NTNU
Norwegian University of
Science and Technology

Multi target tracking

Using random finite sets with a hybrid state space and approximations

Lars-Christian Ness Tokle

Master of Science in Cybernetics and Robotics
Submission date: August 2018
Supervisor: Edmund Førland Brekke, ITK

Norwegian University of Science and Technology
Department of Engineering Cybernetics

Abstract

This study provides a thorough investigation into the theoretical framework and background around the standard model used in multi-target tracking (MTT), including probabilistic graphical models (PGMs) with belief propagation (BP), Bayesian state estimation for a hybrid state space, random finite sets (RFSs) and labeling of sets in their union and the Kullback-Leibler divergence (KL-divergence). This serves as a foundation for developing novel derivations of the Poisson multi Bernoulli mixture (PMBM) filter and for how to include a hybrid state space. The models in the interacting multiple models (IMM) are viewed as a discrete state in the hybrid state space. This enables appropriate conditioning and thus it is possible to avoid the increase in computational complexity of having the discrete states in the data association.

Through the derivations it is seen that the components of the target set, i.e. the underlying sets in the union, can be handled independently under a given association. This is used through the prediction and update step to provides track continuity, and hence the possibility for track labeling. Additionally, track labels are seen as being a latent variable pointing to individual sets, being either a single track set or the undetected targets set, in the unionized set of targets. The track labels follow a specific track after first detection, and hence in a manner provide target identities. The only change in the multi target distribution is in the distribution of the union, and no changes are made to the individual distribution components in this union. The only thing changed in the distribution of the union is the addition of a “labeled subset extractor”. Thus, it is seen as being a different approach than the labeling done in the labeled multi Bernoulli (LMB).

A total target information distribution is stated as a compact way of viewing the complete picture within MTT, and furthering insight into identities and tracks. It might lend itself as a starting point for studies of new approximative algorithms on PGMs. Furthermore, the relationship to most of the well known MTT filters, including PMBM, multiple hypotheses tracker (MHT), track-oriented marginal MeMBer-Poisson (TOMB/P), joint integrated probabilistic data association (JIPDA) with its degenerate cases and probability hypothesis density (PHD), are provided, recognizing them as approximations of the PMBM, or as attaining association variables within their distribution. The loopy belief propagation (LBP) way of approximating the association probabilities of Williams and Roslyn is given. Lastly it is shown how much larger the expected number of undetected targets is compared to the expected number of born targets after convergence, as a function of model parameters, along with the relationship between new born targets and clutter in the case of constant initial probability.

Sammendrag

Denne studien gir en grundig innføring i det teoretiske grunnlaget rundt standardmodellen for multippel målfølging. Dette inkluderer grafiske sannsynlighetsmodeller med beskjedsending, Bayesiansk tilstandsestimering for et hybrid tilstandsrom, tilfeldige endelige mengder og merking av mengdene i deres union, samt Kullback-Leibler-divergensen. Dette fungerer som en grunnmur for en ny utledning av Poisson-multippel-Bernoulli-mikstur-filteret og hvordan inkludere et hybrid tilstandsrom i dette. Modellene i interagerende multiple modeller blir så tolket som en diskret tilstand i det hybride tilstandsrommet. Dette gjør at en kan foreta en passende betingelse slik at vi kan unngå å øke regnekompleksiteten ved å ta de diskrete tilstandene med i dataassosiasjonen.

I utlendingen ser en at komponentene i målmengden, altså de underliggende mengdene i unionen, kan bli håndtert uavhengig gitt en assosiasjon. Dette blir brukt i prediksjonen og oppdateringen for å gi kontinuitet i følgingen[1], og derfor mulighet for merking. I tilegg, så er merkingen av følgingen sett til å være en latent variabel som peker til individuelle mengder i unionen av målmengden, hvor de individuelle mengdene enten er en enkeltfølgingsmengde eller den udetekterte målmengden. Merkingene hører til en spesifikk følging etter første deteksjon, og gir derfor på en måte en mulighet til å ha målidentiteter. Den eneste endringen i målmengdedistribusjonen er i distribusjonen over unionen, og det er ingen endringer i de individuelle komponentdistribusjonene. Den eneste forandringen i unionsdistribusjonen er at det er lagt til en "merket delmengdeplukker". Dette er derfor sett som annerledes tilnærming enn hva som er gjort i for eksempel i den merkede multippel Bernoulli, kalt "labeled multi Bernoulli" på englesk.

En total målinformasjonsdistribusjon blir vist som en kompakt måte å se det komplette bildet av multippel målfølging, og tilfører samtidig innsikt i identiteter, merkinger og følgingene. Denne kan også muligens brukes som et utgangspunkt for å finne nye approksimeringsalgoritmer på grafiske sannsynlighetsmodeller. Hvordan man kan få de fleste av de velkjente målfølgingsalgoritmene fra denne blir også gitt (se det engelske sammendraget for en liste over hvilke). De kan bli sett på som approksimasjoner av det utledede filteret eller at de trekker ut assosiasjonshypotesene fra det. Det blir også vist hvordan tilbakekommende beskjedsending for å kalkulere de marginale assosiasjonssannsynlighetene kan bli gjort som beskrevet av Williams and Roslyn. Til slutt vises det hvor mange flere forventede udetekterte mål en vil ha enn forventede tilkommende mål gitt som en funksjon av modellparametere når den har konvergert, sammen med forholdet mellom det forventede antallet tilkommende mål og forventede antallet falske målinger ved approximasjonen av konstant initiell eksistenssannsynlighet.

[1] følging blir her brukt for det engelske ordet "track"

Preface

Problem description

Multi-target tracking
Using random finite sets with a hybrid state space and approximations

Multi-target tracking (MTT) can be decomposed into two tasks; data association and filtering. The former task is about deciding which measurement comes from which target, if any. The latter task is about estimating the kinematic state of a target conditional on such associations. In recent years a new mathematical framework called random finite set (RFS) has gained considerable attention. Data association, filtering and a probabilistic number of targets appears within the mathematical framework, when modeled correctly, and it has therefore given considerable insight into the problem with derivations of both old and new algorithms. It also shows that the problem is still not fully understood, and there are therefore several different schools on how to make algorithms based on the framework.

The association problem can be quite challenging due to its inherent exponential complexity, which not only applies to multi-frame methods such as multiple hypotheses tracker (MHT), but also to single-frame methods, such as joint probabilistic data association (JPDA). One promising approach to circumvent this exponential complexity is belief propagation (BP).

Another challenge in multi-target tracking is that a single kinematic model may not be appropriate all the time. It may for example be desirable to switch between low-maneuverability and high-maneuverability models in an adaptive manner. The standard solution to this is the interacting multiple models (IMM) method, which is relatively straightforward to integrate with the single-target probabilistic data association (PDA).

The penultimate aim of this project would be to investigate the potential for including multiple models (MM) in the RFS framework and combine notions between RFS, MHT and JPDA.

The project involves the following tasks.

1. Generalize the standard IMM to state dependent model transitions and general distributions
2. Derive the RFS Poisson multi Bernoulli mixture (PMBM) filter and show how MM can be incorporated
3. Describe a way to extract track identities from a RFS
4. Relate the PMBM to MHT, joint integrated probabilistic data association (JIPDA) and probability hypothesis density (PHD) along with their inclusion of MM
5. Propose a way to combine notions from MHT and JPDA

Work description

This project started out as a continuation of my semester project, which was about using loopy belief propagation (LBP) on the association when including MM. It was first proposed that I would continue to do some more experiments with IMM-JPDA using LBP while also extending the algorithm to work on Gaussian mixtures using some form of mixture reduction. After a while, I came about the hybrid state interpretation of MM, and realized that what I had done in the semester project was over-complicating the problem. I also signed a contract for a PhD position in sensor fusion right before starting this work, so I wanted this thesis to build a good foundation for future work. At the same time, there were also questions I had that were unanswered around the theoretical framework. Knowing that I had some holes in my understanding from my semester project, wanting to make a good foundation and having these unanswered questions, meant I spent much time 'doodling around' with (or, focused on, if you prefer) some derivations and target identities within RFSs.

I soon realized that the report was going to be quite theoretically heavy, and lengthy if I included all this. Gaussian mixture target tracking was already to some degree studied, and I had in mind some original notions and theoretical developments that I had not seen combined in the literature before, at least in an accessible manner. It was therefore decided to leave the Gaussian mixture implementation, as it would mainly just add pages to the report if it was without thorough simulations and analysis. Some general results are already available in the literature, and not providing it here allows more focus on the description of the multi target framework, how the different algorithms come about, what their relations are and how to interpret the different aspects. Even as it is now, there are quite some subtle details that I do not feel I have been able to portray adequately, although I have certainly tried.

Also, mastering notation in multi target tracking seems to be an art in itself. It often either gets too cluttered or carries too little information, is not general enough, or too abstract. Even after deciding upon a notation style, using it correctly is still a time consuming task. I hope that I have been able to make the notation of this text understandable to the reader.

Acknowledgments

This work has been completed independently. I have to thank my supervisor Edmund Brekke, for the help he has given in form of literature advice and discussions on several topics. Implementations of several algorithms (i.e. JPDA, LBP-association, IMM-PDA) have been given from various sources, but have all been rewritten and amended to fit each other and to work with the IMM-JPDA assumptions used in the simulations done during the work. There are however no results from these algorithms in this report. I also have to thank my partner, Erin Lindsay for doing proof reading and helping me with how to phrase some things.

Lars-Christian Ness Tokle Trondheim, 24.08.2018

Contents

Acronyms

BP	belief propagation
BPP	Bernoulli point process
CPHD	cardinalized probability hypothesis density
DAG	directed acyclic graph
FISST	finite set statistics
i.i.d.	independent identically distributed
IMM	interacting multiple models
IPDA	integrated probabilistic data association
JIPDA	joint integrated probabilistic data association
JPDA	joint probabilistic data association
KF	Kalman filter
KL-divergence	Kullback-Leibler divergence
LBP	loopy belief propagation
LHS	left hand side
LMB	labeled multi Bernoulli
MB	multi Bernoulli
MBM	multi Bernoulli mixture
MHT	multiple hypotheses tracker
MM	multiple models
MOMB/P	measurement-oriented marginal MeMBer-Poisson
MRF	Markov random field
MTT	multi-target tracking
PDA	probabilistic data association
PDF	probability density function
PGM	probabilistic graphical model
PHD	probability hypothesis density
PMB	Poisson multi Bernoulli

PMBM	Poisson multi Bernoulli mixture
PPP	Poisson point process
RFS	random finite set
RHS	right hand side
SDF	set distribution function
TOMB/P	track-oriented marginal MeMBer-Poisson
TOMHT	track oriented multiple hypotheses tracker
VSIMM	variable structure interacting multiple models

Nomenclature

$[1:n]$	The set of integers $\{1, \dots, n\}$
$p_\sigma(\xi)$	A probability density function of the (possibly vector valued) variable σ evaluated at ξ
$\mathrm{P}(A)$	probability of event A occurring
$\mathop{\mathbb{E}}_p[f(x)]$	The expected value of $f(x)$ under the distribution p
$\mathbb{H}[p(x), q(x)]$	$= \mathop{\mathbb{E}}_p[-\ln(q(x))]$, the cross entropy between the distributions $p(x)$ and $q(x)$
$\mathbb{H}[p(x)]$	$= \mathbb{H}[p(x), p(x)] = \mathop{\mathbb{E}}_p[-\ln(p(x))]$, the entropy of the distribution $p(x)$
$\mathcal{X}$	The state space for continuous states
$\mathcal{L}$	The state space for discrete states
x_t^i	The continuous state for target i at time t
l_t^i	The discrete state for target i at time t
z_t^j	Measurment j at time t
X	$= \{x^1, \dots, x^n\}$, the RFS multi target state
Z_t	$= \{z^1, \dots, z^m\}$, the RFS multi target measurement
$f_{t\,\vert\,t'}^{i,l,\theta^i}(x)$	The state PDF in continuous state x and discrete state l of track i under a hypothesis θ^i at time t given information up to and including time t', where typically θ are association hypotheses and $t' \in \{t_-, t\}$. i is suppressed if only one target is considered, θ suppressed if only one hypothesis is considered, l is supressed if there are no discrete state or they are marginalized out and $t\,\vert\,t'$ suppressed if the time and conditioning is clear from context
$f^l(x_t \,\vert\, x_{t_-})$	The continuous state transition PDF under the discrete state hypothesis l, which is suppressed when there are no discrete states or explicitly taken

	into consideration otherwise
$\pi^{l_t}_{l_{t_-}}(x_{t_-})$	The discrete transition probability from l_{t_-} to l_t in continuous state x_{t_-}, which is suppressed if the transition is independent of the continuous states
$\mathrm{P}_d(x)$	Probability of sensor detection in the state, x
$\bar{\mathrm{P}}^{\zeta}_d$	Averaged probability of detection under a hypothesis or distribution ζ
$\mathrm{P}_s(x)$	Probability that the target do not leave the scene before next time step in state, x, known as surviving
$\bar{\mathrm{P}}^{\zeta}_s$	Averaged probability of survival under a hypothesis or distribution ζ
$h^l(z \mid x)$	Conditional measurement PDF at z given the continuous state, x, and discrete state, l
$\eta^{\alpha}_t(x)$	The PPP intensity of a target arriving the scene in state, x, known as birth, and α can allow for multiple birth processes.
$\bar{\eta}^{\alpha}_t$	The expected number of born targets in process α
$\lambda_{t \mid t'}(x)$	The PPP intensity of targets in the scene that are still undetected by a sensor in state, x, at time t given information up to and including time t', and known as undetected target intensity
$\bar{\lambda}_{t \mid t'}$	$= \int_{x \in \mathcal{X}} \lambda_{t \mid t'}(x)\, \mathrm{d}x$, the expected number of unknown targets at time t given information up to and including time t'
$f^{\lambda}_{t \mid t'}(x)$	$= \frac{\lambda_{t \mid t'}(x)}{\bar{\lambda}_{t \mid t'}}$, the state PDF of undetected targets
$\mu_t(z)$	The PPP intensity of false alarms at z
$\bar{\mu}_t$	$= \int_{z \in \mathcal{Z}} \mu_t(z)\, \mathrm{d}z$, the expected number of false alarms
$h^{\mu}_t(z)$	$= \frac{\mu_t(z)}{\bar{\mu}_t}$, the PDF of false measurements
$\mathcal{P}^N_n$	$= \{\pi \subseteq [1:N] : \lvert\pi\rvert = n, \pi^i = j \Rightarrow \pi^{i'} \neq j \,\forall\, i \neq i', i, i' \in [1:n]\}$, the set of all indexings so that it contains all the ways of chosing n objects out of N in order
$\mathcal{Q}^N_n$	$= \{\pi \subseteq [0:N] : \lvert\pi\rvert = n, \pi^i = j > 0 \Rightarrow \pi^{i'} \neq j \,\forall\, i \neq i', i, i' \in [1:n]\} \supset \mathcal{P}^N_n$, the set of all indexings so that it contains all the ways of chosing n objects out of $N+1$ in order, where the 0'th object can be repeated

1 Introduction

1.1 Motivation and problem

Multi-target tracking (MTT) is the field in which one tries to track several targets that are in close proximity to each other. This involves estimating both the number and state of targets from sensor data. Ongoing challenges include firstly determining measurement origin and secondly tracking maneuvering targets. If no specific deterministic features of the targets are captured, the origin of a received measurement may be ambiguous for near targets. Also, some sensors have less than unity detection probability and may detect clutter (targets that do not exist). Hence, upon completion of a sensor scan, measurements could originate from clutter, known targets (previously seen), or new targets (not previously seen). On the other hand, an absence of measurements does not necessarily mean there are no targets present. Furthermore, tracking maneuvering targets (targets varying in speed and direction) is challenging because varying or multiple dynamical models must be considered in order to properly estimate target behaviour. This can be computationally expensive when dealing with several targets, and is further complicated by uncertainty in measurement origin and target position. In order to develop a safely operating autonomous system using such sensors, we need an effective way of retaining valuable information and reducing measurement origin ambiguity as best as possible.

An example of such a system is a ship using radar as a sensor. The standard radar is exactly a sensor of the type described above. Radar data does not give deterministic information on ship identity, other ships might not be detected, and non-existent ships may be falsely detected. To travel safely using only this sensor to avoid collision is therefore not straightforward; the number of other ships in close proximity and where they are moving can only be determined probabilistically. If two ships are moving close to or crossing paths with each other, it can be difficult to determine from the radar data which of the ships gave rise to which measurement in each scan; are the ships travelling in a straight path, or maybe one of them is maneuvering around the other somehow? Also one ship might even be blocking the view to the other ship for the radar, and the other might therefore be temporarily undetectable. Not knowing these things for sure can in some cases lead an algorithm to fail to understand what is going on properly, and therefore lead to dangerous situations if decisions are made based on its output.

Through the years, there has mainly been two approaches to handle the measurement origin uncertainty in MTT. The first, known as multiple hypotheses tracker (MHT), is to form an “ever growing” tree of hypotheses, with some method of discarding or combining hypotheses afterwards. The second, known as joint probabilistic data association (JPDA), is to marginalize over all joint association hypotheses for each target at each time step. In recent years a new statistical framework has emerged using something called random finite set (RFS) with finite set statistics (FISST) as the mathematical tool to describe distributions, giving a new paradigm of understanding and algorithms. This framework has shown to be a good generalized framework for MTT, providing new algorithms and derivations of most of the earlier well known algorithms using different types of restrictions or approximations. However, no matter how we model the targets or receive the measurements, the greatest challenge is how to best reduce the ever-growing set of hypotheses to form both accurate and computationally tractable algorithms for specific applications. Even though decent attempts on this have been proposed, it is still application specific how well these methods perform.

In many real world applications, we do not know which dynamical model the targets follow. Continuing with the example of tracking the behaviour of ships, different dynamical models could describe a ship which is travelling on auto-pilot at a constant speed and direction (non-maneuvering), or which can vary in speed and direction (maneuvering). One therefore often increases the set of hypotheses to include several models for target dynamics. In many cases, this has shown to give better tracking performance when the targets vary in behavior. Again this increases the computational complexity, but the increased set of hypotheses may be necessary to obtain acceptable tracking performance in some scenarios.

Ideally, if one had an infinite amount of computer power and memory, one would track the full distribution over the full history of both model and association (origin) hypotheses. However, due to computational limitations, approximations have to be made. In practice, too many approximations will of course reduce the accuracy of tracking, while too few approximations may not improve performance at all. To find the optimal level of approximation is an ongoing research field. While JPDA is popular algorithm, it has drawbacks when the number of targets become large, as the calculation of the exact marginal probabilities has exponential computational complexity. Loopy belief propagation (LBP) between the measurements and the targets has been shown as a possible way to speed up the calculation of these marginal association probabilities, and algorithms based on this have shown promising tracking performance and computational run time.

One of the standard ways of handling maneuvering targets is to use multiple models, with the algorithm of choice often being the interacting multiple models (IMM) [Bar-Shalom et al., 2001]. When including this in a multi target scenario, it is not consistent in the literature if authors include these models into the data association or not. Authors such as Chen and Tugnait [2001] include the models after claiming that there are interactions among the different models between different targets that need consideration. On the other hand, Musicki and Suvorova [2008] derive a version that does not need this by conditioning, and thereby keep the computational requirements of data association having multiple models, down to a single-target marginalization overhead. We are going to investigate this in terms of a hybrid state space, which both becomes the same as Musicki and Suvorova [2008].

How to label individual targets is something that is highly debated within the field of MTT using RFS. Authors such as Mahler [2007] claim that the problem gets harder when one labels the targets, Vo and Vo [2011] claims that it simplifies the problem, whereas García-Fernández et al. [2016] and Granström et al. [2018] in some way claim that it is not needed due to being inherent in the problem already. We will here argue for the latter, but see if we can keep some of the notions from the two former by labeling the sets instead of elements. Finding proper ways of doing this and interpreting it is important. How we model the real world using mathematics, to subsequently make reliable and tractable algorithms, can depend on the labeling technique and interpretation.

1.2 Report outline

Part I starts with Chapter 2 providing an introduction to probabilistic graphical models (PGMs) and LBP later to be used to calculated marginal association probabilities. Chapter 3 gives the necessary background in Bayesian state estimation including a novel derivation of the Kalman filter (KF), and how to deal with a hybrid state space. Chapter 4 is an introduction to RFS and gives a novel interpretation of labeling. Chapter 5 presents the Kullback-Leibler divergence (KL-divergence) and shows some standard statistical projection results.

Part II dels with the MTT properties under the standard model. Chapter 6 gives the assumption of the standard model and then interprets these in the setting of FISST. Chapter 7 provides a novel derivation of the Poisson multi Bernoulli mixture (PMBM) filter using these assumptions and shows how a hybrid state can be included.

Part III gives relations to other filters and gives some way of doing approximations to the intractable PMBM filter. Chapter 8 shows how the PMBM can be approximated through labeling and using the KL-divergence, to arrive at well known algorithms. Chapter 9 shows Williams and Roslyn [2014] approximative LBP algorithm for calculating marginal association probabilities, and chapter 10 points out some interesting aspects of the undetected targets, before chapter 11 finalizes and concludes this work.

Part I

Background Theory

2 Probabilistic graphical models

Probabilistic graphical models are a very convenient tool when dealing with distributions over several variables. They provide an easy way of understanding connections between variables and also can make inference and marginalization more tractable. Graphical models exploit factorization properties in probability density function (PDF)s that can simplify the problem at hand considerably. This section presents relevant background theory mostly based on Bishop [2016, Chapter 8]. Another reference that treats this in much more detail is Koller and Friedman [2009].

2.1 Bayesian networks

One of the most intuitive models is the so-called Bayesian network, which specifies the dependencies of the variables explicitly. A PDF of the form

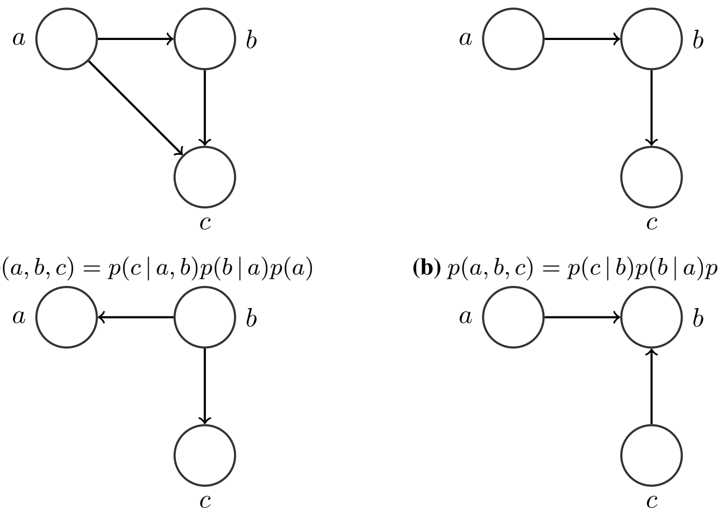
$$p(a, b, c) = p(c \mid a, b)p(b \mid a)p(a), \tag{2.1}$$

can be represented as a graph with three nodes and three edges. The representation uses one node for each variable and one edge per dependency and is a directed acyclic graph (DAG). The above distribution consists of three variables and three dependencies (two for $p(c \mid a, b)$ plus one for $p(b \mid a)$). The corresponding Bayesian network representation can be seen in Figure 2.1a.

The key in using this representation is that if, for instance, c is conditionally independent of a given b, the PDF can be written as

$$p(a, b, c) = p(c \mid a, b)p(b \mid a)p(a) = p(c \mid b)p(b \mid a)p(a), \tag{2.2}$$

and the graph can be simplified to Figure 2.1b. Then we can immediately see the dependencies just from looking at the graph. For instance, if b is now observed, we can get information about a and c, whereas observing a will not give any more information about c. Of course this can also be communicated through the math using Bayes rule, but it is



(a) $p(a, b, c) = p(c \mid a, b)p(b \mid a)p(a)$ **(b)** $p(a, b, c) = p(c \mid b)p(b \mid a)p(a)$

(c) $p(a, b, c) = p(a \mid b)p(c \mid b)p(b)$ **(d)** $p(a, b, c) = p(b \mid a, c)p(a)p(c)$

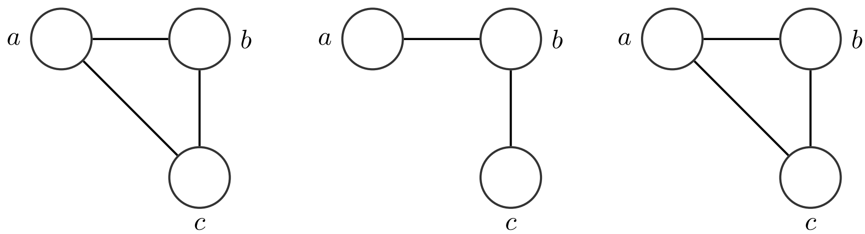
Figure 2.1: Bayesian Networks

usually more convenient to see this in a glance of the graph. It also tells us which variables we need knowledge about, in order to say something about another variable.

The representation is of course not unique, as one can decompose the conditions as one likes, and therefore get other representations and different graphs. For instance, using Bayes rule we have that $p(c \mid b)p(b \mid a)p(a) = p(c \mid b)p(a \mid b)p(b)$ for which the latter factorization can be seen as a graph in Figure 2.1c and is therefore an equivalent representation to Figure 2.1b. The last example demonstrates that if a and c are unconditionally independent, while b is dependent on both, we get the distribution $p(a, b, c) = p(b \mid a, b)p(a)p(b)$ shown by the graph in Figure 2.1d. This last representation describes a different set of dependencies than the others.

2.2 Markov Random Fields

The next model is the Markov random field (MRF) which is represented by undirected graphs in contrast to the directed graphs in Bayesian networks. The PDFs shown in Figure 2.1 are similarly shown as MRFs in Figure 2.2. To describe the distributions in a MRF, we need to form what is called potential functions over something called maximum cliques. A potential function, $\psi(\cdot)$ is simply any non-negative function over the variables, while a clique is a set of nodes where all the nodes are connected to each other (i.e. a fully connected sub graph). A maximum clique is thus just the cliques in the graph that are such that no other node can be added to the clique while still being a clique (adding another node to the subgraph will make the subgraph not fully connected). So for instance,



(a) $\frac{1}{Z_1}\psi_{a,b,c}(a,b,c)$ **(b)** $\frac{1}{Z_2}\psi_{a,b}(a,b)\psi_{b,c}(b,c)$ **(c)** $p(b\,|\,a,c)p(a)p(c)$

Figure 2.2: MRF with the given factorization

in Figure 2.2a the only maximum clique is the full graph, while in Figure 2.2b there are two maximum cliques consisting of (a, b) and (b, c) respectively.

Forming the potential functions for the three cases given in Figure 2.2 gives

$$p(a,b,c) = p(c\,|\,a,b)p(b\,|\,a)p(a) = \frac{1}{Z_1}\psi_{a,b,c}(a,b,c), \tag{2.3}$$

$$p(a,b,c) = \begin{Bmatrix} p(c\,|\,b)p(b\,|\,a)p(a) \\ p(c\,|\,b)p(a\,|\,b)p(b) \end{Bmatrix} = \frac{1}{Z_2}\psi_{a,b}(a,b)\psi_{b,c}(b,c), \tag{2.4}$$

$$p(a,b,c) = p(b\,|\,a,c)p(a)p(c) = \frac{1}{Z_3}\psi_{a,b,c}(a,b,c) \tag{2.5}$$

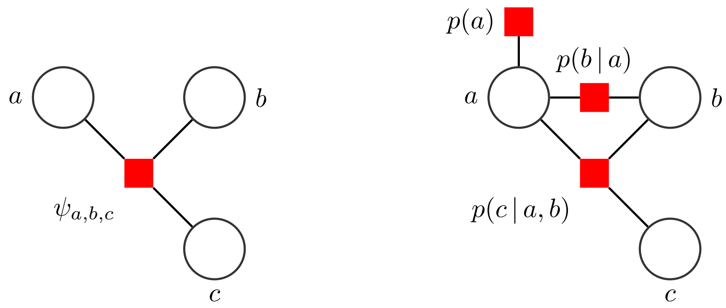
for Figure 2.2a, Figure 2.2b and Figure 2.2c respectively. We need the normalizing constants Z_i because the potential functions $\psi_{\cdot}(\cdot)$ in general do not have to be valid PDFs. The potential functions can in each case be seen to be

$$\psi_{a,b,c}(a,b,c) \propto p(a,b,c), \tag{2.6}$$

$$\psi_{a,b}(a,b) \propto p(b\,|\,a)p(a) = p(a\,|\,b)p(b), \tag{2.7}$$

$$\psi_{b,c}(b,c) \propto p(c\,|\,b), \tag{2.8}$$

and therefore, can be written as the conditional PDFs as for the Bayesian network. This is not always the case as one can use quite general factorizations, but it will always be possible when rewriting a Bayesian network as a MRF. The key difference is found by looking at Figure 2.1a in contrast to Figure 2.2a and seeing that the dependencies clearly shown in the first is no longer obvious in the latter. Representing the factorization given by the Bayesian network in Figure 2.1d as a MRF we lose even more conditional information, as one of the factors includes all three variables. In forming the potential functions we need to take this into account and therefore add an edge as seen in Figure 2.2c. The two slightly different factorizations in Figure 2.1b and Figure 2.1c will both be represented as the same when turned into a MRF, as shown in Figure 2.2c.



(a) Conditionally dependent **(b)** Conditionally independent

Figure 2.3: Factor Graph

2.3 Factor Graphs

Factor graphs are a third graphical description and can be seen as an extension of the MRF to give more explicit modelling of different dependencies. A factor graph is a graph where the factors also are given a node. They are therefore bipartite graphs (the graph nodes can be separated into two different sets that have no internal connections), where the first partition are the random variables (usually represented by circles), and the second are the factors (usually represented by squares). Again the representation is not unique as there are several ways to group the factors. The Markov field shown in Figure 2.2a can be represented as both the graphs shown in Figure 2.2, where the first is factored according to the factors in the MRF and the second is factored according to the Bayesian Network. This shows some of the expressive power of Factor graphs as they can explicitly show the modelled dependencies of a given problem. It also shows that a factor graph can be made easily from both Bayesian networks and MRFs by just considering the given factorization over conditionals or clique factors respectively.

2.4 Belief Propagation

Belief Propagation (BP), also sometimes known as the sum-product algorithm, is an algorithm for making inference in factor graphs when one wants to find marginal probabilities. The name belief propagation stems from the fact that one sends the beliefs as messages from the leaves to the other variables. So one starts at the leaves and sums over the variables in the factor there before making a product of these sums, and then repeating at the next layer. If we consider a case with say 31 binary variables and we want the marginal for one of them; if we naively were to sum over all the other variables that would amount to summing over $2^{30} \approx 10^9$ = one billion values. Compared to many real world problems, 31 binary variables is not 'big', so we can clearly see the need for more efficient algorithms.

BP exploits the graph structure, and hence the factorization of these problems and propagates beliefs from the leaves of the graph to the desired node(s). The limitation is that the graph cannot have loops (or has to be made into one without loops), for this algorithm to give guarantees on convergence and correctness.

Taking a PDF over X with a given factorization, such that the factors, $\psi_s(x_s)$ are taken over the subset of variables $x_s \subset X,\ s \in S$ such that $\bigcup_{s\in S} x_s = X$. Then the PDF can be written as

$$p(x) = \frac{1}{Z}\prod_{s\in S}^{N} \psi_s(x_s). \tag{2.9}$$

Say we are interested in the marginal distribution of the variable x_i. The node corresponding to x_i will have a set of factors that are associated with it, that each is the root of a sub tree disconnected from all the other factors associated with x_i by assumption of no loops. To simplify the notation slightly, we can group the factors to the joint over this kind of sub tree to $\Psi_s(x_i, X_s)$. Here Ψ is the product of all the factors of that subtree including the neighbouring factor of x_i, and X_s is the set of variables that are in that particular sub tree. Using $\mathrm{ne}(x)$ to denote the neighbours of x, $X \setminus x$ to denote the set X with x taken out, marginalizing to give the variable x_i and expanding gives

$$p(x_i) \propto \sum_{X\setminus x_i}\prod_{s\in S}^{N} \psi_s(x_s) = \prod_{s\in\mathrm{ne}(x_i)}\sum_{X_s}\Psi_s(x_i, X_s) \tag{2.10}$$

$$\propto \prod_{s\in\mathrm{ne}(x_i)}\sum_{x_{\mathrm{ne}(\psi_s)\setminus x_i}} \psi_s(x_i, x_{\mathrm{ne}(\psi_s)\setminus x_i}) \prod_{(m\in\mathrm{ne}(\psi_s)\setminus x_i)}\ \prod_{(l\in\mathrm{ne}(x_m)\setminus\psi_s)}\sum_{X_{ml}}\Psi_l(x_m, X_{ml}). \tag{2.11}$$

From the first to the second line, we see that the expansion has led to a repetition in structure as the last factors in the second line are of the same type as the factors of the first line. We have moved out to the neighbouring variables and are considering the new subtrees starting from their factors. As expected we can continue this expansion all the way throughout the graph to get the full marginalization. It also shows that this is indeed a set of products of sums, where the algorithm gets its name sum-product algorithm from.

To develop the algorithm one uses the notion of messages. Two types of messages are used, one from variables to factors and one from factors to variables. The message from the factors $\psi_s(x_s)$ to the variable x_i in it, is defined as the marginal for x_i in the factor over that sub tree with root at ψ_s;

$$\mu_{\psi_s\to x_i}(x_i) = \sum_{X_s}\Psi_s(x_i, X_s) = \sum_{x_{\mathrm{ne}(\psi_s)\setminus x_i}} \psi_s(x_i, x_{\mathrm{ne}(\psi_s)\setminus x_i}) \prod_{m\in\mathrm{ne}(\psi_s)\setminus x_i} \mu_{x_m\to\psi_s}(x_m), \tag{2.12}$$

and the message from a variable x_m to a factor ψ_s is defined as the product of the remaining marginal factors;

$$\mu_{x_m\to\psi_s}(x_m) = \prod_{l\in\mathrm{ne}(x_m)\setminus\psi_s}\sum_{X_{ml}}\Psi_l(x_m, X_{ml}) = \prod_{l\in\mathrm{ne}(x_m)\setminus\psi_s}\mu_{\psi_l\to x_m}(x_m). \tag{2.13}$$

From these two definitions it is clear that calculation of the messages can be done recursively through the graph as the messages are dependent on similar messages. Rewriting the marginal distribution over x_i in terms of the messages, now gives

$$p(x_i) \propto \prod_{s \in \mathrm{ne}(x_i)} \mu_{\psi_s \to x_i}(x_i) = \mu_{x_i \to \psi_d}(x_i)\mu_{\psi_d \to x_i}(x_i) \ \forall\, d \in \mathrm{ne}(x_i), \tag{2.14}$$

where the normalization constant can be found by summing over x_i again. It is here assumed that the variables are discrete, but when there are continuous variables the summations can be replaced by integrals where needed. In addition, x_i is any variable, so the algorithm will also work for any wanted variable. Moreover we have that the messages will be the same for all variables and the only thing needed is that a node needs to have all the other incoming messages before it can pass along its message. We can therefore start at the outer rim of the graph and propagate the messages through the graph until all messages have reached all nodes. In this way we have calculated the messages in the graph and can thus form the marginal of all the variables.

The leaves can either be a variable or a factor. In the case of a variable the message will simply be a 1, and in case of a factor, it will simply be the factor. In case we want to condition on some variables, this is also done simply by not summing over those particular nodes, and using the value assumed or observed. Hence we are able to include the information available in the conditioning.

This algorithm is shown to give the exact marginals as long as the graph does not contain any loops. It has been extended to work exact on graphs with loops, with the resulting algorithm called the junction tree algorithm, where the loops are basically dealt with by clustering them into “bigger” nodes. However if the nodes get too big, the marginalization over them quickly becomes intractable.

2.5 Loopy Belief Propagation

In loopy belief propagation (LPB) one applies the BP even though there are loops. Since we now cannot know the input from all the other messages from before (the messages from one variable in the loop to another will be dependent on the variable itself) we just have to assume some value (like unity) and then try. The message will then be passed around the loop and will most likely come back as something different than assumed in the first place. Since it was based on this value, we now have to pass this new corrected message as the earlier one was wrong. This value was still dependent on what was given by this message in the first place, so it is likely to change again after one round. Hence we have to do several rounds of BP. In general this is not guaranteed to converge, and when it does, it is not even guaranteed that it gives the right value.

There exists some conditions for which LBP converges. For instance, if it can be shown that a full cycle of the message passing is a contraction mapping, we know that it has to converge to one of the fixed points. There are also other conditions, which are not discussed here.

3 Bayesian state estimation

State estimation is a major component of MTT. Multi-target tracking can in many ways be seen as an extension of single target tracking, with the main difference lying in the ambiguity surrounding measurement origin in MTT. Tracking a single target when there are no erroneous measurements [1] and with unity detection probability, becomes standard state estimation when the input to the system is unknown.

We therefore start with the derivation of Bayesian state estimation, and the special case linear Gaussian dynamics and measurement model, which results in the KF. Next we take a look at how to handle hybrid state spaces within the Bayesian framework. This provides the basis for a discussion of MTT in chapter 7

3.1 Bayesian state estimation

A general dynamical system for x with known exogenous inputs u, unknown exogenous inputs v, measurement y and measurement noise w can be written as

$$x_0 = x(t = 0), \qquad x_0 \sim p_{x_0}(x_0) \tag{3.1}$$

$$x_t = f(x_{t_-}, u_{t_-}, v_{t_-}), \qquad v_{t_-} \sim p_v(v_{t_-}) \tag{3.2}$$

$$z_t = h(x_t, u_t, w_t), \qquad w_t \sim p_w(w_t) \tag{3.3}$$

where subscript t on the variables indicates the discrete time, and $p_\sigma(\xi)$ are PDF over σ evaluated at ξ. The subscript σ will sometimes be omitted when it is clear from the context which variable the PDF comes from. Knowing the PDF we can do estimation in this system by using the total probability theorem between time steps and Bayes theorem when measurements are received.

[1] Not to be confused with measurement noise, as this means that the sensor detected something that is not there

3.1.1 Prediction

From the state distribution at a given time we often want to predict the distribution at the next time step. This can be done by using the total probability theorem, where the marginalization is done over the variables from the previous time step.

First we see that given x_{t_-}, the only unknown variable in (3.2) is v_{t_-}. Abusing notation and assuming f in (3.2) is invertible w.r.t v_{t_-} we can use the change of variable formula for PDFs to get

$$p_{x_t \,|\, x_{t_-}}(x_t \,|\, x_{t_-}) = p_v(f^{-1}(x_{t_-}, u_{t_-}, x_t)) \left| \det \left(\frac{\partial v_{t_-}}{\partial x_t} \right) \right|. \tag{3.4}$$

If the function f is not invertible (i.e. non-injective but surjective in v, pointwise in x and u), it will still be piecewise invertible in v. In that case (3.4) has to be replaced by a sum over the piecewise inverses that exist for a particular value of x_t. Intuitively, if two or more different values of v map to the same value in x_t, we need to take into account all the events in v that give the particular x_t.

For clarity and simplicity, the PDF in (3.4) will hence be denoted as

$$f(x_t \,|\, x_{t_-}) \tag{3.5}$$

and is read as the density function of x_t given the state at time $t-1$. The notation

$$f_{t \,|\, t'}(x_t) \tag{3.6}$$

will also be used to denote the PDF of x_t at time t given all information available up to and including t'.

Using this notation and the total probability theorem to form the joint distribution over x_t and x_{t_-}, then marginalizing over x_{t_-} we get the predicted distribution for x_t;

$$f_{t \,|\, t_-}(x_t) = \int\limits_{x_{t_-} \in \mathcal{X}} f(x_t \,|\, x_{t_-})\, f_{t_- \,|\, t_-}(x_{t_-})\, \mathrm{d}x_{t_-}, \tag{3.7}$$

where $\mathcal{X}$, namely the state space, denotes the set of possible states x can have.

3.1.2 Measurement update

When we get a new measurement we want to update our state distribution in a statistically optimal way. This is done using Bayes theorem. Knowing the distribution for the state x_t and the conditional distribution for the measurement z_t given x_t, Bayes rule with our notation becomes

$$f_{t \,|\, t}(x_t) = \frac{h(z_t \,|\, x_t)\, f_{t \,|\, t_-}(x_t)}{p(z_t)} = \frac{h(z_t \,|\, x_t)\, f_{t \,|\, t_-}(x_t)}{\int\limits_{x_t \in \mathcal{X}} h(z_t \,|\, x_t)\, f_{t \,|\, t_-}(x_t)\, \mathrm{d}x_t}. \tag{3.8}$$

Note that the denominator is simply a normalizing constant in terms of x_t. The PDF $h(z_t \mid x_t)$ is calculated in the same manner as (3.4) assuming invertibility, or in the case of non-invertibility is the summands of the piece wise inverses.

3.1.3 State estimation as a probabilistic graphical model

Since the dynamical system is treated using random variables, it can be formulated on a PGM. This is done in Figure 3.1, where the gray variables is signaling that we are conditioning on them. Figure 3.1a is the Bayesian network, and clearly shows how we condition in the problem. The conditional PDF is here written explicitly in the Bayesian network for clarity, whereas this is usually just implicit from the edge directions. Figure 3.1b shows the factor graph where we can clearly see the factors (constraints, or functions) involved, and the difference from the Bayesian network is the factor over the prior $f(x_0)$. These PGMs give a clear and good graphical description of what we were doing in the last two subsections. It is an easy way of understanding the dependencies and the information flow between the involved variables.

Let us see what happens if we apply belief propagation (BP) to this graph. Notice that this graph has no loops, and BP should give an exact answer. The algorithm is given for disrete variables, but can be easily extended to continuous variables by applying integrals instead of sums. The limitation is of course that these integrations may not be solvable parametrically, and in general one has to rely on numerics to solve them. Nevertheless it can be "solved" both in conceptually and theoretically.

The general BP messages, (2.12) and (2.13) are restated here for reference:

$$\mu_{\psi_s \to x_i}(x_i) = \sum_{X_s} \Psi_s(x_i, X_s) = \sum_{x_{\mathrm{ne}(\psi_s)\setminus x_i}} \psi_s(x_i, x_{\mathrm{ne}(\psi_s)\setminus x_i}) \prod_{m \in \mathrm{ne}(\psi_s)\setminus x_i} \mu_{x_m \to \psi_s}(x_m), \tag{3.9}$$

$$\mu_{x_m \to \psi_s}(x_m) = \prod_{l \in \mathrm{ne}(x_m)\setminus \psi_s} \sum_{X_{ml}} \Psi_l(x_m, X_{ml}) = \prod_{l \in \mathrm{ne}(x_m)\setminus \psi_s} \mu_{\psi_l \to x_m}(x_m). \tag{3.10}$$

We can start off by forming the message from the factor $f_{0|0}(x_0)$, with shorthand notation f_0, to the variable x_0 in the factor graph of Figure 3.1b. By looking at (3.9) we can see that the factor f_0 only contains the variable we are sending a message to. Hence the message reduces to the factor itself;

$$\mu_{f_0 \to x_0}(x_0) = f_{0|0}(x_0). \tag{3.11}$$

Continuing with the message from x_0 to $f(x_1|x_0)$ (shorthand: $f_{1,0}$) we see from (3.10) that it is the product of all the messages to that node, except the one coming from the one we are sending a message to. This message therefore becomes the single other incoming message;

$$\mu_{x_0 \to f_{1,0}}(x_0) = \mu_{f_0 \to x_0}(x_0) = f_{0|0}(x_0). \tag{3.12}$$

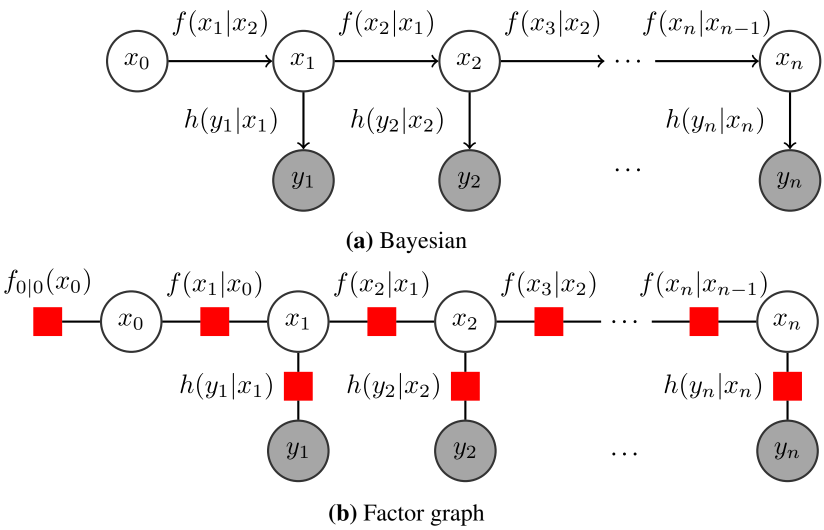


Figure 3.1: General networks for a dynamical system

This may not be a very interesting result, but it makes sense. The contribution from x_0 is its prior.

Again continuing this, we will discover something familiar. Next up is the message from the factor $f_{1,0}$ to the node x_1. The general formula gives that it is the of the product of the factor multiplied with the message coming from all the other connected nodes marginalized with respect to all other nodes, that is the factor $f_{1,0}$ multiplied with the message from x_0 to $f_{1,0}$ marginalized with respect to x_0. Doing this becomes;

$$\mu_{f_{1,0}\to x_1}(x_1) = \int f(x_1|x_0)\mu_{x_0\to f_{1,0}}(x_0)\,\mathrm{d}x_0 = \int f(x_1|x_0)f_{0|0}(x_0)\,\mathrm{d}x_0 = f_{1|0}(x_1), \tag{3.13}$$

and should be recognized as the prediction step, (3.7) of the Bayesian state estimation filter.

The message from y_1 to x_1 through $h_1 = h(y_1|x_1)$ is dependent on if we want the observed version or the marginal version. Since there are no prior over the y's the marginalized version of the message to x_1 will simply be the 1, and hence have no contribution to the belief over x_1. The observed version on the other hand, uses the information contained in observing the variable and will be similar to conditioning, as will become clear next. The variable y_1 has no other factors than h_1, and the message from y_1 to h_1 is thus simply given by

$$\mu_{y_1\to h_1}(y_1) = 1. \tag{3.14}$$

Similarly as for the message from $f_{1,0}$ to x_1 the (marginalized) message from h_1 to x_1 is

$$\mu_{h_1\to x_1}(x_1) = \int h(y_1|x_1)\mu_{y_1\to h_1}(y_1)\,\mathrm{d}y_1 = \int h(y_1|x_1)\cdot 1\,\mathrm{d}y_1 = 1, \tag{3.15}$$

but the observed message from h_1 to y_1 is

$$\mu_{h_1\to x_1}(x_1, y_1) = h(y_1|x_1)\mu_{y_1\to h_1}(y_1) = h(y_1|x_1). \tag{3.16}$$

The next message to compute is the message from x_1 to $f_{2,1} = f(x_2|x_1)$, and will be the product of the incoming messages from $f_{1,0}$ and h_1. For the case of marginalized y_1 this messages becomes

$$\mu_{x_1\to f_{2,1}}(x_1) = \mu_{f_{1,0}\to x_1}(x_1)\mu_{h_1\to x_1}(x_1) = f_{1|0}(x_1), \tag{3.17}$$

and for the case of observing y_1 this becomes

$$\mu_{x_1\to f_{2,1}}(x_1) = \mu_{f_{1,0}\to x_1}(x_1)\mu_{h_1\to x_1}(x_1, y_1) = h(y_1|x_1)f_{1|0}(x_1) \propto f_{1|1}(x_1). \tag{3.18}$$

We can now see that calculating the messages between the states becomes an non-normalized version of the Bayes filter, where the messages from $f_{t,t-1}$ to x_t give the prediction, and the messages from x_t to $f_{t+1,t}$ give the measurement update in the case of including a measurement (this can be generalized in a straightforward manner to the case of several measurements), and doing nothing in the case of not including any more information. The difference in using BP is that the messages are not necessarily normalized and that this has to be done explicitly whenever the true marginal distribution is needed.

3.2 Kalman filter

The standard discrete KF assumes a linear dynamic and measurement model with additive Gaussian noise, that is uncorrelated between dynamics and measurement, between time steps, and an initial state that is also Gaussian distributed

$$x_t = Ax_{t\_} + Bu_{t\_} + v_{t\_}, \qquad \begin{cases} v_{t\_} \sim \mathcal{N}(v_{t\_}; 0, Q_{t\_}) \\ x_{t\_} \sim \mathcal{N}(x_{t\_}; \hat{x}_{t\_}, P_{t\_}) \end{cases}, \tag{3.19}$$

$$y_t = Cx_t + w_t, \qquad w_t \sim \mathcal{N}(w_t; 0, R_t). \tag{3.20}$$

This model can make the KF somewhat limiting. However, it can be extended to work in the correlated case, in exchange for increased complexity, and in the nonlinear case by linearization around the expected values, in exchange for non-optimality.

Here $\mathcal{N}(x; \mu, \Sigma)$ denotes the Gaussian distribution, sometimes called the normal distribution, over x with mean μ and covariance Σ, such that

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \tag{3.21}$$

3.2.1 The product identity

KF assumes a Gaussian distribution, therefore we need to consider some properties of the given model equations. We start with the general model used in both the linear-Gaussian dynamical equation (3.19) and the linear-Gaussian measurement equation (3.20);

$$\gamma = G\lambda + \sigma \tag{3.22}$$

where G is a given transformation matrix and the other variables are Gaussian distributed vectors of appropriate sizes. The expected values are given as $\mathbb{E}[\gamma] = \bar{\gamma}$, $\mathbb{E}[\lambda] = \bar{\lambda}$ and $\mathbb{E}[\sigma] = \bar{\sigma} = 0$, and the covariances given as $\mathrm{Cov}(\gamma) = \Gamma$, $\mathrm{Cov}(\lambda) = \Lambda$ and $\mathrm{Cov}(\sigma) = \Sigma$. λ and σ are assumed to be independent.

In this setting, γ and λ are the vectors of interest, and σ is some kind of (unwanted) noise. To derive the distributions of either of them, we need to find their joint distribution. Moving $G\lambda$ over to left hand side and taking λ as given, we see that the the resulting left hand side follows the distribution of σ;

$$p_{\gamma|\lambda}(\gamma|\lambda) = p_\sigma(\gamma - G\lambda). \tag{3.23}$$

The joint distribution will therefore be

$$\begin{aligned} p_{\gamma,\lambda}(\gamma,\lambda) &= p_{\gamma\,|\,\lambda}(\gamma\,|\,\lambda)p_\lambda(\lambda) = p_\sigma(\gamma - G\lambda)p_\lambda(\lambda) \\ &= \mathcal{N}\left(\gamma - G\lambda;\ 0,\ \Sigma\right)\mathcal{N}\left(\lambda;\ \bar{\lambda},\ \Lambda\right) \\ &= \mathcal{N}\left(\gamma;\ G\lambda,\ \Sigma\right)\mathcal{N}\left(\lambda;\ \bar{\lambda},\ \Lambda\right). \end{aligned} \tag{3.24}$$

It will be shown that

$$\bar{\gamma} = G\bar{\lambda} + \bar{\sigma} = G\bar{\lambda}, \tag{3.25}$$

$$\Gamma = G\Lambda G^T + \Sigma, \tag{3.26}$$

$$\hat{\lambda} = \mathbb{E}[\lambda\,|\,\gamma] = \bar{\lambda} + \Lambda G^T(G\Lambda G^T + \Sigma)^{-1}(\gamma - G\bar{\lambda}), \tag{3.27}$$

$$\hat{\Lambda} = \mathrm{Cov}(\lambda\,|\,\gamma) = \Lambda - \Lambda G^T(G\Lambda G^T + \Sigma)^{-1}G\Lambda. \tag{3.28}$$

and that therefore the Gaussian product identity holds;

$$\mathcal{N}\left(\gamma;\ G\lambda,\ \Sigma\right)\mathcal{N}\left(\lambda;\ \bar{\lambda},\ \Lambda\right) = \mathcal{N}\left(\gamma;\ \bar{\gamma},\ \Gamma\right)\mathcal{N}\left(\lambda;\ \hat{\lambda},\ \hat{\Lambda}\right). \tag{3.29}$$

The product identity can be derived by algebraic manipulation of the quadratic exponential in the Gaussian product. We already know that the normalizing constant must follow, since (3.24) is a valid probability distribution. In any case, one can use the same following derivation to show that the normalization constant is also given by this new covariance matrix. This is done by seeing that the determinants of the covariances will form a product. Knowing that the determinant of a block triangular or diagonal matrix is also the product of the determinants of the diagonal blocks, one can form the same matrix as in the exponent. The steps for finding the means and the covariances of the product identity are shown by (3.30) where the last equality defines the given values. Marginalizing over λ or conditioning on γ will now simply result in the removal of one of the terms. This also shows that the joint, marginal and conditional distributions are Gaussian as well.

$$
\begin{aligned}
&(\gamma - G\lambda)^T \Sigma^{-1} (\gamma - G\lambda) + (\lambda - \bar{\lambda})^T \Lambda^{-1} (\lambda - \bar{\lambda}) \\
&= \begin{bmatrix} \gamma - G\lambda \\ \lambda - \bar{\lambda} \end{bmatrix}^T \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Lambda^{-1} \end{bmatrix} \begin{bmatrix} \gamma - G\lambda \\ \lambda - \bar{\lambda} \end{bmatrix} \\
&= \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix}^T \begin{bmatrix} \mathbb{I} & 0 \\ -G^T & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & \Lambda^{-1} \end{bmatrix} \begin{bmatrix} \mathbb{I} & -G \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \\
&= \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix}^T \left(\begin{bmatrix} \mathbb{I} & G \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ G^T & \mathbb{I} \end{bmatrix} \right)^{-1} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \\
&= \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix}^T \begin{bmatrix} G\Lambda G^T + \Sigma & G\Lambda \\ \Lambda G^T & \Lambda \end{bmatrix}^{-1} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \\
&= \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix}^T \left(\begin{bmatrix} \mathbb{I} & 0 \\ \Lambda G^T (\Sigma + G\Lambda G^T)^{-1} & \mathbb{I} \end{bmatrix} \cdots \right. \\
&\qquad \begin{bmatrix} \Sigma + G\Lambda G^T & 0 \\ 0 & \Lambda - \Lambda G^T (\Sigma - G\Lambda G^T)^{-1} G\Lambda \end{bmatrix} \cdots \\
&\qquad\qquad \left. \begin{bmatrix} \mathbb{I} & (\Sigma + G\Lambda G^T)^{-1} G\Lambda \\ 0 & \mathbb{I} \end{bmatrix} \right)^{-1} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \\
&= \left(\begin{bmatrix} \mathbb{I} & 0 \\ -\Lambda G^T (\Sigma + G\Lambda G^T)^{-1} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \right)^T \cdots \\
&\qquad \begin{bmatrix} \Sigma + G\Lambda G^T & 0 \\ 0 & \Lambda - \Lambda G^T (\Sigma - G\Lambda G^T)^{-1} G\Lambda \end{bmatrix}^{-1} \cdots \\
&\qquad\qquad \left(\begin{bmatrix} \mathbb{I} & 0 \\ -\Lambda G^T (\Sigma + G\Lambda G^T)^{-1} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} \end{bmatrix} \right) \\
&= \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} - \Lambda G^T (\Sigma + G\Lambda G^T)^{-1} (\gamma - G\bar{\lambda}) \end{bmatrix}^T \cdots \\
&\qquad \begin{bmatrix} \Sigma + G\Lambda G^T & 0 \\ 0 & \Lambda - \Lambda G^T (\Sigma - G\Lambda G^T)^{-1} G\Lambda \end{bmatrix}^{-1} \cdots \\
&\qquad\qquad \begin{bmatrix} \gamma - G\bar{\lambda} \\ \lambda - \bar{\lambda} - \Lambda G^T (\Sigma + G\Lambda G^T)^{-1} (\gamma - G\bar{\lambda}) \end{bmatrix} \\
&= \begin{bmatrix} \gamma - \bar{\gamma} \\ \lambda - \hat{\lambda} \end{bmatrix}^T \begin{bmatrix} \Gamma^{-1} & 0 \\ 0 & \hat{\Lambda}^{-1} \end{bmatrix} \begin{bmatrix} \gamma - \bar{\gamma} \\ \lambda - \hat{\lambda} \end{bmatrix},
\end{aligned} \tag{3.30}
$$

3.2.2 Prediction

Predicting the PDF of x_{t_-} for the subsequent time step t, amounts to finding the PDF of x_t based on the PDFs of x_{t_-} and v_{t_-}, and not their true values. This is done using the product identity (3.29) and the prediction rule (3.7) and marginalizing out the variables from the previous time step. By using $x_t = \gamma$, $x_{t_-} = \lambda$, $v = \sigma$, $A = G$, $\Gamma = P_{t\,|\,t}$, $P_{t\,|\,t_-} = \Lambda$ and $Q_{t_-} = \Sigma$, we get

$$
\begin{aligned}
f_{t\,|\,t_-}(x) &= \int\limits_{x_{t_-}\in\mathbb{R}} \mathcal{N}\left(x_t;\ Ax_{t_-} + Bu_{t_-},\ Q_{t_-}\right)\mathcal{N}\left(x_{t_-};\ \hat{x}_{t_-\,|\,t_-},\ P_{t_-\,|\,t_-}\right)\,\mathrm{d}x_t \\
&= \int\limits_{x_{t_-}\in\mathbb{R}} \mathcal{N}\left(x_t;\ \hat{x}_{t\,|\,t_-},\ P_{t\,|\,t_-}\right)\mathcal{N}\Big(x_{t_-};\ \mathbb{E}[x_{t_-}\,|\,x_t]\,,\ \mathrm{Cov}(x_{t_-}|x_t)\Big)\,\mathrm{d}x \\
&= \mathcal{N}\left(x_t;\ \hat{x}_{t\,|\,t_-},\ P_{t\,|\,t_-}\right),
\end{aligned} \tag{3.31}
$$

where $\hat{x}_{t\,|\,t_-}$ denotes the mean and $P_{t\,|\,t_-}$ denotes the covariance of x_t given the information up to and including $t-1$ and so on.

The mean and covariance of x_{t+1} are thus calculated as;

$$
\hat{x}_{t\,|\,t_-} = A\hat{x}_{t_-\,|\,t_-} + Bu_{t_-}, \tag{3.32}
$$

$$
P_{t\,|\,t_-} = AP_{t_-\,|\,t_-}A^T + Q_{t_-}. \tag{3.33}
$$

These equations are the standard discrete KF equations for the prediction step.

3.2.3 Measurement update

If we have a prior distribution of x_t and receive a measurement z_t, by conditioning on this measurement we can obtain a distribution with a smaller covariance, as more information is included. This is also done using the product identity (3.29), however for measurement update we will of course use the measurement update rule (3.8). We now let $z_t = \gamma$, $x_t = \lambda$, $w_t = \sigma$, $\hat{x}_{t\,|\,t_-} = \bar{\lambda}$, $\hat{x}_{t\,|\,t} = \mathbb{E}[x_t\,|\,z_t] = \hat{\lambda}$, $C = G$, $P_{t\,|\,t} = \mathrm{Cov}(x_t\,|\,z_t) = \hat{\Lambda}$, $P_{t\,|\,t_-} = \Lambda$, $R_t = \Sigma$ and $S_t = \Gamma$ and thus get

$$
f_{t\,|\,t}(x_t) = \frac{\mathcal{N}\left(z_t;\ Cx_t,\ R_t\right)\mathcal{N}\left(x_t;\ \hat{x}_{t\,|\,t_-},\ P_{t\,|\,t_-}\right)}{p(z_t)} \tag{3.34}
$$

$$
= \frac{\mathcal{N}\left(z_t;\ \hat{z}, S_t\right)\mathcal{N}\left(x_t;\ \hat{x}_t, \hat{P}_t\right)}{\mathcal{N}\left(z_t;\ \hat{z}_t, S_t\right)} \tag{3.35}
$$

$$
= \mathcal{N}\left(x_t;\ \hat{x}_{t\,|\,t},\ P_{t\,|\,t}\right), \tag{3.36}
$$

with

$$\begin{aligned}\hat{x}_{t\,|\,t} &= \hat{x}_{t\,|\,t_-} + P_{t\,|\,t_-}C^T(CP_{t\,|\,t_-}C^T + R_t)^{-1}(y_t - C\hat{x}_{t\,|\,t_-}) \\ &= \hat{x}_{t\,|\,t_-} + K_t(y_t - C\hat{x}_{t\,|\,t_-}),\end{aligned} \tag{3.37}$$

$$\begin{aligned}P_{t\,|\,t} &= P_{t\,|\,t_-} - P_{t\,|\,t_-}C^T(CP_{t\,|\,t_-}C^T + R_t)^{-1}CP_{t\,|\,t_-} \\ &= P_{t\,|\,t_-} - K_tS_tK_t^T \\ &= (\mathbb{I} - K_tC)P_{t\,|\,t_-}(\mathbb{I} - K_tC)^T + K_tRK_t^T,\end{aligned} \tag{3.38}$$

$$K_t = P_{t\,|\,t_-}C^T(CP_{t\,|\,t_-}C^T + R_t)^{-1} = P_{t\,|\,t_-}C^TS_t^{-1}, \tag{3.39}$$

which are the standard discrete KF measurement update equations. As shown, the covariance update equation (3.38) can be put into several different forms where the last two equations are often used to help an implementation keep the covariance matrix symmetric.

3.3 Hybrid state space

Sometimes the state space consists of a continuous state, $x \in \mathcal{X} \subseteq \mathbb{R}^{d_c}$, and a discrete state, $l \in \mathcal{L} \subseteq \mathbb{N}^{d_d}$, where d_c and d_d are the numbers of dimensions of the continuous and discrete parts of the state space, respectively. Such a state space is called a hybrid state space. Bayesian estimation is optimal on continuous, discrete and hybrid state variables, so the theory above is also valid for discrete and hybrid state spaces, assuming the correct integrals are interchanged with summations. However, mixing continuous and discrete random variables tends not to give simple distributions to work with, and one can treat this in several ways. How to do this in practice is thus not necessarily so obvious, and we will go through one way it can be done here.

We will be tracking two separate PDFs, one for the discrete variables $\mu^l_{t\,|\,t'} \triangleq \mathrm{P}\left(l_t = l\,|\,t'\right)$ and one for the continuous variables conditioned on the discrete variables $f^l_{t\,|\,t'}(x) \triangleq f_{t\,|\,t'}(x_t\,|\,l_t)$. From a joint distribution, $f_{t\,|\,t'}(x, l)$ one can separate it into this conditioning according to

$$\mu^l_{t\,|\,t'} = \int\limits_{x\in\mathcal{X}} f_{t\,|\,t'}(x, l)\,\mathrm{d}x \tag{3.40}$$

$$f^l_{t\,|\,t'}(x) = \frac{f_{t\,|\,t'}(x, l)}{\mu^l_{t\,|\,t'}}. \tag{3.41}$$

The transition function is going to be modeled as

$$f\left(x_t, l_t\,|\,x_{t_-}, l_{t_-}\right) = f^l\left(x_t\,|\,x_{t_-}, l_{t_-}\right)\pi^{l_t}_{l_{t_-}}\left(x_{t_-}\right) \tag{3.42}$$

$$= f^l\left(x_t\,|\,x_{t_-}\right)\pi^{l_t}_{l_{t_-}}\left(x_{t_-}\right), \tag{3.43}$$

where there is and explicit independence between x_t and l_{t_-} when l_t is given. It is possible to relax this independence assumption, but the mixing step will then be problematic, and one might have to do the prediction in one step.

Given these two distributions at time $t-1$ given information up to and including time $t-1$, and applying the prediction step rule, we can write the prediction step for the discrete variables as

$$\mu^l_{t\,|\,t_-} = \sum_{l_{t_-}\in\mathcal{L}} \int_{x_{t_-}\in\mathcal{X}} \pi^{l_t}_{l_{t_-}}(x_{t_-})\, f^l_{t_-\,|\,t_-}(x_{t_-})\, \mu^l_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}. \tag{3.44}$$

To predict the continuous states to get $f^l_{t\,|\,t_-}(x_t)$ we are going to do an intermediate step of conditioning x_{t_-} on l_t to obtain $f_{t_-\,|\,t_-}(x_{t_-}\,|\,l_t)$, which will be called mixing. To see why this is a good idea, we can consider

$$\begin{aligned} f^l_{t\,|\,t_-}(x_t) \triangleq f_{t\,|\,t_-}(x_t\,|\,l_t) &= \sum_{l_{t_-}\in\mathcal{L}} \int_{x_{t_-}\in\mathcal{X}} \frac{f_{t\,|\,t_-}(x_t, l_t, x_{t_-}, l_{t_-})}{\mu^l_{t\,|\,t_-}}\, \mathrm{d}x_{t_-} \\ &= \int_{x_{t_-}\in\mathcal{X}} f^l(x_t\,|\,x_{t_-}) \sum_{l_{t_-}\in\mathcal{L}} \pi^{l_t}_{l_{t_-}}(x_{t_-})\, f^l_{t_-\,|\,t_-}(x_{t_-})\, \frac{\mu^l_{t_-\,|\,t_-}}{\mu^l_{t\,|\,t_-}}\, \mathrm{d}x_{t_-} \\ &= \int_{x_{t_-}\in\mathcal{X}} f^l(x_t\,|\,x_{t_-})\, f_{t_-\,|\,t_-}(x_{t_-}\,|\,l_t)\, \mathrm{d}x_{t_-}, \end{aligned} \tag{3.45}$$

where we have gotten the wanted mixing PDF

$$f_{t_-\,|\,t_-}(x_{t_-}\,|\,l_t) = \sum_{l_{t_-}\in\mathcal{L}} \pi^{l_t}_{l_{t_-}}(x_{t_-})\, f^l_{t_-\,|\,t_-}(x_t)\, \frac{\mu^l_{t_-\,|\,t_-}}{\mu^l_{t\,|\,t_-}}. \tag{3.46}$$

We can clearly see that this distribution is a valid PDF and a mixture. The optimal thing to do is to keep the mixture. However it will grow in size through time and one therefore needs to do mixture reduction to make the estimation algorithm feasible. As pointed out by Blom and Bar-Shalom [1988], doing the hypothesis reduction in the PDF $f_{t_-\,|\,t_-}(x_{t_-}\,|\,l_t)$ is ideal for systems with linear Gaussian continuous part and an independent jump Markov discrete part. This is due to the fact that doing it before the mixture reduces the information available to the continuous state from the specific discrete state that we are going to predict. Whereas doing it later, i.e. after prediction, increases computational cost and due to linearity in the linear Gaussian case, will be equivalent to doing it before.

In the nonlinear or non Gaussian case it is less obvious, but nevertheless a good idea to retain as much information as possible for as long as possible. The main computational load will be in the continuous state prediction and the update step. Thus doing mixture reduction straight after the mixing step (which is also the source of the increasing number of mixture components) does still seem like the best option in the general case, but further investigation is needed to check if this is indeed the case. Special dynamics or distributions have to be considered on their own anyway.

After having done (3.46), one does the prediction and update steps for each discrete state individually. The prediction step is shown in (3.45), and the update is given by

$$f^l_{t\,|\,t}(x)\,\mu^l_{t\,|\,t} = \frac{h^l(z_t\,|\,x)\, f^l_{t\,|\,t_-}(x)\,\mu^l_{t\,|\,t_-}}{\displaystyle\sum_{l_t\in\mathcal{L}}\int_{x\in\mathcal{X}} h^l(z_t\,|\,x)\, f^l_{t\,|\,t_-}(x)\,\mathrm{d}x\mu^l_{t\,|\,t}} = \frac{h^l(z_t\,|\,x)\, f^l_{t\,|\,t_-}(x)\,\mu^l_{t\,|\,t_-}}{\displaystyle\sum_{l_t\in\mathcal{L}} h^l(z_t)\,\mu^l_{t\,|\,t_-}}$$
$$= \left[\frac{h^l(z_t\,|\,x)\, f^l_{t\,|\,t_-}(x)}{h^l(z_t)}\right]\left[\frac{h^l(z_t)\,\mu^l_{t\,|\,t_-}}{\displaystyle\sum_{l_t\in\mathcal{L}} h^l(z_t)\,\mu^l_{t\,|\,t_-}}\right], \tag{3.47}$$

where the bracketed terms on the right hand side (RHS) correspond to the terms on the left hand side (LHS). We have also introduced a discrete-state dependent measurement PDF, $h^l(z_t\,|\,x)$, and the measurement likelihood conditioned on the discrete state, $h^l(z_t)$. Allowing the measurements to be conditioned on the discrete state as well, lets us model that a target for instance can switch between giving biased measurements and not, change measurement noise levels and so on. Say that a sensor is set up to give point estimates of the target, which specific point it is measuring and how exact this point location is, might be dependent on some property of the target. If this property can be described by a discrete variable, we can model this and therefore be able to estimate better measurement properties, which again can result in a more consistent estimate.

This concludes the full step of prediction and update of a hybrid system in what could be called IMM style, due to the mixing step that is a special feature of IMM.

3.3.1 The interacting multiple models as a hybrid state formulation

IMM [Blom, 1984] or even variable structure interacting multiple models (VSIMM) [Li and Bar-Shalom, 1992, 1996] are a special case of the hybrid formulation given here, where l_t, or a subset thereof, represents the discrete event that the continuous state evolves according to a specific dynamic model, and the discrete transition function is of a particular type.

We can let the discrete transition probabilities be independent of the continuous state, such that $\pi^{l_t}_{l_{t_-}}(x_{t_-}) = \pi^{l_t}_{l_{t_-}}$. Then this is the same transition as considered in the IMM for its multiple models, where it here can (if needed) represent more than just different models. With this independence, the discrete state prediction given in (3.44) becomes the model prediction of the IMM, given by [(11.6.6-8)][1], and $\pi^{l_{t_-}}_{x_{t_-}} \frac{\mu^l_{t_-\,|\,t_-}}{\mu^l_{t\,|\,t_-}}$, the mixing probabilities in (3.46), becomes the mixing probabilities of the IMM given by [(11.6.6-7)][1]. The

[1] [·] refers to content on pages 455-456 in Bar-Shalom et al. [2001, pp.455] which describes the IMM algorithm

prediction and update equations for the continuous state given by (3.45) and (3.47) respectively, are just generalizations to general distribution compared to the Gaussian assumption in the prediction and update step of the IMM as given by [step 3][1]. Lastly we have the discrete-state probability updates, also given by (3.47), are exactly the IMM model update given by [(11.6.6-15)][1]. The hybrid state formulation is therefore, just as said, a generalization of the IMM, as given by Bar-Shalom et al. [2001] on pages 455-456, to allow general continuous-state distributions and continuous-state dependent discrete-state transition function.

One might argue that such a generalizations are of little value without any specific algorithms to accommodate them. However, it shows that the IMM is a solution to a specific problem within a broader set of problems, which also for instance includes problems where detectability changes. This tells us that we can look for other types of modeling within this class whenever the IMM assumptions breaks down. If, for instance, the model transitions are dependent on the target position, which in many cases might be the case, one might get into trouble if this is not considered by the algorithm. By placing properly scaled Gaussians in the state space to model varying transition probabilities, the problem is still solvable with a closed form solution. Some caution must be taken however. Since the transitions must sum to one, one will have one transition with a positive weighted Gaussian and one or more of the others will have the same Gaussian with negative weight so that their sum will cancel it. This leads to a Gaussian mixture with negative weights in the mixing step, son one should be careful when performing mixture reduction to ensure that distribution stays non-negative.

The IMM was originally made by Blom [1984] and later rederived by Blom and Bar-Shalom [1988] seen as the optimal timing of hypothesis reduction. The hypothesis reduction step in the standard IMM is to reduce the mixture of (3.46) down to one Gaussian, while we here leave this as a separate problem that can be fine tuned to the application.

VSIMM, as described by for instance Li and Jilkov [2005] or Li and Bar-Shalom [1996], is to the authors understanding a way of doing on-line model parameter estimation and handling the state dependent model transition function in such a way that one can discard unlikely models. The implementations are in general ways to discard unlikely transitions, merge others or estimate new ones in some manner, if there are sufficient evidence to do so. One way this can be done, is by applying a somewhat crude gating to Gaussians in the transition, and not having any uniform transition probability what so ever. Doing this gating in the hybrid state transition can then make it behave like a variable structure, since some discrete state transitions will give zero probability. Otherwise, the VSIMM uses the same assumptions as described here, but then goes on to make ways to prune unlikely discrete states and and use data to estimate new ones.

3.3.2 Detectability and features as states

If one adds an independent discrete state for target detectability, which could be needed if there is an ambiguity in measurement origin or if there is uncertain target existence, the hybrid state space also allows for this through the discrete state measurement likelihood if we allow it to integrate to something less than unity, i.e. having detectability less than unity. For this problem we will defer more elaboration until existence and measurement origin uncertainty has been properly introduced in chapter 6.

With the appropriate modeling, it is possible to include features, both static and dynamic, both discrete and continuous, into this framework so that when the measurement origin is uncertain we can evaluate likelihoods of, for instance, the visual cues to identify targets.

4 Random Finite Sets

In MTT one deals with the situation in which there are an unknown number of targets with unknown positions in the scanned environment. A notion of these unknowns can be given by a probability distribution. From a given number of targets, one needs to estimate the location of the targets, which typically can be dependent on the number. RFSs allow you to model the situation as a set with an unknown number of elements (targets) with unknown positions (states). This uses something similar to PDFs to treat this in a Bayesian framework similar to that of normal state estimation. These distribution-like functions, which we will call set distribution functions (SDFs) here, are treated using a framework known as FISST [Mahler, 2007]. This chapter presents a short description of RFSs, their distributions, how to do integrals in this framework and taking unions both with and without keeping identities.

Sets are unordered

A RFS X consists of a set of random vectors $x^i,\ i \in \{1, \ldots, n\}$, where $n = |X| \in \mathbb{N}$, the cardinality of the set (e.g. the number of targets), is also a random variable. Like any set, a RFS is unordered; $\left\{x^1, x^2\right\} = \left\{x^2, x^1\right\}$, i.e. the indicies are irrelevant and therefore in general are only included for notational purposes [Mahler, 2007].

In the context of MTT, the measurements, for instance, are unordered in nature, they have no identity other than that they maybe arrive or are stored in a certain arbitrary order. When treating the measurements in a scan as a RFS within the FISST framework, the need to do data association shows up as a result of this unorderedness without any more modelling, as will be discussed further in chapter 7.

4.1 The general set distribution

A distribution of a RFS will be called a SDF in this text as it cannot be treated as a PDF. For a RFS the SDF is defined using FISST [Mahler, 2007], and can be represented by

Williams [2015b, (1)]

$$p_X\left(\{x^1,\ldots,x^n\}\right) = \mathrm{P}(n)\sum_{\pi\in\mathcal{P}_n} p_n(x^{\pi_1},\ldots,x^{\pi_n}) \tag{4.1}$$

where $\mathrm{P}(n)$ is the cardinality distribution of the random set, $p_n(\cdot)$ is the cardinality-conditioned joint state distribution, and $\mathcal{P}_n$ is the set of $n!$ possible permutations of the n objects. The reason for using the sum is that any of the "positions" in the argument of $p_n(\cdot)$ could have given any of the vectors x^i, and we thus need to account for of all combinations of vectors in all positions so that the distribution is permutation invariant, as the set is.

Seeing the permutation variable π as latent variable could also be done. This can be viewed as having what could be called a random tuple or random orderer list, if that makes sense in the application and the modeling of it;

$$p_{X,\pi}\left((x^1,\pi^1),\ldots,(x^n,\pi^n)\right) = \mathrm{P}(n)p_n\left(x^{\pi^1},\ldots,x^{\pi^n}\right). \tag{4.2}$$

4.2 The independent identical set distribution

When the targets are independent identically distributed (i.i.d.) with PDF $p_x(x)$ we have that

$$p_n(X) = \prod_{i=1}^{n} p_x(x^i) \triangleq p_x^X, \tag{4.3}$$

where the last equality defines the simplifying set exponent notation. The SDF ot the set is then as (11.121) on page 366 by Mahler [2007]

$$p_X\left(\{x^1,\ldots,x^n\}\right) = n!\mathrm{P}(n)\prod_{i=1}^{n} p_x(x^i) = n!\mathrm{P}(n)p_x^X. \tag{4.4}$$

Here the summation have simply been replaced with a multiplication by $n!$ since it gives the same for all permutations.

Note that (4.1) is much more general than this, since it can handle the case where the set members are correlated (i.e. not independent). Having correlated set members can for instance model cases where there are can be up to two targets, but with zero likelihood for them being in the same area, which cannot be modeled properly with the i.i.d. SDF. The notion of keeping the latent permutation variable does not necessarily make much sense in this case, since there is no extra information obtained by knowing the order in an i.i.d. distribution, unless we have more information from application dependent modeling.

4.3 The set integral

It might at first be a little counter intuitive that the factor $n!$ should be there in (4.4), since integrating out all the x^i would leave $n!\mathrm{P}(n)$ and not just simply $\mathrm{P}(n)$. This comes again from the fact that we are dealing with sets, and looking at (4.1) the PDF $p_n(\cdot)$ is given for a special ordering of the vectors, which is only one of $n!$ equally plausible. By using a random tuple, that makes this ordering specific, we get what is expected. Instead we have to adapt our integral when we marginalize out the state from the SDF to be performed with a set differential so that all the permutations we integrate over are accounted for. Introducing set differential notation and integrating out the set elements to get the cardinality, the cardinality conditioned set integral is given by Mahler [2007, pp. 361]

$$\begin{aligned}
&\int\limits_{x^i \in \mathcal{X}} p_X\left(\{x^1, \ldots, x^n\}\right) \delta\{x^1, \ldots, x^n\} \\
&\qquad = \frac{1}{n!} \int\limits_{x^i \in \mathcal{X}} \mathrm{P}(n) \sum_{\pi \in \mathcal{P}_n} p_n(x^{\pi_1}, \ldots, x^{\pi_n}) \mathrm{d}x^1 \cdots \mathrm{d}x^n \\
&\qquad = \frac{\mathrm{P}(n)}{n!} \sum_{\pi \in \mathcal{P}_n} \int\limits_{x^i \in \mathcal{X}} p_n(x^{\pi_1}, \ldots, x^{\pi_n}) \mathrm{d}x^1 \cdots \mathrm{d}x^n = \frac{\mathrm{P}(n)}{n!} n!.
\end{aligned} \tag{4.5}$$

We can now see that marginalizing out the state will result in the cardinality distribution as intended, and thus the correct way to perform the integral over the set given a cardinality. Without conditioning on the cardinality we also have to sum over n, and we see that the total set integral will then become unity as any distribution should. Without conditioning on the caridinality it is written as

$$\int p_X(X)\, \delta X = \sum_{n=0}^{\infty} \frac{1}{n!} \int p_X\left(\{x^1, \ldots, x^n\}\right) \mathrm{d}x^1 \cdots \mathrm{d}x^n \tag{4.6}$$

4.4 Union of independent sets

Say we have two independent RFSs, $Y \sim p_1$ and $Z \sim p_2$, where p_1 and p_2 are their SDFs. Then their joint distribution can be written as

$$p(Y, Z) = p_1(Y) p_2(Z), \tag{4.7}$$

by the independence assumption. Now, let us say that their elements come from the same space, i.e. $Y \subset \mathcal{X}$ and $Z \subset \mathcal{X}$, so that it is possible to take their union, $X = Y \cup Z$. What can we say about the distribution of their union?

Let us first introduce what the author would call component labeling of the set, X;

$$X^L = \{(x^1, \pi^1), \ldots, (x^n, \pi^n)\} : \bigcup_{i=1}^{n} \{x^i\} = X, \tag{4.8}$$

where π^i, the label, describes which of the sets in the union element i is from. For our example, we have $\pi^i \in \{1, 2\}$, where $\pi^i = 1$ is the event $x^i \in Y$ and $\pi^i = 2$ is the event $x^i \in Z$. We will also use what the author would call the 'component extractor'

$$X^{L(k)} = \{x : (x, k) \in X^L\} \subseteq X, \tag{4.9}$$

which extracts an unlabeled subset of X^L that has component label k, i.e. $X^{L(k)} \not\subseteq X^L$, since unlabeled elements are different from labeled elements.

The labeled union can then be written as $X^L = Y^1 \cup Z^2$ where the numbered superscript denotes the label of the elements in the set, such that $Y^1 = \{(y^1, 1), \dots, (y^n, 1)\}$ etc.. With this we can write the joint distribution over Y and Z as a distribution of their labeled union;

$$p(X^L) = p_1(X^{L(1)})p_2(X^{L(2)}). \tag{4.10}$$

This can be interpreted as having modeled two distinct independent sets, but the only event we can observe is their union. The component labeling is then also an actual event, and may or may not be observed. In the case when the label is not observed, it is only a latent variable. The observed event will then be the unlabeled union, $X = Y \cup Z$. The labeling can then be seen as giving the likelihood of which elements in X came from the set Y or Z. Hence the (unlabeled) distribution of X can then be seen as the marginalization of the labels;

$$p(X) = p(Y \cup Z) = \sum_{\pi^i \in \{1,2\}} p(X^L) \tag{4.11}$$

$$= \sum_{\pi^i \in \{1,2\}} p_1(X^{L(1)})p_2(X^{L(2)}) = \sum_{\hat{Y} \subseteq X} p_1(\hat{Y})p_2(X \setminus \hat{Y}). \tag{4.12}$$

The last formula is equivalent to the set convolution by Mahler [2007, pp. 386], although we have taken a completely different approach to arrive at it. Another way to see that we need this summation is that the permutation invariance means we have to "try all combinations" as in (4.1). The name set convolution is used, since the formula resembles that of a convolution, and there exists integral transformations to turn them into multiplications similar to Fourier transforms [Mahler, 2007].

If one has the union of several sets, the above formula is applied recursively for adding in one set at a time, to give (11.252) in Mahler [2007, pp.385]. One can also keep the latent component pointer, if one is interested in the likelihood of which underlying set an element came from. This could for instance be the case if one is modeling ships as being the union of a cargo ship, cabin cruiser and sailboat. If one is measuring this using a radar it can be hard to distinguish which type of ship a measurement is, but their likelihood might change over time as they behave differently. Another thing this can be used for, is during tracking where new components are added at each prediction and update step, so it is possible to keep track of the likelihood of having a set of targets coming from that component.

4.5 Specific set distribution functions

4.5.1 The Poisson point process

A commonly used distribution in MTT is the Poisson point process (PPP) where the cardinality distribution is given by the Poisson distribution $\mathrm{P}(n) = \frac{\bar{\lambda}^n}{n!}\mathrm{e}^{-\bar{\lambda}}$, with $\bar{\lambda}$ being the expected cardinality, and the objects are i.i.d. according to some PDF $p(x)$. This gives the SDF [Mahler, 2007, (11.122), pp.366]

$$p_X(X) = \frac{\bar{\lambda}^n}{n!}\mathrm{e}^{-\bar{\lambda}} n! p_x^X = \mathrm{e}^{-\bar{\lambda}} \prod_{i=1}^{n} \lambda^X, \tag{4.13}$$

in which we have the PPP intensity, $\lambda(x^i) = \bar{\lambda}\, p(x^i)$, and therefore also $\bar{\lambda} = \int_{x\in\mathcal{X}} \lambda(x)\,\mathrm{d}x$. This tells us that modeling the intensity is all we need if we are dealing with a PPP.

Union of independent PPP

Say we have a union of two independent RFSs, $X = X_1 \cup X_2$, where each is a PPP with intensities λ_1 and λ_2 respectively. Then, since we in general do not know which distribution gave rise to which element, we have that the total distribution is given by the set convolution of (4.12)

$$p(X) = \sum_{\substack{X_1 \subseteq X,\\ X_2 = X\setminus X_1}} p(X_1)p(X_2) = \mathrm{e}^{-\bar{\lambda}_1 - \bar{\lambda}_2} \sum_{\substack{X_1 \subseteq X,\\ X_2 = X\setminus X_1}} \lambda_1^{X_1}\lambda_2^{X_2} = \mathrm{e}^{-(\bar{\lambda}_1+\bar{\lambda}_2)}\left(\lambda_1+\lambda_2\right)^X. \tag{4.14}$$

The last equality is most easily observed by writing the left hand side out for some sets and see that it follows, and also by noting that it integrates to unity using the set integral. Thus the union of two independent PPPs is again a PPP with the summed intensity, which should not be a surprise from standard theory on the Poisson distribution.

An approach to labeling the PPP

By rewriting the sum with a label variable instead of directly over the subsets we can get another insight into unions of several different i.i.d. RFSs

$$\sum_{\substack{X_1 \subseteq X,\\ X_2 = X\setminus X_1}} \lambda_1^{X_1}\lambda_2^{X_2} = \sum_{\pi \in \{1,2\}^{|X|}} \prod_{i=1}^{n} \lambda_{\pi_i}(x_i). \tag{4.15}$$

A RFS that is made up of several different i.i.d. components can (similar to the general SDF, but not quite the same) be seen to be a marginalization over a latent component indicator variable. By including this as an augmentation to the set elements, we can include which component gave rise to the individual elements if wanted;

$$p\left(\left\{\begin{bmatrix}x_i\\ \pi_i\end{bmatrix}\right\}_{i\in\{1,\dots n\}}\right) = \mathrm{e}^{-\bar{\lambda}_1-\bar{\lambda}_2}\prod_{i=1}^{n}\lambda_{\pi_i}(x_i). \tag{4.16}$$

Even more generally, we can perform a mix of these, where we want to know the component for some and not for some others. By letting $\pi_i = 0$ mean it is irrelevant which component and $\lambda_0(x) = \lambda_1(x) + \lambda_2(x)$, the last equation still holds.

Note that this does not suggest that we can perform a labeling to distinguish the individual objects that are realizations of a PPP, but rather have distributions over which components they came from. This follows from the fact that the PPP is inherently unordered and i.i.d., and any individual labeling, as done by for example Vo and Vo [2011], introduces an ordering that is not necessarily described by anything in the PPP itself. If the PPP models a discrete time arrival process derived from a continuous time exponentially distributed arrival rate, a specific ordering could be inferred from the continuous time ordering, an indistinguishable event in the homogeneous case, and not an i.i.d. event in the inhomogeneous case.

4.5.2 The multi Bernoulli point process

Another common distribution in the MTT setting is the multi Bernoulli (MB) [Mahler, 2007, sec 11.3.4.5], which is the union of multiple Bernoulli point processs (BPPs) [Williams, 2015b]. A Bernoulli process describes an event with a binary outcome and can be used to describe the existence probability of a given object. To describe the state (or the point) of the process we also need an existence conditioned state distribution. If the probability of existence is $r \in [0, 1]$ and the existence conditioned state distribution is $p(x)$, the BPP uses both to make single target RFS SDF to be given by [Williams, 2015b, (9)]

$$p_X(X) = \begin{cases} 1-r, & X = \emptyset, \\ rp(x), & X = \{x\}, \\ 0, & \text{otherwise } (X = \left\{x^1, x^2\right\} \text{ etc.}). \end{cases} \tag{4.17}$$

This distribution is seen to either have a cardinality of zero or one as all other sets have zero likelihood.

If we now have a union of independent RFS that follow a Bernoulli point process such that $X = \bigcup_{i=1}^{N} X_i$ with individual existence probabilities r_i and existence conditioned state distributions $p_i(x^i)$ we get the SDF by applying (4.12) recursively and simplifying the set convolutions into a sum over a single variable, to be [Mahler, 2007, (11.132)][Williams, 2015b, (11)]

$$p_X(X) = \sum_{\pi \in \mathcal{P}_n^N} \left[\prod_{i=1}^{n} r_{\pi_i} p_{\pi_i}(x^i) \right] \left[\prod_{\substack{j=1 \\ j \notin \pi}}^{N} (1 - r_j) \right] \tag{4.18}$$

$$= \prod_{i'=1}^{N} (1 - r_{i'}) \sum_{\pi \in \mathcal{P}_n^N} \prod_{i=1}^{n} \frac{r_{\pi_i} p_{\pi_i}(x^i)}{1 - r_{\pi_i}} \tag{4.19}$$

for a set X of $n <= N$ elements, and otherwise 0. $\mathcal{P}_n^N$ is the set of all sets π of indexes so that all possible ways to choose n elements out of N in order (i.e. including permutations) is in the set, where each element $\pi_i \subseteq [1 : N]$ and $|\pi| = n$. Again, this is summed over all possible ways that n objects could have come from the N components. Given the permutation, and hence the origin and existence state of the components, the SDF becomes the product of the individual densities. This SDF has a maximum of N on its cardinality, as higher cardinalities have a probability of zero. Another thing to note is that this set is not i.i.d. as adding another object might change the likelihood of the others.

Labeling of MB

Again we can consider π as a latent component origin variable, and get a random multi Bernoulli distributed tuple instead. For $\pi^i \neq \pi^j \, \forall \, i, j$ this can be written as

$$p_X\left((x^1, \pi^1), \ldots, (x^n, \pi^n)\right) = \left[\prod_{i'=1}^{N} (1 - r_{i'}) \right] \left[\prod_{i=1}^{n} \frac{r_{\pi^i} p_{\pi^i}(x^i)}{1 - r_{\pi^i}} \right]. \tag{4.20}$$

This is exactly what is known as labeled multi Bernoulli (LMB) [Vo and Vo, 2011], but π is not thought of as a label here, but rather as a latent component pointer, which one could form a PDF over. It is essentially just a convenient notation for not having a tuple with one set element per BPP component. One can also, more naturally, write this using the individual BPPs as given by (4.17) as

$$p_X\left(\{(x^1, \pi^1), \ldots, (x^n, \pi^n)\}\right) = p_X\left(X^L\right) = \prod_{i=1}^{N} p_i(X^{L(i)}), \tag{4.21}$$

where $p_i(\cdot)$ now is the single component BPP SDF. That is, all targets $x^{i'}$ that have $\pi^{i'} = i$ are treated as coming from component i, and if there are several components such that $\pi^{i'} = \pi^{i''} = i$, the BPP will give a likelihood of zero and therefore the LMB will also give a zero likelihood.

Note that this type of “labeling” is slightly different from what for example Vo and Vo [2011] is considering, in the sense that here it is a latent variable extracted from the SDF, whereas Vo and Vo [2011] introduces the labeling on top of the SDF and introduces an unnecessary indicator function in (4.20) to achieve (4.21). Indicator functions and delta functions typically mean that there is some sort of over-representation in the distribution. It seems to the author that this is the case with the formulation by Vo and Vo [2011] as there is, as shown here, already a latent variable describing what they want, and the BPP as given here is already capable of handling this from its cardinality distribution.

As previously noted, Vo and Vo [2011] also include a labeling on the PPP that is different to what was done in this text. If they are considering the PPP as the limit of a MB with infinite amount of components, one could possibly conceive a labeling structure similar to what is done here for the MB, but then having to work on the joint of the underlying MB instead, which then reduces to the MB anyway. Since a MB have to arise naturally or be engineered, one should try to avoid handling individual components in these limits. Therefore the only practical individual object labeling of a PPP conceivable to the author is that of the arrival order in continuous time, but will not be discussed further. Of course there could also be other ways of seeing this that the author does not know about.

5 The Kullback-Leibler divergence

If we want to approximate one distribution with another, we need to specify what constitutes a good approximation. A metric could be a good choice, but in statistics one sometimes considers unsymmetrical cases, and these "distance" functions are called divergences. Since one is approximating a function by another, the literature often considers functionals and these methods are called variational inference. The following is mostly based on Koller and Friedman [2009].

5.1 Negative log likelihood divergence

To consider an approximation of $p(x)$ by a usually much simpler distribution $q(x)$, one often considers minimizing the negative log likelihood with respect to N data points x_i sampled from p. By taking the limit of infinite data points we get

$$\lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N} \ln(q(x_i)) = \int_x -p(x)\ln(q(x))\,\mathrm{d}x = \mathop{\mathbb{E}}_p[-\ln(q(x))] = \mathbb{H}[p(x), q(x)], \tag{5.1}$$

where $\mathbb{H}[\cdot,\cdot]$ is the cross entropy between p and q [Goodfellow et al., 2016, (3.51)]. This divergence can also be used for discrete variables or RFSs by changing the integral to a summation or set integral respectively.

5.2 KL-divergence its projections

The cross entropy has the problem of not being zero even with perfect match between the distributions. This is seen from the fact that the entropy, given by the cross entropy between p and itself, is nonzero;

$$\mathbb{H}[p(x)] = -\int_x p(x)\ln(p(x))\,\mathrm{d}x > 0. \tag{5.2}$$

Subtracting this from the negative log likelihood divergence gives the well known KL-divergence;

$$\begin{aligned}\mathrm{D}_{\mathrm{KL}}(p \;||\; q) &= \mathbb{H}[p(x), q(x)] - \mathbb{H}[p(x)] \\ &= \int_x -p(x)\ln(q(x))\,\mathrm{d}x + \int_x p(x)\ln(p(x))\,\mathrm{d}x \\ &= \int_x p(x)\log\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x = \mathop{\mathbb{E}}_p[\ln(p(x)) - \ln(q(x))]\,,\end{aligned} \tag{5.3}$$

which is non negative, and equal to zero if and only if the two distributions are equal on the support of p [Bishop, 2016, sec. 1.6]. This is also known as relative entropy. Performing the minimization of $\mathrm{D}_{\mathrm{KL}}(p \;||\; q)$ with respect to q is known as moment projection, while minimizing $\mathrm{D}_{\mathrm{KL}}(q \;||\; p)$ with respect to q is known as information projection. The moment projection version of this divergence we "derived" used the interpretation that we want to find the best q to approximate data sampled from p in the log-likelihood sense. The information projection can therefore be seen as the opposite; we find the best approximate distribution, q, such that p is the best fit distribution when sampling from q in the log likelihood sense, which is likely to be a different distribution than the other way around.

5.3 Moment projection onto the exponential family

Suppose we want to approximate p by q, where $q = q(x;\theta)$ is a distribution of the exponential family parameterized by θ, with a sufficient statistic given by a function $\tau(x)$ such that

$$q(x;\theta) = \frac{\exp\left(\tau(x)^T t(\theta) + g(x)\right)}{Z(\theta)}. \tag{5.4}$$

Here, $t(\theta)$ and $g(x)$ are appropriate functions and $Z(\theta)$ is a normalizing constant. The moment projection of this can be written as

$$\begin{aligned}\mathrm{D}_{\mathrm{KL}}(p(x) \;||\; q(x;\theta)) &= \mathop{\mathbb{E}}_p[\ln(p) - \ln(q(x;\theta))] && (5.5)\\ &= \mathop{\mathbb{E}}_p\left[\ln(p) - \tau(x)^T t(\theta) - g(x) + \ln(Z(\theta))\right] && (5.6)\\ &= \mathop{\mathbb{E}}_p[\ln(p)] - \mathop{\mathbb{E}}_p\left[\tau(x)^T\right] t(\theta) - \mathop{\mathbb{E}}_p[g(x)] + \ln(Z(\theta)). && (5.7)\end{aligned}$$

Now assume that θ' is a unique parameter such that $\mathop{\mathbb{E}}_{q(x;\theta')}\left[\tau(x)^T\right] = \mathop{\mathbb{E}}_{p}\left[\tau(x)^T\right]$, and add and subtract $\mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta'))$. We then get

$$\begin{aligned}
\mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta)) = \; & \mathop{\mathbb{E}}_{p}[\ln(p)] - \mathop{\mathbb{E}}_{p}\left[\tau(x)^T\right] t(\theta) - \mathop{\mathbb{E}}_{p}[g(x)] + \ln(Z(\theta)) && (5.8)\\
& -\mathop{\mathbb{E}}_{p}[\ln(p)] + \mathop{\mathbb{E}}_{p}\left[\tau(x)^T\right] t(\theta') + \mathop{\mathbb{E}}_{p}[g(x)] - \ln(Z(\theta')) \\
& \quad + \mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta')) \\
= \; & \mathop{\mathbb{E}}_{q(x;\theta')}\left[\tau(x)^T\right] t(\theta') + \mathop{\mathbb{E}}_{q(x;\theta')}[g(x)] - \ln(Z(\theta')) && (5.9)\\
& - \mathop{\mathbb{E}}_{q(x;\theta')}\left[\tau(x)^T\right] t(\theta) - \mathop{\mathbb{E}}_{q(x;\theta')}[g(x)] + \ln(Z(\theta)) \\
& + \mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta')) \\
= \; & \mathop{\mathbb{E}}_{q(x;\theta')}[\ln(q(x;\theta')) - \ln(q(x;\theta))] + \mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta')) \\
& && (5.10)\\
= \; & \mathrm{D}_{\mathrm{KL}}(q(x;\theta') \mid\mid q(x;\theta)) + \mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta')) && (5.11)\\
\geq \; & \mathrm{D}_{\mathrm{KL}}(p(x) \mid\mid q(x;\theta')) \,. && (5.12)
\end{aligned}$$

Where the last line follows from the non negativity of the KL-divergence, and we have that $\theta' : \mathop{\mathbb{E}}_{q}[\tau(x)] = \mathop{\mathbb{E}}_{p}[\tau(x)]$ is the moment projection of p onto this set of distributions parameterized by θ. Note that this holds for any distributions p and all distributions q that can be written as (5.4).

5.3.1 Moment projection onto normal distribution

The normal distribution is clearly of the form (5.4), where we can identify

$$\begin{aligned}
\theta &= \{\mu, \Sigma\} \,, && (5.13)\\
\tau(x) &= \begin{bmatrix} x \\ \mathrm{vec}(xx^T) \end{bmatrix}, && (5.14)\\
t(\theta) &= \begin{bmatrix} \Sigma^{-1}\mu \\ -0.5\mathrm{vec}(\Sigma^{-1}) \end{bmatrix}, && (5.15)\\
Z(\theta) &= \sqrt{\det(2\pi\Sigma)}\exp(0.5\mu^T\Sigma^{-1}\mu), && (5.16)\\
g(x) &= 0. && (5.17)
\end{aligned}$$

The sufficient statistics are the two first moments of x. Thus, making $\theta = \{\mu, \Sigma\}$ such that the two first moments of q match those of p, will be the moment projection of p onto the set of normal distributions. That is

$$\{\mu', \Sigma'\} = \arg\min_{\mu,\Sigma} \mathrm{D}_{\mathrm{KL}}(q \mid\mid p) = \left\{ \mathop{\mathbb{E}}_{p}[x] \,, \mathop{\mathbb{E}}_{p}\left[xx^T\right] - \mathop{\mathbb{E}}_{p}[x]\mathop{\mathbb{E}}_{p}[x]^T \right\}. \quad (5.18)$$

5.3.2 Mixture distributions and their two first moments

A mixture is a distribution consisting of several other weighted distributions. In applications, one often ends up having a mixture. This may be due to having several weighted hypotheses, wherein each hypothesis has its own distribution, given that the hypothesis is correct. A mixture can thus be seen as a marginal of a joint distribution over the random variable of interest and a discrete latent variable of the form

$$p(x) = \sum_{i=1}^{N} p_i(x)\mu_i, \tag{5.19}$$

where i is the latent variable with the subscript indicating the i'th component of the mixture, having the distribution $p_i(x)$, and $\mu_i : \sum_{i=1}^{n} \mu_i = 1$ is probability of the latent variable, called its weight. We can find the mean $\hat{x}$ and covariance P of this type of distribution by calculating the first and second order moments

$$\begin{aligned} \hat{x} &= \mathop{\mathbb{E}}_{i,x}[x] = \sum_{i=1}^{N} \mu_i \mathop{\mathbb{E}}_{x\,|\,i}[x] = \sum_{i=1}^{N} \mu_i \hat{x}_i, && (5.20) \\ P &= \mathop{\mathbb{E}}_{i,x}\left[xx^T\right] - \mathop{\mathbb{E}}_{i,x}[x] \mathop{\mathbb{E}}_{i,x}[x]^T && (5.21) \\ &= \left[\sum_{i=1}^{N} \mu_i \mathop{\mathbb{E}}_{x\,|\,i}\left[xx^T\right]\right] - \hat{x}\hat{x}^T && (5.22) \\ &= \left[\sum_{i=1}^{N} \mu_i (P_i + \hat{x}_i \hat{x}_i^T)\right] - \hat{x}\hat{x}^T, && (5.23) \end{aligned}$$

where the subscript i denotes that the mean $\hat{x}_i$ or covariance P_i is conditioned on coming from mixture i. It is therefore straightforward to project a mixture, in which we know the individual components means and covariances, into a normal distribution according to the last subsection.

5.3.3 Merging components in Gaussian mixture

When one is dealing with mixtures one often wants to reduce the mixture to a mixture with less components. However, the optimal solution to this reduction is not known, and is a challenging problem. This is due to the fact that the components in the new mixture can for instance explain only a part of the probability mass of a component in the mixture to be reduced, while the rest is explained by some of the others. The optimal solution hence consists of finding how much of which component is to be described by each of the new components. This typically lends itself to algorithms like expectation maximization[Bishop, 2016] and using some form of gradient descent. However, partly because a mixture often is multi modal and partly because the new components can be permuted in the optimal solution and still be optimal, this optimization problem is not convex, and therefore a though problem.

However, heuristics can be applied. One heuristic would be to merge similar components, and pruning (throwing away) some others. Ideally, one would just use the KL-divergence to find the components to merge and then merge them using moment matching as described above. Unfortunately no closed form solution exists even to the simplified problem of finding the best components to merge. What exists, however, is an upper bound when the components are Gaussian. This upper bound is due to Runnalls [2007] and has proven to give better results than using the heuristic by Salmond [2009, 1990] and behave slightly different than when using the integral squared error [Williams, 2003]. Runnalls upper bound is basically based on using the log sum inequality on the KL-divergence, and combining component i and j is given by [Runnalls, 2007, (21)]

$$B(i,j) = \frac{1}{2}\left[(w_i + w_j)\ln\left(\det(P)\right) - w_i \ln\left(\det(P_i)\right) - w_j \ln\left(\det(P_j)\right)\right], \tag{5.24}$$

where w_i and w_j are the weights of the components to be combined, and P_i and P_j their covariances respectively. P is the moment matched merged covariance and is given by

$$P = \frac{1}{w_i + w_j}\left[w_i P_i + w_j P_j + \frac{w_i w_j}{w_i + w_j}(\mu_i - \mu_j)(\mu_i - \mu_j)^T\right], \tag{5.25}$$

where μ_i and μ_j are the components means, respectively.

This distance between two Gaussian mixture components can be used to choose components to merge in a greedy fashion.

5.3.4 Moment projection onto Poisson distribution

The Poisson distribution,

$$\mathrm{P}\left(n\right) = \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} = \frac{\exp\left(n\ln(\lambda) - \ln(n!)\right)}{\exp(\lambda)} \tag{5.26}$$

is also of the form (5.4), with expected value

$$\bar{n} = \mathbb{E}[n] = \sum_{n=0}^{\infty} n\frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} = \lambda\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^{n-1}}{(n-1)!} = \lambda. \tag{5.27}$$

We can identify

$$\theta = \lambda, \tag{5.28}$$

$$\tau(x) = n, \tag{5.29}$$

$$t(\theta) = \ln(\lambda), \tag{5.30}$$

$$g(n) = \ln(n!) = \sum_{i=1}^{n}\ln(i), \tag{5.31}$$

$$Z(\theta) = \mathrm{e}^{\lambda}. \tag{5.32}$$

Hence, the moment projection of a distribution, p, over the natural numbers onto a Poisson distribution, q, is achieved with

$$\lambda = \mathop{\mathbb{E}}_{p}[\tau(n)] = \mathop{\mathbb{E}}_{p}[n] \,. \tag{5.33}$$

5.4 Moment projection onto factorized distribution

Sometimes we want approximate a joint distribution p by something much simpler. One such simplification is to moment project the joint $p(x)$ onto a fully factored distribution $q(x) = \prod_{i=1}^{n} q^i(x_i)$, where $\dim(x) = n$. Using $p^i(x_i)$ to denote the marginal distribution of x_i we can write the divergence as.

$$\begin{aligned}
\mathrm{D}_{\mathrm{KL}}(q \,||\, p) &= \mathop{\mathbb{E}}_{p}\left[\ln\left(\frac{p(x)}{\prod_{i=1}^{n} q^i(x_i)}\right)\right] = \mathop{\mathbb{E}}_{p}\left[\ln\left(\frac{p(x)\prod_{i=1}^{n} p^i(x_i)}{\prod_{i=1}^{n} q^i(x_i)\prod_{i=1}^{n} p^i(x_i)}\right)\right] && (5.34)\\
&= \mathop{\mathbb{E}}_{p}\left[\ln\left(\frac{\prod_{i=1}^{n} p^i(x_i)}{\prod_{i=1}^{n} q^i(x_i)}\right) + \ln\left(\frac{p(x)}{\prod_{i=1}^{n} p^i(x_i)}\right)\right] && (5.35)\\
&= \mathop{\mathbb{E}}_{p}\left[\sum_{i=1}^{n} \ln\left(\frac{p^i(x_i)}{q^i(x_i)}\right)\right] + \mathop{\mathbb{E}}_{p}\left[\ln\left(\frac{p(x)}{\prod_{i=1}^{n} p^i(x_i)}\right)\right] && (5.36)\\
&= \sum_{i=1}^{n} \mathop{\mathbb{E}}_{p^i}\left[\ln\left(\frac{p^i(x_i)}{q^i(x_i)}\right)\right] + \mathrm{D}_{\mathrm{KL}}\left(\prod_{i=1}^{n} p^i(x_i) \,||\, p(x)\right) && (5.37)\\
&\geq \sum_{i=1}^{n} \mathrm{D}_{\mathrm{KL}}\left(q^i(x_i) \,||\, p^i(x_i)\right) . && (5.38)
\end{aligned}$$

Since the KL-divergence reaches its minimum, zero, when the distributions are equal, we see that the moment projection of p onto a fully factored distribution is the product of its marginals.

5.5 Moment projection of SDF onto i.i.d. SDF

A general SDF can be written as $p(X) = \mathrm{P}(n)n!p_n(x^1, \dots, x^n)$, where p_n is a permutation-invariant distribution. Permutation invariance can be achieved from non-permutation-invariant distributions, by summing over permutations and dividing by $n!$.

The moment projection of this general SDF, p, onto an i.i.d. SDF, $q(X) = \hat{\mathrm{P}}(n)n!g^{\{x^1,\dots,x^n\}}$, can be written as

$$
\begin{aligned}
\mathrm{D}_{\mathrm{KL}}(p \mid\mid q) = \mathrm{D}_{\mathrm{KL}}\Big(\mathrm{P}(n)n!p_n(x^1,\ldots,x^n) \mid\mid \hat{\mathrm{P}}(n)n!g^{\{x^1,\ldots,x^n\}}\Big) \\
&= \mathop{\mathbb{E}}_p\left[\log\left(\frac{\mathrm{P}(n)n!p_n(x^1,\ldots,x^n)}{\hat{\mathrm{P}}(n)n!g^{\{x^1,\ldots,x^n\}}}\right)\right] \\
&= \mathop{\mathbb{E}}_p\left[\log\left(\frac{\mathrm{P}(n)}{\hat{\mathrm{P}}(n)}\right)\right] + \mathop{\mathbb{E}}_p\left[\log\left(\frac{p_n(\{x^1,\ldots,x^n\})}{g^{\{x^1,\ldots,x^n\}}}\right)\right], \qquad (5.39)
\end{aligned}
$$

where the expectations are taken with respect to the true distribution, $p(X)$. Since we have no restrictions on the form of the cardinality, we can set $\hat{\mathrm{P}}\,(n) = \mathrm{P}\,(n)\ \forall\, n$ to achieve the minimum of zero in the first term. The last term can be better understood through some more massage;

$$
\mathbb{E}\left[\log\left(\frac{p_n(\{x^1,\ldots,x^n\})}{g^{\{x^1,\ldots,x^n\}}}\right)\right] - \mathbb{H}[p_n(X)] \tag{5.40}
$$

$$
= \mathbb{E}\big[\log\big(p_n(\{x^1,\ldots,x^n\})\big)\big] - \mathbb{E}\Big[\log\Big(g^{\{x^1,\ldots,x^n\}}\Big)\Big] - \mathbb{H}[p_n(X)] \tag{5.41}
$$

cancelling the entropy, and expanding the log-product into a sum-log and then writing out the expectation in terms of the set integral;

$$
= -\mathbb{E}\left[\sum_{i=1}^{n}\log(g(x^i))\right] = -\int p(X)\sum_{x'\in X}\log(g(x'))\,\delta X \tag{5.42}
$$

expanding the set integral;

$$
= -\sum_{|X|=0}^{\infty}\frac{1}{|X|!}\int p(X)\sum_{x'\in X}\log(g(x'))\,\mathrm{d}x^1\cdots\mathrm{d}x^{|X|} \tag{5.43}
$$

noting that this is zero for $|X|=0$, and that the summands will be equal after taking the integral, thereby simply turning into a factor of $|X|$ and the log of an arbitrary $x'\in X$;

$$
= -\sum_{|X|=1}^{\infty}\frac{1}{|X|!}|X|\int\log(g(x'))p(X)\,\mathrm{d}x^1\cdots\mathrm{d}x^{|X|} \tag{5.44}
$$

noting that this can be rewritten as a set integral of $X\setminus\{x\}$ as an inner integral and over x' as an outer integral;

$$
= -\int\log(x')\int p(X\cup\{x'\})\delta\,X\,\mathrm{d}x' \tag{5.45}
$$

recognizing the formula for the probability hypothesis density (PHD), $D(x)$ [Mahler, 2007, (16.26)];

$$
= -\int\log(g(x'))D(x')\,\mathrm{d}x' \tag{5.46}
$$

substituting $D(x) = \bar{n}s(x)$, the factorization into a proper distribution and a scaling factor[1];

$$= -\bar{n} \int \log(g(x'))s(x')\,\mathrm{d}x'. \tag{5.47}$$

The last line is known to attain its minimum for $g(x) = s(x)$, which can for instance, be shown using calculus of variations with Lagrange multipliers for the normalization constraint.

It is thus shown that the moment projection onto an i.i.d. SDF, q, is achieved with

$$q(X) = \hat{\mathrm{P}}(|X|)g^X = \hat{\mathrm{P}}(|X|) \prod_{x \in X} g(x), \tag{5.48}$$

$$\hat{\mathrm{P}}(|X|) = \mathrm{P}(|X|), \tag{5.49}$$

$$g(x) = \frac{D(x)}{\int D(x)\,\mathrm{d}x}, \tag{5.50}$$

$$D(x) = \int p(X \cup \{x\})\,\delta X, \qquad \text{[Mahler, 2007, (16.26)]} \tag{5.51}$$

where $D(x)$ is the PHD of the set X, described by Mahler [2007, sec. 16.2], under the SDF p.

[1] This scaling factor is known to be the expected number of targets, and hence the notation $\bar{n} = \mathbb{E}[n]$.

Part II

The Multi Target Filter

6 Properties multiple targets

6.1 Assumptions

For MTT we need to know the underlying properties of the targets and measurements, or at least the assumptions we make about them. The literature there has become a set of appropriate assumptions that are regarded as the standard model. These assumptions can be summarized as follows [Williams, 2015b]:

Assumption A.

1. There is a “scene” of interest, described by a subset of the target state space $\mathcal{X}_t$, where at any time t there exists an unknown number $n_t \in \mathbb{N}$ of targets.
2. Between each time step, each target $x^i_{t_-}$, $i \in [1 : n_{t_-}]$ at time $t_- < t$ follows some i.i.d. Markovian dynamics, with transition PDF
$$f\left(x^i_t \,|\, x^i_{t_-}\right). \tag{6.1}$$
3. Between each time step, each target $i \in [1 : n_{t_-}]$ stays, or survives if you like, in the scene according to an i.i.d. Markovian process with possibly both a time- and a time step size- dependent probability of survival, given by
$$\mathrm{P}_s\left(x^i_{t_-}\right), \tag{6.2}$$
4. At each time t, there is an unknown number of new targets n^η_t arriving, where n^η_t is a random number following a Poisson distribution
$$n^\eta_t \sim \frac{\mathrm{e}^{-\bar{\eta}_t}\bar{\eta}_t^{n^\eta_t}}{n^\eta_t!}, \tag{6.3}$$
and each newly arriving target follows an i.i.d. PDF, referred to as the “arrival distribution” over the state space, independent of the pre-existing targets, given by
$$f^\eta_t(x). \tag{6.4}$$

5. At each time step t, the measurement device will receive at most one measurement from a specific target, i, in state x_t^i with a possibly time and state dependent probability of detection

$$\mathrm{P}_d\left(x_t^i\right). \tag{6.5}$$

6. Each received measurement z_t^j, $j \in [1 : m_t]$, conditioned on its corresponding detected target x_t^i, independent of all other measurements and targets, follows the measurement PDF

$$h\left(z_t^j \mid x_t^i\right). \tag{6.6}$$

7. At each time t, a unknown number m_t^μ of false alarms, e.g. erroneous measurements, are received, with m_t^μ being a random variable following a Poisson distribution

$$m_t^\mu \sim \frac{\mathrm{e}^{-\bar{\mu}_t} \bar{\mu}_t^{m_t^\mu}}{m_t^\mu!}. \tag{6.7}$$

Each false alarm received, follows an i.i.d. PDF over the measurement space, independent of the targets and target related measurements, given by

$$h_t^\mu(z). \tag{6.8}$$

8. Each measurement can come from at most one target, and each target can give at most one measurement.

There are applications where some of these assumptions are relaxed or slightly different. For example, relaxing the last part of assumption A-8 will result in extended object tracking[Granstrom et al., 2016], where targets can give more than one measurement. Some other authors, such as for instance Vo et al. [2014], consider the arrival of new targets to be that of a MB instead of assumption A-4. This might be appropriate for some applications, given that it is modeled appropriately.

6.2 The multi target transition function

Assumption A-4 basically states that targets arrive according to a PPP with intensity given by $\eta_t(x) = \bar{\eta}_t f_t^\eta(x)$. There are several reasons for this to be a good approximation. Firstly the targets are assumed to move i.i.d., so one should expect targets to also arrive i.i.d., and therefore at least the i.i.d. assumption of the PPP is valid. Secondly, the Poisson distribution is the discrete time cardinality distribution of objects arriving with intervals that are, not necessarily uniformly, exponentially distributed in continuous time. Thirdly, a good approximation of the cardinality of a multi Bernoulli process with bounds on the sum of absolute error in distribution is given by Le Cam [1960] to be $\sum_i^n r_i^2$ (very small when the r_i's are small), where r_i is the individual Bernoulli probabilities.

In terms of mathematics, assumptions A-2 and A-3 basically state that the single target transition is that of a BPP, and therefore that the transition of multiple targets is that of a MB. Together with the birth model, we get that the multi target transition density is that of a Poisson multi Bernoulli (PMB). This is also in accordance with Mahler [2007, pp. 472], and we have that combining assumptions A-2, A-3 and A-4 is given in mathematical terms by the conditional SDF

$$f(X_t \mid X_{t_-}) = \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t_-}}} \mathrm{e}^{-\bar{\eta}_t} \eta_t^{X_t \setminus X^s} \sum_{\omega \in \mathcal{P}_{|X^s|}^{|X_{t_-}|}} \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t_-}|} \left(1 - \mathrm{P}_s\left(x_{t_-}^i\right)\right) \quad . \tag{6.9}$$
$$\times \prod_{i=1}^{|X^s|} f\left(x^{s,i} \mid x_{t_-}^{\omega_i}\right) \mathrm{P}_s\left(x_{t_-}^{\omega_i}\right)$$

This can be seen from the marginalization of the labeled union individual surviving targets and a set of arriving targets, as was discussed in chapter 4. The component labeling of X_t would then first be describing if it is a newly arrived target or a surviving target, and the labeling of individual elements in the surviving set would mean which of the set elements in X_{t_-} it transitioned from. This is represented in the above as the set convolution over surviving and arriving targets, and the sum over the permutation variable ω, respectively.

6.3 The multi target measurement function

Assumption A-7, similarly as assumption A-4, states that false alarms follow that of a PPP with intensity $\mu_t(z) = \bar{\mu}_t h_t^{\mu}(z)$. Typically a sensor has a resolution in space, and its measurements are then actually over a discrete grid. If one then assumes that each grid cell has a Bernoulli process for generating false alarms, and that the resolution is large enough, we have, again by Le Cam [1960], that the PPP is a good descriptor of the false alarm process.

Also similar to the multi target transition function, assumptions A-5 and A-6 tell us that a target originated singleton set measurement, Z, given a single target, x, is a BPP with existence probability given by $\mathrm{P}_d(x)$ and existence conditioned distribution given by $h(z \mid x)$ such that

$$h(Z \mid x) = \begin{cases} h(z \mid x)\, \mathrm{P}_d(x)\,, & Z = \{z\} \\ 1 - \mathrm{P}_d(x)\,, & Z = \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{6.10}$$

Therefore, using the same labeling and marginalization argument of the transition function, we have that the total multi target measurement function is also a PMB

$$h(Z_t \mid X_t) = \sum_{\substack{Z^d \subseteq Z_t: \\ |Z^d| \leq |X_t|}} \mathrm{e}^{-\bar{\mu}_t} \mu_t^{Z_t \setminus Z^d} \sum_{\omega^d \in \mathcal{P}^{|X_t|}_{|Z^d|}} \prod_{\substack{i=1: \\ i \notin \omega^d}}^{|X_t|} \left(1 - \mathrm{P}_d\left(x_t^i\right)\right) \prod_{j=1}^{|Z^d|} h\left(z^{d,i} \mid x_t^{\omega^d}\right) \mathrm{P}_d\left(x_t^{\omega^d}\right)$$
$$= \mathrm{e}^{-\bar{\mu}_t} \sum_{\omega \in \mathcal{Q}^{|X_t|}_{|Z_t|}} \prod_{\substack{j=1: \\ \omega_j = 0}}^{|Z_t|} \mu_t \left(z_t^j\right) \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_t|} \left(1 - \mathrm{P}_d\left(x_t^i\right)\right) \prod_{\substack{j=1: \\ \omega_j > 0}}^{Z_t} h\left(z_t^j \mid x_t^{\omega_j}\right) \mathrm{P}_d\left(x_t^{\omega_j}\right), \tag{6.11}$$

with

$$\mathcal{Q}_n^N = \{q_j \in [0:N] \,\forall\, j \in [1:n] | q_j = i > 0 \implies q_{j'} \neq i \,\forall\, j, j' \in [1:n]\}, \tag{6.12}$$

which is equivalent to the description by Mahler [2007, pp. 421]. Here, we have also used a measurement to target association variables ω, where one can see that $\omega_j = 0$ implies false alarm, $\omega_j = i > 0$ implies a specific measurement to target association, and $\mathcal{Q}_n^N$ is the set of feasible associations that satisfies assumption A-8. It is also possible to formulate this with a target to measurement variables, σ, as

$$h(Z_t \mid X_t) = \mathrm{e}^{-\bar{\mu}_t} \sum_{\sigma \in \mathcal{Q}^{|Z_t|}_{|X_t|}} \prod_{\substack{j=1: \\ j \notin \sigma}}^{|Z_t|} \mu_t \left(z_t^j\right) \prod_{\substack{i=1: \\ \sigma_i = 0}}^{|X_t|} \left(1 - \mathrm{P}_d\left(x_t^i\right)\right) \prod_{\substack{i=1: \\ \sigma_i > 0}}^{|X_t|} h\left(z_t^{\sigma_i} \mid x_t^i\right) \mathrm{P}_d\left(x_t^i\right), \tag{6.13}$$

Similarly to the meaning of the measurement to target association variables, the target to measurement variables we have that $\sigma_i = 0$ implies no detection, and $\sigma_i = j > 0$ implies that target i is associated with measurement j. It should also be clear that the target to measurement and measurement to target formulations are equivalent. This means that given ω we also know σ, and the the other way around. In other terms

$$\mathcal{Q}^{|Z_t|}_{|X_t|} \iff \mathcal{Q}^{|X_t|}_{|Z_t|}: \tag{6.14}$$
$$\sigma_i = j > 0 \iff \omega_j = i > 0, \,\forall\, i, j. \tag{6.15}$$

or, perhaps even simpler;

$$\sigma_i = j > 0 \implies \omega_{\sigma_i} = i \tag{6.16}$$

Where the last line in some sense indicates that σ can be thought of as the inverse of ω since it has the same information, but points in the other direction.

The last formulation we are going to look at is the over-representation where we use both σ and ω, and we introduce a function $\gamma(\sigma, \omega)$ to indicate if (6.15) is true;

$$\gamma(\sigma, \omega) = \mathbf{1}[\sigma_i = j > 0 \iff \omega_j = i > 0, \,\forall\, i, j]. \tag{6.17}$$

With this we write the multi target measurement function as

$$h(Z_t \mid X_t) = \mathrm{e}^{-\bar{\mu}_t} \sum_{\substack{\sigma \in [0:|Z_t|]^{|X_t|}, \\ \omega \in [0:|X_t|]^{|Z_t|}}} \gamma\left(\sigma, \omega\right) \prod_{\substack{j=1: \\ \omega_j = 0}}^{|Z_t|} \mu_t\left(z_t^j\right) \prod_{\substack{i=1: \\ \sigma_i = 0}}^{|X_t|} \left(1 - \mathrm{P}_d\left(x_t^i\right)\right) \prod_{\substack{(i,j) \in [1:|X_t|] \times [1:|Z_t|]: \\ \sigma_i = j,\ \omega_j = i}} h\left(z_t^j \mid x_t^i\right) \mathrm{P}_d\left(x_t^i\right), \tag{6.18}$$

where the set to be summed over is extended to include non feasible associations, since $\gamma = 0$ in those cases anyway.

As we shall see later, and also stated by Mahler [2007], the sum in the above equations, over what could be interpreted as measurement origin, can be seen as one of, if not the, greatest difficulties in MTT.

7 The multi target filter

Using the notions discussed in chapter 6, we are now going to have a look at what the optimal filter will look like, first starting with a look on the birth process and the distribution it induces, then on incorporating measurements and how that changes the distribution. It is also going to be pointed out how labeling can be achieved and how to handle a hybrid state space. At last we are going to describe some implications of the modeling towards the initial distribution of targets.

7.1 The birth process and undetected targets

Let us first consider that there are no targets at time 0. From assumption A-4, we then know that the prediction to time 1 must give a PPP for the targets, since it will be a union of the arriving targets, that follows a PPP with intensity η, and no targets (can be seen as a PPP with intensity zero). At the time 2, the targets that arrived at time 1 survive with a certain probability and a new set of targets arrive. We then want the distribution of the union of the surviving targets and the newly arrived targets. Note that there is no detection or any sensor information involved yet. We are simply following the assumptions regarding target arrival, which apparently tell us to consider distributions over targets that are yet to be detected by any sensor. This quantity will be denoted by $\lambda_{t\,|\,t'}(x)$; read as the intensity of undetected targets at time t given information up to and including time t'. If we have no targets at time 0 we have $\lambda_{0\,|\,0}(x) = 0$ and $\lambda_{1\,|\,0}(x) = \eta_1(x)$ from this reasoning.

So, assumption A-4 tells us that the arrival process is a PPP. We know that the union of several PPPs is again a PPP, and are therefore going to assume that the arrival process induces a PPP that we need to handle. This will be called the SDF of undetected targets since it comes from the distribution of arriving targets, which may or may not be detected, and will be seen later to have prediction and update steps that only deal with the undetected portion of the distribution.

We are going to see what the prediction step, and then subsequently the update step will do with this undetected distribution. We shall see that the prediction step keeps the PPP form, which is no surprise if one knows of the PHD filter of cahpter 16 and its connection to the PPP given in secton 16.2.1.2 by Mahler [2007], or the Poisson distributions relationship to the binomial distribution. The update step on the other hand, will not keep the form and will induce a MB in addition to the PPP. This will be seen as the result of the measurements being finitely many, unordered, and possibly false alarms, and will automatically provide us with descriptions of the measurement origin uncertainty and target existence uncertainty.

7.1.1 Prediction of undetected targets

Say we have estimated an undetected target PPP to have an intensity of $\lambda_{t_- \mid t_-}(x)$ for some t_-, for brevity simply denoted by $\lambda(x)$ in this section. We will also use

$$\lambda^s(x_t, x_{t_-}) = f\left(x \mid x_{t_-}^{r,i}\right) \mathrm{P}_s\left(x_{t_-}^{r,i}\right) \lambda\left(x_{t_-}^{r,i}\right), \tag{7.1}$$

$$\mathrm{P}_s\left(x_{t_-}^i\right) \lambda\left(x_{t_-}^i\right) = \int_{x' \in \mathcal{X}} \lambda^s\left(x', x_{t_-}^i\right) \mathrm{d}x' = \int_{x' \in \mathcal{X}} f\left(x' \mid x_{t_-}^i\right) \mathrm{P}_s\left(x_{t_-}^i\right) \lambda\left(x_{t_-}^i\right) \mathrm{d}x' \tag{7.2}$$

and

$$\begin{aligned} \lambda^s(x_t) &= \int_{x_{t_-} \in \mathcal{X}} \lambda^s(x_t, x_{t_-}) \, \mathrm{d}x_{t_-} \\ &= \int_{x_{t_-} \in \mathcal{X}} f\left(x \mid x_{t_-}^{r,i}\right) \mathrm{P}_s\left(x_{t_-}^{r,i}\right) \lambda\left(x_{t_-}^{r,i}\right) \mathrm{d}x_{t_-} \end{aligned} \tag{7.3}$$

to try and make life a bit simpler. We are interested to see what our new multi target estimate should be at the next time step, in a similar manner to the standard Bayesian state estimation.

To predict the SDF of undetected targets forward, we first form the joint over the two time steps. By using superscript s to denote surviving targets at time t, the joint distribution over X_{t_-} and X_t will be the multiplication of the PPP of undetected targets and the conditional PMB for the transition of multiple target, given by (6.9);

$$\begin{aligned} p\left(X_t, X_{t_-}\right) = \mathrm{e}^{-\bar{\lambda}} \lambda^{X_{t_-}} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t_-}}} \mathrm{e}^{-\bar{\eta}_t} \eta_t^{X_t \setminus X^s} \sum_{\omega \in \mathcal{P}_{|X^s|}^{|X_{t_-}|}} \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t_-}|} \left(1 - \mathrm{P}_s\left(x_{t_-}^i\right)\right) \\ \times \prod_{i=1}^{|X^s|} f\left(x^{s,i} \mid x_{t_-}^{\omega_i}\right) \mathrm{P}_s\left(x_{t_-}^{\omega_i}\right) \end{aligned}$$

distributing $\lambda^{X_{t_-}}$ and $\eta_t^{X^s}$ into the products;

$$= \mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t_-}}} \sum_{\omega \in \mathcal{P}_{|X^s|}^{|X_{t_-}|}} \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t_-}|} \left(1 - \mathrm{P}_s\left(x_{t_-}^i\right)\right) \lambda\left(x_{t_-}^i\right)$$

$$\times \prod_{i=1}^{|X^s|} \frac{f\left(x^{s,i} \mid x_{t_-}^{\omega_i}\right)}{\eta_t(x_{t_-}^{\omega_i})} \mathrm{P}_s(x_{t_-}^{\omega_i})\, \lambda(x_{t_-}^{\omega_i})$$

substituting in (7.1) and (7.2), and multiplying out the parenthesis;

$$= \mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t_-}}} \sum_{\omega \in \mathcal{P}_{|X^s|}^{|X_{t_-}|}} \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t_-}|} \left(\lambda\left(x_{t_-}^i\right) - \int_{x' \in \mathcal{X}} \lambda^s\left(x', x_{t_-}^i\right) \mathrm{d}x'\right) \tag{7.4}$$

$$\times \prod_{i=1}^{|X^s|} \frac{\lambda^s\left(x^{s,i}, x_{t_-}^{\omega_i}\right)}{\eta_t(x^{s,i})}.$$

Now we want to marginalize over X_{t_-} in (7.4) using the set integral to get the predicted multi target state SDF. First we see that we can split the integral over the surviving and non-surviving components. We also have that the targets in X_{t_-} are i.i.d., so the permutations after marginalization will give the same outcome and can therefore be treated by multiplying in a combinatorial constant. After integration, the last sum will simply turn into the number of ways to choose the surviving elements in X^s out of X_{t_-} in order, which are $\frac{|X_{t_-}|!}{(|X_{t_-}|-|X_t^s|)!}$. Lastly we have that the integral of the different non-surviving targets will become the same constant. This will give

$$p(X_t) = \int p(X_t, X_{t_-})\, \delta X_{t_-} \tag{7.5}$$

inserting (7.4) for the SDF and (4.6) for the set integral;

$$= \sum_{|X_{t_-}|=0}^{\infty} \frac{\mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t}}{|X_{t_-}|!} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t_-}}} \sum_{\omega \in \mathcal{P}_{|X^s|}^{|X_{t_-}|}} \tag{7.6}$$

$$\prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t_-}|} \int_{x_{t_-}^i \in \mathcal{X}} \left(\lambda\left(x_{t_-}^i\right) - \int_{x' \in \mathcal{X}} \lambda^s\left(x', x_{t_-}^i\right) \mathrm{d}x'\right) \mathrm{d}x_{t_-}^i$$

$$\times \prod_{i=1}^{|X^s|} \int_{x_{t_-}^{\omega_i} \in \mathcal{X}} \frac{\lambda^s\left(x^{s,i}, x_{t_-}^{\omega_i}\right)}{\eta_t(x^{s,i})} \mathrm{d}x_{t_-}^{\omega_i}$$

performing the integrals using (7.3) and $\bar{\lambda} = \int\limits_{x \in \mathcal{X}} \lambda(x)\,\mathrm{d}x$;

$$= \sum_{|X_{t\text{-}}|=0}^{\infty} \frac{\mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t}}{|X_{t\text{-}}|!} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t\text{-}}}} \sum_{\omega \in \mathcal{P}^{|X_{t\text{-}}|}_{|X^s|}} \prod_{\substack{i=1: \\ i \notin \omega}}^{|X_{t\text{-}}|} \left(\bar{\lambda} - \bar{\lambda}^s\right) \times \prod_{i=1}^{|X^s|} \frac{\lambda^s(x^{s,i})}{\eta_t(x^{s,i})} \tag{7.7}$$

inserting the combinatorial constant for the last sum and rewriting the products;

$$= \sum_{|X_{t\text{-}}|=0}^{\infty} \frac{\mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t}}{|X_{t\text{-}}|!} \sum_{\substack{X^s \subseteq X_t: \\ |X^s| \leq X_{t\text{-}}}} \frac{|X_{t\text{-}}|!}{(|X_{t\text{-}}|-|X^s|)} \left(\bar{\lambda} - \bar{\lambda}^s\right)^{(|X_{t\text{-}}|-|X^s|)} \left(\frac{\lambda^s}{\eta}\right)^{X^s} \tag{7.8}$$

interchanging the sums and cancelling $|X_{t\text{-}}|!$;

$$= \mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t} \sum_{X^s \subseteq X_t} \left(\frac{\lambda^s}{\eta}\right)^{X^s} \sum_{|X_{t\text{-}}|=0}^{\infty} \frac{1}{(|X_{t\text{-}}| - |X^s|)} \left(\bar{\lambda} - \bar{\lambda}^s\right)^{(|X_{t\text{-}}|-|X^s|)} \tag{7.9}$$

recognising the power series of the exponential function;

$$= \mathrm{e}^{-\bar{\lambda}-\bar{\eta}_t}\eta_t^{X_t} \sum_{X^s \subseteq X_t} \left(\frac{\lambda^s}{\eta}\right)^{X^s} \mathrm{e}^{\bar{\lambda}-\bar{\lambda}^s} = \mathrm{e}^{-\bar{\lambda}^s-\bar{\eta}_t} \sum_{X^s \subseteq X_t} \eta_t^{X_t \setminus X^s} \left(\lambda^s\right)^{X^s} \tag{7.10}$$

recognizing that this is union of two PPPs, given by (4.14) to be a new PPP;

$$= \mathrm{e}^{-\bar{\lambda}^s-\bar{\eta}_t} \left(\lambda^s + \eta_t\right)^{X_t}. \tag{7.11}$$

This gives us the total predicted distribution to time t for the undetected targets conditioned on information up $t' < t$ as

$$\lambda_{t\,|\,t'}(x) = \eta_t(x) + \lambda^s_{t\,|\,t'}(x) = \eta_t(x) + \int\limits_{x' \in \mathcal{X}} f(x\,|\,x')\,\mathrm{P}_s(x')\,\lambda_{t\text{-}\,|\,t'}(x')\,\mathrm{d}x'. \tag{7.12}$$

This is the same result as Williams [2015b, (37)] has, and equivalent to the PHD predictor of Mahler [2007, (16.95) pp.589] without spawning.

Another interpretation on why we get back a PPP and not a PMB, which one might expect since the multi target transition function indeed is a PMB, is that the Poisson cardinality can be seen as the limit of the binomial distribution with infinitely many “trials” that give rise to i.i.d. targets. The i.i.d. MB transition of these targets are essentially just i.i.d. BPP updates of all the BPPs in the binomial, and hence the limit still applies. Bounds on the relations between a multi Bernoulli probability mass function[1] and the Poisson probability mass function are given with the original credits to Le Cam [1960], but perhaps simpler described by for instance Serfling [1978], in which the binomial probability mass function is of course a special case.

[1] Not meaning the point process right here, but the distribution over the integers, and better known as Poisson binomial

Hybrid state space

Having a hybrid state space in the undetected targets can be dealt with in a straightforward fashion. Tracking one intensity per discrete state can gained by expanding the transition and changing an integration to a summation, and gives

$$\lambda^{l}_{t\,|\,t_-}(x) = \eta^{l}_{t}(x) + \sum_{l' \in \mathcal{L}} \int_{x' \in \mathcal{X}} f^{l}(x\,|\,x')\, \pi^{l}_{l'}(x')\, \lambda^{l'}_{t_-\,|\,t_-}(x')\, \mathrm{d}x'. \tag{7.13}$$

One now has to do this for each of the discrete states l. The the number of targets is now gained by summing over the discrete states and integrating out the continuous state

$$\bar{\lambda}_{t\,|\,t_-} = \sum_{l \in \mathcal{L}} \int_{x \in \mathcal{X}} \lambda^{l}_{t\,|\,t_-}(x)\, \mathrm{d}x \tag{7.14}$$

and the discrete state probability as

$$\mu^{l,\lambda}_{t\,|\,t_-} = \int_{x \in \mathcal{X}} \frac{\lambda^{l}_{t\,|\,t_-}(x)}{\bar{\lambda}_{t\,|\,t_-}}\, \mathrm{d}x. \tag{7.15}$$

7.1.2 Measurements of the undetected targets

We now want to see what happens to the undetected target SDF when we receive measurements. We know from the previous section that the prediction step keeps the form when it is a PPP. By this we proceed to assume that the target SDF prior to measurement is a PPP.

To simplify notation through the derivation, we will use the current undetected target intensity

$$\lambda(x) = \lambda_{t\,|\,t_-}(x)\,, \tag{7.16}$$

the updated undetected target intensity

$$\lambda^{+}(x) = \lambda_{t\,|\,t}(x) = (1 - \mathrm{P}_d(x))\,\lambda(x)\,, \tag{7.17}$$

the joint undetected target and measurement intensity

$$\Lambda(z, x) = h(z\,|\,x)\, \mathrm{P}_d(x)\, \lambda(x)\,, \tag{7.18}$$

the undetected target measurement intensity

$$\Lambda(z) = \int_{x \in \mathcal{X}} \Lambda(z, x)\, \mathrm{d}x = \int_{x \in \mathcal{X}} h(z\,|\,x)\, \mathrm{P}_d(x)\, \lambda(x)\, \mathrm{d}x \tag{7.19}$$

the expected number of detections of the undetected targets

$$\bar{\Lambda} = \int_{z \in \mathcal{Z}} \Lambda(z)\, \mathrm{d}z, \tag{7.20}$$

and the expected number of undetected targets after update

$$\bar{\lambda}^{+} = \int\limits_{x\in\mathcal{X}} \lambda^{+}(x)\,\mathrm{d}x = \int\limits_{x\in\mathcal{X}} (1-\mathrm{P}_d(x))\,\lambda(x)\,\mathrm{d}x = \bar{\lambda} - \bar{\Lambda}. \tag{7.21}$$

The joint of the undetected target and measurements is given by multiplying the undetected target SDF, $f^{\lambda}(X)$ by the multi target measurement equation (6.11), and is given by

$$\begin{aligned} p_u(Z_t, X_t) &= h(Z_t \mid X_t)\, f^{\lambda}(X_t) && (7.22)\\ &= \mathrm{e}^{-\bar{\lambda}}\,\lambda^{X_t}\mathrm{e}^{-\bar{\mu}_t} \sum_{\omega\in\mathcal{Q}^{|X_t|}_{|Z_t|}} \prod_{\substack{j=1:\\ \omega_j=0}}^{|Z^t|} \mu_t\left(z_t^j\right) \prod_{\substack{i=1:\\ i\notin\omega}}^{|X_t|} \left(1-\mathrm{P}_d(x_t^i)\right) \prod_{\substack{j=1:\\ \omega_j>0}}^{|Z_t|} h\!\left(z_t^j \mid x_t^{\omega_j}\right) \mathrm{P}_d\!\left(x_t^{\omega_j}\right), \end{aligned}$$

distributing λ^{X_t} and $\left\{\mu_t\left(z_t^j\right) | \omega_j = 0\right\}$ into the products;

$$= \mathrm{e}^{-\bar{\lambda}}\mathrm{e}^{-\bar{\mu}_t}\mu_t^Z \sum_{b_t\in\mathcal{Q}^{|X_t|}_{|Z_t|}} \prod_{\substack{i=1:\\ i\notin\omega}}^{|X_t|} \left(1-\mathrm{P}_d(x_t^i)\right)\lambda(x_t^i) \prod_{\substack{j=1:\\ \omega_j>0}}^{|Z_t|} \frac{h\!\left(z_t^j \mid x_t^{\omega_j}\right)\mathrm{P}_d\!\left(x_t^{\omega_j}\right)\lambda\!\left(x_t^{\omega_j}\right)}{\mu_t\left(z_t^j\right)},$$

using (7.17) and (7.18);

$$= \mathrm{e}^{-\bar{\lambda}}\mathrm{e}^{-\bar{\mu}_t}\mu_t^Z \sum_{b_t\in\mathcal{Q}^{|X_t|}_{|Z_t|}} \prod_{\substack{i=1:\\ i\notin b_t}}^{|X_t|} \lambda^{+}(x_t^i) \prod_{\substack{j=1:\\ \omega_j>0}}^{|Z_t|} \frac{\Lambda\!\left(z_t^j, x_t^{\omega_j}\right)}{\mu_t\left(z_t^j\right)}. \tag{7.23}$$

In general we are interested in $f(X \mid Z)$, which needs the invocation of Bayes theorem along with the marginal of the joint with respect to X. We are going to pursue the marginal, but will begin first with some notes. The states are i.i.d., so when taking the marginal, there is no difference in which target the measurement originated from, and the sum over target permutations can be interchanged with a combinatorial multiplication. This will be the number of ways to choose target originated measurements, $|Z_t^d|$ from the targets $|X_t|$ in order, namely $\frac{|X_t|!}{(|X_{tim}|-|Z_t^d|)!}$. With this, the set integral of (7.23) with respect to X_t becomes

$$h^{\lambda}(Z_t) = \int p_u(Z_t, X_t)\,\delta X_t \tag{7.24}$$

$$= \sum_{|X_t|=0}^{\infty} \frac{\mathrm{e}^{-\bar{\lambda}}\mathrm{e}^{-\bar{\mu}_t}\mu_t^Z}{|X_t|!} \sum_{\omega\in\mathcal{Q}^{|X_t|}_{|Z_t|}} \prod_{\substack{i=1:\\ i\notin\omega}}^{|X_t|} \int\limits_{x\in\mathcal{X}} \lambda^{+}(x_t^i)\,\mathrm{d}x \prod_{\substack{j=1:\\ \omega_j>0}}^{|Z|} \int\limits_{x\in\mathcal{X}} \frac{\Lambda\!\left(z_t^j, x_t^{\omega_j}\right)}{\mu_t\left(z_t^j\right)}\,\mathrm{d}x \tag{7.25}$$

using (7.19) and (7.21) and inserting for the combinatorial for a part of the sum;

$$= \sum_{|X_t|=0}^{\infty} \frac{\mathrm{e}^{-\bar{\lambda}}\mathrm{e}^{-\bar{\mu}_t}\mu_t^{Z_t}}{|X_t|!} \sum_{\substack{Z^d\subseteq Z:\\ |Z^d|\le|X_t|}} \frac{|X_t|!}{(|X_t|-|Z_t^d|)!} (\bar{\lambda}^{+})^{(|X_t|-|Z_t^d|)} \left(\frac{\Lambda}{\mu_t}\right)^{Z_t^d} \tag{7.26}$$

canceling $|X_t|!$ and taking the sum over $|X_t|$, recognizing the power series for the exponential function;

$$= \mathrm{e}^{-\bar{\lambda}} \mathrm{e}^{-\bar{\mu}_t} \mu_t^{Z_t} \sum_{Z_t^d \subseteq Z_t} \mathrm{e}^{\bar{\lambda}^+} \left(\frac{\Lambda}{\mu_t} \right)^{Z_t^d} \tag{7.27}$$

using the last part of (7.21) and factoring togheter the exponentials in e;

$$= \mathrm{e}^{-\bar{\lambda} - \bar{\mu}_t + \bar{\lambda} - \bar{\Lambda}} \mu_t^{Z_t} \sum_{Z_t^d \subseteq Z_t} \left(\frac{\Lambda}{\mu_t} \right)^{Z_t^d} = \mathrm{e}^{-\bar{\mu}_t - \bar{\Lambda}} \mu_t^{Z_t} \sum_{Z_t^d \subseteq Z_t} \left(\frac{\Lambda}{\mu_t} \right)^{Z_t^d} \tag{7.28}$$

recognizing this as the union of two PPPs to use (4.14) ;

$$= \mathrm{e}^{-(\bar{\mu}_t + \bar{\Lambda})} (\mu_t + \Lambda)^{Z_t}, \tag{7.29}$$

and we see that the total measurement SDF is a new PPP. This should not come as a surprise as this is very similar to the setup of the prediction equations where the surviving targets would take on the role of the measurements and new targets would take the role of false alarms in measurement space.

Conditioning (7.23) on Z_t will now give

$$f^{\lambda}(X_t | Z_t) = \frac{h(Z_t \,|\, X_t)\, f^{\lambda}(X_t)}{h^{\lambda}(Z_t)} \tag{7.30}$$

$$= \frac{\mathrm{e}^{-\bar{\lambda}} \mathrm{e}^{-\bar{\mu}_t}}{\mathrm{e}^{-(\bar{\mu}_t + \bar{\Lambda})}} \left(\frac{\mu_t}{\mu_t + \Lambda} \right)^{Z_t} \sum_{\omega \in \mathcal{Q}_{|Z_t|}^{|X_t|}} \prod_{\substack{i=1:\\ i \notin \omega}}^{|X_t|} \lambda^+(x_t^i) \prod_{\substack{j=1:\\ \omega_j > 0}}^{|Z_t|} \frac{\Lambda\left(z_t^j, x_t^{\omega_j}\right)}{\mu_t\left(z_t^j\right)}, \tag{7.31}$$

simplifying the exponential, expanding the first product and distributing it;

$$= \mathrm{e}^{-\bar{\lambda}^+} \sum_{\omega \in \mathcal{Q}_{|Z_t|}^{|X_t|}} \prod_{\substack{j=1:\\ \omega_j = 0}}^{|Z_t|} \frac{\mu_t\left(z_t^j\right)}{\mu_t\left(z_t^j\right) + \Lambda\left(z_t^j\right)} \prod_{\substack{i=1:\\ i \notin \omega}}^{|X_t|} \lambda^+(x_t^i) \prod_{\substack{j=1:\\ \omega_j > 0}}^{|Z_t|} \frac{\Lambda\left(z_t^j, x_t^{\omega_j}\right)}{\mu_t\left(z_t^j\right) + \Lambda\left(z_t^j\right)}, \tag{7.32}$$

this can also be written using σ instead of ω as described earlier ;

$$= \mathrm{e}^{-\bar{\lambda}^+} \sum_{\sigma \in \mathcal{Q}_{|X_t|}^{|Z_t|}} \prod_{\substack{j=1:\\ j \notin \sigma}}^{|Z_t|} \frac{\mu_t\left(z_t^j\right)}{\mu_t\left(z^j\right) + \Lambda(z^j)} \prod_{\substack{i=1:\\ \sigma_i = 0}}^{|X_t|} \lambda^+(x_t^i) \prod_{\substack{i=1:\\ \sigma_i > 0}}^{|X_t|} \frac{\Lambda(z_t^{\sigma_i}, x_t^i)}{\mu_t\left(z_t^{\sigma_i}\right) + \Lambda(z_t^{\sigma_i})}, \tag{7.33}$$

identifying $r^{(t,\sigma_i)}$ from (7.37) and splitting the sum;

$$= \mathrm{e}^{-\bar{\lambda}^+} \sum_{\substack{X_t^d \subseteq X_t:\\ |X_t^d| \le |Z_t|}} (\lambda^+)^{X_t \setminus X_t^d} \sum_{\sigma \in \mathcal{P}_{|X_t^d|}^{|Z_t|}} \prod_{\substack{j=1:\\ j \notin \sigma}}^{|Z_t|} \left(1 - r^{(t,j)}\right) \prod_{\substack{i=1:\\ \sigma_i > 0}}^{|X_t|} r^{(t,\sigma_i)} \frac{\Lambda\left(z_t^{\sigma_i}, x_t^{d,i}\right)}{\Lambda(z_t^{\sigma_i})}, \tag{7.34}$$

also identifying $f^{(t,j)}(x)$ in (7.37) ;

$$= \mathrm{e}^{-\bar{\lambda}^+} \sum_{\substack{X_t^d \subseteq X_t: \\ |X_t^d| \le |Z_t|}} (\lambda^+)^{X_t \setminus X_t^d} \sum_{\sigma \in \mathcal{P}^{|Z_t|}_{|X_t^d|}} \prod_{\substack{j=1: \\ j \notin \sigma}}^{|Z_t|} \left(1 - r^{(t,j)}\right) \prod_{\substack{i=1: \\ \sigma_i > 0}}^{|X_t|} r^{(t,\sigma_i)} f^{(t,\sigma_i)}\left(x_t^{d,i}\right), \tag{7.35}$$

which should be recognized as a PMB, since the first part is clearly a PPP, the latter a MB and the first sum corresponds to the set convolution. This shows that having a prior PPP with intensity $\lambda_{t\,|\,t_-}(x)$ for undetected targets, and then receiving measurements will create a posterior that is a union of the still undetected targets that follow a PPP with intensity $\lambda_{t\,|\,t}(x)$, and a MB for possible target detections given by

$$\text{undetected} \left\{ \lambda_{t\,|\,t}(x) = \lambda^+(x) = (1 - \mathrm{P}_d(x))\lambda_{t\,|\,t_-}(x) \right., \tag{7.36}$$

$$\text{detected hypotheses} \begin{cases} f^{(t,j)}(x) = \dfrac{\Lambda\left(z_t^j, x\right)}{\Lambda\left(z_t^j\right)} = \dfrac{h\left(z_t^j \mid x\right) \mathrm{P}_d(x)\, \lambda_{t\,|\,t_-}(x)}{\displaystyle\int_{x \in \mathcal{X}} h\left(z_t^j \mid x\right) \mathrm{P}_d(x)\, \lambda_{t\,|\,t_-}(x)\, \mathrm{d}x}, \\[2ex] r^{(t,j)} = \dfrac{\Lambda(z^j)}{\mu_t\,(z^j) + \Lambda(z^j)} = \dfrac{\displaystyle\int_{x \in \mathcal{X}} h\left(z_t^j \mid x\right) \mathrm{P}_d(x)\, \lambda_{t\,|\,t_-}(x)\, \mathrm{d}x}{\mu_t\left(z_t^j\right) + \displaystyle\int_{x \in \mathcal{X}} h\left(z_t^j \mid x\right) \mathrm{P}_d(x)\, \lambda_{t\,|\,t_-}(x)\, \mathrm{d}x} \end{cases} \tag{7.37}$$

respectively. This result is the same as Williams [2015b, (42),(56),(57)]. We will later introduce a better superscript notation to accommodate a growing set of hypotheses and possible targets. In fact, it will later be shown that we have to consider that every measurement could be a previously undetected target, as well as any of the previously detected targets.

Also, note that there is an existence probability which is less than unity in the case of false alarm. This means that the notion of detected targets might give the wrong impression, since it is only a potential target. We shall therefore call every Bernoulli component of the detected targets as a track, which potentially could be related to a real target and potentially be a false target, i.e. clutter. In writing, this distinction might sometimes slip, but the a Bernoulli component should always be thought of as a potentially erroneous target. Hence, a true target can be related to any of these tracks, and we only know which probabilistically. Since we have to consider that each new measurement could be a never target that is never seen before, a natural indexing of the BPP is to use a linear index over the measurements ordered in time and according to some, arbitrary but specific, ordering within each scan. We shall adopt that the Bernoulli component i was created by measurement j at time t such that

$$i = j + \sum_{\tau=0}^{t-1} |Z_\tau|, \tag{7.38}$$

sometimes abbreviated

$$i = (t, j). \tag{7.39}$$

A guess now would be that the total target distribution also has this form, a PMB. We will later see that this is not the case, due to data associations at later sensor scans and the MB part. Instead of a PMB we will have a PMBM with weights that are given by the data association probabilities.

Labeling

Like before we can interpret the permutation sums as a marginalization of a latent component label and include this into the SDF. Using π as the latent component label, and the component labeled multi target state as $X_t^L = \left\{(x_t^i, \pi^i)\right\}$ where $\pi^i = 0$ meaning still undetected, we can write the component labeled distribution as

$$f^{\lambda}\left(X_t^L \mid Z_t\right) = \mathrm{e}^{-\bar{\lambda}^+} \left(\lambda^+\right)^{X_t^{L(0)}} \prod_i f^i\left(X_t^{L(i)}\right). \tag{7.40}$$

Here $f^i(X)$ is the BPP component coming from measurement $i = (t, j)$, and its cardinality distribution takes care of the zero likelihood case of multiple targets having the same BPP label. It is important to note that marginalizing out π will give the distribution derived earlier, as seen by the former actually just being the sum over π.

7.2 Detected targets

The last subsection pointed to that the true distribution of the targets under the current assumptions is that of a union of a PPP and a MB. To proceed we are going to assume that we have the more general distribution of a PMBM, which is the union of a PPP and a multi Bernoulli mixture (MBM). These components will be denoted by

$$f_{t\,|\,t'}^{\mathrm{ppp}}(X^u) = \mathrm{e}^{-\bar{\lambda}_{t\,|\,t'}} \lambda_{t\,|\,t'}^{X^u} \qquad f_{t\,|\,t'}^{\mathrm{mbm}}\left(X^d\right) = \sum_{\theta \in \Theta_t} w_{t\,|\,t'}^{\theta} f_{t\,|\,t'}^{\theta}\left(X^d\right) \tag{7.41}$$

respectively, where Θ are the set of hypotheses in the mixture, w^{θ} is the hypothesis probability, or weight, and $f^{\theta}\left(X^d\right)$ is the hypothesis conditioned MB. Note that a MB is a special case of a MBM with only a single component in the mixture. With this the PMBM SDF is given as

$$f_{t\,|\,t'}^{\mathrm{pmbm}}(X_t) = \sum_{X_t^d \subseteq X_t} f_{t\,|\,t'}^{\mathrm{ppp}}\left(X_t \setminus X_t^d\right) f_{t\,|\,t'}^{\mathrm{mbm}}\left(X_t^d\right). \tag{7.42}$$

In what follows we are going to see that the prediction and update step will maintain this form, but have an increase in the number of hypotheses after the update step, and therefore not be a true conjugate prior[1]. Nevertheless, having the same form makes algorithm development and approximations easier, and it is therefore a remarkable result, even though it is computationally unfeasible due to the expanding parameter space over longer times or larger problems.

The BPP in the MB can be thought of as a single track or a potential target, and we will therefore use the hypothesis conditioned notation of

$$f^{i,\theta^i}_{t\,|\,t'}(x) \tag{7.43}$$

to denote the existence conditioned state PDF for track/component i, and

$$r^{i,\theta^i}_{t\,|\,t'} \tag{7.44}$$

for the existence probability of track/component i, such that the BPP SDF becomes

$$f^{i,\theta^i}_{t\,|\,t'}(X) = \begin{cases} r^{i,\theta^i}_{t\,|\,t'} f^{i,\theta^i}_{t\,|\,t'}(x), & X = \{x\}, \\ 1 - r^{i,\theta^i}_{t\,|\,t'}, & X = \emptyset, \\ 0, & \text{otherwise}. \end{cases} \tag{7.45}$$

The difference between the BPP and the state distribution is seen by the argument, as the former takes sets and the latter vectors.

7.2.1 Prediction of the detected targets.

Since the undetected and the detected components are independent in the posterior at t_-, their respective predictions are also independent due to i.i.d. Markovian dynamics. The prediction of the detected components MBM can therefore be derived independently of the undetected components PPP. Furthermore, since the joint over two time steps is factored into the multi target transition PMB and a mixture, it is itself a mixture over the very same hypotheses. Hence we only have to treat the prediction of a MB in detail, which according to intuition should be another MB with existence probabilities and state PDFs given by the intuitive formulas. To see all this we extract a part of the set convolution of the multi target

[1] A conjugate prior is when one can simply update the parameters of the prior distribution to get the posterior. Here we need to add new ones, and even though they are of the same form, it does not make a true conjugate prior.

transition density and get

$$f^{\mathrm{pmbm}}_{t\,|\,t_-}(X_t, X_{t_-}) = f(X_t \,|\, X_{t_-})\, f^{\mathrm{pmbm}}_{t_-\,|\,t_-}(X_{t_-}) \tag{7.46}$$

$$= \sum_{\substack{X^d_t \subseteq X_t,\\ X^d_{t_-} \subseteq X_{t_-}}} f\left(X_t \setminus X^d_t \,|\, X_{t_-} \setminus X^d_{t_-}\right) f^{\mathrm{ppp}}_{t_-\,|\,t_-}\left(X_{t_-} \setminus X^d_{t_-}\right) \tag{7.47}$$

$$\sum_{\theta\in\Theta} w^\theta_{t_-}\, f^s\left(X^d_t \,|\, X^d_{t_-}\right) f^\theta_{t_-\,|\,t_-}\left(X^d_{t_-}\right)$$

$$\triangleq \sum_{\left\{X^d_t \quad X^d_{t_-}\right\} \subseteq \left\{X_t \quad X_{t_-}\right\}} f^{\mathrm{ppp}}_{t\,|\,t_-}\left(X_t \setminus X^d_t,\, X_{t_-} \setminus X^d_{t_-}\right) \sum_{\theta\in\Theta} w^\theta_{t_-}\, f^\theta_{t\,|\,t_-}\left(X^d_t,\, X^d_{t_-}\right). \tag{7.48}$$

Here we see that this is the set convolution over the joint multi target state over two consecutive time steps. At the same time the prediction is in some sense unaware of the mixture and predicts it for each component independently. $f^s(\cdot\,|\,\cdot)$, with s denoting surviving, is introduced as the multi target transition function without birth, or $\eta = 0$ if you like. Thus the PPP birth process is just included in the prediction of the undetected targets, as is appropriate.

In fact, we can do the same exercise for the MB as well since it also is the union of independent components under the hypothesis. One does the exact same thing as above, and then only needs to worry about a single BPP prediction. This simplifies to

$$f^{i,\theta^i}_{t\,|\,t_-}(X_t, X_{t_-}) =$$
$$\begin{cases} \left(1 - r^{i,\theta^i}_{t_-\,|\,t_-}\right), & X_{t_-} = X_t = \emptyset, \\ (1 - \mathrm{P}_s(x_{t_-}))\, r^{i,\theta^i}_{t_-\,|\,t_-} f^{i,\theta^i}_{t_-\,|\,t_-}(x_{t_-}), & X_{t_-} = \{x_{t_-}\}, X_t = \emptyset \\ \mathrm{P}_s(x_{t_-})\, f(x_t \,|\, x_{t_-})\, r^{i,\theta^i}_{t_-\,|\,t_-} f^{i,\theta^i}_{t_-\,|\,t_-}(x_{t_-}), & X_{t_-} = \{x_{t_-}\}, X_t = \{x_t\} \\ 0, & \text{otherwise.} \end{cases} \tag{7.49}$$

Marginalizing out X_{t_-} using the set integral can be done by first integrating out the state when $X_{t_-} = \{x_{t_-}\}$ and then summing over the cardinality. There are only two parts in this sum, due to the cardinality distribution of the BPP giving zero for any more than one target. To do this we will use the predicted existence probability which is the marginal of the joint existence, survival and state distribution

$$r^{i,\theta^i}_{t\,|\,t_-} = \int_{x\in\mathcal{X}} \mathrm{P}_s(x)\, f^{i,\theta^i}_{t_-\,|\,t_-}(x)\, r^{i,\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x \tag{7.50}$$

and the existence conditioned predicted state distribution, given by Bayes theorem and the joint over the predicted existence and state distribution

$$f^{i,\theta^i}_{t\,|\,t_-}(x_t) = \frac{\int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\,\mathrm{P}_s(x)\,f^{i,\theta^i}_{t_-\,|\,t_-}(x)\,r^{i,\theta^i}_{t_-\,|\,t_-}\,\mathrm{d}x}{r^{i,\theta^i}_{t\,|\,t_-}}$$
$$= \frac{\int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\,\mathrm{P}_s(x)\,f^{i,\theta^i}_{t_-\,|\,t_-}(x)\,\mathrm{d}x}{\int\limits_{x\in\mathcal{X}} \mathrm{P}_s(x)\,f^{i,\theta^i}_{t_-\,|\,t_-}(x)\,\mathrm{d}x}. \tag{7.51}$$

With this, we can identify the two first cases in (7.49) are the only cases that can give $X_t = \emptyset$ with nonzero likelihood and have the predicted BPP for the empty set as

$$f^{i,\theta^i}_{t\,|\,t_-}(X_t = \emptyset) = \int f^{i,\theta^i}_{t\,|\,t_-}(\emptyset, X_{t_-})\ \delta X_{t_-},$$

inserting the two first cases of (7.49);

$$= \left(1 - r^{i,\theta^i}_{t_-\,|\,t_-}\right) + \int\limits_{x\in\mathcal{X}} (1 - \mathrm{P}_s(x))\,r^{i,\theta^i}_{t_-\,|\,t_-} f^{i,\theta^i}_{t_-\,|\,t_-}(x)\ \mathrm{d}x,$$

distributing the PDF, noting that it integrates to unity and inserting (7.50);

$$= 1 - r^{i,\theta^i}_{t_-\,|\,t_-} + r^{i,\theta^i}_{t_-\,|\,t_-} - r^{i,\theta^i}_{t\,|\,t_-}$$
$$= 1 - r^{i,\theta^i}_{t\,|\,t_-}. \tag{7.52}$$

Identifying case 3 in (7.49) as the only case that give nonzero likelihood for $X_t = \{x_t\}$, we get

$$f^{i,\theta^i}_{t\,|\,t_-}(X_t = x_t) = \int f^{i,\theta^i}_{t\,|\,t_-}(\{x_t\}, X_{t_-})\ \delta X_{t_-},$$

inserting for case 3 in (7.49)

$$= \int\limits_{x\in\mathcal{X}} \mathrm{P}_s(x)\,f(x_t\,|\,x)\,r^{i,\theta^i}_{t_-\,|\,t_-} f^{i,\theta^i}_{t_-\,|\,t_-}(x)\ \mathrm{d}x,$$

inserting (7.50) and (7.51)

$$= r^{i,\theta^i}_{t\,|\,t_-} f^{i,\theta^i}_{t\,|\,t_-}\left(x^t\right). \tag{7.53}$$

Since all other cases of X_t give zero likelihood, we are done. This is clearly a new BPP, and we have that taking the union of all the predicted BPPs for detected and PPP for undetected targets we get back the same form as we started with, also with the same number of global hypotheses with the same weights. Along with (7.12) this concludes a derivation, or a proof if you like, of theorem 1 by Williams [2015b], which is the prediction step of the PMBM filter. Note that we did this first through deriving component independence in the prediction, such that the multi target state can be seen as consisting of a union of independent sets, and then deriving the prediction for the separated components/sets. As such it is seen that multi target prediction step attains the identities of the sets in the underlying union and hence track continuity of the BPP in the sense that the predicted BPP would be the same target as it was on the last time step if it actually exists.

Hybrid state

If we were to consider a hybrid state, there is nothing in the preceding derivations that hinders that to be the case. To be more explicit, discrete states can be treated just in the same way as continuous states, only that one have to do summations instead of integrals. As we did in the section on hybrid state spaces we are going to additionally condition the continuous states on the discrete states while the discrete states are conditioned on the same variables as the continuous states was for the non hybrid case.

Writing out the state distribution including discrete states in the above, we get the existence probability conditioned on the global hypothesis, θ^i, to be

$$r^{i,\theta^i}_{t\,|\,t_-} = \sum_{l_{t_-}\in\mathcal{L}} \int\limits_{x_{t_-}\in\mathcal{X}} \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, r^{i,\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}, \tag{7.54}$$

the existence conditioned discrete-state probability

$$\begin{aligned} \mu^{i,l^i_t,\theta^i}_{t\,|\,t_-} &= \frac{\sum\limits_{l_{t_-}\in\mathcal{L}} \int\limits_{x_{t_-}\in\mathcal{X}} \pi_{l^i_{t_-}}^{l^i_t}(x_{t_-})\, \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, r^{i,\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}}{r^{i,\theta^i}_{t\,|\,t_-}} \\ &= \frac{\sum\limits_{l_{t_-}\in\mathcal{L}} \int\limits_{x_{t_-}\in\mathcal{X}} \pi_{l^i_{t_-}}^{l^i_t}(x_{t_-})\, \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}}{\sum\limits_{l_{t_-}\in\mathcal{L}} \int\limits_{x_{t_-}\in\mathcal{X}} \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}}, \end{aligned} \tag{7.55}$$

the existence and predicted discrete-state conditioned mixing distribution

$$\begin{aligned} f^{i,\theta^i}_{t_-\,|\,t_-}\big(x_{t_-}\,|\,l^i_t\big) &= \frac{\sum_{l_{t_-}\in\mathcal{L}} \pi_{l^i_{t_-}}^{l^i_t}(x_{t_-})\, \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, r^{i,\theta^i}_{t_-\,|\,t_-}}{\mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, r^{i,a^i_t}_{t\,|\,t_-}} \\ &= \frac{\sum_{l_{t_-}\in\mathcal{L}} \pi_{l^i_{t_-}}^{l^i_t}(x_{t_-})\, \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}}{\sum_{l_{t_-}\in\mathcal{L}} \int\limits_{x_{t_-}\in\mathcal{X}} \mathrm{P}_s^{l^i_{t_-}}(x_{t_-})\, f^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}(x_{t_-})\, \mu^{i,l^i_{t_-},\theta^i}_{t_-\,|\,t_-}\, \mathrm{d}x_{t_-}}, \end{aligned} \tag{7.56}$$

and at last the existence and predicted discrete-state conditioned predicted countinuous-state distribution

$$f^{i,l^i_t,\theta^i}_{t\,|\,t_-}(x_t) = \int\limits_{x_{t_-}\in\mathcal{X}} f^{l^i_t}(x_t\,|\,x_{t_-})\, f^{i,\theta^i}_{t_-\,|\,t_-}\big(x_{t_-}\,|\,l^i_t\big)\, \mathrm{d}x_{t_-}. \tag{7.57}$$

The discrete state conditioned survival probability, $\mathrm{P}_s^l(x)$, has been introduced to include the possibility for modeling target survival as discrete state dependent. These equations are completely analogous to the equations derived in the section on hybrid state in the Bayesian estimation chapter. The exception is the inclusion of existence probability and survival probability. The existence probability is seen to disappear in state updates as are only target related so these equations are only differing in that they have survival probability. The survival probability is also in some sorts a discrete state transition and can therefore be combined with the discrete transition function in all the state update equations. If the survival and transition probability is modeled to be consisting of a uniform component and weighted Gaussian components they can be combined into a single transition by combining Gaussian components.

The author again suggests that hypothesis reduction within the single target PDF should be done in the mixing distribution, right before the last step above, or as a part of combining the last two steps if the distributions used allows for simplifications through doing that.

This is in fact a generalization of the prediction step of the IMM-joint integrated probabilistic data association (JIPDA) of Musicki and Suvorova [2008], in the sense that if we let $\mathrm{P}_s^l(x)$ be constant, $\pi_{l_{t_-}}^{l_t}(x) = \pi_{l_{t_-}}^{l_t}$, and $f(x_t \,|\, x_{t_-})$ and $f_{t_- \,|\, t_-}^{i,\theta_t^{\pi_i}}(x_{t_-} \,|\, l_t)$ be approximated by Gaussians, we get the exact same prdiction algorithm. If $\pi_{l_{t_-}}^{l_t}(x) = \pi_{l_{t_-}}^{l_t}$, making it equivalent to the first line of [(9)][1], (7.54) becomes [(10)][1] with $\Pi_{21} = 0$[2] and the existence probability ψ in their notation equivalent to r in ours. Similarly, with the given assumptions, (7.55) and (7.56) are equivalent to [(21)-(24)][1] and (7.57) equivalent to [(25)][1] under the Gaussian assumption.

Labeling

An important thing to notice is that the predicted distribution consists of the same BPPs in the sense that they keep their identity, while they change the existence probability and existence conditioned state distribution. Using component labeling will therefore keep the same labeling as before the prediction step, and this can therefore also be seen as a derivation of the LMB prediction step as well. The MB is in some sense already labeled, in that its individual BPP models an independent individual target.

7.2.2 Detection of all targets

Similar to the joint over time steps in the prediction, the joint over measurements and targets can be split up since the components are all independent. However, since we are conditioning on the measurements in the end, we cannot simply take the union of all the individual components updated. We will see this later. It will anyway simplify things later if we take a look at what happens to a single BPP without any clutter in an update.

[1] [·] Referring to content by Musicki and Suvorova [2008]

[2] $\Pi_{21} = 0$ signifies no reentry here which was allowed by Musicki and Suvorova [2008], and is modeled by the PPP birth process here giving reentering targets “a new identity” or by letting the scene go beyond our field of view to not treat those targets as dead, but simply having no detection probability.

Update of a BPP

The joint over the i'th BPP component and its corresponding BPP measurement conditioned on the global hypothesis θ can be written as

$$p(Z_t, X_t^i) = \begin{cases} 1 - r_{t\,|\,t_-}^{i,\theta^i}, & X_t^i = Z_t = \emptyset, \\ \left(1 - \mathrm{P}_d(x_t^i)\right) f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\, r_{t\,|\,t_-}^{i,\theta^i}, & X_t^i = \{x_t^i\}, Z_t = \emptyset, \\ h(z \,|\, x_t^i)\, \mathrm{P}_d(x_t^i)\, f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\, r_{t\,|\,t_-}^{i,\theta^i}, & X_t^i = \{x_t^i\}, Z_t = \{z\}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.58}$$

There are two cases that correspond to no detection; either the target does not exist, or it was not detected. The marginal for $Z_t = \emptyset$ is given by the set integral, and is the sum of the first two cases where the state is marginalized out in the second case;

$$\begin{aligned} h(Z_t = \emptyset) &= \int p(\emptyset, X_t^i)\ \delta X_t &(7.59)\\ &= 1 - r_{t\,|\,t_-}^{i,\theta^i} + \int\limits_{x_t^i \in \mathcal{X}} \left(1 - \mathrm{P}_d(x_t^i)\right) f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\, r_{t\,|\,t_-}^{i,\theta^i}\, \mathrm{d}x_t^i &(7.60)\\ &= 1 - r_{t\,|\,t_-}^{i,\theta^i} + r_{t\,|\,t_-}^{i,\theta^i} - \int\limits_{x_t^i \in \mathcal{X}} \mathrm{P}_d(x_t^i)\, f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\, r_{t\,|\,t_-}^{i,\theta^i}\, \mathrm{d}x_t^i &(7.61)\\ &= 1 - r_{t\,|\,t_-}^{i,\theta^i} \int\limits_{x_t^i \in \mathcal{X}} \mathrm{P}_d(x_t^i)\, f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\ \mathrm{d}x_t^i. &(7.62) \end{aligned}$$

Case three is the only case corresponding to detection, so the marginal is just the integral with respect to x_t^i

$$\begin{aligned} h(Z_t = \{z\}) &= \int p(\{z\}, X_t^i)\ \delta X_t &(7.63)\\ &= \int\limits_{x_t^i \in \mathcal{X}} h(z \,|\, x_t^i)\, \mathrm{P}_d(x_t^i)\, f_{t\,|\,t_-}^{i,\theta^i}(x_t^i)\, r_{t\,|\,t_-}^{i,\theta^i}\, \mathrm{d}x_t^i. &(7.64) \end{aligned}$$

We now condition the target BPP on the corresponding measurements

$$f_{t\,|\,t_-}^{i,\theta^i}(X_t|Z_t) = \begin{cases} \dfrac{1 - r_{t\,|\,t_-}^{i,\theta^i}}{1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i}, & X_t^i = Z_t = \emptyset, \\ \dfrac{\big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, r_{t\,|\,t_-}^{i,\theta^i}}{1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i} & X_t^i = \big\{x_t^i\big\}, Z_t = \emptyset, \\ \dfrac{h\big(z \mid x_t^i\big)\, \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, r_{t\,|\,t_-}^{i,\theta^i}}{\displaystyle\int_{x_t^i \in \mathcal{X}} h\big(z \mid x_t^i\big)\, \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, r_{t\,|\,t_-}^{i,\theta^i}\, \mathrm{d}x_t^i}, & X_t^i = \big\{x_t^i\big\}, Z_t = \{z\}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.65}$$

We have to work a bit more on case 2 to get it into a form with a state distribution and a probability of existence;

$$\frac{\big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, r_{t\,|\,t_-}^{i,\theta^i}}{1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i} =$$
$$\frac{r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i}{1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i} \frac{\big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)}{\displaystyle\int_{x_t^i \in \mathcal{X}} \big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i}. \tag{7.66}$$

In the above we can identify the updated existence probability when no detection occurred as (also seen by taking 1 minus case one)

$$r_{t\,|\,t}^{i,\bar{\theta}^i} = \frac{r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i}{1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i} \tag{7.67a}$$

and the updated existence conditioned state distribution when no detection occured as

$$f_{t\,|\,t}^{i,\bar{\theta}^i}\big(x_t^i\big) = \frac{\big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)}{\displaystyle\int_{x_t^i \in \mathcal{X}} \big(1 - \mathrm{P}_d\big(x_t^i\big)\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i} \tag{7.67b}$$

and in the case of detections, we get the existence probability and existence conditioned state distribution as

$$r^{i,\hat{\theta}^i}_{t\,|\,t} = 1, \tag{7.67c}$$

and

$$f^{i,\hat{\theta}^i}_{t\,|\,t}\left(x^i_t\right) = \frac{h\!\left(z \mid x^i_t\right) \mathrm{P}_d\!\left(x^i_t\right) f^{i,\theta^i}_{t\,|\,t_-}\left(x^i_t\right) r^{i,\theta^i}_{t\,|\,t_-}}{\displaystyle\int\limits_{x^i_t \in \mathcal{X}} h\!\left(z \mid x^i_t\right) \mathrm{P}_d\!\left(x^i_t\right) f^{i,\theta^i}_{t\,|\,t_-}\left(x^i_t\right) r^{i,\theta^i}_{t\,|\,t_-}\, \mathrm{d}x^i_t}, \tag{7.67d}$$

respectively. With this we have that the updated distribution is also a BPP, although a degenerate one in the case of detection since probability of existence is one. The bar over θ^i is introduced temporarily to mean undetected, while the hat means detected hypothesis. It is encouraging to see that the existence probabilities sum to one for the two cases.

Update of the joint multi target

We will now proceed to see what having multiple targets and multiple measurements together will add in complexity over the single BPP case. The joint SDF over the multi target state and the measurements is given by the multiplication of the PMBM and the multi target measurement function ($N_{t\,|\,t_-}$ denotes the number of BPP components in the MBM)

$$p_{t\,|\,t_-}\left(X_t, Z_t\right) = h\!\left(Z_t \mid X_t\right) f^{\mathrm{pmbm}}_{t\,|\,t}\left(X_t\right) \tag{7.68}$$

$$= \mathrm{e}^{-\bar{\mu}_t} \sum_{\sigma \in \mathcal{Q}^{|Z_t|}_{|X_t|}} \prod_{\substack{j=1:\\ j \notin \sigma}}^{|Z_t|} \mu_t\left(z^j_t\right) \prod_{\substack{i=1:\\ \sigma_i = 0}}^{|X_t|} \left(1 - \mathrm{P}_d\!\left(x^i_t\right)\right) \prod_{\substack{i=1:\\ \sigma_i > 0}}^{|X_t|} h\!\left(z^{\sigma_i}_t \mid x^i_t\right) \mathrm{P}_d\!\left(x^i_t\right)$$

$$\sum_{\theta \in \Theta_{t_-}} w^{\theta}_{t\,|\,t_-}\, \mathrm{e}^{-\bar{\lambda}_{t\,|\,t_-}} \sum_{\pi \in \mathcal{Q}^{N_{t\,|\,t_-}}_{|X_t|}} \prod_{\substack{i=1:\\ \pi_i = 0}}^{|X_t|} \lambda_{t\,|\,t_-}\left(x^i_t\right) \prod_{\substack{i=1:\\ i \notin \pi}}^{N_{t\,|\,t_-}} \left(1 - r^{i,\theta^i}_{t\,|\,t_-}\right) \prod_{\substack{i=1:\\ \pi_i > 0}}^{|X_t|} r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-} f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-}\left(x^i_t\right). \tag{7.69}$$

Combining products over the same range;

$$
\begin{aligned}
&= \sum_{\theta\in\Theta_{t_-}} w^{\theta}_{t\,|\,t_-}\, \mathrm{e}^{-\bar{\mu}_t}\mathrm{e}^{-\bar{\lambda}_{t\,|\,t_-}} \sum_{\sigma\in\mathcal{Q}^{|Z_t|}_{|X_t|}} \sum_{\pi\in\mathcal{Q}^{N_{t\,|\,t_-}}_{|X_t|}} \prod_{\substack{j=1:\\ j\notin\sigma}}^{|Z_t|} \mu_t\left(z^j_t\right) \prod_{\substack{i=1:\\ i\notin\pi}}^{N_{t\,|\,t_-}} \left(1-r^{i,\theta^i}_{t\,|\,t_-}\right)\\
&\qquad \prod_{\substack{i=1:\\ \sigma_i=0\\ \pi_i=0}}^{|X_t|} \left(1-\mathrm{P}_d(x^i_t)\right)\lambda_{t\,|\,t_-}(x^i_t) \prod_{\substack{i=1:\\ \sigma_i>0\\ \pi_i=0}}^{|X_t|} h(z^{\sigma_i}_t \,|\, x^i_t)\,\mathrm{P}_d(x^i_t)\,\lambda_{t\,|\,t_-}(x^i_t)\\
&\prod_{\substack{i=1:\\ \sigma_i=0,\\ \pi_i>0}}^{|X_t|} \left(1-\mathrm{P}_d(x^i_t)\right) r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-} f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-}\left(x^i_t\right) \prod_{\substack{i=1:\\ \sigma_i>0,\\ \pi_i>0}}^{|X_t|} h(z^{\sigma_i}_t \,|\, x^i_t)\,\mathrm{P}_d(x^i_t)\, r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-} f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t_-}\left(x^i_t\right).
\end{aligned}
\tag{7.70}
$$

Here, the updated undetected targets given by (7.36) can be found as the first product on the second line, the scaled versions of the newly detected BPP components in (7.37) can be identified as the last product on the second line, and scaled versions of (7.67) as the product over the non-existent components on the first line as well as the two products over undetected but existent and detected components on third line. Since we see the BPPs for new tracks, we claim that the amount of components in the MB will increase by $|Z_t|$ such that the updated number of BPP components will be $N_{t\,|\,t} = N_{t\,|\,t_-} + |Z_t|$.

One can also see that σ equivalently can be seen to hypothesize to update the BPP components or the PPP component directly instead of specific target states. This is especially so, since a single target does not have a particular distribution in the RFS framework[1], and it is then rather more correctly the track/components in the multi target distribution that get updated under some hypothesis instead of the target states. To accommodate the more correct update hypothesis, we introduce the track to measurement association variable (or measurement pointer if you want) for already existing components (previously detected target hypotheses);

$$
a_t = \left\{ a^i_t \in [0:|Z_t|] \middle| a^i_t = j > 0 \implies a^{i'}_t \neq j, i \neq i', \forall\, i, i' \in \left[1 : N_{t\,|\,t_-}\right] \right\} \tag{7.71}
$$

where, in the same way as the target to measurement pointer, it has $a^i_t = j > 0$ for component i giving rise to measurement j, and $a^i_t = 0$ for not giving rise to any measurement.

Note that in the above, there are two $\{\sigma_{i'}, \pi_{i'}\}$ combined hypotheses that correspond to $a^i_t = 0$, namely the track hypothesis $\{i' \notin \pi\}$ and the target hypothesis $\{\pi_{i'} = i, \sigma_{i'} = 0\}$. All other $\{\sigma_{i'} = j, \pi_{i'} = i\}$ target hypotheses are equivalently described by by the track to measurement hypothesis a_t, along with an updated target to component hypothesis,

$$
\bar{\pi} \in \mathcal{R}^{N_{t\,|\,t}}_{|X_t|}(a_t, N_{t\,|\,t_-}) = \left\{ \pi \in \mathcal{Q}^{N_{t\,|\,t}}_{|X_t|} = \mathcal{Q}^{(N_{t\,|\,t_-}+|Z_t|)}_{|X_t|} \middle| j \in a_t \implies N_{t\,|\,t_-} + j \notin \pi \right\} \tag{7.72}
$$

[1]Targets are elements of a set, and as such have no specific description other than having a state described by the distribution over the set of targets

that incorporates the components of possible detections of never before detected targets when the measurement is not hypothesized to be associated with an already existing component.

With this we rewrite the joint multi target and measurement equation in terms of the previously derived update equations

$$p_{t\,|\,t\text{-}}\left(X_t, Z_t\right) = \tag{7.73}$$

$$\sum_{\theta\in\Theta_{t\text{-}}} w^{\theta}_{t\,|\,t\text{-}} \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t\text{-}}} \sum_{a_t\in\mathcal{Q}^{|Z_t|}_{|N_{t\,|\,t}|}} \sum_{\bar{\pi}\in\mathcal{R}^{N_{t\,|\,t}}_{|X_t|}(a_t,N_{t\,|\,t\text{-}})} \prod_{\substack{i=1:\\ \bar{\pi}_i=0}}^{|X_t|} \left(1-\mathrm{P}_d\left(x^i_t\right)\right)\lambda_{t\,|\,t\text{-}}\left(x^i_t\right)$$

$$\prod_{\substack{i=N_{t\,|\,t\text{-}}+1:\\ i\notin\bar{\pi},\\ i-N_{t\,|\,t\text{-}}\notin a_t}}^{Nt\,|\,t} \left(\mu_t\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)+\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)\right)\left(1-\frac{\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)}{\mu_t\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)+\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)}\right)$$

$$\prod_{\substack{i=N_{t\,|\,t\text{-}}+1:\\ i\in\bar{\pi}}}^{Nt\,|\,t} \left(\mu_t\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)+\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)\right)\frac{\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)}{\mu_t\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)+\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)}\frac{\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t, x^i_t\right)}{\Lambda\left(z^{i-N_{t\,|\,t\text{-}}}_t\right)}$$

$$\prod_{\substack{i=1:\\ a^i_t=0,\\ \bar{\pi}_i>0}}^{|X_t|} \left(1-r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\int_{x^i_t\in\mathcal{X}}\mathrm{P}_d\left(x^i_t\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t\right)\frac{r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\int_{x^i_t\in\mathcal{X}}\left(1-\mathrm{P}_d\left(x^i_t\right)\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t}{1-r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\int_{x^i_t\in\mathcal{X}}\mathrm{P}_d\left(x^i_t\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t}$$

$$\frac{\left(1-\mathrm{P}_d\left(x^i_t\right)\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)}{\int_{x^i_t\in\mathcal{X}}\left(1-\mathrm{P}_d\left(x^i_t\right)\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t}$$

$$\prod_{\substack{i=1:\\ a^i_t>0,\\ \bar{\pi}_i>0}}^{|X_t|} r^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\int_{x^i_t\in\mathcal{X}} h\left(z^{a^i_t}_t\,|\,x^i_t\right)\mathrm{P}_d\left(x^i_t\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t \frac{h\left(z^{a^i_t}_t\,|\,x^i_t\right)\mathrm{P}_d\left(x^i_t\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)}{\int_{x^i_t\in\mathcal{X}} h\left(z^{a^i_t}_t\,|\,x^i_t\right)\mathrm{P}_d\left(x^i_t\right) f^{\pi_i,\theta^{\pi_i}}_{t\,|\,t\text{-}}\left(x^i_t\right)\mathrm{d}x^i_t}. \tag{7.74}$$

It should be made clear that this is just a rewriting of the previous equation where it is just multiplied and divided by the same number and restructured. If one multiplies out the fractions and inserts the old permutation variables, one obtains the exact same equation as the previous.

We see in the two last products that there is only one hypothesis corresponding to the bar hypothesis in the single component update of (7.67) but several that correspond to the hat hypothesis. We want to include this into the global hypothesis structure Θ_t such that we can update Θ_{t_-} to Θ_t by incorporating a_t as well. We do this by letting $\theta_t^i = (\theta_{t_-}^i, a_t^i)$ for previously detected tracks that have nonzero existence probability, $\theta_t^i = (\theta_{t_-}^i, \emptyset)$ for tracks that have zero existence probability, and $\theta_t^i = j$ for new tracks when the measurements are available for it. The new set of global hypotheses then becomes

$$\Theta_t = \bigcup_{\substack{\theta_{t_-} \in \Theta_{t_-}, \\ a_t \in \mathcal{A}_t}} (\theta_t^i)_{i=1}^{N_t}, \tag{7.75}$$

with the single component hypotheses defined as

$$\theta_t^i = \begin{cases} (\theta_{t_-}^i, \emptyset), & i \in [1 : N_{t_-}], r_{t\,|\,t_-}^{i,\theta_{t_-}^i} = 0, \\ (\theta_{t_-}^i, a_t^i), & i \in [1 : N_{t_-}], r_{t\,|\,t_-}^{i,\theta_{t_-}^i} > 0, \\ \emptyset, & i \in [(N_{t_-} + 1) : N_t], i - N_{t_-} \in a_t, \\ i - N_{t_-}, & i \in [(N_{t_-} + 1) : N_t], i - N_{t_-} \notin a_t. \end{cases} \tag{7.76}$$

(7.77)

The two first lines are the new track hypotheses for existing tracks, while the two last lines are for new tracks where we include hypotheses for having zero existence probability from the beginning, if the measurement it belongs to is used by another component.

The BPP components are scaled in the above compared to what was derived earlier in (7.37) and (7.67). To accommodate this, we are going to introduce variables, w_t^{i,a_t^i} to represent this scaling. The scaling variables needed are seen to be

$$w_t^{i,(\theta_{t_-}^i, a_t^i)} = w_t^{i,\theta_t^i} = \begin{cases} 1 - r_{t\,|\,t_-}^{i,\theta^i} \displaystyle\int_{x_t^i \in \mathcal{X}} \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, \mathrm{d}x_t^i, & i \in [1 : N_{t\,|\,t_-}], a_t^i = 0\,, \\ \displaystyle\int_{x_t^i \in \mathcal{X}} h\Big(z_t^j \mid x_t^i\Big)\, \mathrm{P}_d\big(x_t^i\big)\, f_{t\,|\,t_-}^{i,\theta^i}\big(x_t^i\big)\, r_{t\,|\,t_-}^{i,\theta^i}\, \mathrm{d}x_t^i & i \in [1 : N_{t\,|\,t_-}], a_t^i = j > 0, \\ 1, & i = N_{t\,|\,t_-} + j \notin \bar{\pi}, \\ \mu_t\left(z_t^j\right) + \Lambda\Big(z_t^j\Big)\,, & i = N_{t\,|\,t_-} + j \in \bar{\pi}. \end{cases} \tag{7.78}$$

The first case in the above, corresponds to no detection of already established tracks, the second to detection of already established tracks, the fourth to new tracks and the third to this tracks being non-existent due to the measurement being associated with another component. Case three has a BPP with an existence probability of zero due to assumption A-8, and the existence conditioned state distribution for a zero existence probability component has no effect, so we therefore do not need to consider it. This component does not show up directly in the derivation and denotes a degenerate BPP, but it is included to make notation consistent and also for letting us keep the structure of the problem as will be seen in the upcoming result.

Writing the joint in terms of the scaling and (7.36), (7.37) and (7.67) with the new hypothesis structure simplifies it to [1]

$$p_{t\,|\,t\text{-}}(X_t, Z_t) = \sum_{\theta_{t\text{-}} \in \Theta_{t\text{-}}} w_{t\,|\,t\text{-}}^{\theta_{t\text{-}}} \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t\text{-}}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \sum_{\bar{\pi} \in \mathcal{Q}_{|X_t|}^{N_{t\,|\,t}}} \prod_{\substack{i=1:\\ \bar{\pi}_i = 0}}^{|X_t|} \lambda_{t\,|\,t}(x_t^i)$$
$$\prod_{\substack{i=1:\\ i \notin \bar{\pi}}}^{N_{t\,|\,t}} w_t^{i,(\theta_{t\text{-}}^i, a_t^i)} \left(1 - r_{t\,|\,t}^{i,\theta_t^i}\right) \prod_{\substack{i=1:\\ \bar{\pi}_i > 0}}^{|X_t|} w_t^{\bar{\pi}_i,(\theta_{t\text{-}}^{\bar{\pi}_i}, a_t^{\bar{\pi}_i})} r_{t\,|\,t}^{\bar{\pi}_i,\theta_t^{\bar{\pi}_i}} f_{t\,|\,t}^{\bar{\pi}_i,\theta_t^{\bar{\pi}_i}}(x_t^i)\,, \tag{7.79}$$

extracting the weights out of the products into a new product

$$= \sum_{\theta_{t\text{-}} \in \Theta_{t\text{-}}} w_{t\,|\,t\text{-}}^{\theta_{t\text{-}}} \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t\text{-}}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \sum_{\bar{\pi} \in \mathcal{Q}_{|X_t|}^{N_{t\,|\,t}}} \prod_{\substack{i=1:\\ \bar{\pi}_i = 0}}^{|X_t|} \lambda_{t\,|\,t}(x_t^i) \tag{7.80}$$
$$\prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i} \prod_{\substack{i=1:\\ i \notin \bar{\pi}}}^{N_{t\,|\,t}} \left(1 - r_{t\,|\,t}^{i,\theta_t^i}\right) \prod_{\substack{i=1:\\ \bar{\pi}_i > 0}}^{|X_t|} r_{t\,|\,t}^{\bar{\pi}_i,\theta_t^{\bar{\pi}_i}} f_{t\,|\,t}^{\bar{\pi}_i,\theta_t^{\bar{\pi}_i}}(x_t^i)\,.$$

The last line should be recognized as an non-normalized PMBM. Knowing that using Bayes' theorem on this is simply a scaling by a constant (dependent on Z_t) and that the distributions involved were normalized before, we have our desired result by normalizing the new weights.

The normalization constant is given by the likelihood of the measurements, which again is given by the marginalization with respect to X_t. Taking the set integral of the joint gives us

$$h_{t\,|\,t\text{-}}(Z_t) = \int p_{t\,|\,t\text{-}}(X_t, Z_t)\ \delta X_t \tag{7.81}$$

$$= \sum_{\theta_{t\text{-}} \in \Theta_{t\text{-}}} w_{t\,|\,t\text{-}}^{\theta_{t\text{-}}} \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t\text{-}}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \mathrm{e}^{\bar{\lambda}_{t\,|\,t}} \prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i} \tag{7.82}$$

$$= \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t\text{-}}} \mathrm{e}^{\bar{\lambda}_{t\,|\,t}} \sum_{\theta_{t\text{-}} \in \Theta_{t\text{-}}} w_{t\,|\,t\text{-}}^{\theta_{t\text{-}}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i}. \tag{7.83}$$

To get this we have simply used the fact that under a given hypothesis it is a scaled PMB where the MB are scaled by the weights and the normalization constant is missing for the PPP, The calculations of these sums have exponential complexity, as they are over all

[1] a_t^i does not exist and is arbitrary for $i > N_{t\text{-}}$ since it can only be updated by one measurement, and is only in the equation to simplify notation in calculation towards the final result

feasible associations, and hence not a tractable calculation. Nevertheless, we have that the the multi target state distribution conditioned on the last measurement set can with this be written as

$$f_{t\,|\,t}^{\mathrm{pmbm}}(X_t) = \frac{p_{t\,|\,t_-}(X_t, Z_t)}{h_{t\,|\,t_-}(Z_t)} \tag{7.84}$$

$$= \sum_{\theta_{t_-} \in \Theta_{t_-}} w_{t\,|\,t}^{\theta_{t_-}} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t}} \sum_{\pi \in \mathcal{Q}_{|X_t|}^{N_{t\,|\,t}}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \prod_{\substack{i=1:\\ \pi_i=0}}^{|X_t|} \lambda_{t\,|\,t}(x_t^i)$$

$$\frac{\prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i}}{\sum_{\theta_{t_-} \in \Theta_{t_-}} w_{t\,|\,t_-}^{\theta_{t_-}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i}} \prod_{\substack{i=1:\\ i \notin \pi}}^{N_{t\,|\,t}} \left(1 - r_{t\,|\,t}^{i,\theta^i}\right) \prod_{\substack{i=1:\\ \pi_i>0}}^{|X_t|} r_{t\,|\,t}^{\pi_i,\theta_t^{\pi_i}} f_{t\,|\,t}^{\pi_i,\theta_t^{\pi_i}}\left(x_t^i\right). \tag{7.85}$$

The (normalized) single component hypothesis weights can be combined to into one weight to form a global hypothesis probability for the association at this time step. It can also be included into the global hypothesis weight to make just one single hypothesis probability for each global hypothesis. That is, we can write the updated hypothesis weight/probability as

$$w_{t\,|\,t}^{\theta_t} = w_{t\,|\,t}^{\theta_{t_-}} \frac{\prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i}}{\sum_{\theta_{t_-} \in \Theta_{t_-}} w_{t\,|\,t_-}^{\theta_{t_-}} \sum_{a_t \in \mathcal{Q}_{|N_{t\,|\,t}|}^{|Z_t|}} \prod_{i=1}^{N_{t\,|\,t}} w_t^{i,\theta_t^i}} \tag{7.86}$$

Inserting this into the above gives us the final result

$$f_{t\,|\,t}^{\mathrm{pmbm}}(X_t) = \sum_{\theta \in \Theta_t} w_{t\,|\,t}^{\theta} \mathrm{e}^{-\bar{\mu}_t} \mathrm{e}^{-\bar{\lambda}_{t\,|\,t}} \sum_{\pi \in \mathcal{Q}_{|X_t|}^{N_{t\,|\,t}}} \prod_{\substack{i=1:\\ \pi_i=0}}^{|X_t|} \lambda_{t\,|\,t}(x_t^i) \tag{7.87}$$

$$\prod_{\substack{i=1:\\ i \notin \pi}}^{N_{t\,|\,t}} \left(1 - r_{t\,|\,t}^{i,\theta^i}\right) \prod_{\substack{i=1:\\ \pi_i>0}}^{|X_t|} r_{t\,|\,t}^{\pi_i,\theta_t^{\pi_i}} f_{t\,|\,t}^{\pi_i,\theta_t^{\pi_i}}\left(x_t^i\right).$$

With this we have derived the normalized version of theorem 2 by Williams [2015b] without going through the more abstract derivation using probability generating functionals as he did, and which is also a slightly different approach to the recently direct derivation by Garcia-Fernandez et al. [2018]. An implementation can use a non normalized distribution since the weights have the appropriate relative magnitude, and hence avoid calculating the normalizing constant.

Hybrid state update

The hybrid states are independent across targets and can be conditioned on the global hypothesis in the update so that one do not need to consider the discrete states, or the continuous states, in the measurement association across targets. If the state is indeed hybrid, the predicted state PDF is written as

$$f_{t\,|\,t_-}^{i,\theta_{t_-}^i}(x,l) = f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x)\,\mu_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}, \tag{7.88}$$

and therefore (7.67) for the full joint hybrid state and existence update for detections becomes

$$\begin{aligned}
f_{t\,|\,t}^{i,\theta_t^i}(x,l)\,r_{t\,|\,t}^{i,\theta_t^i} &= f_{t\,|\,t}^{i,l,\theta_t^i}(x)\,\mu_{t\,|\,t}^{i,l,\theta_t^i} r_{t\,|\,t}^{i,\theta_t^i} && (7.89)\\
&= \frac{h^l\Big(z_t^{a_t^i}\,|\,x\Big)\,\mathrm{P}_d^l(x)\,f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x)\,\mu_{t\,|\,t_-}^{i,l,\theta_{t_-}^i} r_{t\,|\,t_-}^{i,\theta_{t_-}^i}}{\sum\limits_{l'\in\mathcal{L}}\int\limits_{x'\in\mathcal{X}} h^{l'}\Big(z_t^{a_t^i}\,|\,x'\Big)\,\mathrm{P}_d^{l'}(x')\,f_{t\,|\,t_-}^{i,l',\theta_{t_-}^i}(x')\,\mu_{t\,|\,t_-}^{i,l',\theta_{t_-}^i} r_{t\,|\,t_-}^{i,\theta^i}\,\mathrm{d}x'} && (7.90)\\
&= \frac{h^l\Big(z_t^{a_t^i}\,|\,x\Big)\,\mathrm{P}_d^l(x)\,f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x)}{\int\limits_{x'\in\mathcal{X}} h^l\Big(z_t^{a_t^i}\,|\,x'\Big)\,\mathrm{P}_d^l(x')\,f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x')\,\mathrm{d}x'} && (7.91)\\
&\quad \frac{\mu_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}\int\limits_{x'\in\mathcal{X}} h^l\Big(z_t^{a_t^i}\,|\,x'\Big)\,\mathrm{P}_d^l(x')\,f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x')\,\mathrm{d}x'}{\sum\limits_{l'\in\mathcal{L}}\mu_{t\,|\,t_-}^{i,l',\theta_{t_-}^i}\int\limits_{x'\in\mathcal{X}} h^{l'}\Big(z_t^{a_t^i}\,|\,x'\Big)\,\mathrm{P}_d^{l'}(x')\,f_{t\,|\,t_-}^{i,l',\theta_{t_-}^i}(x')\,\mathrm{d}x'}\cdot 1,
\end{aligned}$$

and for undetected target it becomes

$$f_{t\,|\,t}^{i,\theta_t^i}(x,l)\,r_{t\,|\,t}^{i,\theta_t^i} = f_{t\,|\,t}^{i,l,\theta_t^i}(x)\,\mu_{t\,|\,t}^{i,l,\theta_t^i} r_{t\,|\,t}^{i,\theta_t^i} \tag{7.92}$$

$$= \frac{\left(1-\mathrm{P}_d^l(x)\right) f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x)\,\mu_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i} r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}}{1-r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}\sum\limits_{l'\in\mathcal{L}}\mu_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}}\mathrm{P}_d^{l'}(x')\,f_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'} \tag{7.93}$$

$$= \frac{\left(1-\mathrm{P}_d^l(x)\right) f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x)}{\int\limits_{x'\in\mathcal{X}}\left(1-\mathrm{P}_d^l(x')\right) f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'} \tag{7.94}$$

$$\frac{\mu_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}}\left(1-\mathrm{P}_d^l(x')\right) f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'}{\sum\limits_{l'\in\mathcal{L}}\mu_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}}\left(1-\mathrm{P}_d^{l'}(x')\right) f_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'}$$

$$\frac{r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}\sum\limits_{l'\in\mathcal{L}}\mu_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}}\left(1-\mathrm{P}_d^{l'}(x')\right) f_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'}{1-r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}\sum\limits_{l'\in\mathcal{L}}\mu_{t\,|\,t\text{-}}^{i,l',\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}}\mathrm{P}_d^{l'}(x')\,f_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'}$$

under the new global hypothesis θ_t. It should be noted, that it is just to multiply out the fractions to get the original fraction. In the above we can now identify the update equation for the existence probability, in case of a component detection, as

$$r_{t\,|\,t}^{i,\theta_t^i} = 1, \tag{7.95}$$

the discrete-state probability conditioned on a global hypothesis and existence, as

$$\mu_{t\,|\,t}^{i,l_t^i,\theta_t^i} = \frac{\mu_{t\,|\,t\text{-}}^{i,l_t^i,\theta_{t\text{-}}^i}\int\limits_{x\in\mathcal{X}} h^{l_t^i}\!\left(z_t^{a_t^i}\,|\,x\right)\mathrm{P}_d^{l_t^i}(x)\,f_{t\,|\,t\text{-}}^{i,l_t^i,\theta_{t\text{-}}^i}(x)\,\mathrm{d}x}{\sum\limits_{l\in\mathcal{L}}\mu_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}\int\limits_{x'\in\mathcal{X}} h^{l}\!\left(z_t^{a_t^i}\,|\,x'\right)\mathrm{P}_d^{l}(x')\,f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x')\,\mathrm{d}x'}, \tag{7.96}$$

and, by using this, the continuous-state distribution conditioned on a global hypothesis, existence and the discrete-state, as

$$f_{t\,|\,t}^{i,l_t^i,\theta_t^i}(x) = \frac{\mu_{t\,|\,t\text{-}}^{i,l_t^i,\theta_{t\text{-}}^i}\,h^{l_t^i}\!\left(z_t^{a_t^i}\,|\,x\right)\mathrm{P}_d^{l_t^i}(x)\,f_{t\,|\,t\text{-}}^{i,l_t^i,\theta_{t\text{-}}^i}(x)}{\mu_{t\,|\,t}^{i,l_t^i,\theta_t^i}}. \tag{7.97}$$

For the undetected component hypothesis, we get the update for the existence probability to be

$$r_{t\,|\,t}^{i,\theta_t^i} = \frac{r_{t\,|\,t_-}^{i,\theta_{t_-}^i} \sum_{l'\in\mathcal{L}} \mu_{t\,|\,t_-}^{i,l',\theta_{t_-}^i} \int\limits_{x'\in\mathcal{X}} \left(1 - \mathrm{P}_d^{l'}(x')\right) f_{t\,|\,t_-}^{i,l',\theta_{t_-}^i}(x')\,\mathrm{d}x'}{1 - r_{t\,|\,t_-}^{i,\theta_{t_-}^i} \sum_{l'\in\mathcal{L}} \mu_{t\,|\,t_-}^{i,l',\theta_{t_-}^i} \int\limits_{x'\in\mathcal{X}} \mathrm{P}_d^{l'}(x')\, f_{t\,|\,t_-}^{i,\theta_{t_-}^i}(x')\,\mathrm{d}x'}, \tag{7.98}$$

the discrete-state probability conditioned on a global hypothesis and existence, as

$$\mu_{t\,|\,t}^{i,l_t^i,\theta_t^i} = \frac{\mu_{t\,|\,t_-}^{i,l,\theta_{t_-}^i} \int\limits_{x'\in\mathcal{X}} \left(1 - \mathrm{P}_d^{l}(x')\right) f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x')\,\mathrm{d}x'}{\sum_{l'\in\mathcal{L}} \mu_{t\,|\,t_-}^{i,l',\theta_{t_-}^i} \int\limits_{x'\in\mathcal{X}} \left(1 - \mathrm{P}_d^{l'}(x')\right) f_{t\,|\,t_-}^{i,l',\theta_{t_-}^i}(x')\,\mathrm{d}x'} \tag{7.99}$$

and, by using this, the continuous-state distribution conditioned on a global hypothesis, existence and the discrete-state, as

$$f_{t\,|\,t}^{i,l_t^i,\theta_t^i}(x) = \frac{\left(1 - \mathrm{P}_d^{l}(x)\right) f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x)}{\int\limits_{x'\in\mathcal{X}} \left(1 - \mathrm{P}_d^{l}(x')\right) f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x')\,\mathrm{d}x'} \tag{7.100}$$

In the above we have used the notion that there is no special difference in how to handle discrete states in contrast to continuous states in theory, and then just used (7.67) with discrete probabilities as well. We then proceeded by restructuring to find a structure that is more practical in the hybrid case.

The measurement hypothesis weights were a marginalization over the continuous states, but as there is in principle not any difference in the discrete states, the integrals simply become sums in the case of discrete states. This lets us write the weights for detection hypotheses as

$$\begin{aligned} w_t^{i,\theta_t^i} &= r_{t\,|\,t_-}^{i,\theta_{t_-}^i} \sum_{l\in\mathcal{L}} \mu_{t\,|\,t_-}^{i,l,\theta_{t_-}^i} \int\limits_{x'\in\mathcal{X}} h^l\!\left(z_t^{a_t^i} \,|\, x'\right) \mathrm{P}_d^{l}(x')\, f_{t\,|\,t_-}^{i,l,\theta_{t_-}^i}(x')\,\mathrm{d}x' && (7.101)\\ &\triangleq \sum_{l\in\mathcal{L}} w_t^{i,l,\theta_t^i}, && (7.102) \end{aligned}$$

and undetected hypotheses as

$$w_t^{i,\theta_t^i} = 1 - r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i} \sum_{l \in \mathcal{L}} \mu_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i} \int_{x' \in \mathcal{X}} \mathrm{P}_d^l(x')\, f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x')\, \mathrm{d}x' \tag{7.103}$$

$$= \sum_{l \in \mathcal{L}} \mu_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i} \left(1 - r_{t\,|\,t\text{-}}^{i,\theta_{t\text{-}}^i} \int_{x' \in \mathcal{X}} \mathrm{P}_d^l(x')\, f_{t\,|\,t\text{-}}^{i,l,\theta_{t\text{-}}^i}(x')\, \mathrm{d}x' \right) \tag{7.104}$$

$$\triangleq \sum_{l \in \mathcal{L}} w_t^{i,l,\theta_t^i}, \tag{7.105}$$

where we have defined w_t^{i,l,a_t^i} as the summand in the two different cases. Using this we can also write the discrete-state probability update for detections as

$$\mu_{t\,|\,t}^{i,l_t^i,\theta_t^i} = \frac{w_t^{i,l,\theta_t^i}}{w_t^{i,\theta_t^i}}. \tag{7.106}$$

From this we see, in the same way as for the continuous states, that the weights do not carry any explicit information regarding the discrete states. However, the discrete-state distribution give higher weights for measurements that are likely according the discrete-state hypothesis in the exact same way as the continuous-states distribution does, and the weights therefore carry information from the full hybrid state needed to distinguish which target is likely to have given rise to which measurement. It is therefore not necessary to include the discrete states in the data association as some authors, such as Chen and Tugnait [2001] and myself [Tokle, 2017], have done previously. Other authors have understood this before, where for instance Musicki and Suvorova [2008, appendix B] derive the IMM-JIPDA using this type of conditioning, while it is a bit more unclear how de Feo et al. [1997] or Bar-Shalom et al. [1991], for instance, have treated this.

Labeling

A key thing to note here is that the update step under an association hypothesis essentially is a union of independent updates of the components. The measurement origin ambiguity and the need to handle data association then shows up as a result of the targets and measurements being sets, and hence unordered. Since the sets are unions of independent objects, the set convolution has to be performed in the joint SDF. These convolutions can be seen as the marginalization over association hypotheses for the measurements, and marginalization over component origin hypotheses for the targets.

As seen here, all BPP components have their origin in in a certain measurement associated with a potential target that has never been detected. This track, conditioned on a measurement association hypothesis, is then independently predicted and updated through time. As such, each BPP in the MB can be seen as a potential target track that started at some measurement,

$$i = j + \sum_{\tau'=0}^{\tau-1} |Z_{\tau'}|, \ \tau \in [0 : t], \ j \in [1 : |Z_t|],$$

where the predicted and updated track existence is given by $r^{i,\theta^i}_{t\,|\,t'}$. Williams [2015b] made the same type of note, but phrased it differently, and maybe hard for a reader to understand through the derivation using probability generating functionals.

The component labeling approach can still be applied as the tracks can still be seen as independent sets. This is the same procedure Meyer et al. [2018] proposes for labeling in the 2nd paragraph in section XI, whereas they use the equivalent but perhaps intuitively different labeling of the LMB. This is also the same notion of labeling used in the classical algorithms, such as JPDA and MHT [Bar-Shalom et al., 2011], and hence provides target identification in a “classical sense”. This is in contrast to for instance Vo et al. [2014], where targets are labeled already at possible arrival. Since the arrival of these targets are probabilistic, one needs to label non-existing and never detected targets as well in this setup. This may or may not be what one actually wants in a given application. The author would argue that deferring identification until one has target-specific information is preferable in most cases. In some applications one might have information regarding target identity already at possible arrival, and then labeling the birth process can of course be a good approach. Otherwise, the target identification done here is probably preferable, if one needs to assign target identities. The component labeling also lends itself to for instance labeling different birth PPP components such that one can do inference on the target origin and so on.

There are several reasons to stress this. One is that the BPP track existence is only known probabilistically. For instance if one wants to see if there is safe passage for an autonomous ship in a region, one needs to know if there are targets in that region. The targets can come from any of the BPP tracks or still be undetected, and hence one has to marginalize over the component origin of the targets to get the full likelihood for having a set of targets, i.e. no labels. On the other hand, if one is interested to see what the targets are doing through time, one is essentially interested in what is going on with specific BPP tracks. Another reason is that labeling can in some cases cause unnecessarily higher computational loads, as for instance pointed out by Garcia-Fernandez et al. [2018, sec. IV] regarding the generalized LMB [Vo et al., 2014] just being a different and less efficient parameterization of the MBM than the parameterization discussed here.

7.3 The multi target initial distribution

We have now discussed how the prediction step and the update step are affecting the distribution of targets in the RFS multi target state space. If it is a PMBM, both the prediction step and the update step will maintain a form such that we still have a PMBM afterwards. One question remains, how should we set the initial distribution? We hinted at this in the beginning of section 7.1.1 where we discussed the need to predict targets that we have no measurements of. What happens to this distribution if we do not measure anything, i.e. setting $\mathrm{P}_d(x) = 0$ and $\mu_t\,(z) = 0$ in the update equations?

With no measurements, there will be no update, and the distribution remains the same as the prior. However, we clearly have to update our beliefs about the target distribution in the prediction anyway, even if we are not measuring anything. In reality, there are probably targets moving around in the scene, even if we are not watching. Presumably our probabilistic model of births and deaths should also be valid for that time. Say we assume the targets have been going about their business without us watching, for a long time before we started taking measurements. Then we should probably consider that we should apply the prediction step for that time period as well.

If one has modeled this correctly, it will not diverge, as that would mean one expects to find an infinite amount of targets in the scene, which does not make much sense in real life. $\mathrm{P}_s(x)$ can be seen to act as a sink, and $\eta_t(x)$ as a source. If these are stationary, this will in fact converge to a stationary intensity. This intensity, or an approximation thereof, is the intensity of the initial distribution one should use when one does not have any other information available.

By this argument we may say that our initial distribution is the PPP predicted for an infinite time (or long enough to be a good approximation). So our initial distribution should be such that the PPP intensity is given by

$$\lambda_{0\,|\,0}(x) = \eta_t(x) + \int\limits_{x' \in \mathcal{X}} f(x \mid x')\,\mathrm{P}_s(x')\,\lambda_{0\,|\,0}(x')\,\mathrm{d}x', \tag{7.107}$$

which can be solved numerically in general, or explicitly for some special models, when the model is stationary. The initial multi target SDF is thus given by

$$f_{0\,|\,0}(X_0) = \mathrm{e}^{-\bar{\lambda}_{0\,|\,0}} \lambda_{0\,|\,0}^{X_0}. \tag{7.108}$$

Part III

Mutli Target Tracking: Relations and approximations

8 Relation of PMBM to other filters

8.1 The total target information distribution

We have seen that the PMBM filter can look quite complex and it can be a bit hard to grasp what is really going on. In this chapter we will present a new, possibly uglier but hopefully more transparent, distribution of the problem. This involves forming the joint over all variables at all time steps. At the same time, we will extend the birth process to consist of several components. However, we will begin with some interpretations.

Both a_τ and b_τ describe hypotheses between different measurements at different time steps. Specific a_t and b_t can be seen as describing tracks in measurement space that relate old and new measurements through time. Where $a^i_{\tau'} = j' > 0$ can be seen as measurement $i = (\tau, j)$, with $\tau < \tau'$, hypothesizing that measurement $i' = (\tau', j')$, originates from the same target as i. While $b^{j'}_{\tau'} = i > 0$ can be seen as measurement $i' = (\tau', j')$ hypothesizing that measurement $i = (\tau, j)$ originated from the same target as i'. That is, a_τ are hypotheses of old measurements being associated with the same targets as a new measurement, while b_τ are hypotheses of new measurements being associated with the same target as an old measurement. I.e. they point to data forward or backward in time respectively.

Furthermore, $\pi^k = i > 0$ indicates that the k'th target state-trajectory, $(x^k_{[t^k_0:t^k]}, l^k_{[t^k_0:t^k]})$, was first detected by measurement $i = (\tau, j)$, and therefore the measurement track that starts at measurement (τ, j). $\pi^k = 0$ denotes that the target has never been detected. This is exactly the labeling we have been discussing in previous chapters.

Target related variables of interest include;

- the initial time the target came into the scene, t^i_0,
- the time it was last believed to be in the scene, t^k,
- the birth component it came from, α^k,
- the hybrid state-trajectory, $(x^k_\tau, l^k_\tau) \in \mathcal{X} \times \mathcal{L}$, $\tau \in [t^k_0 : t^k]$,

- the measurement track it is associated with, π^k,
- all the measurements included in a track $i = (\tau, j)$, indexed using $\theta^i_{t^i} = (a^i_{t^i_0}, \ldots, a^i_{t^i})$,
- all measurements, Z_τ, and
- each of their measurement-track origins, b^j_τ.

To include measurements to track variables, as well as track to measurement variables, is an over parameterization and not needed. Nevertheless, we will include it here as it can be useful to show how some algorithms came about.

We will model that component $\alpha \in A$ gives rise to new targets according to a PPP with intensity $\eta^\alpha_\tau(x)$, where we will also use $\eta_\tau(x) = \sum_{\alpha \in A} \eta^\alpha_\tau(x)$. Other birth processes, such as a LMB [Vo et al., 2014] can also be considered, but they are not treated here.

To form the joint, we also need a consistency factor so that we ensure that existence hypotheses, and measurement association hypotheses, are compatible with each other. This factor is given by

$$\gamma(a_\tau, b_\tau, \pi, t_0^{[1:n]}, t^{[1:n]}) = \mathbf{1}\left[\begin{matrix} \pi^k \in \left[0{:}\sum_{\tau=1}^{t} |Z_\tau|\right] \\ a^i_\tau = j > 0 \iff b^j_\tau = i > 0, \\ b^j_\tau \in \pi \cup \{0\}, \\ \pi^k = i > 0{:}\, i = (t, j) \implies t^k_0 \leq t \leq t^k \\ a^{\pi^k}_t = 0\, \forall\, t > t^k, \pi^k > 0 \end{matrix}\right]. \tag{8.1}$$

The first constraint says that the target has to be related to either one of the measuement tracks, or none. The second, that the measurement track has to be consistent. Third, new measurements can only be related to tracks that come from targets, and implicitly also that π is the hypothesis over which measurements are related to "new" targets. The fourth line says that if target k is related to measurement i then it has to have existed at that time. The fifth line says that a target that is dead, can not have any measurements associated with it, and this is also an implicit limitation on b^j_τ through line 2.

With these constraints, we now state the full PDF of the joint of all these variables, noting that the measurements are now ordered in relation to each other, and hence are not necessarily permutation invariant anymore;

$$
\begin{aligned}
&p\left(\left\{\begin{bmatrix} x^k_{t^k_0:t^k} \\ l^k_{t^k_0:t^k} \\ \pi^k \\ \alpha^k \\ t^k \\ t^k_0 \end{bmatrix}\right\}_{k\in[1:|X|]}, \left\{\begin{bmatrix} z^j_\tau \\ b^j_\tau \\ a^{(\tau,j)}_{[\tau+1:t]} \end{bmatrix}_{j\in[1:|Z_\tau|]}\right\}_{\tau\in[1:t]}\right) = \\
&\prod_{(\alpha,\tau)\in A\times[0:t]} \mathrm{e}^{-\bar{\eta}^\alpha_\tau} \prod_{\substack{k=1:\\ t^k_0=\tau\\ \alpha^k=\alpha}}^{n} \eta^{\alpha^k}_{t^k_0}\left(x^k_{t^k_0}, l^k_{t^k_0}\right) \prod_{\substack{k=1:\\ t^k=\tau}}^{n} \left(1 - \mathrm{P}_s^{l^k_{t^k}}\left(x^k_{t^k}\right)\right) && \text{(births/deaths)} \\
&\times \prod_{k=1}^{n} \prod_{\tau=t^k_0+1}^{t^k} f^{l^k_\tau}\left(x^k_\tau \mid x^k_{\tau\text{-}}\right) \pi^{l^k_\tau}_{l^k_{\tau\text{-}}}\left(x^k_{\tau\text{-}}\right) \mathrm{P}_s^{l^k_{\tau\text{-}}}\left(x^k_{\tau\text{-}}\right) && \text{(predictions)} \\
&\times \prod_{\tau=0}^{t} \gamma(a_\tau, b_\tau, \pi, t^{1:n}_0, t^{1:n}) \left[\frac{\mathrm{e}^{-\bar{\mu}_\tau}}{|Z_\tau|!} \prod_{\substack{j=1:\\ b^j=0}}^{|Z_\tau|} \mu_\tau\left(z^j_\tau\right)\right] && \text{(false alarms)} \\
&\times \left[\prod_{\substack{k\in[1:|X|],\\ j\in[1:|Z_\tau|]:\\ a^{\pi^k}_\tau=j,\\ b^j_\tau=\pi^k,\\ t^k_0\le\tau\le t^k}} \mathrm{P}_d^{l^k_\tau}\left(x^k_\tau\right) h^{l^k_\tau}\left(z^j_\tau \mid x^k_\tau\right)\right] \left[\prod_{\substack{k=1:\\ a^{\pi^k}_\tau=0,\\ t^k_0\le\tau\le t^k}}^{|X|} \left(1 - \mathrm{P}_d^{l^k_\tau}\left(x^k_\tau\right)\right)\right]. && \text{(detections)}
\end{aligned}
\tag{8.2}
$$

The first line corresponds to the likelihoods of arriving targets and leaving targets (births and deaths). The very first part should be identified as a PPP at every time step, for every birth component. The last part of the the first line is the probabilities that targets left the scene between time step t^i and t^i_+ being in continuous and discrete state $x^i_{t^i}$ and $l^i_{t^i}$, respectively.

The second line, describes the likelihoods and probabilities that the targets first survive in state $(x^i_{\tau\text{-}}, l^i_{\tau\text{-}})$, the discrete state, conditioned on survival, updates to l^i_τ and then the continuous state updates to x^i_τ, conditioned on survival and the discrete state.

The two last lines describe the measurements conditioned on a given association hypothesis. The first of which, has the consistency factor γ so that only consistent association and existence events are given nonzero likelihood, and also the likelihood for the false measurements through its PPP. The final line, describes the probability of the given association between targets and measurements, while also giving the likelihood of the locations in measurement-space and state-space under this hypothesis.

The above distribution is what the author would call total target information, as this is the distribution over all targets that could have been in the scene between time 0 and t, and is over all variables that are related to the targets in some way. All algorithms that are based on assumption A and have hybrid state space, work with this distribution in one way or another. It may be seen as an extension of the PGM formulation of Meyer et al. [2018], although it is not obvious how to extend a PGM to incorporate the inference of a PPP, which in some way is needed for considering target trajectories that is undetected. Nevertheless, PGMs formulations lend themselves to the investigation of approximations by applying some form of KL-divergence and/or BP Koller and Friedman [2009], and can as such be a way of looking for new algorithms. Chow-Liu trees, local linearization, marginalization and key-frames might be schemes of interest when looking into this possibility.

8.1.1 Total target information in relation to RFS of trajectories and labeling

Similar thoughts as the total target information given above have been made by other authors such as Granström et al. [2018] and García-Fernández et al. [2016] considering RFSs of trajectories, where the set elements are target trajectories instead of target states. One of their goal is to show that the PMBM filter actually do create tracks such that the components can be thought of as being the same through time.

However, these authors seems to dislike the use of association variables and labels as part of their distribution. This may be appropriate of them, but the thoughts presented here does seem to show that we can consider the set convolution as a marginalization of a latent variable relating targets to tracks, and association variables essentially are between measurements at different time steps.

This is in no way claiming a rigorous proof of this, but the thought is intriguing, and in some way seems intuitive. Observed measurements does come from targets that are moving around, and they do in reality form tracks of measurements from these specific targets, even though we do not observe this directly. It does therefore seem that one should be able to extract these as latent hypothesis variables as done here and have probabilities of these being real tracks or not. In the above we have essentially split this into two separate hypotheses; one making tracks using a_t^i and b_t^j, and the other on which of these are true and which of these are false using the target to track/component label π.

It does also seem intuitive as all the components of the PMBM is predicted and updated independently if the association hypothesis is given, and can then be treated as the individual sets they in reality are. Also keeping the undetected targets unlabeled makes intuitive sense. If one were to consider a region in the nearest ocean right here and now without looking, one might be able to picture a likelihood of where one expects to find boats and how many, but starting to identify them with labels without observing them does not seem like an intuitive task.

8.1.2 Total target information in relation to PMBM

The previously derived, so called PMBM filter, is just a marginalization within this distribution, over; the earlier states, birth component (including time of birth), time of death, data association and first detection hypotheses. As such the PMBM filter does not consider, at the last time step, targets that might previously have been there, but are now gone. Rather, they are marginalized out. Note that it was slightly arbitrary that we considered track to measurement variables, and not measurement to track variables, to form association hypotheses in the PMBM. Since they are equivalent, we could just as well formulate the hypotheses in the PMBM using the measurement to target variables instead, and then make a global hypothesis structure on them.

8.2 Relationship to MHT

The MHT of Reid [1979] can similarly be seen to be a marginalization over some of the variables, while also adding some assumptions. Reid [1979] does not explicitly state how to handle the birth process, but uses a new-target intensity that can be interpreted as the undetected target intensity, and the original MHT therefore only deals with what we have called detected targets. He also does not consider target death, and considers the targets to follow a Gaussian linear assumption. The relationship between MHT and FISST/PMBM is more rigorously studied by Brekke and Chitre [2018, 2017].

The variables that are marginalized out in Reid's MHT are the previous states, birth component, birth time, and the track to measurement association variables, while the target death is treated to never occur, i.e. $\mathrm{P}_s(x) = 1$. The original hypotheses of MHT can thus be seen to make hypotheses over b_t^j and π, relating measurements to components/tracks and targets to tracks. The hypothesis tree described by Reid [1979] is exactly that of b_t^j "augmented" by π, such that $b_t^j = i > 0$ relates measurement j at time t to the track i, and $b_t^j = 0$ along with some $(j,t) = i' \in \pi$ or $(j,t) = i' \notin \pi$ describing the hypothesis of new track or clutter, respectively. But as such, it is under some assumptions equivalent to the labeled PMBM we described earlier, where the hypothesis structure is taken as the alternative equivalent hypothesis structure, of measurement to tracks, than we used in the derivations, being tracks to measurements.

A difference is of course that one usually seeks the most probable association hypothesis in MHT, instead of marginalizing it out as in PMBM. What one should seek, is of course, a matter of application. If one is interested in the association itself, one might want to take the most probable hypothesis. Whereas if one is interested in target location, marginalization might be better, as this takes all known information into consideration.

8.3 Relationship to TOMHT

We can also take a look at the track oriented multiple hypotheses tracker (TOMHT) of Kurien [1990] within this framework. As is stated in its name, the TOMHT formulates the global hypotheses using the track to measurement variables a_t^i. The association hypotheses are also exactly the same as in the PMBM we used in the derivation, but also including π to label the track and specify explicit existence hypotheses of the tracks. This relation is originally described by Williams [2015b]. Again, as with MHT, the difference is that one typically seeks the most probable hypothesis.

8.4 Relationship to TOMB/P

With component labels, the PMBM can be written as

$$f_{t\,|\,t'}^{\mathrm{Lpmbm}}\left(X_t^L\right) = f_{t\,|\,t'}^{\lambda}\left(X_t^{L(0)}\right) \sum_{\theta \in \Theta_t} \prod_{i=1}^{N_{t\,|\,t'}} w_{t\,|\,t'}^{i,\theta^i} f_{t\,|\,t'}^{i,\theta^i}\left(X_t^{L(i)}\right), \tag{8.3}$$

where $f_{t\,|\,t'}^{\lambda}(X)$ is the undetected target PPP at time t after the measurement scan at time t' is taken into consideration, and $f_{t\,|\,t'}^{i,\theta^i}(X)$ is the i'th Bernoulli component under hypothesis θ^i.

The only thing making the Bernoulli components coupled is the marginalization over the hypotheses. The global hypotheses make the distribution hard to handle as there is a combinatorial number of them, in the number of tracks and measurements, and are hence intractable. One way to increase tractability is to approximate the distribution as factored into independent components and then only have single track hypotheses to deal with. To do this, we can consider a minimization of the KL-divergence between the component labeled PMBM, and a component labeled PMB. The approximation SDF will be given by a component labeled PMB

$$f_{t\,|\,t'}^{\mathrm{TOMBP}}\left(X_t^L\right) = f_{t\,|\,t'}^{\lambda}\left(X_t^{L(0)}\right) \prod_{i=1}^{N_{t\,|\,t'}} f_{t\,|\,t'}^{i}\left(X_t^{L(i)}\right), \tag{8.4}$$

where $f_{t\,|\,t'}^{i}(X_t)$ is the i'th marginal Bernoulli component of the PMBM. This follows from the derivations of section 5.4, which are also valid for RFS, and states that the approximation of a distribution by a fully factored distribution is the product of its marginals. Since we have labeled the components they can be seen as different set variables, and hence we can marginalize with respect to them. Under the current labeling system, this is the marginal for each component. That is

$$f^{\text{TOMBP}}_{t\,|\,t'}\big(X^L_t\big) = \arg\min_q \mathrm{D}_{\mathrm{KL}}\big(f^{\mathrm{pmbm}}(\cdot)\;||\;q\big)\,, \tag{8.5}$$

$$\text{s.t.} \qquad q(X^L) = \mathrm{e}^{-\bar{\lambda}}\lambda^{X^{L(0)}}\prod_{i=0}^{N_{t\,|\,t'}} q^i(X^{L(i)}), \tag{8.6}$$

$$\int q^i(X)\delta X = 1.$$

In words; the component labeled SDF constrained to be fully factored into valid SDFs attains the minimum of the KL-divergence when its factors are the marginal for the labeled components. This also applies if the distribution consists of other types of components than PPPs and BPPs as well, but we will not go further into that.

The PPP is already a factor in the PMBM, so that will simply remain after the approximation. The BPPs are just coupled through the associations, as previously stated, so the marginalization is in fact only needed to be carried out over the association hypotheses (the existence probability and state PDF are both conditioned on the association hypothesis). This is therefore also equivalent to just approximating the association probabilities to be factored, as done recursively by Williams [2015b] to arrive at his track-oriented marginal MeMBer-Poisson (TOMB/P). As such, the TOMB/P can be seen as labeling the components and then approximating the labeled SDF as factored using the KL-divergence after each measurement scan. Williams [2015a, sec.II.B] also states this implicitly, but in relation to the JPDA, which is again an approximation of the TOMB/P into that of a single Gassian.

A similar analysis can be done with the measurement-oriented marginal MeMBer-Poisson (MOMB/P) of Williams [2015b], where the component labels point to the last measurement that it is related to, instead of its first measurement, and then proceeds to approximate this distribution by fully factorizing over the labeled components. This change can be seen as changing the association parameterization in the PMBM from using a^i_t to b^j_t.

Williams [2015a] has also been working on approximating the unlabeled distribution directly, but the set convolution (marginalization of component labels) is seen to be a largely complicating factor as several BPPs in the original distribution can be be a weighted part of several BPPs in the approximate. He interprets the labels as missing data and uses expectation maximization to find these weights. This gives an additional layer of complexity, that is alleviated by minimizing an upper bound instead.

The marginal BPP can be found by using the set integral, and is shown as

$$
\begin{aligned}
f^{i}_{t\,|\,t'}(\{x\}) &= \int_{(x,i)\in X^L} f^{\mathrm{TOMBP}}_{t\,|\,t'}\left(X^L_t\right)\delta X^L && (8.7)\\
&= \sum_{\theta\in\Theta_t} w^{\theta}_{t\,|\,t'} f^{i,\theta^i}_{t\,|\,t'}(\{x\}) && (8.8)\\
&= \sum_{\theta^i\in\Theta^i_t} f^{i,\theta^i}_{t\,|\,t'}(\{x\}) \left[w^{i,\theta^i}_{t\,|\,t'} \sum_{\substack{\theta\in\Theta_t:\\ \theta^i\in\theta}} \prod_{\substack{i'=1:\\ i'\neq i}}^{N_{t\,|\,t'}} w^{i,\theta^i}_{t\,|\,t'} \right], && (8.9)
\end{aligned}
$$

for existence, and similarly for non-existence. If this has been done recursively, and the association hypotheses are only summed over one time step, the bracketed term should be recognized as the marginal association probabilities of JPDA [Bar-Shalom et al., 2011, sec.6.2]. If hypotheses are summed over several time steps, it becomes the marginal track association hypothesis in the TOMHT.

We let

$$
\beta^{i,\theta^i}_{t\,|\,t'} = w^{i,\theta^i}_{t\,|\,t'} \sum_{\substack{\theta\in\Theta_t:\\ \theta^i\in\theta}} \prod_{\substack{i'=1:\\ i'\neq i}}^{N_{t\,|\,t'}} w^{i,\theta^i}_{t\,|\,t'} \tag{8.10}
$$

denote the marginal hypothesis probability of hypothesis θ^i for track i. Having calculated this we can see that the marginal BPP is given by

$$
\begin{aligned}
&f^{i}_{t\,|\,t'}(X) = \sum_{\theta^i\in\Theta^i_t} f^{i,\theta^i}_{t\,|\,t'}(X)\,\beta^{i,\theta^i}_{t\,|\,t'}\\
&= \begin{cases}
\displaystyle\sum_{\theta^i\in\Theta^i_t} f^{i,\theta^i}_{t\,|\,t'}(x)\, r^{i,\theta^i}_{t\,|\,t'}\beta^{i,\theta^i}_{t\,|\,t'} = \sum_{\varphi^i\in\Theta^i_t} r^{i,\varphi^i}_{t\,|\,t'}\beta^{i,\varphi^i}_{t\,|\,t'} \sum_{\theta^i\in\Theta^i_t} f^{i,\theta^i}_{t\,|\,t'}(x)\, \frac{r^{i,\theta^i}_{t\,|\,t'}\beta^{i,\theta^i}_{t\,|\,t'}}{\displaystyle\sum_{\varphi^i\in\Theta^i_t} r^{i,\varphi^i}_{t\,|\,t'}\beta^{i,\varphi^i}_{t\,|\,t'}}\,, & X=\{x\},\\
\displaystyle\sum_{\theta^i\in\Theta^i_t}\left(1-r^{i,\theta^i}_{t\,|\,t'}\right)\beta^{i,\theta^i}_{t\,|\,t'} = 1-\sum_{\theta^i\in\Theta^i_t} r^{i,\theta^i}_{t\,|\,t'}\beta^{i,\theta^i}_{t\,|\,t'}\,, & X=\emptyset,
\end{cases}\\
&= \begin{cases}
r^{i}_{t\,|\,t'} f^{i}_{t\,|\,t'}(x)\,, & X=\{x\},\\
1-r^{i}_{t\,|\,t'}, & X=\emptyset.
\end{cases} \qquad (8.11)
\end{aligned}
$$

Fortunately, this is again a BPP, with state distribution given by a mixture. The mixture weights are the hypothesis-probabilities weighted by the existence-probabilities and then normalized, with the hypothesis-conditioned distributions as the corresponding distributions in the mixture. The existence-probabilities, are quite naturally, given by hypothesis-weighted averaged existence-probability.

The TOMB/P is therefore still a mixture, but forgoes the need to consider global hypotheses after the approximation has been made. This mixture will also increase as the components gets updated through time, and there will be a need for mixture reduction within the single track mixtures as well.

8.4.1 Mixture reduction in GM-MM-TOMPB/P

After approximating the true multi target distribution by a component-labeled factored one, we have a single target distribution that is a mixture. If we now assume that this is a Gaussian mixture, or at least has been approximated to be one, we can use Runnall's distance described in section 5.3.3, to reduce this mixture down to a certain number of components, or until some maximum cumulative divergence has been reached. This is done by finding the pairwise distance between all components in the mixture, selecting the two components that give the smallest distance and merging them by moment matching. Moment matching is described in section 5.3.2. This is then done recursively until a certain number in the mixture is reached or the cumulative distance will go over a given threshold if further reduced.

The first use of Gassian mixture for target tracking using mixture reduction algorithms is often attributed Salmond [1990, 2009] in the single target case and Pao [1994] in the multi target case, where probabilistic data association (PDA) and JPDA probabilities were used, respectively.

If a hybrid state space is used, such as for multiple models, multiple target identities, detection levels and so on, one does this procedure for each of the discrete states in the discrete state space.

Whether if there is a hybrid state space or not, this should be applied right before the prediction step to get as good an approximation as possible when the mixture is reduced. For a hybrid state space, this is therefore done after the mixing step (prediction of discrete states) so that the weight change and new components are considered in the reduction. If there is no discrete state, right after measurements will be the same time as right before prediction, as there are usually no other steps, other than maybe computing an estimate for an operator.

8.4.2 JIPDA, JPDA, IPDA and PDA

If the mixture in the preceding subsection is reduced down to a single Gaussian, i.e. moment matching the entire distribution, one effectively gets a JIPDA filter of Mušicki and Evans [2002]. However one also tracks the undetected target quantity that the update equations of JIPDA neglects [Williams, 2015b]. If the target existence is known, i.e. the existence probability is one or zero and no births or undetected target quantities are present: JIPDA reduces to JPDA. Hence JIPDA can be seen as an extension of the JPDA filter of Fortmann et al. [1983] to incorporate uncertain target existence. If target existence is unknown, but there is at most one target in the scene, JIPDA reduces to the integrated probabilistic data association (IPDA) filter of Musicki et al. [1994]. This again reduces to the PDA filter of Bar-Shalom and Tse [1975] if there is known to be one, and only one, target in the scene. JPDA and PDA are more recently and perhaps better explained by Bar-Shalom et al. [2011].

In case of a hybrid state space, if one reduces the mixture down to one component for each discrete hypothesis, the algorithm reduces to IMM-JIPDA, IMM-JPDA, IMM-IPDA and IMM-PDA, which are described by Musicki and Suvorova [2008], Chen and Tugnait [2001], Musicki and Suvorova [2008], and Houles and Bar-Shalom [1989], respectively.

With this we have seen how several of the key filters, that marginalize over the measurement hypothesis and treat targets as independent, can be derived from the true component labeled PMBM. This involved applying the appropriate assumptions and using the moment projection of the KL-divergence onto a factored distribution. Individual distributions consisted of the discrete state distribution and one Gaussian for each discrete state. This can in some way be seen as a type of verification of these algorithms, but it also implies which directions one should go, if one of these simpler algorithms should be insufficient in some aspects.

For instance, if one is using a JPDA but struggling with many appearing and disappearing targets, one can look toward the JIPDA along with an appropriate modeling of the birth intensity, $\eta_t(x)$, and survival probability, $\mathrm{P}_s(x)$. If on the other hand, the issue is in too much coalescence, one can look at having a Gaussian mixture instead, so that later measurements hopefully will decrease the weights of the components that cause coalescence. The coalescence phenomena is treated to some degree by Blom and Bloem [2000, 2002], but their approach is not robust to track-switching as they prune association hypotheses. Keeping the associations, but merging similar components, might be a way to avoid a bit of both track coalescence and track switching.

8.5 Relationship to CPHD and PHD

The cardinalized probability hypothesis density (CPHD) and PHD filter [Mahler, 2007, ch.16] can also be seen as a KL-divergence approximation to the PMBM filter [Williams, 2015a]. One can get the CPHD filter by approximating the unlabeled PMBM posterior recursively by an i.i.d. target state distribution along with the cardinality distribution [Williams, 2015a, appendix A.A]. The PHD filter takes this one step further and approximates the cardinality distribution to be Poisson.

Section 5.5 showed that the i.i.d. SDF moment projection indeed uses the normalized PHD as its distribution. Thus we have already shown that approximating any SDF by an i.i.d. SDF with an unconstrained cardinality distribution, one gets the CPHD. The PHD filter is achieved if one approximates the cardinality distribution to be Poisson. The cardinality distribution can be handled separately from the spatial distribution, as we saw in section 5.5, so for the Poisson approximation it is just to find the expected number of targets, as we saw in section 5.3.4, with the distribution being given by the PHD. Since the PHD integrates to the expected number of targets, we have that the PHD is in fact the same as the PPP intensity.

What remains to be seen is if the filter has the same prediction and update equations as the moment projection gives. The PHD is given by

$$D(x) = \int p(X \cup \{x\})\, \delta X. \tag{8.12}$$

We know that p contains a permutation in some way or another. In the case where p is a PMBM the only permutation is over the components, i.e. the PPP and the different BPPs. We then know that x is going to be permuted between the components, and can hold it fixed in one component when we do the integral over the others, and then sum over the permutation of x afterwards. The MBs integrates to one when x is not fixed to it, and so does the PPP. Fixing x to one BPP enables us to take that out of the set integral, and the integral is therefore over a PMB which integrates to unity. The same can be said about the PPP where we then take the intensity out of the integral. One might be skeptic about the $n!$ and that the cardinality will be erroneous when taking a BPP or an intensity out of the integral, but note that the integral is taken over X only, and the cardinality of $X \cup \{x\}$ is $|X| + 1$, so it is all fine.

Taking the permutation after the integration, the PHD of a PMB with n BPP components is the sum of the PPP intensity and all the BPP with the case of existence;

$$D(x) = \lambda(x) + \sum_{i=1}^{n} r^i f^i(x). \tag{8.13}$$

Using this on (7.37) the multi target update step shown in section 7.1.2, will then be exactly equivalent to (16.108)-(16.109) by Mahler [2007], where $D_{k+1|k+1}(x) = \lambda_{t\,|\,t}(x)$, $D_{k+1|k}(x) = \lambda_{t\,|\,t\_}(x)$, $L_z(x) = h(z\,|\,x)$ and $p_D(x) = \mathrm{P}_d(x)$. Since we have already shown that the prediction steps of a PPP is equivalent to the prediction step of the PHD filter in section 7.1.1, the PHD filter can be seen as a recursive moment projection of the true multi target filter onto a PPP, after each update step.

In the time of writing it is not obvious to the author if the CPHD filter, as for instance shown in section 16.7 of Mahler [2007], is also equivalent to the KL-divergence of the true prediction and update steps, but that should be the case since it should be the true cardinality distribution and PHD.

8.6 Are there ways to combine MHT and JPDA?

We have seen that both MHT and JPDA are closely related to the true multi target distribution. We have also seen that marginalizing the joint association in a JPDA fashion, but keeping the mixture will end up as a Gaussian mixture, which in many ways can be seen as a single target TOMHT. A question that is still unanswered is if we can combine some of the global hypothesis structure of MHT with some marginalization along the lines of JPDA, while maybe also keeping the single target mixtures.

One approach to this might be to see if there are some associations that have less ambiguity and marginalize those out, while keeping the ones with higher ambiguity still inside the global hypothesis tree of the MHT. One would probably have to condition the marginals on the different global hypotheses that are remaining in the tree, and hence having to perform several marginalizations. There might be a lot of redundancy in these marginalizations so that there exists shortcuts instead of starting from scratch when conditioning on a new hypothesis. This approach can probably be simplified if analyzed along the lines of the efficient hypothesis management 2 of Horridge and Maskell [2006] which is an enhancement of an algorithm of Maskell et al. [2004]. The algorithm uses the mutual exclusion principle to make computational graphs that can be seen to be an instance of BP in an appropriate PGM and association related variables.

Unfortunately this idea is not appropriately developed, and will not be investigated further here due to time limits. It will however remain a topic of future investigation.

9 LBP association

The calculation of the JPDA quantity is still combinatorial in the number of targets and measurements, which has led to several approximations [Romeo et al., 2010; Maskell et al., 2004; Williams and Roslyn, 2014]. A part of the problem is that one actually have to track and consider associations on false track as well as true tracks. The more ambiguous the measurements are, in terms on low detection probability and high clutter intensity, the more false tracks one have to track for a longer time to get a sense on which of them are true or false. Since the marginal association probabilities are combinatorial the number of operations needed, one therefore quickly run into trouble. We are here going to look at a recently developed approximation based on LBP by Williams and Roslyn [2014].

This chapter is largely taken from Tokle [2017]. However, the LBP model incorporating multiple models, or more generally discrete states, has been shown here to be an unnecessary complication, as the discrete states can be correctly conditioned on the association prior to marginalization. As such the problems found by Tokle [2017] regarding the LBP algorithm giving results similar to that of a MAP estimate is circumvented, and the LBP algorithm should be just as good to use with a hybrid state space as with only a continuous state space.

9.1 The joint association probability and its factors

Finding the marginal association probabilities between two sets, e.g. targets and measurements, can be represented as inference on a graph, were the nodes are latent association variables. One way to do this is by using BP on the graph. There are several parameterizations of the association, and hence there are also several ways of making the graph. Following Williams and Roslyn [2014] we will over-parameterize this by using both target to measurement probabilities a_t^i and measurement to target probabilities b_t^j with constraints to make the association consistent.

The joint posterior after measurement over the association variables are given by the product of the single target hypothesis weights by the assumption that the targets are independent according to

$$p(a_t^1, \dots, a_t^{n_t}, b_t^1, \dots, b_t^{m_t} \mid Z^t) \propto \gamma(a_t, b_t) \prod_{i=1}^{n_t} w_{t\,|\,t}^{i,a_t^i} \prod_{j|b^j=0} w_{t\,|\,t}^{n_t+j,1} \tag{9.1}$$

$$\propto \prod_{i=1}^{n_t} \left[\psi_i(a_t^i) \prod_{j=1}^{m_t} \psi_{i,j}(a_t^i, b_t^j) \right] \prod_{j=1}^{m_t} \psi_j(b_t^j), \tag{9.2}$$

where the notation used is given by

$$\psi_i(a_t^i) = w_{t\,|\,t}^{i,a_t^i}, \tag{9.3}$$

$$\gamma(a_t, b_t) = \prod_{j=1}^{m_t} \psi_{i,j}(a_t^i, b_t^j), \tag{9.4}$$

$$\psi_{i,j}(a_t^i, b_t^j) = \begin{cases} 0, & a_t^i = j > 0 \not\Leftrightarrow b_t^j = i > 0 \\ 1, & \text{otherwise} \end{cases}, \tag{9.5}$$

$$\psi_j(b_t^j) = \begin{cases} 1, & b_t^j > 0 \\ w_{t\,|\,t}^{n_t+j,1}, & b_t^j = 0 \end{cases}. \tag{9.6}$$

Here the factor $\gamma(a_t, b_t) \in \{0, 1\}$ assures a consistent association, i.e. $a_t^i = j \iff b_t^j = i$, and is factorized by $\psi_{i,j}$. From (9.2) it is evident that the joint distribution of the association variables factorizes into factors of at most two components, and that these factors can be used in a factor graph.

9.2 The PGM and messages of the association distribution

The MRF and the factor graph for this factorization is a bipartite graph between the target to measurement and the measurement to target variables as shown in Figure 9.1a and Figure 9.1b respectively. The messages in the factor graph (Figure 9.1b) can be formulated as

$$\mu_{\psi_i \to a_t^i}(a_t^i) = \psi_i(a_t^i), \tag{9.7}$$

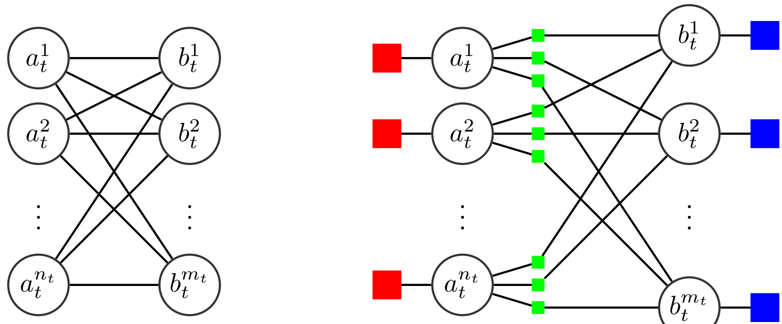
$$\mu_{\psi_{i,j} \to a_t^i}(a_t^i) = \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j) \mu_{b_t^j \to \psi_{i,j}}(b_t^j), \tag{9.8}$$

$$\mu_{a_t^i \to \psi_{i,j}}(a_t^i) = \mu_{\psi_i \to a_t^i}(a_t^i) \prod_{j' \neq j} \mu_{\psi_{i,j'} \to a_t^i}(a_t^i), \tag{9.9}$$

$$\mu_{\psi_j \to b_t^j}(b_t^j) = \psi_j(b_t^j), \tag{9.10}$$

$$\mu_{b_t^j \to \psi_{i,j}}(b_t^j) = \mu_{\psi_j \to b_t^j}(b_t^j) \prod_{i' \neq i} \mu_{\psi_{i,j} \to b_t^j}(b_t^j), \tag{9.11}$$

$$\mu_{\psi_{i,j} \to b_t^j}(b_t^j) = \sum_{a_t^i} \psi_{i,j}(a_t^i, b_t^j) \mu_{a_t^i \to \psi_{i,j}}(a_t^i). \tag{9.12}$$



(a) Markov random field

(b) Factor graph. (ψ_i, $\psi_{i,j}$, ψ_j).

Figure 9.1: Graphs of the overparameterized association

This set of messages is written very explicitly and it is seen that no factor has more than two variables, and hence the factor to variable message is only the marginalization of the factor multiplied with the incoming message. We can therefore pass the messages through the factors and just consider the messages as going between the variables, that is, simply following the edges in Figure 9.1a.

$$
\begin{aligned}
\mu_{a_t^i \to b_t^j}(b_t^j) &= \sum_{a_t^i} \psi_{i,j}(a_t^i, b_t^j)\mu_{a_t^i \to \psi_{i,j}}(a_t^i) && (9.13)\\
&= \sum_{a_t^i} \psi_{i,j}(a_t^i, b_t^j)\mu_{\psi_i \to a_t^i}(a_t^i) \prod_{j' \neq j} \mu_{\psi_{i,j'} \to a_t^i}(a_t^i) && (9.14)\\
&= \sum_{a_t^i} \psi_{i,j}(a_t^i, b_t^j)\psi_i(a_t^i) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(a_t^i), && (9.15)\\
\mu_{b_t^j \to a_t^i}(a_t^i) &= \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j)\mu_{b_t^j \to \psi_{i,j}}(b_t^j) && (9.16)\\
&= \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j)\mu_{\psi_j \to b_t^j}(b_t^j) \prod_{i' \neq i} \mu_{\psi_{i,j} \to b_t^j}(b_t^j) && (9.17)\\
&= \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j)\psi_j(b_t^j) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(b_t^j). && (9.18)
\end{aligned}
$$

These equations only give two different types of message, namely from targets to measurements, and measurements to targets. Again, following Williams and Roslyn [2014] these messages can be rewritten into two different cases due to the consistency factors;

$$\mu_{a_t^i \to b_t^j}(b_t^j) = \sum_{a_t^i} \psi_{i,j}(a_t^i, b_t^j)\psi_i(a_t^i) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(a_t^i) \tag{9.19}$$

$$= \begin{cases} \psi_i(j) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(j), & b_t^j = i \\ \sum_{a_t^i \neq j} \psi_i(a_t^i) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(a_t^i), & b^j \neq i \end{cases}, \tag{9.20}$$

$$\mu_{b_t^j \to a_t^i}(a_t^i) = \sum_{b_t^j} \psi_{i,j}(a_t^i, b_t^j)\psi_j(b_t^j) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(b_t^j) \tag{9.21}$$

$$= \begin{cases} \psi_j(i) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(i), & a_t^i = j \\ \sum_{b_t^j \neq i} \psi_j(b_t^j) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(b_t^j), & a_t^i \neq j \end{cases}. \tag{9.22}$$

9.3 Simpliying the messages by scaling

Notice that there are only two separate values that the messages between the target to measurement pairs can take. As the messages are not normalized, they can be scaled at each iteration. Scaling by dividing by $\mu_{a_t^i \to b_t^j}(b_t^j \neq i)$ and $\mu_{b_t^j \to a_t^i}(a_t^i \neq j)$ respectively will now leave us with many messages that are unity, and only one that is non-unity per edge;

$$\mu_{a_t^i \to b_t^j}(b_t^j) = \begin{cases} \dfrac{\psi_i(j) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(j)}{\sum_{a_t^i \neq j} \psi_i(a_t^i) \prod_{j' \neq j} \mu_{b_t^{j'} \to a_t^i}(a_t^i)} = \dfrac{\psi_i(j)}{\sum_{a_t^i = j' \neq j} \psi_i(j')\mu_{b_t^{j'} \to a_t^i}(j')}, & b_t^j = i \\ 1, & b^j \neq i \end{cases} \tag{9.23}$$

$$\mu_{b_t^j \to a_t^i}(a_t^i) = \begin{cases} \dfrac{\psi_j(i) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(i)}{\sum_{b_t^j \neq i} \psi_j(a_t^j) \prod_{i' \neq i} \mu_{a_t^{i'} \to b_t^j}(b_t^j)} = \dfrac{1}{\sum_{b_t^j = i' \neq i} \psi_j(a_t^j)\mu_{a_t^{i'} \to b_t^j}(i')}, & a_t^i = j \\ 1, & a_t^i \neq j \end{cases} \tag{9.24}$$

We have now essentially reduced the equations to only send messages between the consistent association pairs, and neglect the messages without real information (it is not useful information what a variable does if the listening variable does not receive this). The marginal probabilities of the variables can again be found by multiplying all the incoming factors (mostly unity due to the scaling) and normalizing

$$p(a_t^i = j \mid Z^t) = \frac{\psi_i(j)\mu_{b_t^j \to a_t^i(j)}}{\sum_{i'=0}^{n_t} \psi_i(j)\mu_{b_t^j \to a_t^i}(j)}, \tag{9.25}$$

$$p(b_t^j = i \mid Z^t) = \frac{\psi_j(i)\mu_{a_t^i \to b_t^j}(i)}{\sum_{j'=0}^{m_t} \mu_{a_t^i \to b_t^j}(i)}. \tag{9.26}$$

9.4 Remarks

Williams and Roslyn [2014] also show that iterating these two sets of messages converges and they also give a convergence criterion that gives bounds on the distance from the true converged values. They also pointed to the fact that it will converge to a valid PDF over the variables, but not necessarily the correct one.

This procedure can be seen to solve a constrained optimization problem and minimizes something called the Bethe free energy while being constrained to be a valid distribution [Williams and Lau, 2018]. The clutter and non detection cases can be seen to help convergence of the LBP algorithm through both the Bethe free energy and experimental results. Williams and Lau [2018] has now extended this method to handle multi scan problems using something called fractional free energy, where one scales a part of the objective in the Bethe free energy, and choosing the correct but yet unknown scaling is proven to give the correct result. The resulting procedure is shown to outperform other BP based multi scan procedures on a limited set of experiments. A more thorough look into this approach is another future research topic.

The LBP scheme shown here provides a good way to overcome the combinatorial complexity of calculating these marginal association probabilities. However, as pointed out by Williams and Roslyn [2014], it is still unknown how this compares to the efficient hypothesis management 2 of Horridge and Maskell [2006] that provides the exact probabilities rather than approximations. The latter relies on the problem being sparse, while the LBP does not, so it is believed that the two algorithms will be appropriate for complementary situations.

10 Some aspects of the undetected target intensity

10.1 Implications of not estimating undetected targets

If the modeling assumptions are done correctly, and in addition the birth intensity is stationary, the undetected target intensity will converge with time, but possibly stay nonuniform in space. One might therefore be tempted to approximate this as constant, especially after a while. We will therefore take a look at what are the implications are on the birth intensity when assuming either the existence probability or expected number of undetected targets have a specific value. In this section we will use r_0 to denote the initial existence probability.

The use of other initialization procedures, for instance cascaded logic track formation such as $M_1/N_1\&M_2/N_2$ section 3.3 & 7.3 of Bar-Shalom et al. [2011], probably both can and should be considered in this framework as well, so that one can infer what it assumes on the relation between the birth process and target existence probability acceptance threshold.

10.1.1 Birth process when assuming a constant expected number for the undetected targets

We shall now express $\bar{\lambda}_{t\,|\,t_-}$ in terms of a stationary birth process, $\bar{\eta}_{t-1} = \bar{\eta}$, assuming that the estimation procedure has been going on for a long time so that λ has converged and that we have $\lambda^{\infty}_{t\,|\,t_-} \triangleq \lambda_{t\,|\,t_-} = \lambda_{t_-\,|\,t_{-2}}$. We can write a full time update as

$$\lambda^{\infty}_{t\,|\,t_-}(x_t) = \eta(x_t) + \int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\,\mathrm{P}_s(x)\,(1-\mathrm{P}_d(x))\,\lambda^{\infty}_{t\,|\,t_-}(x)\,\mathrm{d}x, \tag{10.1}$$

$$\begin{aligned}
\bar{\lambda}^{\infty}_{t\,|\,t_-} f^{\lambda}_{t\,|\,t_-}(x_t) &= \bar{\eta}\, f^{\eta}(x_t) + \bar{\lambda}^{\infty}_{t\,|\,t_-} \int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\,\mathrm{P}_s(x)\,(1-\mathrm{P}_d(x))\, f^{\lambda}_{t\,|\,t_-}(x)\,\mathrm{d}x, &(10.2)\\
&= \bar{\eta}\, f^{\eta}(x_t) + (1-\bar{\mathrm{P}}^{\lambda}_d)\bar{\lambda}^{\infty}_{t\,|\,t_-} \int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\,\mathrm{P}_s(x)\, f^{\lambda}_{t\,|\,t}(x)\,\mathrm{d}x, &(10.3)\\
&= \bar{\eta}\, f^{\eta}(x_t) + \bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)\bar{\lambda}^{\infty}_{t\,|\,t_-} \int\limits_{x\in\mathcal{X}} f(x_t\,|\,x)\, f^{\lambda,\mathrm{s}}_{t\,|\,t}(x)\,\mathrm{d}x, &(10.4)
\end{aligned}$$

where“s” reads surviving, and the average detection probability of the undetected targets is given by

$$\bar{\mathrm{P}}^{\lambda}_d = \int\limits_{x\in\mathcal{X}} \mathrm{P}_d(x)\, f^{\lambda}_{t\,|\,t_-}(x)\,\mathrm{d}x, \tag{10.5}$$

the average survival probability of the undetected targets as

$$\bar{\mathrm{P}}^{\lambda}_s = \int\limits_{x\in\mathcal{X}} \mathrm{P}_s(x)\, f^{\lambda}_{t\,|\,t}(x)\,\mathrm{d}x, \tag{10.6}$$

and the state distribution of undetected surviving targets as

$$\begin{aligned}
f^{\lambda,\mathrm{s}}_{t\,|\,t}(x) &= \frac{\mathrm{P}_s(x)\, f^{\lambda}_{t\,|\,t}(x)}{\bar{\mathrm{P}}^{\lambda}_s} = \frac{\mathrm{P}_s(x)\,(1-\mathrm{P}_d(x))\, f^{\lambda}_{t\,|\,t_-}(x)}{\bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)}. &(10.7)\\
&= \frac{\mathrm{P}_s(x)\,(1-\mathrm{P}_d(x))\,\lambda^{\infty}_{t\,|\,t_-}(x)}{\bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)\bar{\lambda}^{\infty}_{t\,|\,t_-}}
\end{aligned}$$

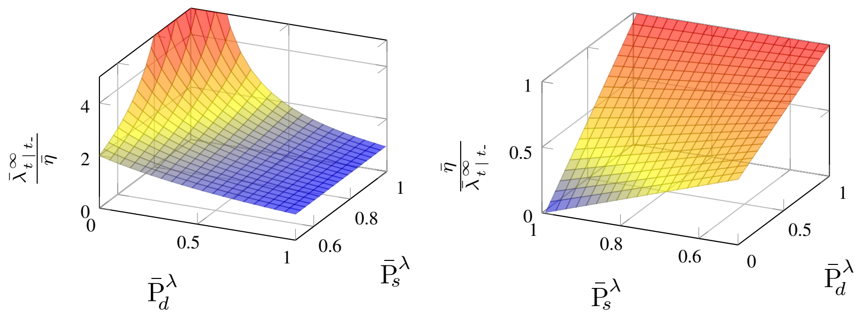
Integrating with respect to x_t and rearranging gives the expected converged number of undetected targets as a function of the expected number of born targets

$$\bar{\lambda}_{t\,|\,t_-} = \bar{\eta} + \bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)\bar{\lambda}_{t\,|\,t_-} = \frac{\bar{\eta}}{1-\bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)}, \tag{10.8}$$

rearranging again to end up at the number of born targets as a function of the number of undetected targets

$$\bar{\eta} = \Big(1-\bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)\Big)\,\bar{\lambda}^{\infty}_{t\,|\,t_-} \leq \bar{\lambda}^{\infty}_{t\,|\,t}. \tag{10.9}$$

These functions are presented in Figure 10.1. For $\bar{\mathrm{P}}^{\lambda}_d$ close to unity or $\bar{\mathrm{P}}^{\lambda}_s$ close to zero we get that $\bar{\lambda}_{t\,|\,t_-} \approx \bar{\eta}$, and even more so if both occur. This is also intuitive, as low probability of survival means that the targets in the scene previously, are more likely to have left, and high detectability means that the targets previously in the scene are very likely to have



(a) undetected from birth

(b) birth from undetected (note the reversed order and direction on axis)

Figure 10.1: The relationship between the converged expected number of undetected targets and the expected number of arriving targets, as a function of detection probability and survival probability.

already been detected, so almost all undetected targets are the newly born ones. This, of course, also goes the other way around.

As $\bar{\lambda}_{t\,|\,t_-}$ converges after some time for proper modeling, with stationary birth process and non-moving sensor, we can model it as constant. For high detectability and/or low survival we even have $\bar{\lambda}_{t\,|\,t_-} \approx \bar{\eta}$, but one makes a mistake by assuming this if these conditions do not apply, although the error may be small. One important thing to consider is that the actual number of born targets is always less than the number of undetected targets, which should also be evident since there are some surviving in addition to the newly born ones. If one approximates $\bar{\lambda}_{t\,|\,t_-} = \bar{\eta}$ and models $\bar{\eta}$ directly to use in an algorithm, $\bar{\lambda}_{t\,|\,t_-}$ is in fact always an underestimation of what one believes. It is therefore probably better to use the relation shown above and get the appropriate approximation of $\bar{\lambda}_{t\,|\,t_-}$ according to ones assumptions.

Note that this does not make any claims on the distribution of the undetected targets, only on the number. For a non-uniform birth and death process, which one in general should at least consider, there will naturally be a gradient in $\lambda(x)$ towards places of higher birth intensity from high detectability and low survival places. For moving sensors this becomes more complicated, as shown by Williams [2012].

Also note that the proper initial distribution, discussed earlier, corresponds to setting $\bar{\mathrm{P}}_d = 0$ since that is the same as having no sensor. If the survival probability is high, say for instance $\bar{\mathrm{P}}^{\lambda}_s = 0.95$, the initial expected number will be $20\bar{\eta}$, a much larger number than if one assumes the converged number. However, this will quite quickly reduce as one can se from the fact that this converges with a discrete time eigenvalue of $\bar{\mathrm{P}}^{\lambda}_s \left(1 - \bar{\mathrm{P}}^{\lambda}_d\right)$ which in the case of $\bar{\mathrm{P}}^{\lambda}_d = 0.95$ is less than 0.05. This means that after just a few scans one can reasonably approximate a constant number of undetected targets, but one should be a bit more cautious when the detection probability is low and survival probability is high.

10.1.2 The birth process when one assumes a constant initial existence probability

An even simpler approximation is to use a constant initial existence probability. From (7.37) we have

$$r_0 = \frac{\int\limits_{x\in\mathcal{X}} h(z\,|\,x)\,\mathrm{P}_d(x)\,\lambda_{t\,|\,t\_}(x)\,\mathrm{d}x}{\int\limits_{x\in\mathcal{X}} h(z\,|\,x)\,\mathrm{P}_d(x)\,\lambda_{t\,|\,t\_}(x)\,\mathrm{d}x + \mu_t\,(z)} \tag{10.10}$$

$$= \frac{h^{\lambda}_{t\,|\,t\_}(z)\,\bar{\mathrm{P}}^{\lambda}_d\,\bar{\lambda}_{t\,|\,t\_}}{h^{\lambda}_{t\,|\,t\_}(z)\,\bar{\mathrm{P}}^{\lambda}_d\,\bar{\lambda}_{t\,|\,t\_} + \bar{\mu}_t h^{\mu}_t\,(z)}, \tag{10.11}$$

with

$$h^{\lambda}_{t\,|\,t\_}(z) = \int\limits_{x\in\mathcal{X}} h(z\,|\,x)\,\frac{\mathrm{P}_d(x)\,\lambda_{t\,|\,t\_}(x)}{\bar{\mathrm{P}}^{\lambda}_d\,\bar{\lambda}_{t\,|\,t\_}}\,\mathrm{d}x = \int\limits_{x\in\mathcal{X}} h(z\,|\,x)\,f^{\lambda,d}_{t\,|\,t\_}(x)\,\mathrm{d}x, \tag{10.12}$$

and can get $\bar{\lambda}$ in terms of an assumed constant r_0,

$$\bar{\lambda}_{t\,|\,t\_} = \frac{r_0}{1-r_0}\,\frac{\bar{\mu}_t h^{\mathrm{fa}}(z)}{h^{\lambda}_{t\,|\,t\_}(z)\,\mathrm{P}^{\lambda}_d} = \frac{r_0}{1-r_0}\,\frac{\bar{\mu}_t}{\mathrm{P}^{\lambda}_d}. \tag{10.13}$$

The last equality follows from $\bar{\lambda}_{t\,|\,t\_}$ not depending on z and $h^{\cdot}(\cdot)$ are valid PDFs. Assuming a constant r_0 thus also assumes $h^{\lambda}_{t\,|\,t\_}(z) = h^{\mu}_t(z)$, which of course is generally not the case, and shows that this should probably not be done, or at least should only be done with great care in considering the implications on the tracking algorithm. For instance, the uniform clutter model $h^{\mu}_t(z) = V^{-1}_{\mathcal{Z}}$, almost certainly implies $f^{\lambda}_{t\,|\,t\_}(x) = V^{-1}_{\mathcal{X}}$ (at least for the states that are directly related to the measurement), and therefore almost certainly also implies $f^{\eta}_{t\,|\,t\_}(x) = V^{-1}_{\mathcal{X}}$, which, as previously stated, one should probably not have in most applications.

Furthermore, inserting (10.9) into (10.13) and rearranging we arrive at

$$\bar{\eta} = \frac{r_0}{1-r_0}\,\frac{1-\bar{\mathrm{P}}^{\lambda}_s(1-\bar{\mathrm{P}}^{\lambda}_d)}{\bar{\mathrm{P}}^{\lambda}_d}\,\bar{\mu}_t = \frac{r_0}{1-r_0}\left(1 + \frac{(1-\bar{\mathrm{P}}^{\lambda}_s)(1-\bar{\mathrm{P}}^{\lambda}_d)}{\bar{\mathrm{P}}^{\lambda}_d}\right)\bar{\mu}_t \tag{10.14}$$

$$\geq \frac{r_0}{1-r_0}\bar{\mu}_t, \ \forall\,\bar{\mathrm{P}}^{\lambda}_s,\,\bar{\mathrm{P}}^{\lambda}_d \in [0,1], \tag{10.15}$$

and see that assuming the value of r_0 also assumes a birthrate after steady state has been reached. Since this relation is there implicitly in choosing r_0, one is probably better off by making it explicit by seeing if this $\bar{\eta}$ is reasonable for the application.

10.2 Some potentials in the undetected target intensities

Williams [2012] pointed out the potential gain of tracking this the undetected targets, and implemented a grid based approximation of $\lambda(x)$ which he updated through time. He showed with this, especially for sensors with varying field of view, that this gave an overall increase in tracking performance. Horridge and Maskell [2011] also did similar experiments with a particle approximation. Both of the latter seem to be incorporating a relatively large value on uniform distribution for birth, which is often not the case in many real scenarios where targets typically arrive at the border of the surveillance region, whereas for example Mori et al. [1986] seems to have done this.

It was also mentioned the possibility to consider the birth intensity as a union of PPPs, $\eta_t(x) = \sum_{\alpha \in \mathcal{A}} \eta_t^\alpha(x)$. Having it as a union, we can extract the latent birth component variable α and consider where the target came from. Tracking that distribution through time can be done by having separate $\lambda^\alpha(x)$'s. Why would one do this, one might ask. The answer is application dependent, but one could for instance have birth components at different spatial locations, and one might be interested in where the targets originated, or one might model different birth densities for different "types" of targets. The latter can be done by augmenting the state with another discrete variable that does not change through time and then estimate this. Tracking discrete states in the PPP can, as has been pointed out earlier, be done by tracking one intensity per discrete state.

The undetected target intensity and the false alarm intensity, naturally have a tight relationship when it comes to the initial track existence. This relationship is point-wise in measurement space. Having a locally too high or too low ratio between the undetected target measurement intensity and false alarm intensity will over estimate or under estimate the initial existence probability, respectively. If one then think of false tracks as erroneous tracks with an existence probability greater than a given threshold and missed tracks as true tracks having existence probability lower than a given threshold, one would likely create more of the former by over estimating and the latter by under estimating the initial existence probability, and hence by having too high or too low ratio, respectively.

11 Summary, Discussion and Conclusions

11.1 Summary

In part I, this report first introduced PGMs and the related BP concept. It then provided an outline of Bayesian state estimation. A novel derivation of the KF equation was provided through considering the special case of a Gaussian linear model, where it was shown that one only needs to handle the quadratic form in the exponential when deriving the Gaussian product identity. Description of how to handle a hybrid state space was provided, showing that it is analogous to having only continuous states, and conditioning these on the discrete states, in a manner similar to that of the IMM. The IMM was seen to be a special case of this when applying the appropriate assumptions. Next, a brief introduction to RFS was provided describing the inherent notions. The novel way of seeing the set convolution as a marginalization provided a way of intuitively extracting what the author called component labels, in order to infer which underlying set a specific element came from. The last part of the theory background was the chapter on the KL-divergence. This described some standard results on how one distribution can be approximated by another, with some concrete examples.

Part II first provided the assumptions of the standard model of multi target tracking, and how these are translated into the RFS framework. Then the derivation of the PMBM filter followed with the inclusion of hybrid state space and component labeling.

Part III then introduced the total target information filter, along with an interpretation of the association hypotheses and component labeling. It was seen that the associations could be viewed as being between measurements at different time steps, and therefore make what we called measurement-tracks. The component labeling was seen to indicate which of these measurement-tracks a target belonged to and could therefore also be considered to form a track existence hypothesis. It was also pointed out how the PMBM filter was a marginalization within this distribution. Further, it was shown that it was related to the random finite set of trajectory, as for instance given by Granström et al. [2018] and García-Fernández et al. [2016]. Garcia-Fernandez et al. [2018] pointed out how the generalized LMB filter of Vo and Vo [2011] was a less efficient parameterization of the same events as in the labeled PMBM, which is very similar to the component labeled PMBM we have discussed here. The same reasoning of measurement-tracks and existence hypothesis also applies to the component labeled PMBM.

Next the relationship between PMBM and MHT was discussed, where we demonstrated that our component labeling along with the measurement-tracks created the same hypotheses as Reid [1979]. The same type of reasoning lead us to the TOMHT of Kurien [1990]. Furthermore, we moment projected the PMBM onto a distribution, factored according to the component labels, and naturally arrived at the TOMB/P of Williams [2015b]. A procedure for mixture reduction within the single track state space by using Gaussian mixtures was described, namely Runnalls [2007] algorithm for merging components using moment matching. It was pointed out how reducing this down to one Gaussian per discrete state would result in (a corrected version of) IMM-JIPDA and its degenerate cases, by applying assumptions. Using the i.i.d. target assumption was also looked into, and it was seen that the PHD filter CPHD filter could also be seen as a recursive moment projection of the unlabeled PMBM onto an i.i.d. SDF with Poisson and unconstrained cardinality, respectively. A slightly crude approach to combining JPDA and MHT was lastly described.

We then tackled the combinatorial complexity of calculating the marginal association probabilities by using LBP on the over-parameterized association distribution. The discrete states could be seen as states and hence conditioned on the association before the association probabilities were calculated. This was in contrast to what Tokle [2017] was considering, and his (or my if you like) problems with having erroneous marginals by forcing the discrete states to be a part of LBP was circumvented.

Finally, we looked at the undetected target intensity and saw what different approximations would mean in terms of birth intensity, and when one should be cautious. We have also seen that the relative size of the undetected target intensity to false alarm intensity is the initial existence probability and point-wise erroneous modeling of this relationship could therefore cause too many false tracks or miss too many true tracks, in terms of a given existence threshold, if too high or too low, respectively.

11.2 Concluding remarks

This study provided a thorough investigation into the theoretical framework and background around the standard model used in MTT, which served as a foundation for developing novel derivations of the PMBM filter and how to include a hybrid state space. Through the derivations it was seen that the components, i.e. the underlying sets in the union, could be handled independently through the prediction and update step under a given association to provide track continuity, and hence the possibility for track labeling. Additionally, track labeling could be seen as being a latent variable pointing to individual sets, being either a single track or the undetected targets, in the union of targets and will follow a specific track after detection and hence in a manner provide target identities. A total target information distribution was stated as a way of considering the problem, and further giving insight into identities and tracks. Furthermore, the relationship to most of the well known MTT filters were provided, recognizing them as an approximation of the PMBM or as attaining association variables within their distribution. Lastly, ways of approximating the association probabilities or the undetected target quantity were provided.

The results presented here is limited by the assumptions of the standard model. Many applications can have multiple measurements per target Granstrom et al. [2016], some might have targets with coordinated motion where targets are avoiding each other or moving together, some might have different birth processes and some might have measurements coming from more than one targets. These are not covered by the standard model and therefore also not by our derivations. One can use the same principles as discussed here, but one needs to incorporate appropriate modeling and re-derive at least parts of what is presented here if these assumptions are not met.

Viewing multiple models of IMM as being a discrete state allows one to treat them in a similar manner to the continuous states. This enables appropriate conditioning and thus possible to avoid the increase in computational complexity of having them in the association in a similar manner as Musicki and Suvorova [2008] did. It also provides ways to treat for instance varying detectability [Mušicki and Evans, 2002] as this also can be seen as being a discrete state and hence provide a different measurement model instead of dynamic model. Treating these together simplifies derivations to the same framework, but how to efficiently treat this in an implementation was not investigated here. Using IMM along with JPDA type algorithms is not straight forward, as different authors have done it in different ways. Some have included the models in the association [Chen and Tugnait, 2001; Blom and Bloem, 2002; Tokle, 2017], whereas others have not [de Feo et al., 1997; Musicki and Suvorova, 2008].

The possibility to handle the independent sets, being the components of the MTT SDF, independently through the prediction and update step provides a new way to establish track to establish track continuity and as such gain the possibility of labeling the track sets. To properly identify tracks and establish track continuity, may seem like a trivial and self explaining task. However, when for instance Mahler [2007], one of the founders or inventors of RFS for MTT, claims that labeling will make the estimation more complex, whereas Vo et al. [2014], another significant contributer to the field, claims that it makes it simpler, caution is needed. Hence, the insight that is given here, that goes in the same direction as Granström et al. [2018] and García-Fernández et al. [2016] although with another interpretation, can hopefully be of value towards the goal of understanding this.

The labeling interpretation given here can be phrased as: *a union of independent target sets will forever be a union of independent targets sets when the sets move and give measurements independently. Hence, we can do inference regarding which is which, thereby also providing a way of doing inference in the individual independent sets*. Detected track sets were seen as singleton sets following BPPs, whereas the still undetected target set was seen as a union of one or several undetected target sets coming from different birth processes following PPPs. Labeling the individual sets in the unuonized target set instead of the set elements differs from how others do it [Vo and Vo, 2011]. It is also intriguing when seeing the little extra notation needed, as one simply extracts the subset that is labeled according to the given component without changing anything withing the component. This is in contrast to having to introduce delta functions and indicator functions as in the LMB Vo and Vo [2011].

The total target information distribution provides a compact way of viewing the complete picture within MTT, although admittedly looking a bit nasty to handle. It could be seen as and extension of the PGM formulation of Meyer et al. [2018], although in a non obvious manner, but might nevertheless lend itself as a framework for studies of new approximative algorithms. Showing the relationship between PMBM, MHT, TOMB/P, JIPDA with its degenerate cases and PHD through statistical divergences clearly shows what type of information there is to gain or lose if opting for a different algorithm. This type of insight is good to have when performing the art of engineering and follows the principle *"Everything should be made as simple as possible, but not simpler"*.

Similarly, having good knowledge on how to simplify the marginalization when calculating the JPDA probabilities is valuable when ones' problem gets too big to solve using exact methods. The simplifying assumption of stationary undetected target intensity or initial existence probability and its relation the the birth intensity under ones' modeling parameters is also a good tool for a practical engineer.

All this provides a good theoretical foundation for MTT. Derivations and descriptions of the RFS filter incorporating hybrid state space and how to extract target identities from it provides a good insight into the problem at hand. Also the expansion into the full total target information distribution may provide future insight into new approaches to solving the MTT problem. Giving the relationship between this and several well-known algorithms can hopefully be of help when deciding on which algorithm to use and what approximations to apply in a given application.

11.3 Topics of future work

There are multiple paths of future work leading from this. One theoretical path is to look more rigorously into the labeling provided here, and see if there is an appropriate mathematical foothold to define it formally. A practical path using this type of labeling, could be an attempt at applying it in a PHD filter fashion, and attempt to establish a connection to the linear multi target IPDA [Musicki and La Scala, 2008], which treats other targets as clutter through the false alarm intensity.

Another practical path is to look into new approximation schemes on a PGM formulation of the total target information, using appropriate graph-techniques, free energy techniques and/or variational inference. Where the association is a subproblem remaining after marginalization of all the other variables and therefore also has the multi/single scan BP and free energy association techniques [Williams and Lau, 2018] as subproblems. Additionally, one could look into the efficient hypotheses management [Horridge and Maskell, 2006] as an instance of BP, and see firstly how this performs computationally compared to the LBP approach given here. Secondly, to look into possible ways of extending this to multi scan problems or to the measurement to track formulation. As previously pointed out, seeing if these algorithms can provide insight into combining notions from MHT and JPDA is also topic of future interest.

An even more practical path is that of evaluating an implementation of the Gaussian mixture IMM-JPDA and comparing performance gains in different types of scenarios. Also, investigating if the number of components in the mixture can be adapted automatically to deal with varying scenarios.

Bibliography

Bar-Shalom, Y., Chang, K., and Blom, H. (1991). Tracking of splitting targets in clutter using an interacting multiple model joint probabilistic data association filter. *Proc. 30th IEEE Conf. Decis. Control*, pages 2043–2048.

Bar-Shalom, Y., Li, X.-R., and Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., New York, USA.

Bar-Shalom, Y., Tian, X., and Willett, P. K. (2011). *Tracking and data fusion : a handbook of algorithms*. YBS Publishing.

Bar-Shalom, Y. and Tse, E. (1975). Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460.

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. Springer.

Blom, H. A. (1984). An efficient filter for abruptly changing systems. In *23rd IEEE Conf. Decis. Control*, pages 656–658. IEEE.

Blom, H. A. and Bar-Shalom, Y. (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Trans. Automat. Contr.*, 33(8):780–783.

Blom, H. A. and Bloem, E. A. (2000). Probabilistic data association avoiding track coalescence. *IEEE Trans. Automat. Contr.*, 45(2):247–259.

Blom, H. A. P. and Bloem, E. A. (2002). Combining IMM and JPDA for tracking multiple maneuvering targets in clutter. In *Proc. 5th Int. Conf. Inf. Fusion, FUSION 2002*, volume 1, pages 705–712. Int. Soc. Inf. Fusion.

Brekke, E. and Chitre, M. (2018). Relationship between Finite Set Statistics and the Multiple Hypothesis Tracker. *IEEE Trans. Aerosp. Electron. Syst.*, pages 1–1.

Brekke, E. F. and Chitre, M. (2017). The multiple hypothesis tracker derived from finite set statistics. In *20th Int. Conf. Inf. Fusion, Fusion 2017 - Proc.*, pages 1–8. IEEE.

Chen, B. and Tugnait, J. K. (2001). Tracking of multiple maneuvering targets in clutter using IMM/JPDA filtering and fixed-lag smoothing. *Automatica*, 37(2):239–249.

de Feo, M., Graziano, A., Miglioli, R., and Farina, A. (1997). IMMJPDA versus MHT and Kalman filter with NN correlation: performance comparison. *IEE Proc. - Radar, Sonar Navig.*, 144(2):49.

Fortmann, T. E., Bar-Shalom, Y., and Scheffe, M. (1983). Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association. *IEEE J. Ocean. Eng.*, 8(3):173–184.

García-Fernández, Á. F., Svensson, L., and Morelande, M. R. (2016). Multiple target tracking based on sets of trajectories.

Garcia-Fernandez, A. F., Williams, J. L., Granstrom, K., and Svensson, L. (2018). Poisson multi-Bernoulli mixture filter: direct derivation and implementation. *IEEE Trans. Aerosp. Electron. Syst.*, 54(4):1883 – 1901.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Granstrom, K., Baum, M., and Reuter, S. (2016). Extended Object Tracking: Introduction, Overview and Applications.

Granström, K., Svensson, L., Xia, Y., Williams, J., and García-Fernández, A. F. (2018). Poisson multi-Bernoulli mixture trackers: continuity through random finite sets of trajectories. *Proc. FUSION-18*, pages 986–994.

Horridge, P. and Maskell, S. (2006). Real-time tracking of hundreds of targets with efficient exact JPDAF implementation. In *2006 9th Int. Conf. Inf. Fusion, FUSION*, pages 1–8. IEEE.

Horridge, P. and Maskell, S. (2011). Using a Probabilistic Hypothesis Density filter to confirm tracks in a multi-target environment. *Proc. SDF-11*.

Houles, A. and Bar-Shalom, Y. (1989). Multisensor Tracking of a Maneuvering Target in Clutter. *IEEE Trans. Aerosp. Electron. Syst.*, 25(2):176–189.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Kurien, T. (1990). Issues in the design of practical multitarget tracking algorithms. In Bar-Shalom, Y., editor, *Multitarget-Multisensor Track. Adv. Appl.*, chapter Chapter 3, pages 43–83. Artech-House, Norwood, Massachusetts.

Le Cam, L. (1960). An Approximation Theorem for the Poisson Binomial Distribution. *Pacific J. Math.*, 10(4):1181–1197.

Li, X. R. and Bar-Shalom, Y. (1992). Mode-Set Adaptation in Multiple-Model Estimators for Hybrid Systems. *1992 Am. Control Conf.*, pages 1794–1799.

Li, X. R. and Bar-Shalom, Y. (1996). Multiple-model estimation with variable structure. *IEEE Trans. Automat. Contr.*, 41(4):478–493.

Li, X.-R. and Jilkov, V. P. (2005). Survey of maneuvering target tracking. Part V: Multiple-model methods. *IEEE Trans. Aerosp. Electron. Syst.*, 41(4):1255–1321.

Mahler, R. P. S. (2007). *Statistical multisource-multitarget information fusion*. Artech House.

Maskell, S., Briers, M., and Wright, R. (2004). Fast mutual exclusion. volume 5428, pages 526–536. International Society for Optics and Photonics.

Meyer, F., Kropfreiter, T., Williams, J. L., Lau, R., Hlawatsch, F., Braca, P., and Win, M. Z. (2018). Message Passing Algorithms for Scalable Multitarget Tracking. *Proc. IEEE*, 106(2):121–259.

Mori, S., Tse, E., Chong, C. Y., and Wishner, R. P. (1986). Tracking and Classifying Multiple Targets Without A Priori Identification. *IEEE Trans. Automat. Contr.*, 31(5):401–409.

Mušicki, D. and Evans, R. (2002). Joint Integrated Probabilistic Data Association - JIPDA. *Proc. 5th Int. Conf. Inf. Fusion, FUSION 2002*, 2(3):1120–1125.

Musicki, D., Evans, R., and Stankovic, S. (1994). Integrated probabilistic data association. In *IEEE Trans. Automat. Contr.*, volume 39, pages 1237–1241. IEEE.

Musicki, D. and La Scala, B. (2008). Multi-target tracking in clutter without measurement assignment. *IEEE Trans. Aerosp. Electron. Syst.*, 44(3):877–896.

Musicki, D. and Suvorova, S. (2008). Tracking in clutter using IMM-IPDA-based algorithms. *IEEE Trans. Aerosp. Electron. Syst.*, 44(1):111–126.

Pao, L. Y. (1994). Multisensor multitarget mixture reduction algorithms for tracking. *J. Guid. Control. Dyn.*, 17(6):1205–1211.

Reid, D. B. (1979). An Algorithm for Tracking Multiple Targets. *IEEE Trans. Automat. Contr.*, 24(6):843–854.

Romeo, K., Crouse, D. F., Bar-Shalom, Y., and Willett, P. (2010). The JPDAF in Practical Systems: Approximations. In Drummond, O. E., editor, *Signal Data Process. Small Targets 2010*, pages 76981I–76981I–10, Orlando.

Runnalls, A. R. (2007). Kullback-Leibler approach to Gaussian mixture reduction. *IEEE Trans. Aerosp. Electron. Syst.*, 43(3):989–999.

Salmond, D. J. (1990). Mixture reduction algorithms for target tracking in clutter. In *Signal Data Process. Small Targets 1990*, volume 1305, page 37. SPIE.

Salmond, D. J. (2009). Mixture reduction algorithms for point and extended object tracking in clutter. *IEEE Trans. Aerosp. Electron. Syst.*, 45(2):667–686.

Serfling, R. J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials. *SIAM Rev.*, 20(3):567–579.

Tokle, L.-C. N. (2017). Multi-target tracking using loopy belief propagation and multiple models. Technical report, Norwegian University of Science and Technology, Trondheim.

Vo, B. N., Vo, B. T., and Phung, D. (2014). Labeled random finite sets and the bayes multi-target tracking filter. *IEEE Trans. Signal Process.*, 62(24):6554–6567.

Vo, B. T. and Vo, B. N. (2011). A random finite set conjugate prior and application to multi-target tracking. *Proc. 2011 7th Int. Conf. Intell. Sensors, Sens. Networks Inf. Process. ISSNIP 2011*, pages 431–436.

Williams, J. and Roslyn, L. A. (2014). Approximate evaluation of marginal association probabilities with belief propagation. *IEEE Trans. Aerosp. Electron. Syst.*, 50(4):2942–2959.

Williams, J. L. (2003). *Gaussian Mixture Reduction for Tracking Multiple Maneuvering Targets in Clutter*. PhD thesis, Air Force Institute of Technology.

Williams, J. L. (2012). Hybrid Poisson and multi-Bernoulli filters. *Inf. Fusion (FUSION), 2012 15th Int. Conf.*, pages 1103–1110.

Williams, J. L. (2015a). An efficient, variational approximation of the best fitting multi-bernoulli filter. *IEEE Trans. Signal Process.*, 63(1):258–273.

Williams, J. L. (2015b). Marginal multi-bernoulli filters: RFS derivation of MHT, JIPDA, and association-based member. *IEEE Trans. Aerosp. Electron. Syst.*, 51(3):1664–1687.

Williams, J. L. and Lau, R. A. (2018). Multiple Scan Data Association by Convex Variational Inference. *IEEE Trans. Signal Process.*, 66(8):2112–2127.