal{N}(x; m_{t-1}^i, P_{t-1}^i). \quad (5.30)$$

The prediction and update steps are given as follows

Prediction:

$$p_{t|t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_{t|t-1}} w_{t|t-1}^i \mathcal{N}(x; m_{t|t-1}^i, P_{t|t-1}^i). \quad (5.31)$$

where

$$m_{t|t-1}^i = F_{t-1} m_{t-1}^i, \quad P_{t|t-1}^i = F_{t-1} P_{t-1}^i F_{t-1}^{Tr} + Q_t.$$

Update: Assume that the predicted density is of the form in (5.31) then $\hat{g}_t(Z_t|x)$ has the form in [172, Eq 3.39, p.50]. Consequently the posterior distribution $p_t(x|Z_{1:t})$ at time t is also a Gaussian mixture of the form

$$p_t(x|Z_{1:t}) = \sum_{i=1}^{J_t} w_t^i \mathcal{N}(x; m_t^i, P_t^i) \quad (5.32)$$

where $w_t^i = \frac{\bar{w}_t^i}{\sum_{i=1}^{J_t} \bar{w}_t^i}$, $\sum_{i=1}^{J_t} \bar{w}_t^i$ is the normalizing constant in the RFS single-target Bayes recursion and

$$\hat{g}_t(Z_t|x) p_{t|t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_t} \bar{w}_t^i \mathcal{N}(x; m_t^i, P_t^i). \quad (5.33)$$

If the density in (5.27) is Gaussian mixture, p_0 is Gaussian mixture, then all the predicted $p_{t|t-1}$ and p_t are also Gaussian mixtures.

Later in 2011 [181] proposed a forward-backward smoother to improve the performance of this method.

5.2.2 Multi-target Tracking

Unlike conventional multi-target techniques, the formulation of the multi-target tracking problem in RFS framework allows the multi-target posterior distribution to be propagated using the Bayes recursion as done in the single-target Bayes filter. As discussed in Chapter 3, the definitions of the set derivative, set integral and the global density make the computation of the multi-target posterior distribution possible. Unlike the conventional multi-target tracking approach, the number of targets can be estimated along with their target states because the multi-target posterior distribution of set-valued parameters also capture the uncertainty in target numbers. Due to the computational intractability of the full multitarget Bayesian filter in practice, some researchers use sequential Monte Carlo (SMC) approximation of the full multitarget Bayes filter to track a small number of targets in simple application such as tracking three vehicles in terrain without clutter [159], tracking pedestrians with laser range scanners [146]; and locating the small unknown, time-varying number of active speakers and the voice activity interval for each speaker based on the time-difference-of interval measurements [89, 178]. At the same time, a more general approximation of a full multi-target Bayesian filter is proposed by [112].

In most applications, the full multitarget Bayes filter is computationally intractable so a drastic but principled approximation was derived as a probability hypothesis density (PHD) filter in 2000 by Mahler [96, 104]. The PHD filter was derived under the assumption that the clutter is a Poisson process and the predicted multi-target distribution is approximately a Poisson distribution. The PHD filter recursively propagates the first-order multi-target moment density or intensity and provides the number of targets in the region by integration over the region. The estimates of the target states are the peaks in the PHD. The advantage of the PHD filter is that the computational complexity is at order $O(mn)$ where n is the number of targets and m is the number of measurements and it does not require data association like the traditional techniques. Furthermore, at each time step an estimate of the target number, which cannot be achieved in the traditional techniques, is computed directly from data. Section 5.2.2.1 summarizes the derivation of PHD.

5.2.2.1 Probability Hypothesis Density (PHD) Filter

The Probability Hypothesis Density (PHD) is the first-order multi-target moment density D_t which is an approximation and is developed to overcome the computational intractability of the multi-target Bayes filter. Instead of propagating the posterior distribution as in multi-target Bayes filter,

The PHD filter only propagate the posterior intensity, a first order statistical moment. This strategy propagates the first moment (mean) of the single target state and is illustrated in Figure 5.1¹

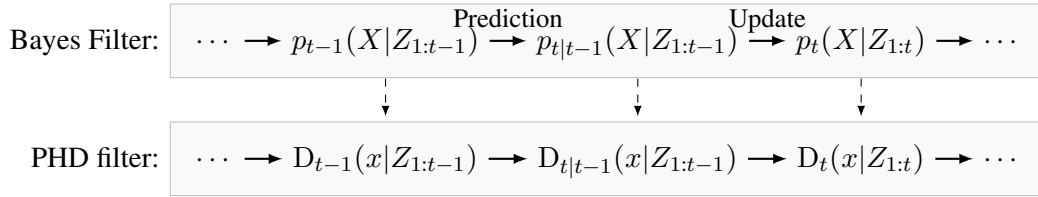


Figure 5.1: Multi-target Bayes Filter and its first-order multi-target moment D_t

For a RFS X on \mathcal{X} with $p_t(X_t|Z_{1:t})$ is given in (3.91), the definition of PHD [96, p.1154] or intensity [39, 162] is

Definition 5.1 (Probability Hypothesis Density (PHD)): *The PHD is the density $D_t(x|Z_{1:t})$ whose integral*

$$\int_S D_t(x|Z_{1:t}) dx \quad (5.34)$$

on any region S of state space is the expected number of targets contained in S

$$N_t(S) = \int |X \cap S| p_t(X_t|Z_{1:t}) \mu_s(dX). \quad (5.35)$$

where μ_s is given in (3.73).

By definition $D_t(x|Z_{1:t}) dx$ is the expected number of targets in an infinitesimally small region dx of x i.e. $D_t(x|Z_{1:t})$ is the intensity (or expected target density) at x . The PHD recursive filter is derived in the following.

PHD Filter Equations:

Given the PHD update $D_{t-1}(x|Z_{1:t-1})$ at time $t-1$, by Definition 3.14, at time t the PHD prediction $D_{t|t-1}(x|Z_{1:t-1})$ and the PHD update $D_t(x|Z_{1:t})$ are given by

$$D_{t|t-1}(x|Z_{1:t-1}) = \int K_x^{-1} p_{t|t-1}(\{x\} \cup W | Z_{1:t-1}) \mu_s(dW) \quad (5.36)$$

$$D_t(x|Z_{1:t}) = \int K_x^{-1} p_t(\{x\} \cup W | Z_{1:t}) \mu_s(dW) \quad (5.37)$$

where $p_{t|t-1}(X_t|Z_{1:t-1})$ and $p_t(X_t|Z_{1:t})$ are given by (3.89) and (3.91) respectively. The assumptions made are analogous to the assumption given in Chapter 3.2 as follows

- Each target evolves and generates measurements independently of one another.
- Clutter is Poisson distributed with intensity κ_t at time t and independent of target-originated measurements.
- Spawned targets, existing targets and new born targets are statistically independent when conditioned on the previous states.
- The predicted multi-target RFS is Poisson.

¹This figure is based on [96, p.1154]

The first three assumptions are common in multi-target tracking [8, 14]. The last assumption is made to simplify the complicated formula for updating the density $p_t(X_t|Z_{1:t})$ by approximating $p_t(X_t|Z_{1:t})$ with a Poisson distribution having intensity $D_{t|t-1}(x|Z_{1:t-1})$ such that the mean is $\bar{N}_t = \int D_{t|t-1}(x|Z_{1:t-1})$. [96] showed that $D_{t|t-1}(x|Z_{1:t})$ in (5.36) and $D_t(x|Z_{1:t+1})$ in (5.37) can be expanded as follows

$$D_{t|t-1}(x|Z_{1:t-1}) = \int p_{S_t}(\zeta) \bar{f}_{t|t-1}(x|\zeta) D_{t-1}(\zeta|Z_{1:t-1}) d\zeta + \int \beta_{t|t-1}(x|\zeta) D_{t-1}(\zeta|Z_{1:t-1}) d\zeta + \gamma_t(x) \quad (5.38)$$

$$D_t(x|Z_{1:t}) \cong (1 - p_{D_t}(x)) D_{t|t-1}(x|Z_{1:t-1}) + \sum_{z \in Z_{t+1}} \frac{p_{D_t}(x) \bar{g}_t(z|x) D_{t|t-1}(x|Z_{1:t-1})}{\kappa_t(z) + \int p_{D_t}(\zeta) \bar{g}_t(z|\zeta) D_{t|t-1}(\zeta|Z_{1:t-1}) d\zeta} \quad (5.39)$$

Note that

- $\bar{f}_{t|t-1}(x|\zeta)$ is the probability density that the target state ζ at time $t - 1$ moves to the target state at time t with surviving probability $p_{S_t}(\zeta)$.
- $\beta_{t|t-1}(\cdot|\zeta)$ is the intensity of the RFS $\mathcal{B}_{t|t-1}(\zeta)$ spawned at time t from target ζ at time $t - 1$
- $\gamma_t(x)$ is intensity of the birth RFS Γ_t at time t
- κ_t is the intensity of the clutter RFS Λ_t

Although the PHD recursion requires less computation and is simpler than the multi-target Bayes filter, it also involves multiple integral which has no closed form expression in general and hence it is difficult to implement. Some approximation of the PHD filter are derived [158, 177, 198] by using particle filter. Sidenbladh [158] implements the PHD filter by using particle filter to track multiple vehicles in terrain and compared the results with [159]. [158] shows that SMC implementation of the PHD filter is much cheaper computationally than the SMC approximation of the full multi-target Bayes filter, and it performs as well as SMC implementation of the full multi-target Bayes filter in term of target locations. However, the target number error is high under low signal to noise ratio (SNR).

At the same time, the relationship between conventional probability theory and FISST was established and led to the development of a principled Sequential importance resampling (SIS) implementation of the PHD filter (SMC-PHD) under a moderate level of measurement noise and false alarm rates [171]. This technique is more general than techniques in [158, 198]. Later the convergence of the SMC implementation of the PHD Filter was established [32, 76, 76]. More recently, another SMC implementation of the PHD filter was proposed by using an auxiliary particle filter with the point process model [190, 191].

Under the assumption of a linear Gaussian system model, the closed form solution of the PHD filter, namely Gaussian mixture PHD (GM-PHD) filter, was established in [169, 170]. The closed form solution overcomes not only the problem with a large number of particles but also the unreliability of clustering techniques for extracting state estimates which is the main drawbacks of the particle filter implementation of the PHD filter. The GM-PHD is presented next

5.2.2.2 Gaussian Mixture PHD Filter

In order to obtain the closed form solution for the PHD filter, the following assumptions are made

1. The system model are linear given by (2.17) and (2.18), i.e.

$$\bar{f}_{t|t-1}(x|x') = \mathcal{N}(x; F_{t-1}x', Q_t) \quad (5.40)$$

$$\bar{g}_t(z|x) = \mathcal{N}(z; H_t x, R_t). \quad (5.41)$$

2. The survival and detection probabilities are both state independent, i.e for all $x \in \mathcal{X}$ and for all $t = 1, 2, \dots$

$$p_{S_t} = p_{S_t}(x) \quad (5.42)$$

$$p_{D_t} = p_{D_t}(x) \quad (5.43)$$

3. The intensities of spontaneous birth and spawned RFSs are both Gaussian mixtures of the form

$$\gamma_t(x) = \sum_{i=1}^{J_{\gamma_t}} w_{\gamma_t}^i \mathcal{N}(x; m_{\gamma_t}^i, P_{\gamma_t}^i) \quad (5.44)$$

$$\beta_{t|t-1}(x|x') = \sum_{j=1}^{J_{\beta_t}} w_{\beta,t}^j \mathcal{N}(x; F_{\beta,t-1}^j x' + d_{\beta,t-1}^j, Q_{\beta,t-1}^j) \quad (5.45)$$

where J_{γ_t} , $w_{\gamma_t}^i$, $m_{\gamma_t}^i$ and $P_{\gamma_t}^i$ for $i = 1, \dots, J_{\gamma_t}$ are given model parameters that determine the shape of the spontaneous birth intensity; and similarly J_{β_t} , $w_{\beta,t}^j$, $F_{\beta,t-1}^j$, $d_{\beta,t-1}^j$ and $Q_{\beta,t-1}^j$ for $j = 1, \dots, J_{\beta_t}$ determine the shape of the spawning intensity of a target with previous state x' .

Detailed explanations of these assumptions can be found in [169]. Under these assumptions, the GM-PHD filter expands the PHD filter as follows

Assume that the posterior PHD at time $t - 1$ is a Gaussian mixture of the form

$$D_{t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_{t-1}} w_{t-1}^i \mathcal{N}(x; m_{t-1}^i, P_{t-1}^i). \quad (5.46)$$

The prediction and update steps are given as follows

Prediction:

$$D_{t|t-1}(x|Z_{1:t-1}) = D_{S,t|t-1}(x|Z_{1:t-1}) + D_{\beta,t|t-1}(x|Z_{1:t-1}) + D_{\gamma,t|t-1}(x|Z_{1:t-1}) \quad (5.47)$$

where the spontaneous birth is given in (5.44), the surviving PHD $D_{S,t|t-1}(x|Z_{1:t-1})$ is

$$D_{S,t|t-1}(x|Z_{1:t-1}) = p_{S_t} \sum_{i=1}^{J_{t-1}} w_{t-1}^i \mathcal{N}(x; m_{S,t|t-1}^i, P_{S,t|t-1}^i) \quad (5.48)$$

and where

$$m_{S,t|t-1}^i = F_{t-1}m_{t-1}^i, \quad P_{S,t|t-1}^i = F_{t-1}P_{t-1}^iF_{t-1}^{Tr} + Q_t;$$

the spawning PHD $D_{\beta,t|t-1}(x|Z_{1:t-1})$ is

$$D_{\beta,t|t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_{t-1}} \sum_{j=1}^{J_{\beta,t}} w_{t-1}^i w_{\beta,t}^j \mathcal{N}(x; m_{\beta,t|t-1}^{i,j}, P_{\beta,t|t-1}^{i,j}) \quad (5.49)$$

and where

$$m_{\beta,t|t-1}^{i,j} = F_{\beta,t-1}^j m_{t-1}^i + d_{\beta,t-1}^j, \quad P_{\beta,t|t-1}^{i,j} = F_{\beta,t-1}^j P_{\beta,t-1}^j (F_{\beta,t-1}^j)^{Tr} + Q_{\beta,t-1}^j;$$

Update: Assume that the predicted PHD is of the form

$$D_{t|t-1}(x|Z_{1:t-1}) = \sum_{i=1}^{J_{t|t-1}} w_{t|t-1}^i \mathcal{N}(x; m_{t|t-1}^i, P_{t|t-1}^i). \quad (5.50)$$

Then the posterior PHD $D_t(x|Z_{1:t})$ at time t is

$$D_t(x|Z_{1:t}) = (1 - p_{D_t})D_{t|t-1}(x|Z_{1:t-1}) + \sum_{z \in Z_t} D_{D,t}(x; z) \quad (5.51)$$

where

$$D_{D,t}(x; z) = \sum_{i=1}^{J_{t|t-1}} w_t^i(z) \mathcal{N}(x; m_{t|t}^i(z), P_{t|t}^i) \quad (5.52)$$

and where

$$\begin{aligned} w_t^i(z) &= \frac{p_{D_t} w_{t|t-1}^i q_t^i(z)}{\kappa_t(z) + p_{D_t} \sum_{l=1}^{J_{t|t-1}} w_{t|t-1}^l q_t^l(z)}, \\ q_t^i(z) &= \mathcal{N}(z; H_t m_{t|t-1}^i, R_t + H_t P_{t|t-1}^i H_t^{Tr}), \\ m_{t|t}^i(z) &= m_{t|t-1}^i + K_t^i (z - H_t m_{t|t-1}^i), \\ P_{t|t}^i &= (I - K_t^i H_t) P_{t|t-1}^i, \\ K_t^i &= P_{t|t-1}^i H_t^{Tr} (H_t P_{t|t-1}^i H_t^{Tr} + R_t)^{-1} \end{aligned}$$

Given that the initial PHD $D_0(x)$ at time $t = 0$ is a Gaussian mixture, the posterior density $D_t(x|Z_{1:t})$ is also Gaussian mixture PHD (GM-PHD) from which the individual target states can be extracted. The expected number of target $\hat{N}_{t|t-1}$ and \hat{N}_t associated with $D_{t|t-1}(x|Z_{1:t-1})$ and

$D_t(x|Z_{1:t})$ respectively are obtained by summing the appropriate mixture weights as follows

$$\hat{N}_{t|t-1} = \hat{N}_{t-1}(p_{S_t} + \sum_{i=1}^{J_{\beta_t}} w_{\beta,t}^i) + \sum_{i=1}^{J_{\gamma_t}} w_{\gamma_t}^i \quad (5.53)$$

$$\hat{N}_t = \hat{N}_{t|t-1}(1 - p_{D_t}) + \sum_{z \in Z_t} \sum_{i=1}^{J_{t|t-1}} w_t^i(z) \quad (5.54)$$

The number of Gaussian components increases exponentially so a pruning procedure ([169, p.7]) was proposed to reduce the number of Gaussian components which are propagated to the next time step. The multi-target states are extracted from the means of the Gaussian component with weights larger than some weight threshold. The GM-PHD filter is simple and effective under linear assumptions [67]. A technique for multi-sensor multi-object tracking, a more challenging problem than a single-sensor multi-object problem, employing GM-PHD filter is proposed in [137]. Another implementation of the PHD filter for the class of conditionally linear/Gaussian models was proposed [111]. numerical approximation with exact computation.

At each time step, GM-PHD filter only provides the state estimates of individual targets that may be in the surveillance region, but does not give the target identities or labels. Thus the GM-PHD tracker [129] was proposed by partitioning the outputs of the GM-PHD filter into the tracks by performing track-to-estimate association or using the GM-PHD filter as a clutter filter to eliminate some of the clutter from the measurement set before applying the data association technique. In general, the PHD filter and its variants GM-PHD and SMC-PHD do not provide information about the target label (or identity). In order to track multiple targets, the target labels are added to the target states. The target labels make it possible to distinguish between tracks (trajectories of targets). Another possibility is to associate target labels directly with each Gaussian in GM-PHD. Yet another possibility is to propagate the target labels with the target states. Thus the trajectories of targets can be obtained [26, 31, 33, 34, 49, 71, 86, 128–131, 188, 200]. The GM-PHD filter is applied in many fields such as tracking motion cells [78] where the cells neither move close nor cross each other, tracking obstacles in forward-looking sonar data [27], tracking sonar images [26, 28, 30], tracking multiple objects in a large video surveillance dataset [192], tracking with video data [90, 135, 138], tracking vehicles in terrain [158], tracking with acoustic sensors [193], tracking multiple groups of targets [25, 186, 187] and tracking a variable number of humans [68, 137]. Other applications of MTT problem have been surveyed and analyzed in [100, 101, 103].

PHD filter or one of its implementations is explored to track multiple manoeuvring targets in clutter by using multiple model methods in [132, 134, 143, 176]. The filter in [132, 134] is a generalized version of the GM-PHD filter in [169, 170] and is extended to deal with a broader class of problems using linear fractional transformations [133]. The PHD filter and its implementation such as GM-PHD and SMC-PHD are being investigated by many researchers in order to improve the performance of multi-target tracking algorithms [148, 194]. Another type of performance improvement is to estimate unknown clutter intensity for PHD Filter in [85]. The GM-PHD filter is used to derive the PHD-SLAM filter for the feature-based simultaneous localization and mapping

(SLAM) problem [113–115] and is applied to automotive imagery sensor data for constructing a map of stationary objects which is essential for autonomous vehicles [88].

A generalization of the PHD filter called the group PHD filter was derived by Mahler [94, 105] for detecting and tracking group objects such as squads, platoons, and brigades. For tracking in high target density, tracking closely spaced targets and detecting targets of interest in a dense multi-target background, the Gaussian mixture PHD filter is applied to group the targets according to a certain attributes [25, 59]. So far not many applications of the group PHD filter have been reported in the literature .

As mentioned in [49, 158], the estimate of target numbers is inconsistent in the presence of false alarms and/or missed detection. In 2006 Mahler derived a new approximation, called the cardinality PHD (CPHD) filter which propagates not only the PHD but also the entire cardinality distribution [98–101]. The CPHD, the second order moment, is a generalization of the PHD in the sense that the false alarms can be a general identically independent distributed cluster process rather than a Poisson process. However, the spawned targets cannot be modeled in the CPHD filter. Similar to the PHD filter, the CPHD filter avoids the data association. The advantage of the CPHD compared to the PHD filter is that it reliably estimates the number of targets directly from data. The disadvantages of the CPHD filter are that the computational complexity is at order $O(m^3n)$ compared to $O(mn)$ for the PHD filter, and that it does not take into account spawning targets. Similar to the PHD filter, the CPHD filter is inherently computationally intractable in general so the Gaussian mixture CPHD (GM-CPHD) filter, which is a closed form expression for the CPHD filter under linear Gaussian multi-target models, is proposed by [173]. The GM-CPHD filter for tracking a fixed number of targets outperforms the standard JPDA filter in simulations [174]. Furthermore, the GM-CPHD filter performs accurately and shows a dramatic reduction in the variance of the estimated number of targets compared to the GM-PHD filter [173]. Similar to the GM-PHD, the GM-CPHD filter is also suitable for mildly nonlinear system model as shown by simulations in [173]. The GM-CPHD filter is applied to track ground moving targets in [166] and to track multiple speakers in [136, 140]. The GM-CPHD is more responsive to changes in target number compared to the MHT algorithm [165]. A new GM-CPHD filter for passive bearings-only tracking was derived in [199]. A labeled version of the GM-CPHD was proposed in [141]. Similarly to the PHD filter and its variants, the CPHD filter and its variant GM-CPHD filter have been explored and applied to various problems in [50, 136, 140, 166].

The multi-target multi-Bernoulli (MeMBeR) filter was derived by Mahler 2007 [101] based on the assumption that every multitarget posterior is the probability law of a multi-target multi-Bernoulli process. The MeMBeR filter has advantages such as easy implementation of the birth model provided it is not too dense, a formal Poisson false alarm model, the number of targets which is estimated directly rather than inferred and no measurement-to-track association. Furthermore it is more accurate than a PHD or CPHD filter albeit more computationally demanding [101]. Similar to the CPHD filter, the MeMBeR filter does not have spawning model. A new MeMBeR filter, namely the cardinality balanced MeMBeR (CBMeMBeR) filter, was derived in [172, 183] to reduce the cardinality bias from the MeMBeR filter which overestimates the target number. The advantage of the CBMeMBeR filter is that it has smaller computational complexity than the CPHD filter and a

similar computational complexity to the PHD while the MeMBeR filter has a higher computational complexity than the CPHD filter. The authors of [172, 183] implement the CBMeMBeR filter by using SMC and Gaussian mixture under low clutter and high probability of detection with the following results: The Gaussian mixture implementation of CBMeMBeR (GM-CMMeMBeR) filter is superior for linear system and mild non-linearities. If the non-linearity is severe, the SMC implementation of the CBMeMBeR (SMC-CBMeMBeR) filter outperforms the CPHD and the PHD filter. The CBMeMBeR is applied to address the mobile multiple target tracking problem in [189]. The CBMeMBeR is employed to track speakers in three audio-visual sequences in [72]. Since the development of the CBMeMBeR filter, many studies have been devoted to approximating it by particle filters such as the Gaussian particle MeMBeR (GP-MeMBeR) filter proposed to handle a non-linear system with Gaussian noises [195, 197], a new multi-target filtering solution proposed in [184] to accommodate non-linear target model and unknown nonhomogeneous clutter intensity and sensor field-of-view and a polynomial predictive particle MeMBeR filter derived in [196] to deal with situation where the target dynamics are not modeled accurately. An overview of the approximations of the full multi-target Bayesian filter is given in Figure 5.2. The original paper and some important papers are listed under each filter.

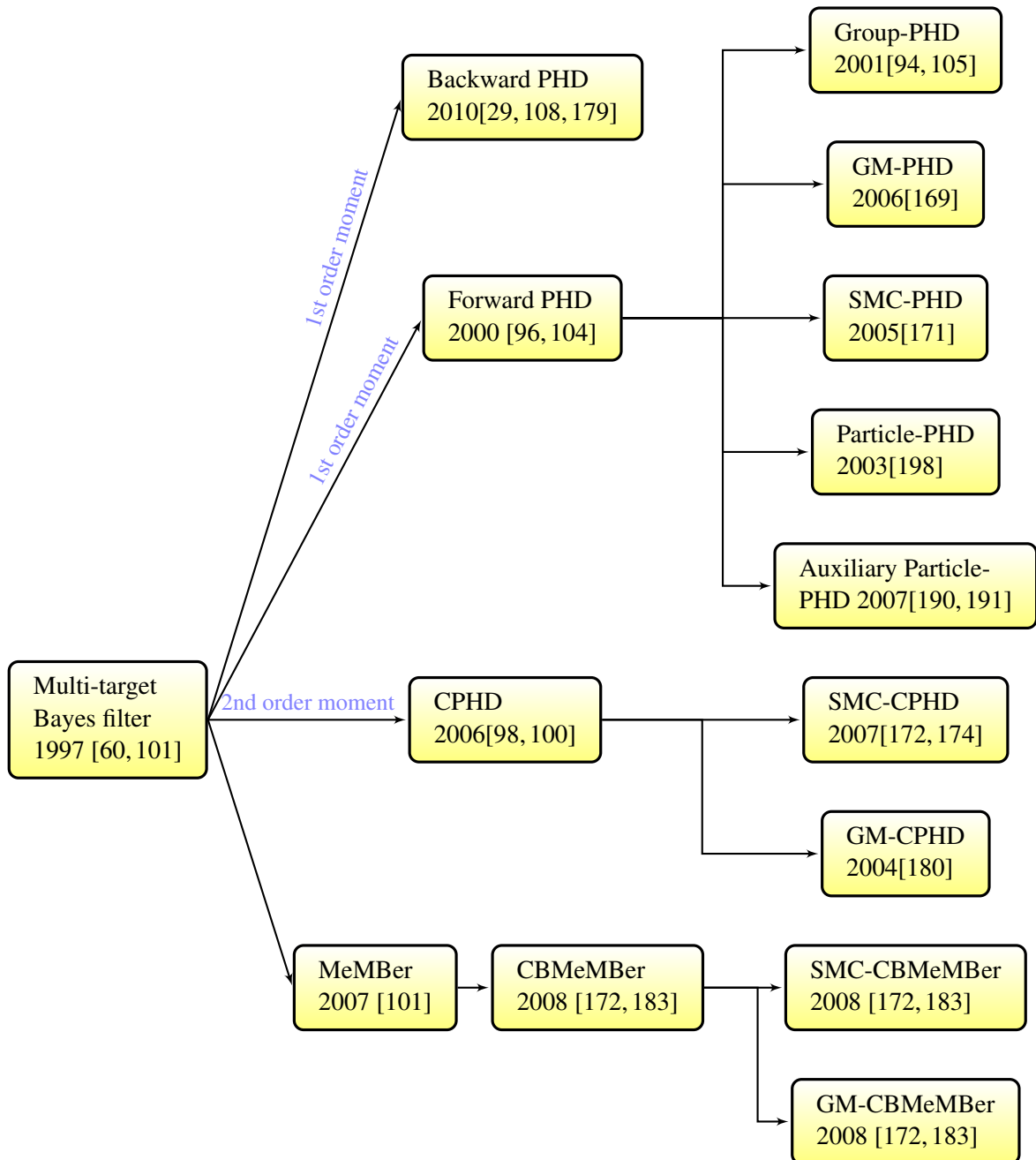


Figure 5.2: Overview of the approximations of the multi-target Bayes filter and their development together with the original works and some important papers which contributed to the development of the filters

5.3 Conclusion and Discussion

In this chapter, An overview of the development of target tracking techniques were discussed. Both conventional techniques and RFS-based techniques were covered. These two techniques can be applied to both single target tracking and multiple target tracking, and they are still under development especially the RFS-based techniques. When a large number of unknown targets move close together and cross each other or spawn other targets in a highly dense environment such as biological cells, the existing filtering techniques do not give reliable results [22, p.191-228] or [101, chapter 10 and 16]. Only if the SNR is high then the PHD filter and its variants estimate the states of the targets quite well but are unreliable when estimating the number of targets. Neither the CPHD filter nor the MeMBeR filter is suitable for this problem because none of them consider spawning targets in its model. A solution for this problem is to use the batch processing to estimate a set of tracks (the trajectories of targets) from the multi-target posterior distribution obtained from Bayesian recursive framework.

Chapter 6

PMCMC Method for RFS based Multi-target Tracking

6.1 Introduction

The cell tracking problem described in Chapter 1.1 is characterized by high target density and high clutter. For the problems with these features, techniques such as Multiple Hypothesis Tracking (MHT), Joint Probabilistic Data Association (JPDA), Joint Integrated Probabilistic Data Association - JIPDA do not give reliable solution for the reasons given in Chapter 1.1. It is however possible to use PMMH technique. In order to apply such technique, we must derive the posterior distribution for a set of the tracks (the trajectories of targets) since it is used in the MH algorithm. The main purpose of this chapter is to derive the posterior distribution for a sequence of augmented multi-target states that is equivalent to the posterior distribution for a set of tracks. The second objective of this chapter is to derive the Particle Marginal Metropolis-Hastings (PMMH) algorithm for an RFS based Multi-target tracking.

In the multi-target tracking problem, the number of targets and the number of measurements are variable and unknown. Moreover, the order of the target states and the measurements is irrelevant, e.g. the measurements (z_1, z_2) contains the same information as the measurements (z_2, z_1) . There is also the possibility that there is no measurement or target state at a time instance. Due to these features of the multi-target state and the multi-target measurement, RFSs are a natural way to represent the collection of target states and measurements at a time instance. This representation allows the multi-target tracking problem to be formulated in a Bayesian framework.

The first key contribution of this chapter is the formulation of the problem in the RFS framework in Chapter 6.2. A possible set of different tracks (trajectories of targets) with the property that no two different tracks share any state at any time is defined as a track hypothesis. There is a one-to-one correspondence between the track hypothesis and the sequence of augmented multi-target states. Thus conditional on a sequence of noisy multi-target measurements, the posterior distribution for a track hypothesis is equivalent to the posterior distribution for the corresponding sequence of augmented multi-target states.

Due to the complicated nature of the posterior distribution, the only viable option in order to compute it, is to use numerical methods such as Markov Chain Monte Carlo (MCMC). However, applying MCMC method directly is impractical because the computation of the likelihood function

in the posterior distribution involves considering all possible combinations of target states and noisy multi-target measurements. For problems such as cell tracking problem this is intractable. In order to reduce the number of possible combinations of multi-target states and multi-target measurements such that the problem becomes computationally tractable, at each instance time, an auxiliary variable will be introduced to represent the relationship between target labels and measurements indices. Furthermore, an augmented auxiliary variable is constructed to represent the relationship between the augmented multi-target states and the multi-target measurements. For the duration of the time scans, a *sequence* of augmented auxiliary variables represents the relationship between a *sequence* of augmented multi-target states and a *sequence* of multi-target measurements. Computation of the joint distribution is tractable using sampling techniques such as the PMMH algorithm which is described in Section 6.3.1.

The second contribution of this chapter is the derivation in Section 6.3 of a new algorithm, namely the PMMH algorithm for RFS based Multi-target tracking, for sampling from the joint distribution given the sequence of ordered multi-target measurements. This new algorithm combines the PMMH algorithm in Section 6.3.1 with the proposal moves (based on [127]) which are designed to consider all possibilities of a sequence of augmented auxiliary variables.

Section 6.2 formulates the problem in RFS framework and then derives the posterior distribution using Bayes recursion. Section 6.3.1 derives the new PMMH algorithm to solve the problem formulated in Section 6.2.

6.2 Formulation of the MTT problem in an RFS framework

6.2.1 Multi-target System Model in Random Finite Set Framework

The multi-target system model in Chapter 3.2 is reproduced for convenience. At time t , a multi-target state and a multi-target measurement are respectively represented as finite sets X_t and Z_t . If n_t targets are present at time t , the multi-target state $X_t = \{x_1, x_2, \dots, x_{n_t}\} \subset \mathcal{X}$ where $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ is the single-target state space and n_x is the dimension of a single target state. Similarly, if there are m_t observations at time t , the multi-target observation $Z_t = \{z_1, \dots, z_{m_t}\} \subset \mathcal{Z}$ where $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ is the measurement space and n_z is the dimension of a single-target measurement.

6.2.1.1 Multi-target State

Let T be the number of measurement scans. Then $\mathcal{T} = \{1, \dots, T\}$ is the set of time indices. Each state $x' \in X_{t-1}$ is assumed to follow a Markov process in the following sense. The target either continues to exist at time $t \in \mathcal{T}, t > 1$ with probability $p_{S_t}(x')$ and moves to the new state x according to the probability density $\bar{f}_{t|t-1}(x|x')$ or dies with probability $1 - p_{S_t}(x')$ and takes on the value \emptyset . Thus, given a single state $x' \in X_{t-1}$ at time $t - 1$, its behavior at time t is modeled by the Bernoulli RFS

$$S_{t|t-1}(x')$$

that is either $\{x\}$ when the target survives or \emptyset when the target dies. The survival or death of all existing target from time $t - 1$ to time t is hence modeled by

$$\mathcal{S}_{t|t-1}(X_{t-1}) = \bigcup_{x' \in X_{t-1}} \mathcal{S}_{t|t-1}(x').$$

In order to express the probability density $\pi_{\mathcal{S},t|t-1}(\cdot|X_{t-1})$ of the RFS $\mathcal{S}_{t|t-1}(X_{t-1})$ we introduce the following notation. Let $\mathbb{T}(U, V)$ denote the set of all one-to-one functions taking a finite set U to a finite set V . The set of all 1-1 function $\mathbb{T}(U, V) = \emptyset$ if $|U| > |V|$ and we use the convention that the sum over the empty set is zero ($|A|$ denotes the cardinality of the set A). A 1-1 function $\alpha \in \mathbb{T}(X_t, X_{t-1})$ is used to associate the targets at time t with the targets at time $t - 1$. Specifically, $x' = \alpha(x)$ means that the target state x' at time $t - 1$ has evolved to the state x at time t (i.e. $\alpha(x)$ represents the previous state at time $t - 1$ of the target state x). A target state x' at time $t - 1$ not associated with any target state at time t is dead. With this notation, $\pi_{\mathcal{S},t|t-1}(\cdot|X_{t-1})$ can be expressed as

$$\begin{aligned} \pi_{\mathcal{S},t|t-1}(X_t|X_{t-1}) &= K_x^{|X_t|} \sum_{\alpha \in \mathbb{T}(X_t, X_{t-1})} \prod_{x' \in X_{t-1} - \alpha(X_t)} (1 - p_{S_t}(x')) \\ &\quad \times \prod_{x \in X_t} p_{S,t}(\alpha(x)) \bar{f}_{t|t-1}(x|\alpha(x)) \end{aligned} \quad (6.1)$$

where $X_{t-1} - \alpha(X_t)$ means set difference, K_x is the unit volume on space \mathcal{X} and the sum is $\prod_{x' \in X_{t-1}} (1 - p_{S_t}(x'))$ if $X_t = \emptyset$.

A new target at time t may result from either the spontaneous birth (independent of the surviving targets) which is modeled by an RFS of spontaneous births Γ_t or spawning from a target state x' at time $t - 1$ which is modeled by an RFS of spawning $B_{t|t-1}(x')$. Thus the multi-target state at time t is the union of the surviving targets, the spawned targets and the spontaneous births

$$X_t = \mathcal{S}_{t|t-1}(X_{t-1}) \cup \mathcal{B}_{t|t-1}(X_{t-1}) \cup \Gamma_t \quad (6.2)$$

where $\mathcal{B}_{t|t-1}(X_{t-1}) = \bigcup_{x' \in X_{t-1}} B_{t|t-1}(x')$. The actual forms of $B_{t|t-1}$ and Γ_t are problem dependent. Assume that Γ_t is a Poisson RFS with intensity function γ_t and that $\mathcal{B}_{t|t-1}$ is a Poisson RFS with intensity function $\beta_{t|t-1}(\cdot|x')$ spawned by the target state x' at time $t - 1$, then we have that

$$\begin{aligned} \pi_{\Gamma,t}(X_t) &= e^{-\langle \gamma_t, 1 \rangle} K_x^{|X_t|} \prod_{x \in X_t} \gamma_t(x), \\ \pi_{\mathcal{B},t|t-1}(X_t|X_{t-1}) &= e^{-\sum_{x' \in X_{t-1}} \langle \beta_{t|t-1}(\cdot|x'), 1 \rangle} K_x^{|X_t|} \prod_{x \in X_t} \sum_{x' \in X_{t-1}} \beta_{t|t-1}(x|x') \end{aligned}$$

where $\langle u, v \rangle = \int u(x)v(x)dx$, $\langle \gamma_t, 1 \rangle$ is the expected number of spontaneously generated new targets, $\langle \beta_{t|t-1}(\cdot|x), 1 \rangle$ is the expected number of new targets spawned from the target state x . Assuming the three RFSs on the right hand side of (6.2) are mutually independent conditional on X_{t-1} , the RFS transition density of (6.2) can be described in the form of the *multi-target transition density* $f_{t|t-1}(\cdot|X_{t-1})$ which gives the probability density that the multi-target state moves from

X_{t-1} at time $t - 1$ to X_t at time t . Let $\pi_{\mathcal{B},t|t-1}(\cdot|X_{t-1})$ and $\pi_{\Gamma,t}$ be the probability densities of the RFS of spawning from X_{t-1} and spontaneous birth Γ_t respectively, the multi-target transition density (3.80) is rewritten as

$$f_{t|t-1}(X_t|X_{t-1}) = \sum_{\uplus_{i=1}^3 U_i = X_t} \pi_{\mathcal{S}_{t|t-1}}(U_1|X_{t-1}) \pi_{\mathcal{B},t|t-1}(U_2|X_{t-1}) \pi_{\Gamma,t}(U_3) \quad (6.3)$$

Note that X_t in (6.3) considers the new spontaneous birth and spawning compared to surviving targets only in (6.1). (6.2) describes the time evolution of the multi-target state and incorporates the model of target motion, spontaneous birth and spawning which are captured in the multi-target transition density (6.3).

The transition density $f_{t|t-1}(X_t|X_{t-1})$ in (6.3) can be expanded as follows.

$$f_{t|t-1}(X_t|X_{t-1}) = \sum_{W \subseteq X_t} \sum_{\alpha \in \mathcal{T}(W, X_{t-1})} e^{-\mu_f(X_{t-1})} \prod_{\bar{x} \in X_t - W} b(\bar{x}|X_{t-1}) \times \prod_{x' \in X_{t-1} - \alpha(W)} (1 - p_{S_t}(x')) \prod_{x \in W} p_{S_t}(\alpha(x)) \bar{f}_{t|t-1}(x|\alpha(x)) \quad (6.4)$$

where α is given in Section 3.2.1.1 and

$$\begin{aligned} \mu_f(X_{t-1}) &= \langle \gamma_t, 1 \rangle + \sum_{x' \in X_{t-1}} \langle \beta_{t|t-1}(\cdot|x'), 1 \rangle, \\ b(x|X_{t-1}) &= \gamma_t(x) + \sum_{x' \in X_{t-1}} \beta_{t|t-1}(x|x'). \end{aligned}$$

Here given X_{t-1} , $\mu_f(X_{t-1})$ is the expected number of new targets (spontaneous birth or spawning) and $b(\cdot|X_{t-1})$ is intensity function of a new target state. Each $W \subset X_t$ is the set of surviving targets which is evolved from the previous state at time $t - 1$ and the second sum is $e^{-\mu_f(X_{t-1})} \prod_{\bar{x} \in X_t} b(\bar{x}|X_{t-1}) \prod_{x' \in X_{t-1}} (1 - p_{S_t}(x'))$ if $W = \emptyset$.

6.2.1.2 Multi-target Measurement

At time t , each single-target state $x \in X_t$, is either detected with probability $p_{D_t}(x)$ and generates an observation z with likelihood $\bar{g}_t(z|x)$, or missed with probability $1 - p_{D_t}(x)$. Thus, at time t , each single-target state $x \in X_t$ generates an RFS $D_t(x)$ that can take either the value $\{z\}$ when the target is observed by a sensor or \emptyset when the target is not detected. The detection and generation of measurements for all targets at time t is hence given by the RFS

$$\mathfrak{D}_t(X_t) = \bigcup_{x \in X_t} D_t(x).$$

We assume that

- (A.1) No two different targets share the same measurement at any time.

Assumption (A.1) can be interpreted as follows: if more than two targets generate the same measurement, then this measurement will be arbitrarily associated with one of the targets and the other target will be considered as not detected. Similar to the RFS of the surviving targets, the probability density of the RFS $\mathfrak{D}_t(X_t)$ is given by

$$\pi_{\mathfrak{D},t}(Z_t|X_t) = K_z^{|Z_t|} \sum_{\alpha \in \mathcal{T}(Z_t, X_t)} \prod_{x \notin \alpha(Z_t)} (1 - p_{D_t}(x)) \prod_{z \in Z_t} p_{D_t}(\alpha(z)) \bar{g}_t(z|\alpha(z)) \quad (6.5)$$

where K_z is the unit volume on \mathcal{Z} . Assumption (A.1) allows us to consider 1-1 function between Z_t and X_t . If $Z_t = \emptyset$ the sum is $\prod_{x \in X_t} (1 - p_{D_t}(x))$

Apart from target-originated measurements, the sensor also receives a set of false/spurious measurements or clutter which is modeled by an RFS Λ_t . Consequently, at time t , the multi-target measurement Z_t is the union of target-generated measurements and clutter,

$$Z_t = \mathfrak{D}_t(X_t) \cup \Lambda_t. \quad (6.6)$$

By (3.80), the multi-target likelihood function $g_t(Z_t|X_t)$ is given by

$$g_t(Z_t|X_t) = \sum_{U \subseteq Z_t} \pi_{\mathfrak{D},t}(U|X_t) \pi_{\Lambda,t}(Z_t - U). \quad (6.7)$$

When Λ_t is a Poisson RFS with intensity κ_t ,

$$\pi_{\Lambda,t}(Z) = e^{-\langle \kappa_t, 1 \rangle} K_z^{|Z|} \prod_{z \in Z} \kappa_t(z),$$

and the multi-target likelihood function $g_t(Z_t|X_t)$ in (6.7) has the following form [172]

$$g_t(Z_t|X_t) = K_z^{|Z_t|} \sum_{W \subseteq Z_t} \sum_{\alpha \in \mathcal{T}(W, X_t)} e^{-\langle \kappa_t, 1 \rangle} \prod_{z' \in Z_t - W} \kappa_t(z') \prod_{x \in X_t - \alpha(W)} (1 - p_{D_t}(x)) \prod_{z \in W} p_{D_t}(\alpha(z)) \bar{g}_t(z|\alpha(z)). \quad (6.8)$$

where the second sun is $e^{-\langle \kappa_t, 1 \rangle} \prod_{z' \in Z_t} \kappa_t(z') \prod_{x \in X_t} (1 - p_{D_t}(x))$ if $W = \emptyset$. The terms in the second sum have their following meanings: the first two terms describe the clutter, the third term (the second product) expresses the missed detections and the last product describe the target-generated measurements. The multi-target measurement in (6.6) incorporates not only target generated measurements but also clutter which are captured in the multi-target likelihood function (6.8).

6.2.2 Track Hypothesis in RFS Framework

The purpose of this section is to define the track hypothesis which is a set of the trajectories of the target states. We begin by defining a track (trajectory of single target states) which is a path of a target over time. In terms of the states, a track is a collection of at least m^* single states on

consecutive times with the same label where m^* is called a track gate. Denote $\mathcal{T} = \{1, 2, \dots, T\}$ as the set of time indices; and $\mathcal{K} = \{1, 2, \dots, K\}$ as the set of target labels where T is the number of measurement scans, and K denotes the maximum number of target for the duration \mathcal{T} . Mathematically, a track is defined as follows

Definition 6.1 (Track): Given a track gate m^* , a track τ is an array of the form

$$\tau = (k, t, x_0, \dots, x_m), \quad m \geq m^* - 1 \quad (6.9)$$

where $k \in \mathcal{K}$ is the track label or identity, $t \in \mathcal{T}$ is the initial time of the track, $x_i \in \mathcal{X}$ is state of the track at time $t + i$ for $i = 0, \dots, m$. For the track τ in (6.9), we denote the instances of the track existence, the initial time of the track, the last existing time of the track, and the track label respectively by

$$\begin{aligned} \mathfrak{T}(\tau) &= \{t, t + 1, \dots, t + m\}, \\ \mathfrak{T}_0(\tau) &= t, \quad \mathfrak{T}_f(\tau) = t + m \\ \mathcal{L}(\tau) &= k. \end{aligned}$$

For $t' \in \mathfrak{T}(\tau)$, we denote the state at time t' by

$$\mathbf{x}_{t'}(\tau) = x_{t'-t}.$$

A collection of tracks in which no two tracks share the same state at any time is called a track hypothesis.

Definition 6.2 (Track hypothesis): A track hypothesis ω is a set of tracks such that no two tracks share the same label and no two tracks share the same state at any time i.e. for all $\tau, \tau' \in \omega$ such that $\tau \neq \tau'$

1. $\mathcal{L}(\tau) \neq \mathcal{L}(\tau')$ and
2. $\mathbf{x}_t(\tau) \neq \mathbf{x}_t(\tau')$ for any $t \in \mathfrak{T}(\tau) \cap \mathfrak{T}(\tau')$.

For a track hypothesis ω , we denote the multi-target state at time t by

$$\mathbf{X}_t(\omega) = \{\mathbf{x}_t(\tau) : \tau \in \omega\}.$$

Each element $\mathbf{x}_t(\tau)$ is the state of the target label $\mathcal{L}(\tau)$ at time t . In order to capture the label of the target state, each single state is augmented with the target label. Thus the augmented single-target state space is a hybrid space

$$\tilde{\mathcal{X}} = \mathcal{X} \times \mathcal{K} \quad (6.10)$$

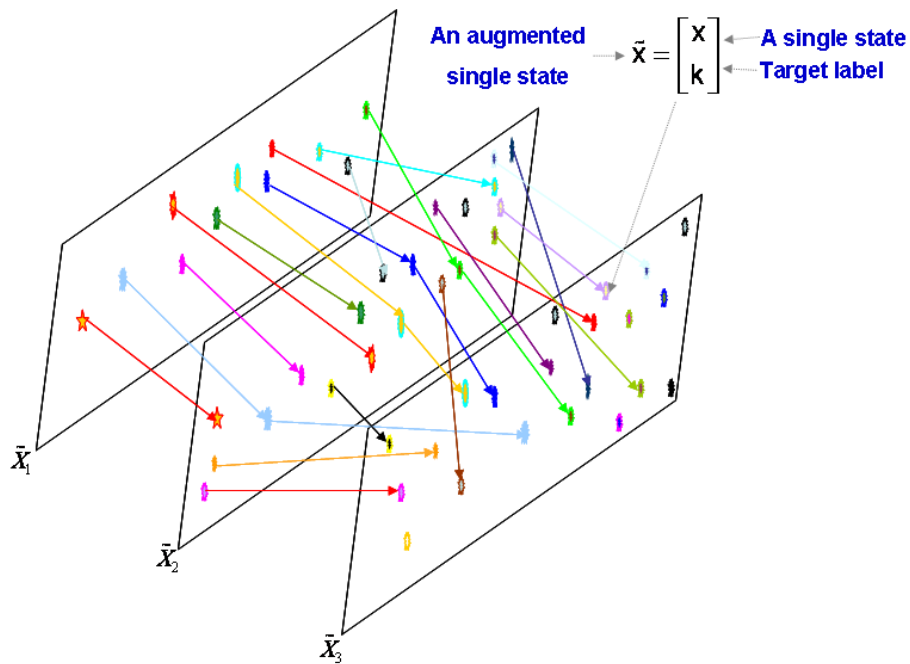


Figure 6.1: The augmented single-target states \tilde{x} live in an augmented multi-target state \tilde{X}_t at time $t = 1, 2, 3$. The augmented single-target states at different time steps which are connected by a line represents a track. The augmented single-target states at time step $t = 3$ which do not connect to other augmented single-target states at the previous time steps $t = 1, 2$ are new single augmented target states.

Hereafter, if there is no ambiguity the state space and augmented state space are used interchangeably when referring to $\tilde{\mathcal{X}}$. At time t , we denote the augmented multi-target state by \tilde{X}_t (note that $\tilde{X}_t \in \mathcal{F}(\tilde{\mathcal{X}})$) where $\mathcal{F}(A)$ denotes the collection of all finite subsets of the set A . Let τ be given in (6.9). Denote the augmented single-target state (illustrated in Figure 6.1) of track τ at time $t \in \mathfrak{T}(\tau)$ by

$$\tilde{x}_t(\tau) = (\mathbf{x}_t(\tau), k)$$

and the augmented multi-target state of track hypothesis ω at time t by

$$\tilde{X}_t(\omega) = \{\tilde{x}_t(\tau) : \tau \in \omega\}. \quad (6.11)$$

Let $\tilde{x} = (x, k)$. We denote the single target state of \tilde{x} and the label of \tilde{x} respectively by

$$\mathbf{x}(\tilde{x}) = x, \quad \mathcal{L}(\tilde{x}) = k.$$

Furthermore, the set of the labels of an augmented multi-target state \tilde{X}_t is denoted by

$$\mathcal{L}(\tilde{X}_t) = \{\mathcal{L}(\tilde{x}) : \tilde{x} \in \tilde{X}_t\}.$$

6.2.3 Posterior Distribution

Our goal is to estimate the tracks from a sequence of noisy multi-target measurements. We are therefore interested in the posterior distribution $p(\omega|Z_{1:T})$. In this section we derive the expressions for the posterior distribution $p(\omega|Z_{1:T})$ given by

$$p(\omega|Z_{1:T}) = p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$$

where $\tilde{X}_{1:T} = \tilde{\mathbf{X}}_{1:T}(\omega) = (\tilde{X}_1, \dots, \tilde{X}_T)$ and $\tilde{X}_t = \tilde{\mathbf{X}}_t(\omega)$ for $t = 1, \dots, T$. We will propagate the posterior distribution $p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$ via Bayes recursion as follows.

Assume that we have calculated the posterior distribution up to time $t - 1$. $p_{1:t}(\tilde{X}_{1:t}|Z_{1:t})$ the posterior distribution at time t can be calculated using the Bayesian recursion

$$p_{1:t}(\tilde{X}_{1:t}|Z_{1:t}) = p_{1:t-1}(\tilde{X}_{1:t-1}|Z_{1:t-1}) \frac{f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(Z_t|\tilde{X}_t)}{p(Z_t|Z_{1:t-1})}.$$

starting with $p_1(\tilde{X}_1|Z_1) = p_0(\tilde{X}_1)g_1(Z_1|\tilde{X}_1)/p(Z_1)$ where p_0 is the prior distribution of \tilde{X}_1 . Denote $f_{1|0}(\tilde{X}_1|\tilde{X}_0) = p_0(\tilde{X}_1)$, the posterior distribution $p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$ can be written as follows

$$p_{1:T}(\tilde{X}_{1:T}|Z_{1:T}) = \frac{\prod_{t=1}^T f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(Z_t|\tilde{X}_t)}{p(Z_{1:T})}. \quad (6.12)$$

The augmented multi-target transition density $f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})$ and the likelihood function $g_t(Z_t|\tilde{X}_t)$ will be discussed next.

The multi-target transition density $f_{t|t-1}(X_t|X_{t-1})$ has already been defined in (6.4). We are now considering the augmented multi-target states which also include the target labels and hence contains the information about the tracks. This simplifies the expression for the transition density. Given X_t and X_{t-1} ($t > 1$), and the multi-target transition density $f_{t|t-1}(X_t|X_{t-1})$ in (6.4), the relationship between \tilde{X}_{t-1} and \tilde{X}_t can be expressed as follows: At time t the set of surviving targets from the previous time step $t - 1$ is denoted by $W^* = \{\tilde{x} \in \tilde{X}_t : \mathcal{L}(\tilde{x}) \in \mathcal{L}(\tilde{X}_{t-1})\}$, then α in (6.4) is the 1-1 mapping α^* from $W^* \subseteq \tilde{X}_t$ to \tilde{X}_{t-1} with the property $\alpha^*(\tilde{x}) = \tilde{x}'$ if $\mathcal{L}(\tilde{x}) = \mathcal{L}(\tilde{x}')$ for $\tilde{x} \in \tilde{X}_t$. $\tilde{X}_t - W^*$ is the set of targets which are either born spontaneously or spawned from a previous state $\tilde{x}' \in \tilde{X}_{t-1}$. Intuitively, the augmented target state $\tilde{x}' \in \tilde{X}_{t-1}$ dies if its label does not belong to the set of target labels at time t ; or it survives and moves to the state $\tilde{x} \in \tilde{X}_t$ if \tilde{x} and \tilde{x}' have the same label. Furthermore, the target state $\tilde{x} \in \tilde{X}_t$ is a new target if its label does not belong to a set of target labels at time $t - 1$. Thus for $f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})$ the first two sum in (6.4) reduces to a single term corresponding to $W = W^*$ and $\alpha = \alpha^*$ and (6.4) can be written as follows

$$f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1}) = e^{-\mu_f(\tilde{X}_{t-1})} \prod_{\tilde{x} \in \tilde{X}_t - W^*} b(\tilde{x}|\tilde{X}_{t-1}) \prod_{\tilde{x}' \in \tilde{X}_{t-1} - \alpha^*(W^*)} (1 - p_{S_t}(\tilde{x}')) \times \left(\prod_{\tilde{x} \in W^*} p_{S_t}(\alpha^*(\tilde{x})) \bar{f}_{t|t-1}(\tilde{x}|\alpha^*(\tilde{x})) \right) \quad (6.13)$$

where $b(\tilde{x}|\tilde{X}_{t-1}) = b(\mathbf{x}(\tilde{x})|X_{t-1})$ is the intensity of a new target \tilde{x} (spontaneous birth or spawning), $p_{S_t}(\tilde{x}') = p_{S_t}(\mathbf{x}(\tilde{x}'))$ is the surviving probability of $\tilde{x}' \in \tilde{X}_{t-1}$ and $\mu_f(\tilde{X}_{t-1}) = \mu_f(X_{t-1})$ is the expected number of new targets. As in (6.4) the first term and the first product on the right hand side of (6.13) describes the presence of the new targets, the second product explains the dead targets and the last product explains the surviving targets.

$g_t(Z_t|\tilde{X}_t)$, $t \geq 1$ is the likelihood that a set of measurements Z_t will be collected given the set of augmented target states \tilde{X}_t at time t which is independent of the target labels so $g_t(Z_t|\tilde{X}_t) = g_t(Z_t|X_t)$. For intuitive notation, we denote $p_{D_t}(\tilde{x}) = p_{D_t}(\mathbf{x}(\tilde{x}))$, $\bar{g}_t(z|\tilde{x}) = \bar{g}_t(z|\mathbf{x}(\tilde{x}))$. (6.8) can therefore be written as

$$g_t(Z_t|\tilde{X}_t) = \sum_{W \subseteq Z_t} e^{-\langle \kappa_t, 1 \rangle} \prod_{z \in Z_t - W} \kappa_t(z) \times \left(\sum_{\alpha \in \mathcal{T}(W, \tilde{X}_t)} \prod_{\tilde{x} \in \tilde{X}_t - \alpha(W)} (1 - p_{D_t}(\tilde{x})) \prod_{z \in W} p_{D_t}(\alpha(z)) \bar{g}_t(z|\alpha(z)) \right) \quad (6.14)$$

where the second sum is $\prod_{\tilde{x} \in \tilde{X}_t} (1 - p_{D_t}(\tilde{x}))$ if $W = \emptyset$. The posterior distribution given by (6.12) has no closed-form expression so numerical methods such as MCMC must be used. However, direct application of MCMC to the above form of the posterior distribution is intractable when the set of measurements and/or the number of target states at time t is large because computation of the likelihood function $g_t(Z_t|\tilde{X}_t)$ in (6.12) which is given by (6.14) involves sum over all combinations of elements of Z_t and elements of \tilde{X}_t . To overcome this problem, at each time instance we introduce an auxiliary variable which describes a possible relationship between target labels and the measurements. The likelihood function given in (6.14) can be rewritten as an alternative form of the multi-target likelihood given in [101]

$$g_t(Z_t|\tilde{X}_t) = \sum_{\theta_t} e^{-\langle \kappa_t, 1 \rangle} \prod_{j: j \notin \theta_t(\mathcal{L}(\tilde{X}_t))} \kappa_t(z_j) \prod_{\tilde{x}' \in \tilde{X}_t: \theta_t(\mathcal{L}(\tilde{x}'))=0} (1 - p_{D_t}(\tilde{x}')) \times \prod_{\tilde{x} \in \tilde{X}_t: \theta_t(\mathcal{L}(\tilde{x}))>0} p_{D_t}(\tilde{x}) \bar{g}_t(z_{\theta_t(\mathcal{L}(\tilde{x}))}|\tilde{x}) \quad (6.15)$$

where θ_t is a mapping from $\mathcal{L}(\tilde{X}_t)$ to $\{0, 1, \dots, |Z_t|\}$ with the following property: $\theta_t(k) = \theta_t(k') > 0$ implies $k = k'$ that is, no two targets share the same measurement at any time (assumption [A.1]) and $\theta_t = \emptyset$ if $\tilde{X}_t = \emptyset$. θ_t assigns the target labels to the measurement indices if the targets are detected, and θ_t assigns 0 if the measurement is not coming from a target. θ_t in (6.15) plays an auxiliary role for calculating the likelihood and therefore θ_t is called an auxiliary variable of \tilde{X}_t . (6.15) is the sum of all possible relations between collected measurements and augmented single target states and each possibility is represented by a particular auxiliary variable θ_t . The measurements $z_j \in Z_t$ on the right hand side of (6.15) are arranged in an particular order

so we denote $\tilde{Z}_t = z_{1:|Z_t|} = (z_1, \dots, z_{|Z_t|})$ and denote

$$g_t(\tilde{Z}_t|\tilde{X}_t, \theta_t) = \prod_{j:j \notin \theta_t} \frac{\kappa_t(z_j)}{\langle \kappa_t, 1 \rangle} \prod_{\substack{\tilde{x} \in \tilde{X}_t: \\ \theta_t(\mathcal{L}(\tilde{x}))=0}} (1 - p_{D_t}(\tilde{x})) \prod_{\substack{\tilde{x} \in \tilde{X}_t: \\ \theta_t(\mathcal{L}(\tilde{x}))>0}} p_{D_t}(\tilde{x}) \bar{g}_t(z_{\theta_t(\mathcal{L}(\tilde{x}))}|\tilde{x}) \quad (6.16)$$

where $\frac{\kappa_t(z_j)}{\langle \kappa_t, 1 \rangle}$ is the density of clutter, $g(\tilde{Z}_t|\tilde{X}_t, \theta_t)$ in (6.16) is 1 if $Z_t = \emptyset$ (i.e. all targets are undetected if $\tilde{X}_t \neq \emptyset$) or $\prod_{z \in Z_t} \frac{\kappa_t(z)}{\langle \kappa_t, 1 \rangle}$ if $\tilde{X}_t = \emptyset$ (i.e. all measurements are clutter if $Z_t \neq \emptyset$). Let

$$w(\theta_t) = e^{-\langle \kappa_t, 1 \rangle} \langle \kappa_t, 1 \rangle^{|\{1, \dots, |Z_t| - \{j:j \in \theta_t(\mathcal{L}(\tilde{X}_t))\}\}|}$$

where $w(\theta_t) = e^{-\langle \kappa_t, 1 \rangle} \langle \kappa_t, 1 \rangle^{|Z_t|}$ if $\tilde{X}_t = \emptyset$. Conditional on \tilde{X}_t and θ_t , target-generated measurements and clutter in Z_t are known, then $g_t(\tilde{Z}_t|\tilde{X}_t, \theta_t)$ is the product of the density of clutter, the densities of target-generated measurements and the probabilities of undetected target states. Then (6.15) can be rewritten as

$$g_t(Z_t|\tilde{X}_t) = \sum_{\theta_t} g_t(\tilde{Z}_t|\tilde{X}_t, \theta_t) w(\theta_t). \quad (6.17)$$

We extend θ_t to an augmented auxiliary variable $\tilde{\theta}_t$ by adding the target label

$$\tilde{\theta}_t(k) = (\theta_t(k), k) \quad (6.18)$$

where $k \in \mathcal{L}(\tilde{X}_t)$ if $\theta_t \neq \emptyset$ or \emptyset if $\theta_t = \emptyset$. Hence (6.17) can be rewritten in terms of $\tilde{\theta}_t$ as follows

$$g_t(Z_t|\tilde{X}_t) = \sum_{\tilde{\theta}_t} g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t) w(\tilde{\theta}_t). \quad (6.19)$$

The posterior distribution $p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$ in (6.12) can now be rewritten using (6.19) as follows: Given $\mu_0(\tilde{X}_1)$, at time $t = 1$, denote $f_{1|0}(\tilde{X}_1|\tilde{X}_0) = \mu_0(\tilde{X}_1)$ and by (6.19) we have

$$p_1(\tilde{X}_1|Z_1) = \frac{\sum_{\tilde{\theta}_1} f_{1|0}(\tilde{X}_1|\tilde{X}_0) g_1(\tilde{Z}_1|\tilde{X}_1, \tilde{\theta}_1) w(\tilde{\theta}_1)}{p(Z_1)}. \quad (6.20)$$

Denote $\tilde{\theta}_{1:t} = (\tilde{\theta}_1, \dots, \tilde{\theta}_t)$ ($t > 1$). Assume that $p_{1:t-1}(\tilde{X}_{1:t-1}|Z_{1:t-1})$ is calculated in term of $\tilde{\theta}_{1:t-1}$ and given by

$$p_{1:t-1}(\tilde{X}_{1:t-1}|Z_{1:t-1}) = \frac{\sum_{\tilde{\theta}_{1:t-1}} \prod_{i=1}^{t-1} f_{i|i-1}(\tilde{X}_i|\tilde{X}_{i-1}) g_i(\tilde{Z}_i|\tilde{X}_i, \tilde{\theta}_i) w(\tilde{\theta}_i)}{p(Z_{1:t-1})},$$

then $p_{1:t}(\tilde{X}_{1:t}|Z_{1:t})$ is recursively propagated as follows

$$\begin{aligned}
p_{1:t}(\tilde{X}_{1:t}|Z_{1:t}) &= p_{1:t-1}(\tilde{X}_{1:t-1}|Z_{1:t-1}) \frac{f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_{t-1})}{p(Z_t|Z_{1:t-1})} \\
&= \frac{\sum_{\tilde{\theta}_{1:t-1}} \prod_{i=1}^{t-1} f_{i|i-1}(\tilde{X}_i|\tilde{X}_{i-1})g_i(\tilde{Z}_i|\tilde{X}_i, \tilde{\theta}_i)w(\tilde{\theta}_i)}{p(Z_{1:t-1})} \times \\
&\quad \frac{\sum_{\tilde{\theta}_t} f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)w(\tilde{\theta}_t)}{p(Z_t|Z_{1:t-1})} \\
&= \frac{\sum_{\tilde{\theta}_{1:t}} \prod_{i=1}^t f_{i|i-1}(\tilde{X}_i|\tilde{X}_{i-1})g_i(\tilde{Z}_i|\tilde{X}_i, \tilde{\theta}_i)w(\tilde{\theta}_i)}{p(Z_{1:t})}.
\end{aligned}$$

Denote $w(\tilde{\theta}_{1:T}) = \prod_{t=1}^T w(\tilde{\theta}_t)$. Then $p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$ can be written as

$$\begin{aligned}
p_{1:T}(\tilde{X}_{1:T}|Z_{1:T}) &= \frac{\sum_{\tilde{\theta}_{1:T}} \prod_{t=1}^T f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)w(\tilde{\theta}_t)}{p(Z_{1:T})} \\
&= \sum_{\tilde{\theta}_{1:T}} \left(\prod_{t=1}^T w(\tilde{\theta}_t) \right) \frac{\prod_{t=1}^T f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)}{p(Z_{1:T})} \\
&= \frac{\sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T}) \prod_{t=1}^T f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)}{p(Z_{1:T})} \tag{6.21}
\end{aligned}$$

Hereafter, we denote $\tilde{Z}_{1:t} = (\tilde{Z}_1, \dots, \tilde{Z}_t)$. Bayes recursion also give us

$$p_{1:T}(\tilde{X}_{1:T}|\tilde{Z}_{1:T}, \tilde{\theta}_{1:T}) = \frac{\prod_{t=1}^T f_{t|t-1}(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)}{p(\tilde{Z}_{1:T}|\tilde{\theta}_{1:T})}. \tag{6.22}$$

Then (6.21) can be rewritten as

$$p_{1:T}(\tilde{X}_{1:T}|Z_{1:T}) = \frac{\sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T})p_{1:T}(\tilde{X}_{1:T}|\tilde{Z}_{1:T}, \tilde{\theta}_{1:T})p(\tilde{Z}_{1:T}|\tilde{\theta}_{1:T})}{p(Z_{1:T})} \tag{6.23}$$

$\tilde{\theta}_{1:T}$ is not random, but for the algorithmic development in the next section it is useful to treat $\tilde{\theta}_{1:T}$ as random, and the probability of $\tilde{\theta}_{1:T}$ can be defined as follows

$$p_{1:T}(\tilde{\theta}_{1:T}) = \frac{w(\tilde{\theta}_{1:T})}{\sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T})}. \tag{6.24}$$

Then

$$\frac{w(\tilde{\theta}_{1:T})}{\sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T})} p(\tilde{Z}_{1:T}|\tilde{\theta}_{1:T}) = p(\tilde{\theta}_{1:T}|\tilde{Z}_{1:T})p(\tilde{Z}_{1:T})$$

or

$$p(\tilde{\theta}_{1:T}|\tilde{Z}_{1:T}) = \frac{p(\tilde{Z}_{1:T}|\tilde{\theta}_{1:T})w(\tilde{\theta}_{1:T})}{p(\tilde{Z}_{1:T}) \sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T})} \tag{6.25}$$

$$p(\tilde{\theta}_{1:T}|\tilde{Z}_{1:T}) \propto w(\tilde{\theta}_{1:T})p(\tilde{Z}_{1:T}|\tilde{\theta}_{1:T}) \tag{6.26}$$

so (6.23) becomes

$$\begin{aligned} p_{1:T}(\tilde{X}_{1:T}|Z_{1:T}) &= \sum_{\tilde{\theta}_{1:T}} p_{1:T}(\tilde{X}_{1:T}|\tilde{Z}_{1:T}, \tilde{\theta}_{1:T}) p(\tilde{\theta}_{1:T}|\tilde{Z}_{1:T}) \frac{p(\tilde{Z}_{1:T})}{p(Z_{1:T})} \sum_{\tilde{\theta}_{1:T}} w(\tilde{\theta}_{1:T}) \\ &\propto \sum_{\tilde{\theta}_{1:T}} p_{1:T}(\tilde{X}_{1:T}|\tilde{Z}_{1:T}, \tilde{\theta}_{1:T}) p(\tilde{\theta}_{1:T}|\tilde{Z}_{1:T}) \end{aligned} \quad (6.27)$$

$$\propto \sum_{\tilde{\theta}_{1:T}} p_{1:T}(\tilde{X}_{1:T}, \tilde{\theta}_{1:T}|\tilde{Z}_{1:T}). \quad (6.28)$$

Hereafter, we denote $\tilde{Z}_{1:t}$ as $(z_{1:|Z_1|}, \dots, z_{1:|Z_t|})$. For notational simplicity Z , \tilde{Z} , \tilde{X} , $\tilde{\theta}$ and $p(\tilde{X}_{1:T}|Z_{1:T})$ are used in place of $Z_{1:T}$, $\tilde{Z}_{1:T}$, $\tilde{X}_{1:T}$, $\tilde{\theta}_{1:T}$ and $p_{1:T}(\tilde{X}_{1:T}|Z_{1:T})$ respectively if there is no ambiguity.

The variable $\tilde{\theta}$ is in essence a nuisance variable being marginalized out. Our aim is to sample $\tilde{\theta}$, \tilde{X} from $p(\tilde{X}, \tilde{\theta}|\tilde{Z})$. The right hand side of (6.27) suggests that we can use MC methods as follows. For each MC iteration we first sample $\tilde{\theta}$ conditional on \tilde{Z} and then we sample \tilde{X} conditional on $\tilde{\theta}$ and \tilde{Z} . This approach is called Marginal Metropolis-Hastings (MMH) sampling. The Particle Marginal - Metropolis Hastings (PMMH) sampler [4] is an improvement of MMH by using SMC approximation as a proposal distribution for the Metropolis-Hastings (MH) sampler.

6.3 PMMH Algorithm for RFS-based Multi-target Tracking

The method which combines MCMC method to sample $\tilde{\theta}$ from $p(\cdot|\tilde{Z})$ and SMC method to sample \tilde{X} from $p(\tilde{X}|\tilde{Z}, \tilde{\theta})$ to get a sample $(\tilde{\theta}, \tilde{X})$ from $p(\tilde{X}, \tilde{\theta}|\tilde{Z})$ is called PMMH and was described in Chapter 4.2.3. Using MCMC to sample directly from $p(\tilde{\theta}|\tilde{Z})$ is difficult because the denominator in (6.25) is extremely difficult to compute. However, the Metropolis-Hastings (MH) algorithm is able to generate a sample without knowing this constant of proportionality. The MH algorithm generates a MC by using a proposal distribution in which each new proposed sample only depends on the current sample. The construction of the proposal distribution is described in detail with illustrated figures in Subsection 6.3.2. In Subsection 6.3.3 the PMMH Algorithm for RFS based Multi-target tracking is derived by combining the PMMH sampler described in 6.3.1 and the proposal distribution described in Subsection 6.3.2 to generate samples from the posterior distribution $p(\tilde{X}, \tilde{\theta}|\tilde{Z})$.

6.3.1 PMMH Algorithm

The PMMH given in Chapter 4.2.3 is partly reproduced for convenience. An augmented multi-target state \tilde{X} and ordered multi-target measurement \tilde{Z} are used in place of X and Z respectively.

Given $\tilde{\theta}$ and \tilde{Z} , the SMC algorithm propagates the particle $\{\tilde{X}_{1:t}^n\}$ and updates the weights $\{W_{1:t}^n\}$ for $n = 1, \dots, N$ and $t = 1, \dots, T$ as follows:

At time $t = 1$: Importance sampling (IS) is used to approximate $p(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$ by using an importance density $q(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$ as follows: N particles $\{\tilde{X}_1^n\}$ are sampled from $q(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$

and assigned importance weight $\{W_1^n\}$ which represents the discrepancy between the two densities. Then a resampling step is used to sample N times from the IS approximation $\hat{p}(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$ of $p(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$. The N particles $\{\tilde{X}_{(1)}^n\}_{n=1}^N$ which are obtained from the resampling step are approximately distributed according to $p(\tilde{X}_1|\tilde{Z}_1, \tilde{\theta}_1)$.

At time $t = 2, \dots, T$: The posterior distribution

$$p(\tilde{X}_{1:t}|\tilde{Z}_{1:t}, \tilde{\theta}_{1:t}) \propto p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1})f(\tilde{X}_t|\tilde{X}_{t-1})g(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t), \quad (6.29)$$

suggests that the samples at the previous time $t - 1$ which approximate the posterior distribution

$$p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1})$$

can be used at time step t by extending each of these particles through the IS distribution

$$q(\tilde{X}_t|\tilde{Z}_t, \tilde{X}_{t-1}, \tilde{\theta}_t)$$

to produce samples approximately distributed according to $p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1})q(\tilde{X}_t|\tilde{X}_{t-1}, \tilde{\theta}_t)$ where $q(\tilde{X}_t|\tilde{Z}_t, \tilde{X}_{t-1}, \tilde{\theta}_t)$ is an IS distribution for $f(\tilde{X}_t|\tilde{X}_{t-1})g_t(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t)$. The pseudocode for the SMC algorithm is given in Algorithm 11 below. $\mathbf{W}_t = (W_t^1, \dots, W_t^N)$ is the array of normalized importance weights at time t and defines a probability distribution on $\{1, \dots, N\}$ denoted by $\mathfrak{F}(\cdot|\mathbf{W}_t)$.

Algorithm 11 : SMC Algorithm

Input: Given $\tilde{Z}, \tilde{\theta}, p_{S_t}, p_{D_t}, \kappa_t$, the birth intensity γ_t , for $t = 1, \dots, T$ and sample number N .

Output: $\tilde{X}_{1:T}^n, W_T^n$, and $w_t(\tilde{X}_{1:t}^n)$ for $n = 1, \dots, N$ such that $\sum_{n=1}^N W_T^n \delta(\tilde{X}_{1:T}^n - \tilde{X}_{1:T})$ approximate $p(\tilde{X}|\tilde{\theta}, \tilde{Z})$

At time $t = 1$:

- sample $\tilde{X}_1^n \sim q(\cdot|\tilde{Z}_1, \tilde{\theta}_1)$ (resampling step). Then compute

$$w_1(\tilde{X}_1^n) = \frac{p(\tilde{X}_1^n, \tilde{Z}_1|\tilde{\theta}_1)}{q(\tilde{X}_1^n|\tilde{Z}_1, \tilde{\theta}_1)} = \frac{p_0(\tilde{X}_1^n)g(\tilde{Z}_1|\tilde{X}_1^n, \tilde{\theta}_1)}{q(\tilde{X}_1^n|\tilde{Z}_1, \tilde{\theta}_1)} \quad (6.30)$$

and normalize $W_1^n = w_1(\tilde{X}_1^n) / \sum_{m=1}^N w_1(\tilde{X}_1^m)$.

At $t = 2, \dots, T$:

- sample $A_{t-1}^n \sim \mathfrak{F}(\cdot|\mathbf{W}_{t-1})$ (resampling step), then $\tilde{X}_t^n \sim q(\cdot|\tilde{X}_{t-1}^{A_{t-1}^n}, \tilde{Z}_t, \tilde{\theta}_t)$ and set $\tilde{X}_{1:t}^n = (\tilde{X}_{1:t-1}^{A_{t-1}^n}, \tilde{X}_t^n)$. Then compute

$$\begin{aligned} w_t(\tilde{X}_{1:t}^n) &= \frac{p(\tilde{X}_{1:t}^n, \tilde{Z}_{1:t}|\tilde{\theta}_{1:t})}{p(\tilde{X}_{1:t-1}^{A_{t-1}^n}, \tilde{Z}_{1:t-1}|\tilde{\theta}_{1:t-1})q(\tilde{X}_t^n|\tilde{X}_{t-1}^{A_{t-1}^n}, \tilde{Z}_t, \tilde{\theta}_t)} \\ &= \frac{f(\tilde{X}_t^n|\tilde{X}_{t-1}^{A_{t-1}^n})g(\tilde{Z}_t|\tilde{\theta}_t, \tilde{X}_t^n)}{q(\tilde{X}_t^n|\tilde{X}_{t-1}^{A_{t-1}^n}, \tilde{Z}_t, \tilde{\theta}_t)} \end{aligned} \quad (6.31)$$

and normalize $W_t^n = w_t(\tilde{X}_{1:t}^n) / \sum_{m=1}^N w_t(\tilde{X}_{1:t}^m)$.

In Algorithm 11, for $n = 1, \dots, N$ the variable A_{t-1}^n is the index of the 'parent' at time $t - 1$ of particle $\tilde{X}_{1:t}^n$ for $t = 2, \dots, T$. The variables $B_{1:T}^n$ is introduced as the ancestral lineage of the particle $\tilde{X}_{1:T}^n$ such that $B_T^n = n$ and $B_t^n = A_t^{B_{t+1}^n}$ for $t = T - 1, \dots, 1$. Therefore, particle $\tilde{X}_{1:T}^n = (\tilde{X}_1^{B_1^n}, \dots, \tilde{X}_T^{B_T^n})$ for $n = 1, \dots, N$.

The SMC algorithm provides us an approximation of the posterior distribution $p(\tilde{X}|\tilde{Z}, \tilde{\theta})$ as follows (recall that $\tilde{X}^n = \tilde{X}_{1:T}^n$)

$$\hat{p}(\tilde{X}|\tilde{Z}, \tilde{\theta}) = \sum_{n=1}^N W_T^n \delta(\tilde{X} - \tilde{X}^n)$$

where $\delta(\cdot)$ is the dirac delta function. In addition, the estimate of the marginal likelihood $p(\tilde{Z}|\tilde{\theta})$ is

$$\hat{p}(\tilde{Z}|\tilde{\theta}) = \prod_{t=1}^T \hat{p}(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t})$$

where $\hat{p}(\tilde{Z}_1|\tilde{Z}_0, \tilde{\theta}_1) = \hat{p}(\tilde{Z}_1|\tilde{\theta}_1)$ and

$$\hat{p}(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}) = \frac{1}{N} \sum_{n=1}^N w_t(\tilde{X}_{1:t}^n)$$

is an estimate at time t of

$$p(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}) = \int w_t(\tilde{X}_{1:t}) q(\tilde{X}_t|\tilde{Z}_t, \tilde{X}_{t-1}, \tilde{\theta}_t) p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1}) d\tilde{X}_{1:t}. \quad (6.32)$$

(6.32) can be explained as follows

$$\begin{aligned} p(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}) &= \int p(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}, \tilde{X}_{1:t}) p(\tilde{X}_{1:t}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}) d\tilde{X}_{1:t} \\ &= \int p(\tilde{Z}_t|\tilde{X}_t, \tilde{\theta}_t) p(\tilde{X}_t|\tilde{X}_{t-1}) p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1}) d\tilde{X}_{1:t} \\ &= \int w_t(\tilde{X}_{1:t}) q(\tilde{X}_t|\tilde{Z}_t, \tilde{X}_{t-1}, \tilde{\theta}_t) p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1}) d\tilde{X}_{1:t} \text{ (by (6.31)).} \end{aligned}$$

As $\tilde{X}_{1:t}^n \sim q(\tilde{X}_t|\tilde{Z}_t, \tilde{X}_{t-1}, \tilde{\theta}_t) p(\tilde{X}_{1:t-1}|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t-1})$, then

$$p(\tilde{Z}_t|\tilde{Z}_{1:t-1}, \tilde{\theta}_{1:t}) \approx \frac{1}{N} \sum_{n=1}^N w_t^n(\tilde{X}_{1:t}^n)$$

When $\tilde{\theta}$ is unknown, estimating both \tilde{X} and the unknown $\tilde{\theta}$ from the posterior distribution $p(\tilde{\theta}, \tilde{X}|\tilde{Z})$ is required. The MH algorithm is employed with a proposal distribution of the following form

$$\mathbf{q}(\tilde{X}^*, \tilde{\theta}^*|\tilde{X}, \tilde{\theta}, \tilde{Z}) = \mathbf{q}(\tilde{\theta}^*|\tilde{\theta}, \tilde{Z}) p(\tilde{X}^*|\tilde{Z}, \tilde{\theta}^*). \quad (6.33)$$

This leads to an MMH algorithm with acceptance rate

$$\frac{p(\tilde{X}^*, \tilde{\theta}^* | \tilde{Z}) \mathbf{q}(\tilde{X}, \tilde{\theta} | \tilde{X}^*, \tilde{\theta}^*, \tilde{Z})}{p(\tilde{X}, \tilde{\theta} | \tilde{Z}) \mathbf{q}(\tilde{X}^*, \tilde{\theta}^* | \tilde{X}, \tilde{\theta}, \tilde{Z})} = \frac{p(\tilde{Z} | \tilde{\theta}^*) w(\tilde{\theta}^*) \mathbf{q}(\tilde{\theta} | \tilde{\theta}^*, \tilde{Z})}{p(\tilde{Z} | \tilde{\theta}) w(\tilde{\theta}) \mathbf{q}(\tilde{\theta}^* | \tilde{\theta}, \tilde{Z})}. \quad (6.34)$$

By using $\hat{p}(\tilde{X} | \tilde{Z}, \tilde{\theta})$ and $\hat{p}(\tilde{Z} | \tilde{\theta})$ in place of $p(\tilde{X} | \tilde{Z}, \tilde{\theta})$ and $p(\tilde{Z} | \tilde{\theta})$ respectively in the MMH update on the right hand side of (6.34), the PMMH sampler is given in Algorithm 12 for $l = 1, \dots, L$.

Algorithm 12 : PMMH Algorithm

Input: Given \tilde{Z} , p_{S_t} , p_{D_t} , κ_t , the birth intensity γ_t for $t = 1, \dots, T$ and sample number L .

Output: $S_X(l)$, $S_{\tilde{\theta}}(l)$, and $\gamma_{\theta}(l)$ for $l = 1, \dots, L$.

At iteration $l = 1$

- Set $\tilde{\theta}$ arbitrarily. Denote $S_{\tilde{\theta}}(l) = \tilde{\theta}$, then
- run an SMC algorithm targeting $p(\cdot | \tilde{Z}, \tilde{\theta})$, sample $\tilde{X} \sim \hat{p}(\cdot | \tilde{Z}, \tilde{\theta})$ and calculate $\hat{p}(\tilde{Z} | \tilde{\theta})$. Assign $S_X(l) = \tilde{X}$ and $\gamma_{\theta}(l) = \hat{p}(\tilde{Z} | \tilde{\theta})$.

At iteration $l > 1$

- Propose $\tilde{\theta}^* \sim \mathbf{q}(\cdot | S_{\tilde{\theta}}(l-1), \tilde{Z})$,
- run an SMC algorithm targeting $p(\cdot | \tilde{Z}, \tilde{\theta}^*)$, sample $\tilde{X}^* \sim \hat{p}(\cdot | \tilde{Z}, \tilde{\theta}^*)$ and calculate $\hat{p}(\tilde{Z} | \tilde{\theta}^*)$.
- calculate an acceptance rate

$$\alpha = \min \left\{ 1, \frac{\hat{p}(\tilde{Z} | \tilde{\theta}^*) w(\tilde{\theta}^*) \mathbf{q}(S_{\tilde{\theta}}(l-1) | \tilde{\theta}^*, \tilde{Z})}{\gamma_{\theta}(l-1) w(S_{\tilde{\theta}}(l-1)) \mathbf{q}(\tilde{\theta}^* | S_{\tilde{\theta}}(l-1), \tilde{Z})} \right\}$$

- if $\alpha \geq u$, set $S_X(l) = \tilde{X}^*$, $\gamma_{\theta}(l) = \hat{p}(\tilde{Z} | \tilde{\theta}^*)$ and $S_{\tilde{\theta}}(l) = \tilde{\theta}^*$. Otherwise $S_X(l) = S_X(l-1)$, $S_{\tilde{\theta}}(l) = S_{\tilde{\theta}}(l-1)$, $\gamma_{\theta}(l) = \gamma_{\theta}(l-1)$ where $u \sim \text{Unif}[0, 1]$.
-

In order to apply the PMMH, we need to construct the proposal distribution $\mathbf{q}(\cdot | \tilde{\theta}, \tilde{Z})$ in (6.34) which is discussed in the next Subsection.

6.3.2 Design and Construction of Proposal Distribution

(6.27) suggests us to sample $\tilde{\theta}$ from the conditional probability distribution $p(\cdot | \tilde{Z})$. Here each sample $\tilde{\theta}_{1:T}$ from $p(\cdot | \tilde{Z}_{1:T})$ is a sequence of auxiliary variables associated with a track hypothesis. Let Θ be the collection of all sequences of auxiliary variables $\tilde{\theta}$ where each sequence correspond to a track hypothesis. Then a sample from the distribution $p(\cdot | \tilde{Z})$ is an element of Θ . Since sampling from this distribution is difficult because the denominator in (6.25) is extremely difficult to compute, an alternative is to use the Metropolis Hastings algorithm with the proposal distribution of the form in (6.33) to generate an MC with $p(\tilde{\theta} | \tilde{Z})$ as its stationary distribution. Constructing the proposal distribution which makes the MC converges quickly to its stationary distribution $p(\tilde{\theta} | \tilde{Z})$ is the main goal of this subsection. Instead of constructing the MC on the space Θ , we construct it on an equivalent space.

6.3.2.1 Track Hypothesis Auxiliary Variable

The space containing the track information can be constructed as follows. For a given $\tilde{\theta} \in \Theta$, a track auxiliary variable $\tilde{\theta}_\tau$ is defined as follows

$$\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \quad (6.35)$$

where $k = \mathcal{L}(\tau)$, $t = \mathfrak{T}_0(\tau)$ and $\tilde{\theta}_{t+i}(k) = (j_i, k)$ for $i = 0, \dots, m$. Hence, the track auxiliary variable $\tilde{\theta}_\tau$ contains information about the measurements associated with a track τ . $\tilde{\theta}_\tau$ inherits the following properties from track τ : 1) label i.e. $\mathcal{L}(\tilde{\theta}_\tau) = \mathcal{L}(\tau)$, 2) the instances of the track existence i.e. $\mathfrak{T}(\tilde{\theta}_\tau) = \mathfrak{T}(\tau)$, 3) the initial time of appearance $\mathfrak{T}_0(\tilde{\theta}_\tau) = \mathfrak{T}_0(\tau)$ and 4) the last time of existence $\mathfrak{T}_f(\tilde{\theta}_\tau) = \mathfrak{T}_f(\tau)$. We denote the measurement index of $\tilde{\theta}_\tau$ at time $t' \in \mathfrak{T}(\tilde{\theta}_\tau)$ by

$$\mathfrak{I}_{t'}(\tilde{\theta}_\tau) = j_{t'-t}.$$

Hence the target (labeled) $\mathcal{L}(\tau)$ is undetected at time t' if $\mathfrak{I}_{t'}(\tilde{\theta}_\tau) = 0$ or it generates the measurement $z_{\mathfrak{I}_{t'}(\tilde{\theta}_\tau)}$ if $\mathfrak{I}_{t'}(\tilde{\theta}_\tau) > 0$. We also define the track hypothesis auxiliary variable

$$\tilde{\theta}_\omega = \{\tilde{\theta}_\tau : \tau \in \omega\}. \quad (6.36)$$

$\tilde{\theta}_\omega$ and $\tilde{\theta}$ are equivalent representations of the association between tracks and measurements. Given $\tilde{\theta}_\omega$, for $t = 1, \dots, T$, $\tilde{\theta}_t$ is defined by \emptyset if $t \notin \bigcup_{\tilde{\theta}_\tau \in \tilde{\theta}_\omega} \mathfrak{T}(\tilde{\theta}_\tau)$ otherwise

$$\tilde{\theta}_t(\mathcal{L}(\tilde{\theta}_\tau)) = (\mathfrak{I}_t(\tilde{\theta}_\tau), \mathcal{L}(\tilde{\theta}_\tau)), \quad \tilde{\theta}_\tau \in \tilde{\theta}_\omega \quad (6.37)$$

Thus constructing an MC on the space of $\tilde{\theta}$ is equivalent to constructing an MC on the space of $\tilde{\theta}_\omega$ denoted by Θ_ω . Denote the probability going from $\tilde{\theta}_\omega$ to $\tilde{\theta}_{\omega^*}$ given \tilde{Z} by $q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_\omega)$, then $q(\tilde{\theta}^* | \tilde{Z}, \tilde{\theta}) = q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_\omega)$.

6.3.2.2 Proposal Distribution Construction

First we make the following assumptions which are reasonable for MTT.

(A.2) The maximum speed of any target is \bar{v} .

(A.3) The maximum number of consecutive missed detection for any track is \bar{d} , ($\bar{d} \geq 1$).

\bar{d} in Assumption (A.3) can e.g. be chosen such that the probability of \bar{d} consecutive missed detections is below an acceptable threshold.

Given a track hypothesis ω , at time t we denote the clutter associated with track hypothesis ω by

$$\Lambda_t(\omega) = \left\{ z_j \in Z_t : j \notin \bigcup_{\tau \in \omega} \{\mathfrak{I}_t(\tilde{\theta}_\tau)\} \right\}. \quad (6.38)$$

The proposal distribution $q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_{\omega})$ is constructed using fourteen moves called move m , $m = 1, \dots, 14$ (see Figure 6.2) which are classified into eleven groups

List of Proposal Moves		
Group	Type	m
I	Birth (B)	1
	Death (D)	12
II	Split (S)	7
	Merge	5
III	Extension (E)	2
	Reduction (R)	8
IV	Extension Merge (EM)	4
	Birth Merge (BM)	14
V	Switch (Sw)	6
VI	Extension Merge (EM)	4
	Delete Split (DS)	13
VII	Extension Merge (EM)	4
	Delete Split (DS)	13
VIII	Extension Merge (EM)	4
IX	Birth Merge (BM)	14
X	Update (Up)	10
XI	Point Update (PUp)	11

The moves in groups I, II, III, V, and X are from [127] while the moves of the remaining groups are derived to speed up the convergence of the MC on the space of $\tilde{\theta}_{\omega}$. If a group consists of two moves, then one move is the reverse move of the other. If a group includes only one move, the move and its reverse move are the same. Now we will build the proposal distribution $q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_{\omega})$. Let $\mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m)$ be the set of all possible new track hypothesis auxiliary variables which are constructed from move m , $m = 1, \dots, 14$. Specifically,

- If $\tilde{\theta}_{\omega} = \emptyset$, only a Birth move is proposed i.e. $\mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m) = \emptyset$ for $m \neq 1$.
- If $|\tilde{\theta}_{\omega}| = 1$, neither Merge, Extension Merge nor Switch move occurs i.e. $\mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m) = \emptyset$ for $m = 4, 5, 6$.

Based on this construction $\tilde{\theta}_{\omega^*}$ is chosen uniformly at random (u.a.r) from $\bigcup_{m=1}^{14} \mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m)$. Let N_P be the number of new possible track hypothesis auxiliary variables in $\bigcup_{m=1}^{14} \mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m)$. Then the proposal distribution is

$$q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_{\omega}) = \begin{cases} \frac{1}{N_P}, & \text{if } \tilde{\theta}_{\omega^*} \in \bigcup_{m=1}^{14} \mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m); \\ 0, & \text{otherwise.} \end{cases} \quad (6.39)$$

One $\tilde{\theta}_{\omega^*}$ is chosen u.a.r from $\bigcup_{m=1}^{14} \mathbf{P}(\tilde{Z}, \tilde{\theta}_{\omega}, m)$, $\tilde{\theta}^*$ is found by (6.37).

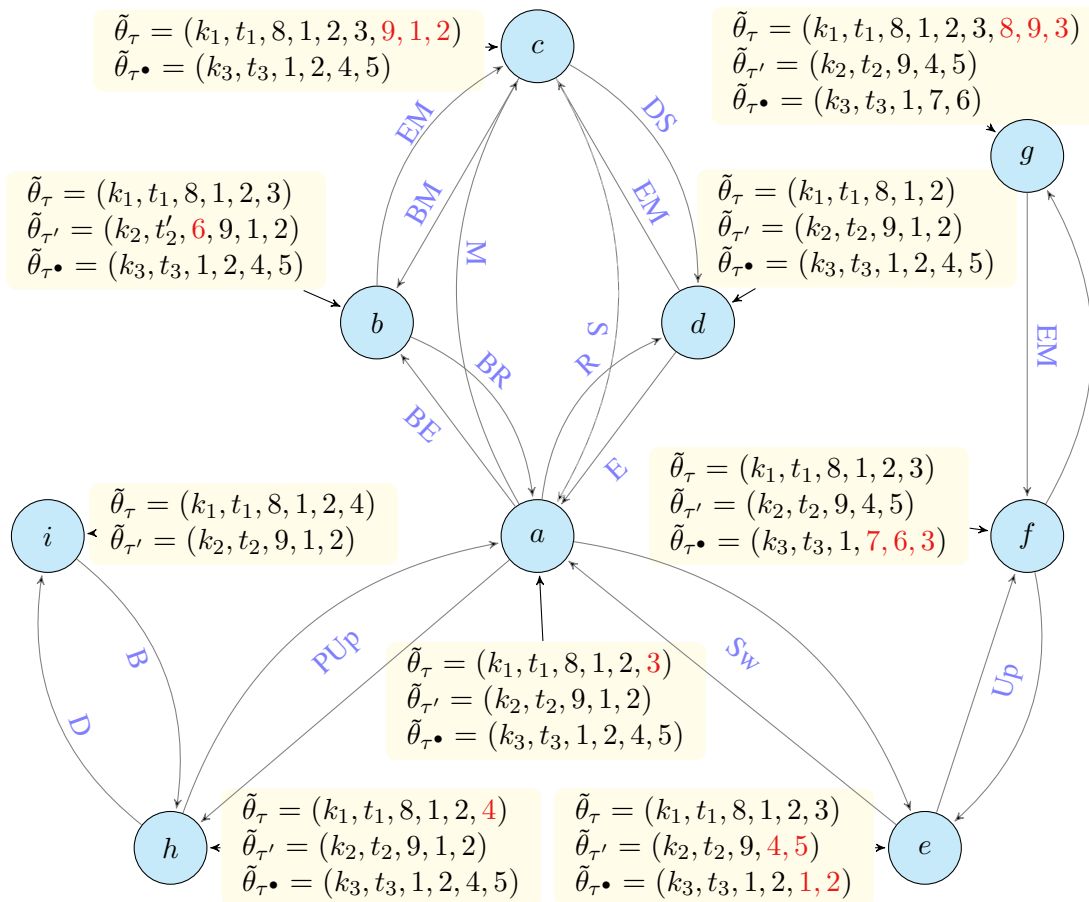


Figure 6.2: Fourteen moves of the MC on the space of $\tilde{\theta}_\omega$ with track gate $m^* = 3$ and $\bar{d} = 2$ where $t_3 = t_1 + 3$, $t_2 = t_3 + 1$ and $t'_2 = t_2 - 1$. Each move proposes a new track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*}$ that modifies the current track hypothesis auxiliary variable $\tilde{\theta}_\omega$. The Birth (B) move ($i \rightarrow h$) adds $\tilde{\theta}_{\tau^\bullet}$ which is constructed from the set of clutter $\bigcup_{t \in \mathcal{T}} \Lambda_t(\omega)$ to node (i) while the Death (D) move ($h \rightarrow i$) removes $\tilde{\theta}_{\tau^\bullet}$ at node (h) where $\Lambda_t(\omega)$ is given in (6.38). The Split (S) move ($c \rightarrow a$) splits $\tilde{\theta}_\tau$ at node (c) while the Merge (M) move ($a \rightarrow c$) combines $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ at node (a). The Extension (E) move ($d \rightarrow a$) adds measurement index 3 after the last measurement index of $\tilde{\theta}_\tau$ at node (d) while the Reduction (R) move ($a \rightarrow d$) removes the last measurement index 3 from $\tilde{\theta}_\tau$ at node (a). Similarly, the Backward Extension (BE) move ($a \rightarrow b$) adds measurement index 6 before the first measurement index of $\tilde{\theta}_{\tau'}$ at node (a) while the Backward Reduction (BR) move ($b \rightarrow a$) removes the first measurement index 6 from $\tilde{\theta}_{\tau'}$ at node (b). The Switch (Sw) move ($a \leftrightarrow e$) exchanges measurement indices between $\tilde{\theta}_{\tau'}$ and $\tilde{\theta}_{\tau^\bullet}$. The Extension Merge (EM) move ($b \rightarrow c$) merges $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ at node (b) but removes the first measurement index at $\tilde{\theta}_{\tau'}$ while the Birth Merge (BM) move ($c \rightarrow b$) adds $\tilde{\theta}_{\tau'}$ at node (b) starting at measurement index 6 then merging to $\tilde{\theta}_\tau$ at node (c) starting from measurement index 9. The Extension Merge (EM) move ($d \rightarrow c$) applies to $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ at node (d) while Delete Split (DS) move ($c \rightarrow d$) applies to $\tilde{\theta}_\tau$ at node (c). The Extension Merge (EM) move ($f \leftrightarrow g$) applies to $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau^\bullet}$. The Update (Up) move ($e \leftrightarrow f$) applies $\tilde{\theta}_{\tau^\bullet}$ while the Point Update (PUp) move ($a \leftrightarrow h$) applies to $\tilde{\theta}_\tau$.

By the construction of this proposal distribution, $\tilde{\theta}_{\omega^*}$ specifies some track hypothesis ω^* and hence $\tilde{\theta}^*$ is the sequence of augmented auxiliary variables of $\tilde{\mathbf{X}}_{1:T}(\omega^*)$. Hence whenever $\tilde{X} \sim q(\cdot | \tilde{Z}, \tilde{\theta}^*)$, there exists a track hypothesis ω^* such that $\tilde{X} = \tilde{\mathbf{X}}_{1:T}(\omega^*)$.

In order to sample from the proposal distribution $q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_{\omega})$, knowing that a measurement is clutter or potentially target-generated measurement will reduce the computation. The next subsection will explain this idea in more detail.

6.3.2.3 Neighborhoods of measurements

In multi-target tracking, the association between the states at scans are of importance to determine the trajectories of targets. However, the states are hidden Markov and are only observed indirectly through the noisy measurements. This association can be transformed equally into the association of measurements at different time scans which can be found in neighborhoods of measurements. This subsection will introduce a set which contains all measurements potentially generated from the same target. Note that the introduction of this set will reduce number of possible track auxiliary variable associated with one of the fourteen proposal moves but it does not affect the estimate of target number as well as the RFS concept.

From now on, time scan or time index are used interchangeably. Given a measurement z at time t , a measurement z' at time $t + d$, $d \in \mathbf{d}$ where $\mathbf{d} = \{1, 2, \dots, \bar{d} + 1\}$. z' is called a d -neighbor of z (*neighbor* at time scan $t + d$ of z) if $|z' - z| \leq d\bar{v}$. A set of these elements is called d -neighborhood of z (or *neighborhood* at time scan $t + d$ of z) and is denoted by $L_d(z, t)$. i.e.

$$L_d(z, t) = \{z' \in Z_{t+d} : \|z' - z\| \leq d\bar{v}\}, \quad d \in \mathbf{d}. \quad (6.40)$$

where $\|\cdot\|$ is the Euclidean norm on \mathbb{R}^{n_z} . This explains the idea that if a measurement z is generated from a target labeled i at time t then $z' \in L_d(z, t)$ is a possible measurement generated by target i at time $t + d$.

The introduction of $L_d(z, t)$ reduces the computation of the proposal distribution by choosing only neighbors of z as the potential target-generated measurements from the same target which generates the measurement z . Consider a $z \in Z_t$ if $L_d(z, t) = \emptyset$ for all $d \in \mathbf{d}$ then z may be the last measurement generated from a target if z is a d' -neighbor of any measurement $z' \in Z_{t-d'}$ where $d' \in \mathbf{d}$, $t - d' > 0$ (i.e. $z \in L_{d'}(z', t - d')$) otherwise z is a clutter. If there exist $d \in \mathbf{d}$ such that $L_d(z, t) \neq \emptyset$, the target which generated measurement z potentially survives at time $t + d$.

The union of all $L_d(z, t)$, $d \in \mathbf{d}$ is called neighborhood of z and denoted by $L(z, t)$. Mathematically

$$L(z, t) = \bigcup_{d \in \mathbf{d}} L_d(z, t).$$

An element of $L(z, t)$ is called a neighbor of z . If $L(z, t) = \emptyset$, z may be the last measurement generated from a target if z is a neighbor of any measurement in the previous \mathbf{d} time scan otherwise

z is a clutter. If $L(z, t) \neq \emptyset$, the target which generated measurement z potentially survives in the next time scan.

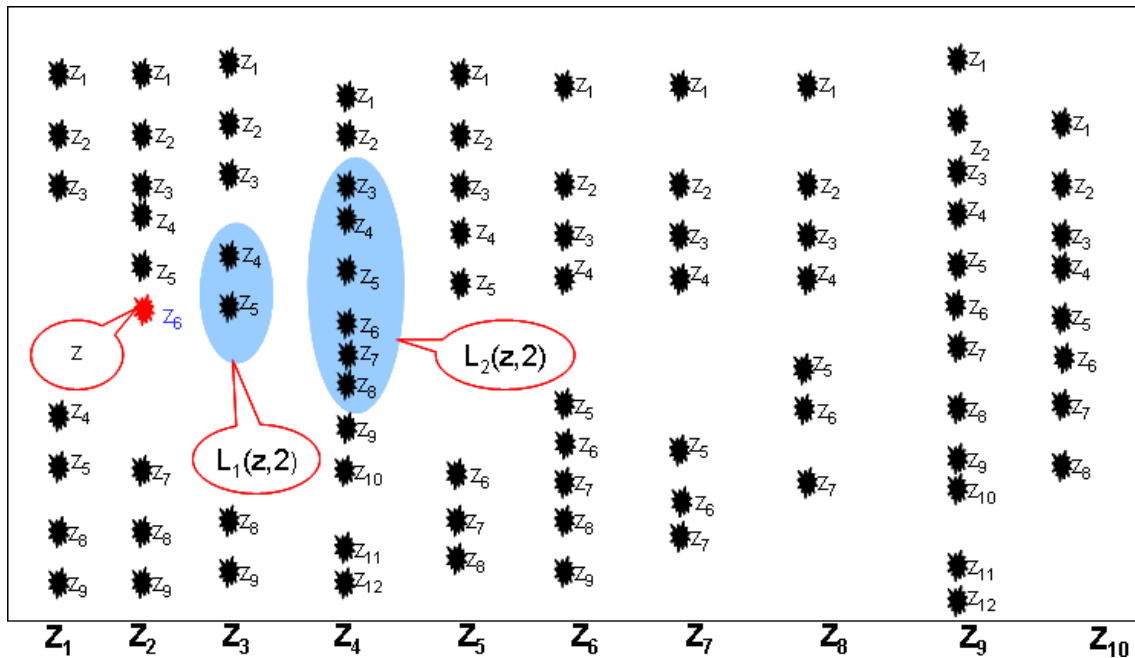


Figure 6.3: Given $z \in Z_t$, the neighborhood of z at the next consecutive time scan is $L_1(z, 2)$ and Neighborhood of z at the second consecutive time scan is $L_2(z, 2)$ where $\bar{d} = 1$.

Similarly, denote by $L_d^B(t)$ the set of measurements at time t which neighborhood at time $t + d$ is not empty i.e.

$$L_d^B(t) = \{z \in Z_t : L_d(z, t) \neq \emptyset\}; \quad (6.41)$$

$L^B(t)$ is the set of all possible target-generated measurement at time t which survives in the future i.e.

$$L^B(t) = \{z \in L_d^B(t) : d \in \mathbf{d}\}; \quad (6.42)$$

At time t if $L^B(t) = \emptyset$, all measurements are clutter or the last measurement of a track, otherwise any element of $L^B(t)$ is a potential target-generated measurement. In particular, any measurement of a non-empty $L_d^B(t)$ and a neighbor at time scan $t + d$ may be generated from a target. For example, in Figure 6.3, z_6 at time $t = 2$ and any element of $L(z_6, 2) = L_1(z_6, 2) \cup L_2(z_6, 2)$ may be generated from a target where $\bar{d} = 2$.

The next Subsection will detail how proposal distribution associated with the fourteen proposal moves is constructed using the neighborhoods.

6.3.2.4 Proposal Moves

In this Subsection, we will discuss the construction of the moves (in groups) with illustrated figures.

The following notations will be used throughout.

Denote by $\rho_j(\tilde{Z}_t)$, $t \in \mathcal{T}$ the projection mapping that takes an element $\tilde{Z}_t = (z_1, \dots, z_n)$ to the value $\rho_j(\tilde{Z}_t) = z_j$.

The measurement of a target is denoted by either the symbol \emptyset if the target is not detected by a sensor or by z if the target generate measurement z . The first is called empty measurement and the latter is called target-generated measurement. Furthermore, given $\tilde{\theta}_\tau$, τ_z is the sequence of measurements (empty measurement or target-generated measurement) of target $\mathcal{L}(\tilde{\theta}_\tau)$ and is called the track measurement of target $\mathcal{L}(\tilde{\theta}_\tau)$. For example given $\tilde{\theta}_\tau = (k, t, j_0 \dots, j_m)$, then $\tau_z = (y_0, \dots, y_m)$ where

$$y_i = \begin{cases} \emptyset, & \text{if } j_i = 0; \\ z_{j_i} \in Z_{t+i}, & \text{if } j_i > 0 \end{cases}$$

for $i = 0, \dots, m$.

As discussed in Subsection 6.3.2, the properties (label, initial time, the last time and the duration of existence) of $\tilde{\theta}_\tau$ and τ are the same apart from the properties of the track states in τ . Thus, the track and the track auxiliary variable are used interchangeably for the properties other than the states of the track. In order to construct the fourteen moves in detail, we first explain the purpose of proposing these fourteen moves for solving the problem and describe briefly the characteristic of these fourteen moves with illustrated figures.

Sketch of Proposal moves

1. Birth and death moves:

The purpose of the Birth move and the Death move is to deal with unknown number of targets [127]. These moves are illustrated in Figure 6.4. The track hypothesis ω reduces the number of tracks by one for the Death move and increases the number of tracks by one for the Birth move.

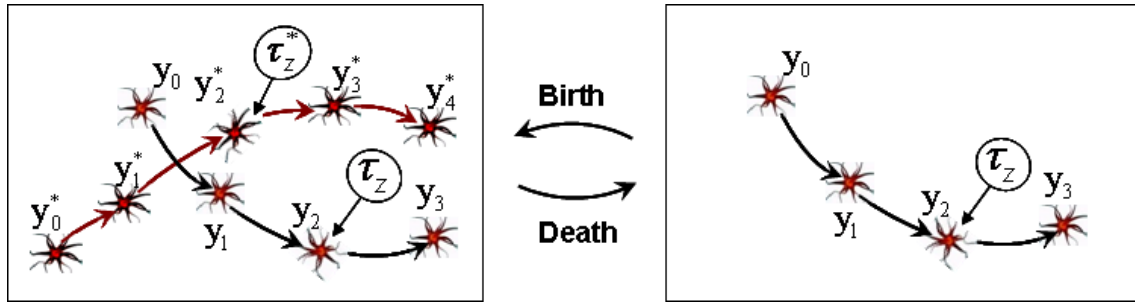


Figure 6.4: A Birth move is proposed from $\tilde{\theta}_\omega$ to $\tilde{\theta}_{\omega^*}$ by adding a track auxiliary variable $\tilde{\theta}_{\tau^*}$ with its track measurement $\tau_z^* = (y_0^*, \dots, y_4^*)$ to $\tilde{\theta}_\omega$ and its reverse move, a Death move is proposed from $\tilde{\theta}_{\omega^*}$ to $\tilde{\theta}_\omega$ by removing the track auxiliary variable $\tilde{\theta}_{\tau^*}$.

2. Split and Merge moves:

When $\tilde{\theta}_\omega \neq \emptyset$, Split and Merge moves [127] are a reversible pair of moves. This pair of moves also change the number of tracks. In a split move a track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $|\mathfrak{T}(\tilde{\theta}_{\tau^*})| \geq 2m^*$ is split into two track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ where $\mathfrak{T}_0(\tilde{\theta}_{\tau'}) = \mathfrak{T}_f(\tilde{\theta}_\tau) + 1$. If $\mathfrak{T}_0(\tilde{\theta}_{\tau'}) > \mathfrak{T}_f(\tilde{\theta}_\tau) + 1$ this move become the Delete Split move which is discussed later in point 5 of this subsection. The reverse, a Merge move is applied to any two track auxiliary variables $\tilde{\theta}_\tau, \tilde{\theta}_{\tau'} \in \tilde{\theta}_\omega$ in which the first target-generated measurement of the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is in the d -neighborhood ($d = 1$) of the last target-generated measurement of the target $\mathcal{L}(\tilde{\theta}_\tau)$. If $d > 1$ the move is called an Extension Merge move discussed later in point 5 of this section. The Split and Merge moves are sketched in Figure 6.5.

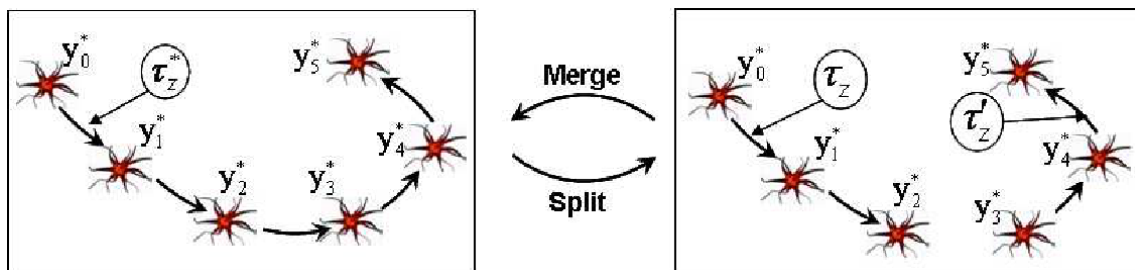


Figure 6.5: The Split move divides track auxiliary variable $\tilde{\theta}_{\tau^*} \in \omega$ with $\tau_z^* = (y_0^*, y_1^*, \dots, y_5^*)$ into two new track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ with $\tau_z = (y_0^*, y_1^*, y_2^*)$ and $\tau_z' = (y_3^*, y_4^*, y_5^*)$ respectively. Its reverse move, the Merge move, is applied to the track auxiliary variables $\tilde{\theta}_{\tau'}$ and $\tilde{\theta}_\tau$ to form a proposed track auxiliary variable $\tilde{\theta}_{\tau^*}$. For the merge move it is required that $y_3^* \in L_1(y_2^*, \mathfrak{T}_f(\tilde{\theta}_\tau))$.

3. Extension and Reduction move:

The objective of the Extension move [127] is to extend the duration of a track by one or more time scans. The Reduction move [127] reduces the duration with one or more time scans but not below m^* . These moves are sketched in Figure 6.6.

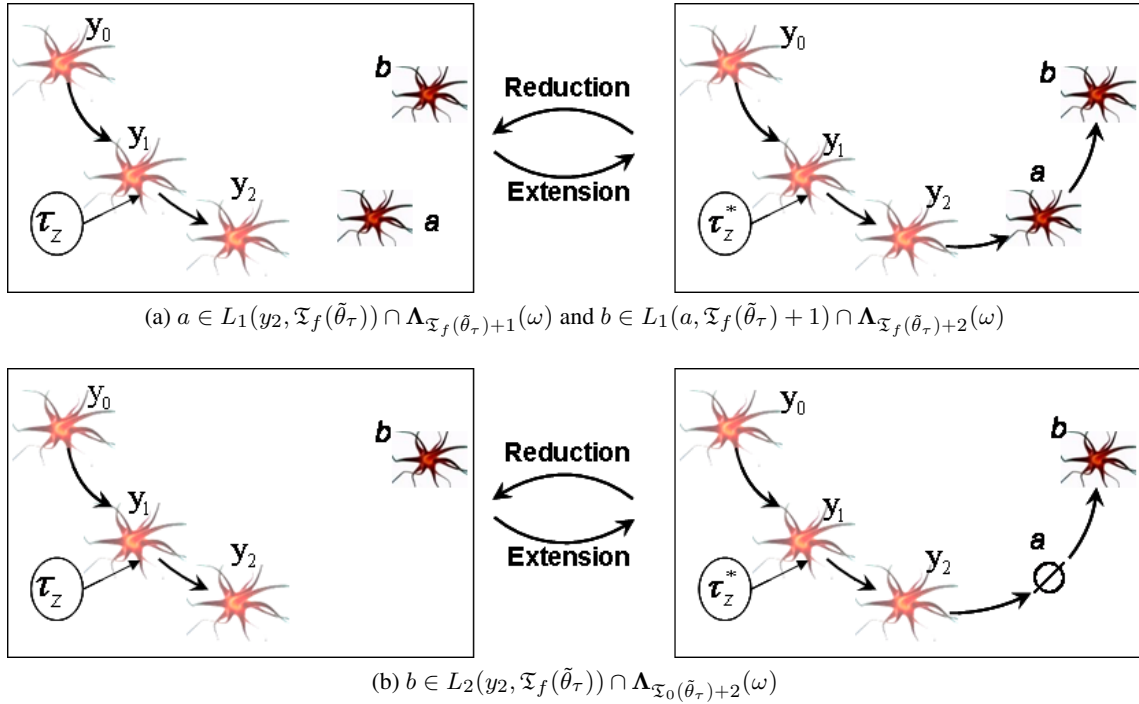


Figure 6.6: The Extension move extends the track auxiliary variable $\tilde{\theta}_\tau$ with $\tau_z = (y_0, y_1, y_2)$ by adding a, b to τ_z where a, b are shown in Figure 6.6a and Figure 6.6b to form a track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, y_2, a, b)$. In reverse, the reduction move is applied to track auxiliary variable $\tilde{\theta}_{\tau^*}$ by removing a and b from $\tilde{\theta}_{\tau^*}$ to form a track auxiliary variable $\tilde{\theta}_\tau$.

4. Switch move:

This move [127] considers the possibility that the measurements from two targets moving close to each other may be switched. This move is self-reversible. The switch move is to exchange some measurements between targets $\mathcal{L}(\tilde{\theta}_\tau)$ and $\mathcal{L}(\tilde{\theta}_{\tau'})$ while keeping the measurements from all other targets as before (see Figure 6.7).

5. Extension Merge move/Birth Merge move and Extension Merge move/Delete Split move:

The purpose of the Extension Merge move is to allow the track measurement of a current target to be extended before merging it with other track. The Extension Merge move is a combination of the Extension move and the Merge move. This move is proposed to increase the probability of proposing the Extension move and then the merge move. It may be self-reversible (see Figure 6.8). In the reverse of the Extension Merge move there is a possibility that a track measurement from a new born target may merge with track measurement from the current targets. This possibility is called a Birth Merge move and is a combination of a Birth move and a Merge move (see Figure 6.9). The Birth Merge move may not change the number of tracks and it may be self-reversible (see Figure 6.10).

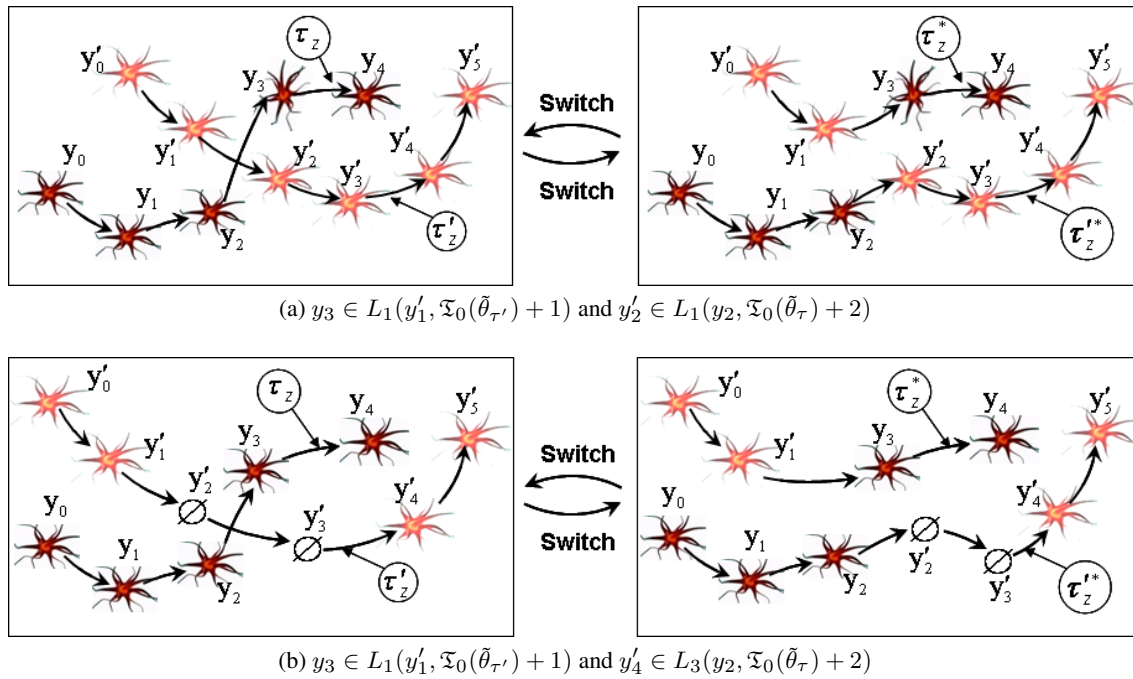


Figure 6.7: Given $\tilde{\theta}_{\tau}$ with track measurement $\tau_z = (y_0, \dots, y_4)$ and $\tilde{\theta}_{\tau'}$ with track measurement $\tau'_z = (y'_0, \dots, y'_5)$ where $\mathfrak{T}_0(\tilde{\theta}_{\tau'}) = \mathfrak{T}_0(\tilde{\theta}_{\tau}) + 1$. A Switch move exchanges the measurements (y_3, y_4) from the target $\mathcal{L}(\tau)$ with the measurement (y'_2, \dots, y'_5) from the target $\mathcal{L}(\tau')$. Thus $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$ are formed with the sequences of measurements $\tau_z^* = (y_0, y_1, y_2, y'_2, \dots, y'_5)$ and $\tau'^*_z = (y'_0, y'_1, y_3, y_4)$ respectively.

Another reverse of the Extension Merge move is the Delete Split move (see point 2 above) in which a track measurement from a current target is split into two track measurement after deleting some measurements (see Figure 6.11).

6. Backward Extension move and Backward Reduction move:

Backward Extension move and Backward Reduction move are proposed by first applying the Death move, and then the Birth move. Thus Backward Extension move and Backward Reduction move are derived in this thesis to increase the probability of proposing the combination of Birth move and Death move. A Backward Extension considers the possibility that the target $\mathcal{L}(\tilde{\theta}_{\tau})$ may appear earlier by adding new measurements to the beginning of the track measurement of $\tilde{\theta}_{\tau}$ when the target $\mathcal{L}(\tilde{\theta}_{\tau})$ does not appear at the first time scans of the sensor. Its reverse is a Backward Reduction move. This move considers that the target may appear later. Equivalently, this move considers the first few measurements in a track as false alarms which are deleted to form a new track auxiliary variable (see Figure 6.12).

7. Update move and Point Update move:

The Update move and the Point Update move are proposed to deal with dense targets and dense measurements by considering different possibilities for a target $\mathcal{L}(\tilde{\theta}_{\tau})$ whose target-generated measurements have many neighbors. Particularly, the Update move [127] modifies the measurements since time t_0 of the track auxiliary variable $\tilde{\theta}_{\tau} \in \tilde{\theta}_{\omega}$ where t_0 is

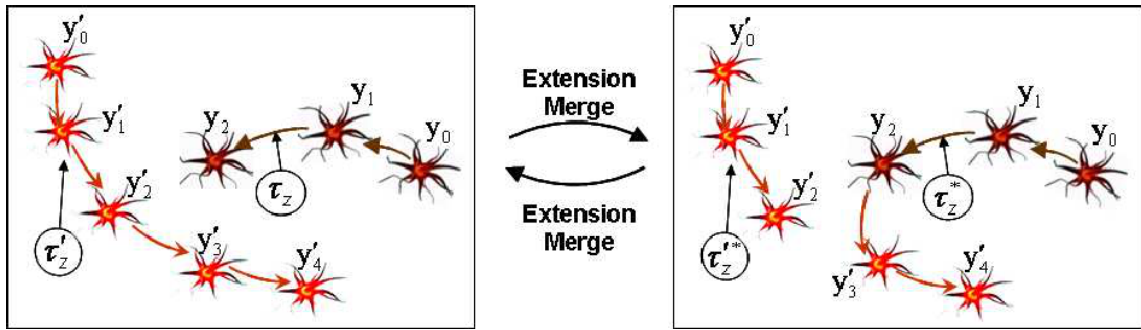


Figure 6.8: Given $\tilde{\theta}_\tau$ with $\tau_z = (y_0, y_1, y_2)$ and $\tilde{\theta}_{\tau'}$ with $\tau'_z = (y'_0, \dots, y'_4)$ where $\mathfrak{T}_0(\tilde{\theta}_\tau) = \mathfrak{T}_0(\tilde{\theta}_{\tau'})$. The Extension Merge move is proposed by merging y_2 to y'_3 of the track measurement from target $\mathcal{L}(\tau')$ where $y'_3 \in L_1(y_2, \mathfrak{T}_f(\tilde{\theta}_{\tau'}))$. Thus track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, y_2, y'_3, y'_4)$ and $\tilde{\theta}_{\tau'^*}$ with $\tau'_z^* = (y'_0, y'_1, y'_2)$ are formed. Its reverse move, an Extension Merge move, is applied to the track auxiliary variables $\tilde{\theta}_{\tau'^*}$ and $\tilde{\theta}_{\tau^*}$ to form track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$.

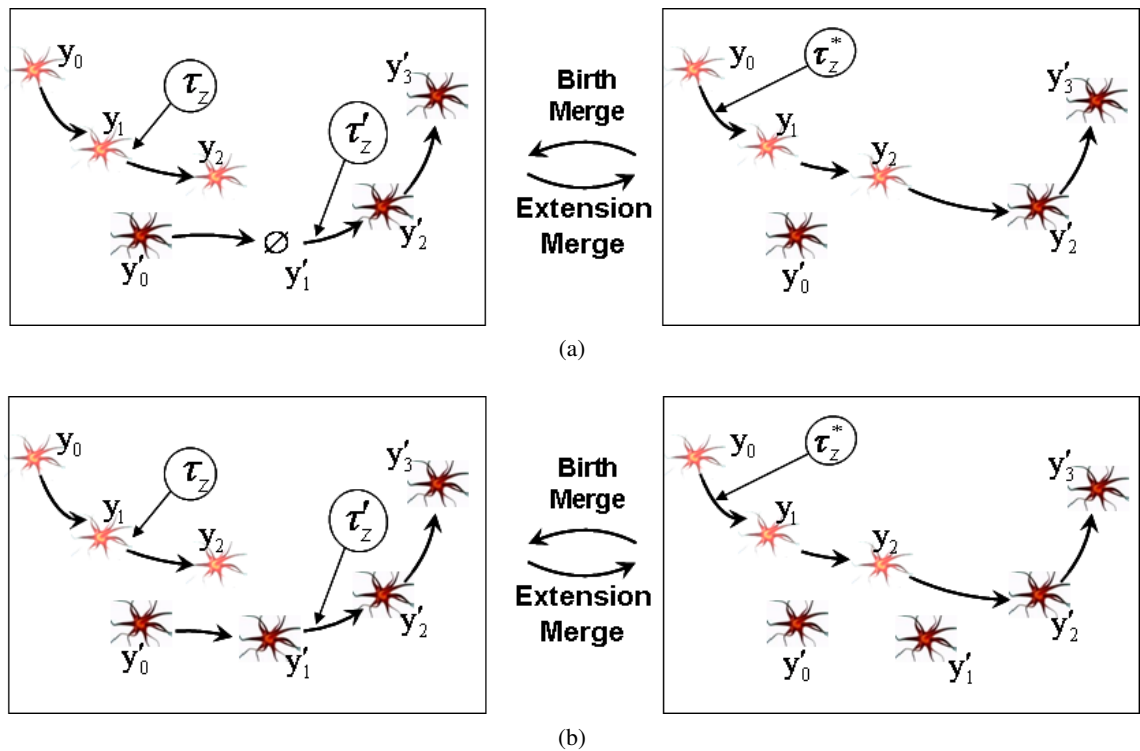


Figure 6.9: An Extension Merge move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ with $\mathfrak{T}_f(\tilde{\theta}_\tau) = \mathfrak{T}_f(\tilde{\theta}_{\tau'}) + 1$ and $y'_2 \in L_1(y_2, \mathfrak{T}_f(\tilde{\theta}_{\tau'}))$ to form track auxiliary variables $\tilde{\theta}_{\tau^*}$. Its reverse move, a Birth Merge move starts at time $\mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + 1$ with the measurement $y'_0 \in \Lambda_{\mathfrak{T}_0(\tilde{\theta}_{\tau^*})+1}(\omega^*)$, add the next measurement $y'_1 \in L_1(y'_0, \mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + 1)$ and then merges to (y'_2, y'_3) where $y'_2 \in L_1(y'_1, \mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + 2)$ to form two track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ with $\tau_z = (y_0, y_1, y_2)$ and $\tau'_z = (y'_0, \dots, y'_3)$ respectively where $m^* = 3$. Note that ω^* is a track hypothesis of $\tilde{\theta}_{\tau^*}$.

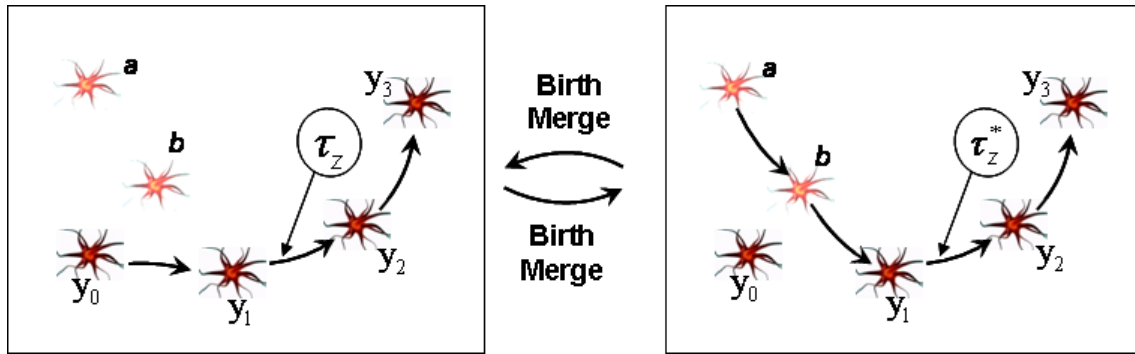
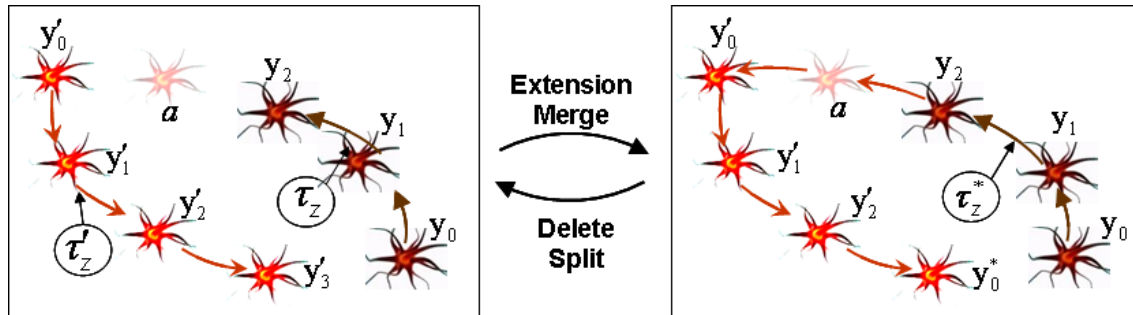
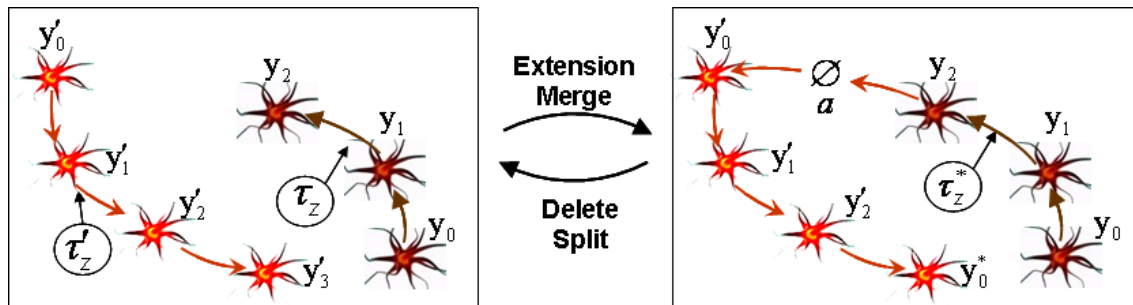


Figure 6.10: A Birth Merge move is proposed for track auxiliary variable $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_3)$ to form a track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tilde{\theta}_{\tau^*} = (a, b, y_1, \dots, y_3)$ where $a \in \Lambda_{\mathfrak{X}_0(\tilde{\theta}_\tau)-1}(\omega)$ $b \in L_1(a, \mathfrak{X}_0(\tilde{\theta}_\tau) - 1) \cap \Lambda_{\mathfrak{X}_0(\tilde{\theta}_\tau)}(\omega)$ and $y_1 \in L_1(b, \mathfrak{X}_0(\tilde{\theta}_\tau))$. Its reverse move is also a Birth Merge move.



(a) $a \in L_1(y_2, \mathfrak{X}_f(\tilde{\theta}_\tau)) \cap \Lambda_{\mathfrak{X}_f(\tilde{\theta}_\tau)+1}(\omega)$ and $y'_0 \in L_1(a, \mathfrak{X}_f(\tilde{\theta}_\tau) + 1)$



(b) $y'_0 \in L_2(y_2, \mathfrak{X}_f(\tilde{\theta}_\tau))$

Figure 6.11: An Extension Merge move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ with $\tau_z = (y_0, y_1, y_2)$ and $\tilde{\theta}_{\tau'}$ with $\tau'_z = (y'_0, \dots, y'_3)$ where $\mathfrak{X}_0(\tilde{\theta}_{\tau'}) = \mathfrak{X}_f(\tilde{\theta}_\tau) + 2$ to form a track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, y_2, a, y'_0, \dots, y'_3)$. Its reverse move, a Split Delete move, is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$ to form the track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$.

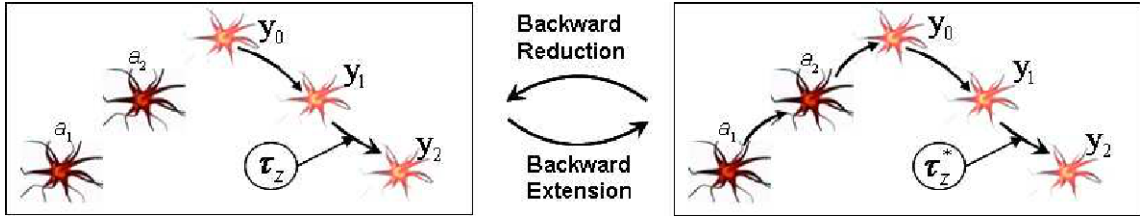


Figure 6.12: The Extension Backward move is applied to the track auxiliary variable $\tilde{\theta}_\tau$ with $\tau_z = (y_0, y_1, y_2)$ by adding two more measurement $a_1 \in \Lambda_{\mathfrak{T}_0(\tilde{\theta}_\tau)-2}(\omega)$, $a_2 \in L_1(a_1, \mathfrak{T}_0(\tilde{\theta}_\tau) - 2) \cap \Lambda_{\mathfrak{T}_0(\tilde{\theta}_\tau)-1}(\omega)$, $y_0 \in L_1(a_2, \mathfrak{T}_0(\tilde{\theta}_\tau) - 1)$ (in this example $\mathfrak{T}_0(\tilde{\theta}_\tau) > 2$) to form the track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (a_1, a_2, y_0, y_1, y_2)$. Its reverse move, the Backward Reduction move, is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$ to form a track auxiliary variable $\tilde{\theta}_\tau$ by removing the measurements $a_l \in \tau_z^*$, $l = 1, 2$.

not the first existing time of the target $\mathcal{L}(\tilde{\theta}_\tau)$ (see Figure 6.13) while a Point Update move modifies a single measurement of a track measurement (see Figures 6.14, 6.15, 6.16 and 6.17). This Point Update move is derived in this thesis to deal with problems where targets move close and/or cross each other.

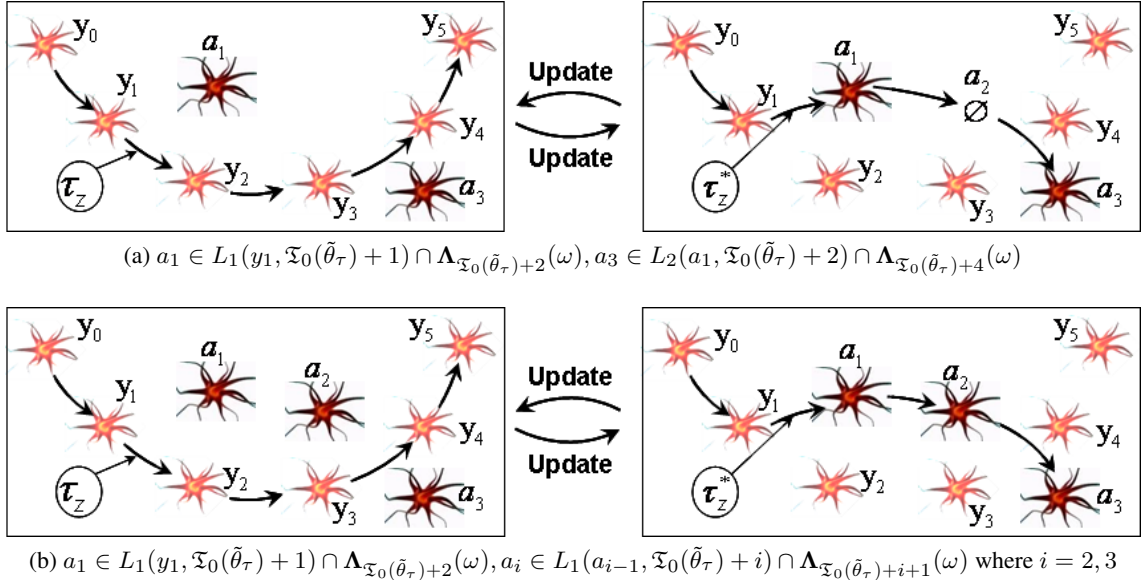


Figure 6.13: The Update move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_5)$ from time $\mathfrak{T}_0(\tilde{\theta}_\tau) + 2$ by deleting measurement y_l , $l = 2, \dots, 5$ and adding the new measurements a_r , $r = 1, \dots, 3$ where a_r shown in Figure 6.13a and 6.13b to the track measurement τ_z to form the track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, a_1, a_2, a_3)$. Its reverse, the Update move is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$.

After having introduced the purpose of these fourteen moves, the next subsection will describe the construction of these fourteen moves.

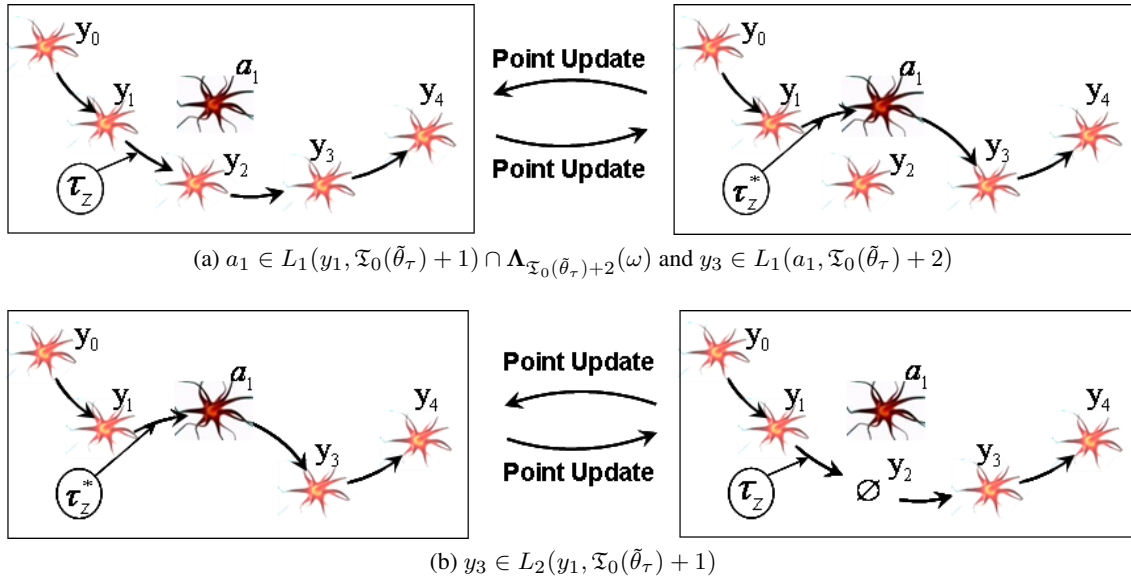


Figure 6.14: A Point Update move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_4)$ at time $\mathfrak{T}_0(\tilde{\theta}_\tau) + 2$ by exchanging the measurement y_2 by the measurement a_1 given in Figure 6.14a or Figure 6.14b to form the track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, a_1, y_3, y_4)$. Its reverse the Point Update move is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$.

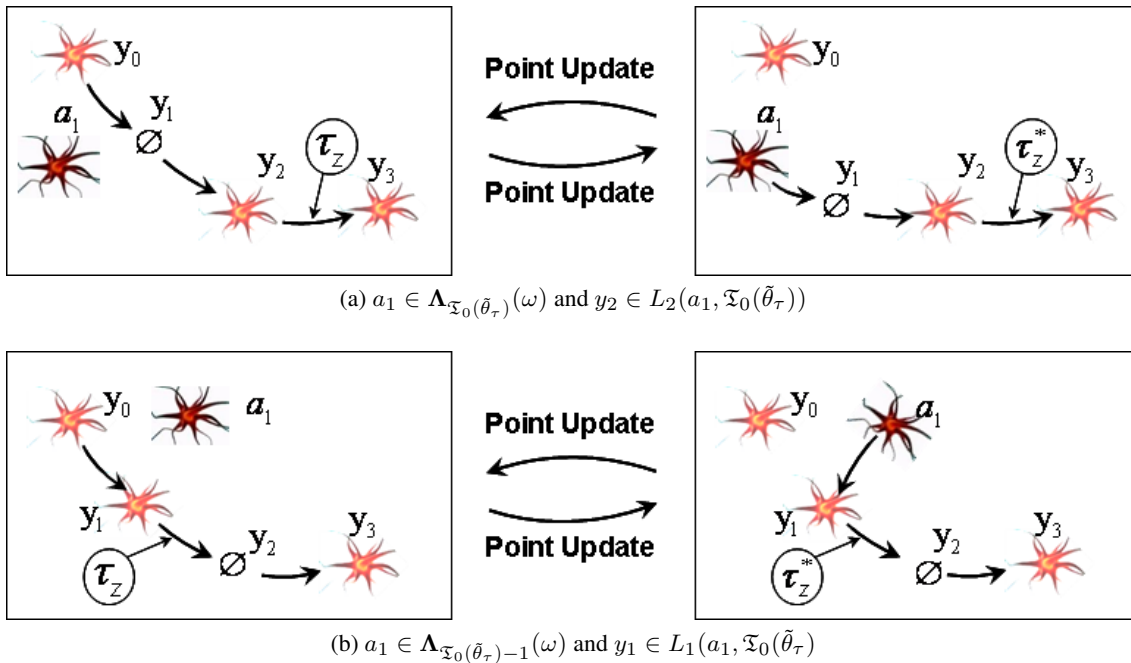


Figure 6.15: A Point Update move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_3)$ at the first existing time scan $\mathfrak{T}_0(\tilde{\theta}_\tau)$ by replacing y_0 by a_1 shown in Figure 6.15a or 6.15b to form the track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (a_1, y_1, y_2, y_3)$. Its reverse Point Update move is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$.

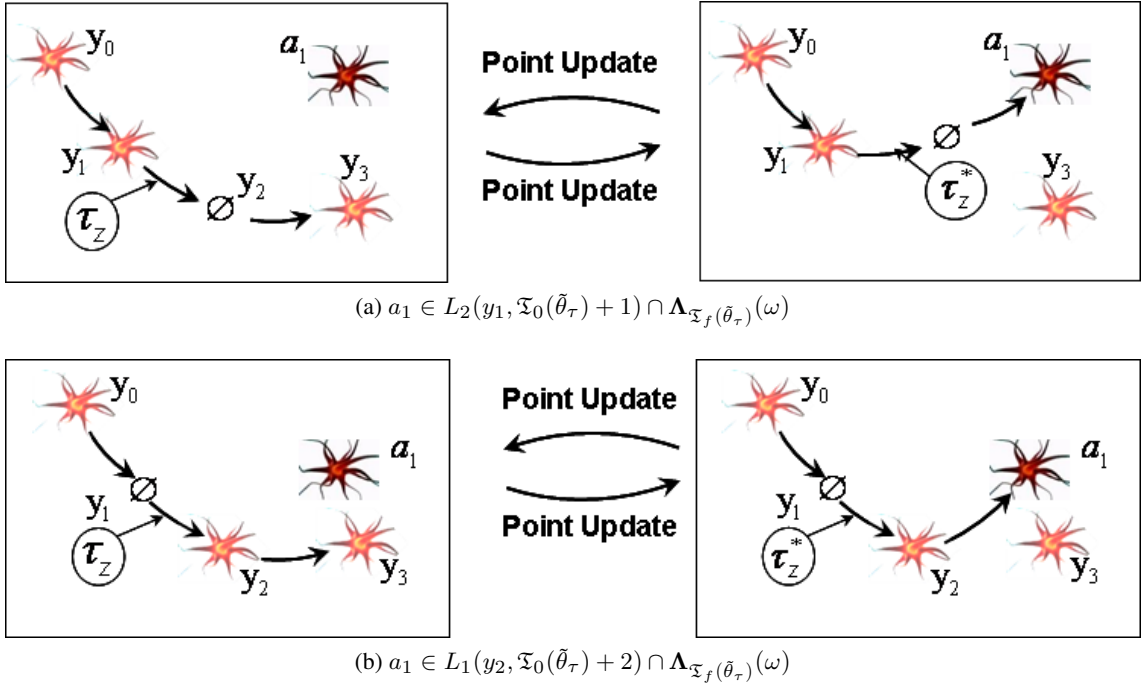


Figure 6.16: A Point Update move is proposed for track auxiliary variable $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_3)$ at the last existence time scan $\mathfrak{T}_f(\tilde{\theta}_\tau)$ of track τ by replacing y_3 by a_1 shown in Figure 6.16a or 6.16b to form the track auxiliary variable $\tilde{\theta}_{\tau^*}$ with $\tau_z^* = (y_0, y_1, y_2, a_1)$. Its reverse Point update move is applied to the track auxiliary variable $\tilde{\theta}_{\tau^*}$.

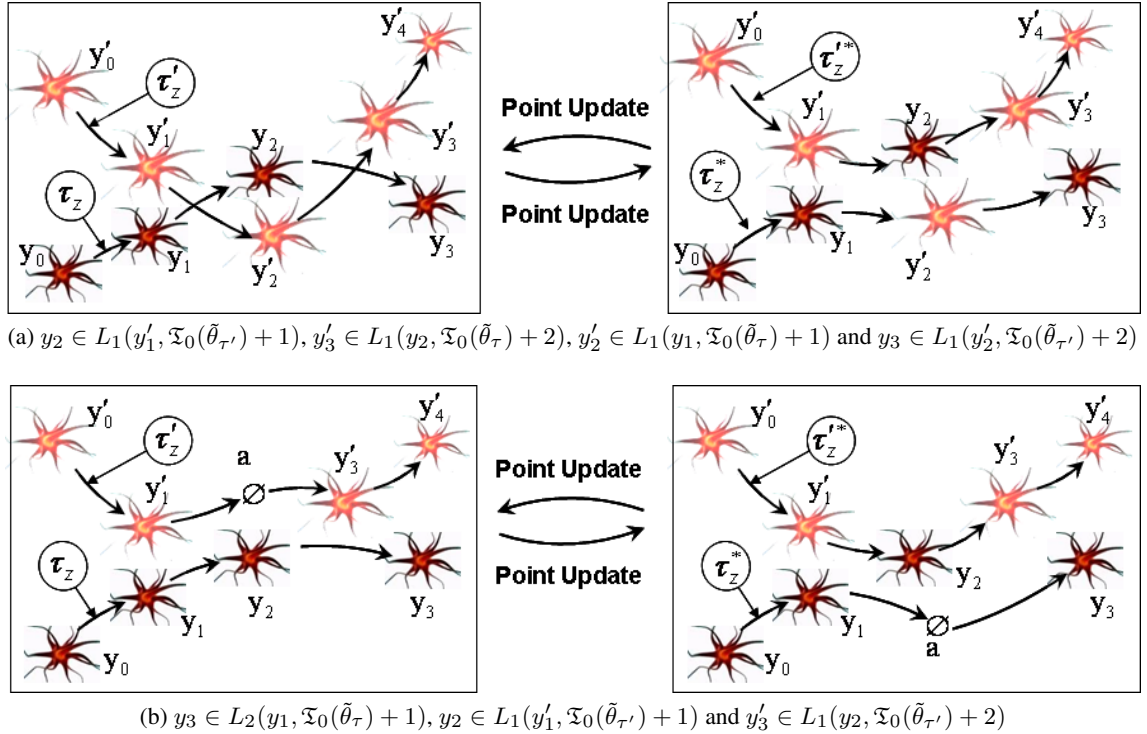


Figure 6.17: A Point Update move is proposed for track auxiliary variables $\tilde{\theta}_\tau$ with $\tau_z = (y_0, \dots, y_3)$ and $\tilde{\theta}_{\tau'}$ with $\tau'_z = (y'_0, \dots, y'_4)$ at the time scan $\mathfrak{T}_0(\tilde{\theta}_\tau) + 2$ to form the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$ with $\tau_z^* = (y_0, y_1, y'_2, y_3)$ and $\tau'_z{}^* = (y'_0, y'_1, y_2, y'_3, y'_4)$ respectively. Its reverse the Point update move is applied to the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$.

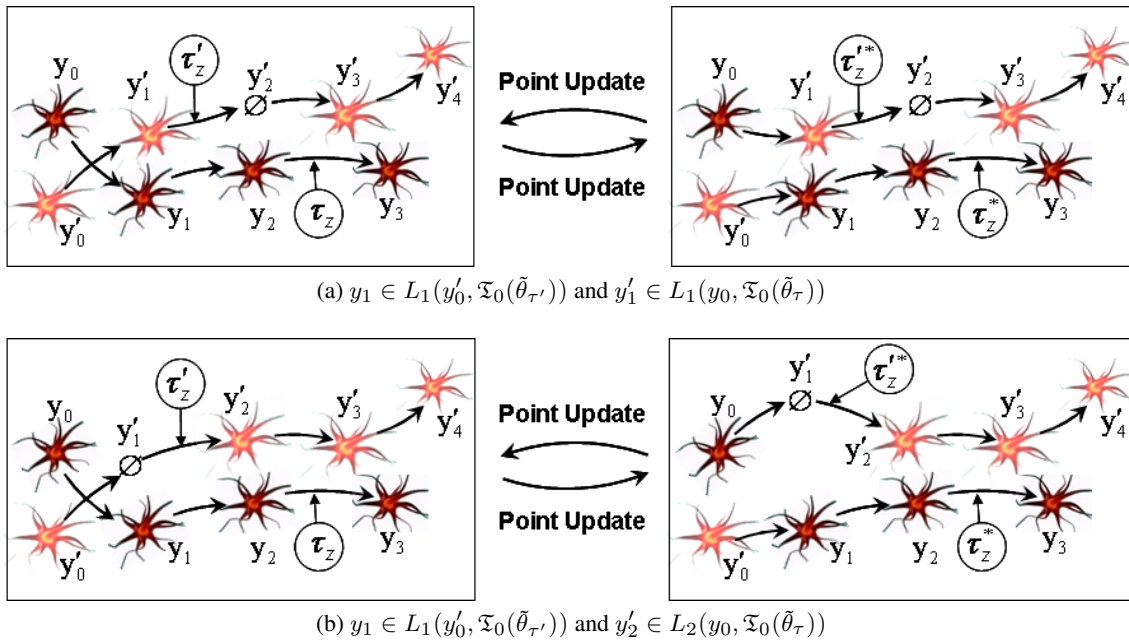


Figure 6.18: A Point Update move is proposed for track auxiliary variables $\tilde{\theta}_{\tau}$ with $\tau_z = (y_0, \dots, y_3)$ and $\tilde{\theta}_{\tau'}$ with $\tau'_z = (y'_0, \dots, y'_4)$ at the time scan $\mathfrak{T}_0(\tilde{\theta}_{\tau})$ to form the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$ with $\tau_z^* = (y_0, y_1, y_2, y_3)$ and $\tau'_z{}^* = (y_0, y'_1, \dots, y'_4)$ respectively. Its reverse, the Point update move is applied to the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$.

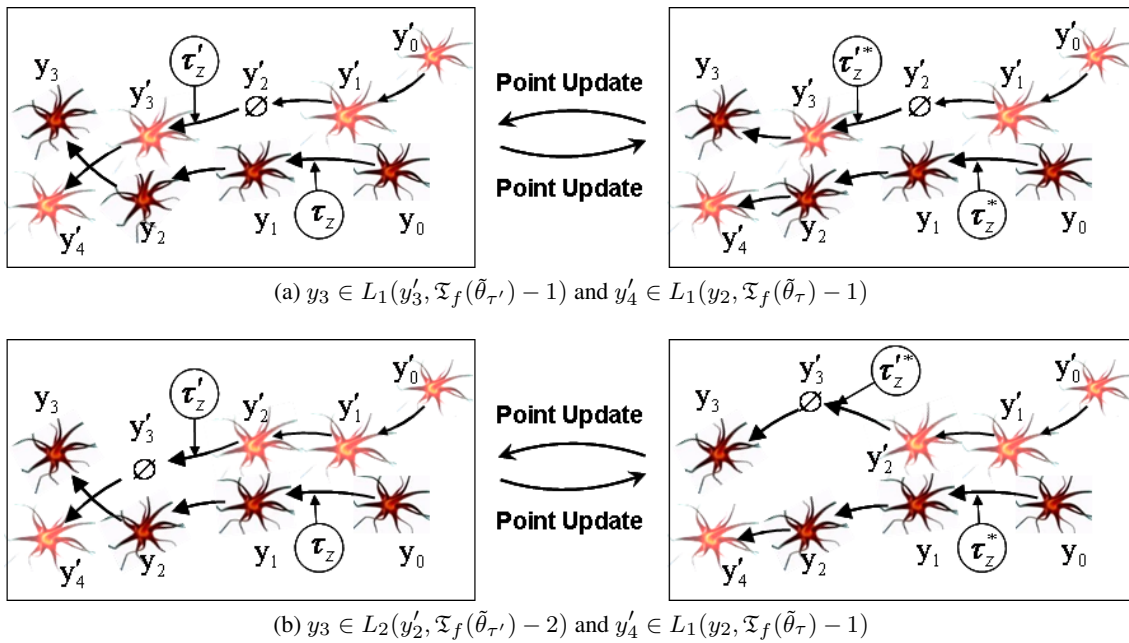


Figure 6.19: A Point Update move is proposed for track auxiliary variables $\tilde{\theta}_{\tau}$ with $\tau_z = (y_0, \dots, y_3)$ and $\tilde{\theta}_{\tau'}$ with $\tau'_z = (y'_0, \dots, y'_4)$ at the time scan $\mathfrak{T}_f(\tilde{\theta}_{\tau}) = \mathfrak{T}_f(\tilde{\theta}_{\tau'})$ to form the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$ with $\tau_z^* = (y_0, y_1, y_2, y_4)$ and $\tau'_z{}^* = (y'_0, \dots, y'_3, y_3)$ respectively. Its reverse, the Point update move is applied to the track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$.

Construction of Proposal moves: Let $K_{\omega^*} = \max_{\tau \in \omega} \mathcal{L}(\tau) + 1$ be the target label of the new target for the moves which increase the number of tracks in the track hypothesis ω .

1. **Birth and death moves:**

Birth move:

A Birth move adds a new track auxiliary variable $\tilde{\theta}_{\tau^*}$ such that $|\mathfrak{T}(\tilde{\theta}_{\tau^*})| \geq m^*$ to the track hypothesis auxiliary variable $\tilde{\theta}_{\omega}$ while keeping all other track auxiliary variables as before, forming a proposed track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = \tilde{\theta}_{\omega} \cup \{\tilde{\theta}_{\tau^*}\}$. The track auxiliary variable $\tilde{\theta}_{\tau^*}$ is constructed as follows

Before constructing the Birth move, we introduce the following notations.

At time t and for any $z' \in L^B(t) \cap \mathbf{\Lambda}_t(\omega)$ (i.e. the measurement z' has not been assigned to any existing tracks), we denote by

$$L_{\omega}(z', t) = \{(z, d) \in (\mathcal{Z}, \mathbf{d}) : z \in L_d(z', t) \cap \mathbf{\Lambda}_{t+d}(\omega)\} \quad (6.43)$$

the set of (z, d) , z is d -neighbors of z' , $d \in \mathbf{d}$. If $L_{\omega}(z', t) \neq \emptyset$, then there exist at least a measurement $z \in Z_{t+d}$ which is d -neighbor of an element z' . We also denote by

$$\begin{aligned} \bar{\mathbf{Z}}_t(\omega) = \{z \in L^B(t) \cap \mathbf{\Lambda}_t(\omega) : (z', d') \in L_{\omega}(z, t), \\ L_d(z', t + d') \cap \mathbf{\Lambda}_{t+d+d'}(\omega) \neq \emptyset, d \in \mathbf{d}\} \end{aligned} \quad (6.44)$$

the set of elements at time t for which the element z and two other measurements $z' \in L_d(z, t)$, $z'' \in L_{d'}(z', t + d + d')$ are not assigned to any existing tracks at time t , $t + d$ and $t + d + d'$ respectively where $d, d' \in \mathbf{d}$. We choose at least three consecutive target-generated measurements because any measurement and its d -neighbor are always possibly generated from the same target. By this notation, any element in this set can potentially be the initial target-generated measurement of a new target. Then we denote by $T_B(\omega)$ the set of the time scans at which a new target may appear conditional on the current track hypothesis ω as follows

$$T_B(\omega) = \{t \in \{1, \dots, T - m^* + 1\} : \bar{\mathbf{Z}}_t(\omega) \neq \emptyset\}. \quad (6.45)$$

A possible new target K_{ω^*} may enter at any time scan $t_0 \in T_B(\omega)$. Next we will describe how to construct

$$\begin{aligned} \mathbf{P}_{t_0}(\tilde{\mathbf{Z}}, \tilde{\theta}_{\omega}) = \{(K_{\omega^*}, t_0, j_0, \dots, j_m) : j_0, j_m > 0; j_i = 0 \text{ or } z_{j_i} \in Z_{t_0+i} \cap \mathbf{\Lambda}_{t_0+i}(\omega), \\ \text{for } i = 1, \dots, m; m \geq m^* - 1\} \end{aligned}$$

which is a set of new track auxiliary variables starting at time index t_0 as follows

- **Initiation:** Denote $\mathbf{P}_{t_0}(\tilde{\mathbf{Z}}, \tilde{\theta}_{\omega}, 1) = \emptyset$ as a set of new track hypothesis auxiliary variables starting at time index t_0 . An initial time of a new target t_0 is chosen from $T_B(\omega)$ given in (6.45). At initial time t_0 , track auxiliary variables is assigned to a measurement index $j_0^n > 0$, for $n = 1, \dots, |\bar{\mathbf{Z}}_{t_0}(\omega)|$. Then denote the set of the new track

auxiliary variables at time t_0 by

$$S_{t_0} = \{(K_{\omega^*}, t_0, j_0^n) : z_{j_0^n} \in \bar{\mathbf{Z}}_{t_0}(\omega), n = 1, \dots, |\bar{\mathbf{Z}}_{t_0}(\omega)|\}.$$

We must extend each element of S_{t_0} with more measurement indices.

- Existence:

(E.1) At time $t = t_0 + 1$: For each $(K_{\omega^*}, t_0, j_0^n) \in S_{t_0}$, $n = 1, \dots, |S_{t_0}|$. We extend this track auxiliary variable as follows.

The neighborhood at time t of $z_{j_0^n}$ which is not assigned to any existing tracks is $\mathbf{N}_1^\omega(z_{j_0^n}, t_0) = L_d(z_{j_0^n}, t_0) \cap \Lambda_t(\omega)$. Assign

$$\begin{aligned} S_{t_0:t_0+1}^n &= \{(K_{\omega^*}, t_0, j_0^n, j_1^{s_n}) : s_n = 0, \dots, |\mathbf{N}_1^\omega(z_{j_0^n}, t_0)|, \\ & j_1^0 = 0, z_{j_1^i} \in \mathbf{N}_1^\omega(z_{j_0^n}, t_0), i \neq 0\}. \end{aligned} \quad (6.46)$$

(E.2) At time $t > t_0 + 1$: Denote $S_{t_0:t-1} = \bigcup_{n=1}^{|S_{t_0:t-2}|} S_{t_0:t-1}^n$ as the set of the new track auxiliary variables up to time $t - 1$. If $S_{t_0:t-1} \neq \emptyset$, we consider three steps as follows

Step 1: For each $\tilde{\theta}_{\tau^*}^n = (K_{\omega^*}, t_0, j_0^n, \dots, j_k^n) \in S_{t_0:t-1}$, $n = 1, \dots, |S_{t_0:t-1}|$. We want to extend one more measurement by finding the last time $t_0 + l$ where the target is detected i.e. $l = \max\{i : j_i^n > 0, i = 0, \dots, k\}$. We need to look for a new measurement in the d -neighborhood of the last measurement $z_{j_l^n}$ where $d = k - l + 1$ and d is always less than or equal to $\bar{d} + 1$. Its neighborhood at time t which is not assigned to any existing tracks at time t is $\mathbf{N}_d^\omega(z_{j_l^n}, t_0 + l) = L_d(z_{j_l^n}, t_0 + l) \cap \Lambda_t(\omega)$. Then assign

$$\begin{aligned} S_{t_0:t}^n &= \{(K_{\omega^*}, t_0, j_0^n, \dots, j_k^n, j_{k+1}^{s_n}) : s_n = 0, \dots, |\mathbf{N}_d^\omega(z_{j_l^n}, t_0 + l)|, \\ & j_{k+1}^0 = 0, z_{j_{k+1}^i} \in \mathbf{N}_d^\omega(z_{j_l^n}, t_0 + l), i \neq 0\}. \end{aligned}$$

Step 2: Denote $S_{t_0:t} = \bigcup_{n=1}^{|S_{t_0:t-1}|} S_{t_0:t}^n$ as the set of the new track auxiliary variables up to time t . If $S_{t_0:t} \neq \emptyset$ and $t \leq \max T_B(\omega)$, we consider an element $\tilde{\theta}_{\tau^*} = (K_{\omega^*}, t_0, j_0^*, \dots, j_{t-t_0}^*) \in S_{t_0:t}$ (note that $t - t_0 = k + 1$). If $\tilde{\theta}_{\tau^*}$ has more than \bar{d} consecutive zeros (i.e. $j_i^* = 0$ for $i \geq t - t_0 - \bar{d} - 1$), remove $\tilde{\theta}_{\tau^*}$ from $S_{t_0:t}$. Otherwise if $z_{j_{t-t_0}^*}$ is the last measurement generated by the target and the duration time of the target is larger than or equal to m^* (i.e. $j_{t-t_0}^* > 0$ and $t - t_0 \geq m^* - 1$), then assign $\tilde{\theta}_\omega \cup \{\tilde{\theta}_{\tau^*}\}$ to $\mathbf{P}_{t_0}(\tilde{\mathbf{Z}}, \tilde{\theta}_\omega)$ and keep $\tilde{\theta}_{\tau^*}$ for further extension.

Step 3: repeat (E.2) until $\max T_B(\omega)$.

Then the set of all proposed track hypothesis auxiliary variables for a Birth move ($m = 1$) is

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{\mathbf{Z}}, 1) = \prod_{t \in T_B(\omega)} \mathbf{P}_t(\tilde{\mathbf{Z}}, \tilde{\theta}_\omega) \quad (6.47)$$

Note that a possible Birth move could also been formed backwards in time starting with the final time t_f and finishing with the initial time t_0 . Such constructions are useful if we would like to construct a feasible association which ends at a particular measurement (e.g. Birth Merge move and Backward extension moves). In practice, when there are a large possible new track hypotheses in the set $\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 1)$, the computation is very expensive. The culling is required to reduce the size of the set.

Culling:

Culling is an implementation issue when dealing with a large number of a set. This culling is done recursively in time as the measurement association is being built up. The culling is based on the likelihood of the measurements associated with a track. As these are calculated in terms of the transition density $\bar{f}_{t|t-1}$, we will mainly consider forward constructions of the set of the measurement association. The exceptions are the cases when the set of possible new track-measurement associations is so small that no culling is needed. In those cases, either a forward or a backward in time construction can be used.

We introduce the following notations

- Let $r(z'_{j_0}, z_j, d)$ denote a suitable function to estimate the initial target state from the measurement z'_{j_0} and its neighbor z_j . In this thesis, we assume our measurements are the position measurements of the targets. However, this formula can be extend to other type of measurements. For example, when $x = [\xi, \zeta, v_\xi, v_\zeta]^T$ where (ξ, ζ) denotes the true target position in the two dimensional Cartesian plane, (v_ξ, v_ζ) is its velocity and $z = [\xi', \zeta']^T = [\xi, \zeta]^T + v_{noise}$ the position of the target observed by a sensor with a two dimensional vector noise v_{noise} . Note that y^{Tr} is the transpose of y . Then a possible function is $r(z'_{j_0}, z_j, d) = [z'_{j_0}, (z_j - z'_{j_0})/d]^T$.

There are many possibilities for constructing a new track hypothesis auxiliary variable starting with a measurement $z_{j_0} \in \bar{\mathbf{Z}}_{t_0}(\omega)$ because the number of elements in the set $L_\omega(z_{j_0}, t_0)$ may be larger than 1 and in general the d -neighborhood of z_{j_0} or its neighbors is large. The Birth move is reconstructed as follows with a given measurement gate threshold g_z .

Denote $\mathbf{P}_{t_0}(\tilde{Z}, \tilde{\theta}_\omega, z_{j_0})$ as a set of samples starting from measurement $z_{j_0} \in \bar{\mathbf{Z}}_{t_0}(\omega)$ at time $t_0 \in T_B(\omega)$.

- Initiation: We denote a collection of the sets which consists of the track auxiliary variable and sequence of states starting from the initial measurement $z_{j_0} \in \bar{\mathbf{Z}}_{t_0}(\omega)$ at time $t_0 \in T_B(\omega)$ by

$$S_{t_0}(z_{j_0}) = \left\{ \{ \tilde{\theta}_{\tau^*}^n, x_{t_0}^n \} : \tilde{\theta}_{\tau^*}^n = (K_{\omega^*}, t_0, j_0), x_{t_0}^n = r(z_{j_0}, z_{j^n}, d^n), \right. \\ \left. (z_{j^n}, d^n) \in L_\omega(z_{j_0}, t_0) \right\}$$

where $\tilde{\theta}_{\tau^*}^n = (K_{\omega^*}, t_0, j_0)$ are the same for all $n = 1, \dots, |L_\omega(z_{j_0}, t_0)|$ but the initial states $x_{t_0}^n$ are different in velocity, acceleration or etc. Next we will find the next measurements which generate from this new target starting from z_{j_0}

- **Existence:** Similar to the previous construction for this step, we consider two cases as follows.

(F.1) At time $t = t_0 + 1$: We construct the next single target states of $x_{t_0}^n$ where $\{\tilde{\theta}_{\tau^*}^n, x_{t_0}^n\} \in S_{t_0}(z_{j_0})$ for $n = 1, \dots, |S_{t_0}(z_{j_0})|$. We consider two steps as follows. Similar to the previous construction, for each $x_{t_0}^n$, a collection of new sets up to time t is

$$\begin{aligned} S_{t_0:t}^n(z_{j_0}) &= \left\{ \{\tilde{\theta}_{\tau^*}^{i_n}, x_{t_0:t}^{i_n}\} : \tilde{\theta}_{\tau^*}^{i_n} = (K_{\omega^*}, t_0, j_0, j_1^{i_n}); i_n = 0, \dots, |L_1(z_{j_0}, t_0)|; \right. \\ &\quad \left. x_{t_0:t-1}^{i_n} = x_{t_0:t-1}^n; j_1^0 = 0, z_{j_1^{i_n}} \in L_1(z_{j_0}, t_0), i_n \neq 0; \right. \\ &\quad \left. x_t^{i_n} \sim f_{t|t-1}(\cdot | x_{t-1}^n); \bar{g}_t(z_{j_1^{i_n}} | x_t^{i_n}) \geq g_z \right\} \end{aligned}$$

where a measurement gate threshold g_z is given to make this set $S_{t_0:t}^n(z_{j_0})$ smaller than the set in (6.46) by select the samples $\{(K_{\omega^*}, t_0, j_0, j_1^{i_n}), x_{1:t}^{i_n}\}$ which have the most likely chance to occur ($\bar{g}_t(z_{j_1^{i_n}} | x_t^{i_n}) \geq g_z$).

(F.2) At time $t > t_0 + 1$: Let $S_{t_0:t-1}(z_{j_0}) = \bigcup_{n=1}^{|S_{t_0:t-2}(z_{j_0})|} S_{t_0:t-1}^n(z_{j_0})$.

If $S_{t_0:t-1}^n(z_{j_0}) \neq \emptyset$, we consider three steps similar to the ones in the previous construction as follows.

Step 1: Similar to the previous construction, we want to extend one more measurement by finding the last time $t_0 + l$ where the target is detected i.e. $l = \max\{i : j_i^n > 0, i = 0, \dots, k\}$. We need to look for a new measurement in the set d -neighborhood of the last measurement $z_{j_l^n}$ where $d = k - l + 1$. Its neighborhood at time t which is not assigned to any existing tracks at time t is $\mathbf{N}_d^\omega(z_{j_l^n}, t_0 + l) = L_d(z_{j_l^n}, t_0 + l) \cap \Lambda_t(\omega)$. Then assign

$$\begin{aligned} S_{t_0:t}^n(z_{j_0}) &= \left\{ \{\tilde{\theta}_{\omega^*}^{i_n}, x_{t_0:t}^{i_n}\} : \tilde{\theta}_{\omega^*}^{i_n} = (K_{\omega^*}, t_0, j_0^n \dots, j_k^n, j_{k+1}^{i_n}), j_{k+1}^0 = 0, \right. \\ &\quad \left. z_{j_{k+1}^{i_n}} \in \mathbf{N}_d^\omega(z_{j_l^n}, t_0 + l), i_n \neq 0, x_{t_0:t-1}^{i_n} = x_{t_0:t-1}^n, \right. \\ &\quad \left. x_t^{i_n} \sim f_{t|t-1}(\cdot | x_{t-1}^n), \bar{g}_t(z_{j_{k+1}^{i_n}} | x_t^{i_n}) \geq g_z \right\}. \end{aligned}$$

Step 2: Let $S_{t_0:t}(z_{j_0}) = \bigcup_{n=1}^{|S_{t_0:t-1}(z_{j_0})|} S_{t_0:t}^n(z_{j_0})$.

If $S_{t_0:t}(z_{j_0}) \neq \emptyset$ and $t \leq \max T_B(\omega)$, we consider an element

$$\tilde{\theta}_{\tau^*} = (K_{\omega^*}, t_0, j_0^*, \dots, j_{t-t_0}^*)$$

where $\{\tilde{\theta}_{\tau^*}, x_{t_0:t}^*\} \in S_{t_0:t}(z_{j_0})$ (note that $t - t_0 = k + 1$). If $\tilde{\theta}_{\tau^*}$ has more than \bar{d} consecutive zeros (i.e. $j_i^* = 0$ for $i \geq t - t_0 - \bar{d} - 1$), remove $\{\tilde{\theta}_{\tau^*}, x_{t_0:t}^*\}$ from $S_{t_0:t}(z_{j_0})$. Otherwise if $z_{j_{t-t_0}^*}$ is the last measurement generated by the target and duration time of the target is larger than or equal to m^* (i.e. $j_{t-t_0}^* > 0$ and $t - t_0 \geq m^* - 1$). Then assign $\tilde{\theta}_{\omega} \cup \{\tilde{\theta}_{\tau^*}\}$ to $\mathbf{P}_{t_0}(\tilde{Z}, \tilde{\theta}_{\omega}, z_{j_0})$. The element $\tilde{\theta}_{\tau^*}$ still keeps for further extension.

Step 3: repeat (F.2) until $\max T_B(\omega)$.

With this new construction, the set $\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 1)$ in (6.47) can be reduced in size and is rewritten as follows

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 1) = \prod_{t \in T_B(\omega)} \prod_{z \in \tilde{Z}_t(\omega)} \mathbf{P}_t(\tilde{Z}, \tilde{\theta}_\omega, z).$$

Death move

A death move is the reverse of a birth move. The death move is constructed so that it may revert to the initial track hypothesis auxiliary variable after a birth move (see Figure 6.4). A track auxiliary variable $\tilde{\theta}_{\tau^*}$ is removing from $\tilde{\theta}_\omega$ while keeping all other track auxiliary variables as before, forming a proposed track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = \tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}$. Then the set of all proposed track hypothesis auxiliary variables for a Death move ($m = 12$) is

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 12) = \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = \tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}, \tilde{\theta}_{\tau^*} \in \tilde{\theta}_\omega\}.$$

2. Split and Merge moves:

Split move:

The Split move is proposed for a track auxiliary variable $\tilde{\theta}_{\tau^*}$ by dividing $\tilde{\theta}_{\tau^*}$ into two track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ if the duration time of the target $\mathcal{L}(\tilde{\theta}_{\tau^*})$ is larger than or equal to $2m^*$ i.e. $|\mathfrak{T}(\tilde{\theta}_{\tau^*})| \geq 2m^*$ and the following conditions hold:

- (SP1) The last existing time scan of the proposed target $\mathcal{L}(\tilde{\theta}_\tau)$ and the first existing time scan of the proposed target $\mathcal{L}(\tilde{\theta}_{\tau'})$ are chosen such that the target $\mathcal{L}(\tilde{\theta}_{\tau^*})$ is detected at those time scans.
- (SP2) The duration of existence for proposed targets $\mathcal{L}(\tilde{\theta}_\tau)$ and $\mathcal{L}(\tilde{\theta}_{\tau'})$ are larger than or equal to m^* .

Denote by t_1 and t_2 the last existing time scan of the proposed target $\mathcal{L}(\tilde{\theta}_\tau)$ and the first existing time scan of the proposed target $\mathcal{L}(\tilde{\theta}_{\tau'})$ respectively (i.e. $t_1 = \mathfrak{T}_f(\tilde{\theta}_\tau)$ and $t_2 = \mathfrak{T}_0(\tilde{\theta}_{\tau'})$). Mathematically, (SP1) and (SP2) can be written as follows.

- $\exists t_1 \in \{\mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + m^* - 1, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^*\}$, such that $\mathfrak{I}_{t_1}(\tilde{\theta}_{\tau^*}) > 0$ and
- $\exists t_2 \in \{t_1 + 1, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^* + 1\}$ such that $\mathfrak{I}_{t_2}(\tilde{\theta}_{\tau^*}) > 0$.

If $t_2 \in \{t_1 + 2, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^* + 1\}$, the move is called a Delete Split move.

The Split/Delete Split move is applied to the track auxiliary $\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m)$ to propose two new track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ as follows

$$\begin{aligned} \tilde{\theta}_\tau &= (k, t, j_0, \dots, j_{t_1-t}) \\ \tilde{\theta}_{\tau'} &= (K_{\omega^*}, t_2, j_{t_2-t}, \dots, j_m). \end{aligned}$$

The set of all proposed track hypothesis auxiliary variables for the Split move ($m = 7$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 7) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}) \cup \{\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}\}; \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, \\ & m \geq 2m^* - 1; \tilde{\theta}_\tau = (k, t, j_0, \dots, j_{t_1-t}), \tilde{\theta}_{\tau'} = (K_{\omega^*}, t_2, j_{t_2-t}, \dots, j_m), \\ & t_1 \in \{t + m^* - 1, \dots, t + m - m^*\}, j_{t_1-t} > 0, \\ & t_2 \in \{t_1 + 1, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^* + 1\}, j_{t_2-t} > 0\}. \end{aligned}$$

Merge move:

The Merge move is the reverse of the Split move. When $K_\omega > 1$, the merge move can take place if there exists a pair of track auxiliary variables $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ such that the first target-generated measurement from the target $\mathcal{L}(\tau')$ is a neighbor of the last target-generated measurement from the target $\mathcal{L}(\tau)$ i.e. $z_{j'} \in L_d(z_j, \mathfrak{T}_f(\tilde{\theta}_\tau))$ where $z_{j'} \in Z_{\mathfrak{T}_0(\tilde{\theta}_{\tau'})}$ and $z_j \in Z_{\mathfrak{T}_f(\tilde{\theta}_\tau)}$ for $d \in \mathbf{d}$. Mathematically, the following set of possible merge move pairs is given by

$$\begin{aligned} \mathbf{M} = & \left\{ (\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}) \in (\tilde{\theta}_\omega, \tilde{\theta}_\omega) : d = \mathfrak{T}_0(\tilde{\theta}_{\tau'}) - \mathfrak{T}_f(\tilde{\theta}_\tau), d \in \mathbf{d}, \right. \\ & \left. \mathfrak{T}_{\mathfrak{T}_0(\tilde{\theta}_{\tau'})}(\tilde{\theta}_{\tau'}) = j' > 0, \mathfrak{T}_{\mathfrak{T}_f(\tilde{\theta}_\tau)}(\tilde{\theta}_\tau) = j > 0, z_{j'} \in L_d(z_j, \mathfrak{T}_f(\tilde{\theta}_\tau)) \right\}. \end{aligned}$$

where $d = \mathfrak{T}_0(\tilde{\theta}_{\tau'}) - \mathfrak{T}_f(\tilde{\theta}_\tau)$ is the distance between the first time index of the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ and the last time index of the target $\mathcal{L}(\tilde{\theta}_\tau)$. This distance must be positive. $z_{j'} \in L_d(z_j, \mathfrak{T}_f(\tilde{\theta}_\tau))$ means that the first target-generated measurement of target $\mathcal{L}(\tilde{\theta}_{\tau'})$ must be in the d -neighbor of the last target-generated measurement of target $\mathcal{L}(\tilde{\theta}_\tau)$. Note that the order of $(\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}) \in \mathbf{M}$ means that the track auxiliary variable $\tilde{\theta}_\tau$ merges to the track auxiliary variable $\tilde{\theta}_{\tau'}$.

For any pair $(\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}) \in \mathbf{M}$, the Merge move is constructed by combining two track auxiliary variables $\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m)$ and $\tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n)$ to form a single track auxiliary variable

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, \underbrace{0, \dots, 0}_{d-1}, j'_0, \dots, j'_n), d = t' - t - m \in \mathbf{d}.$$

Then the set of all proposed track hypothesis auxiliary variables for the Merge move ($m = 5$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 5) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}\}, (\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}) \in \mathbf{M}, \\ & \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m), \tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n), d = t' - t - m \in \mathbf{d}, \\ & \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, \underbrace{0, \dots, 0}_{d-1}, j'_0, \dots, j'_n)\}. \end{aligned}$$

3. Extension and Reduction move:

Extension: The Extension move is proposed for $\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega$ where the last existing time scan of the target $\mathcal{L}(\tilde{\theta}_\tau)$ is less than the last scan T i.e. $\mathfrak{T}_f(\tilde{\theta}_\tau) < T$. A sequence (j_1^*, \dots, j_n^*) is added to $\tilde{\theta}_\tau$ to form the proposed track auxiliary variable

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, j_1^*, \dots, j_n^*)$$

where $j_n^* \in \mathbf{\Lambda}_{t+m+n}(\omega)$, $j_i^* = 0$ or $j_i^* \in \mathbf{\Lambda}_{t+m+i}(\omega)$ for $i = 1, \dots, n-1$; and $A^* = (k, t+m, j_m, j_1^*, \dots, j_n^*)$ is constructed in the same as in the Birth move where the initial measurement of the new target in the Birth move is the last measurement generated by the target $\mathcal{L}(\tilde{\theta}_\tau)$ at time $\mathfrak{T}_f(\tilde{\theta}_\tau) = t+m$. The set of all proposed track hypothesis auxiliary variables for the Extension move ($m=2$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 2) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}; \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m), t+m < T; \\ &\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, j_1^*, \dots, j_n^*); j_n^* \in \mathbf{\Lambda}_{t+m+n}(\omega); \\ &j_i^* = 0 \text{ or } j_i^* \in \mathbf{\Lambda}_{t+m+i}, i = 1, \dots, n-1\}. \end{aligned}$$

Reduction: The Reduction move is proposed for $\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega$ if $m > m^*$ by deleting the index measurement j_{t_0-t+1}, \dots, j_m where t_0 is chosen from $\{\mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + m^* - 1, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - 1\}$ such that $\mathfrak{T}_{t_0}(\tilde{\theta}_{\tau^*}) > 0$. The new track auxiliary variable $\tilde{\theta}_\tau$ is

$$\tilde{\theta}_\tau = (k, t, j_0, \dots, j_{t_0-t}).$$

The set of all proposed track hypothesis auxiliary variables for the Reduction move ($m=8$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 8) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}) \cup \{\tilde{\theta}_\tau\}; \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, m \geq m^*; \\ &\tilde{\theta}_\tau = (k, t, j_0, \dots, j_{t_0-t}), t_0 - t \geq m^* - 1\}. \end{aligned}$$

4. Switch move:

The switch move is proposed to exchange some measurement indices between targets $\mathcal{L}(\tilde{\theta}_\tau)$ and $\mathcal{L}(\tilde{\theta}_{\tau'})$ while keeping all other the measurement indices from other targets fixed (see Figure 6.7) as follows.

Firstly, we define a set M_s which collects all pairs of measurement indices and their time indices which can be switched between two track auxiliary variables

$$\begin{aligned} M_s &= \{(\tilde{\theta}_\tau, t_0, \tilde{\theta}_{\tau'}, t'_0) : \mathfrak{T}_{t_0}(\tilde{\theta}_\tau), \mathfrak{T}_{t'_0}(\tilde{\theta}_{\tau'}) > 0; \tilde{\theta}_\tau, \tilde{\theta}_{\tau'} \in \tilde{\theta}_\omega; \\ &\mathfrak{T}_{t_0+d}(\mathcal{L}(\tilde{\theta}_\tau)) = j'_d > 0 \text{ and } \mathfrak{T}_{t_0+l}(\mathcal{L}(\tilde{\theta}_\tau)) = 0, l = 1, \dots, d-1, d \geq 1; \\ &\mathfrak{T}_{t'_0+d'}(\mathcal{L}(\tilde{\theta}_{\tau'})) = j_d > 0 \text{ and } \mathfrak{T}_{t'_0+s}(\mathcal{L}(\tilde{\theta}_{\tau'})) = 0, s = 1, \dots, d'-1, d' \geq 1 \\ &z_{j'_d} \in L_d(z_{j'}, t'_0); z_{j_d} \in L_{d'}(z_j, t_0)\}. \end{aligned} \quad (6.48)$$

where $z_j \in Z_{t_0}, z_{j'} \in Z_{t'_0}, z_{j'_d} \in Z_{t'_0+d}, z_{j_{d'}} \in Z_{t_0+d}$. The time scan $t_0 + d$ is the first time after t_0 for which the target $\mathcal{L}(\tilde{\theta}_\tau)$ is detected and $t'_0 + d'$ is the first time after t'_0 for which the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is detected. For example, in Figure 6.7, $t_0 = \mathfrak{T}_0(\tilde{\theta}_\tau) + 2 = t'_0 = \mathfrak{T}_0(\tilde{\theta}_{\tau'}) + 1$. In Figure 6.7a, $d = d' = 1$ while in Figure 6.7b $d = 1, d' = 3$. The Switch move is applied to the two track auxiliary variables $\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m)$ and $\tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n)$ at time scans $t_0 + d$ and $t'_0 + d$ respectively where $(\tilde{\theta}_\tau, t_0, \tilde{\theta}_{\tau'}, t'_0) \in M_s$ to form new track auxiliary variables $\tilde{\theta}_{\tau^*}$ and $\tilde{\theta}_{\tau'^*}$ (see Figure 6.7) as follows

$$\begin{aligned}\tilde{\theta}_{\tau^*} &= (k, t, j_0, \dots, j_{t_0-t}, \overbrace{0, \dots, 0}^{d'-1}, j'_{t'_0-t'+d}, \dots, j'_n) \\ \tilde{\theta}_{\tau'^*} &= (k', t', j'_0, \dots, j'_{t'_0-t'}, \underbrace{0, \dots, 0}_{d-1}, j_{t_0-t+d}, \dots, j_m).\end{aligned}$$

The set of all proposed track hypothesis auxiliary variables for the Switch move ($m = 6$) is

$$\begin{aligned}\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 6) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}; \\ &|\mathfrak{T}(\tilde{\theta}_{\tau^*})|, |\mathfrak{T}(\tilde{\theta}_{\tau'^*})| \geq m^*; (\tilde{\theta}_\tau, t_0, \tilde{\theta}_{\tau'}, t'_0) \in M_s\}.\end{aligned}$$

5. Extension Merge move/Birth Merge move and Extension Merge move/Delete Split move:

Extension Merge move: If $|\tilde{\theta}_\omega| > 1$, two track auxiliary variables $\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m)$ and $\tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n) \in \tilde{\theta}_\omega$ are used in the Extension Merge move as follows.

Firstly, the track auxiliary variable $\tilde{\theta}_\tau$ is extended in the same way as for an Extension move up to time $t + m'$ to form

$$\tilde{\theta}_{\tau^\bullet} = (k, t, j_0, \dots, j_m, j_{m+1}, \dots, j_{m'}), \quad j_{m'} > 0 \text{ and } m' > m.$$

where $j_i = 0$ or $z_{j_i} \in \Lambda_{t+i}(\omega)$ for $i = m + 1, \dots, m' - 1$ and $z_{j_{m'}} \in \Lambda_{t+m'}(\omega)$. Secondly, if at time $t_0 = t + m' + d \in \mathfrak{T}(\tilde{\theta}_{\tau'})$, $d \in \mathbf{d}$, the target-generated measurement of target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is in the d -neighborhood of the last target-generated measurement of track $\mathcal{L}(\tilde{\theta}_{\tau^\bullet})$ (i.e. $z_{j'_{t_0-t'}} \in L_d(z_{j_{m'}}, t + m')$ where $z_{j'_{t_0-t'}} \in Z_{t_0}$, $\mathfrak{T}_{t_0}(\tilde{\theta}_{\tau'}) = j'_{t_0-t'}$), then $\tilde{\theta}_{\tau^\bullet}$ and $\tilde{\theta}_{\tau'}$ are merged measurements as follows. Here, we consider two situations for d .

Case 1: When $d > 1$ (see Figure 6.11 for illustration)

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, \dots, j_{m'}, \underbrace{0, \dots, 0}_{d-1}, j'_{t_0-t'}, \dots, j'_n).$$

Case 2: When $d = 1$ (see Figures 6.9 and 6.8 for illustration)

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m, j_{m+1}, \dots, j_{m'}, j'_{t_0-t'}, \dots, j'_n).$$

The construction of this Extension Merge move leaves us with the remaining elements of $\tilde{\theta}_{\tau'}$ which are not merged into $\tilde{\theta}_{\tau}$ i.e.

$$\tilde{\theta}_{\tau'^*} = (k', t', j'_0, \dots, j'_{t_1-t'}) \quad (6.49)$$

where t_1 is the latest time before t_0 at which the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is observed by a sensor. For example, when $m^* = 3$, in Figure 6.11, $t_0 = t$ and there is no measurement left; in Figure 6.9b, there are two measurements j'_0 and j'_1 left, i.e. $\tilde{\theta}_{\tau'^*} = (k', t', j'_0, j'_1)$; in Figure 6.9a, there are one measurement j'_0 left, i.e. $\tilde{\theta}_{\tau'^*} = (k', t', j'_0)$; and in Figure 6.8, there are three measurements j'_0, j'_1 and j'_2 left i.e. $\tilde{\theta}_{\tau'^*} = (k', t', j'_0, \dots, j'_2)$. If $|\mathfrak{T}(\tilde{\theta}_{\tau'^*})| \geq m^*$, the proposed track hypothesis auxiliary variable is $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}$ (see Figure 6.8). In this case, the set of proposed track hypothesis auxiliary variables for the Extension Merge move ($m = 4$) is

$$\begin{aligned} \mathbf{E}^1(\tilde{\theta}_{\omega}, \tilde{Z}) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}; \tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m), \\ & \tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n) \in \tilde{\theta}_{\omega}; \tilde{\theta}_{\tau'^*} = (k', t', j'_0, \dots, j'_{t_1-t'}), \\ & \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_{m'}, \underbrace{0, \dots, 0}_{d-1}, j'_{t_0-t'}, \dots, j'_n), d \geq 1, m' > m; \\ & z_{j_{m'}} \in \Lambda_{t+m'}(\omega); j_i = 0 \text{ or } z_{j_i} \in \Lambda_{t+i}(\omega) \text{ for } i = m+1, \dots, m'-1; \\ & t_1 = \max\{i : i < t_0, j_{i-t'} > 0\}, t_1 - t' \geq m^* - 1\}. \end{aligned}$$

Otherwise, the proposed track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}\}$ (see Figures 6.11 and 6.9). In this case, the set of proposed track hypothesis auxiliary variables for the Extension Merge move ($m = 4$) is

$$\begin{aligned} \mathbf{E}^2(\tilde{\theta}_{\omega}, \tilde{Z}) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}\}; \tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m), \\ & \tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n) \in \tilde{\theta}_{\omega}; \\ & \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_{m'}, \underbrace{0, \dots, 0}_{d-1}, j'_{t_0-t'}, \dots, j'_n), d \geq 1, m' > m; \\ & z_{j_{m'}} \in \Lambda_{t+m'}(\omega); j_i = 0 \text{ or } z_{j_i} \in \Lambda_{t+i}(\omega), i = m+1, \dots, m'-1\}. \end{aligned}$$

The set of all proposed track hypothesis auxiliary variables for the Extension Merge move ($m = 4$) is

$$\mathbf{P}(\tilde{\theta}_{\omega}, \tilde{Z}, 4) = \mathbf{E}^1(\tilde{\theta}_{\omega}, \tilde{Z}) \cup \mathbf{E}^2(\tilde{\theta}_{\omega}, \tilde{Z}).$$

Birth Merge move:

A Birth Merge move is a combination of a Birth move and a Merge move. Thus, this move is divided in two steps. The first step is a Birth move to propose $\tilde{\theta}_{\tau^{\bullet}} = (K_{\omega^*}, t^{\bullet}, j_0^{\bullet}, \dots, j_n^{\bullet})$ with $j_n^{\bullet} > 0$ ($n \geq 0$). If a d -neighbor of j_n^{\bullet} is assigned to an existing track τ , the second step is to merge $\tilde{\theta}_{\tau^{\bullet}}$ to the existing track auxiliary variable $\tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_{\omega}$ at

time $t_0 = t^\bullet + n + d > t$ as done in the Extension Merge move. If $t_0 = t$, the move is called Extension Backward move which is discussed later. Similar to the Extension Merge move, this construction also leaves us with the remaining elements of $\tilde{\theta}_\tau$ which are not merged into $\tilde{\theta}_{\tau^\bullet}$ i.e

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_{t_1-t}) \quad (6.50)$$

where t_1 is the latest time before t_0 at which the target $\mathcal{L}(\tilde{\theta}_\tau)$ is observed by the sensor. There are 2 cases:

Case 1: If $|\mathfrak{I}(\tilde{\theta}_{\tau^*})| < m^*$ (see Figure 6.10), then

$$\tilde{\theta}_{\tau^*} = (k, t^\bullet, j_0^\bullet, \dots, j_n^\bullet, \underbrace{0, \dots, 0}_{d-1}, j_{t_0-t}, \dots, j_m).$$

Thus the track hypothesis auxiliary variable is $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}$. The set of all proposed track hypothesis auxiliary variables for the Birth Merge move ($m = 14$) in this case is then

$$\begin{aligned} \mathbf{BM}^1 = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}; \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, t_0 > 1, \\ & \tilde{\theta}_{\tau^*} = (k, t^\bullet, j_0^\bullet, \dots, j_n^\bullet, \underbrace{0, \dots, 0}_{d-1}, j_{t_0-t}, \dots, j_m), d \geq 1, z_{j_0^\bullet} \in \Lambda_{t^\bullet}(\omega), \\ & z_{j_l^\bullet} \in \Lambda_{t^\bullet+l}(\omega) \text{ or } j_l^\bullet = 0, l = 1, \dots, n; \max\{i : j_i > 0, i < t_0 - t\} < m^* - 1\}. \end{aligned}$$

Case 2: If $|\mathfrak{I}(\tilde{\theta}_{\tau^*})| \geq m^*$ (see Figure 6.8), then

$$\tilde{\theta}_{\tau^*} = (K_{\omega^*}, t^\bullet, j_0^\bullet, \dots, j_n^\bullet, j_{t_0-t}, \dots, j_m).$$

Thus the track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}$, and the set of all proposed track hypothesis auxiliary variables for the Birth Merge move ($m = 14$) in this case is

$$\begin{aligned} \mathbf{BM}^2 = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}; \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, \\ & \tilde{\theta}_{\tau^*} = (k, t^\bullet, j_0^\bullet, \dots, j_n^\bullet, j_{t_0-t}, \dots, j_m), \tilde{\theta}_{\tau'^*} = (k, t, j_0, \dots, j_{t_1-t}) \\ & z_{j_0^\bullet} \in \Lambda_{t^\bullet}(\omega), z_{j_l^\bullet} \in \Lambda_{t^\bullet+l}(\omega) \text{ or } j_l^\bullet = 0, l = 1, \dots, n; \\ & t_1 = \max\{i : j_{i-t} > 0\}, t_1 - t \geq m^* - 1\}. \end{aligned}$$

The set of all proposed track hypothesis auxiliary variables for the Birth Merge move ($m = 14$) is

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 14) = \mathbf{BM}^1 \cup \mathbf{BM}^2.$$

Delete Split move:

As mentioned in point 2 of this Subsection, the Delete Split move delete some measurement indices of a track auxiliary variable $\tilde{\theta}_{\tau^*}$ before dividing it into two track auxiliary variables $\tilde{\theta}_{\tau}$ and $\tilde{\theta}_{\tau'}$ if the duration time of the target $\mathcal{L}(\tilde{\theta}_{\tau^*})$ is larger than $2m^*$ with the same conditions as for Split move. Thus the set of all proposed track hypothesis auxiliary variables for the Delete Split move ($m = 13$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_{\omega}, \tilde{Z}, 13) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau^*}\}) \cup \{\tilde{\theta}_{\tau}, \tilde{\theta}_{\tau'}\}; \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_{\omega}, \\ & m \geq 2m^* - 1; \tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_{t_1-t}), \tilde{\theta}_{\tau'} = (K_{\omega^*}, t_2, j_{t_2-t}, \dots, j_m), \\ & t_1 \in \{t + m^* - 1, \dots, t + m - m^*\}, j_{t_1-t} > 0, \\ & t_2 \in \{t_1 + 2, \dots, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^* + 1, j_{t_2-t} > 0\}\}. \end{aligned}$$

6. Backward Extension move and Backward Reduction move:

Backward Extension move: The Backward Extension move is proposed for a track auxiliary variable $\tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_{\omega}$ where the first existing time scan of the target $\mathcal{L}(\tilde{\theta}_{\tau})$ is not the first scan of the sensor (i.e. $\mathfrak{T}_0(\tilde{\theta}_{\tau}) > 1$) and the first measurement generated from the target $\mathcal{L}(\tilde{\theta}_{\tau})$ is in the neighborhood of a measurement $z'_{j_l} \in Z_{\mathfrak{T}_0(\tilde{\theta}_{\tau})-d} - \mathbf{Z}_{\mathfrak{T}_0(\tilde{\theta}_{\tau})-d}(\omega)$ at time scan $\mathfrak{T}_0(\tilde{\theta}_{\tau}) - d \geq 1$, $d \in \mathbf{d}$.

This move is a special situation of case 1 in a Birth Merge move when $t^{\bullet} + n + 1 = t$ (see Figure 6.12). Specifically, a Birth Merge move is applied to a track auxiliary variable $\tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m)$ to propose the track auxiliary variable

$$\tilde{\theta}_{\tau^*} = (k, t - d, j_0^{\bullet}, \dots, j_{d-1}^{\bullet}, j_0, \dots, j_m)$$

where $A = (k, t - d, j_0^{\bullet}, \dots, j_{d-1}^{\bullet}, j_0)$ is proposed in the same way as in the Birth move. Then the track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}\}) \cup \{\tilde{\theta}_{\tau^*}\}$. The set of all proposed track hypothesis auxiliary variables for the Backward Extension move ($m = 3$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_{\omega}, \tilde{Z}, 3) = & \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega} - \{\tilde{\theta}_{\tau}\}) \cup \{\tilde{\theta}_{\tau^*}\}, \tilde{\theta}_{\tau} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_{\omega}, t > 1; \\ & \tilde{\theta}_{\tau^*} = (k, t - d, j_0^{\bullet}, \dots, j_{d-1}^{\bullet}, j_0, \dots, j_m), t - d \geq 1, \\ & z_{j_0^{\bullet}} \in \Lambda_{t-d}(\omega), z_{j_l^{\bullet}} \in \Lambda_{t-d+l}(\omega) \text{ or } j_l^{\bullet} = 0, l = 1, \dots, d - 1\}. \end{aligned}$$

Backward Reduction move: The backward reduction move is proposed for track auxiliary variables $\tilde{\theta}_{\tau^*}$ whose duration of existence is larger than m^* ($|\mathfrak{T}(\tilde{\theta}_{\tau^*})| > m^*$) to form a track auxiliary variable $\tilde{\theta}_{\tau}$ with $|\mathfrak{T}(\tilde{\theta}_{\tau})| \geq m^*$. The time index t_0 for the Backward Reduction is chosen in the set $\{\mathfrak{T}_0(\tilde{\theta}_{\tau^*}) + 1, \mathfrak{T}_f(\tilde{\theta}_{\tau^*}) - m^* + 1\}$ such that $\mathfrak{T}_{t_0}(\tilde{\theta}_{\tau^*}) > 0$. Then the track auxiliary variable $\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m)$ is shorten by removing the first $t_0 - \mathfrak{T}_0(\tilde{\theta}_{\tau^*})$ measurement indices from track auxiliary variable $\tilde{\theta}_{\tau^*}$ to form the track aux-

iliary variable

$$\tilde{\theta}_\tau = (k, t_0, j_{t_0-t}, \dots, j_m).$$

Thus the proposed track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}) \cup \{\tilde{\theta}_\tau\}$. Then the set of all proposed track hypothesis auxiliary variables for the Backward Reduction move ($m = 9$) is

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 9) = \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_{\tau^*}\}) \cup \{\tilde{\theta}_\tau\}; \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, m \geq m^*; \tilde{\theta}_\tau = (k, t_0, j_{t_0-t}, \dots, j_m), m - t_0 - t \geq m^* - 1\}.$$

7. Update move and Point Update move:

Update move: The Update move is applied to track auxiliary variable

$$\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega$$

at the time index $t_0 \in \mathfrak{T}(\tilde{\theta}_\tau)$ to form a new track auxiliary variable

$$\begin{aligned} \tilde{\theta}_{\tau^*} &= (k, t, j_0, \dots, j_{t_0-t-1}, j_0^\bullet, \dots, j_n^\bullet), & \text{if } t_0 > t \\ \tilde{\theta}_{\tau^*} &= (k, t, j_0^\bullet, \dots, j_n^\bullet), & \text{if } t_0 = t \end{aligned}$$

where the update $A_s = (k, t_0 - 1, j_{t_0-t-1}, j_0^\bullet, \dots, j_n^\bullet)$ is constructed the same as in the Extension move if $t_0 > t$ or $A_s = (k, t_0, j_{t_0-t-1}, j_0^\bullet, \dots, j_n^\bullet)$ is constructed the same as in the Birth move if $t_0 = t$. Thus the proposed track hypothesis auxiliary variable is $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}$. Then the set of all proposed track hypothesis auxiliary variables for the Update move ($m = 10$) is

$$\begin{aligned} \mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 10) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}, \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega, \\ & l \in \{0, \dots, m\}, \tilde{\theta}_{\tau^*} = (k, t, j_0^*, \dots, j_n^*); \exists s, l < s \leq m : j_s^* \neq j_s; \\ & j_0^* > 0, j_i^* = j_i, i = 0, \dots, l, \text{ if } l > 0, n \geq m^* - 1; j_r^* = 0 \text{ or} \\ & z_{j_r^*} \in \mathbf{\Lambda}_{t+r}(\omega) \cup \{z_{\mathfrak{I}_{t+r}(\tilde{\theta}_\tau)} : \mathfrak{I}_{t+r}(\tilde{\theta}_\tau) > 0\}, r = l + 1, \dots, n\}. \end{aligned}$$

Point Update move: The Point Update move is applied to the track auxiliary variable $\tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega$ at the time index $t_0 \in \mathfrak{T}(\tilde{\theta}_\tau)$ to form a new track auxiliary variable $\tilde{\theta}_{\tau^*}$ as follows.

If t_0 is not the first existing time of the target $\mathcal{L}(\tilde{\theta}_\tau)$, let t_1 be the latest time scan before t_0 at which the target $\mathcal{L}(\tilde{\theta}_\tau)$ is observed by the sensor i.e. $t_1 = \max\{i \in \mathfrak{T}(\tilde{\theta}_\tau) : i < t_0, j_{t_0-i} > 0\}$ and generates measurement $z_{j_{t_1-t}} \in Z_{t_1}$. Let $d_1 = t_0 - t_1$. If t_0 is not the last existing time of the target $\mathcal{L}(\tilde{\theta}_\tau)$, then let t_2 be the earliest time scan after t_0 at which the target $\mathcal{L}(\tilde{\theta}_\tau)$ is observed by the sensor i.e. $t_2 = \min\{i \in \mathfrak{T}(\tilde{\theta}_\tau) : i > t_0, j_{i-t_0} > 0\}$ and generates measurement $z_{j_{t_2-t}} \in Z_{t_2}$. Let $d_2 = t_2 - t_0$ and $d_0 = t_2 - t_1$.

Thus, a proposed track auxiliary variable $\tilde{\theta}_{\tau^*}$ is formed as follows

$$\tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_{t_0-t-1}, j^\bullet, j_{t_0-t+1}, \dots, j_m)$$

where $j^\bullet \neq j_{t_0-t}$ is chosen from one of the following situations.

- If t_0 is not in the first scan (see Figure 6.14 for illustration), we choose either
 - $j^\bullet = 0$ if t_0 is not the last exiting time of the target $\mathcal{L}(\tilde{\theta}_\tau)$ and the next target-generated measurement $z_{j_{t_2-t}}$ of target $\mathcal{L}(\tilde{\theta}_\tau)$ is in the d_0 -neighborhood of the previous target-generated measurements $z_{j_{t_1-t}}$ (see Figure 6.14b) i.e. $j^\bullet = 0$ if $t_0 < \mathfrak{T}_f(\tilde{\theta}_\tau)$ and $z_{j_{t_2-t}} \in L_{d_0}(z_{j_{t_1-t}}, t_1)$, $d_0 \in \mathbf{d}$; or
 - $j^\bullet > 0$ if the measurement z_{j^\bullet} at time t_0 is a d_1 -neighbor of the previous target-generated measurement $z_{j_{t_1-t}} \in Z_{t_1}$ and provided t_0 is not the last exiting time of the track $\mathcal{L}(\tilde{\theta}_\tau)$, the next target-generated measurement $z_{j_{t_2-t}} \in Z_{t_2}$ is a d_2 -neighbor of measurement $z_{j^\bullet} \in Z_{t_0}$ (see Figure 6.14a) i.e. $j^\bullet > 0$ if $z_{j^\bullet} \in L_{d_1}(z_{j_{t_1-t}}, t_1)$ and if $t_0 < \mathfrak{T}_f(\tilde{\theta}_\tau)$, $z_{j_{t_2-t}} \in L_{d_2}(z_{j^\bullet}, t_0)$, $d_2 \in \mathbf{d}$.
- If t_0 is in the first scan, $j^\bullet > 0$ is chosen such that the next target-generated measurement $z_{j_{t_2-t}}$ is a d_2 -neighbor of the measurement $z_{j^\bullet} \in Z_{t_0}$ i.e. $z_{j_{t_2-t}} \in L_{d_2}(z_{j^\bullet}, t_0)$ (see Figure 6.15 for illustration).

Then the track hypothesis $\tilde{\theta}_{\omega^*}$ is proposed as follows.

At time t_0 , if j^\bullet is chosen either as zero or $\Lambda_{t_0}(\omega)$ then $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}$. The set of proposed track hypothesis auxiliary variables for the Point Update move ($m = 11$) in this case is

$$\begin{aligned} \text{PU}^1(\tilde{\theta}_\omega, \tilde{Z}) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_\omega - \{\tilde{\theta}_\tau\}) \cup \{\tilde{\theta}_{\tau^*}\}; \tilde{\theta}_\tau = (k, t, j_0, \dots, j_m) \in \tilde{\theta}_\omega; \\ &\quad \tilde{\theta}_{\tau^*} = (k, t, j_0, \dots, j_{t_0-t-1}, j^\bullet, j_{t_0-t+1}, \dots, j_m), j^\bullet \neq j_{t_0-t}, t_0 \in \mathfrak{T}(\tilde{\theta}_\tau); \\ &\quad j^\bullet = 0 \text{ or } j^\bullet \in \Lambda_{t_0}(\omega)\}. \end{aligned}$$

Otherwise j^\bullet has already been assigned to a track. Let $\tilde{\theta}_{\tau'} = (k', t', j'_0, \dots, j'_n) \in \tilde{\theta}_\omega$ be the track auxiliary variable such that $\mathfrak{T}_{t_0}(\tilde{\theta}_{\tau'}) = j^\bullet = j'_{t_0-t'}$. Then $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ exchange their measurement indices at time t_0 as follows (see Figure 6.17).

If t_0 is not the first existing time of the target $\mathcal{L}(\tilde{\theta}_{\tau'})$, let t'_1 be the latest time before t_0 at which the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is observed by the sensor (i.e. $j'_{t'_1-t'} > 0$) and generates measurement $z_{j'_{t'_1-t'}}$. Let $d'_1 = t_0 - t'_1$. If t_0 is not the last existing time of the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ (i.e. $t_0 < \mathfrak{T}_f(\tilde{\theta}_{\tau'})$), let t'_2 be the earliest time after t_0 at which the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ is observed by the sensor and generates measurement $z_{j'_{t'_2-t'}}$. Let $d'_2 = t'_2 - t_0$ and $d'_0 = t'_2 - t'_1$. The measurement indices can be exchanged between the targets $\mathcal{L}(\tau)$ and $\mathcal{L}(\tau')$ at time t_0 to form

$$\begin{aligned} \tilde{\theta}_{\tau^*} &= (k, t, j_0, \dots, j_{t_0-t-1}, j'_{t_0-t'}, j_{t_0-t+1}, \dots, j_m), \\ \tilde{\theta}_{\tau'^*} &= (k', t', j'_0, \dots, j'_{t'_0-t'-1}, j_{t_0-t}, j'_{t'_0-t'+1}, \dots, j'_n). \end{aligned}$$

(see Figures 6.17, 6.18 and 6.19) provided that one of the following conditions holds:

- (a) The followings hold a) The target $\mathcal{L}(\tilde{\theta}_\tau)$ is not observed by the sensor ($j_{t_0-t} = 0$); b) the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ both exists before time index t_0 (i.e. $\mathfrak{X}_0(\tilde{\theta}_{\tau'}) < t_0$) and after time index t_0 (i.e. $t_0 < \mathfrak{X}_f(\tilde{\theta}_{\tau'})$); and c) $d'_0 = t'_2 - t'_1 \in \mathbf{d}$ and $z_{j'_{t'_2-t'}}$ is a d'_0 -neighbor of $z_{j'_{t'_1-t'}}$ (i.e. $z_{j'_{t'_2-t'}} \in L_{d'_0}(z_{j'_{t'_1-t'}}, j'_1)$) (see Figure 6.17b).
- (b) The target $\mathcal{L}(\tilde{\theta}_\tau)$ is detected by the sensor (i.e. $j_{t_0-t} > 0$) and if the initial target of $\mathcal{L}(\tilde{\theta}_{\tau'})$ is equal to t_0 (see Figure 6.18); and $z_{j'_{t'_2-t'}}$ is a neighbor of $z_{j_{t_0-t}}$ at time t_0 (i.e. $z_{j'_{t'_2-t'}} \in L_{d'_2}(z_{j_{t_0-t}}, t_0)$) (see Figure 6.18 for illustration).
- (c) The target $\mathcal{L}(\tilde{\theta}_\tau)$ is detected by the sensor (i.e. $j_{t_0-t} > 0$) and if the final time of the track $\mathcal{L}(\tau')$ is equal to t_0 and $z_{j_{t_0-t}}$ is a d'_1 -neighbor of $z_{j'_{t'_1-t'}}$ (i.e. $z_{j_{t_0-t}} \in L_{d'_1}(z_{j'_{t'_1-t'}}, t'_1)$) (see Figure 6.19 for illustration).
- (d) The target $\mathcal{L}(\tilde{\theta}_\tau)$ is detected by the sensor (i.e. $j_{t_0-t} > 0$) and if the target $\mathcal{L}(\tilde{\theta}_{\tau'})$ both exists after time t_0 (i.e. $\mathfrak{X}_f(\tilde{\theta}_{\tau'}) = t_0$) and before time t_0 (i.e. $\mathfrak{X}_0(\tilde{\theta}_{\tau'}) < t_0$); $z_{j_{t_0-t}}$ at time t_0 is a d'_1 -neighbor of $z_{j'_{t'_1-t'}}$ (i.e. $z_{j_{t_0-t}} \in L_{d'_1}(z_{j'_{t'_1-t'}}, t'_1)$); and $z_{j'_{t'_2-t'}}$ is a d'_2 -neighbor of $z_{j_{t_0-t}}$ (i.e. $z_{j'_{t'_2-t'}} \in L_{d'_2}(z_{j_{t_0-t}}, t_0)$) (see Figure 6.17a for illustration).

Then the proposed track hypothesis auxiliary variable for updating the track auxiliary variable $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ is $\tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega^*} - \{\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}$.

The set of all proposed track hypothesis auxiliary variables for the Point Update move ($m = 11$) for this case is

$$\begin{aligned} \mathbf{PU}^2(\tilde{\theta}_\omega, \tilde{Z}) &= \{\tilde{\theta}_{\omega^*} : \tilde{\theta}_{\omega^*} = (\tilde{\theta}_{\omega^*} - \{\tilde{\theta}_\tau, \tilde{\theta}_{\tau'}\}) \cup \{\tilde{\theta}_{\tau^*}, \tilde{\theta}_{\tau'^*}\}; \\ \tilde{\theta}_{\tau^*} &= (k, t, j_0, \dots, j_{t_0-t-1}, j^\bullet, j_{t_0-t+1}, \dots, j_m), \\ \tilde{\theta}_{\tau'^*} &= (k', t', j'_0, \dots, j'_{t'_0-t'-1}, j'_{t_0-t}, j'_{t'_0-t'+1}, \dots, j'_n). \} \end{aligned} \quad (6.51)$$

The set of all proposed track hypothesis auxiliary variables for Point Update move ($m = 11$) is

$$\mathbf{P}(\tilde{\theta}_\omega, \tilde{Z}, 11) = \mathbf{PU}^1(\tilde{\theta}_\omega, \tilde{Z}) \cup \mathbf{PU}^2(\tilde{\theta}_\omega, \tilde{Z}).$$

6.3.2.5 Property of the Markov chain

These fourteen proposal moves were constructed to generate samples conditional on \tilde{Z} and $\tilde{\theta}_\omega$. A sample $\tilde{\theta}_{\omega^*} \sim q(\cdot | \tilde{\theta}_\omega, \tilde{Z})$ is one of the proposal moves constructed in Section 6.3.2.4. After we have obtained a track hypothesis auxiliary variable $\tilde{\theta}_{\omega^*}$ using a MC constructed in the previous section, a sequence of augmented auxiliary variables $\tilde{\theta}^*$ corresponding to $\tilde{\theta}_{\omega^*}$ can be obtained using (6.37). We denote the corresponding distribution of $\tilde{\theta}^*$ by $q(\cdot | \tilde{\theta}, \tilde{Z})$. By the construction of the proposal moves, $\tilde{\theta}_{\omega^*}$ specifies a track hypothesis ω^* and $\tilde{\theta}^*$ is the corresponding sequence of auxiliary variables of $\tilde{\mathbf{X}}_{1:T}(\omega^*)$. Hence whenever $\tilde{X} \sim q(\cdot | \tilde{Z}, \tilde{\theta}^*)$, there exist some track hypothesis ω^* such that $\tilde{X} = \tilde{\mathbf{X}}_{1:T}(\omega^*)$.

In practice, there are a large number of track hypothesis auxiliary variables and some of them do not represent possible associations between measurements and true targets. Thus, reducing the size of $\mathbf{P}(\omega, \tilde{Z})$ is very important and depends on the system model and the birth locations. Another issue is that the computation of all track hypothesis auxiliary variables is very time consuming for the problems with dense clutter and targets, so reusing proposal moves not applied in the previous time steps is an option for reducing the computations. Another option is to use culling as described in Section 6.3.2.4 on page 135.

Proposition 6.1: *Assume that the moves for the proposal distribution $q(\tilde{\theta}_{\omega^*} | \tilde{\theta}_{\omega}, \tilde{Z})$ have been constructed as in Subsections 6.3.2.2 and 6.3.2.4 and that there exists an aperiodic state of a MC. Then the MC generated from $q(\tilde{\theta}_{\omega^*} | \tilde{Z}, \tilde{\theta}_{\omega})$ is ergodic.*

Proof. The MC which is generated from the proposal moves is irreducible because any two states can be connected through a series of birth and death moves. Thus starting from $\tilde{\theta}_{\omega} \in \Theta_{\omega}$ the MC can reach any $\tilde{\theta}_{\omega^*} \in \Theta_{\omega}$.

By Theorem 4.3 and assumption that there exists an aperiodic state, the irreducible MC is aperiodic.

Furthermore, the space Θ_{ω} is finite so by Theorem 4.5, the irreducible and aperiodic MC on the space Θ_{ω} is positive recurrent and then by Theorem 4.6 the MC is ergodic. \square

The assumption that there exists an aperiodic state is very mild. It is easily satisfied as shown in the following example in Figure 6.20.

The MC return to state a in 2 time steps with a Death move to a state c followed by a Birth move back to state a . It can also return in 3 time steps with a Death move to state c followed by a Birth move to state b and followed a Reduction move back to state a .

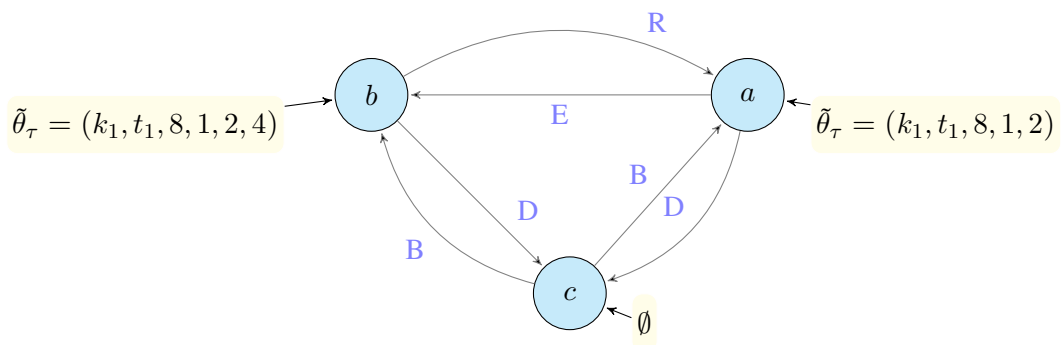


Figure 6.20: Example of aperiodic state

After a new track hypothesis auxiliary variable $\tilde{\theta}_{1:T}^*$ has been obtained from the proposal distribution $q(\tilde{\theta}_{1:T}^* | \tilde{Z}_{1:T}, \tilde{\theta}_{1:T})$, $\tilde{X}_{1:T}^*$ can be sampled from $p(\cdot | \tilde{Z}_{1:T}, \tilde{\theta}_{1:T}^*)$ in (6.27) using the SMC Algorithm 11. The PMMH algorithm for RFS based Multi-target Tracking is described in the following subsection by combining these two sampling techniques.

6.3.3 PMMH Algorithm for RFS based Multi-target tracking

Initializing $\tilde{\theta}$ arbitrarily in Algorithm 12 makes the computation expensive. This can be alleviated by using an estimate from the Gaussian Mixture Probability Hypothesis Density (GM-PHD) tracker as the initial estimate. Using a good estimate from this popular technique may reduce the computational cost significantly. We also keep the estimate \tilde{X}^G from the GM-PHD tracker such that the GM-PHD only need to sample $N - 1$ instead of N samples from $q(\cdot|\tilde{Z}, \tilde{\theta})$. The SMC modified to suit this situation is called the conditional SMC [4].

The pseudocode for the SMC Algorithm 13 below provides us with the parameters $B_{1:T}^n$ as the ancestral lineage of the particle $\tilde{X}_{1:T}^n$. The SMC algorithm conditional on $\tilde{X}_{1:T}^k = (\tilde{X}_1^{B_1^k}, \dots, \tilde{X}_T^{B_T^k})$ described in Algorithm 13 samples $N - 1$ particles.

Algorithm 13 : Conditional SMC Algorithm

At time $t = 1$:

- if $n \neq B_1^k$, sample $\tilde{X}_1^n \sim q(\cdot|\tilde{Z}_1, \tilde{\theta}_1)$ and compute $w_1(\tilde{X}_1^n)$ by using (6.30) and normalize $W_1^n \propto w_1(\tilde{X}_1^n)$.

At $t = 2, \dots, T$:

- if $n \neq B_t^k$, sample $A_{t-1}^n \sim \mathfrak{F}(\cdot|\mathbf{W}_{t-1})$,
 - then sample $\tilde{X}_t^n \sim q(\cdot|\tilde{X}_{t-1}^{A_{t-1}^n}, \tilde{Z}_t, \tilde{\theta}_t)$, set $\tilde{X}_{1:t}^n = (\tilde{X}_{1:t-1}^{A_{t-1}^n}, \tilde{X}_t^n)$ and
 - compute $w_t(\tilde{X}_{1:t}^n)$ by using (6.31) and normalize $W_t^n \propto w_t(\tilde{X}_{1:t}^n)$.
-

Based on the PMMH Algorithm 12, the algorithm of the PMMH for MTT is summarized in Algorithm 14 below.

Algorithm 14 : PMMH Algorithm for MTT

Input: Given \tilde{Z} , p_{S_t} , p_{D_t} , κ_t , the birth intensity γ_t for $t = 1, \dots, T$ and sample number L .

Output: $S_X(l)$, $S_{\tilde{\theta}}(l)$, and $\gamma_{\theta}(l)$ for $l = 1, \dots, L$.

At iteration $l = 1$

- Run GM-PHD tracker to obtain \tilde{X}^G , then obtains $\tilde{\theta}(l)$ from \tilde{X}^G and denote $B_{1:T} = \underbrace{(1, \dots, 1)}_T$.
- Run a conditional SMC algorithm targeting $p(\tilde{X}|\tilde{Z}, \tilde{\theta}(l))$ conditional on \tilde{X}^G and $B_{1:T}$. Then sample $\tilde{X}^* \sim \hat{p}(\cdot|\tilde{Z}, \tilde{\theta}(l))$ and calculate $\gamma_{\theta}(l) = \hat{p}(\tilde{Z}, \tilde{\theta}(l))$. Then denote $S_X(l) = \tilde{X}^*$.

At iteration $l > 1$

- Propose $\tilde{\theta}^* \sim \mathbf{q}(\cdot|\tilde{\theta}(l-1), \tilde{Z})$ (see Subsection 6.3.2.4)
- Run an SMC algorithm targeting $p(\tilde{X}|\tilde{Z}, \tilde{\theta}^*)$. Then sample $\tilde{X}^* \sim \hat{p}(\cdot|\tilde{Z}, \tilde{\theta}^*)$; calculate $\hat{p}(\tilde{Z}, \tilde{\theta}^*)$ and the probability

$$\alpha = \min \left\{ 1, \frac{\hat{p}(\tilde{Z}, \tilde{\theta}^*)w(\tilde{\theta}^*)\mathbf{q}(\tilde{\theta}(l-1)|\tilde{\theta}^*, \tilde{Z})}{\gamma_{\theta}(l-1)w(\tilde{\theta})\mathbf{q}(\tilde{\theta}^*|\tilde{\theta}(l-1), \tilde{Z})} \right\}$$

if $\alpha \geq u$, set $S_X(l) = \tilde{X}^*$, $\gamma_{\theta}(l) = \hat{p}(\tilde{Z}, \tilde{\theta}^*)$. Otherwise $S_X(l) = S_X(l-1)$, $\tilde{\theta}(l) = \tilde{\theta}(l-1)$, $\gamma_{\theta}(l) = \gamma_{\theta}(l-1)$ where $u \sim \text{Unif}[0, 1]$.

6.4 Summary and Discussion

The MTT problem was formulated in a random finite set framework. Particularly, a track was defined as a trajectory of target states equipped with a target label and the appearance time; a track hypothesis was also defined as a set of different tracks such that no two different tracks share any states at any time. Furthermore, augmented multi-target states were formulated as a collection of augmented single target states such that each augmented single target state is extended from single target state by adding a target label. With the augmented multi-target states formulated in a RFS framework, the posterior distribution of a sequence of augmented multi-target states was derived via Bayes recursive framework. Furthermore, we also showed that conditional on a sequence of noisy multi-target measurement, the posterior distribution of a track hypothesis is the same as the posterior distribution of its corresponding sequence of augmented multi-target states.

There is no-closed form expression for the posterior distribution so numerical methods such as MCMC are the only feasible option. However, directly applying this method is computationally intractable when the number of targets and measurements are large because the likelihood function in the posterior distribution considers all combinations between the target states and the measurements at a time instance. An auxiliary variable at a time instance was introduced to overcome this problem by mapping the target labels to the measurement indices. Any target label mapped to a 0 is undetected. Each auxiliary variable represents a combination between target states and measurements. The augmented auxiliary variable was subsequently established to show the correlation between the augmented target states and the measurements at a time instance. Thus, a sequence of augmented auxiliary variables was derived to capture the relationship between the sequence of augmented multi-target states and the sequence of the multi-target measurements.

A new algorithm, the PMMH algorithm for RFS based Multi-target Tracking which is a combination of PMMH [4] and the proposal moves in Section 6.3.2.2, was derived to numerically solve for the joint distribution $p(\tilde{X}, \tilde{\theta}|\tilde{Z})$. In the next chapter we will illustrate the PMMH algorithm for RFS based Multi-target Tracking in a simulation example.

Chapter 7

Simulation and Performance

The PMMH algorithm for RFS based Multi-target tracking is simulated and evaluated in this chapter. The multi-object metric for evaluating the performance of the algorithm is discussed in Section 7.1. This metric is called Optimal Subpattern Assignment (OSPA). Simulation results and performance evaluation of the PMMH algorithm for RFS based Multi-target Tracking are given in section 7.2. The results are discussed in Section 7.3

7.1 Multi-object Miss-distance

Let X and Y be two finite sets where $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$ and assume that $m < n$. The set X with smaller cardinality is initially chosen as a reference. We want to determine the assignment between the m points of X and the n points of Y that minimizes the sum of distances, subject to the constraint that distances are capped at a preselected maximum or cut-off value c . This minimum sum of distances can be interpreted as the total localization error, which is assigned to the points in Y by giving the points in X as reference. All points which remain unassigned are charged with c the maximum error value. These errors can be interpreted as cardinality errors which are penalized at the maximum rate. The total error committed is then the sum of the localization error and the cardinality error. Remarkably, the per target error obtained by normalizing total error by n (the largest cardinality of the two given sets) is a proper metric [155].

The OSPA metric $\bar{d}_p^{(c)}$ is defined as follows. Let $\bar{d}^{(c)}(x, y) := \min(c, \|x - y\|)$ for $x, y \in \mathcal{X}$, and Π_k denotes the set of permutations on $\{1, 2, \dots, k\}$ for any positive integer k . Then, for $p \geq 1, c > 0$, and $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$,

- if $m \leq n$:

$$\bar{d}_p^{(c)}(X, Y) := \left[\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^n \bar{d}^{(c)}(x_i, y_{\pi(i)})^p + c^p(n - m) \right) \right]^{\frac{1}{p}}$$

- if $m > n$: $\bar{d}_p^{(c)}(X, Y) := \bar{d}_p^{(c)}(Y, X)$; and
- if $m = n = 0$: $\bar{d}_p^{(c)}(X, Y) := \bar{d}_p^{(c)}(Y, X) = 0$

The OSPA distance is interpreted as a p -th order per-target error, comprised of a p -th order per-target localization error and a p -th order per-target cardinality error. Precisely, for $p < \infty$ these components are given by

- if $m \leq n$:

$$\bar{e}_{p,\text{loc}}^{(c)}(X, Y) := \left(\frac{1}{n} \min_{\pi \in \Pi_n} \sum_{i=1}^n \bar{d}^{(c)}(x_i, y_{\pi(i)})^p \right)^{\frac{1}{p}} \quad \bar{e}_{p,\text{card}}^{(c)}(X, Y) := \left(\frac{c^p(n-m)}{n} \right)^{\frac{1}{p}}$$
- if $m > n$:

$$\bar{e}_{p,\text{loc}}^{(c)}(X, Y) = \bar{e}_{p,\text{loc}}^{(c)}(Y, X), \quad \bar{e}_{p,\text{card}}^{(c)}(X, Y) = \bar{e}_{p,\text{card}}^{(c)}(Y, X)$$

They can thus be interpreted as contributions due to localization only (within the optimal subpattern assignment) and cardinality only (penalized at maximal distance). The decomposition of the OSPA metric into separate components is usually not necessary for performance evaluation, but may provide valuable additional information.

The order parameter p determines the sensitivity of the metric to outliers, and the cut-off parameter c determines the relative weighting of the penalties assigned to cardinality and localization errors. When $p = 1$, the OSPA distance can be interpreted exactly as the sum of the "per-target localization error" and the "per-target cardinality error". For details see [155].

This metric is suitable for evaluating the multi-target tracking problem because at each time it considers not only the error between the number of estimated targets and the number of true targets but also the error between the position of estimated targets and the position of the true targets.

7.2 Simulation and Performance

In this section, we demonstrate the multi-target PMMH algorithm with a simulated sample and evaluate its performance using the Optimal Sub-pattern Assignment distance (OSPA) [155]. In order to apply the OSPA metric, we choose $p = 2$ and $c = \max_{t \in \mathcal{T}} \max_{x \in X_t, x' \in X_t^{\text{true}}} d(x, x')$ where X_t^{true} is the set of true multi-target. The surveillance area is the square region $\mathcal{R} = [-1000m, 1000m] \times [-1000m, 1000m]$. We use the surveillance duration of $T = 50$ scans with sampling interval $T_s = 1$ second. We denote x^{Tr} as the transpose of x . The state vector is $x_t = [\xi_t, \zeta_t, v_{\xi_t}, v_{\zeta_t}]^{Tr}$ where (ξ_t, ζ_t) denotes the target position on 2D Cartesian plane and (v_{ξ_t}, v_{ζ_t}) is its velocity $t = 1, \dots, T$. Linear state and measurement models are used

$$x_t = Ax_{t-1} + v_{t-1}, z_t = Cx_t + w_t \quad (7.1)$$

where

$$A = \begin{bmatrix} 1 & 0 & T_s & 0 \\ 0 & 1 & 0 & T_s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^{Tr},$$

v_t and w_t are zero mean Gaussian process with covariance Q and R , respectively; where

$$Q = \sigma_v^2 \begin{bmatrix} \frac{T_s^2}{4} I_2 & \frac{T_s}{2} I_2 \\ \frac{T_s}{2} I_2 & I_2 \end{bmatrix}, \quad R = \sigma_w^2 I_2$$

σ_v is the standard deviation of the velocity process noise; σ_w is the standard deviation of the measurement noise. The level of density was chosen to be moderate where the number of targets varies from 1 to 50. The parameters were chosen according to general real tracking examples. Targets move with initial speeds uniformly distributed between 30 and 150 meters per second so the maximum speed is $150(m/s)$ and \bar{v} in (6.40) is 150, $\sigma_v = 5m/s$ and $\sigma_w = 10m$. In order to demonstrate the closely space track, the targets appear from $J = 24$ possible locations and can be born at any time in these J possible locations (see Figure 7.1) with intensity

$$\gamma_t(x) = \sum_{i=1}^J \frac{1}{J} \mathcal{N}(x; m_{\gamma}^{(i)}, P_{\gamma})$$

where $P_{\gamma} = \text{diag}(Pu_m^2)$, $P = [100, 100, 25, 25]$ and $u_m^2 = u_m^T u_m$, $u_m = [m, m, \frac{m}{s}, \frac{m}{s}]$ are used to model spontaneous births in the vicinity of $m_{\gamma}^{(i)}$, $i = 1, \dots, J$. Target spawning is not considered in this example. The track is confirmed if a target exists at least 3 consecutive times so $m^* = 3$. The ground truth from 19 of these J birth locations is plotted together with false alarms in Figure 7.2. These targets moves from top right to bottom left, or from the middle to either top right or bottom left. Each target survives with probability $P_S = 0.99$ and is detected

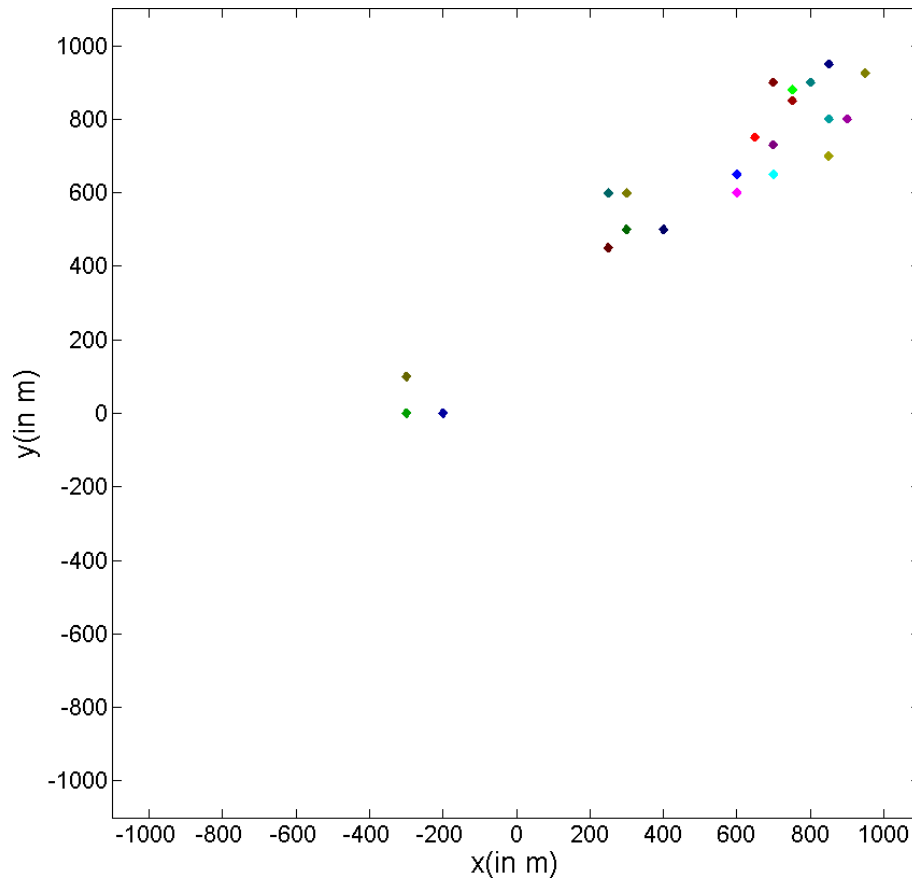


Figure 7.1: Location of the appearance of targets with mean $m_{\gamma}^{(i)}$ and $P_{\gamma}^{(i)}$, $i = 1, \dots, J$

with probability $P_D = 0.8$, and the maximum number of consecutive missed detections of any track is chosen as $\bar{d} = 2$. The detected measurements are immersed in clutter that is modeled as a

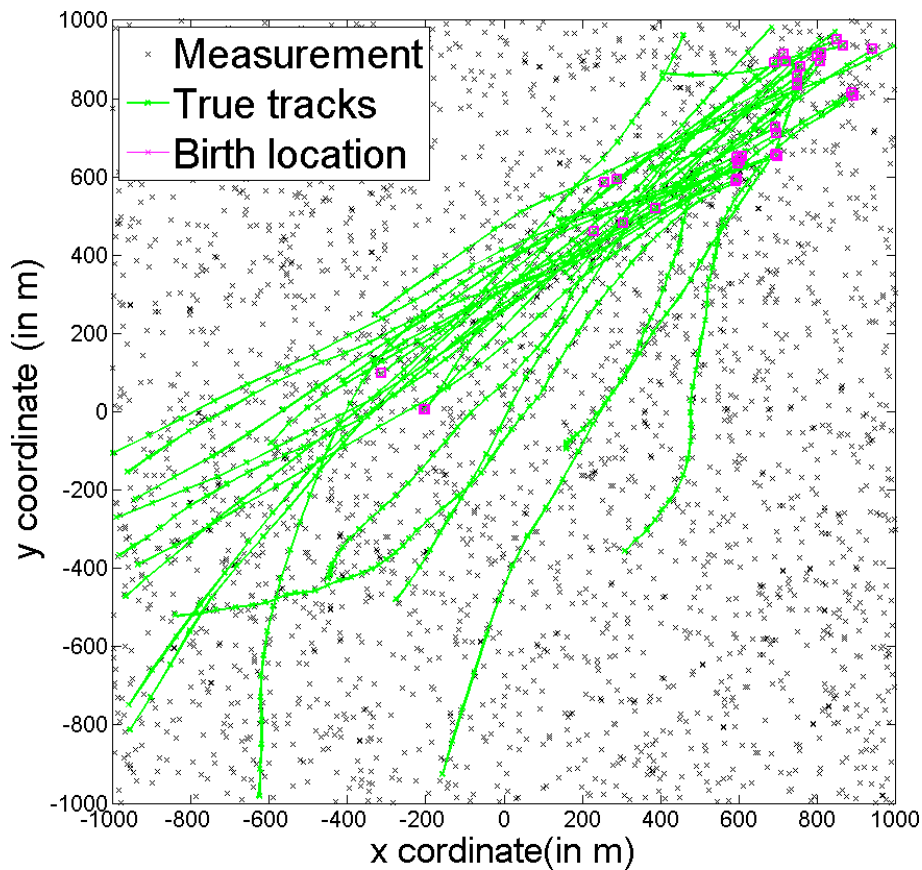


Figure 7.2: Ground truth tracks with appearance location plotted together with noisy measurements.

Poisson RFS Λ_t with intensity

$$\lambda_c = \kappa_t V u$$

where u is the uniform density over the surveillance region, $V = 4 \times 10^6 m^2$, $\kappa_t = 12.5 \times 10^{(-6)} m^{-2}$ is the intensity function and λ_c is the average number of clutter returns per unit volume (i.e. $\lambda_c = 50$ clutter returns per scan over the surveillance region \mathcal{R}). In this thesis, we modeled the clutter as a Poisson process which is general form for unpredictable clutter. We haven't simulate the scenario in which the clutter is not Poisson process however the performance may be as good as the same for other type of clutter model. In the case the clutter is unpredictable, the clutter can be estimated at each time scan as in filtering algorithm [102]. In general, the parameters can be chosen according to the underlying targets.

7.3 Numerical Result and Discussion

The problem of closely spaced and crossing targets cannot be solved reliably by popular filtering techniques. MHT algorithm is known to break down with a large number of targets and a large number of measurements. Our algorithm, PMMH for multi-target tracking, is designed to deal with this problem. In the absence of a proper estimator we use the PMMH algorithm 13 and we

stop when the last 50 accepted samples are associated with the Point Update move, indicating that the algorithm has "settled" on a fixed number of targets. The tracks from our algorithm are plotted against ground truth in Figure 7.3. The algorithm search and compare all possibilities of track hypothesis with the general Poisson assumption for clutter. The variances for parameter can be initiated large if there is uncertainty of new born target. The algorithm may not be sensitive with the choice of parameters since the samples are drawn from the important sampling distribution such that the support of this important sampling contains the support of posterior distribution. Thus the performance may not be degrade much if the measurement followed slightly different statistics to those assumed by the filter.

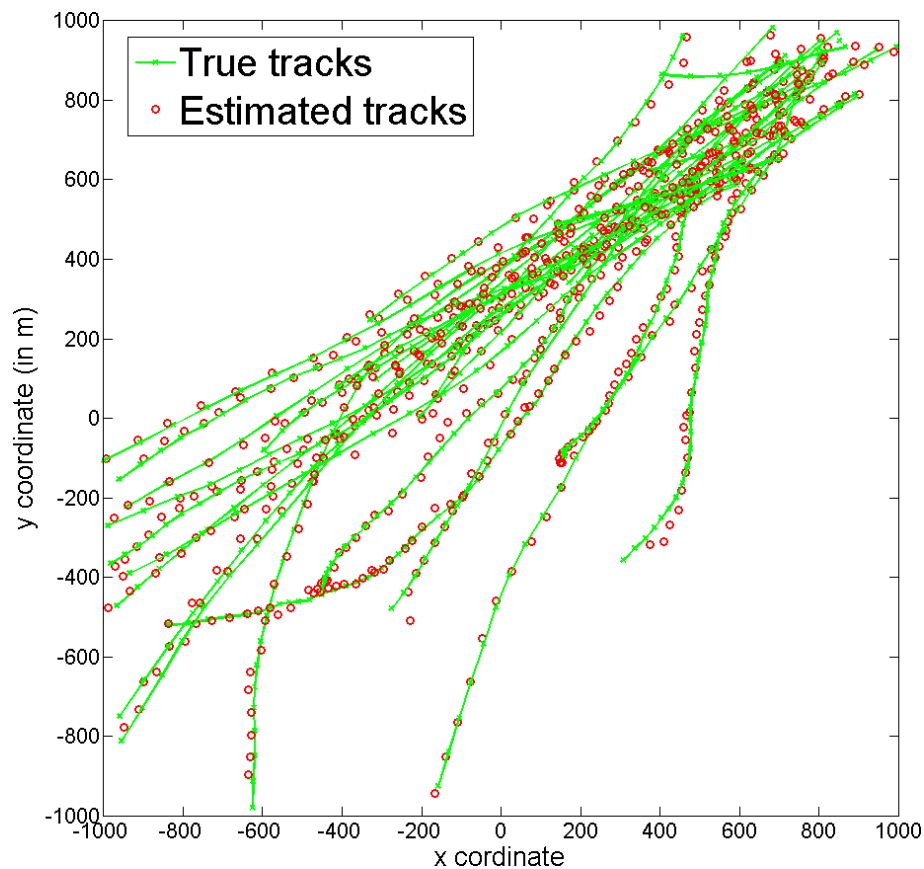


Figure 7.3: The true tracks and estimated tracks from PMMH for MTT a with GMPHD estimate as the initial state of a Markov chain.

The performance of PMMH for MTT is evaluated using the OSPA metric in Figure 7.4. In this figure, there are some large errors which occurred at six different time scan periods, more specifically $t = 1, 5, 39$, and the time intervals $9 - 10, 41 - 42$ and $47 - 50$. Figures 7.5 and 7.6 explain the origin of these errors. These errors result from the miss-detections of the targets when targets first appear or before the targets disappear from the surveillance area. Figure 7.7 shows the targets whose states were not tracked by our algorithm. The targets which are not tracked are labeled and their trajectories are drawn in dashed line with cyan color. For examples, at time $t = 1$ the targets 3 and 4 are born but not observed by the sensor. The same happens for target 10 at time

$t = 5$. At time $t = 9$ and $t = 10$, the sensor does not detect target 4 before the target disappears from the surveillance area. This is also the case for the target 10 at time $t = 39$ and $t = 40$.

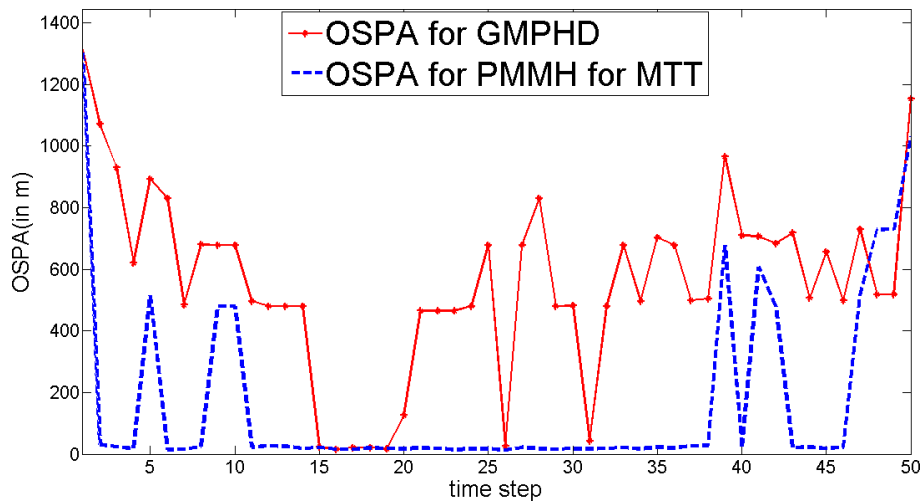


Figure 7.4: The error using estimates from a GM-PHD filter and using the estimates from PMMH for MTT.

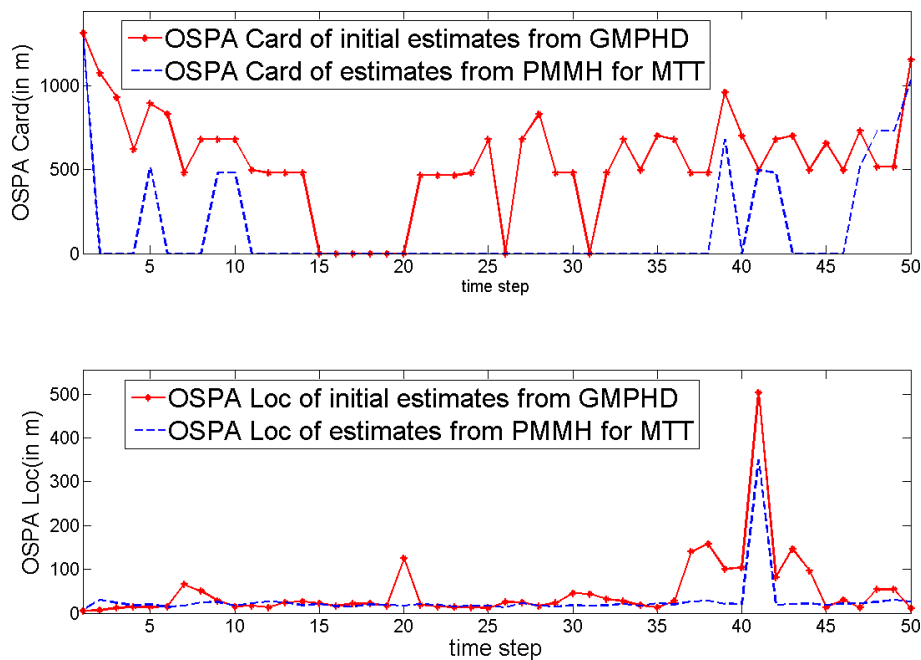


Figure 7.5: Multi-target estimation errors (cardinality error and localization error) for GMPHD and the PMMH for MTT.

The PMMH algorithm 13 confirmed a false alarm before the true appearance time of target 30 as an initial state of the target. The "OSPA Loc" in Figure 7.5 also shows that whenever targets are detected during their existence period the location error seem to be small. However, the "OSPA" in Figure 7.4 shows that there is an error during the time period between 47 and the last scan time $T = 50$. This happens because the target 33 only exists during time $t = 47$ to the last scan $T = 50$

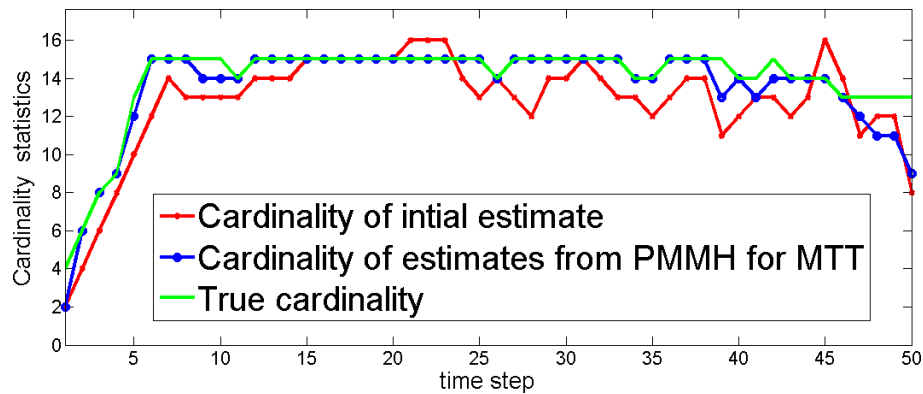


Figure 7.6: True cardinality (green line) shown versus estimated cardinality from GM-PHD filter (red line), and PMMH for MTT (blue line).

but is only detected at every second time instant from time $t = 48$. Therefore the PMMH for MTT algorithm does not have enough information to distinguish this target from clutter.

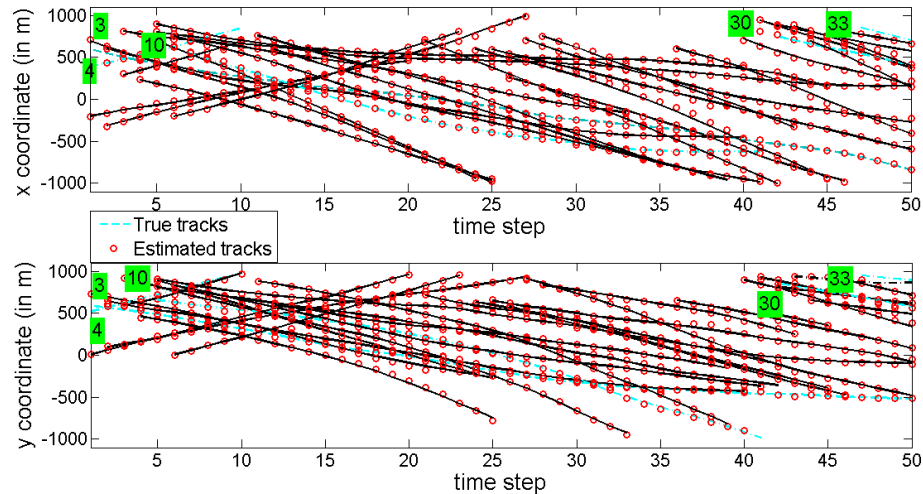


Figure 7.7: Ground truth tracks and their estimates. The states of labeled and cyan colored targets were not detected by PMMH for MTT algorithm.

7.4 Conclusion

A batch formulation and solution based on random finite sets for the MTT problem in a cluttered environment with low detection probabilities has been proposed. A simulation was successfully carried out on a moderately difficult scenario with medium probability detection ($P_D = 0.8$). The trajectories of a variable number of targets were tracked successfully. Tracking performance was reliable compared to standard filtering based MTT methods. However, the computational cost is high for the batch method.

Chapter 8

Conclusion

This Chapter closes the dissertation by giving a summary and conclusion of the contributions, and some suggestions for further researches.

8.1 Summary and Conclusion

TIn this dissertation we have considered the multi target tracking problem where many targets move close together, and may cross each other. This problem is motivated by the cell tracking problem in medicine where a large unknown number of cells move very close to each other, and they may also cross each other. In addition, they may spawn other targets or die unpredictable. The environment where the cell move may be noisy and may be heavily cluttered. Tracking the trajectories of the targets (cells) in such environment is a most difficult problem. Conceptually this problem can be formulated in a Bayesian setting, and the multi target Bayes filter can be used for estimating the target states from the observed measurements. However, due to the large number of targets and measurements the multi target Bayes filter is computationally intractable since all possible combinations of targets and measurements must be considered, and hence computationally feasible approximations must be required. Commonly used methods based on the Bayes filter such as MHT, JPDA, JIPDA, PHD filter, GM-PHD etc, all fail to varying degrees on problems of the type considered here.

In this thesis we have proposed a batch processing method for the MTT problem. The problem has been approached using the RFS framework. This is a natural framework for this type of problems since it can easily handle an unknown number of targets and measurements which in addition also vary over time. The problem has been rigorously formulated in the RFS framework and the MTT Bayes filter has been derived. In order to overcome the computational difficulties with the MTT Bayes filter an auxiliary variable which associates target labels and measurements indices at a time step has been introduced. In order to find this association between targets and measurements we have constructed a Markov Chain based on 14 proposal moves. The Bayes filter has then been approximated using a PMMH algorithm where an SMC method is combined with an MCMC method. The reason for this choice is that the variables involved are strongly correlated and SMC and MCMC on their own do not give reliable results in such cases. In the proposed approach the samples of the target states are drawn conditionally on the auxiliary variable using

an SMC algorithm. The PMMH algorithm has been implemented on a simulation example with very promising results.

As illustrated by the simulation example the proposed method is a very promising batch method for the MTT problem. The algorithm has several strengths: It is formulated in the RFS framework which is a natural framework for dealing with an unknown and time varying number of targets and measurements. Moreover the computational burden is greatly reduced by the introduction of the auxiliary variable without sacrificing accuracy. Finally the proposed PMMH algorithm for approximating the posterior distribution combines the strengths of MCMC and SMC methods thus enabling efficient sampling of strongly correlated variables. The computational cost of the algorithm is still high and it is therefore important to choose good initial estimates and proposal distributions in order to achieve fast convergence. In this thesis this has been achieved by initializing the algorithm using the estimate from the GM-PHD filter and constructing a MC based on 14 proposal moves for finding the association between targets and measurements.

Even though the results are very promising there are still many open questions and room for improvements in the algorithms. The most important ones are briefly discussed next under topics for further research.

8.2 Future Research

The main motivation for this work has been the cell tracking problem in medicine. It would therefore be of great interest to apply the developed algorithm to real data. The data are given in the form of cell images and therefore require image processing before measurements of the type considered in this thesis can be obtained. In addition to the actual application to cell data it is also of interest to develop image processing methods which takes into account that the processed data will subsequently be used in a target tracking algorithm. The results from the target tracking could also be fed back to the image processing algorithm, thus creating an integrated image processing and tracking algorithm.

The computation cost is high and finding algorithm improvements which reduce the computational burden is an important practical problem. Improvements can e.g. be sought in the areas of better initial estimates, better proposal distributions, or parallelizing the algorithm for implementation on multi-core processors.

On a more fundamental level the development of a track estimator would be a significant contribution to the field of multi-target tracking. This is a difficult problem as both the number of targets and their trajectories need to be estimated. A cost function would have to include both the errors in the number of targets and the errors in the target state.

Appendix A

Background Mathematics

This Appendix presents some definitions and results of an probability, measures and integration. The results are needed especially in this thesis. More details can be found in [19, 39, 48, 63]

A.1 Probability and measures

Definition A.1: *The set of all possible outcomes of an experiment is called the sample space and is denoted by Ω .*

Definition A.2: *A collection $\sigma(\Omega)$ of subsets of Ω is called a σ -algebra if it satisfies*

- (a) $\emptyset \in \sigma(\Omega)$,
- (b) if $A_1, A_2, \dots \in \sigma(\Omega)$ then $\cup_{i=1}^{\infty} A_i \in \sigma(\Omega)$,
- (c) if $A \in \sigma(\Omega)$ then $A^c \in \sigma(\Omega)$

where $A^c = \Omega - A$ is the complement of A where $\Omega - A$ denotes the difference operation between the two sets Ω and A .

Definition A.3 (Filtration): *A sequence of σ -fields $\mathbf{F}_1, \mathbf{F}_2, \dots$ on Ω such that*

$$\mathbf{F}_1 \subset \mathbf{F}_2 \subset \dots \subset \mathbf{F}_n \subset \mathbf{F}_{n+1} \subset \dots$$

is called a filtration.

Definition A.4: *A measure μ on $(\Omega, \sigma(\Omega))$ is a function $\mu : \sigma(\Omega) \rightarrow [0, \infty)$ satisfying*

- (a) $\mu(A) \geq 0$ for all $A \in \sigma(\Omega)$, $\mu(\emptyset) = 0$,
- (b) if $A_1, A_2, \dots \in \sigma(\Omega)$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$ then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

The triple $(\Omega, \sigma(\Omega), \mu)$ is called a measure space. The pair $(\Omega, \sigma(\Omega))$ is called a measurable space. An element $A \in \sigma(\Omega)$ is called a measurable set. A probability measure is a measure with total measure one (i.e., $\mu(\Omega) = 1$); a probability space is a measure space with a probability measure. Several further properties of a measure can be derived from the definition

- (Monotonicity) A measure is monotonic, i.e. if $A_1 \subseteq A_2, A_1, A_2 \in \sigma(\Omega)$ then $\mu(A_1) \leq \mu(A_2)$
- (Measures of infinite unions of measurable sets)

- A measure μ is countably subadditive: If A_1, A_2, \dots is a countable sequence of sets in Ω , not necessarily disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

- A measure μ is continuous from below: If $A_1, A_2, \dots \in \sigma(\Omega)$ and $A_1 \subseteq A_2 \subseteq \dots$, then the union of the sets A_i is measurable, and

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mu(A_i)$$

- (Measures of infinite intersections of measurable sets): If $A_1, A_2, \dots \in \sigma(\Omega)$ and $A_1 \supseteq A_2 \supseteq \dots$, then the intersection of the sets A_i is measurable; furthermore, if at least one of the A_i has finite measure, then

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mu(A_i)$$

If $\Omega = \bigcup_{i=0}^{\infty} A_i$ for some countable sequence $A_i \in \sigma(\Omega)$ with $\mu(A_i) < \infty$, then μ is said to be σ -finite.

Definition A.5 (Measurable function): Let $(\mathcal{X}_1, \sigma(\mathcal{X}_1))$ and $(\mathcal{X}_2, \sigma(\mathcal{X}_2))$ be two measurable spaces. A mapping $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ is said to be measurable if the inverse images of a measurable set is measurable i.e. $f^{-1}(A) = \{x \in \mathcal{X}_1 : f(x) \in A\} \in \sigma(\mathcal{X}_1)$ for $A \in \sigma(\mathcal{X}_2)$.

Definition A.6: Let $(\Omega, \sigma(\Omega), \mathbb{P})$ be a probability space and (S, \mathcal{S}) be a measurable space. Then a **random variable** is a measurable function $X : \Omega \rightarrow S$ with the property that $\{\omega \in \Omega : X(\omega) \in B\} \in \sigma(\Omega)$ for any $B \in \mathcal{S}$. Such a function is said to be $\sigma(\Omega)$ -measurable

Definition A.7: Let (S, \mathcal{S}) be a measurable space. The **distribution function** of a random variable $X : \Omega \rightarrow S$ is the probability measure $\mu : \mathcal{S} \rightarrow [0, 1]$ given by $\mu(B) = (\mathbb{P} \circ X^{-1})(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ for any $B \in \mathcal{S}$

We denote the distribution function of a random variable X by \mathbf{P}_X .

Definition A.8: Let $(\mathcal{X}, \sigma(\mathcal{X}), \mu)$ be a measure space where μ is a (nonnegative, countably additive) measure. A set $A \in \sigma(\mathcal{X})$ will be called an **atom** for μ [66, 168], if

1. $\mu(A) > 0$ and
2. for any proper subset B of A i.e. $B \subset A$, $\mu(B) = 0$.

We shall say that μ is **purely atomic** or **simply atomic** if every measurable set of positive measure contains an atom. We shall say that μ is **nonatomic** (or **atomless**) if there are no atoms for μ .

A.2 Topology

Definition A.9: A **topology** \mathbb{T} on Ω is a family of subsets of Ω such that

- (conventions on empty set) $\emptyset, \Omega \in \mathbb{T}$
- (arbitrary union) if $A_i \in \mathbb{T}, i \in I$ then $\cup_{i \in I} A_i \in \mathbb{T}$ where I is an arbitrary set.
- (finite intersection) if $A_1, A_2, \dots, A_n \in \mathbb{T}$ then $\cap_{i=1}^n A_i \in \mathbb{T}$.

The pair (Ω, \mathbb{T}) is called a topological space. The open sets in Ω are defined to be the members of \mathbb{T} . A subset of Ω is said to be closed if its complement is in \mathbb{T} (i.e., its complement is open). A subset of Ω may be open, closed, both, or neither.

A Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. For a topological space \mathbb{T} on Ω , the collection of all Borel sets on Ω forms a σ -algebra, known as the Borel algebra or **Borel σ -algebra** $\mathcal{B}(\Omega)$. The Borel algebra on Ω is the smallest σ -algebra containing all open sets (or, equivalently, all closed sets).

Definition A.10: A collection of subset $\mathbf{B} \subseteq \mathbb{T}$ is a base for the topological space if each non-empty set $A \in \mathbb{T}$ can be represented as a union of a subfamily \mathbf{B}_i of \mathbf{B} , i.e. $A = \cup_{C \in \mathbf{B}_i} C$ where $\mathbf{B}_i \subseteq \mathbf{B}$

Definition A.11: A topological space $(\mathcal{X}, \mathbb{T})$ is said to be a **Hausdorff space** if for any $x, y \in \mathcal{X}$ $x \neq y$ there are disjoint open set U_x, U_y containing x and y respectively.

Definition A.12: A topological space $(\mathcal{X}, \mathbb{T})$ is called **compact** if each of its open covers has a finite subcover. Explicitly, this means that for every arbitrary collection $\{U_i \in \mathbb{T} : i \in I\}$ such that $\mathcal{X} = \cup_{i \in I} U_i$, there is a finite subset $J \subset I$ such that $\mathcal{X} = \cup_{i \in J} U_i$. A set $A \subset X$ is compact in X if each of its open covers has a finite subcover i.e. if $A \subset \cup_{i \in I} U_i$ where $\{U_i \in \mathbb{T} : i \in I\}$ is a collection of open sets, then there is a finite subset $J \subset I$ such that $A \subset \cup_{i \in J} U_i$.

Definition A.13: A topological space $(\mathcal{X}, \mathbb{T})$ is called **locally compact** if for each $x \in \mathcal{X}$ there is an open set $U \in \mathbb{T}$ that $x \in U$ and the closure of U is compact. The closure of U is $Cl(U) = \{x \in \mathcal{X} : V \cap U \neq \emptyset \text{ for each open set } V \text{ containing } x\}$

Definition A.14: In a topological space $(\mathcal{X}, \mathbb{T})$, a subset A of \mathcal{X} is said to be a **dense** subset of \mathcal{X} if $Cl(A) = \mathcal{X}$. A topological space $(\mathcal{X}, \mathbb{T})$ is **separable** if it contains a countable dense subset.

A.3 Integration of measurable function

Let $(\mathcal{X}, \sigma(\mathcal{X}))$ be a measurable space. Let $\mathbf{1}_A$ be the indicator function of subset $A \subseteq \mathcal{X}$, i.e. $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. Consider a measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$ in the following cases

- If f is a non-negative simple function i.e. $f = \sum_{i=1}^n c_i \mathbf{1}_{A_i}$ where $\{A_i\}$ is a finite decomposition of $\sigma(\mathcal{X})$, then the integral of f with respect to the measure μ is

$$\int_{\mathcal{X}} f(x) \mu(dx) = \sum_{i=1}^n c_i \mu(A_i)$$

- If f is a non-negative function i.e. $f : \mathcal{X} \rightarrow [0, \infty]$ and there exist a sequence of simple function such that $0 \leq f_n < f_{n+1} < f$ for all n and $\lim_{n \rightarrow \infty} f_n = f$, then the integral of f with respect to μ is defined as the limit of the integral of simple functions

$$\int_{\mathcal{X}} f(x) \mu(dx) = \lim_{n \rightarrow \infty} \int_{\mathcal{X}} f_n(x) \mu(dx)$$

- The integral of a general measurable function $f : \mathcal{X} \rightarrow [-\infty, \infty]$ with respect to μ is

$$\int_{\mathcal{X}} f(x) \mu(dx) = \int_{\mathcal{X}} f^+(x) \mu(dx) + \int_{\mathcal{X}} f^-(x) \mu(dx)$$

where the positive part of f is

$$f^+(x) = \begin{cases} f(x), & \text{if } 0 \leq f(x) \leq \infty; \\ 0, & \text{if } -\infty \leq f(x) \leq 0. \end{cases}$$

and the negative part of f is

$$f^-(x) = \begin{cases} -f(x), & \text{if } -\infty \leq f(x) \leq 0; \\ 0, & \text{if } 0 \leq f(x) \leq \infty. \end{cases}$$

The integral of f over any measurable set $A \in \sigma(\mathcal{X})$ is

$$\int_A f(x) \mu(dx) = \int \mathbf{1}_A(x) f(x) \mu(dx)$$

Definition A.15 (Absolutely continuous): Let $\mu_i, i = 1, 2$ be σ -finite μ_1 and μ_2 on the same measurable space $(\mathcal{X}, \sigma(\mathcal{X}))$. μ_1 is absolutely continuous with respect to a μ_2 , denoted $\mu_1 \preceq \mu_2$ if $\mu_2(A) = 0$ implies that $\mu_1(A) = 0$ for $A \in \sigma(\mathcal{X})$.

The Radon-Nikodým theorem says that μ_1 is *absolutely continuous* with respect to a μ_2 if there exist a measurable function $f : \mathcal{X} \rightarrow [0, \infty)$ such that

$$\mu_1(A) = \int_A f(x) \mu_2(dx).$$

Then f is called the Radon-Nikodým derivative or density of μ_2 with respect to μ_1 and is denoted by $g = d\mu_1/d\mu_2$

A.4 Markov Chains

This section gives some definitions and basic results for Markov chains. More details can be found in [39, 63].

Kronecker Delta δ_{ij} . If $i, j \in \mathbb{N}$, the Kronecker Delta is given by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$$

Dirac measure δ_x . Let $x \in S$, a Dirac measure δ_x on a set S (with any σ -algebra of subset of S) is defined for any $A \subseteq S$

$$\delta_x(A) = \begin{cases} 0, & \text{if } x \notin A \\ 1, & \text{if } x \in A. \end{cases}$$

Definition A.16: A generating-function of a sequence $a = \{z_i : i = 0, 1, \dots\}$ of a real numbers is the function G_a defined by

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}$$

The sequence a_i may in principle be reconstructed from the function G_a by setting

$$a_i = \frac{G_a^{(i)}(0)}{i!}$$

where $G^{(i)}$ denotes the i th derivative of the function G_a . In many circumstances it is easier to work with the generating function G_a than with the original sequence.

Definition A.17 (dth order Markov process): A hidden state sequence $\{X_t\}_{t \geq 1}$ is a d order Markov process when conditional distribution of X_k given the past values X_l with $1 \leq l < k$ depends on the d tuple X_{k-d}, \dots, X_{k-1} i.e.

$$\mathbb{P}(X_k | X_1, \dots, X_{k-1}) = \mathbb{P}(X_k | X_{k-d}, \dots, X_{k-1})$$

Definition A.18: [(State space)]The **state space** S is called

- (i) countable if S is discrete, with a finite or countable number of elements, and with \mathcal{S} the σ -field of all subsets of S .
- (ii) general if it is equipped with a countably generated σ -field¹ \mathcal{S}

¹Countably generated σ -field is a σ -algebra that can be generated by a countable collection of sets.

Appendix B

Mathematical Proofs

Now we prove (4.24) for $t \geq 1$

$$p_\theta(Z_t|Z_{1:t}) \approx \frac{1}{N} \sum_{n=1}^N w_t(X_{t-1:t}^n) \quad (\text{B.1})$$

with the convention that $p_\theta(Z_1|Z_{1:0}) = p_\theta(Z_1)$ and $X_{0:1} = X_1$.

Let $t = 2$, we have

$$\begin{aligned} p_\theta(Z_{1:2}) &= \int p_\theta(X_{1:2}, Z_{1:2}) dX_{1:2} = \int p_\theta(X_2, Z_2|X_1, Z_1) p_\theta(X_1, Z_1) dX_{1:2} \\ &\int p_\theta(X_1, Z_1) dX_1 \int p_\theta(X_2, Z_2|X_1, Z_1) dX_2 \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} &= \int g_\theta(Z_2|X_2) f_\theta(X_2|X_1) p_\theta(X_1, Z_1) dX_{1:2} \quad \text{by (4.17)} \\ &= \int g_\theta(Z_2|X_2) f_\theta(X_2|X_1) \left(\int w(X_1) q_\theta(X_1|Z_1) dX_1 \right) dX_2 \quad \text{(by (4.19))} \\ &= \int w(X_1) q_\theta(X_1|Z_1) dX_1 \int g_\theta(Z_2|X_2) f_\theta(X_2|X_1) dX_2 \\ &= \int w(X_1) q_\theta(X_1|Z_1) dX_1 \int w(X_{1:2}) q_\theta(X_2|Z_2, X_1) dX_2 \quad \text{(by (4.21)).} \end{aligned} \quad (\text{B.3})$$

From (B.2), we have

$$\begin{aligned} p_\theta(Z_{1:2}) &= \int p_\theta(X_1, Z_1) dX_1 \int p_\theta(X_2, Z_2|X_1, Z_1) dX_2 \\ &= p_\theta(Z_1) \int p_\theta(X_2, Z_2|X_1, Z_1) dX_2 = p_\theta(Z_1) p_\theta(Z_2|X_1, Z_1) \\ &\stackrel{(a)}{=} p_\theta(Z_1) p_\theta(Z_2|Z_1) \end{aligned} \quad (\text{B.4})$$

where (a) holds by Z_2 are statistically independent of X_1 conditional on Z_1 . Hence by (B.3) and (B.4), we have

$$p_\theta(Z_1) = \int w(X_1) q_\theta(X_1|Z_1) dX_1 \quad (\text{B.5})$$

$$p_\theta(Z_2|Z_1) = \int w(X_{1:2}) q_\theta(X_2|Z_2, X_1) dX_2. \quad (\text{B.6})$$

Let $t = T$, use the same argument for (B.3) and (B.4), we have

$$p_\theta(Z_{1:T}) = p_\theta(Z_1) \prod_{i=1}^T p_\theta(Z_i | Z_{1:i-1})$$

where $p_\theta(Z_1)$ is given in (B.5) and for $t = 2, \dots, T$ it follows from (B.6), we have

$$p_\theta(Z_t | Z_{1:t-1}) = \int w(X_{1:t}) q_\theta(X_t | Z_t, X_{t-1}) dX_t.$$

From Algorithm 4 and for $t = 1, \dots, T$, we have $X_t^n \sim q_\theta(X_t | Z_t, X_{t-1})$, $n = 1, \dots, N$ with convention that $q_\theta(X_1 | Z_1, X_0) = q_\theta(X_1 | Z_1)$. Thus the approximation of $p_\theta(Z_t | Z_{1:t-1})$ is

$$\hat{p}_\theta(Z_t | Z_{1:t-1}) = \frac{1}{N} \sum_{n=1}^N w_t(X_{t-1:t}^n)$$

Index

σ -algebra, 161

absolutely continuous, 164

absorbing, 61

aperiodic, 60

belief functional, 35

Borel σ -algebra, 163

closed set, 61

communicate, 60

distribution function, 162

EAP estimator, 10

filtration, 58

global density, 31

initial distribution, 57

intercommunicate, 60

irreducible, 61

MAP, 10

Markov chain, 56

maximum a posterior probability (MAP), 10

measurable space, 161, 162

measure, 161

measure space, 161

Minimum Mean Square error (MMSE), 10

random finite set, 28

random variable, 162

set integral, 34

state space, 165

topological space, 163

topology, 162

transition probability, 57

Bibliography

- [1] B. W. Ahn, J. W. Choi, and T. L. Song, “An adaptive interacting multiple model with probabilistic data association filter using variable dimension model,” in *Proceedings of the 41st SICE Annual Conference*, vol. 2, Aug. 2002, pp. 713 – 718.
- [2] A. Alouani, T. Rice, and N. Auger, “On the generalized nearest neighbor filter,” in *IEEE, 22nd Southeastern Symposium on System Theory*, 11-13 March 1990, pp. 261 –264.
- [3] B. Anderson and J. Moore, *Optimal filtering*. Prentice-Hall, 1979.
- [4] C. Andrieu, A. Doucet, and R. Holenstein, “Particle Markov chain Monte Carlo methods,” *Journal of the Royal statistical society series B-statistical methodology*, vol. 72, Part 3, pp. 269–342, 2010.
- [5] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174 –188, Feb. 2002.
- [6] K. Åström, *Introduction to stochastic control theory*. Academic Press, New York, 1970.
- [7] Y. Bar-Shalom, “Tracking methods in a multitarget environment,” *IEEE Transactions on Automatic Control*, vol. 23, no. 4, pp. 618 – 626, Aug. 1978.
- [8] Y. Bar-Shalom and T. Fortmann, *Tracking and data association*. Academic Press, 1988.
- [9] Y. Bar-Shalom and X. Li, *Estimation and tracking- Principles, techniques, and software*. Norwood, MA: Artech House, Inc, 1993.
- [10] ———, *Multitarget-multisensor Tracking: Principles and Techniques*. Urbana, IL., YBS, 1995.
- [11] Y. Bar-Shalom and E. Tse, “Tracking in a cluttered environment with probabilistic data association,” *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.
- [12] T. Bayes, “An Essay towards solving a Problem in the Doctrine of Chances.” *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, 1764. [Online]. Available: <http://www.stat.ucla.edu/history/essay.pdf>
- [13] Y. Bishop, S. Fienberg, and P. Holland, *Discrete multivariate analysis*. Springer Verlag, 2007.

- [14] S. Blackman, *Multiple-target tracking with radar applications*. Norwood, MA: Artech House, 1986.
- [15] ———, “Association and fusion of multiple sensor data,” in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. B. Shalom, Ed. Norwood MA: Artech House, 1990, pp. 187–218.
- [16] ———, “Multiple hypothesis tracking for multiple target tracking,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.
- [17] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Artech House Norwood, MA, 1999.
- [18] H. Blom and E. Bloem, “Probabilistic data association avoiding track coalescence,” *IEEE Transactions on Automatic Control*, vol. 45, no. 2, pp. 247–259, 2000.
- [19] Z. Brzeźniak and T. Zastawniak, *Basic stochastic processes: A course through exercises*. Springer, 1999.
- [20] O. Cappe, S. Godsill, and E. Moulines, “An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.
- [21] B. Carlin and T. Louis, “Bayes and empirical Bayes methods for data analysis,” *Statistics and Computing*, vol. 7, no. 2, pp. 153–154, 1997.
- [22] S. Challa, R. Evans, M. Morelande, and D. Musicki, *Fundamentals of object tracking*. Cambridge University Press, 2011.
- [23] D. Chan, D. Harrison, and D. Langan, “Tracking in a high-clutter environment: simulation results characterizing a bi-level MHT algorithm,” in *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 1954, 1993, pp. 540–551.
- [24] D. Chan and D. Langan, “Performance results of the bi-level MHT tracking algorithm for two crossing targets in a high clutter environment,” in *Proceedings of SPIE-Signal and Data Processing of Small Targets*, 1994, pp. 406–416.
- [25] D. Clark and S. Godsill, “Group Target Tracking with the Gaussian Mixture Probability Hypothesis Density Filter,” in *Proceeding of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, Dec. 2007, pp. 149–154.
- [26] D. Clark, I. Ruiz, Y. Petillot, and J. Bell, “Particle PHD filter multiple target tracking in sonar image,” *IEEE Transactions on Aerospace and Electronic System*, vol. 43, no. 1, pp. 409–416, January 2007.
- [27] D. Clark, B.-N. Vo, and J. Bell, “GM-PHD filter multi-target tracking in sonar images,” in *Proceedings of the SPIE on Signal Processing, Sensor Fusion, and Target Recognition XV*, 2006.

- [28] D. Clark, “Multiple target tracking with the probability hypothesis density filter,” Ph.D. dissertation, Department of Electrical, Electronic and Computing Engineering, Heriot-Watt University, Oct. 2006.
- [29] —, “First-moment multi-object forward-backward smoothing,” in *Proceeding of the 13th Conference on Information Fusion*, Jul. 26-29, 2010, pp. 1 –6.
- [30] D. Clark and J. Bell, “Bayesian multiple target tracking in forward scan sonar images using the PHD filter,” *Proceedings of IEE - Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 327 – 334, October 2005.
- [31] —, “Data Association for the PHD Filter,” in *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing Conference*, Dec. 2005, pp. 217 – 222.
- [32] —, “Convergence results for the particle PHD filter,” *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2652 –2661, July 2006.
- [33] —, “Multi-target state estimation and track continuity for the particle PHD filter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1441 –1453, October 2007.
- [34] D. Clark, K. Panta, and B.-N. Vo, “The GM-PHD Filter Multiple Target Tracker,” in *Proceeding of the 9th International Conference on Information Fusion*, July 2006, pp. 1 –8.
- [35] S. Colegrove and S. Davey, “On using nearest neighbours with the probabilistic data association filter,” in *The Record of the IEEE International on Radar*, 2000, pp. 53 –58.
- [36] S. Colgrove, A. Davis, and J. Ayliffe, “Track initiation and nearest neighbours incorporated into probabilistic data association,” *Journal of Electrical and Electronics Engineering, Australia*, pp. 191–8, 1986.
- [37] I. Cox and S. Hingorani, “An efficient implementation of Reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 138 –150, Feb. 1996.
- [38] I. Cox and M. Miller, “On finding ranked assignments with application to multitarget tracking and motion correspondence,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 486–489, 1995.
- [39] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Springer, 1988.
- [40] —, *An introduction to the theory of point processes*. Springer, 2007, vol. 2.
- [41] G. Demos, R. Ribas, T. Broida, and S. Blackman, “Applications of MHT to dim moving targets,” in *Proceedings of the SPIE-The International Society for Optical Engineering*, vol. 1305, 1990, pp. 297–309.

- [42] R. Douc and O. Cappé, “Comparison of resampling schemes for particle filtering,” in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, 2005, pp. 64–69.
- [43] A. Doucet, N. de Freitas, and N. Gordon, “An introduction to sequential Monte Carlo methods,” in *Sequential Monte Carlo methods in practice*, A. Doucet, N. De Freitas, and N. Gordon, Eds. New York: Springer-Verlay, 2001, pp. 3–14.
- [44] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.
- [45] A. Doucet, S. Godsill, and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering,” *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [46] A. Doucet and A. Johansen, “A tutorial on particle filtering and smoothing: Fifteen years later,” in *Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009.
- [47] O. Drummond, “Target tracking with retrodicted discrete probabilities,” in *Proceedings of the SPIE-The International Society for Optical Engineering*, vol. 3163, 1997, pp. 249–268.
- [48] R. Engelking, *General topology*. Heldermann Verlag, 1989, vol. 6.
- [49] O. Erdinc, P. Willett, and Y. Bar-Shalom, “Probability hypothesis density filter for multitarget multisensor tracking,” in *In Proceedings of the 8th International Conference on Information Fusion*, vol. 1, July 2005.
- [50] O. Erdinc, P. Willett, and S. Coraluppi, “The Gaussian Mixture Cardinalized PHD tracker on MSTWG and SEABAR’07 datasets,” in *Proceedings of the 11th International Conference on Information Fusion*, July 3 2008, pp. 1–8.
- [51] J. Fisher and D. Casasent, “Fast JPDA multitarget tracking algorithm,” *Applied optics*, vol. 28, no. 2, pp. 371–376, 1989.
- [52] G. Fishman, *Monte Carlo: concepts, algorithms, and applications*. Springer, 1996.
- [53] R. Fitzgerald, “Track Biases and Coalescence with Probabilistic Data Association,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-21, no. 6, pp. 822–825, Nov. 1985.
- [54] ———, *Multitarget-Multisensor Tracking: Advanced Applications*. Norwood MA: Artech House, 1990, ch. Development of practical PDA logic for multitarget tracking by microprocessor, pp. 1–23.
- [55] O. Frank, J. Nieto, J. Guivant, and S. Scheduling, “Multiple target tracking using Sequential Monte Carlo Methods and statistical data association,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, Oct. 2003, pp. 2718–2723.

- [56] H. Gauvrit, J. Le Cadre, and C. Jauffret, “A formulation of multitarget tracking as an incomplete data problem,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1242–1257, Oct. 1997.
- [57] J. Geweke, “Bayesian inference in econometric models using Monte Carlo integration,” *Econometrica: Journal of the Econometric Society*, vol. 57, no. 6, pp. 1317–1339, 1989.
- [58] W. Gilks, S. Richardson, and D. Spiegelhalter, “Introducing markov chain monte carlo,” in *Markov chain Monte Carlo in practice*, W. Gilks, S. Richardson, and D. Spiegelhalter, Eds. Chapman and Hall, 1996, pp. 1–19.
- [59] A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, “Evolving networks for group object motion estimation,” in *Proceedings of IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications*, April 2008, pp. 99–106.
- [60] I. Goodman, R. Mahler, and H. Nguyen, *Mathematics of data fusion*. Boston: Kluwer Academic Publishers, 1997.
- [61] N. Gordon, D. Salmond, and A. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” in *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, 1993, pp. 107–113.
- [62] J. Goutsias, R. Mahler, and H. Nguyen, *Random sets: theory and applications*. Springer New York, Berlin, Heidelberg, 1997.
- [63] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford University Press, USA, 2001.
- [64] ———, *Probability and random processes, 3rd*. Oxford University Press, USA, 2001, ch. Markov chains, pp. 213–239.
- [65] D. Hall and J. Llinas, *Handbook of multisensor data fusion*. CRC press, 2001.
- [66] P. Halmos, *Measure theory*. Springer-Verlag, New York., 1974, vol. 18.
- [67] Y. Hao, F. Meng, W. Zhou, F. Sun, and A. Hu, “Gaussian mixture probability hypothesis density filter algorithm for multi-target tracking,” in *Proceedings of the IEEE International Conference on Communications Technology and Applications*, Oct. 2009, pp. 738–742.
- [68] D. Hauschildt and N. Kirchhof, “Improving indoor position estimation by combining active TDOA ultrasound and passive thermal infrared localization,” in *Proceedings of 8th Workshop on Positioning Navigation and Communication (WPNC)*, Apr. 2011, pp. 94–99.
- [69] Y. Ho and R. Lee, “A Bayesian approach to problems in stochastic estimation and control,” *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.
- [70] R. Holenstein, “Particle Markov Chain Monte Carlo,” Ph.D. dissertation, The Faculty of Graduate Studies (Computer Science), The University Of British Columbia, 2009.

- [71] Y. L. Hongyan Zhu, Chongzhao Han, “Particle Labeling PHD Filter for Multi-Target Track-Valued Estimates,” in *Proceedings of 14th Conference on Information Fusion*, July 2011.
- [72] R. Hoseinnezhad, B.-N. Vo, B.-T. V. Vo, and D. Suter, “Bayesian integration of audio and visual information for multi-target tracking using a CB-member filter,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2011, pp. 2300–2303.
- [73] A. Houles and Y. Bar-Shalom, “Multisensor tracking of a maneuvering target in clutter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 2, pp. 176–189, Mar. 1989.
- [74] C. Hue, J.-P. Le Cadre, and P. Perez, “Tracking multiple objects with particle filtering,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 791–812, Jul. 2002.
- [75] A. Jazwinski, *Stochastic processes and filtering theory*. Academic Press, 1970.
- [76] A. Johansen, S. Singh, A. Doucet, and B.-N. Vo, “Convergence of the SMC Implementation of the PHD Filter,” *Methodology and Computing in Applied Probability*, vol. 8, no. 2, pp. 265–291, 2006.
- [77] N. Johnson, S. Kotz, and N. Balakrishnan, *Discrete multivariate distributions*. Wiley, 1997, vol. 165.
- [78] R. Juang, A. Levchenko, and P. Burlina, “Tracking cell motion using GM-PHD,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Jun.28-Jul.1, 2009, pp. 1154–1157.
- [79] S. Julier and J. Uhlmann, “A general method for approximating nonlinear transformations of probability distributions,” *Dept. of Engineering Science, University of Oxford, Tech. Rep*, 1996.
- [80] ———, “A new extension of the Kalman filter to nonlinear systems,” in *Proceedings of SPIE - Signal Processing, Sensor Fusion, and Target Recognition VI*, vol. 3, 1997, pp. 182–193.
- [81] R. Kalman, “A new approach to linear filtering and prediction problems,” *Transaction of the ASME - Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [82] R. Karlsson and F. Gustafsson, “Monte Carlo data association for multiple target tracking,” in *Proceedings of IEE on Target Tracking: Algorithms and Applications*, vol. 1, Oct. 2001, pp. 13/1 – 13/5 vol.1.
- [83] G. Kitagawa, “Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models,” *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. pp. 1–25, 1996. [Online]. Available: <http://www.jstor.org/stable/1390750>

- [84] T. Kurien, “Multi-Target Multi-Sensor Tracking: Advanced Applications, chapter: Issues in the Design of Practical Multi-Target Tracking Algorithms,” *Artech House, Norwood, MA*, 1990.
- [85] F. Lian, C. Han, and W. Liu, “Estimating Unknown Clutter Intensity for PHD Filter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 4, pp. 2066–2078, Oct. 2010.
- [86] L. Lin, Y. Bar-Shalom, and T. Kirubarajan, “Data association combined with the probability hypothesis density filter for multitarget tracking,” in *Proceedings of the SPIE - Signal and Data Processing of Small Targets*, 2004, pp. 464–475.
- [87] J. Liu and R. Chen, “Sequential monte carlo methods for dynamic systems,” *Journal of the American statistical association*, vol. 93, no. 443, pp. 1032–1044, 1998.
- [88] C. Lundquist, L. Hammarstrand, and F. Gustafsson, “Road Intensity Based Mapping Using Radar Measurements With a Probability Hypothesis Density Filter,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1397–1408, Apr. 2011.
- [89] W.-K. Ma, B.-N. Vo, S. Singh, and A. Baddeley, “Tracking an unknown time-varying number of speakers using TDOA measurements: a random finite set approach,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3291–3304, Sept. 2006.
- [90] E. Maggio, M. Taj, and A. Cavallaro, “Efficient Multitarget Visual Tracking Using Random Finite Sets,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1016–1027, Aug. 2008.
- [91] R. Mahler, “Global integrated data fusion,” in *Proc. 7th Nat. Symp. on Sensor Fusion*, vol. 1 (Unclassified). Sandia National Laboratories, Albuquerque, ERIM Ann Arbor MI, Mar. 16-18, 1994, pp. 187–199.
- [92] ———, *An Introduction to Multisource-Multitarget Statistics and Applications*. Lockheed Martin Technical Monograph, 104 pages, Mar. 15, 2000.
- [93] ———, “The search for tractable Bayesian multitarget filters,” in *Proceedings of SPIE - Conference on Signal and Data Processing of Small Targets*, Drummond, OE, Ed., vol. 4048, Apr. 24-27, 2000, pp. 310–320.
- [94] ———, “Detecting, tracking, and classifying group targets: a unified approach,” in *Proceedings of the SPIE - Conference on Signal Processing, Sensor Fusion, and Target Recognition X*, vol. 4380, Apr. 16-18, 2001, pp. 217–228.
- [95] ———, “Engineering statistics for multi-object tracking,” in *Proceedings of IEEE Workshop on Multi-Object Tracking*, Jul. 8 2001, pp. 53–60.
- [96] ———, “Multitarget Bayes filtering via first-order multitarget moments,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

- [97] ——. (2003, Jun.) Short tutorial: Engineering Statistics for Multi-object Tracking. [Online]. Available: <http://research.microsoft.com/en-us/um/redmond/events/womot03/mahler-tutorial.pdf>
- [98] ———, “A theory of PHD filters of higher order in target number,” in *Proceedings of the SPIE - Conference on Signal Processing, Sensor Fusion, and Target Recognition XV*, vol. 62350, Apr.17-19, 2006.
- [99] ———, “PHD filters of second order in target number,” in *Proceedings of the SPIE - the 18th Conference on Signal and Data Processing of Small Targets*, vol. 6236, Apr.18-20, 2006.
- [100] ———, “PHD filters of higher order in target number,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, Oct. 2007.
- [101] ———, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, Inc., 2007.
- [102] R. P. S. Mahler, B. T. Vo, and B. N. Vo, “CPHD Filtering With Unknown Clutter Rate and Detection Profile,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3497–3513, Aug. 2011.
- [103] R. Mahler, “A survey of PHD filter and CPHD filter implementations ,” in *Proceedings of SPIE - Signal Processing, Sensor Fusion, and Target Recognition XVI*, Kadar, I, Ed., vol. 6567, 2007.
- [104] ———, “A Theoretical Foundation for the Stein-Winter ’Probability Hypothesis Density (PHD)’ Multitarget Tracking Approach,” in *Proc. 2000 MSS National Symposium on Sensor and Data Fusion*, vol. I (Unclassified) . San Antonio, TX, Jun. 20-22, 2000.
- [105] ———, “Extended first-order Bayes filter for force aggregation,” in *Proceedings of the SPIE - Signal and Data Processing of Small Targets*, vol. 4728, 2002, pp. 196–207.
- [106] ———, “The Random-set approach to data fusion,” in *Proceedings of SPIE -Automatic Object Recognition IV*, vol. 2234, 1994, pp. 287 – 295.
- [107] ———, ““Statistics 101” for multisensor, multitarget data fusion,” *IEEE on Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 53–64, 2004.
- [108] R. Mahler, B.-N. Vo, and B.-T. Vo, “The Forward-Backward Probability Hypothesis Density Smoother,” in *Proceedings of the 13th Conference on Information Fusion*, Jul. 26-29, 2010.
- [109] G. Matheron, *Random sets and integral geometry*. New York: John Wiley & Sons In, 1975.
- [110] S. Meyn, R. Tweedie, and J. Hibey, *Markov chains and stochastic stability, 2nd edition*. Springer London et al., 2009.

- [111] M. Morelande, “A sequential Monte Carlo method for PHD approximation with conditionally linear/Gaussian models,” in *Proceedings of the 13th Conference on Information Fusion*, Jul. 26-29, 2010, pp. 1 –7.
- [112] Morelande, M.R. and Challa, S., “A multitarget tracking algorithm based on random sets,” in *Proceedings of the Sixth International Conference on Information Fusion*, vol. 2, 2003, pp. 807 – 814.
- [113] J. Mullane, B.-N. Vo, and M. Adams, “Rao-Blackwellised PHD SLAM,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2010, pp. 5410 –5416.
- [114] J. Mullane, B.-N. Vo, M. Adams, and B.-T. Vo, “A Random-Finite-Set Approach to Bayesian SLAM,” *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 268 –282, Apr. 2011.
- [115] J. Mullane, B.-N. Vo, M. Adams, and W. Wijesoma, “A random set formulation for Bayesian SLAM,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept. 2008, pp. 1043 –1049.
- [116] D. Musicki and R. Evans, “Integrated probabilistic data association-finite resolution,” *Automatica*, vol. 31, no. 4, pp. 559–570, 1995.
- [117] ———, “Joint Integrated Probabilistic Data Association - JIPDA,” in *Proceedings of the Fifth International Conference on Information Fusion*, vol. 2, 2002, pp. 1120 – 1125 vol.2.
- [118] ———, “Linear joint integrated probabilistic data association - LJIPDA,” in *Proceedings of the 41st IEEE Conference on Decision and Control*, vol. 3, Dec. 2002, pp. 2415 – 2420.
- [119] ———, “Joint integrated probabilistic data association: JIPDA,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1093 – 1099, Jul. 2004.
- [120] D. Musicki, R. Evans, and S. Stankovic, “Integrated probabilistic data association (IPDA),” in *Proceedings of the 31st IEEE Conference on Decision and Control*, 1992, pp. 3796 –3798 vol.4.
- [121] ———, “Integrated probabilistic data association,” *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1237 –1241, Jun. 1994.
- [122] D. Musicki and R. Evans, “Tracking in clutter using probabilistic data association,” in *Proceedings of the International Conference on Radar*, Oct. 1992, pp. 82 –85.
- [123] D. Musicki and S. Suvorova, “Tracking in clutter using IMM-IPDA-based algorithms,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 111 –126, Jan. 2008.
- [124] C. Musso, N. Oudjane, and F. LeGland, “Improving regularised particle filters,” in *Sequential Monte Carlo methods in practice*. New York: Springer-Verlag, 2001, pp. 247–271.

- [125] V. Nagarajan, M. Chidambara, and R. Sharma, "Combinatorial problems in multitarget tracking - a comprehensive solution," *IEE Proceedings F on Communications, Radar and Signal Processing*, vol. 134, no. 1, pp. 113–118, 1987.
- [126] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, vol. 1, 2004, pp. 735–742.
- [127] —, "Markov Chain Monte Carlo Data Association for Multi-Target Tracking," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 481–497, Mar. 2009.
- [128] K. Panta, "Multi-target tracking using 1st moment of random finite sets," Ph.D. dissertation, Department of Electrical and Electronic Engineering, Univerisity of Melbourne, February 2007.
- [129] K. Panta, D. Clark, and B.-N. Vo, "Data association and track management for the Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1003–1016, 2009.
- [130] K. Panta, B.-N. Vo, and S. Singh, "Improved Probability Hypothesis Density (PHD) Filter for Multitarget Tracking," in *Proceedings of the Third International Conference on Intelligent Sensing and Information Processing*, Dec. 2005, pp. 213 –218.
- [131] K. Panta, B.-N. Vo, S. Singh, and A. Doucet, "Probability hypothesis density filter versus multiple hypothesis tracking," in *Proceedings of SPIE - Conference on Signal Processing, Sensor Fusion, and Target Recognition XIII*, vol. 5429, 2004, pp. 284–295.
- [132] A. Pasha, B.-N. Vo, H. Tuan, and W.-K. Ma, "Closed Form PHD Filtering for Linear Jump Markov Models," in *Proceedings of the 9th International Conference on Information Fusion*, Jul. 2006.
- [133] S. Pasha, H. Tuan, and P. Apkarian, "The LFT based PHD filter for nonlinear jump Markov models in multi-target tracking," in *Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 28th Chinese Control Conference*, Dec. 2009, pp. 5478–5483.
- [134] S. Pasha, B.-N. Vo, H. D. Tuan, and W.-K. Ma, "A Gaussian Mixture PHD Filter for Jump Markov System Models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 919 –936, Jul. 2009.
- [135] N. T. Pham, "Tracking of multiple objects using the PHD filter," Ph.D. dissertation, Department of Electrical and Computer Engineering, National University of Singapore, 2007.
- [136] N. T. Pham, W. Huang, and S. Ong, "Tracking multiple speakers using CPHD filter," in *Proceedings of the 15th international conference on Multimedia*. ACM, Sept.24-29, 2007, pp. 529–532.

- [137] ———, “Multiple sensor multiple object tracking with GMPHD filter,” in *Proceedings of the 10th International Conference on Information Fusion*, Jul. 2007.
- [138] ———, “Tracking Multiple Objects using Probability Hypothesis Density Filter and Color Measurements,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, Jul. 2007, pp. 1511–1514.
- [139] M. Pitt and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.
- [140] A. Plinge, D. Hauschildt, M. H. Hennecke, and G. A. Fink, “Multiple speaker tracking using a microphone array by combining auditory processing and a gaussian mixture cardinalized probability hypothesis density filter,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2011, pp. 2476–2479.
- [141] E. Pollard, B. Pannetier, and M. Rombaut, “Hybrid Algorithms for Multitarget Tracking using MHT and GM-CPHD,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 832–847, April 2011.
- [142] G. Pulford and R. Evans, “A multipath data association tracker for over-the-horizon radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1165–1183, Oct. 1998.
- [143] K. Punithakumar, T. Kirubarajan, and A. Sinha, “A multiple model probability hypothesis density filter for tracking maneuvering targets,” in *Proceedings of SPIE - Conference on Signal and Data Processing of Small Targets*, 2004, pp. 113–121.
- [144] C. Rago, P. Willett, and R. Streit, “A comparison of the JPDAF and PMHT tracking algorithms,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, May 1995, pp. 3571–3574.
- [145] D. Reid, “An algorithm for tracking multiple targets,” *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.
- [146] S. Reuter and K. Dietmayer, “Pedestrian Tracking Using Random Finite Sets,” in *Proceedings of the 14th International Conference on Information Fusion*, Jul. 2011.
- [147] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*. Artech House Publishers, 2004.
- [148] B. Ristic, D. Clark, and B.-N. Vo, “Improved SMC implementation of the PHD filter,” in *Proceedings of the 13th International Conference on Information Fusion*, Jul. 2010.
- [149] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Verlag, 2004.
- [150] ———, *Monte Carlo statistical methods, 2nd*. Springer, 2004, ch. Markov Chains, pp. 205–265.

- [151] G. Roberts, “Markov chain concepts related to sampling algorithms,” *Markov chain Monte Carlo in practice*, vol. 57, pp. 45–57, 1996.
- [152] S. Ross, *Stochastic processes*. Wiley New York, 1996.
- [153] ———, *Stochastic processes, 2nd*. Wiley New York, 1996, ch. Markov chains, pp. 163–230.
- [154] D. Salmond, “Mixture reduction algorithms for target tracking in clutter,” in *Proceedings of SPIE - Conference on Signal and Data Processing of Small Targets*, vol. 1305, 1990, pp. 434–445.
- [155] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, “A Consistent Metric for Performance Evaluation of Multi-Object Filters,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.
- [156] R. Serfozo, *Basics of Applied Stochastic Processes*. Springer, 2009, ch. Markov Chains, pp. 1–98.
- [157] A. Shiryaev, *Probability, 2nd*. Springer, 1996, ch. Sequences of Random Variables That form Markov Chains, pp. 564–586.
- [158] H. Sidenbladh, “Multi-target particle filtering for the probability hypothesis density,” in *Proceedings of the Sixth International Conference on Information Fusion*, vol. 2, 2003, pp. 800 – 806.
- [159] H. Sidenbladh and S.-L. Wirkander, “Tracking Random Sets of Vehicles in Terrain,” in *Proceedings of the IEEE Workshop on Computer Vision and Pattern Recognition*, vol. 9, Jun.16-22 2003.
- [160] R. Singer, “Estimating optimal tracking filter performance for manned maneuvering targets,” *IEEE Transactions on Aerospace and Electronic Systems*, no. 4, pp. 473–483, 1970.
- [161] P. Smith and G. Buechler, “A branching algorithm for discriminating and tracking multiple objects,” *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 101–104, 1975.
- [162] D. Stoyan, W. Kendall, J. Mecke, and D. Kendall, *Stochastic geometry and its applications*. Wiley New York, 1995.
- [163] R. Streit, “Probabilistic multi-hypothesis tracking,” DTIC Document, Tech. Rep., 1995.
- [164] R. Streit and T. Luginbuhl, “Maximum likelihood method for probabilistic multihypothesis tracking,” in *Proceedings of SPIE*, vol. 2235, 1994, pp. 394–405.
- [165] D. Svensson, J. Wintenby, and L. Svensson, “Performance evaluation of MHT and GM-CPHD in a ground target tracking scenario,” in *Proceedings of the 12th International Conference on Information Fusion*, Jul. 2009, pp. 300 –307.

- [166] M. Ulmke, O. Erdinc, and P. Willett, "Gaussian mixture cardinalized PHD filter for ground moving target tracking," in *Proceedings of the 10th International Conference on Information Fusion*, July 2007, pp. 1–8.
- [167] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. Wan, "The unscented particle filter," *Advances in Neural Information Processing Systems*, pp. 584–590, 2001.
- [168] M. Vihola, "Random sets for multitarget tracking and data fusion," Ph.D. dissertation, Department of Information Technology, Tampere University of Technology, 2004.
- [169] B.-N. Vo and W. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, p. 4091, 2006.
- [170] B.-N. Vo and W.-K. Ma, "A closed-form solution for the probability hypothesis density filter," in *Proceedings of the 8th International Conference on Information Fusion*, vol. 2, Jul. 2005.
- [171] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1224–1245, 2005.
- [172] B.-T. Vo, "Random Finite sets in Multi-Object Filtering," Ph.D. dissertation, School of Electrical, Electronic and Computer Engineering, University of Western Australia, 2008.
- [173] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The Cardinalized Probability Hypothesis Density Filter for Linear Gaussian Multi-Target Models," in *Proceedings of the 40th Annual Conference on Information Sciences and Systems*, Mar. 2006, pp. 681–686.
- [174] —, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, p. 3553, 2007.
- [175] B.-N. Vo. (2009, July 28) Random Finite Sets in Stochastic Filtering. [Online]. Available: <http://www.ieeevic.org/docs/slides/IEEEMelbourne1.ppt>
- [176] B.-N. Vo, A. Pasha, and H. D. Tuan, "A Gaussian Mixture PHD Filter for Nonlinear Jump Markov Models," in *Proceedings of the 45th IEEE Conference on Decision and Control*, Dec. 2006, pp. 3162–3167.
- [177] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo implementation of the PHD filter for multi-target tracking," in *Proceedings of the 6th International Conference on Information Fusion*, vol. 2, 2003, pp. 792–799.
- [178] B.-N. Vo, S. Singh, and W. K. Ma, "Tracking multiple speakers using random sets," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 2004, pp. II–357–II–360.

- [179] B.-N. Vo, B.-T. Vo, and R. Mahler, "A closed form solution to the Probability Hypothesis Density Smoother," in *Proceedings of the 13th Conference on Information Fusion*. IEEE, 2010.
- [180] B.-N. Vo, B.-T. Vo, and S. Singh, "Sequential Monte Carlo methods for static parameter estimation in random set models," in *In Proceedings of the 2nd International Conference on Intelligent Sensors, Sensor Networks, and Information Processing*, Dec.5-8 2004, pp. 313 – 318.
- [181] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic, "Bernoulli forward-backward smoothing for joint target detection and tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 99, pp. 4473–4477, 2011.
- [182] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Bayesian Filtering With Random Finite Set Observations," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1313 –1326, April 2008.
- [183] ———, "The Cardinality Balanced Multi-Target Multi-Bernoulli Filter and Its Implementations," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 409 –423, Feb. 2009.
- [184] B.-T. Vo, B.-N. Vo, R. Hoseinnezhad, and R. Mahler, "Multi-Bernoulli filtering with unknown clutter intensity and sensor field-of-view," in *Proceedings of the 45th Annual Conference on Information Sciences and Systems*, Mar.23-25, 2011.
- [185] A.-T. Vu, B.-N. Vo, and R. Evans, "Particle Markov Chain Monte Carlo for Bayesian Multi-target Tracking," in *Proceedings of the 14th International Conference on Information Fusion*, Jul.5-8, 2011.
- [186] Y.-D. Wang, J.-K. Wu, W. Huang, and A. Kassim, "Gaussian mixture probability hypothesis density for visual people tracking," in *Proceedings of the 10th International Conference on Information Fusion*, Jul.9-12, 2007, pp. 1 –6.
- [187] Y.-D. Wang, J.-K. Wu, A. Kassim, and W. Huang, "Data-Driven Probability Hypothesis Density Filter for Visual Tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1085 –1095, Aug. 2008.
- [188] Y. Wang, Z. Jing, and S. Hu, "Data association for PHD filter based on MHT," in *Proceedings of the 11th International Conference on Information Fusion*, Jun.30 -Jul.3, 2008.
- [189] J. Wei and X. Zhang, "Efficient Node Collaboration for Mobile Multi-Target Tracking Using Two-Tier Wireless Camera Sensor Networks," in *Proceedings of IEEE International Conference on Communications*, May 23-27, 2010.
- [190] N. Whiteley, S. Singh, and S. Godsill, "Auxiliary Particle Implementation of the Probability Hypothesis Density Filter," in *Proceedings of the 5th International Symposium on Image and Signal Processing and Analysis*, Sept.27-29 2007, pp. 510 –515.

- [191] ———, “Auxiliary Particle Implementation of Probability Hypothesis Density Filter,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1437–1454, Jul. 2010.
- [192] J. Wu and S. Hu, “PHD filter for multi-target visual tracking with trajectory recognition,” in *Proceedings of the 13th Conference on Information Fusion*, Jul.26-29, 2010.
- [193] L. Xiaodong, Z. Linhu, and L. Zhengxin, “Probability hypothesis densities for multi-sensor, multi-target tracking with application to acoustic sensors array,” in *Proceedings of the 2nd International Conference on Advanced Computer Control*, vol. 5, Mar.27-29, 2010, pp. 218–222.
- [194] T. Xu and W. Ping, “Improved peak extraction algorithm in SMC implementation of PHD filter,” in *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems*, Dec. 2010.
- [195] J. J. Yin and J. Q. Zhang, “The nonlinear multi-target multi-bernoulli filter using polynomial interpolation,” in *Proceedings of IEEE 10th International Conference on Signal Processing*, Oct.24-28, 2010, pp. 2551–2554.
- [196] J. Yin, J. Zhang, and L. Ni, “The polynomial predictive particle MeMber filter,” in *Proceedings of the 2nd International Conference on Mechanical and Electronics Engineering*, vol. 1, Aug.1-3, 2010, pp. V1–18–V1–22.
- [197] J. Yin, J. Zhang, and J. Zhao, “The Gaussian Particle multi-target multi-Bernoulli filter,” in *Proceedings of the 2nd International Conference on Advanced Computer Control*, vol. 4, Mar.27-29, 2010, pp. 556–560.
- [198] T. Zajic and R. Mahler, “A particle-systems implementation of the PHD multitarget tracking filter,” in *Proceedings of SPIE - Signal Processing, Sensor Fusion, and Target Recognition XII*, vol. 5096, Apr.21-23, 2003, pp. 291–299.
- [199] J. Zhang, H. Ji, and C. Ouyang, “A new Gaussian mixture particle CPHD filter for multi-target tracking,” in *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems*, Dec.6-8, 2010.
- [200] S. Zhang, J. Li, B. Fan, and L. Wu, “Data association for GM-PHD with track oriented PMHT,” in *Proceedings of the 3rd International Symposium on Systems and Control in Aeronautics and Astronautics*, June 2010, pp. 386–391.
- [201] B. Zhou and N. Bose, “An efficient algorithm for data association in multitarget tracking,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 458–468, Jan. 1995.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Vu, Tuyet Thi Anh

Title:

A Particle Markov Chain Monte Carlo algorithm for random finite set based multi-target tracking

Date:

2011

Citation:

Vu, T. T. A. (2011). A Particle Markov Chain Monte Carlo algorithm for random finite set based multi-target tracking. PhD thesis, National ICT Australia and Department of Electrical and Electronic Engineering, The University of Melbourne.

Persistent Link:

<http://hdl.handle.net/11343/36875>

File Description:

A Particle Markov Chain Monte Carlo algorithm for random finite set based multi-target tracking

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.