Stochastic models and methods for multi-object tracking

Michele Pace

N^o d'ordre : 4291

INRIA

UNIVERSITÉ DE BORDEAUX

THÈSE

présentée à

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

par **Michele Pace**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : **Mathématiques Appliquées**

****

Méthodes et modèles stochastiques pour le suivi multi-objets

****

Soutenue le 13 juillet 2011 à l'Institut de Mathématiques de Bordeaux

Devant la commission d'examen composée de :

M. Pierre Del Moral	Directeur de Recherche, INRIA	Directeur de thèse
M. François Caron	Chargé de Recherche, INRIA	Co-directeur de thèse
M. Emmanuel Duflos	Professeur, École Centrale de Lille	Rapporteur
Mme Josiane Zerubia	Directrice de Recherche, INRIA	Rapporteur
M. Jean-Charles Noyer	Professeur Université du Littoral Côte d'Opale	Examinateur
M. François Septier	MCF, TELECOM Lille 1	Examinateur
M. Sumeetpal S. Singh	Professeur, University of Cambridge	Examinateur

INRIA Bordeaux - Sud-Ouest et
Institut de Mathématiques de Bordeaux
Université Bordeaux 1
352, cours de la Libération -F 33405
TALENCE cedex

École doctorale de Mathématiques
et Informatique de Bordeaux
U.F.R de Mathématiques Informatique
Bat A33, 1er étage, 352, cours de la
Libération -F 33405 TALENCE cedex

Méthodes et modèles stochastiques pour le suivi multi-objets

Résumé:

La poursuite multi-cibles a pour objet le suivi d'un ensemble de cibles mobiles à partir de données obtenues séquentiellement. Ce problème est particulièrement complexe du fait du nombre inconnu et variable de cibles, de la présence de bruit de mesure, de fausses alarmes, d'incertitude de détection et d'incertitude dans l'association de données. Les filtres PHD (Probability Hypothesis Density) constituent une nouvelle gamme de filtres adaptés à cette problématique. Ces techniques se distinguent des méthodes classiques (MHT, JPDAF, particulaire) par la modélisation de l'ensemble des cibles comme un ensemble fini aléatoire et par l'utilisation des moments de sa densité de probabilité.
Dans la première partie, on s'intéresse principalement à la problématique de l'application des filtres PHD pour le filtrage multi-cibles maritime et aérien dans des scénarios réalistes et à l'étude des propriétés numériques de ces algorithmes. Dans la seconde partie, nous nous intéressons à l'étude théorique des processus de branchement liés aux équations du filtrage multi-cibles avec l'analyse des propriétés de stabilité et le comportement en temps long des semi-groupes d'intensités de branchements spatiaux. Ensuite, nous analysons les propriétés de stabilité exponentielle d'une classe d'équations à valeurs mesures que l'on rencontre dans le filtrage non-linéaire multi-cibles. Cette analyse s'applique notamment aux méthodes de type Monte Carlo séquentielles et aux algorithmes particulaires dans le cadre des filtres de Bernoulli et des filtres PHD.

Mots-clés : Processus de branchements, filtres particulaires, filtrage non-linéaire multi-cibles, systèmes de particules de type champ moyen, semi-groupes de Feynman-Kac, filtre PHD, propriétés de concentration exponentielle, inégalités de contraction fonctionnelles.

Stochastic models and methods for multi-object tracking

Abstract:

The problem of multiple-object tracking consists in the recursive estimation of the state of several targets by using the information coming from an observation process. The objective of this thesis is to study the spatial branching processes and the measure-valued systems arising in multi-object tracking. We focus on a class of filters called Probability Hypothesis Density (PHD) filters by first analyzing their performance on simulated scenarii and then by studying their properties of stability and convergence. The thesis is organized in two parts: the first part overviews the techniques proposed in the literature and introduces the Probability Hypothesis Density filter as a tractable approximation to the full multi-target Bayes filter based on the Random Finite Sets formulation. A series of contributions concerning the numerical implementation of PHD filters are proposed as well as the analysis of their performance on realistic scenarios.
The second part focuses on the theoretical aspects of the PHD recursion in the context of spatial branching processes. We establish the expression of the conditional distribution of a latent Poisson point process given an observation process and propose an alternative derivation of the PHD filter based on this result. Stability properties, long time behavior as well as the uniform convergence of a general class of stochastic filtering algorithms are discussed. Schemes to approximate the measure-valued equations arising in nonlinear multi-target filtering are proposed and studied.

Keywords : Measure-valued equations, non-linear multi-target filtering, Bernoulli filter, Probability Hypothesis Density filter, interacting particle systems, particle filters, Sequential Monte Carlo methods, exponential concentration inequalities, semigroup stability, functional contraction inequalities.

Acknowledgements

First and foremost, I would like to sincerely express my deepest gratitude to my supervisors Prof. Pierre Del Moral and Dr. François Caron for their continued support, guidance, willingness to share their knowledge, and above all for their patience.
Prof. Del Moral with his deep mathematical intuitions and his truly scientific reasoning has been a constant point of reference and an inspiring guide. He is the best supervisor I could have ever asked for.
Dr. Caron is a backbone of the research contained in this dissertation and I'm much indebted for his valuable advice, his supervision, and for all the time he spent reading the manuscript and for his constructive comments and corrections.
I'm very grateful to them and I sincerely hope to continue our collaboration in the future.
I would also like to thank one of the most prolific researcher in the field of multi-object filtering, Prof. Ba Ngu Vo. It has been a genuine pleasure to work and learn from him. The discussions and feedback during the course of my studies were immensely appreciated. I admire him not only as a researcher but also for the sincere and friendly person he is.
I gratefully thank Pr. Josiane Zerubia, Pr. Emmanuel Duflos, Pr. Jean-Charles Noyer, Pr. Sumeetpal S. Singh and Dr. François Septier for having accepted to be members of the defense committee in the midst of all their activities and for their constructive comments and suggestions during the finalization of the manuscript.
My gratitude also goes to Pr. Arnaud Doucet, Pr. Huilong Zhang, Dr. Dan Laneuville; it has been a pleasure to collaborate with you.

Many special thanks to all the friends met during my years at INRIA and at University of Bordeaux for the wonderful time we had together: Damiano, Cédric, Peng, Frank, Adrien, Frédéric (who also pushed me to improve my chess game), Denis, Marco, Jérémie, Johanna, Aurélie, Jade, Mohamed. Eva reserves special mention for the wonderful person she is, for her support and intelligence, her welcoming smile at the end of the day and for the patience she had while I was writing this dissertation; I cannot imagine these last years without her presence.
Words fail me to express my gratitude to my parents for their inseparable support, their sacrifices and for their teaching.

Finally, I convey special acknowledge to INRIA as a whole for the support and the material given to researchers and students and I would like to thank everybody who was important for the successful realization of the this thesis as well as expressing my apology that I could not mention personally one by one.

Contents

List of Algorithms

List of Tables

List of Figures

List of Acronyms

CBMeMBer	Cardinality Balanced Multi-Object Multi-Bernoulli
CPHD	Cardinalized Probability Hypothesis Density (Filter)
EAP	Expected A Posteriori
EKF	Extended Kalman Filter
FFT	Fast Fourier Transform
FFT-PHD	Fast Fourier Transform PHD (filter)
FOV	Field of view of a sensor
GM	Gaussian Mixture
GM-PHD	Gaussian Mixture Probability Hypothesis Density (Filter)
i.i.d	Independent and identically-distributed
IFFT	Inverse Fast Fourier Transform
IGP-PHD	Interacting Group Particle PHD (filter)
IKF-PHD	Interacting Kalman filters PHD (filter)
IMM	Interacting Multiple Model
IPDA	Integrated Probabilistic Data Association
JDA	Joint Probabilistic Data Association
JoM	Joint Multi-Object
FFT	Fast Fourier Transform
KF	Kalman Filter
MaM	Marginal Multi-Object Estimator
MAP	Maximum A Posteriori
MeMBer	Multi-Object Multi-Bernoulli
MCMC	Markov Chain Monte Carlo
MHT	Multiple Hypothesis Tracking
OMAT	Optimal Mass Transfer
OSPA	Optimal SubPattern Assignment
PDA	Probabilistic Data Association
pdf	Probability density function
PHD	Probability Hypothesis Density (Filter)
PGF	Probability Generating Function
PGFl	Probability Generating Functionals
PMHT	Probabilistic Multiple Hypothesis Tracking
RFS	Random Finite Set
SMC	Sequential Monte Carlo
SMC-PHD	Sequential Monte Carlo Probability Hypothesis Density (Filter)
TBD	Track-before-detect
UKF	Unscented Kalman Filter

Notations

Ω	Sample space
$\emptyset$	Empty set
$\sigma(F)$	Sigma-algebra generated by F
$\mathbb{E}$	Expectation operator
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{N}$	Non-negative integers
$\mathbb{N}^+$	Positive integers
$\mathbb{P}$	Probability measure
$p(\cdot)$	Probability density
$\delta_x(\cdot)$	Dirac delta function centered on x
1_A	Indicator function of the set A
$\lvert\cdot\rvert$	Cardinality of a set
E	General measurable space
$\mathcal{M}(E)$	Set of all finite positive measures on a measurable space
$\mathcal{P}(E)$	Set of all probability measures on a measurable space
$\mathcal{B}(E)$	Space of all bounded and measurable real-valued functions
$\mathcal{F}(E)$	Space of finite subsets of E
$\mathcal{C}(E)$	Collection of closed subsets of E
E_s	Target state space
E_o	Measurement space
x_t	Vector at time t
$x_{1:t}$	Sequence of vectors $(x_1, \ldots, x_t)$
X	Random Finite Set
$\mathcal{X}_n$	point process associated to the targets $\mathcal{X}_n \equiv \sum_{i=1}^{N_n} \delta_{X_n^i}$
$\mathcal{N}(\cdot; m, P)$	Gaussian density with mean m and covariance matrix P
$\mu^{\otimes p}$	Product measure of $\mu \in \mathcal{M}(E)$ on the product space E^p
$G : x \in E \mapsto G(x)$	Bounded non-negative potential function
$\Psi_G(\eta) \in \mathcal{P}(E)$	Density $G(x)/\eta(G)$ with respect to a measure η
$(dF)_x$	Set derivative of a F at the point x
$(dF)_{\{x_1,\cdots x_n\}}(S)$	Set derivative at a finite set $X = \{x_1, \cdots x_n\}$
p_X	Probability density of a RFS X
P_X	Probability distribution a RFS X
$\beta_X(S)$	Belief mass function of the RFS X $\beta_X(S) = \mathbb{P}(\{X \subseteq S\})$
$\varsigma_X(S)$	Void probability of the RFS X $\varsigma_X(S) = \mathbb{P}(\{\omega : \lvert X(\omega) \cap S\rvert = 0\})$
$\delta_X(x)$	Sum of Dirac delta functions located at points in the RFS X
$N_X(S)$	Number of points of the RFS X in the set S

$\gamma(x)$	Point process intensity
$\gamma(f)$	Bounded function f integrated over the intensity
$p_{s,t}(\cdot)$	Target survival probability
$p_{d,t}(\cdot)$	Target detection probability
$\beta(M)$	Dobrushin coefficient of a bounded integral operator M
$\|\mu\|_{\mathrm{tv}}$	Total variation norm
$\mathrm{osc}(f)$	Oscillation of a measurable function f
$\mu(f)$	Lebesgue integral of f w.r.t. to the measure μ
f^-	The infimum of a function f
f^+	The supremum of a function f

Chapter 1

Introduction

1.1 Motivation

Many decisions we take every day are based on beliefs concerning the likelihood of uncertain events such as the possibility of rain, the expected time required to complete a task, or the level of traffic in the streets. Other long term decision may involve events such as the outcome of an election, the symptoms of a disease or the future value of a currency. These beliefs are usually expressed in statements such as "I think that...", "chances are that...", and so forth. How do people assess the probability of uncertain events or the value of an uncertain quantity is the subject of cognitive psychology and seminal papers such as [126, 127] have contributed to shed some light onto this difficult question.
Outside psychology, many real-world problems involve the task of estimating unknown quantities from diverse and uncertain observations.

Because uncertainty is so pervasive, rules of choice are necessary.
One of the first estimation problems that were methodologically studied by Laplace, Legendre and Gauss, was the determination of the parameters of planet orbits from a series of observations which were known to be noisy and inaccurate.
Starting from philosophical considerations a series of ideas were gradually developed into mathematical tools. More specifically it was realized that:

- a description of the model and parameters of interest is necessary,
- dealing with inaccuracies in the model and in the observations necessitate a probabilistic reasoning,
- redundant data and observations should reduce the uncertainty and the effects of errors,

- in order to satisfy all the observations in the most accurate way, the residuals (the difference between the observations and the calculated value) should be as small as possible,
- the combination of initial knowledge and subsequent observations generates an iterative procedure.

These observations stand at the base of estimation theory.

In engineering applications the dynamics of the parameter of interest and the uncertainty about the measurements are modelled probabilistically and the estimation problem is generally oriented toward the selection of a point from a continuous space (the best estimate according to some criterion). When the observations arrive sequentially, the new information is used to refine the uncertainty on the parameters at each time step and the process of inference is performed recursively.

Estimation methods have a wide range of applications in different domains such as surveillance systems, control systems, mapping and navigation, signal and image processing, biomedical engineering, military applications and many others. Classical problems concern for example the localization of moving objects (generally called *targets*), the determination of messages in communication networks, the analysis of model parameters for the prediction of the state of production plants, or again the estimation of the volatility of financial instruments.

This thesis focuses on the problem of estimating recursively in time the state of multiple objects given *sets* of observations. In this context, the terms that are commonly used are *tracking*, which refers to the estimation of the state of moving objects based on observations, and *filtering* which is generally used to refer to the elimination of as much noise as possible from the signal in order to obtain a reliable estimate of the state of a dynamical system. As we will see, the problem of *multi-object* tracking, requires all the tools of estimation theory as well as the extensive use of additional statistical techniques to solve issues of data association and measurement validation.

Single-object filtering

To give a concrete example and further introduce the subject we shall consider a very simplified version of an air-traffic control system where a radar reports measurements about the aircrafts in his field of view (FOV) at each time step. Assume at first that only a single aircraft is in the radar's field of view and that perfect weather conditions allow the radar to report only one observation during each scan. In this case the problem of localizing the aircraft is a single-target tracking problem as the objective is to characterize the uncertainty on its state by using the measurement available at each time step. In order to make the inference at least

two models are required: the model describing the dynamics of the aircraft and the model relating the radar measurements to the target's state. Under the assumption that only one target is present and that it generates a single measurement, the problem is relatively simple: the recursive filtering approach consists in using the prior information about the object's dynamics to predict its future state and then updating the uncertainty on the predicted state by using the observation.

The most widely used filtering technique is the ubiquitous Kalman filter [58], a simple and elegant algorithm formulated in 1960 as an optimal recursive Bayesian estimator for a restricted class of linear Gaussian problems. The Kalman filter provides an efficient recursive way to estimate the state of a process by minimizing the variance of the estimation errors. Since the time of its introduction, it has been the subject of extensive research and application, particularly in the area of assisted navigation and in the aerospace industry where its cheap computational requirements made it the *de-facto* standard in years where the computational power was very expensive. However, the conditions necessary for the optimality of the Kalman filter rarely exist in real-life situations where target dynamics are typically nonlinear and the measurement process very complex.

In order to address these limitations without renouncing to the Kalman filter, techniques based on model linearization and Gaussian assumptions [57] have been proposed; when the non-linearities are severe however, these solutions performs poorly.
The main substantial alternative to the Kalman filter consist of a set of sequential Monte Carlo estimation tools, collectively referred to as *particle filters*. Since the seminal paper [43] they have generated a wide interest and they have become a popular method for dynamic estimation problems. The key idea of particle filters is to approximate the probability distribution of the target state by a set of weighted random samples. As the number of samples becomes very large, the estimate obtained approaches the optimal Bayesian estimate. Their computational cost, however, has long been considered as the main disadvantage.

From the purely mathematical point of view, particle filters can be considered as discrete approximation techniques for a flow of measures evolving in time described by a rather general set of equations called Feynman-Kac models [28]. This connection allows for the study of their stability and convergence properties with advanced mathematical techniques developed in the context of the analysis of stochastic processes.

Multi-object filtering

To introduce the multi-object filtering problem, let's return for a moment to our previous example. Imagine that multiple aircrafts are now crossing the FOV of the

radar and that weather conditions have deteriorated such that some of the targets are not properly detected during the scan and that false observations are reported. Can we still perform the tracking of all the targets? To which extent? How can the filter distinguish between a false alarm and a true observation? How should a measurement be associated to a target? When a target can be considered as disappeared and when it is simply undetected? The answer to these questions constitute the core of the multi-target tracking (MTT) problem which has proven to be far more difficult than its single-object counterpart.
The basic principles of MTT were formulated in the mid '50s in early papers by Wax [139] and Sittler [121] but modern developments to extend techniques from single to multiple-object tracking began only in the '70s with papers by Bar-Shalom [3] and Singer [119].
The last decades have witnessed an intensive research interest with an increasing sophistication of the methods proposed, facilitated by the increasing computational capability of modern systems. Despite these attempts however, a standard approach still does not exist and most techniques involve modifications of single-target filtering algorithms.

The fundamental distinction between single and multiple-target tracking is that MTT requires a complex data association phase to discriminate if a measurement has to be considered valid and, if yes, which target has generated it. Many efforts have been directed towards the development of a general method to perform this association in order to treat MTT as a set of independent single-object problems. However, in realistic applications the exhaustive search for the correct association between measurements and targets remains computationally daunting. In addition, as many techniques are still based on linearized models, the effect of inaccuracies introduced by linearization are often summed to errors caused by incorrect associations.
Particle filtering techniques capable of handling nonlinear and non-Gaussian dynamical models are equally problematic due to the large number of particles that are required in realistic applications.

The effects of the combinatorial nature of the so-called measurement-to-track approach to MTT can be experienced in one of the most widely used algorithms for multiobject filtering: the Multiple Hypothesis Tracking filter (MHT) [10]. This algorithm consists in exhaustively searching for all possible associations between tracks and measurements over a number of time steps with the possibility of letting future measurements resolve the uncertainties. Despite the fact that many strategies have been developed over the years to render practical MHT implementations feasible, the combinatorial explosion is unavoidable. Alternative approaches that aim at eliminating the association problem have been proposed, but none of these formulations is general and systematic enough to be adopted as the foundation.

To complicate matters, realistic applications need to combine and exploit different sources of information, and an MTT approach should be general enough to allow for the incorporation of heterogeneous data. The methodologies employed in those cases go under different names, such as ˝*data fusion*˝ or ˝*information fusion*˝. Information fusion has long been more a set of heuristics and ad-hoc solutions before its maturation into a scientifically founded discipline. The reason is obviously not the lack of intuition of researchers, but rather the major problems that had to be solved, such as the *high disparate and ambiguous form that information can have* [80], the fact that *systems are characterized by a varying number of objects of various kinds*, and finally the most crucial obstacle: the *the prohibitive combinatorial complexity* of multi-object, multi-sensor systems.

Random finite sets in multi-object filtering

The last decade saw a practical and theoretical revolution in multi-object filtering thanks to the introduction of the Random Finite Set framework (RFS) [74, 75] which offers a mean to integrate different aspects of data fusion together and seems capable of putting under a single probabilistic umbrella different important aspects of the multi-object problem.

The first systematic treatment of multi-object filtering based on random set theory was proposed by Ronald Mahler with the development of Finite Set Statistics (FISST) [42, 86]. Although the rigorous mathematical foundation for point process theory has been in existence for decades, the theory was, quoting Mahler: ˝*...traditionally formulated with the requirements of mathematicians rather than engineers in mind*˝; Finite Set Statistics on the contrary constitutes an ˝engineering friendly˝ version of the theory of point processes that treats multi-target systems as visualizable images (set of points) maintaining largely the Bayes formalism understood by signal processing engineers.

One of its main goal is to extend seamlessly the formal Bayes modeling to non-traditional multi-object problems by generalizing probability densities and calculus methods so that ideas from statistics and information theory can be extended to random finite sets.
By using the tools of FISST, Mahler developed original and elegant multi-object filtering algorithms; one of them is known as the *Probability Hypothesis Density filter* (PHD filter) [76, 80]. One of the important aspects of the PHD filter is the fact that it replicates to a certain extent the simplicity of the Kalman filter approach to the multi-object, multi-sensor case. The elegance of the Kalman filter, in fact, relies greatly in the way in which the prior and posterior distributions are characterised by a small set of sufficient statistics that are easily propagated in time. When the tracking is generalised to the multi-object, multi-sensor scenario however, no simple analogous implementation seemed evident, reason that made the measurement-to-

track association approach so ubiquitous.

The idea proposed by Mahler and Zajic with the PHD filter is to propagate the first moment of a function that maps a set of targets into a continuous function space: this functional mapping is essential to overcome the fact that the expectation of set-valued random variables is not defined. The function they proposed places Dirac deltas at the target positions and its first moment function is called Probability Hypothesis Density.
Like the mean and variance of the Kalman filter, the PHD is propagated forward via Bayesian predictions and update steps. The recursion is, of course, more complicated than the Kalman filter since an entire function, not just the mean vector and covariance matrix of a Gaussian distribution is being propagated.

Although the conditions used in its mathematical derivation are not usually met in practice, the PHD filter has been proved to be a viable method to perform multi-object filtering thanks to the seamless treatment of fundamental aspects such as the appearance and disappearance of targets and the presence of false measurements and misdetections. Moreover, even if its formulation involves integrals that have no closed form solution in general, numerical approximations are easily implemented and show attractive computational performance. On the other hand, the PHD filter does not identify individual targets as it generally provides a set of points corresponding to the regions of the state space where the highest concentration of objects is expected. In this respect it can be considered as a very efficient filter that eliminates false observations, while further analysis is required to obtain individual target trajectories.
Two main implementations have been proposed: the Sequential Monte Carlo (particle-system) [132, 143] and the Gaussian-mixture implementation [130].

Notable instances of MTT problems where PHD filters have been successfully applied concern the tracking of vehicles on different terrains [118], where reports from human observers were integrated with information from map databases, the tracking of targets by using passive radars that exploit FM radio transmitters [124], group-target tracking [23] and the filtering based on sonar images for the development of self-navigating underwater robots (AUVs) [19, 22]. Computer vision applications include the tracking of faces and people [72], and the tracking of vehicles from video sequences in urban areas [105]; a method for the filtering of feature points trajectories in image sequences is investigated in [49]. Example of applications to radar tracking are described in [64] where the authors use a PHD filter for the localisation of targets in marine environments from 3D-LIDAR measurements and in [97] where 3D naval and aerial scenarios are considered.

Concerning theoretical aspects, works by Clark, Vo, Doucet and Singh in [21, 24, 53] have demonstrated the convergence of the Sequential Monte Carlo approximation and the uniform convergence of the Gaussian mixture PHD filter. Despite

many advances in recent years though, many questions are still open.
This thesis aims at bringing contributions to the theoretical understanding and to the practical application of PHD filtering by showing that various instances of single and multi-object filters belong to a common class of measure-valued equations which can be studied with advanced mathematical tools from the theory of stochastic processes for what concerns convergence properties, asymptotic behaviour and stability.
Additionally, in the more applied part of the dissertation, attempts have been made to address known weakness of existing PHD filter implementations and to increase their accuracy by exploiting ideas coming from the measure theoretic formulation and from industrial problems. Two novel implementations have also been proposed and studied.
Finally, this work aims to bring some contributions to the thesis that PHD filters can be effectively exploited in real-world instances of multi-object filtering and not only on academic examples.

1.2 Organization of the thesis and contributions

This thesis is organised in two parts. The division reflects the choice to present both the numerical aspects related to multi-object tracking with PHD filters as well as the theoretical analysis of the PHD recursion considered as a measure-valued dynamical system. The first part is organised as follows.

Chapter 1 provides an overview of the thesis, outlines the motivation and summarizes the major contributions.

Chapter 2 is intended as a reference for the reader as it introduces the main concepts and formalisms used to define filtering problems and discusses the typographical conventions adopted in the dissertation.

Chapter 3 reviews the most common approaches to the problem of single-object tracking and introduces the random finite set approach to multi-object tracking.

Chapter 4 presents the Probability Hypothesis Density (PHD) filter as a tractable sub-optimal approximation to the full multi-target Bayes filter and describes the *Sequential Monte Carlo* (SMC-PHD) and the *Gaussian Mixture* (GM-PHD) implementations as well as recent developments appeared in the literature.

Chapter 5 concludes the first part with a series of numerical studies on different approximations of the PHD recursion and their application to realistic multi-

tracking problems. The contributions of the chapter can be resumed as follows:

- Section 5.1 compares the performances of different implementations of the PHD filter applied to realistic three-dimensional aerial and naval scenarios provided by the French naval defence provider DCNS.

- Section 5.2 investigates the adoption of a dynamic pruning strategy for the resampling step of the GM-PHD and proposed a target-tracker adapted pruning strategy to mitigate the problem of the track deactivation after repeated misdetections.

- Section 5.3 proposes a novel approach for the resolution of the PHD recursion by using numerical grids.

- Section 5.4 proposes an approximation algorithm for the PHD recursion based on the sampling of the associations between the terms used to approximate the PHD intensity function and the observations at each time step.

The contributions of this chapter have been presented at the *13th International Conference on Information Fusion* [98] and *2010 IEEE Radar Conference* [97] and accepted to the *14th International Conference on Information Fusion* [99].

Theoretical contributions are reported in Part 2.

Chapter 6 considers the problem of estimating a hidden point process and establishes an expression of the conditional distribution of a latent Poisson point process given an observation point process by using a random measure approach combined with reversed Markov kernel techniques. This result has been accepted and and is due to appear as a journal article in *Advances in Applied Probability* (June 2011).

Chapter 7 describes a mean-field and interacting-particle interpretation of a class of spatial branching intensity models arising in multi-object filtering and investigates their difference with respect to traditional Feynman-Kac particle models. The stability properties and the long time behaviour of these distribution flows are studied as well as their asymptotic behaviour. This result has been accepted as a journal article to *SIAM Journal on Control and Optimization* .

Finally, in **Chapter 8** we analyse the exponential stability properties of a class of measure-valued equations arising in nonlinear multi-object filtering and illustrate the results in the context of the Bernoulli and the Probability Hypothesis Density filter. This result has been accepted and is due to appear as a journal article in

Stochastic Analysis and Applications. Conclusions are reported in **Chapter 9**.

Figure 1.2 illustrates the relationship between the chapters. Chapter 2, discussing the conventions and the notations used in the dissertation, should be read first. Part I and part II, can be read independently. The concluding chapter and the annexes are not shown in the picture.

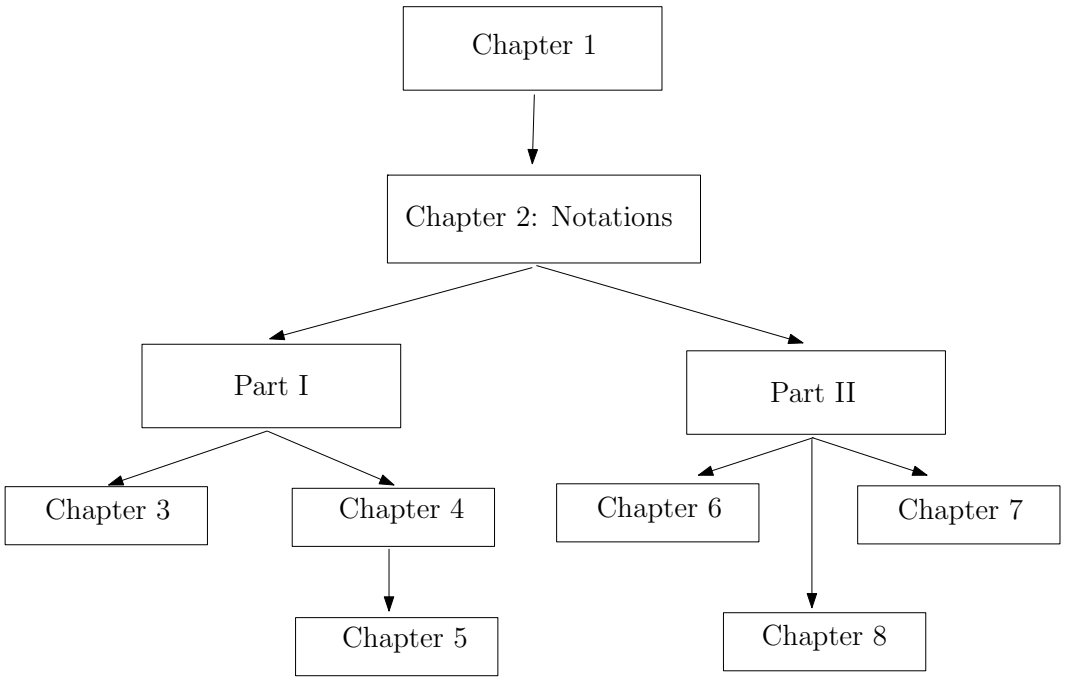


Figure 1.1: Chapter organization

1.3 Publications

The following papers have been published or submitted based on works contained in this dissertation:

1. F. Caron, P. Del Moral, M. Pace, B.-N. Vo
   On the Stability and the Approximation of Branching Distribution Flows, with Applications to Nonlinear Multiple Target Filtering
   To appear in *Stochastic Analysis and Applications* (2011).

2. F. Caron, P. Del Moral, M. Pace, A. Doucet
   Particle approximation of the intensity measures of a spatial branching point process arising in multi-target tracking
   To appear in *SIAM Journal on Control and Optimization* (2011).

3. F. Caron, P. Del Moral, A. Doucet, and M. Pace
   On the Conditional Distributions of Spatial Point Processes
   To appear in *Advances in Applied Probability* (June 2011).

4. M. Pace
   Comparison of PHD Based Filters for the Tracking of 3D Aerial and Naval Scenarios
   in *Proc. 2010 IEEE International Radar Conference, Washington DC, USA.*

5. M. Pace, F. Caron.
   Comparison of implementations of Gaussian Mixture PHD Filters
   in *Proc. 13th International Conference on Information Fusion*, Edinburgh, July 2010.

6. M. Pace, Zhang H.
   Convolution PHD Filter by using the Fast Fourier Transform
   To appear in *Proc. 14th International Conference on Information Fusion*, Chicago, July 2011.

7. F. Caron, P. Del Moral and M. Pace.
   Etude des filtres particulaires PHD
   Rapport de contrat n 2009-DCNS-01 avec DCNS-SIS, 2009.

Chapter 2

Notations and conventions

Applied probability, engineering, financial mathematics, biology and advanced signal processing are just an example of the domains where it is required to obtain some kind of estimates for the values of a dynamic system, given some noisy observations.
In order to facilitate the analysis of a particular aspect of the problem each community tends to use a different set of notations to describe similar problems: in the probabilistic interpretation, for example, the pair signal-observation is generally modelled as a two component Markov chain, while in the engineering literature the dynamical equations of the system and the observations are treated separately. A third set of notations comes from the Bayesian literature. This section reviews the main formalisms used to define filtering problems and establishes the correspondence between them.
Before delving into the details, an overview of the basic notation used in the dissertation is in order:

- A capital letter such as X, Y is generally used to denote random variables or random finite sets. In case of ambiguity it will be clearly specified if we are dealing with random variables or random finite sets.

- A lowercase letter such as x, y denotes the value taken by the corresponding random variable. However, to avoid unhelpfully redundant expressions, the random variable and its realization will be sometimes denoted by a lowercase letter.

- The probability of an event is denoted by the letter $\mathbb{P}$ as for example $\mathbb{P}(X = x)$ or $\mathbb{P}(X \in dx)$.

- Probability mass functions are denoted by $\mathbb{P}(x) := \mathbb{P}(X = x)$.

- Similarly, in case of continuous distributions the notation $\mathbb{P}(dx)$ will be often used instead of $\mathbb{P}(X \in dx)$. If those distributions admit density functions the latter will be denoted by $p(x)$.

- **A note on Bayesian notation** One of the often discussed and criticized issues of Bayesian notation is the overloading of the symbol $p(\cdot)$ for every probability function. For example, the usual expression of the Bayes's rule: $p(x|y) = p(y|x)p(x)/p(y)$ tacitly assumes that the reader knows that each p corresponds to a different function. To disambiguating, it is often written as: $p_{X|Y}(x|y) = p_{Y|X}(y|x)p_X(x)/p_Y(y)$. Clearly when we have dozens of parameters in multivariate densities, this convention gets quickly unreadable; the distinction between discrete and continuous distributions makes things even worse.
  We chose to avoid the use of subscripts in the probability densities and to use the conventional Bayesian notation. In case of ambiguity a note will clarify the nature of the operation discussed. Subscripts, however, will be always used in multi-object densities. In this context for example $p_X(\cdot)$ denotes the probability density of the random finite set X.

- **Markov chain notations** A Markov chain is a sequence of random variables X_t defined in some measurable space E_t, indexed by the parameter $t \in \mathbb{N}$. The Markov property dictates that the future states are independent on the past when the present state is given. Three formulations are commonly used: the first can be considered as a stochastic version of a control system where the Markov chain is defined by an equation of the form:

  $$X_t = F_t(X_{t-1}, U_t)$$

  with a fixed initial condition and an additional, problem dependent, control parameter U_t. The second, more probabilistic way, is to consider the elementary transitions of the chain:

  $$\mathbb{P}(X_t \in dx_t | X_{t-1} = x_{t-1}) = M_t(x_{t-1}, dx_t)$$

  Finally, in the Bayesian literature the elementary transitions are described in term of a probability density function:

  $$\mathbb{P}(X_t \in dx_t | X_{t-1} = x_{t-1}) = p(x_t|x_{t-1})dx_t$$

  The term $p(x_t|x_{t-1})$ represents the density of the Markov transition with respect to a some reference probability measure dx_t.

In the **probabilistic formulation** the filtering model is defined by a two-component Markov chain $(X, Y) = \{(X_t, Y_t);\ \ t \geq 0\}$ taking value in some measurable product

spaces $\{(E_t \times F_t);\ \ t \geq 0\}$ where $t \in \mathbb{N}$ and with initial distribution:

$$\nu_0(d(x_0, y_0)) = g_0(x_0, y_0)\eta_0(dx_0)q_0(dy_0)$$

and Markov transition:

$$T_t((x_{t-1}, y_{t-1}), d(x_t, y_t)) = g_t(x_t, y_t)M_t(x_{t-1}, dx_t)q_t(dy_t)$$

where $g_t : E_t \times F_t \mapsto (0, \infty)$ is a strictly positive function, $q_t \in \mathcal{P}(F_t)$, $\eta_0 \in \mathcal{P}(E_0)$ and M_t is a Markov transition from the space E_{t-1} into E_t. This rather abstract formulation models the evolution of the chain by specifying the joint initial probability $\nu_0(d(x_0, y_0))$ at time $t = 0$, which is the probability that the pair object-observation is in the infinitesimal region $d(x_0, y_0) \in (E_0 \times F_0)$.

A different and common way to define the pair signal/observation is the **engineering formulation**. In this case, X_t denotes a Markov chain whose evolution is described by the dynamical equation:

$$X_t = F_t(X_{t-1}, V_t)$$

where V_t represents a sequence of independent random variables modeling the noise (or the uncertainty) on the evolution process. The observation process is modelled separately by a so-called sensor equation:

$$Y_t = H_t(X_t, W_t)$$

where the sequence of random variables W_t, taking value in an auxiliary measurable space S_t, is independent of X_t and models the noise on the measurement process. The collection of measurable functions $H_t : E_t \times S_t \mapsto F_t$ is chosen so that the laws of $H_t(X_t, W_t)$ and W_t are absolutely continuous with respect to the reference measure $q(dy_t)$ with density $g_t(x_t, \cdot)$:

$$\mathbb{P}(H_t(x_t, W_t) \in dy_t) = g_t(x_t, y_t)q(dy_t) \ \ \forall x_t \in E_t \tag{2.0.1}$$

In the **Bayesian framework** the uncertainty on the value of the hidden process x_t conditional to a series of measurements $y_{1:t}$ is quantified by the conditional probability density $p(x_{0:t}|y_{0:t})$ called *posterior* density. The notation $x_{0:t}$ denotes the sequence of hidden states of the chain from time zero to time t. Similarly $y_{0:t}$ denotes the sequence of observations collected up to time t. The computation, or more often the approximation of this density is done by using the Bayes rule:

$$p(x_{0:t}|y_{0:t}) \propto p(y_{0:t}|x_{0:t})p(x_{0:t}) \tag{2.0.2}$$

where $p(x_{0:t})$ is the *prior* density, modeling the uncertainty on $x_{0:t}$. The general assumptions used in the Bayesian estimation is that the evolution of a target from

time $t-1$ to time t is defined by the density $p(x_t|x_{t-1})$ and that the measurement process can be described by the density $p(y_t|x_t)$. These assumptions imply that the measurements are independent conditionally on the trajectory $x_{0:t}$:

$$p(y_{0:t}|x_{0:t}) = \prod_{i=0}^{t} p(y_i|x_i) \tag{2.0.3}$$

and that the target state at time t depends exclusively on the value of the state at time $t-1$. The conditional distributions of the observations given the hidden states are written as:

$$\begin{aligned} p(y_t|x_t)dy_t &= \mathbb{P}(Y_t \in dy_t|X_t = x_t) \\ &= \mathbb{P}(H_t(x_t, W_t) \in dy_t) \end{aligned}$$

In connection with equation (2.0.1) dy_t stands for $q(dy_t)$ and $p(y_t|x_t)$ represent the likelihood potential function $p(y_t|x_t) = g_t(x_t, y_t)$.
The elementary transitions of the chain X_t are written as:

$$p(x_t|x_{t-1})dx_t = \mathbb{P}(X_t \in dx_t|X_{t-1} = x_{t-1}) \tag{2.0.4}$$

and this leads to the following equivalent expressions:

$$\mathbb{P}((X_0, \ldots, X_t) \in d(x_0, \ldots, x_t)) = p(x_0)p(x_1|x_0) \ldots p(x_t|x_{t-1})dx_0 \ldots dx_t$$

and

$$\mathbb{P}((Y_0, \ldots, Y_n) \in d(y_0, \ldots, y_t)|X_{0:t} = x_{0:t}) = p(y_0|x_0) \ldots p(y_t|x_t)dy_0 \ldots dy_t$$

The conditional distributions of the signal given the observations can be expressed in yet another way in terms of Feynman-Kac formulae. These expressions are based on a change of probability measures on path-space, according to a given potential function and are general enough to describe a great variety of phenomena. Feynman-Kac distributions and their particle approximation play a major role in the theory of non-linear filtering: Monte Carlo methods, for instance, can be interpreted as stochastic numerical approximations of the flow of measures defined by these formulae.
The most comprehensive treatise about the structure and the properties of Feynman-Kac formulae is [28], while an introductory coverage with examples can be found in [29].
As this thesis deals with both applied and theoretical aspects of multi-object filtering, we have chosen to avoid the choice of a unique notation and we have adopted the conventional Bayesian-like notation in the first part and the measure theoretic

notation in the second part.
In order to better clarify the connection between the Bayesian formalism and the more abstract formulation, this chapter is concluded with an overview of the Feynman-Kac distributions in path space and discrete time. The relevant concepts will be denoted as follows: the set of all finite positive measures on some measurable space $(E, \mathcal{E})$ are denoted by $\mathcal{M}(E)$, the subset of all probability measures by $\mathcal{P}(E)$ and the Banach space of all bounded and measurable functions f equipped with the uniform norm $||f||$ are denoted by $\mathcal{B}(E)$.
For conciseness, we will often use the notation:

$$\mu(f) = \int f(x)\ \mu(dx)$$

For measurable subsets $A \in \mathcal{E}$, we will sometimes slightly abuse notation and write $\mu(A)$ instead of $\mu(1_A)$. The Dirac measure at $a \in E$ will be denoted by $\delta_a(\cdot)$; in addition we will use:

$$\delta_a(f) = f(a)$$

$$\delta_a(A) = 1_A(a)$$

In the context of particle filtering the notation $\delta_{x^i}(x)$ denotes the Dirac delta conceptualized as an idealized point mass located at $x^i \in E$. This abuse of notation is used as an abbreviation of $\delta(x - x^i)$. When necessary, the state spaces will be augmented with additional states, called *cemetery states* and denoted by c, c^* or c'. The functions $f \in \mathcal{B}(E)$ are extended to the augmented state by setting $f(c) = f(c') = f(c^*) = 0$.

A bounded positive integral operator Q from a measurable space E_1 into a measurable measurable space E_2 is an operator $f \mapsto Q(f)$ from $\mathcal{B}(E_2)$ into $\mathcal{B}(E_1)$ such that the functions:

$$x \mapsto Q(f)(x) = \int_{E_2} Q(x, dy) f(y)$$

are measurable and bounded for some measure $Q(x, .) \in \mathcal{M}(E_2)$. These operators induce a dual operator $\mu \to \mu Q$ from $\mathcal{M}(E_1)$ into $\mathcal{M}(E_2)$ defined by $(\mu Q)(f) = \mu(Q(f))$.

Let $G : x \in E \mapsto G(x) \in (0, \infty)$ be a bounded positive potential function, we refer to the following change of probability measures with the term Boltzmann-Gibbs transformation:

$$\Psi_G\ :\ \eta \in \mathcal{M}(E) \mapsto \Psi_G(\eta) \in \mathcal{P}(E) \quad \text{with} \quad \Psi_G(\eta)(dx) = \frac{1}{\eta(G)}\, G(x)\, \eta(dx) \quad (2.0.5)$$

provided $\eta(G) > 0$ and recall that $\Psi_G(\eta)$ can be expressed in terms of a Markov transport equation

$$\Psi_G(\eta) = \eta S_\eta \tag{2.0.6}$$

for some selection type transition $S_\eta(x, dy)$.

2.1 Feynman-Kac representation

In discrete time, Feynman-Kac path measures are traditionally defined by the following equation [28]:

$$\mathbb{Q}_t(d(x_0, \ldots, x_t)) = \frac{1}{\mathcal{Z}_t} \left\{ \prod_{p=0}^{t-1} G_p(x_p) \right\} \mathbb{P}_t(d(x_0, \ldots, x_t)) \tag{2.1.1}$$

where the measure $\mathbb{P}_t$ represents the probability measure of the path sequence $(X_0, \ldots, X_t)$ of a Markov chain taking values in a measurable space E and the functions G_t are non-negative, measurable functions such that the normalizing constant are well-defined. They can be seen as the "potential" of the states where the Markov chain transitates. The normalization constant for each time step is:

$$\mathcal{Z}_t = \int_{E^{t+1}} \prod_{p=0}^{t-1} G_p(x_p) \mathbb{P}_t(d(x_0, \ldots, x_t)) \qquad \in (0, \infty) \tag{2.1.2}$$

(**Note**: *Although the subscript t in the notation of* $\mathbb{P}_t$ *and* $\mathbb{Q}_t$ *is not strictly necessary, we adhere to the usual notation and use it to stress the fact that the probability measures are defined on the paths from time 0 to time t*).

The time marginals $\eta_t, t \in \mathbb{N}$ of these path distributions are defined as:

$$\gamma_t(f) = \mathbb{E}\left(f(X_t) \prod_{p=0}^{t-1} G_p(X_p) \right) \tag{2.1.3}$$

$$\eta_t(f) = \frac{\gamma_t(f)}{\gamma_t(1)} \tag{2.1.4}$$

for any bounded measurable function $f(\cdot)$ on E. The continuous-time version is defined in [28]. In the context of non-linear filtering this abstract formulation represents the solution of the non-linear, measure-valued system governing the evolution of the conditional distributions of the target state given the observations.
The relationship between the Feynman-Kac formulae and the problem of filtering may not be immediately obvious, especially considering their rather abstract representation. However, it can be intuitively understood by considering the recursive

particles approximation of the flow of measures $\eta_t(\cdot)$ and by thinking to the likelihoods with respect to the observations as the potential functions of each state where the Markov chains modeling the targets transitate. In this context particles exploring regions with a low potential tend to disappear while particles in regions associated to an high potential tend to reproduce; at each time step the occupation measure of the surviving particles provides a discrete representation of the posterior distribution density. Figure 2.1 illustrates the concept applied to the estimation of a single, mono-dimensional trajectory.

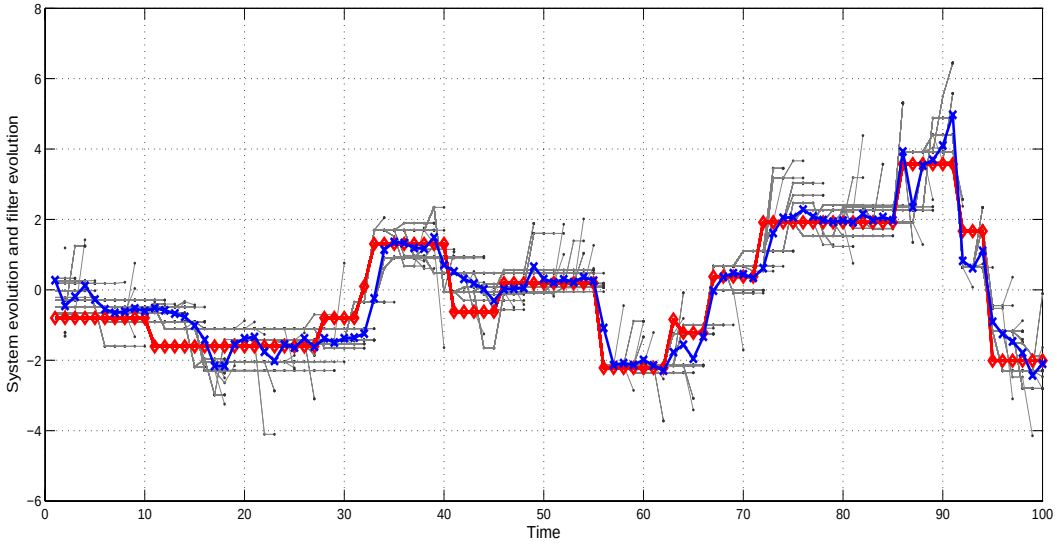


Figure 2.1: Branching-type evolution of a particle system used to estimate the trajectory of a jump Markov model. The occupation measure of the genealogical tree of the particles converges, as the population size tends to infinity, to the conditional distribution of the states of the signal given the observations. The particles and their ancestral lines are shown in gray and marked with $(\cdot)$, whereas the hidden signal is marked with $(\Diamond)$ and the estimated trajectory obtained with a population of 30 particles is marked with $(\times)$. At each time step the set of particles provides an approximation of the conditional distribution of the target state given the sequence of observation from time 1 to time t.

The Feynman-Kac representations for the single-target filtering equations (one-step predictor and optimal filter) is as follow. Consider a fixed sequence of observations $Y_{0:t} = y_{0:t}$ and let G_t be the non-homogeneous functions on E defined for every $x_t \in E$ by:

$$G_t(x_t) = g_t(x_t, y_t) \tag{2.1.5}$$

With this notation the conditional distribution of the path $(X_0, \ldots, X_t)$ (under the standard assumption of independence of the observations) given the observations $y_{0:t}$ is given by the path Feynman-Kac measure:

$$\begin{aligned}\widehat{\mathbb{Q}}_t(d(x_0, \ldots, x_t)) &= \mathbb{P}(d(x_0, \ldots, x_t)|Y_{0:t} = y_{0:t}) \\ &= \frac{1}{\widehat{\mathcal{Z}}_t}\left\{\prod_{p=0}^{t} G_p(x_p)\right\}[\eta_0(dx_0)M_1(x_0, dx_1)\ldots M_t(x_{t-1}, dx_t)]\end{aligned}$$

with the normalizing constants:

$$\widehat{\mathcal{Z}}_t = \int_{E^{t+1}} \left\{\prod_{p=0}^{t} G_p(x_p)\right\}[\eta_0(dx_0)M_1(x_0, dx_1)\ldots M_t(x_{t-1}, dx_t)] \tag{2.1.6}$$

where $M_t(x_{t-1}, dx_t)$ is the Markov chain describing the evolution of the target and η_0 its distribution at time $t = 0$:

$$\begin{aligned}\eta_0(dx_0) &= \mathbb{P}(X_0 \in dx_0) \\ M_t(x_{t-1}, dx_t) &= \mathbb{P}(X_t \in dx_t|X_{t-1} = x_{t-1})\end{aligned}$$

In other words these formulae express the conditional distributions of the path sequence $(X_0, \ldots, X_t)$ as the distribution of the signal paths weighted by the product of the likelihood functions from the time 0 up to time t.
Due to the choice of the potential functions (2.1.5), for any bounded integrable function f the one step predictor and optimal filter can be written as:

$$\eta_t(f) = \mathbb{E}(f(X_t)|Y_{0:t-1} = y_{0:t-1}) \tag{2.1.7}$$

$$\widehat{\eta}_t(f) = \mathbb{E}(f(X_t)|Y_{0:t} = y_{0:t}) \tag{2.1.8}$$

with the following functional representation:

$$\eta_t(f) = \frac{\gamma_t(f)}{\gamma_t(1)} \quad \text{and} \quad \widehat{\eta}_t(f) = \frac{\widehat{\gamma}_t(f)}{\widehat{\gamma}_t(1)} \tag{2.1.9}$$

with:

$$\gamma_t(f) = \mathbb{E}(f(X_t)\prod_{p=0}^{t-1} G_p(X_p)) \quad \text{and} \quad \widehat{\gamma}_t(f) = \gamma_t(G_t f) \tag{2.1.10}$$

This functional representation shows that the filtering equations belong to the same class of Feynman-Kac distributions models. The relationship can be clarified by writing:

$$\eta_t = \text{Law}(X_t|Y_{0:t-1} = y_{0:t-1}) \tag{2.1.11}$$

$$\widehat{\eta}_t = \mathrm{Law}(X_t | Y_{0:t} = y_{0:t}) \tag{2.1.12}$$

The normalizing constant $\widehat{\mathcal{Z}}_t$ coincide with the quantities $\widehat{\gamma}_t(1) = \gamma_t(G_t)$ and they can be expressed as:

$$\widehat{\mathcal{Z}}_t = \widehat{\gamma}_t(1) = \prod_{p=0}^{t} \eta_p(G_p) \tag{2.1.13}$$

Chapters 7 and 8 will consider the extension of these models to the multi-object filtering problem.

Final remarks about notations and conventions

The objective of this section was to provide an overview of the different notations and conventions commonly used to define non-linear filtering problems.
The generality of interacting particle methods and their efficiency in solving complex filtering problems and in approximating a large class of measure-valued processes has made them very popular among different scientific and engineering communities, the result is a lack of a uniform terminology and notation. Of course, this is not necessarily a negative fact, as the different conventions have been established in order to simplify the work and the research in the respective domains. In the first part of the thesis, for the definition of the models and the presentation of the applied results we find more convenient to use the Bayesian notation, while the theoretical aspects studied in the second part are better expressed with a measure-theoretic, Feynman-Kac notation.
The definition of the mathematical objects, however, will be done progressively throughout the thesis, especially in the second part where details and definitions will be purposely repeated at the beginning of each chapter for the sake of clarity. Table 2.1 resumes the correspondence between the notations used in Bayesian and Feynman-Kac modeling.

	Bayesian modeling	**Feynman-Kac models**
State space "quality" measure	*Likelihood* $p(y_t\|x_t)$	*Positive potential function* $G_t(x_t), g_t(x_t, y_t)$
Space exploration	*Markov transition density* $p(x_t\|x_{t-1})$	*Markov kernel* $M_t(x_{t-1}, dx_t)$
Normalizing constant	*Marginal likelihood* $p(y_{0:t})$	*Normalizing constant* $\widehat{\mathcal{Z}}_t$
Full posterior	$p(x_{0:t}\|y_{0:t})$ (density)	$\widehat{\mathbb{Q}}_t(d(x_{0:t}))$ (prob. measure)
Full predictor	$p(x_{0:t}\|y_{0:t-1})$ (density)	$\mathbb{Q}_t(d(x_{0:t}))$ (prob. measure)
Marginal predictive density	*predictive density* $p(x_t\|y_{0:t-1})$	η_t
Marginal posterior density	*filtering density* $p(x_t\|y_{0:t})$	$\widehat{\eta}_t$
Predictor over a function	$\int f(x_t)p(x_t\|y_{0:t-1})dx_t$	$\eta_t(f)$
Corrector over a function	$\int f(x_t)p(x_t\|y_{0:t})dx_t$	$\widehat{\eta}_t(f)$

Table 2.1: Correspondence between terms and notations used in Bayesian and Feynman-Kac modeling.

Part I

Stochastic Models and Algorithms

Chapter 3

Background: single object tracking

3.1 Introduction

Object tracking deals with the recursive estimation of the state of one or more objects by using the information coming from an observation process. The objects are generally referred to as *targets*, the region of the space under observation is generally called *surveillance zone* and the observation process is made up of a series of *sensors*. As the terminology suggests, the problem has been originally formulated in military terms and studied in the context of military and defence applications, but it has rapidly found applications in many different domains such as security systems, biology, computer vision, imagery, etc. In this thesis we will use the usual terminology, by referring to the tracked objects mainly with the term "targets", without necessarily implying military targets. For example, targets may be aircrafts or ships (as in aerial or naval monitoring systems) but they may be as well cells, persons, animals, pixels or cell phones. The observation process may consist of data coming from radars, sonars, cameras, satellites, microphones, etc.
In order to characterize the uncertainty on the dynamics of the target, its state is modelled as a random variable (in single-object filtering) or as a random finite set (in RFS multi-object filtering). In the general case the target state only includes kinematic characteristics such as position, velocity and acceleration but it may as well consist of a set of discrete attributes (type of target, level of danger, activity state etc.).
The information coming from the sensors is generally imperfect: targets may be undetected during one or more time steps, the observations are noisy and spurious observations generated by false alarms have to be taken into account.

This chapter introduces the main concepts and methods for single-object filtering in the Bayesian framework and presents the classical tracking algorithms such as the Kalman filter, the Extended and Unscented Kalman filter and the Particle Filter. The concepts and the approaches for multi-object tracking will be introduced and discussed in Chapter 4.

3.2 Bayesian Estimation

The Bayes filter is at the core of the problem of recursive estimation where the objective is to obtain a characterization of the uncertainty on the value of a hidden parameter given the observation of an experimental outcome. Usually, the experiment is a measurement of a physical phenomenon, and the parameter of interest a physical quantity. When the parameters have dynamical properties that change over time, or when new measurements become available, the estimation process is performed recursively in order to incorporate the new information or to account for the modification of the underlying parameters.
Random noises or imprecisions in the measurement process are taken into account by modeling the conditional observations as a random vector with probability density function $p_{Y|X}(y|x)$. The prior density $p_X(x)$ models everything known, and unknown, about underlying stochastic process before the observation of the experimental outcome. The densities are considered w.r.t some underlying reference measure (e.g. Lebesgue measure).
We remind that we will use the Bayesian notation described in Chapter 2 and avoid the use of subscripts $Y|X$, X and Y.
The probability density of the parameter x after the observation of the experimental outcome is obtained by means of Bayes' law:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{3.2.1}$$

In single-object filtering, this posterior probability density function is used to infer the state of the hidden target.
The Bayes recursion as well as the Kalman filter and its common implementations are introduced in the next sections. An in-depth treatise of the foundations can be found in [37, 108].

3.2.1 Single-object tracking

The simplest and probably most studied case of object tracking is the problem of single-target tracking with no false alarms and no misdetection. In this case the Bayesian approach provides the framework to incorporate at each time step the information coming from the sensor into the target distribution.

This section focuses on the discrete-time model and adopts the usual signal-processing notation: the pair signal-observation are modelled by a Markov chain and by a sensor equation respectively and the object of interest is the conditional probability of the target's state given the observations.
The target state at time t is modelled as a n_x dimensional random vector x_t taking value in a state space E_s. Its evolution is described by a stochastic model which specifies the transition from the state x_t at time t to the new state vector x_{t+1} at time $t+1$. The target dynamics is modelled by:

$$x_t = \phi_t(x_{t-1}, v_{t-1}) \tag{3.2.2}$$

where the nonlinear function $\phi_t(\cdot)$ specifies the transformation of any given state x_{t-1} at time $t-1$ and v_{t-1} denotes the system noise or the uncertainty on the target's dynamical model. In a similar way the observations are modeled as random vectors y_t taking value in an observation space E_o and generated by the measurement model:

$$y_t = h_t(x_t, w_t) \tag{3.2.3}$$

where w_t represents the uncertainty of the measurement process described by the nonlinear function $h_t(\cdot)$. The dimension of the observation vector n_y is generally lower than n_x. The evolution of the target state and the measurement vector are alternatively described by their probability densities:

$$f_{t|t-1}(x_t|x_{t-1}) \tag{3.2.4}$$

$$g_t(y_t|x_t) \tag{3.2.5}$$

called *Markov transition density* and *likelihood function* respectively.
In other terms, the likelihood function of equation (3.2.5) represents the probability density that the target with state vector x_t generates the observation y_t, while the Markov transition density of equation (3.2.4) is the probability density that a target with state vector x_{t-1} at time $t-1$ moves to the state x_t at time t. Figure 3.1 shows a schematic representation of the model.

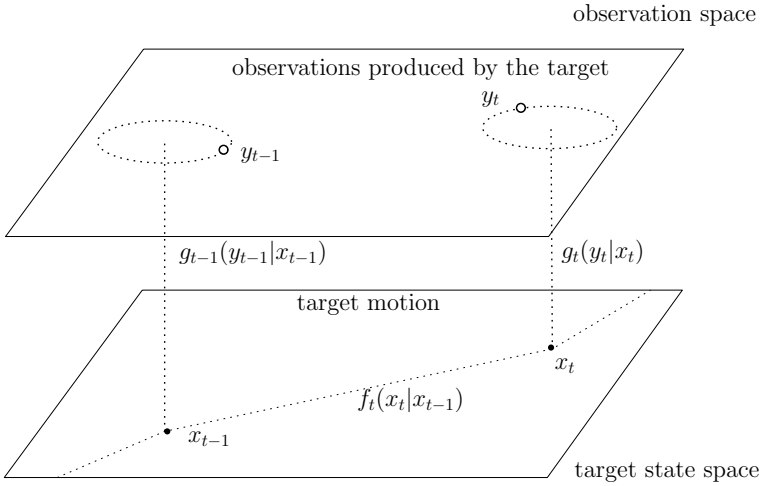


Figure 3.1: Single object tracking schematic representation.

As introduced in Chapter 2, at each time step the measurement y_t is supposed conditionally independent given the target state vector x_t. The probability density $p(y_{0:t}|x_{0:t})$ can thus be expressed as:

$$p(y_{0:t}|x_{0:t}) = g_t(y_t|x_t)g_{t-1}(y_{t-1}|x_{t-1}) \cdots g_1(y_0|x_0) \tag{3.2.6}$$

As introduced in chapter 1, the entity of interest is the *posterior density* or *filtering density* $p(x_{0:t}|y_{0:t})$ or, more often its marginal $p(x_t|y_{0:t})$. Beside filtering, there exists several inference problems that involve computing the posterior distribution of a collection of state variables conditional on a series of observations:

- fixed lag smoothing, where the entity of interest is: $p(x_{t-l}|y_{0:t})$, for $0 \leq l \leq t-1$
- fixed interval smoothing, where the entity of interest is: $p(x_{l:k}|y_{0:t})$ for $1 \leq l < k \leq t$
- prediction, where the entity of interest is: $p(x_{l:k}|y_{0:t})$ for $k > t$ and $1 \leq l \leq k$.

The first two problems reduce to marginalisation of the full distribution $p(x_{0:t}|y_{0:t})$, whereas the third reduces to marginalisation of:

$$p(x_{0:k}|y_{0:t}) = p(x_{0:t}|y_{0:t}) \prod_{i=t+1}^{k} f_{t|t-1}(x_i|x_{i-1}) \tag{3.2.7}$$

The posterior densities $p(x_{t-l}|y_{0:t})$ and $p(x_{l:k}|y_{0:t})$ are called *posterior smoothing densities*.

The conditional posterior pdf $p(x_t|y_{0:t})$ completely defines the uncertainty on the target state, hence it can be used to infer the state of the target; two commonly used estimators for x_t are the maximum a posteriori (MAP):

$$\hat{x}_t^{\text{MAP}} = \arg\sup_{x_t} p(x_t|y_{0:t}) \tag{3.2.8}$$

and the expected a posteriori (EAP):

$$\hat{x}_t^{\text{EAP}} = \int x_t p(x_t|y_{0:t}) dx_t \tag{3.2.9}$$

3.2.2 The Bayes Filter

The Bayes filter is a recursion consisting in two steps, commonly called *prediction* or *propagation* and *update*. At each time step the model of the target dynamics is used to propagate in time the current posterior density via the Chapman-Kolgomorov equation:

$$p(x_t|y_{0:t-1}) = \int f_{t|t-1}(x_t|x_{t-1}) p(x_{t-1}|y_{0:t-1}) dx_{t-1} \tag{3.2.10}$$

while the update step consists in the application of the Bayes rule to incorporate the information coming from the observation at time t into the propagated density:

$$p(x_t|y_{0:t}) = \frac{g_t(y_t|x_t) p(x_t|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{3.2.11}$$

The Bayes filter allows the construction of the exact posterior pdf recursively in time. However, the integrations in equations (3.2.10) and (3.2.11) are intractable in practice, except in very limited special cases. In real world applications, for example, it is rarely possible to adopt the analytical solution, hence approximate solutions based on space discretization techniques and numerical integrations have been proposed. However, due to the complexity of the problem, these approaches are useful only when the dimension of the state and observation spaces are relatively low. The next sections outline the closed form solution (Kalman Filter) as well as the Particle Filter approximation to equations (3.2.10) and (3.2.11).

3.2.3 The Kalman Filter

The Kalman filter [58] constitutes the closed form solution to the Bayes recursion under a set of specific assumptions on the target dynamics and observation model. More specifically, the Kalman filter assumes that the target dynamics and the observation process are linear transformations and that the noises are independent zero-mean Gaussian noises:

$$x_t = F_{t-1} x_{t-1} + C_{t-1} u_{t-1} + v_{t-1} \tag{3.2.12}$$

$$y_t = H_t x_t + w_t \tag{3.2.13}$$

where the matrices F_{t-1} and H_t are $n_x \times n_x$ and $n_y \times n_x$ respectively (n_x and n_y being the dimension of the target state and observation vectors), v_{t-1}, w_t are independent zero-mean Gaussian noises with covariance matrices denoted by Q_{t-1} and R_t respectively and C_{t-1} is the control-input model which is applied to the control vector u_{t-1}. Under these assumptions the transition density and observation likelihood can be written as:

$$f_{t|t-1}(x_t|x_{t-1}) = \mathcal{N}(x_t; F_{t-1}x_{t-1} + C_{t-1}u_{t-1}, Q_{t-1}) \tag{3.2.14}$$
$$g_t(y_t|x_t) = \mathcal{N}(y_t; H_t x_t, R_t) \tag{3.2.15}$$

where $\mathcal{N}(x; m, P)$ denotes the Gaussian pdf with mean m and covariance matrix P, evaluated at x. Under linear and Gaussian assumptions, at each time step the posterior density is Gaussian. Assume that at time $t-1$ it is:

$$p(x_{t-1}|y_{0:t-1}) = \mathcal{N}(x_{t-1}; m_{t-1}, P_{t-1}) \tag{3.2.16}$$

The predicted density at time t is Gaussian with mean $m_{t|t-1}$ and covariance matrix $P_{t|t-1}$:

$$p(x_t|y_{0:t-1}) = \mathcal{N}(x_t; m_{t|t-1}, P_{t|t-1}) \tag{3.2.17}$$

with

$$m_{t|t-1} = F_{t-1}x_{t-1} + C_{t-1}u_{t-1} \tag{3.2.18}$$
$$P_{t|t-1} = F_{t-1}P_{t-1}F_{t-1}^T + Q_{t-1} \tag{3.2.19}$$

The updated density, after the arrival of the observation y_t is also a Gaussian:

$$p(x_t|y_{0:t}) = \mathcal{N}(x_t; m_t, P_t) \tag{3.2.20}$$

with

$$S_t = R_t + H_t P_{t|t-1} H_t^T \tag{3.2.21}$$
$$K_t = P_{t|t-1} H_t^T S_t^{-1} \tag{3.2.22}$$
$$m_t = m_{t|t-1} + K_t(y_t - H_t m_{t|t-1}) \tag{3.2.23}$$
$$P_t = [I - K_t H_t] P_{t|t-1} \tag{3.2.24}$$

The matrices K_t and S_t are the *Kalman gain* and *innovation covariance* respectively. The residual $y_t - H_t x_{t|t-1}$ is referred to as the *innovation*. When the assumptions on the linearity and gaussianity of the model do not hold it is not possible to use the Kalman filter directly. However there exists approximations based on local linearisations of the target and observation models (Extended Kalman Filter) [52] or based on deterministic methods to propagate the first and second moments of the predicted and updated Gaussians (Unscented Kalman Filter)[57, 55].

3.2.4 The Kalman Smoother

So far, only the filtering density $p(x_t|y_{0:t})$ taking into account the observations $y_{0:t}$ has been considered. By incorporating the *future* observations we can obtain a more refined state estimates. With the term future observations we refer to the applications where the filtering is done off-line, once all the observations have been received. In this case the posterior smoothing density $p(x_t|y_{0:T})$ with $T > t$ can be obtained as follows [1]:

$$p(x_t|y_{0:T}) = \int p(x_t, x_{t+1}|y_{0:T})dx_{t+1} \tag{3.2.25}$$

$$= \int p(x_{t+1}|y_{0:T})p(x_t|x_{t+1}, y_{0:T})dx_{t+1} \tag{3.2.26}$$

$$= p(x_t|y_{0:t}) \int \frac{f_{t|t-1}(x_t|x_{t-1})p(x_{t+1}|y_{0:T})}{p(x_{t+1}|y_{0:T})}dx_{t+1} \tag{3.2.27}$$

The smoothing density $p(x_t|y_{0:T})$ can be calculated by a forward-backward algorithm which computes the mean and the covariance matrices of the filtering density at each time step and then, back in time, incorporates the information of future observations. The update steps, for $t = T-1, \ldots, 1$ are done with the following equations:

$$p(x_{t+1}|y_{0:T}) = \mathcal{N}(x_{t+1}; m_{t+1|T}, P_{t+1|T}) \tag{3.2.28}$$

$$p(x_{t+1}|y_{0:t}) = \mathcal{N}(x_{t+1}; m_{t+1|t}, P_{t+1|t}) \tag{3.2.29}$$

$$p(x_t|y_{0:t}) = \mathcal{N}(x_t; m_t, P_t) \tag{3.2.30}$$

$$p(x_t|y_{0:T}) = \mathcal{N}(x_t; m_{t|T}, P_{t|T}) \tag{3.2.31}$$

and

$$m_{t|T} = m_{t|t} + L_t(m_{t+1|T} - m_{t+1|t}) \tag{3.2.32}$$

$$P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T \tag{3.2.33}$$

$$L_t = P_{t|t}F_t^T P_{t+1|t}^{-1} \tag{3.2.34}$$

Further details can be found in [12, 6].

3.2.5 The Extended Kalman Filter

When the target dynamics and the observation process are not linear, the process is modelled by the general equations (3.2.2),(3.2.3) and the Kalman filter in not directly applicable. In this case the Extended Kalman filter, which consists in linearising the model about the current mean and variance before applying the Kalman filter equations, can be used. When the non-linearities in the model are

relatively weak, in fact, it is possible to approximate the predicted and posterior densities by Gaussian densities:

$$p(x_{t-1}|y_{0:t-1}) \approx \mathcal{N}(x_{t-1}; m_{t-1}, P_{t-1}) \tag{3.2.35}$$

$$p(x_t|y_{0:t-1}) \approx \mathcal{N}(x_t; m_{t|t-1}, P_{t|t-1}) \tag{3.2.36}$$

$$p(x_t|y_{0:t}) \approx \mathcal{N}(x_t; m_t, P_t) \tag{3.2.37}$$

where:

$$m_{t|t-1} = \phi_t(x_{t-1}, 0) \tag{3.2.38}$$

$$P_{t|t-1} = G_{t-1}Q_{t-1}G_{t-1}^T + F_{t-1}P_{t-1}F_{t-1}^T \tag{3.2.39}$$

$$m_t = m_{t|t-1} + K_t(y_t - h(m_{t|t-1}, 0)) \tag{3.2.40}$$

$$P_t = [I - K_tH_t]P_{t|t-1} \tag{3.2.41}$$

$$K_t = P_{t|t-1}H_t^T S_t^{-1} \tag{3.2.42}$$

$$S_t = U_tR_tU_t^T + H_tP_{t|t-1}H_t^T \tag{3.2.43}$$

where the F_{t-1}, G_{t-1}, H_t and U_t are the local linearisations of the functions ϕ_{t-1} and h_t as follows:

$$F_{t-1} = \left.\frac{\partial\phi_t(x,0)}{\partial x}\right|_{x=m_{t-1}} \quad , \quad G_{t-1} = \left.\frac{\partial\phi_t(m_{t-1},\nu)}{\partial\nu}\right|_{\nu=0}$$

$$H_t = \left.\frac{\partial h_t(x,0)}{\partial x}\right|_{x=m_{t|t-1}} \quad , \quad U_t = \left.\frac{\partial h_t(m_{t|t-1},\epsilon)}{\partial\epsilon}\right|_{\epsilon=0}$$

When the system and observation models have strong non-linearities, the higher order terms of the Taylor expansion become significant and they cannot be safely ignored. In these cases the performance of the Extended Kalman filter becomes very poor.

3.2.6 The Unscented Kalman Filter

The Unscented Kalman Filter has been proposed by Julier and Uhlmann [57, 54] in order to partially mitigate the problems of the Extended Kalman filter. The *Unscented Transform* is based on the intuition that it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation. The approach consists in choosing a set of points (sigma points) so that their mean and covariance are m_t and P_t. The nonlinear function is applied to each point, yielding to a set of transformed points. The statistics of the transformed points can then be calculated to obtain an estimate of the mean and covariance matrix. At time $t-1$, the posterior density is assumed to be Gaussian:

$$p(x_{t-1}|y_{0:t-1}) \approx \mathcal{N}(x_{t-1}; m_{t-1}, P_{t-1})$$

the procedure starts by using an augmented mean and covariance:

$$\mu_t = [m_{t-1}\ 0^T\ 0^T\ ]^T$$

$$C_t = diag(P_{t-1}, Q_{t-1}, R_t)$$

By using the Unscented Transform a set of $L = 2n_U + 1$ weighted points are generated. The points are commonly called *sigma points* and denoted by $\{\sigma_t^l\}$. Denote by $(\sqrt{P})_l$ the the l-th row of the matrix square root of P; the sigma points and their weights $\{w_t^l\}$ are obtained by the formulas:

$$\begin{array}{ll} \sigma_t^0 = \mu_t & w_t^0 = \frac{\kappa_U}{n_U+\kappa_U} \\ \sigma_t^l = \mu_t + \left(\sqrt{(n_U+\kappa_U)C_t}\right)_l & w^l = \frac{1}{2(n_U+\kappa_U)} \quad l = 1,\ldots,n_U \\ \sigma_t^l = \mu_t - \left(\sqrt{(n_U+\kappa_U)C_t}\right)_l & w^l = \frac{1}{2(n_U+\kappa_U)} \quad l = n_U+1,\ldots,L \end{array}$$

n_U is the dimension of the augmented state ($n_U = n_x + n_v + n_w$) and κ_U is a scaling parameter such that $n_U + \kappa_U \neq 0$. Then the sigma points are partitioned into:

$$\sigma_t^l = [(x_{t-1}^l)^T, (v_{t-1}^l)^T, (w_t^l)^T]^T$$

Figure 3.2 provides an illustration of the sigma points in a simple case.

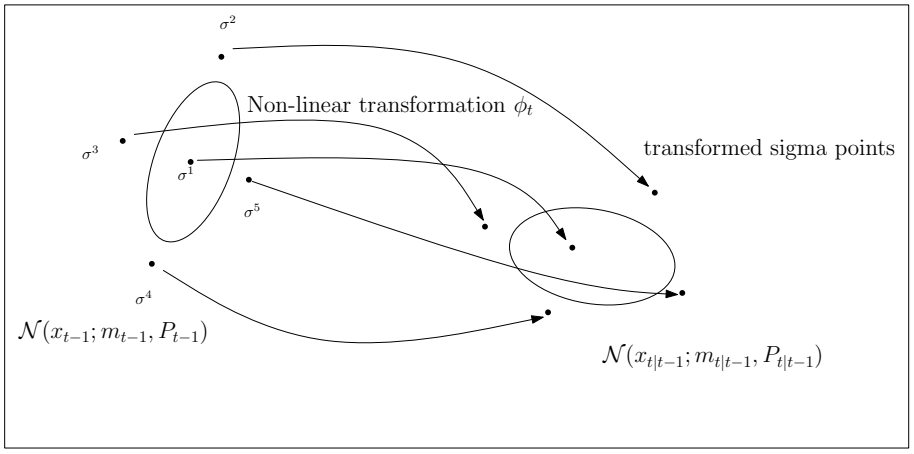


Figure 3.2: The principle of the unscented transform: a set of sigma-points are propagated through the non-linear function (arrows) and used to build a Gaussian term approximating the propagated distribution.

The transformation procedure is as follows: for the prediction, the sigma points are propagated through the transition function according to $x_{t|t-1}^l = \phi_{t-1}(x_{t-1}^l, v_{t-1}^l)$ and the predicted density is approximated by the Gaussian:

$$p(x_t|y_{0:t-1}) \approx \mathcal{N}(x_t; m_{t|t-1}, P_{t|t-1})$$

where:

$$m_{t|t-1} = \sum_{i=1}^{L} w_t^l x_{t|t-1}^l \tag{3.2.44}$$

$$P_{t|t-1} = \sum_{i=1}^{L} w_t^l (x_{t|t-1}^l - m_{t|t-1})(x_{t|t-1}^l - m_{t|t-1})^T \tag{3.2.45}$$

In the update, the sigma points are propagated through the measurement function: $y_{t|t-1}^l = h_t(x_{t|t-1}^l, w_{t-1}^l)$ for $l = 0, \ldots, L$ and the updated density at time t is approximated by the Gaussian:

$$p(x_t|y_{0:t}) \approx \mathcal{N}(x_t; m_t, P_t)$$

where:

$$m_t = m_{t|t-1} + K_t(y_t - y_{t|t-1}) \tag{3.2.46}$$

$$P_t = P_{t|t-1} - G_t S_t^{-1} G_t^T \tag{3.2.47}$$

$$y_{t|t-1} = \sum_{l=1}^{L} w_t^l y_{t|t-1}^l \tag{3.2.48}$$

$$K_t = G_t S_t^{-1} \tag{3.2.49}$$

$$S_t = \sum_{l=1}^{L} w_t^l (y_{t|t-1}^l - y_{t|t-1})(y_{t|t-1}^l - y_{t|t-1})^T \tag{3.2.50}$$

$$G_t = \sum_{l=1}^{L} w_t^l (x_{t|t-1}^l - m_{t|t-1})(x_{t|t-1}^l - m_{t|t-1})^T \tag{3.2.51}$$

The property of the algorithm have been studied in [55] and [56]. The UKF usually performs better than the EKF since the UKF is accurate up to the second order of the Taylor series expansion of the transformation, but it still perform poorly if the non-linearities in the models are severe.

3.2.7 Particle Filters

The particle filter, also known as Sequential Monte Carlo filter, is based on the idea of approximating the pdf $p(\cdot|y_{0:t})$ by a weighted set of particles $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ and propagate them in time by using an *importance distribution*. A detailed discussion on the characteristics of Sequential Monte Carlo filters can be found in the book [37] and in numerous papers such as [108, 43, 2]. Nonetheless, as the techniques employed in Sequential Monte Carlo filters will be used extensively in the following, a summary is provided.

The basic idea of Monte Carlo sampling is that it is possible to approximate complex distributions by a large set of independently and identically distributed samples. Let for example $p(\cdot)$ be a probability density function; its discrete approximation by using N-samples $\{x^{(i)}\}_{i=1}^{N}$ is:

$$p(x) \approx \frac{1}{N}\sum_{i=1}^{N}\delta_{x^{(i)}}(x) \tag{3.2.52}$$

For any arbitrary p-integrable function $f(\cdot)$ the almost sure asymptotic convergence is verified:

$$\frac{1}{N}\sum_{i=1}^{N}f(x^{(i)}) \xrightarrow[N\to\infty]{a.s.} \int f(x)p(x)dx \tag{3.2.53}$$

The rate of convergence does not dependent on the dimension of the integral, but primarily on number of particles.
In many practical cases sampling directly from the desired density is impossible, as in the Bayes filter where the normalizing constant is generally difficult to compute. In those cases the technique of *importance sampling* is used: samples are drawn from a known density called *importance distribution* or *proposal density* and then weighted to take into account the fact that importance distribution differs from the desired distribution.
More precisely, suppose that the target density $p(x)$ is known up to a normalising constant: $p(x) \propto q(x)$. A proposal density $\pi(x)$ such that $support(p) \subseteq support(\pi)$ is chosen and N samples: $\{x_t^{(i)}\}_{i=1}^{N}$ are drawn from it. The target density $p(x)$ is then approximated by the set of weighted particles:

$$p(x) \approx \sum_{i=1}^{N} w^{(i)}\delta_{x^{(i)}}(x) \tag{3.2.54}$$

where the weights are:

$$\tilde{w}^{(i)} = \frac{q(x^{(i)})}{\pi(x^{(i)})} \tag{3.2.55}$$

$$w^{(i)} = \frac{\tilde{w}^{(i)}}{\sum_{j=1}^{N}\tilde{w}^{(j)}} \tag{3.2.56}$$

The weights $w^{(i)}$ are generally called *importance weights* or *normalised importance weights*.

The technique can be straightforwardly applied in the context of filtering.
Suppose the goal is to have an estimator of the trajectory of a target at time t conditioned to all the observations up to time t:

$$\mathbb{E}_{p(x_{0:t}|y_{0:t})}[h] = \int_{{E_s}^{t+1}} h(x_{0:t})p(x_{0:t}|y_{0:t})dx_{0:t} \tag{3.2.57}$$

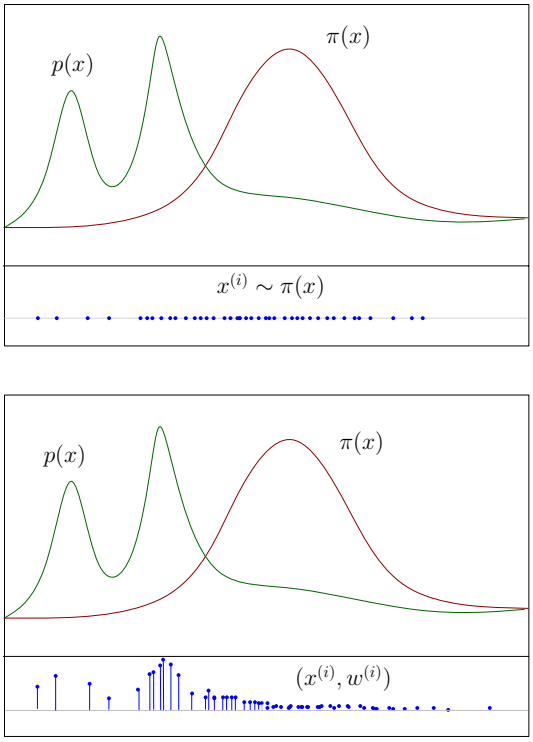


Figure 3.3: Importance sampling. The density $p(x)$ is approximated by a set of weighted particles sampled from an importance density $\pi(x)$. The quality of the approximation obviously depends on the importance density that has been chosen.

By using importance sampling the integral (3.2.57) can be rewritten as:

$$\mathbb{E}_{p(x_{0:t}|y_{0:t})}[h] = \int_{E_s{}^{t+1}} h(x_{0:t})p(x_{0:t}|y_{0:t})dx_{0:t} \tag{3.2.58}$$

$$= \int_{E_s{}^{t+1}} h(x_{0:t})\frac{p(x_{0:t}|y_{0:t})}{\pi(x_{0:t}|y_{0:t})}\pi(x_{0:t}|y_{0:t})dx_{0:t} \tag{3.2.59}$$

$$= \int_{E_s{}^{t+1}} h(x_{0:t})w(x_{0:t})\pi(x_{0:t}|y_{0:t})dx_{0:t} \tag{3.2.60}$$

The approximation of the integral can be done by sampling the target trajectories from $\pi(x_{0:t}|y_{0:t})$ and computing the weights $w(x_{0:t})$. With N samples $x_{0:t}^{(i)}$, $i = 1, \ldots, N$ the integral is approximated by:

$$\mathbb{E}_{p(x_{0:t}|y_{0:t})}[h] \approx \sum_{i=1}^{N} w(x_{0:t}^{(i)})h(x_{0:t}^{(i)}) \tag{3.2.61}$$

where the trajectories are sampled according to $\pi(x_{0:t}|y_{0:t})$ and the weight of each trajectory, denoted by $w(x_{0:t}^{(i)})$, is given by $w(x_{0:t}^{(i)}) = p(x_{0:t}^{(i)}|y_{0:t})/\pi(x_{0:t}^{(i)}|y_{0:t})$.

The technique of importance sampling to recursively obtain an approximate representation of the posterior density is known as *sequential importance sampling* (SIS)

[2, 38, 30]. The importance of this methods derives from the fact that it becomes possible to compute an estimate of $p(x_{0:t}|y_{0:t})$ without modifying the previously simulated trajectories, i.e. recursively in time. The posterior density at time t, in fact, can be obtained from the density at time $t-1$ by using:

$$p(x_{0:t}|y_{0:t}) = p(x_{0:t-1}|y_{0:t-1})\frac{p(x_t|x_{t-1})p(y_t|x_t)}{p(y_t|y_{0:t-1})} \tag{3.2.62}$$

which is used to calculate the weight of the particles recursively in time [37].
In the Particle Filter the posterior density at time $t-1$ is represented by the set of weighted particles $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$:

$$p(x_{t-1}|y_{0:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{x_{t-1}^{(i)}}(x_{t-1}) \tag{3.2.63}$$

while at time t it is approximated by a new set of particles $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^{N}$

$$p(x_t|y_{0:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{x_t^{(i)}}(x_t) \tag{3.2.64}$$

whose weights are obtained via importance sampling. More precisely, suppose that the complete proposal density $\pi_t(x_{0:t}|y_{0:t})$ can be written recursively in time by using the sequential proposal densities $q_t(x_t|x_{t-1}, y_{0:t})$:

$$\pi_t(x_{0:t}|y_{0:t}) = \pi_0(x_0)\prod_{j=1}^{t} q_j(x_j|x_{j-1}, y_j) \tag{3.2.65}$$

$$\tag{3.2.66}$$

The proposal density $q_t(x_t|x_{t-1}, y_{0:t})$ allows the particles to be propagated from time $t-1$ to time t and must verify $support(p(x_t|y_{0:t})) \subseteq support(q_t(\cdot|x_{t-1}^{(i)}, y_t))$.
The weights of the particles at time t are computed by:

$$w_t^{(i)} = \frac{p(x_{0:t}^{(i)}|y_{0:t})}{\pi(x_{0:t}^{(i)}|y_{0:t-1})} \tag{3.2.67}$$

$$= w_{t-1}^{(i)}\frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{p(y_t|y_{0:t-1})q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} \tag{3.2.68}$$

$$\propto w_{t-1}^{(i)}\frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} \tag{3.2.69}$$

As the normalization constant $p(y_t|y_{0:t-1})$ is common to each particle its computation is not necessary as soon as the particle's weights are normalised at each time step to obtain $\sum_{i=1}^{N} w_t^{(i)} = 1$.

The selection of the importance density is very important for the performance of the algorithm; a discussion about the conditionally optimal importance density as well as practical strategies can be found in [38] and [12].
A well known problem of this approach is that the variance of the importance weights increases over time until the complete degeneracy where only one particle has a non-null weight. This problem is mitigated by the inclusion of a resampling step in which the particles are resampled and duplicated according to their weights. Practically the resampling step is done when some condition on the empirical variance of the weights is verified. For instance, given the threshold κ [61] the resampling occurs when:

$$V_{eff} = \frac{1}{\sum_{i=1}^{N}(w_t^{(i)})^2} \leq \kappa \tag{3.2.70}$$

After the resampling the weights are set to $w_t^{(i)} = 1/N,\ i = 1, \ldots, N$.
The basic idea of the particle filter methodology has been widely extended over the years in order to improve the performances and to address some problems. One way, for example, is the use of the so-called Rao-Blackwellization [15] that consists in partitioning the state vector into a linear-Gaussian part and a non-linear non-Gaussian component; a particle filter is then used to solve the non-linear component, while the linear Gaussian part is solved analytically by using a Kalman filter.

An example of the sequential importance sampling filter (SIS) in the estimation of a target trajectory is given in Figures 3.4 to 3.7 where a target moves along the side of a rectangle and the angular measurements are reported by three sensors. The clouds of weighted particles is used to compute the estimate of the trajectory recursively in time.

The Sequential Importance Sampling and Resampling (SISR) filter, commonly called particle filter, is described in algorithm 1.

Single object tracking with misdetections and clutter

When only a single-object is potentially present in the surveillance region but the observation process may fail to detect it or may report spurious observations (clutter) the standard Bayesian filtering is not directly applicable, since it is not known which measurement should be used to update the state of the target. In this case the early approaches proposed in the literature consist in first associating one measurement to the active target and then to apply the standard single-target filter. The **Nearest Neighborhood** method (NN) [4], for example, constitutes the simplest approach and works by associating the closest observation (in a statistical sense) to the active target at each time step. Once the hypotheses on the association is done, the filter uses a Kalman update step to compute the sufficient statistics of the posterior density. Due to the drastic hypotheses, however, this approach performs very poorly in presence of a high clutter intensity or high misdetection probability.

The **Probabilistic Data Association** (PDA) filter [4] calculates, by using the Bayes rule, the probability that a given measurement was generated by the active target. The update step, is performed by using the Kalman update equation with a pseudo-measurement build by taking the average of all the observations weighted by their association probability. This technique is commonly used for single target tracking in clutter and shows reasonable performance [5]. The PDA filter, however, as well as the Nearest Neighborhood method, assumes some prior knowledge on the existence of the track as well as on its mean and covariance matrix. When there is no evidence of the existence of the target, a method for initiating single tracks is provided by the **Integrated Probabilistic Data Association filter** (IPDA). Further generalisations and refinements of the original algorithm have been proposed in [92, 106] and [113].

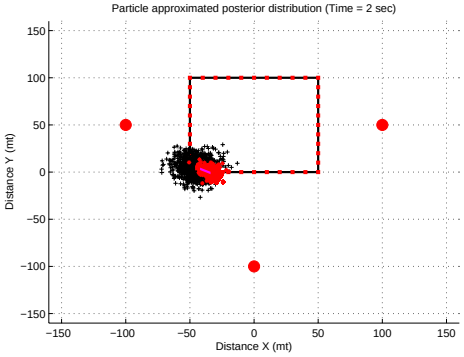


Figure 3.4: Propagated particles (black +), resampled particles (red ·) and estimated trajectory at time $t = 5$.

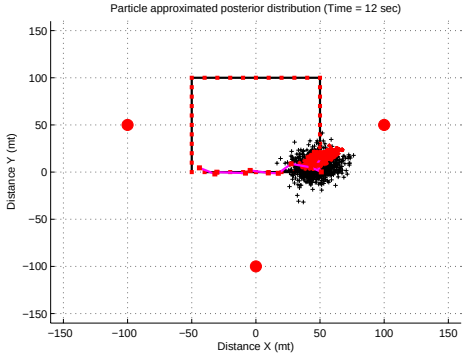


Figure 3.5: Propagated particles (black +), resampled particles (red ·) and estimated trajectory at time $t = 20$.

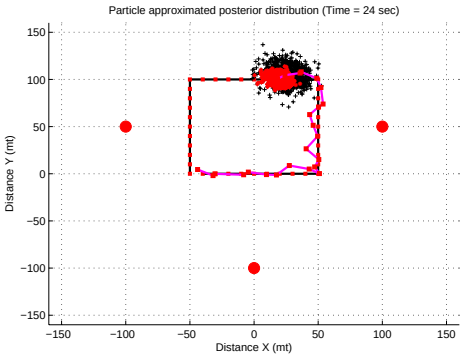


Figure 3.6: Propagated particles (black +), resampled particles (red ·) and estimated trajectory at time $t = 40$.

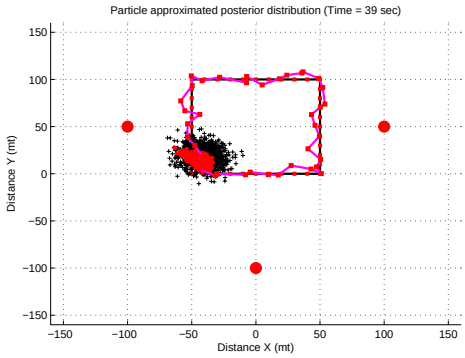


Figure 3.7: Propagated particles (black +), resampled particles (red ·) and estimated trajectory at time $t = 60$.

Algorithm 1 Generic Particle Filter

Initialization
N number of particles
for $i = 1$ to N **do**
  Sample N particles: $x_0^{(i)} \sim \pi_0(x_0)$
  Compute the initial weights: $w_0^{(i)} = p(x_0^{(i)})/\pi_0(x_0^{(i)})$
end for
$W_0 = \sum_{i=1}^{N} w_0^{(i)}$
For all the particles $w_0^{(i)} \leftarrow w_0^{(i)}/W_0$
for time $t \geq 1$ **do**
  for $i = 1$ to N **do**
    Sample $x_t^{(i)} \sim q_t(x_t|x_{t-1}^{(i)}, y_t)$
    Update the weights:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}$$

    with $\sum_{i=1}^{N} w_t^{(i)} = 1$
  end for
  Compute $V_{eff} = [\sum_{i=1}^{N} (w_t^{(i)})^2]^{-1}$
  if $V_{eff} \leq \kappa$ **then**
    Sample N particles according to the normalized distribution of their weights. Once the new particles have been sampled their weights are set to $w_t^{(i)} = 1/N$ and they are used for the next iteration.
  end if
end for

Chapter 4

Multi-object filtering in the Random Finite Set framework

4.1 Overview of multi-target filtering

Multi-object filtering aims at estimating the states of a possibly unknown number of targets in a given surveillance region. The sensors generally collect several measurements that may correspond to clutter or to real targets and typically lack the ability to associate each measurement to a specific target or to distinguish between real targets and false observations. Hence, one has to solve simultaneously three problems: the problem of data association, the estimation of the number of targets and the estimation of their states [4, 8]. Traditional approaches treat the problem by dividing the association from the estimation: observations are first associated to existing tracks and then the estimation is done by using single-target filtering techniques. Different algorithms have been proposed in the literature [5, 4, 8]. The **Global Nearest Neighbour** (GNN), for example, refines the association technique of the Nearest Neighborhood method (sec. 3.2.7) by choosing the associations between the active tracks and the observations that minimize in some statistical sense the total distance. The approach however suffers from the same limitations of the NN filter. The **Joint Probabilistic Data Association** (JPDA) filter [4] is an extension of the PDA filter when a fixed and known number of targets is present. At each time step the JPDA considers the set of all possible hypotheses that survive to a gating operation and combines the associations in proportion to their likelihood by avoiding conflicting measurement-to-track associations. However, the complexity of the computation of joint association probabilities grows exponentially with the number of targets and measurements. As the basic formulation of the algorithm shows exponential complexity and deals with a fixed and known number of targets, nu-

merous enhancements have been proposed: the **Joint Integrated PDA** (JIPDA) filter, for example, extends the technique to an unknown and time-varying number of objects [91] while an extension of the JPDA using Monte Carlo techniques [51] deals non-linear non-Gaussian models.
A widely used technique is the **Multiple Hypothesis Tracking** (MHT) [107] which maintains different association hypotheses between targets and observations and let future measurements resolve the uncertainty. More precisely the MHT filter works by performing an exhaustive search over previous time steps for the possible combinations of associations between active tracks and measurements. Past and present measurements are tested against various possibilities by computing their posterior probabilities using the Bayes rule. Then, a measurement is associated either to clutter or to a single active target and the set of hypotheses with the highest posterior probability is kept and propagated by using a standard Kalman filter. Initiation and termination of tracks are implicitly accommodated by performing the measurement association.
Clearly, it is necessary to adopt techniques to prevent the number of hypotheses from growing exponentially in time and for this reason the method results more computationally demanding than the JPDA.
The biggest drawback of the MHT is indeed its combinatorial nature; practical implementations [9, 10] use validation distances to limit the number of calculation and heuristic *pruning/merging* strategies to eliminate hypotheses with low probability. Variants of MHT have been proposed: the probabilistic MHT (PMHT), for example, uses probabilistic decisions for measurement-to-target associations [123, 140] while [93] proposes an approach to take care of target existence probabilities.

4.2 The Random Finite Set Framework

The Random Finite Set (RFS) approach to multi-object tracking provides a Bayesian framework for the recursive update of the multi-target posterior density. The main idea behind the RFS approach is to model the objects and the observations at each time step as *set-valued random variables* and then to characterize the relative uncertainties by using the probabilistic tools of Finite Set Statistics. The modelling of states and observations as random finite sets (the so called *multi-target state* and *multi-observation*) constitutes in fact the first step toward the generalization of the Bayes filter from the single to the multi-object case.
To give an example of the difficulties arising when dealing with multi-object problems it is sufficient to note that most of the basic concepts such as standard Bayes-optimal state estimators, expected value, least-squares optimization, are not even defined. Finite-set statistics addresses the conceptual gaps arising when the multisensor-multitarget problems are treated in a Bayesian perspective by providing systematic

techniques for generalizing the concepts of derivatives and integrals to *set-derivatives* and *set-integrals*. Similarly, probability-mass functions and likelihood functions are generalized to multisensor-multitarget *belief-mass functions* and *multi-target likelihood* functions.
More generally, FISST allows to extend seamlessly the formal Bayes modeling to nontraditional information and to multi-sensor multi-object problems in order to have systematic procedures to:

- construct multi-object measurement models
- construct multi-object motion models
- transform the multi-object measurement model into likelihood functions
- transform multi-object motion models into Markov densities

This section reviews the basic concepts of Random Finite Set and the main tools of Point Process theory. More advanced FISST concepts such, belief-mass functions, set-derivatives, set-integrals and probability generating functionals are reported in Appendix A. The full theoretic treatment as well as most of the proofs can be found in [88] and [128], and in monographs such as [26, 80].

Historical note

Random set theory was first systematically examined in connection with statistical geometry by Kendall [59] and Matheron [87] in the mid-1970s and then applied to two-dimensional image analysis by Serra [116]. Since then it has become a basic tool in theoretical statistics and it has inspired early practical works in data fusion by Mori, Chong, Tse, and Wisher. However, it is with the work of Mahler on FISST [42, 74] and on the approximations for the Bayesian multi-object filter that the random finite set approach has been consistently adopted in the filtering community.

Random Finite Sets

A random finite set can be described as a finite, set-valued random variable where not only the elements are random but also the cardinality of the set [26]. In other words, the realization of a RFS is an unordered and finite set of elements distributed according to a common probability distribution. This concept is particularly useful in the context of multi-object tracking where both the number of targets and their state are unknown and the uncertainty on both quantities (the cardinality and the state vectors) has to be characterized. The random finite set formulation is thus a natural way to model multi-object tracking problems and constitutes the base from which the Bayesian formulation for multi-object tracking can be built.

A RFS can be completely specified by the (discrete) distribution of its cardinality and a family of symmetric joint distributions that characterize the distributions of the points, conditional on the cardinality of the set [26, 122]. The formal definition of a RFS is as follows:

Definition 4.2.1. *A random finite set X on $E \subseteq \mathbb{R}^d$ is a measurable mapping*

$$X : \Omega \to \mathcal{F}(E)$$

*where Ω is a sample space with a probability measure $\mathbb{P}$ defined on a σ-algebra of events $\sigma(\Omega)$, and $\mathcal{F}(E)$ is the space of **finite** subsets of E, which is equipped with the Matheron topology [87].*

The probability measure $\mathbb{P}$ induces a probability law for the random finite set X. The most natural descriptor of the probability law for the RFS X is the probability distribution P on $\mathcal{F}(E)$ defined for any Borel subset $\mathcal{T}$ of $\mathcal{F}(E)$ as:

$$P(\mathcal{T}) = \mathbb{P}(\{X \in \mathcal{T}\})$$

where $\{X \in \mathcal{T}\}$ denotes the measurable subset $\{\omega \in \Omega : X(\omega) \in \mathcal{T}\}$.
Alternatively, the probability law for a RFS can also be given in terms of *belief mass function* β_X [133], (Appendix (A.1)) defined for any closed subset $S \subseteq E$:

$$\beta_X(S) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \subseteq S\}) = \mathbb{P}(\{X \subseteq S\})$$

or by using the notion of *void probability* [26, 122] which is defined for any $S \subseteq E$ as:

$$\varsigma_X(S) = \mathbb{P}(\{\omega : |X(\omega) \cap S| = 0\}) = \beta_X(S^c)$$

In analogy to the random vector case, a very useful descriptor of an RFS is the probability density. However, as the operation of sum is not defined for subsets, the space $\mathcal{F}(E)$ does not inherit the usual operation of integration. Nonetheless a mathematically consistent notion of probability density on $\mathcal{F}(E)$ is available from point process theory [76, 80]. The characterization of the probability density for random finite sets in the measure theoretic formulation and in FISST are different but closely related. Their relationship has been clarified in [133] where it is shown that the unitless set derivative of a belief mass function is a probability density, and that a set-integral is closely related to the conventional (measure theoretic) integral. In multi-object filtering applications the computation of the probability density of a RFS from the dynamical and observation models is generally much easier with FISST than with the measure theoretic formulation. However, since central FISST concepts such as set-integral and set-derivative are not conventional probabilistic concepts care has to be taken when performing operations concerning multi-object densities.
The next section introduces the multi-object Bayes filter by using FISST and Radon-Nikodým derivatives.

4.3 Multi-object Bayes filter

Once the observations and the targets are modeled as a single *meta-state* and *meta-observation*, the multi-object filtering problem can be posed as a Bayesian filtering problem with state space $\mathcal{F}(E_s)$ and observation space $\mathcal{F}(E_o)$, where E_s and E_o denote the target and observation space respectively.
For instance, for any Borel subset $\mathcal{U} \subseteq \mathcal{F}(E_s)$ and $\mathcal{V} \subseteq \mathcal{F}(E_o)$ let the posterior probability measure of the RFS X_t given all the observation RFSs $Y_{1:t} = (Y_1, \ldots, Y_t)$ be:

$$P_{t|t}(\mathcal{U}|Y_{1:t}) \equiv P(X_t \in \mathcal{U}|Y_{1:t}) \tag{4.3.1}$$

the prior probability of $\mathcal{U}$ given the RFS X_{t-1}:

$$P_{t|t-1}(\mathcal{U}|X_{t-1}) \equiv P(X_t \in \mathcal{U}|X_{t-1}) \tag{4.3.2}$$

and the probability measure of $\mathcal{V}$:

$$P_t(\mathcal{V}|Y_t) \equiv P(Y_t \in \mathcal{V}|Y_t) \tag{4.3.3}$$

Let μ_s and μ_o be the dominating measures of the form (A.2.2). Then the *multi-target posterior density* $p_{t|t}(\cdot|Y_{1:t})$, the multi target transition density $f_{t|t-1}(\cdot|X_{t-1})$ and the multi-target likelihood $g_t(\cdot|X_t)$ are the Radon-Nikodým derivatives of $P_{t|t}(\cdot|Y_{1:t})$ w.r.t μ_s, $P_{t|t-1}(\cdot|X_{t-1})$ w.r.t μ_s and $P_t(\cdot|X_t)$ w.r.t μ_o respectively:

$$P_{t|t}(\mathcal{U}|Y_{1:t}) = \int_{\mathcal{U}} p_{t|t}(X_t|Y_{1:t})\mu_s(dX_t) \tag{4.3.4}$$

$$P_{t|t-1}(\mathcal{U}|X_{t-1}) = \int_{\mathcal{U}} f_{t|t-1}(X_t|X_{t-1})\mu_s(dX_t) \tag{4.3.5}$$

$$P_t(\mathcal{V}|Y_t) = \int_{\mathcal{V}} g_t(Y_t|X_t)\mu_o(dY_t) \tag{4.3.6}$$

The optimal multi-object Bayes recursion is given by the recursion:

$$p_{t|t-1}(X_t|Y_{1:t-1}) = \int f_{t|t-1}(X_t|X)p_t(X_t|Y_{1:t-1})\mu_s(dX) \tag{4.3.7}$$

$$p_t(X_t|Y_{1:t-1}) = \frac{g_t(Y_t|X_t)p_{t|t-1}(X_t|Y_{1:t-1})}{\int g_t(Y_t|X)p_{t|t-1}(X|Y_{1:t-1})\mu_s(dX)} \tag{4.3.8}$$

Obviously, the main difference with the standard single-target, clutter-free Bayes recursion is that X_t and Y_t are RFS that may change dimension as t changes.

In the FISST framework the Bayes recursion is formulated in a conceptually different way by using belief mass functions. For any closed subsets $S \subseteq E_s$ and $T \subseteq E_o$ let the posterior belief mass function of the RFS X_t given all the observation

sets, the belief mass function of the RFS X_t, and the belief mass function of the observation set be:

$$\beta_{t|t-1}(S|Y_{1:t}) \equiv P(X_t \subseteq S|Y_{1:t}) \tag{4.3.9}$$

$$\beta_t(S|Y_{t-1}) \equiv P(X_t \subseteq S|Y_{t-1}) \tag{4.3.10}$$

$$\beta_t(T|X_t) \equiv P(Y_t \subseteq T|X_{t-1}) \tag{4.3.11}$$

The FISST multi-target posterior density $\pi_t(X_t|Y_{1:t})$, FISST multi-target transition density $\phi_{t|t-1}(\cdot|X_{t-1})$ and FISST multi-target likelihood $\rho_t(\cdot|X_t)$ are the set-derivatives of $\beta_{t|t-1}(S|Y_{1:t}), \beta_t(S|Y_{t-1}), \beta_t(T|X_t)$ respectively. The FISST multi-target Bayes filter proposed in [42] is given by the recursion:

$$\pi_{t|t-1}(X_t|Y_{1:t-1}) = \int \phi_{t|t-1}(X_t|X)\pi_t(X|Y_{1:t-1})\delta X \tag{4.3.12}$$

$$\pi_t(X_t|Y_t) = \frac{\rho_t(Y_t|X_t)\pi_{t|t-1}(X_t|Y_{1:t-1})}{\int \rho_t(Y_t|X)\pi_{t|t-1}(X|Y_{1:t-1})\delta X} \tag{4.3.13}$$

Although these equations look very similar to (4.3.7) and (4.3.8) the difference is that (4.3.12) and (4.3.13) are set-integrals and the functions involved have units whereas the former are unitless. The validity of the FISST-Bayes equations is verified by using the result in A.2.1 as described in [133]. An insight on the issues of unit-dependency in RFS density functions, as opposed to conventional densities is given in appendix A.5.

4.4 Point processes

Many important results concerning RFS theory have been stated with terms from point process theory. This section briefly reviews the terminology of point process theory and the relationship with the RFS formulation.

- A **counting measure** n on a space E is a measure taking values in $\mathbb{N} \cup \{\infty\}$ such that $n(B)$ is finite for any bounded subset B of E (see Appendix B.1.1). It can be identified with a countable collection of points of E such that $n(B)$ equals the number of points that fall in B.

- A **point process** $\mathcal{X}$ on the space E is a measurable mapping from a sample space Ω, with a probability measure $\mathbb{P}$ defined on a sigma algebra $\sigma(\Omega)$ to the space of counting measures on E. A point process $\mathcal{X}$ is *simple* if $\mathcal{X}(\{x\}) \in \{0, 1\}, \forall\ x \in E$, with probability 1, *finite* if $\mathcal{X}(E) < \infty$, with probability 1, and *simple-finite* if it is simple and finite.

- A **RFS** X can be identified with a *simple-finite* point process $\mathcal{X}$ in the sense that $\mathcal{X}(B) = |X \cap B|$, for all subsets B of E, where $|X|$ denotes the number of elements in X.

As a consequence the terms *random finite set* and *simple-finite point process* are used to refer to the same mathematical object. Alternative and mathematically equivalent representations for the RFS $X = \{x_1, \ldots, x_n\}$ are shown in figure 4.1 and resumed as follows.

- Dirac sum notation (sum of Dirac delta functions located at points in X):

$$\delta_X(x) = \delta_{x_1}(x) + \ldots + \delta_{x_n}(x) \tag{4.4.1}$$

- Random counting measure notation $N_X(S)$: number of points in a subset S:

$$N_X(S) = \int_S \delta_X(x)dx = |X \cap S| \tag{4.4.2}$$

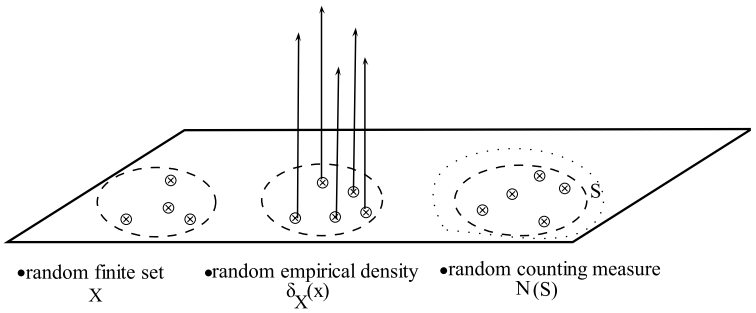


Figure 4.1: Three equivalent representations of a (multidimensional) simple point process.

Poisson Point Process

An important class of random finite sets are the Poisson RFS which are uniquely characterized by their intensity function.
We will use the notation $\gamma(x) \in \mathcal{M}(E)$ for the point process intensity and:

$$\gamma(f) = \int \gamma(x) f(x) dx$$

for the value of a bounded function f over the intensity. When $f(x) = 1$ the integral of the intensity over the domain (sometimes called mass), corresponding to the expected number of points, is denoted by $\gamma(1) \geq 0$.
A realization of a Poisson point process consists in N points i.i.d. with distribution

$\eta(dx) = \gamma(dx)/\gamma(1)$ where N is an integer-valued Poisson random variable with parameter $\gamma(1)$.
One of the main simplifications of Poisson point process comes from the fact that their expectation measure coincides with their intensity measure. Two important properties of Poisson point processes are [26]:

Lemma 4.4.1. ***(Thinning)*** *Let $\sum_{1\leq i\leq N}\delta_{X_i}$ be a Poisson point process with intensity measure $\gamma(x)$ on E. We consider a subset $A \in E$ such that $\gamma(A) > 0$. Then, the restriction $\sum_{1\leq i\leq N} 1_A(X_i)\delta_{X_i}$ of the point process to the set A is a again a Poisson point process with intensity measure $\gamma_A(dx) = 1_A(x)\gamma(dx)$.*

Lemma 4.4.2. ***(Superposition)*** *Let $(\mathcal{X}_i)_{i\geq 1}$ be a sequence of independent Poisson point processes with intensity measure $(\gamma_i)_{i\geq 1}$ on some common measurable state space E. For any $d \geq 1$, $\mathcal{X}$ is a Poisson point process with intensity measure $\sum_{1\leq i\leq d}\gamma_i$ if, and only if, $\mathcal{X}$ is equal in law to the Poisson point process $\sum_{1\leq i\leq d}\mathcal{X}_i$.*

In the RFS context, a multitarget density $f(X)$ is Poisson if:

$$f(X) = e^{-\gamma(1)}\gamma(x_1)\cdots\gamma(x_n) \tag{4.4.3}$$

where $X = \{x_1, \ldots, x_n\}$. Essentially the Poisson RFS characterizes a set of points with no spatial interaction, i.e. complete spatial randomness. In the context of tracking this property is used to model the fact that generally the position of a target and its dynamics is independent from the position of the other targets. Real-world instances where this approximation is not verified are, however, very common. It is easy to verify that a Poisson density $f(X)$ with intensity $\gamma(x)$ and parameter $\gamma(1)$ verifies:

$$\int f(X)\delta X = 1 \tag{4.4.4}$$

The PGFl of a Poisson RFS is:

$$G[h] = e^{I[h]-\gamma(1)} \tag{4.4.5}$$

where $I[h] = \int \gamma(x)h(x)dx$.

Other classes of point processes

Other important classes of RFSs commonly encountered in multi-object filtering are the *independent and identically distributed cluster* RFS, the *Bernoulli* RFS and the *multi-Bernoulli* RFS.

- An independent and identically distributed (i.i.d) cluster RFS X is uniquely characterized by intensity function $\gamma(\cdot)$ and cardinality distribution $\rho(n)$. The cardinality must satisfy $\sum_{n=0}^{\infty} n\rho(n) = \int \gamma(x)dx$ but it has not to be Poisson.

For a given cardinality the elements of X are independent and identically distributed according to $\gamma(\cdot)/\int \gamma(x)dx$.

- A Bernoulli RFS X has probability $1-q$ of being empty, and probability q of containing only one element distributed according to a probability density $p(\cdot)$. It is thus completely described by the parameter pair (q, p).
- A multi-Bernoulli RFS is the union of a fixed number of independent Bernoulli RFSs $X^{(i)}$:
$$X = \bigcup_{i=1}^{M} X^{(i)}$$
where $q^{(i)}$ is the existence probability of $X^{(i)}$ and $p^{(i)}$ the probability density of the element.

More details about the properties of these classes of random finite sets can be found in [135, 26] and references therein.

4.5 Random Finite Set model of multi-object filtering

This section introduces the traditional formulation of the multi-object filtering model in the random finite set framework. The model will be partially redefined and rediscussed in Chapters 7 and 8 in the context of branching processes.
Suppose that at a given time t, N_t targets whose states take values in some measurable space E_s are present. The state space encapsulates all the information about the targets: their kinematic properties (position, velocity, angles) and possibly other information such as the type of target or the dynamical model in use. The target states at time t are denoted by $(x_t^i)_{1\leq i\leq N_t} \in E_s$. At each time step the observation process reports M_t observations in a possibly different space: $(y_t^i)_{1\leq i\leq M_t} \in E_o$. The order in which the targets and observations are listed has no importance, as there is no association between measurements and targets. Moreover, some target may pass undetected during the scan and some measurements may be false alarms caused by noise or errors in the observation process. Therefore, it is natural to model the collections of states and observations at each time step as finite sets [42]:

$$X_t = \{x_{t,1}, \ldots, x_{t,N_t}\} \in \mathcal{F}(E_s)$$

$$Y_t = \{y_{t,1}, \ldots, y_{t,M_t}\} \in \mathcal{F}(E_o)$$

During the transition from time t to $t+1$ some target may disappear, new targets may appear and the surviving targets evolve according to their dynamical model.

From time t to time $t+1$, each target survives with probability $p_{s,t}(x_t)$ or dies with probability $1-p_{s,t}(x_t)$. Conditional on its existence at time t each target is assumed to follow a Markov process modeled by $f_{t+1|t}(x_{t+1}|x_t)$. A new target at time $t+1$ may appear either by spontaneous birth or because created by an existing target (a phenomenon commonly called *spawning*). Thus, the behaviour of a single object $x_{t+1} \in X_{t+1}$ is modeled by a Bernoulli RFS:

$$S_{t+1|t}(x_t)$$

where the target dies and assumes the value $\{\emptyset\}$ with probability $1 - p_{s,t}(x_t)$ or it moves to a new state with probability density $f_{t+1|t}(x_{t+1}|x_t)$.
The RFS of targets generated by an existing target with state x_t is denoted by:

$$B_{t+1|t}(x_t)$$

while the RFS of new targets appeared at time $t+1$ are denoted by Γ_{t+1}. The random finite set of targets at time $t+1$ is thus given by the union of the previous RFS:

$$X_{t+1} = \left(\cup_{x \in X_t} S_{t+1|t}(x)\right) \cup \left(\cup_{x \in X_{t+1}} B_{t+1|t}(x)\right) \cup \Gamma_{t+1} \tag{4.5.1}$$

Similarly, the measurement model, which accounts for detection, uncertainty and clutter is described by defining the set-valued observations for each time step Y_t. A given target is either detected with probability $p_{d,t}(x_t)$ or missed with probability $(1 - p_{d,t}(x_t))$ and conditional on the detection, the probability density of obtaining an observation from x_t is given by the likelihood function $g_t(\cdot|x_t)$ In addition to the target-originated measurements the sensor also receives a set of false measurements, or clutter. The set of measurements at time t: $Y_t = \{y_{t,1}, \ldots, y_{t,M_t}\}$ is given by

$$Y_t = \left(\cup_{x \in E_t} \Theta(x)\right) \cup K_t \tag{4.5.2}$$

where $\Theta(x)$ is the random set of measurements coming from target $x \in X_t$ and K_t is the set of measurements coming from clutter at time. It is worth stressing that, one of the conditions for the derivation of the general PHD filter is that each target contributes to the generation of at most one observation. Moreover, the random finite sets $\Theta(x)$ and K_t are assumed to be mutually independent as well as $S_{t+1|t}$, $B_{t+1|t}$ and Γ_{t+1}. Different models for the single-target transitions $f_{t+1|t}(x_{t+1}|x_t)$ are covered in the survey [65]. Figure 4.2 shows a schematic representation of the multi-target model.

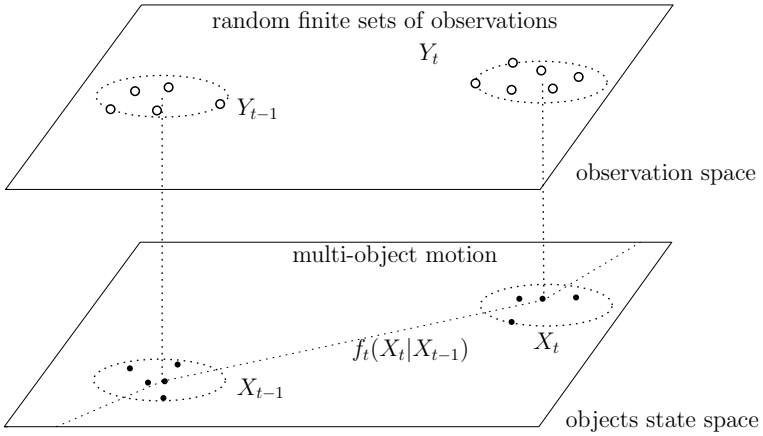


Figure 4.2: Random Finite Set model of multi-object tracking

As discussed in section 4.2, by modeling the states and observations as RFSs the problem can be posed as a Bayesian filtering problem with multi-object state space $\mathcal{F}(E_s)$ and multi object observation space $\mathcal{F}(E_o)$. The posterior probability density of the multi-object state at time t contains all the information about the cardinality and distribution of targets, unfortunately, the multi-target Bayes recursion involves integrals which are intractable in general and even numerical approximations are often too computationally intensive when the number of targets is relatively large [76], [133]. The Probability Hypothesis Density (PHD) filter instead propagates the first moment of the multi-object posterior and constitutes a computationally viable approach to multi-object filtering with random finite sets.
As a reference, table 4.1 provides a summary of the correspondences between the mathematical objects of single and multiple object filtering [42].

Random vector	**Finite Random Set**
state vector x	state set X
observation vector y	observation set Y
sensor model $y_t = h_t(x_t, w_t)$	observation RFS $Y_t = (\cup_{x \in E_t} \Theta(x)) \cup K_t$
motion model $x_{t+1} = m(x_t, v_t)$	target RFS $X_{t+1} = \left(\cup_{x \in X_t} S_{t+1\|t}(x)\right) \cup (\cup_{x \in X_{t+1}} B_{t+1\|t}(x)) \cup \Gamma_{t+1}$
differentiation $\frac{dp_Y}{dy}$	set differentiation $\frac{\delta \beta_\Xi}{\delta Y}(S)$ functional differentiation $\frac{\delta G_\Xi}{\delta Y}[h]$
integration (Lebesgue) $\int_S f_X(x) d\lambda(x) = P_X(S)$	set integration $\int_S f_\Xi(X)\delta(X) = \beta_\Xi(S)$ functional set integration $\int_S h^Y f_\Xi(X)\delta(X) = G_\Xi[h]$
expected value: $\int x f_X(x) dx$	probability hypothesis density (PHD): $\gamma_\Xi(x) = \int f_\Xi(\{x\} \cup Y)\delta(Y)$
probability measure $p_Y(S) = \mathbb{P}(Y \in S)$	belief mass function: $\beta_\Xi(S) = \mathbb{P}(\Xi \subseteq S)$ probability generating functional: $G_\Xi[h]$
density function $f_Y(y) = \frac{dP_Y}{dy}$	multi-object density function $f_\Xi(Y) = \frac{\delta G_\Xi}{\delta Y}[0] = \frac{\delta \beta_\Xi}{\delta Y}(\emptyset)$
prior density $f_X(x)$	multi-object prior density $f_\Xi(X)$
posterior density $f_{t\|t}(x_t\|y_{1:t})$	multi-object posterior $f_{t\|t}(X_t\|Y_{1:t})$
Markov density $f_{t+1\|t}(x_{t+1}\|x_t)$	multi-object Markov density $f_{t+1\|t}(X_{t+1}\|X_t)$
likelihood function $g_t(y_t\|x_t)$	multi-object likelihood $g_t(Y_t\|X_t)$

Table 4.1: Correspondences between the concepts in the random vector and RFS formulation.

4.6 The Probability Hypothesis Density recursion

The Probability Hypothesis Density (PHD) filter is a tractable approximation to the optimal multi-target Bayes filters proposed in [76]. It consists of a prediction and an update step that recursively propagate the first order moment of the target random finite set, from which estimates on the number of targets as well as their states can be obtained.
The derivation of the PHD recursion is based on a set of assumptions:

- each target is assumed to evolve and generate measurements independently of one another,
- the birth RFS is assumed to be a Poisson RFS independent from the current or surviving targets,
- the clutter process is assumed to be a Poisson RFS independent from the measurements RFSs,
- the predicted and posterior multi-target RFS are approximated by Poisson RFSs.

These simplifications allow for an analytical expression for the posterior intensity function, and implicitly make the assumption that the higher order moments are negligible. Although these conditions are not usually met in practice the PHD filter remains a useful method of approximating the posterior intensity. Proofs can be found in [76].
Denoting by $\gamma_t(x)$ and $\gamma_{t|t-1}(x)$ the intensity functions associated to the posterior and predicted multi-object state respectively, the predicted intensity is given by:

$$\gamma_{t|t-1}(x) = \int_{E_s} [p_{s,t}(u) f_{t|t-1}(x|u) + b_{t|t-1}(x|u)] \gamma_{t-1|t-1}(u) du + \mu_t(x) \tag{4.6.1}$$

where $f_{t|t-1}(x|u)$ is the evolution density of a single target at time t, $p_{s,t}(u)$ is the survival probability of a target with state u, $\mu_t(x)$ is the birth intensity function of new targets at time t, and $b_{t|t-1}(x|u)$ is the intensity of the Point process corresponding to the new objects generated from the target with state u at time t. The posterior intensity is obtained by:

$$\gamma_{t|t}(x) = (1 - p_{d,t}(x))\gamma_{t|t-1}(x) + \sum_{y \in Y_t} \frac{p_{d,t}(x) g_t(y|x) \gamma_{t|t-1}(x)}{h_t(y) + \int p_{d,t}(u) g_t(y|u) \gamma_{t|t-1}(u) du} \tag{4.6.2}$$

where $g_t(y|x)$ denotes the single target likelihood function, $p_{d,t}(x)$ is the probability of detection and h_t is the clutter intensity at time t.
Equations (4.6.1) and (4.6.2) show that the PHD recursion is defined on the single target state space, which has a fixed and generally smaller number of dimensions

than the full multiple-target posterior; nevertheless it still involves multiple integrals that have no closed form expressions in general.

Overview of the literature and recent research

The PHD filter has been usually implemented by using sequential Monte Carlo methods, as proposed by Zajic, Mahler, Vo et al. [109, 132, 133, 143] or by using the closed-form Gaussian mixture implementation devised by Vo and Ma [130, 131]. This implementation, under certain simplifying assumptions, greatly improves the computational efficiency of the filter.

Since the derivation of the filter was established in the framework of Finite Set Statistics, its relationship to conventional probability was at first not entirely clear. In [133] Vo, Singh and Doucet established this relationship and investigated a Sequential Monte Carlo multi-target filter. A review of their results is provided in Appendix A.2.

Vo, Singh, Doucet, and Clark also established convergence results for the particle-PHD filter [133], and Clark and Vo proved a strong L1 uniform convergence property for the Gaussian-mixture PHD filter in [24].

Since the PHD filter does not provide label associations, much effort has been spent on researching and developing efficient and reliable peak extraction and peak-to-track association techniques. Panta, Vo, and Singh [104] proposed two different schemes according to which the PHD filter can be enhanced to provide track-valued estimates of individual targets and compared their performance with the MHT. Clark and Bell [20] proposed two methods for the identification of targets whose complexity is lower than the MHT and JPDA filters. Wang, Jing and Hu [142] proposed and studied a data association technique based on MHT in analogy to Panta et al. [103] who used the SMC-PHD filter to remove unlikely measurements before inputting the data to a Multiple Hypothesis Tracker filter. Lin [67] described a technique for peak-to-track association for SMC-PHD filter based on the matching between peaks and tracks by using optimization techniques to minimize the association cost. Panta et al. also discussed various issues regarding initiating, propagating and terminating tracks in the GM-PHD filter and proposed a technique for resolving identities of targets [101].

Practical applications of these methods have included tracking vehicles on different terrains [117], tracking targets with passive radars [125] and sonars [19]. Researchers from Walter of the Swedish Defence Research Agency have employed PHD filters for group-target tracking [117]. Wang, et al. have employed such methods for tracking people in digital videos [138].

More recently Houssineau and Laneuville [63] considered the problem of multi-target tracking with passive data obtained by geographically distributed cameras. Hu and

Wu [141] also investigated the application of the PHD filter to multi-target visual tracking with trajectory recognition.
From the theoretical standpoint the major recent improvements have been provided by Ronald Mahler and Ba-Ngu Vo who recently proposed approximate multi-sensor CPHD and PHD filters [83] that generalize the PHD to the multi-sensor case. Previous approaches to the multi-sensor case where based on the iteration of the corrector equation for each sensor, which was unpleasing because implicitly based on strong simplifying assumptions and not invariant under sensor reordering. Although these solutions remain computationally intractable in the general case, approximations that are invariant under sensor reordering and based on much weaker simplifying assumptions have been proposed [84]. Furthermore, recent research include the derivation of a multi-object first-moment smoother for forward-backward smoothing [18] and the closed form solution to the PHD density smoother [85].
Delande, Duflos et al. [36] proposed an extension of Mahler's work on multi-sensor PHD filtering based on the configuration of the sensors' fields of view (FOVs) and the joint partitioning of both the sensors and the state space. An exhaustive review of the problem of multi-target PHD filtering as well as several important extensions to the multi-sensor case are reported in [35].

Strength and weaknesses of the PHD filter

Potential advantages and disadvantages ([80], pag. 571) of PHD filter are:

Advantages:

- Computational complexity $O(mn)$ where n is the number of targets and m the number of observations.
- Seamless integration of misdetection events, false alarms, and sensor field of view.
- Seamless integration of targets appearence, disappearence and spawning.
- Flexibility in the implementation techniques (Monte Carlo methods or Gaussian approximations)
- Estimation of the number of targets at each time step.

Disadvantages:

- High variance on the estimates of the number of targets (due to the Poisson approximation)
- A great amount of information is lost by replacing the full multi-object posterior with the PHD function

- It doesn't perform well in cases with low signal to noise ratio (SNR).
- It is based on the assumption that targets are sufficiently far away that they can be modeled as mathematical points or close enough that no measurement is generated by more than one target. These assumptions are generally not verified in real-life applications.
- It does not provide peak-to-track association, but only target estimates at each time step.

The rest of this chapter details the two most common implementations of the PHD filter: the Sequential Monte Carlo and the Gaussian Mixture PHD filter. Fundamental for the study of multi-object algorithms is the notion of OSPA distance which is reviewed in section 4.10.

4.7 Sequential Monte Carlo PHD Filter

The generic Sequential Monte Carlo implementation of the PHD particle filter has been proposed by Vo et al. in [132, 133]. The idea is to approximate equations (4.6.1) and (4.6.2) by a set of weighted particles recursively in time.
More precisely, at time $t = 0$ the initial intensity γ_0 is approximated by a set of weighted particles $\left\{w_0^{(i)}, x_0^{(i)}\right\}_{i=1}^{L_0}$ such that:

$$\gamma_0(x) = \sum_{i=1}^{L_0} w_0^{(i)} \delta_{x_0^{(i)}}(x) \tag{4.7.1}$$

where L_0 is the number of particles and $w_0 = \frac{N_0}{L_0}$; N_0 denoting the expected number of targets at time zero. Similarly, at a generic time step $t-1$ the set of weighted particles $\left\{w_{t-1}^{(i)}, x_{t-1}^{(i)}\right\}_{i=1}^{L_{t-1}}$ approximates the intensity function γ_{t-1}, i.e.:

$$\gamma_{t-1}(x) = \sum_{i=1}^{L_{t-1}} w_{t-1}^{(i)} \delta_{x_{t-1}^{(i)}}(x) \tag{4.7.2}$$

The prediction step consists in obtaining a particle approximation of the predicted PHD intensity (4.6.1). Given the set of particles and their weights at time $t-1$ the predicted intensity function is given by:

$$\gamma_{t|t-1}(x_t) = \sum_{i=1}^{L_{t-1}} w_{t|t-1}^{(i)} \phi_t(x_t, x_{t-1}^{(i)}) + \mu_t(x_t) \tag{4.7.3}$$

where $\phi_t(x_t, \xi) = [p_{s,t}(\xi) f_{t|t-1}(x_t|\xi) + b_{t|t-1}(x_t|\xi)]$. Importance sampling is applied to each term of (4.7.3) to obtain a set of $L_{t-1} + J_t$ particles approximating the

predicted intensity.
Two importance distributions $q_t(.|x_{t-1}^{(i)}, Y_t)$ and $p_t(.|Y_t)$ are used to sample the particles associated to the PHD intensity and to the spontaneous birth intensity:

$$x_{t|t-1}^{(i)} \sim \begin{cases} q_t(.|x_{t-1}^{(i)}, Y_t) & i = 1, ..., L_{t-1} \\ p_t(.|Y_t) & i = L_{t-1}+1, ..., L_{t-1}+J_t \end{cases} \tag{4.7.4}$$

J_t particles are sampled from the normalized birth intensity in order to detect potential new targets entering in the scene. Each particle is weighted according to:

$$w_{t|t-1}^{(i)} = \begin{cases} \frac{\phi_t(x_{t|t-1}^{(i)}, x_{t-1}^{(i)})}{q_t(x_{t|t-1}^{(i)}|x_{t-1}^{(i)}, Y_t)} w_{t-1}^{(i)} & i = 1, ..., L_{t-1} \\ \frac{1}{J_t} \frac{\mu_k(x_{t|t-1}^{(i)})}{p_t(x_{t|t-1}^{(i)}|Y_t)} & i = L_{t-1}+1 \dots L_{t-1}+J_t \end{cases} \tag{4.7.5}$$

The update step utilizes the RFS of observations to update the weights of the particles generated by the prediction operator.
The updated intensity function is approximated by:

$$\gamma_t(x) = \sum_{i=1}^{L_{t-1}+J_t} w_{t|t}^{(i)} \delta_{x_{t|t-1}^{(i)}}(x) \tag{4.7.6}$$

where the weights are updated as follows:

$$w_{t|t}^{(i)} = \left[(1 - p_{d,t}(x_{t|t-1}^{(i)})) + \sum_{y \in Y_t} \frac{p_{d,t}(x_{t|t-1}^{(i)}) g_t(y|x_{t|t-1}^{(i)})}{h_t(y) + C_t(y)} \right] w_{t|t-1}^{(i)} \tag{4.7.7}$$

$$C_t(y) = \sum_{j=1}^{L_{t-1}+J_t} p_{d,t}(x_{t|t-1}^{(j)}) g_t(y|x_{t|t-1}^{(j)}) w_{t|t-1}^{(j)} \tag{4.7.8}$$

In equation (4.7.7), $h_t(y)$ is the intensity function of the clutter process.
As in the standard particle filter, resampling is needed after the update step, in order to minimise particle degeneracy. The set of updated particles $\left\{ w_{t|t}^{(i)}, x_{t|t-1}^{(i)} \right\}_{i=1}^{L_{t-1}+J_t}$ are resampled to generate a new set of particles $\left\{ \widetilde{w}_t^{(i)}, \widetilde{x}_t^{(i)} \right\}_{i=1}^{L_t}$.
Generally, in order to redistribute a sufficient number of particles on the important zones of the state space, the number of particles to resample is chosen to be proportional to the expected number of targets [100], i.e. $L_t = \alpha \tilde{N}_t$ where $\tilde{N}_t$ is obtained by summing the particle weights: $\tilde{N}_t = \sum_{i=1}^{L_{t-1}+J_t} w_{t|t}^{(i)}$. However, care must be taken when implementing the resampling step, as the new weights $\left\{ \widetilde{w}_t^{(i)} \right\}_{i=1}^{L_t}$ must sum up to $\tilde{N}_t$, not to one. Figures 4.3 and 4.4 provide an illustration of the particle approximation of the PHD function on a simple tracking scenario. For convenience, the pseudo-code for the SMC-PHD algorithm is reported at the end of the chapter.

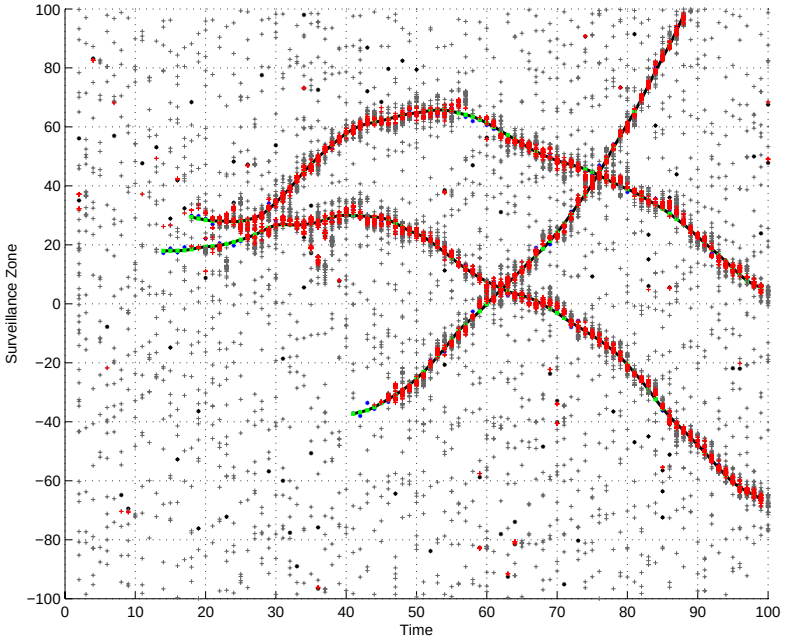


Figure 4.3: At each time step the intensity is approximated by a set of particles. Grey particles are those rejected by the algorithm and red particles are those which survive the resampling step. The intensity in a region of the space is obtained by summing the weights of the particles within that region.

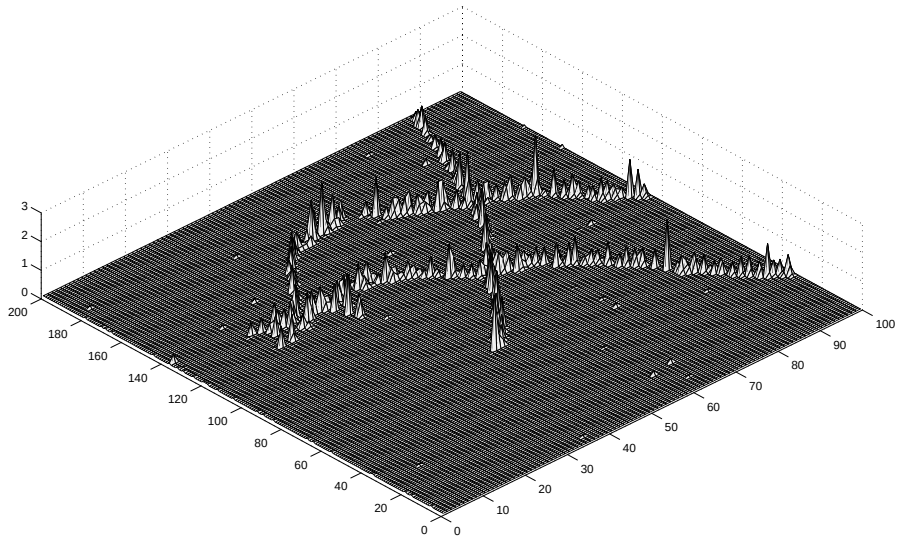


Figure 4.4: The profile of the intensity function can be retrieved by summing the weights of the particles. Peaks of the function correspond to the zone of the state space with the highest concentration of targets.

4.8 Gaussian Mixture PHD Filter

For a limited set of multi-target tracking problems, a closed-form solution exists and it is given by the Gaussian Mixture Probability Hypothesis Density (GM-PHD) filter [130, 129]. This section provides an overview of the linear Gaussian multi-target model for which the equations (4.6.1) and (4.6.2) admit a closed-form solution and describe the GM-PHD recursion. Convergence properties have been analyzed in [24] while an extension of the closed form solution to Jump Markov multi-target models by using the Unscented Transform has been proposed in [131]. A closed form Gaussian mixture solution to the forward-backward Probability Hypothesis Density smoothing recursion has also recently been proposed [85].
The GM-PHD closed form solution is based on the assumption that the target dynamics is linear Gaussian:

$$f_{t|t-1}(x_t|x_{t-1}) = \mathcal{N}(x_t; F_{t-1}x_{t-1}, Q_{t-1}) \tag{4.8.1}$$

as well as the measurement model:

$$g_t(y_t|x_t) = \mathcal{N}(y_t; H_t x_t, R_t) \tag{4.8.2}$$

The matrices F_t and Q_t represent the state transition matrix and process noise covariance matrix respectively, and H_t, R_t the observation matrix and the observation noise covariance matrix, at time t. Moreover, the intensities of the birth and spawn RFS are supposed to be Gaussian mixtures defined as:

$$\mu_t(x_t) = \sum_{i=1}^{J_{\mu,t}} w_{\mu,t}^{(i)} \mathcal{N}(x_t; m_{\mu,t}^{(i)}, P_{\mu,t}^{(i)}) \tag{4.8.3}$$

$$b_{t|t-1}(x_t|\zeta) = \sum_{j=1}^{J_{B,t}} w_{B,t}^{(i)} \mathcal{N}(x_t; F_{B,t-1}^{(j)}\zeta + d_{B,t-1}^{(j)}, Q_{B,t-1}^{(j)}) \tag{4.8.4}$$

where $J_{\mu,t}$, $m_{\mu,t}^{(i)}, P_{\mu,t}^{(i)}$ $w_{\mu,t}^{(i)}$ as well as $J_{B,t}, F_{B,t-1}^{(j)}, d_{B,t-1}^{(j)}, Q_{B,t-1}^{(j)}$, are known model parameters which determine the birth and spawn intensities. A third assumption is that the probabilities of target detection and target survival do not depend on the target state:

$$p_{d,t}(x_t) = p_{d,t} \tag{4.8.5}$$

$$p_{s,t}(x_t) = p_{s,t} \tag{4.8.6}$$

Under the previous assumptions the intensity function at each time step is a Gaussian mixture of the form:

$$\gamma_t(x_t) = \sum_{i=1}^{J_t} w_t^{(i)} \mathcal{N}(x_t; m_t^{(i)}, P_t^{(i)}) \tag{4.8.7}$$

The predicted intensity at time t is also a Gaussian mixture and it is given by the GM-PHD prediction equation:

$$\gamma_{t|t-1}(x_t) = \gamma_{S,t|t-1}(x_t) + \gamma_{B,t|t-1}(x_t) + \mu_t(x_t) \tag{4.8.8}$$

where

$$\gamma_{S,t|t-1}(x_t) = p_{s,t} \sum_{j=1}^{J_{t-1}} w_{t-1}^{(j)} \mathcal{N}(x_t; m_{S,t|t-1}^{(j)}, P_{S,t|t-1}^{(j)}) \tag{4.8.9}$$

$$m_{S,t|t-1}^{(j)} = F_{t-1} m_{t-1}^{(j)} \tag{4.8.10}$$

Moreover:

$$P_{S,t|t-1}^{(j)} = Q_{t-1} + F_{t-1} P_{t-1}^{(j)} (F_{t-1})^T \tag{4.8.11}$$

(4.8.10) and (4.8.11) are the Kalman filter prediction equations.

$$\gamma_{B,t|t-1}(x_t) = \sum_{j=1}^{J_{t-1}} \sum_{l=1}^{J_{B,t}} w_{t-1}^{(j)} w_{B,t}^{(l)} \mathcal{N}(x_t; m_{B,t|t-1}^{(j,l)}, P_{B,t|t-1}^{(j,l)}) \tag{4.8.12}$$

$$m_{B,t|t-1}^{(j,l)} = F_{t-1}^{(l)} m_{t-1}^{(j)} + d_{B,t}^{(l)} \tag{4.8.13}$$

where

$$P_{B,t|t-1}^{(j,l)} = Q_{B,t-1}^{(l)} + F_{B,t-1}^{(l)} P_{B,t|t-1}^{(j,l)} (F_{B,t-1}^{(l)})^T \tag{4.8.14}$$

The posterior intensity is a Gaussian mixture of the form:

$$\gamma_t(x_t) = (1 - p_{d,t}) \gamma_{t|t-1}(x_t) + \sum_{y \in Y_t} \gamma_{D,t}(x_t, y_t) \tag{4.8.15}$$

where

$$\gamma_{D,t}(x, y) = \sum_{j=1}^{J_{t|t-1}} w_t^{(j)}(y) \mathcal{N}(x; m_{t|t}^{(j)}(y), P_{t|t}^{(j)}) \tag{4.8.16}$$

$$w_t^{(j)}(y) = \frac{p_{d,t} w_{t|t-1}^{(j)} q_t^{(j)}(y)}{h_t(y) + p_{d,t} \sum_{h=1}^{J_{t|t-1}} w_{t|t-1}^{(h)} q_t^{(h)}(y)} \tag{4.8.17}$$

$$q_t^{(j)}(y) = \mathcal{N}(y; H_t m_{t|t-1}^{(j)}, R_t + H_t P_{t|t-1}^{(j)} H_t^T) \tag{4.8.18}$$

$$m_{t|t}^{(j)}(y) = m_{t|t-1}^{(j)} + K_t^{(j)} (y - H_t m_{t|t-1}^{(j)}) \tag{4.8.19}$$

$$P_{t|t}^{(j)} = [I - K_t^{(j)} H_t] P_{t|t-1}^{(j)} \tag{4.8.20}$$

$$K_t^{(j)} = P_{t|t-1}^{(j)} H_t^T (H_t P_{t|t-1}^{(j)} H_t^T + R_t)^{-1} \tag{4.8.21}$$

The equations (4.8.19), (4.8.20) and (4.8.21) are the update equations of the Kalman filter. The expected number of targets is obtained by summing the weights of the Gaussian terms [129, 130].

After the update the number of Gaussians in the mixture is $(J_{t-1}(1+J_{B,t})+J_{\gamma,t})(1+|Y_t|)$. Pruning and merging techniques have to be used in order to prevent the number of terms from growing exponentially; deterministic techniques and heuristics have been discussed extensively in [100] where pruning schemes based on a tree-based structure are proposed.
The most basic pruning scheme, however, is based on the deterministic choice of a threshold and on the consequent elimination of the Gaussian terms whose weight is below the threshold value. Section 5.2.1 evaluates different pruning schemes and provides comparions.

In many real-world applications the system and the observation function are not linear Gaussian; in this case the posterior intensity can no longer be represented by a Gaussian Mixture, nonetheless the GM-PHD filter can be adapted to non-linear models by replacing the Kalman filter equations (4.8.10), (4.8.11) and (4.8.19) - (4.8.21) by their linearization as in the Extended Kalman filter (EK) or by their approximation as in the Unscented Kalman filter (UKF) [131]. For convenience, the pseudo-code for the GM-PHD algorithm is reported at the end of the chapter.

4.9 Extensions to the PHD recursion

One limitation of the PHD recursion is the elimination of higher order moments which results in a loss of information on the cardinality causing an imprecise estimation of the number of objects in case of high clutter intensity.
The reason is that the PHD recursion propagates cardinality information with the intensity of a Poisson point process and since the mean and the variance of a Poisson distribution are equal, the variance of the cardinality estimate is expected to be high in presence on an high number of targets. In [79] Mahler introduced a generalization of the PHD recursion, known as the Cardinalized PHD (CPHD), that jointly propagates the posterior intensity function and the posterior cardinality distribution. The closed form solution to the CPHD recursion in case of linear-Gaussian model was published in [136]. The dissertation [135] offers an extensive and detailed study on the CPHD recursion and its implementation.

An extensive amount of research has also been done on tractable approximations to the multi-object Bayes recursion. The Multi-Object Multi-Bernoulli (MeMBer) recursion proposed by Mahler in [80] propagates approximately the multi-object posterior density via the propagation of the parameters of a multi-Bernoulli RFS. However, as demonstrated by Vo et al. in [137], the original MeMBer recursion is affected by a cardinality bias. The problem was addressed by the same authors which proposed a novel unbiased multi-Bernoulli based approximation, called Cardinality Balanced MeMBer (CBMeMBer) filter. The same authors also provided the analytical solution for linear Gaussian models. The Multi-Object Multi-Bernoulli is

outside the scope of this thesis, reference to the original formulation as well as the CBMeMBer filter can be found in [135].

Extended and Unresolved targets

The Probability Hypothesis Density (PHD) filter and the Cardinalized Probability Hypothesis Density (CPHD) filter are single-sensor filters and their multisensor generalizations are computationally intractable. The reason is that the standard multi-target measurement model on which they are based makes two kinds of approximations: the *small target approximation* and *no unresolved targets* approximation respectively. The first assumes that targets are sufficiently far away that they can be modeled as mathematical points. The second assumes that targets are close enough that no measurement is generated by more than a single target.
These assumptions are generally not verified in real-life applications. Two examples are as follows [81, 82]: suppose that a radar radiates a pulse at a ground target that is relatively near the radar. If the wavelength of the pulse is sufficiently small, it will tend to be back-reflected much more strongly by those areas on the target surface that resemble corner reflectors. The result is an image of the target consisting of a large number of point detections that stand out from the rest of the back-reflection signature. Targets that generate measurement-sets of this kind are called *extended targets*. On the other hand, when multiple targets are present in a scene, it is possible that a single radar-signature peak is due to a superposition of the signals of two or more targets that are very close together (relative to the resolution of the sensor). Such targets are said to be *unresolved*. In this case the PHD equations can be derived but they are combinatorial in form and thus computationally intractable. One recent contribution to the practical application of the GM-PHD filter to the tracking of extended targets by approximating Mahler's measurement-update equations has been published in [44].

4.10 Multi-object miss distance

While the concept of a miss-distance is straightforward in single-object problems, this is not the case in the multi-object case.
The concept of miss-distance as the error between the true and the estimated states of the targets plays a fundamental role in filtering as it allows the comparison of different algorithms and establishes an objective criterion to assess the quality of a filter. For a relatively long time a satisfactory notion of a multi-object miss-distance was not available despite an abundance of multi-object filtering techniques. A common practice was to use the optimal assignment paradigms such as that of Drummond and al. [39, 111] but this approach was inherently flawed as the formula is not a

metric and can only consistently measure the distance between multi-object states of the same cardinality. The first rigorous approach to multi-object performance evaluation was proposed by Hoffman and Mahler in [46] based on a Wasserstein construction. However the technique did not admit a physically consistent interpretation and suffered from severe limitations [135].
A key contribution was provided by Schuhmacher, Vo and Vo in [115] with the definition of the OSPA metric as the first meaningful multi-object miss-distance which can take into account the cardinality errors as well as the localization errors in a physically meaningful and mathematically consistent manner. Recently, one of the OSPA original authors described an adaptation of the metric [110] to evaluate the performance of multi-target tracking algorithms by considering the track label error; a second variant of the original metric has been proposed in [45] under the name of Mean OSPA (MOSPA).

As the OSPA metric is extensively used in further chapters to compare numerical results this section provides an overview of its formula, while the discussion of the mathematical properties can be found in [115].

Definition 4.10.1. ***[OSPA]***
Denote by $d_p^{(c)}(x,y) := min(c, d(x,y))$ *the distance between two points* x *and* y *that is cut off at* c*, and by* Π_j *the set of permutations on* $\{1,2,\ldots,j\}$*,* $j \in N_+ = \{1,2,\ldots\}$*. For* $1 \leq p < \infty$*,* $c > 0$ *and two arbitrary finite subsets of ground-truth target positions* $X = \{x_1,\ldots,x_m\}$ *and target estimates* $Y = \{y_1,\ldots,y_n\}$*, the OSPA distance is defined by*

$$\overline{d}_p^{(c)}(X,Y) := \begin{cases} 0, \ \text{if } m = n = 0 \\ \left(\frac{1}{n}\left(\min\limits_{\pi\in\Pi_n}\sum\limits_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m)\right)\right)^{\frac{1}{p}} \ \text{if } m \leq n \\ \bar{d}_p^{(c)}(Y,X) \ \text{if } m > n \end{cases} \tag{4.10.1}$$

The order parameter p determines the sensitivity of the metric in penalizing outlier estimates while the cut-off parameter c determines how the metric penalizes cardinality errors as opposed to localisation errors.
In the context of multi-object performance evaluation, it is useful to interpret the OSPA distance by considering separately two components which account respectively for the localisation and the cardinality errors:

$$\overline{e}_{p,loc}^{(c)}(X,Y) := \left(\frac{1}{n}\min_{\pi\in\Pi_n}\sum_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p\right)^{\frac{1}{p}} \tag{4.10.2}$$

$$\overline{e}^{(c)}_{p,card}(X,Y) := \left(\frac{c^p(n-m)}{n}\right)^{\frac{1}{p}} \tag{4.10.3}$$

if $m \leq n$, and $\bar{e}^{(c)}_{p,loc}(X,Y) := \bar{e}^{(c)}_{p,loc}(Y,X)$, $\bar{e}^{(c)}_{p,card}(X,Y) := \bar{e}^{(c)}_{p,card}(Y,X)$ if $m > n$.

Figure 4.5 shows intuitively the distances considered during the computation of the OSPA metric. Figures 4.6 and following and 4.9 and following show the application of the OSPA to two simple scenarios.

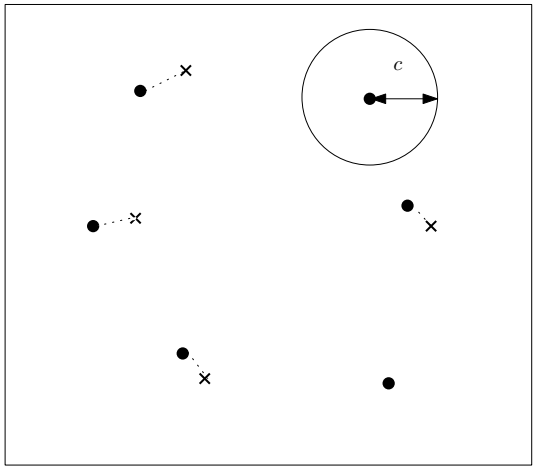


Figure 4.5: An intuitive representation of an optimal subpattern assignment. Black dots represents the ground-truth RFS X while a cross indicates the position of an estimate. The optimal assignment is immediately clear from inspection and indicated with a dotted-line. The cut-off value c is shown for the upper-right point. Two points are not estimated. The corresponding cardinality error is calculated by taking into account the cut-off value and the difference in cardinality between the two RFSs.

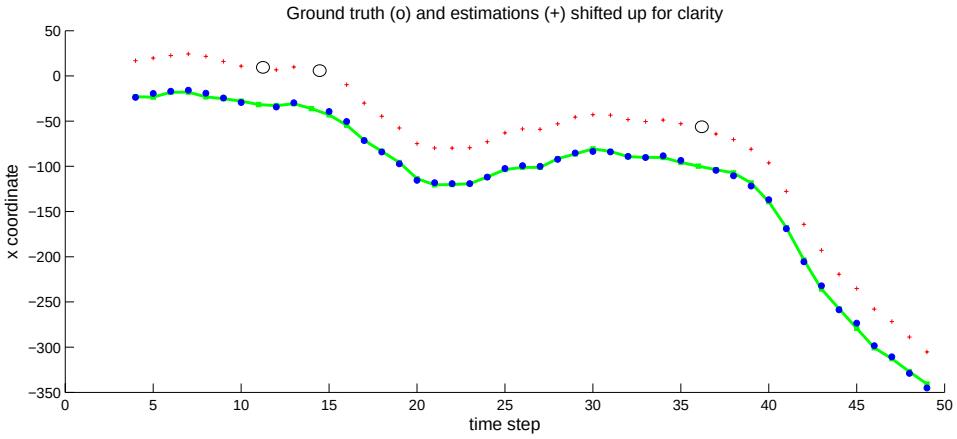


Figure 4.6: Test scenario I. One target evolves in a mono-dimensional region. Its position as estimated by the GM-PHD is shown as a red (+) with an offset of 40 units for the sake of clarity. Misdetections are highlighted by a circle.

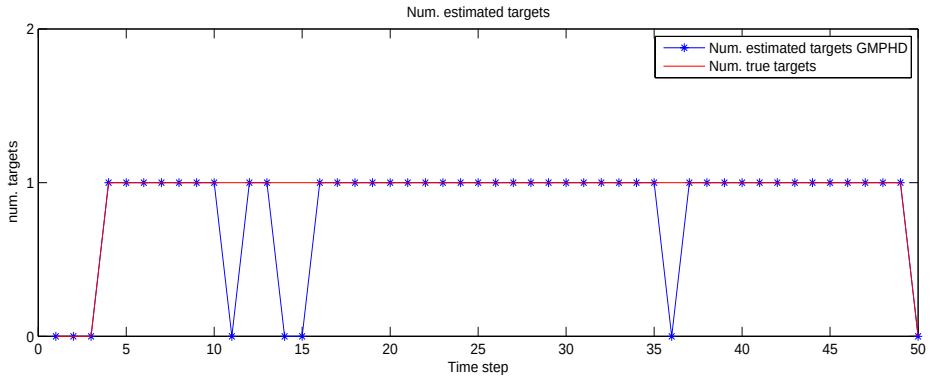


Figure 4.7: Number of ground truth targets and number of estimates.

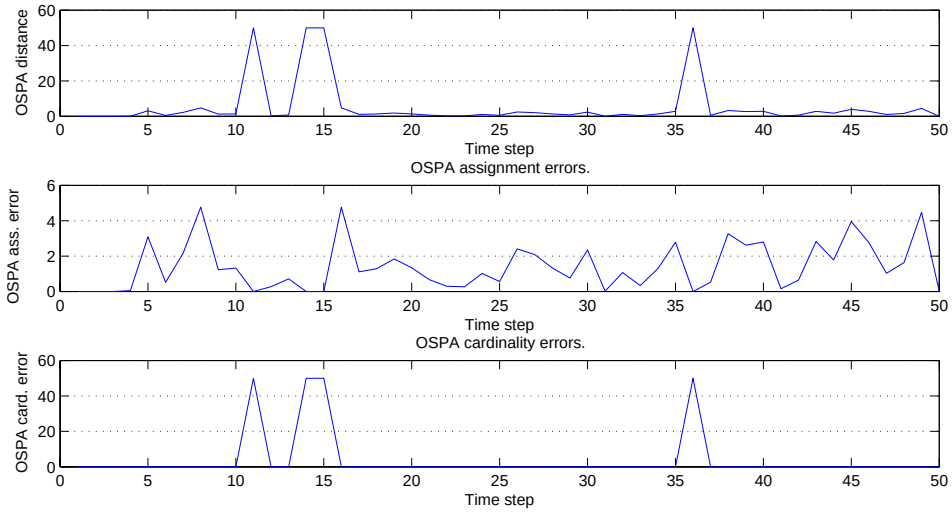


Figure 4.8: Ospa components. Cut-off value $c = 50$, $p = 1$. GM-PHD pruning threshold is set to 0.5

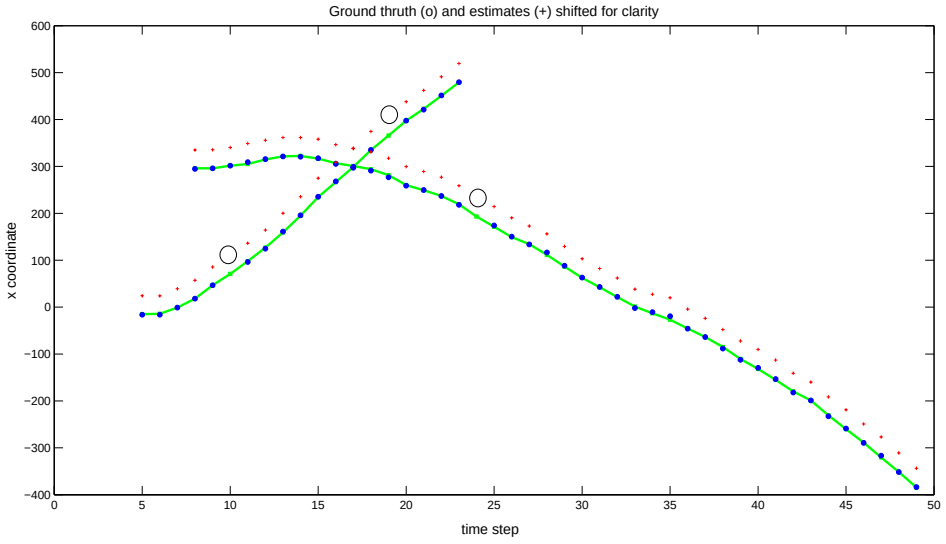


Figure 4.9: Test scenario II. Two targets evolve in a mono-dimensional region. Their positions as estimated by the GM-PHD are shown as a red (+) with an offset of 40 units for the sake of clarity. Misdetections are highlighted by a circle.

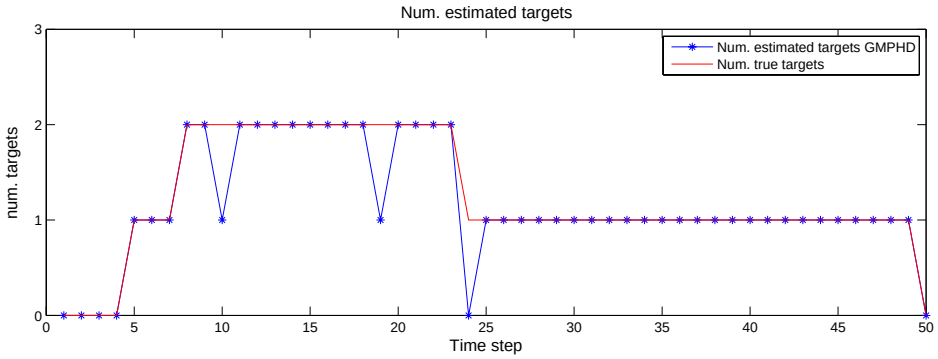


Figure 4.10: Number of ground truth targets and number of estimates.

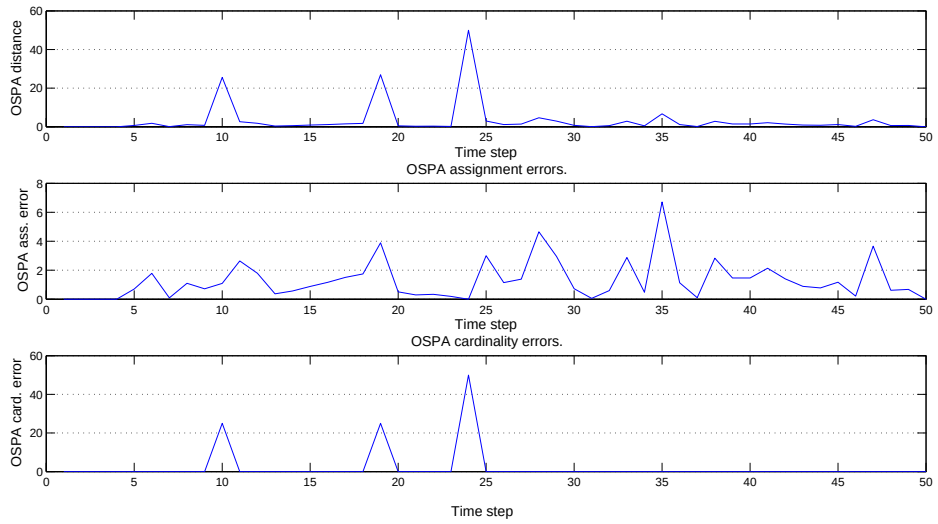


Figure 4.11: OSPA components. Cut-off value $c = 50$, $p = 1$. GM-PHD pruning threshold is set to 0.5

4.11 Multi-object state estimation

As introduced in section 4.2, conventional estimators (such as EAP and MAP) fail at providing target estimates in the multi-target case; moreover, the inconsistencies caused by the mismatch in the units of measurements of the multi-object density cannot be resolved in a consistent way to adapt the EAP or MAP estimators to the multi-object case.
The problem is illustrated by the following example [135, 78]: consider a scenario where there is at most one object located in the interval [0, 2] with units given in metres. Let the object state X be a Bernoulli RFS with probability density:

$$p(X) = \begin{cases} 0.5 & \text{if } X = \emptyset \\ 0.25 & \text{if } X = \{x\}, 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

In other words there is 0.5 probability that the object is absent and if the object is present, it is equally likely to be found anywhere in the interval [0, 2]. In this case the EAP estimate is not defined since the addition of sets is not defined. The MAP estimate, on the contrary would provide $X = \emptyset$ as an estimate, since it is the case with the highest relative probability. Of course this comparison is meaningless because the estimator is considering two quantities with different units of measurement. The inconsistency is revealed for example by changing the units of measurements, from metres to kilometers. In this case the probability density becomes:

$$p(X) = \begin{cases} 0.5 & \text{if } X = \emptyset \\ 250 & \text{if } X = \{x\}, 0 \leq x \leq 0.002 \\ 0 & \text{otherwise} \end{cases}$$

and the naive MAP estimator would give $X = \{x\}$ as the answer.
For this reason it is necessary to rely on alternative solutions for multi-object state estimation. The physical meaning of the intensity function, (i.e. the expected number of objects occurring in a region of the state space) is generally used, as its peaks indicate the regions of the state space with an high concentration of targets.
A short description of the **Marginal Multi-Object** estimator (MaM) and **Joint Multi-Object** (JoM) estimator reported here is intended as a reference; the detailed discussion can be found in [135].
The Marginal Multi-Object estimator is defined as a two-step estimator where the number of objects is first estimated using MAP on the posterior cardinality distribution, and then individual object states are estimated using a MAP estimate on the posterior density:

$$\hat{n}_t = \arg\sup_n \rho_t(n|Y_{1:y}) \tag{4.11.1}$$

$$\hat{X}_t^{\text{MaM}} = \arg \sup_{X:|X|=\hat{n}_t} p_t(X|Y_{1:y}) \tag{4.11.2}$$

where ρ_t denote the posterior cardinality distribution of the target RFS. Given c, a dimensionless constant, the Joint Multi-Object estimator is defined as:

$$\hat{X}_c^{\text{JoM}} = \arg \sup_X p_t(X|Y_{1:y}) \frac{c^X}{|X|!} \tag{4.11.3}$$

The parameter c determines the accuracy for the estimator as well as its rate of convergence. The techniques of multi-object state estimation for the GM-PHD and SMC-PHD used in the thesis are discussed in the next sections.

Multi-object state estimation for the SMC-PHD

From the particle representation of the posterior intensity the state estimates of the individual targets are extracted by locating peaks via clustering. Different algorithms such as the expectation-maximization (EM) or the K-means have been proposed and studied in the literature [117, 132, 22]. The EM method attempts to find the Gaussian mixture that best fits the particles while the K-means [71] partitions the particles of the posterior into the number of clusters given by the integer approximation of intensity integrated over the domain. The performance of both techniques have been studied in [17] with the result that the EM algorithm performs poorly in comparison to the K-means clustering. In the SMC-PHD filter presented in later chapter, K-means clustering is used.

Multi-object state estimation for the GM-PHD

In the Gaussian Mixture PHD filter the posterior intensity function at each time step is represented by a mixture of weighted Gaussians terms. The estimates are extracted by locating the means of all the terms with a weight above a fixed threshold. For instance, given a threshold w_T, the estimates $\hat{X}_t$ are obtained by taking the means $m_t^{(i)}$ of the Gaussian terms:

$$\hat{X}_t = \{m_t^{(i)} : w_t^{(i)} \geq w_T\} \tag{4.11.4}$$

As a result, the GM-PHD filter does not require a computationally demanding clustering technique like the SMC-PHD filter. Enhancements to this basic and widely used strategy are discussed in [100]. Section 5.2 investigates a stochastic pruning strategy and compares it with the standard, deterministic strategies proposed in the literature.

Algorithm 2 Sequential Monte Carlo PHD filter (SMC-PHD)

Initialization.

Sample L_0 initial particles:

$x_0^{(i)} \sim q_0(.)$ (tipically the normalize birth intensity $\mu_0/\mu_0(1)$)

Set the initial weights:

$w_0^{(i)} = N_0/L_0$

Prediction.

At time $t \geq 1$

for time $t = 1, \ldots$ **do**

Sample the prediction particles and compute their weights

for $i = 1, \ldots, L_{t-1}$ **do**

$x_{t|t-1}^{(i)} \sim q_t(.|x_{t-1}^{(i)}, Y_t)$

$w_{t|t-1}^{(i)} = \frac{\phi_t(x_{t|t-1}^{(i)}, x_{t-1}^{(i)})}{q_t(x_{t|t-1}^{(i)}|x_{t-1}^{(i)}, Y_t)} w_{t-1}^{(i)}$

end for

Sample the birth intensity particles and compute their weights

for $i = L_{t-1} + 1, \ldots, L_{t-1} + J_t$ **do**

$x_{t|t-1}^{(i)} \sim p_t(.|Y_t)$

$w_{t|t-1}^{(i)} = \frac{1}{J_t} \frac{\mu_k(x_{t|t-1}^{(i)})}{p_t(x_{t|t-1}^{(i)}|Y_t)}$

end for

Update.

For each observation $y \in Y_t$

$C_t(y) = \sum_{j=1}^{L_{t-1}+J_t} p_{d,t}(x_{t|t-1}^{(j)}) g_t(y|x_{t|t-1}^{(j)}) w_{t|t-1}^{(j)}$

for $i = 1, \ldots, L_{t-1} + J_t$ **do**

$w_{t|t}^{(i)} = \left[(1 - p_{d,t}(x_{t|t-1}^{(i)})) + \sum_{y \in Y_t} \frac{p_{d,t}(x_{t|t-1}^{(i)}) g_t(y|x_{t|t-1}^{(i)})}{h_t(y) + C_t(y)} \right] w_{t|t-1}^{(i)}$

end for

Resampling step.

Compute the total mass $\tilde{N}_t = \sum_{i=1}^{L_{t-1}+J_t} w_{t|t}^{(i)}$

Resample $\left\{ w_{t|t}^{(i)}, x_{t|t-1}^{(i)} \right\}_{i=1}^{L_{t-1}+J_t}$ to obtain $\left\{ \widetilde{w}_t^{(i)}, \widetilde{x}_t^{(i)} \right\}_{i=1}^{L_t}$

The number of particles to resample is chosen $L_t = \alpha \tilde{N}_t$

Continue by using $\left\{ \widetilde{w}_t^{(i)}, \widetilde{x}_t^{(i)} \right\}_{i=1}^{L_t}$

end for

Algorithm 3 Gaussian Mixture PHD filter (GM-PHD)

Given $\{w_{t-1}^{(i)}, m_{t-1}^{(i)}, P_{t-1}^{(i)}\}_{i=1}^{J_t=1}$ and the measurement set Y_t

Prediction of birth targets

$i = 0$

for $j = 1, \ldots, J_{\mu,t}$ **do**

$i = i + 1$

$w_{t|t-1}^{(i)} = w_{\mu,t}^{(j)}, \quad m_{t|t-1}^{(i)} = m_{\mu,t}^{(j)}, \; P_{t|t-1}^{(i)} = P_{\mu,t}^{(j)}$

end for

for $j = 1, \ldots, J_{B,t}$ **do**

for $l = 1, \ldots, J_{t-1}$ **do**

$i = i + 1$

$w_{t|t-1}^{(i)} = w_{t-1}^{(l)} w_{B,t-1}^{(l)}, \; m_{t|t-1}^{(i)} = d_{B,t-1}^{(j)} + F_{B,t-1}^{(j)} m_{t-1}^{(l)},$

$P_{t|t-1}^{(i)} = Q_{B,t-1}^{(j)} + F_{B,t-1}^{(j)} P_{t-1}^{(l)} (F_{B,t-1}^{(j)})^T,$

end for

end for

Prediction of existing targets

for $l = 1, \ldots, J_{t-1}$ **do**

$i = i + 1$

$w_{t|t-1}^{(i)} = p_{s,t} w_{t-1}^{(j)}, \quad m_{t|t-1}^{(i)} = F_{t-1}^{(j)} m_{t-1}^{(j)},$

$P_{t|t-1}^{(i)} = Q_{t-1} + F_{t-1} P_{t-1}^{(j)} (F_{t-1})^T,$

end for

$J_{t|t-1} = i$

PHD update components

for $j = 1, \ldots, J_{t|t-1}$ **do**

$\eta_{t|t-1}^{(j)} = H_t m_{t|t-1}^{(j)}, \; S_t^{(j)} = H_t P_{t|t-1}^{(j)} H_t^T + R_t$

$K_t^{(j)} = P_{t|t-1}^{(j)} H_t^T (S_t^{(j)})^{-1}, \; P_{t|t}^{(j)} = [I - K_t^{(j)} H_t] P_{t|t-1}^{(j)}$

end for

Update

for $j = 1, \ldots, J_{t|t-1}$ **do**

$w_t^{(j)} = 1 - p_{d,t} w_{t|t-1}^{(j)}, \quad m_t^{(j)} = m_{t|t-1}^{(j)}, P_t^{(j)} = P_{t|t-1}^{(j)}$

end for

$l = 0$

for each observation $y \in Y_t$ **do**

$l = l + 1$

for $j = 1, \ldots, J_{t|t-1}$ **do**

$w_t^{(lJ_{t|t-1}+j)} = p_{d,t} w_{t|t-1}^{(j)} \mathcal{N}(z; \eta_{t|t-1}^{(j)}, S_t^{(j)}),$

$m_t^{(lJ_{t|t-1}+j)} = m_{t|t-1}^{(j)} + K_t^{(j)} (z - \eta_{t|t-1}^{(j)}), \; P_t^{(lJ_{t|t-1}+j)} = P_t^{(j)}$

end for

$w_t^{(lJ_{t|t-1}+j)} = \frac{w_t^{(lJ_{t|t-1}+j)}}{h_t(y) + \sum_{i=1}^{J_{t|t-1}} w_t^{(lJ_{t|t-1}+j)}}$, for $j = 1, \ldots, J_{t|t-1}$

end for

$J_t = lJ_{t|t-1} + J_{t|t-1}$

out: $\{w_t^{(i)}, m_t^{(i)}, P_t^{(i)}\}_{i=1}^{J_t}$

Chapter 5

PHD Filters: numerical studies

This chapter concludes the first part of the dissertation by presenting a series of numerical studies on different implementations of the PHD recursion and their performance. Section 5.1 defines a test problem which is very similar to multi-target tracking instances encountered in military applications. In particular, PHD filters are applied to simulated scenarios provided by the French naval defense provider DCNS with the goal of investigating the difficulties on complex, realistic, naval and aerial filtering problems and to measure their performance.

Section 5.2 focuses on the pruning and merging step of the GM-PHD and investigates a stochastic strategy for the automatic determination of the pruning threshold.

Section 5.3 describes a novel technique for the implementation of the PHD recursion on numerical grids.

Section 5.4 presents a mean-field particle algorithm based on the sampling of the associations between particles and obsevations (enriched with virtual states).

5.1 Aerial and Naval Tracking with PHD filters

Introduction

In this section the Probability Hypothesis Density (PHD) filter is applied to realistic three-dimensional aerial and naval scenarios. The study aims at comparing the performances of different implementations of the PHD filter in scenarios that are as similar as possible to those encountered in real-applications.
For instance, in realistic environments the clutter distribution may depend on various factors, such as the geometry of the surveillance area or the physical properties of the measurent process. When this is the case, it is necessary to derive the expression of the clutter intensity in order to apply the PHD recursion. Another source of

difficulty comes from the heterogeneity of the targets which may travel or maneuver with different dynamics in regions with different physical properties. In such cases the implementation of PHD filters has to be adapted and generalized in different ways.
Section 5.1.1 and followings describe the target and clutter models and provide details on the implementation of the filters. Results are presented in section 5.1.3 and conclusions are drawn in section 5.1.3.

5.1.1 Target and measurement model

The kinematic state vector of each target (coordinates, velocity and acceleration at time t) is denoted by $x_t = [\xi_{x,t}, \xi_{y,t}, \xi_{z,t}, \dot{\xi}_{x,t}, \dot{\xi}_{y,t}, \dot{\xi}_{z,t}, \ddot{\xi}_{x,t}, \ddot{\xi}_{y,t}, \ddot{\xi}_{z,t}]^T$ and its evolution is described by an interacting multiple model (IMM) composed by a constant velocity model [65] and a constant-turn model perturbed by random accelerations.
The constant velocity model is defined as follow:

$$x_t = F_t x_{t-1} + G_t a_t \tag{5.1.1}$$

where

$$F_t = \begin{bmatrix} I_3 & T\cdot I_3 & \frac{1}{2}T^2\cdot I_3 \\ 0_3 & I_3 & T\cdot I_3 \\ 0_3 & 0_3 & I_3 \end{bmatrix}, G_t = \begin{bmatrix} 0_3 \\ 0_3 \\ I_3 \end{bmatrix}, a_t \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}\right) \tag{5.1.2}$$

and where T is the sampling period,I_n and 0_n represent the $n \times n$ identity matrix and zero matrix respectively.
The values of $[\sigma_x, \sigma_y, \sigma_z]$ are set to $[20, 20, 10]$ m/s^2 for aerial targets and to $[2, 2, 10^{-5}]$ m/s^2 for naval targets. The constant turn model is defined as:

$$x_{t+1} = \Phi_t x_t + \Gamma_t a_t \tag{5.1.3}$$

where Φ_t , Γ_t and the system noise covariance matrix are:

$$\Phi_t = \begin{bmatrix} I_3 & T\cdot(I_3 + B) & 0_3 \\ 0_3 & T\cdot(I_3 + A) & 0_3 \\ 0_3 & 0_3 & I_3 \end{bmatrix}, \Gamma_t = \begin{bmatrix} 0_3 \\ 0_3 \\ I_3 \end{bmatrix}, a_t \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}\right) \tag{5.1.4}$$

$$A = \begin{bmatrix} c_1 d_1 & -c_2\omega_z - c_1\omega_x\omega_y & c_2\omega_y - c_1\omega_x\omega_z \\ c_2\omega_z - c_1\omega_x\omega_y & c_1 d_2 & -c_2\omega_x - c_1\omega_y\omega_z \\ -c_2\omega_y - c_1\omega_x\omega_z & c_2\omega_x - c_1\omega_y\omega_z & c_1 d_3 \end{bmatrix}$$

$$B = \begin{bmatrix} c_3 d_1 & -c_1\omega_z - c_3\omega_x\omega_y & c_1\omega_y - c_3\omega_x\omega_z \\ -c_1\omega_z - c_3\omega_x\omega_y & c_3 d_2 & c_1\omega_x - c_3\omega_y\omega_z \\ -c_1\omega_y - c_3\omega_x\omega_z & -c_1\omega_x - c_3\omega_y\omega_z & c_3 d_3 \end{bmatrix}$$

$$d_1 = \omega_y^2 + \omega_z^2 \quad d_2 = \omega_x^2 + \omega_z^2 \quad d_3 = \omega_x^2 + \omega_y^2$$

$$c_1 = \frac{cos(\Omega T)-1}{\Omega^2} \quad c_2 = \frac{sin(\Omega T)-1}{\Omega} \quad c_1 = \frac{1}{\Omega^2}(\frac{sin(\Omega T)-1}{\Omega} - T)$$

where $\Omega = [\omega_x, \omega_y, \omega_z]^T$ denotes the angular velocity vector.

The radar, mounted on a ship, completes a 360-degrees scan in $\Delta T = 2$ sec. and collects measurements of targets whose distance is between a *blind distance* of 300 mt. (D_{min}) and the visibility range ($D_{max} = 100$ km) and whose elevation angle is between a maximal and a minimal value ($\theta_{\min}, \theta_{\max}$). The measurements are affected by zero-mean Gaussian random noises with known variance. The observation function represents the passage from Cartesian to spherical coordinates in the platform system of reference. Each observation at time t has the form $z_t = [r_t, \phi_t, \theta_t]^T$, where r_t represents the radial distance, ϕ_t the azimuth angle and θ_t the elevation angle of a point in the surveilled area:

$$z_t = \begin{bmatrix} r_t \\ \phi_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \sqrt{\Delta_x^2 + \Delta_y^2 + \Delta_z^2} \\ \tan^{-1}(\frac{\Delta_y}{\Delta_x}) \\ \tan^{-1}(\frac{\sqrt{\Delta_x^2+\Delta_y^2}}{\Delta_z}) \end{bmatrix} + \begin{bmatrix} w_{r,t} \\ w_{\phi,t} \\ w_{\theta,t} \end{bmatrix} \tag{5.1.5}$$

where $\mu_t = [P_{x,t}, P_{y,t}, P_{y,t}]^T$ denotes the radar position, H_{ant} the vertical position of the antenna with respect to the level of the sea, $\Delta_x = (\xi_{x,t} - P_{x,t})$, $\Delta_y = (\xi_{y,t} - P_{y,t})$, $\Delta_z = [\xi_{z,t} - (P_{z,t} + H_{ant})]$, $w_{r,t} \sim \mathcal{N}(0, \sigma_r^2)$, $w_{\phi,t} \sim \mathcal{N}(0, \sigma_\phi^2)$ and $w_{\theta,t} \sim \mathcal{N}(0, \sigma_\theta^2)$.

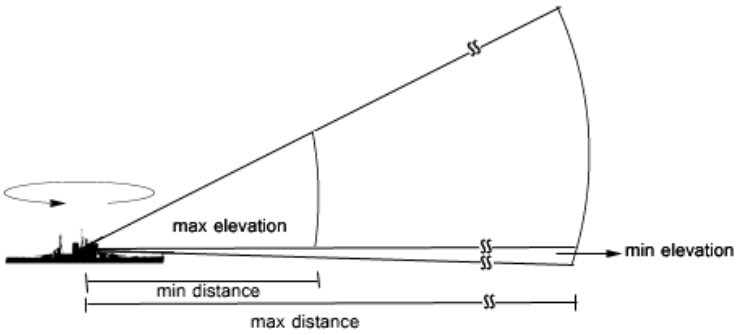


Figure 5.1: Schematic representation of the region observed by the radar.

5.1.2 Clutter Model

The clutter model takes into account the geometry of the surveilled region and, to a lesser degree, its physical properties. It is modeled as the superposition of two Poisson point processes with different intensities. The first models the false alarms

generated by the reflection of the electromagnetic beam on the surface of the sea, the second the spurious measurements generated by atmospheric noises.

Sea-clutter distribution

This kind of false measurements are distributed over a circular region around the radar and have a maximal elevation of 20 meters. Their distance is comprised between a minimal and a maximal value: D_{cmax} and D_{cmin}.
The azimuth angle is uniformly distributed in $[0, 2\pi]$ and the elevation follows a chi-square distribution with parameter $k = 1$. The clutter RFS is modelled as a Poisson point process with intensity: $\kappa_S(z) = \lambda_S \pi_S(r, \phi, \theta) = \lambda_S \pi_S^r(r) \pi_S^\phi(\phi) \pi_S^\theta(\theta)$ where $\pi_S(r, \phi, \theta)$ is the time-invariant clutter probability density over the surveillance region and λ_S the average number of clutter points per scan.

Air-clutter distribution

The false alarms of this type are distributed in the whole region of the sky observed by the radar; their azimuth, elevation and distance are sampled uniformly from the intervals $[0, 2\pi]$, $[\theta_{\min}, \theta_{\max}]$ and $[D_{\min}, D_{\max}]$. As before, the clutter RFS is modeled as a Poisson RFS with intensity: $\kappa_A(y) = \lambda_A \pi_A(r, \phi, \theta)$ where λ_A is the average number of air-clutter points per scan. The intensity of the whole process is given by: $\kappa(y) = \kappa_S(y) + \kappa_A(y)$.
An example of the realization of the clutter process over the surveillance region is given in Fig. 5.2.

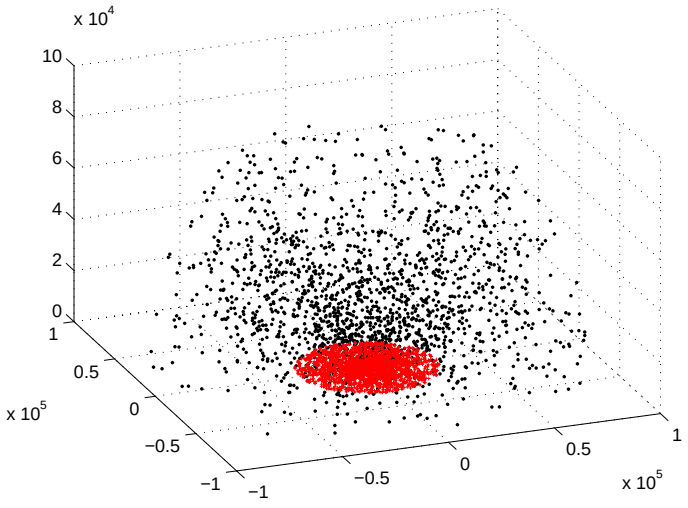


Figure 5.2: Superposition of the point processes modeling the air-clutter (black) and sea-clutter (red) in a region of 100km.

Implementation details for the SMC-PHD Filter

This section provides an overview of the SMC-PHD implementation with a special focus on the prediction step. During the prediction a pre-defined number of particles are sampled from a proposal distribution $p_t(.|Y_t)$ with the purpose of tracking new targets appearing in the scene. As the surveillance region is very large, and targets may appear at any point, sampling uniformly over the search space would be impracticable. The birth intensity particles are instead created around each observations. For instance, let $y_t^{(j)} = [r_{\mathrm{t}}^{(j)}, \phi_{\mathrm{t}}^{(j)}, \theta_{\mathrm{t}}^{(j)}]^T$, $j = 1 \dots |Y_t|$ be an observation in the measurement RFS at time t: the birth intensity particles $\zeta_{j,t}^{(i)} = [r_{j,t}^{(i)}, \sigma_{j,t}^{(i)}, \theta_{j,t}^{(i)}]$ are sampled according to $r_{j,t}^{(i)} \sim r_{\mathrm{t}}^{(j)} + \mathcal{N}(.; 0, \sigma_r^2)$, $\phi_{j,t}^{(i)} \sim \phi_{\mathrm{t}}^{(j)} + \mathcal{N}(.; 0, \sigma_\phi^2)$ and $\theta_{j,t}^{(i)} \sim \theta_{\mathrm{t}}^{(j)} + \mathcal{N}(.; 0, \sigma_\theta^2)$. The velocity component $[\dot{\xi}_{x,t}, \dot{\xi}_{y,t}, \dot{\xi}_{y,t}]$ of each particle takes into account the uncertainty on the initial velocity. The idea has been independently proposed and formalised in the recent paper [47].

Implementation Details for the GM-PHD Filter

Similarly, in the GM-PHD filter when targets may appear at any point of the surveillance region it is easier to model the birth intensity at time t by using the information provided by the observation RFS at time $t-1$ instead of approximating an uniform intensity with Gaussian terms. In this case, at each time step the Gaussian mixture approximating the birth intensity is:

$$b_t(x, Y_{t-1}) = \sum_{i=1}^{|Y_{t-1}|} w_{B,t}^{(i)} \mathcal{N}(x; \mu_{y_{t-1}^{(i)},t}, \Sigma_{B,t}^{(i)}) \tag{5.1.6}$$

where $w_{B,t}^{(k)}$ and $\Sigma_{B,t}^{(k)}$ are carefully chosen to model the intensity and the uncertainty about the evolution of new targets. Each Gaussian term is centered on a previous observation and the velocity components reflect the expected velocity of new targets.

5.1.3 Numerical results

This section reports the results of the SMC-PHD and GM-PHD filters applied to different scenarios. The detailed version of the SMC-PHD algorithm can be found in [132] while the GM-PHD algorithm is discussed in detail in [130]. The target trajectories are simulated by using the parameters provided by DCNS and the measurements are generated by perturbing the ground-truth trajectories with zero-mean Gaussian noises with standard deviations: $\sigma_\phi = 0.3\,\mathrm{deg}$, $\sigma_\theta = 0.5\,\mathrm{deg}$ and $\sigma_r = 15\mathrm{mt}$. The maximal elevation angle for the radar is $\theta_{\max} = 60\,\mathrm{deg}$ and the target detection probability is 0.98. In the case of the SMC-PHD the particles are clustered using the K-Means [71] algorithm while for the GM-PHD the estimates

are extracted by taking the locations of the Gaussian terms having a weight above a predefined threshold. The Optimal Sub-pattern Assignment (OSPA) Metric [115] with cutoff parameter $c = 5000$ is used to determine the accuracy of the filters.

Example 1

In this scenario, 10 aircrafts of the same type travel across the surveillance zone, 20 spurious measurements are registered over the surveilled region of the sky during each scan. The SMC-PHD filter is configured with 300 particles for each active target and it approximates the birth intensity allocating 100 particles to each measurement collected by the radar. The GM-PHD uses a truncation threshold of 0.4 and a merging threshold of 500mt. The results of the GM-PHD and SMC-PHD filters are shown in Figures 5.3 and 5.4.
Table 5.1 reports the average number of detected targets and the average OSPA distance using different levels of clutter and different thresholds for the GM-PHD pruning operator (P.T.). The GM-PHD provides a better estimate of the target number as well as a lower localisation error. Higher values on the OSPA distance reflect the bias introduced by the clustering procedure required by the SMC-PHD.

No clutter	**N. targets**	**OSPA**
GMPHD (P.T. = 0.4)	9.72 (0.001)	779.30 (3410.64)
GMPHD (P.T. = 0.01)	10.91 (0.001)	1236.26 (1201.90)
SMCPHD 300 part.	9.80 (0.0006)	969.66 (2093.82)
10 air-clutter	**N. targets**	**OSPA**
GMPHD (P.T. = 0.4)	9.23 (0.003)	1209.20 (3556.39)
SMCPHD 300 part.	9.08 (0.044)	1639.79 (26457.08)
20 air-clutter	**N. targets**	**OSPA**
GMPHD (P.T. = 0.4)	9.01 (0.003)	1396.24 (2105.21)
GMPHD (P.T. = 0.01)	10.85 (0.004)	1297.86 (2607.72)
SMCPHD 300 part.	8.44 (0.20)	2034.08 (71658.14)

Table 5.1: Average filtering statistics with different configuration parameters (Example 1) (Avg. and var.)

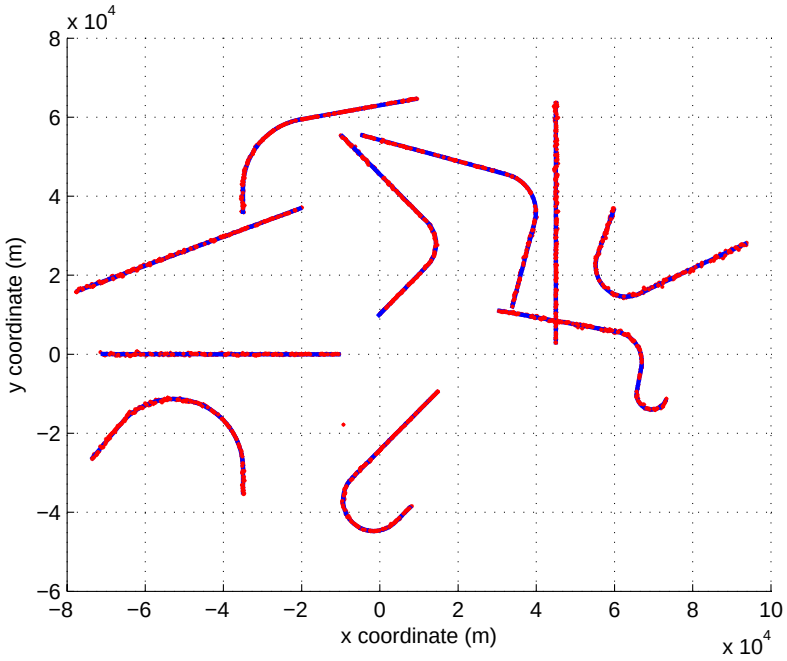


Figure 5.3: (Ex.1) GM-PHD Filter output. Target trajectories (continuous lines) and target estimates (dots).

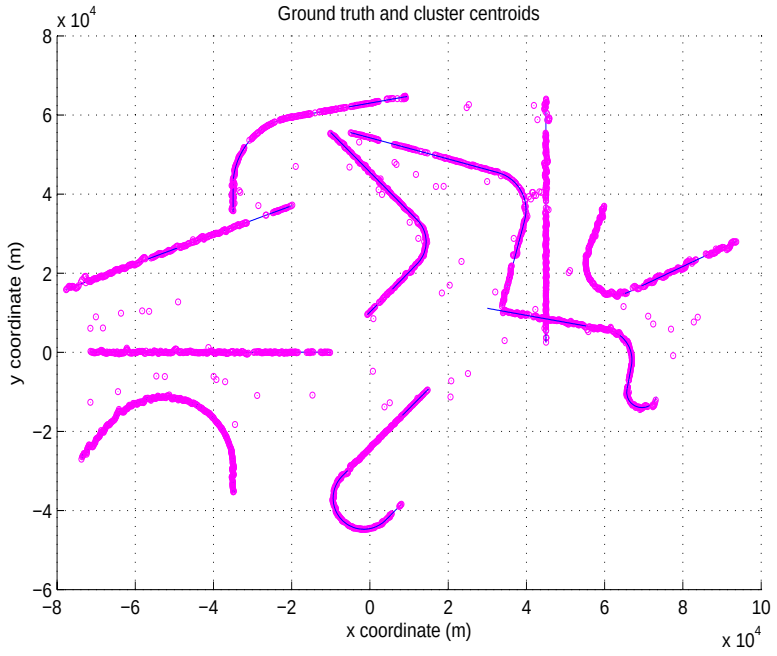


Figure 5.4: (Ex.1) SMC-PHD Filter output. Target trajectories (continuous lines) and cluster centroids corresponding to target estimates (dots).

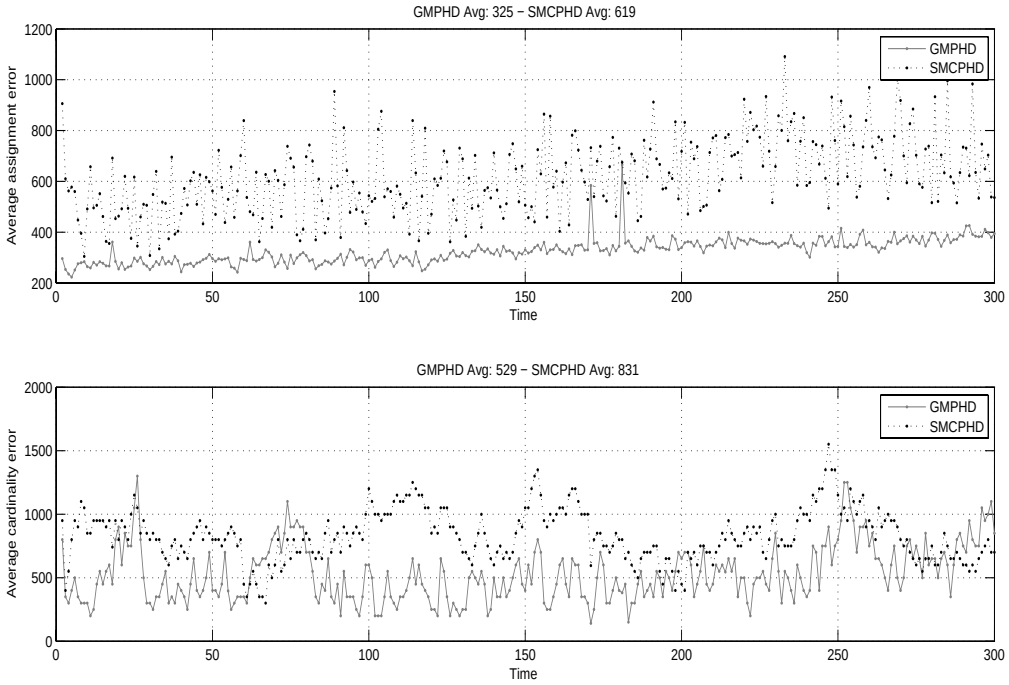


Figure 5.5: (Ex.1) OSPA localization error and cardinality error (averaged over 10 iterations). 300 time steps, 10 aircrafts, 20 air-clutter observations per scan on average, OSPA parameters $p = 2, c = 5000$.

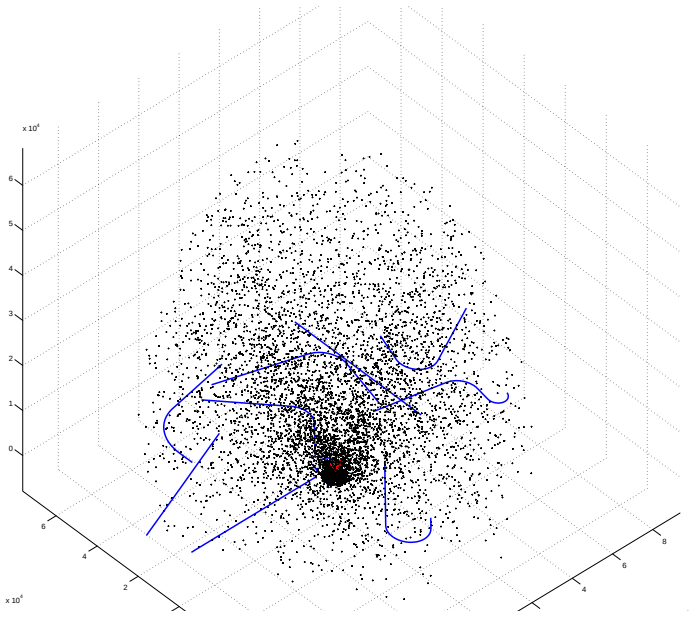


Figure 5.6: (Ex.1) 3D perspective of the filtering scenario.

Example 2

In the second example 10 naval targets are added to the previous scenario. Each one is maneuvering around the platform mounting the radar while 10 aircrafts cross the sky at different altitudes. During each scan 20 spurious measurements are registered close to the surface of the sea and 10 over the observed region of the sky. Both filters are configured with the same parameters as in example 1. The result of the filtering is reported in Figure 5.7 and 5.8. Table 5.2 reports the statistics about the number of targets filtered and the OSPA distance with different levels of clutter and different thresholds for the GM-PHD pruning operator (P.T.).

No clutter	**N. targets**	**OSPA**
GMPHD (P.T. = 0.4)	19.38 (0.006)	869.11 (4009.46)
GMPHD (P.T. = 0.01)	21.14 (0.002)	1041.02 (1030.92)
SMCPHD	19.59 (0.0005)	1169.95 (1138.31)
20 sea-clutter, 10 air-clutter	**N. targets**	**OSPA**
GMPHD (P.T. = 0.4)	18.26 (0.003)	1394.68 (644.35)
GMPHD (P.T. = 0.01)	21.41 (0.015)	1227.67 (2129.91)
SMCPHD 300 particles	18.887 (0.007)	1505.78 (1154.87)

Table 5.2: Average filtering statistics with different configuration parameters (Example 2) (Avg. and var., 10 iterations)

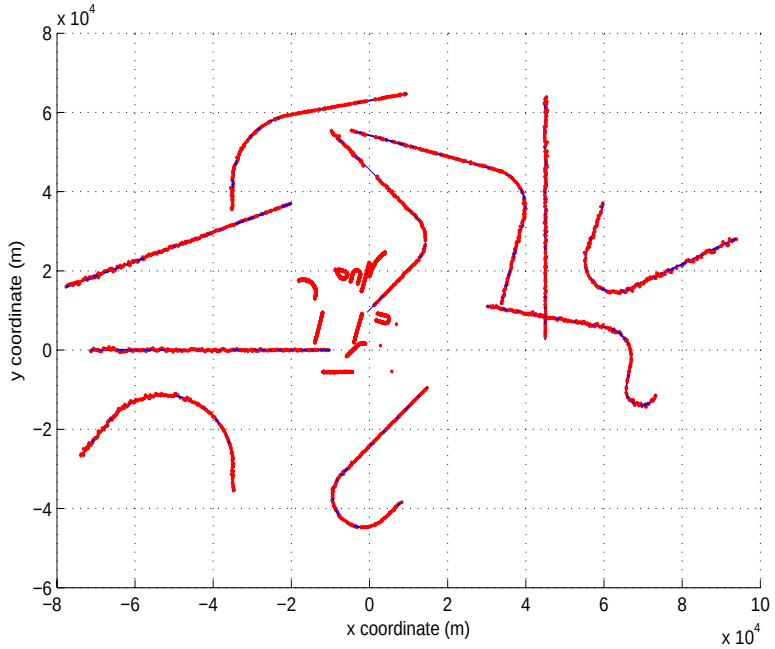


Figure 5.7: (Ex.2) GM-PHD Filter output. Target trajectories (continuous lines) and target estimates (dots).

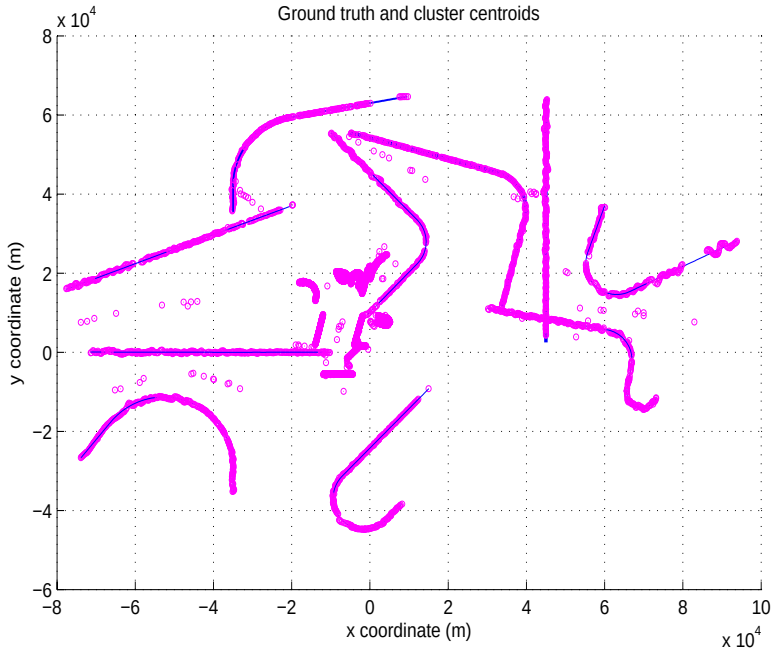


Figure 5.8: (Ex.2) SMC-PHD Filter output. Target trajectories (continuous lines) and cluster centroids corresponding to target estimates (dots).

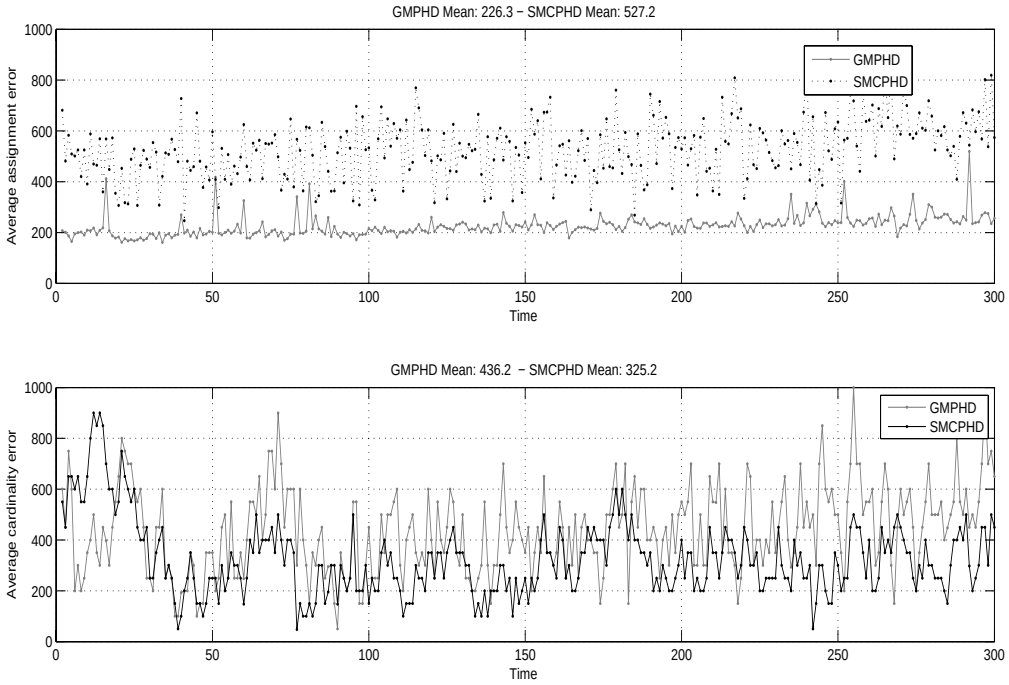


Figure 5.9: (Ex.2) OSPA localization error and cardinality error (averaged over 10 iterations). 300 time steps, 10 aircrafts, 10 naval targets, 20 sea-clutter and 10 air-clutter observations per scan on average, OSPA parameters $p = 2, c = 5000$.

Example 3

In the third example the filtering performance and the execution time is measured on a scenario with 99 naval targets maneuvering in a region of 30km around the platform. 50 spurious measurements are registered during each scan. Due to clutter intensity and to the high number of active targets the filtering of this scenario was impractical with our Matlab implementation of the SMC-PHD filter. On the contrary the Matlab version of the GM-PHD filter had no problem in processing it on a normal personal computer. The scenario is reported in Figure 5.10. The clutter process is not shown to preserve the readability of the figure. A zoom of the surveillance region in the proximity of the radar is shown in Figure 5.11; the circular area of space around the origin without observations corresponds to the radar's blind distance reagion. The radar is configured with the same parameters as in examples 1 and 2. Figure 5.12 shows the estimate of the number of targets during each time step. Despite the high number of clutter points the filter is able to detect, on average, 94.7 targets on the 97 detected on average. The average assignment error was 366 m. which can be considered quite satisfactory given that only one radar is collecting angular measurements of targets distant up to 30km. The assignment error is computed by measuring the average distance between the mean of the Gaussian terms in the posterior intensity and the ground truth positions. A test machine equipped with an Intel 2.4 Ghz processor took 15 seconds on average to process each simulated scan. The time required to run the whole simulation (650 time steps) was 161 minutes.

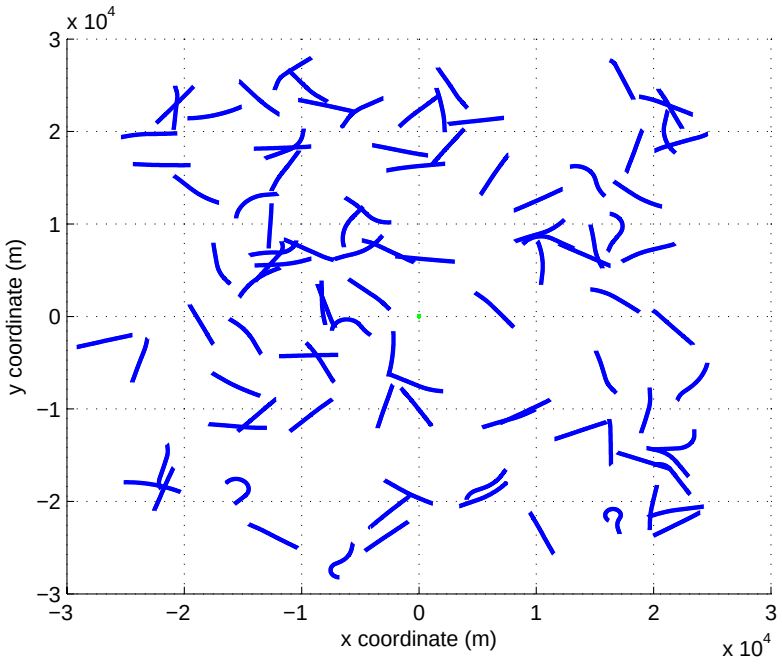


Figure 5.10: (Ex.3) Filtering scenario: 99 naval targets (continuous lines) maneuvering in the surveillance zone.

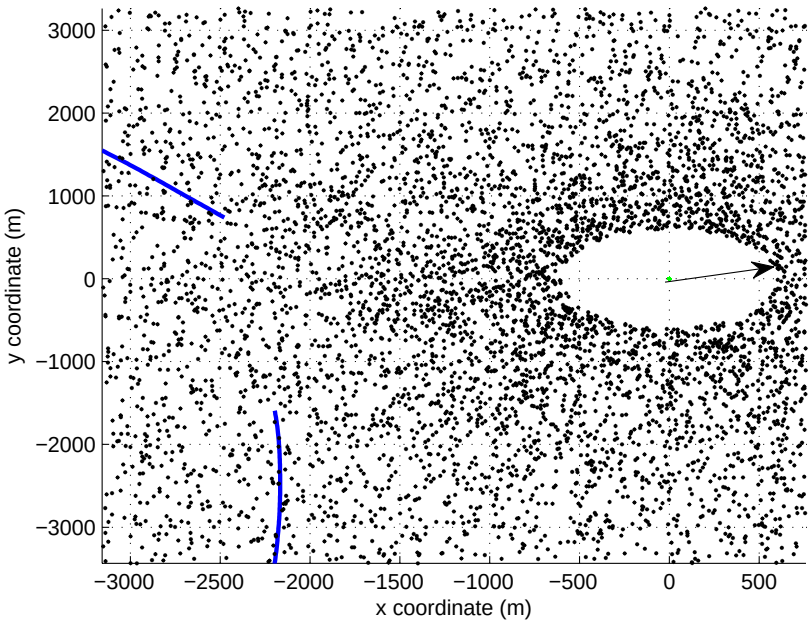


Figure 5.11: (Ex.3) Zoom over the surveillance region near the radar (at the origin). The black arrow indicates the radar's blind distance.

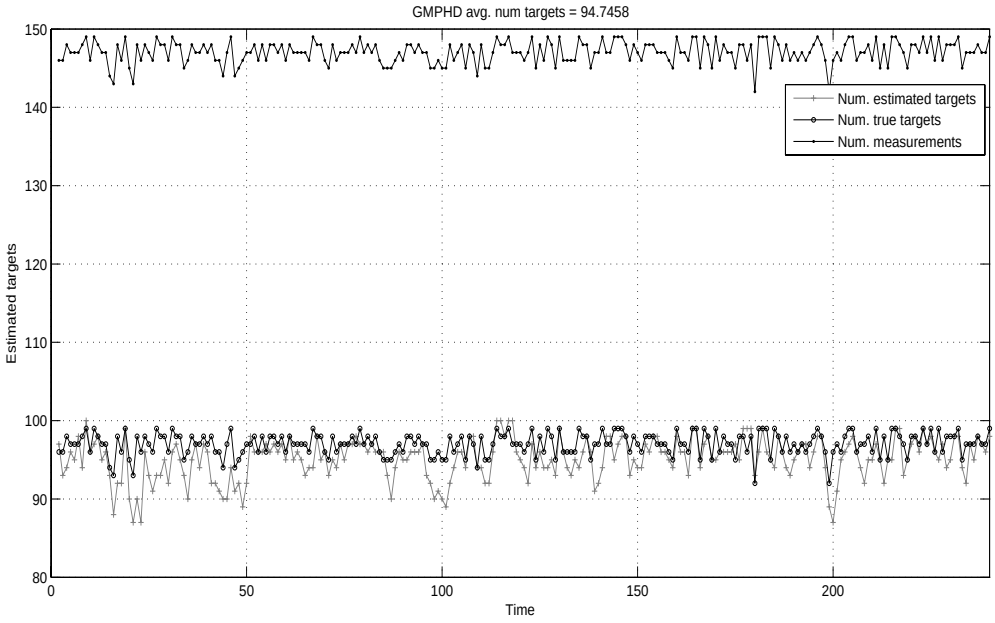


Figure 5.12: The target detection probability is 0.98. The RFS of the observations contains on average 147 measurements. The GM-PHD filter is able to maintain an accurate estimate of the number of targets.

SMC-PHD filter performance w.r.t. the number of particles

The number of particles used in the sequential Monte Carlo PHD filter to approximate the intensity function influences strongly the quality of the result. To evaluate how the filtering performance depends on it, the scenario of example 1 is repeatedly filtered by using an increasing number of particles. Ten iterations are made for each value, starting from 50 particles per active target, up to 2000. The average number of targets filtered and the localisation errors are reported in Figures 5.13 and 5.14. As it is natural to expect, when the number of particle increases the filter becomes much more accurate in the identification of active targets and the overall filtering error reduces. Nevertheless, the computational cost increases linearly on the number of particles (Figure 5.15) and the GM-PHD is generally able to provide good results at a fraction of the computational cost required by the SMC-PHD. The simulation with 2000 particles per target took about 250 minutes to complete on the test machine, and the result is only slightly better than the one obtained with the GM-PHD which required less than 3 minutes.

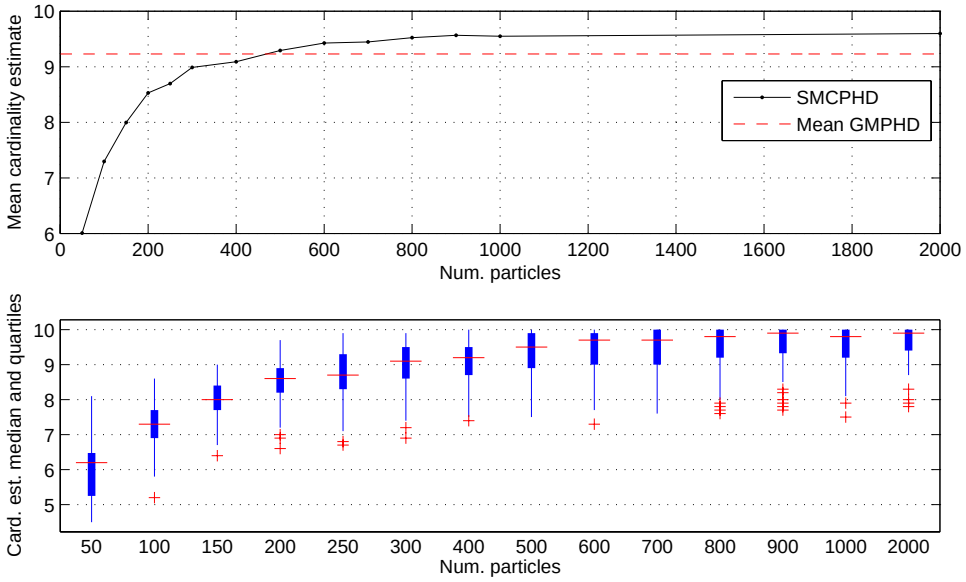


Figure 5.13: Average estimation of the number of targets estimated by the SMC-PHD filter with an increasing number of particles. The average value provided by the GM-PHD is shown as a dashed line.

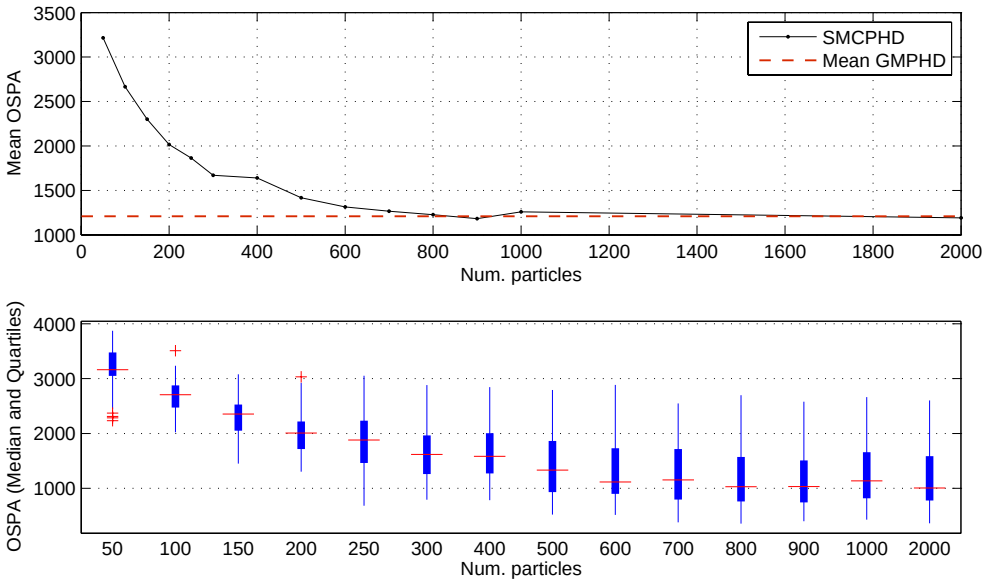


Figure 5.14: Quality of the filtering provided by the SMC-PHD with an increasing number of particles, measured using the average OSPA distance. The average value provided by the GM-PHD is shown as a dashed line.

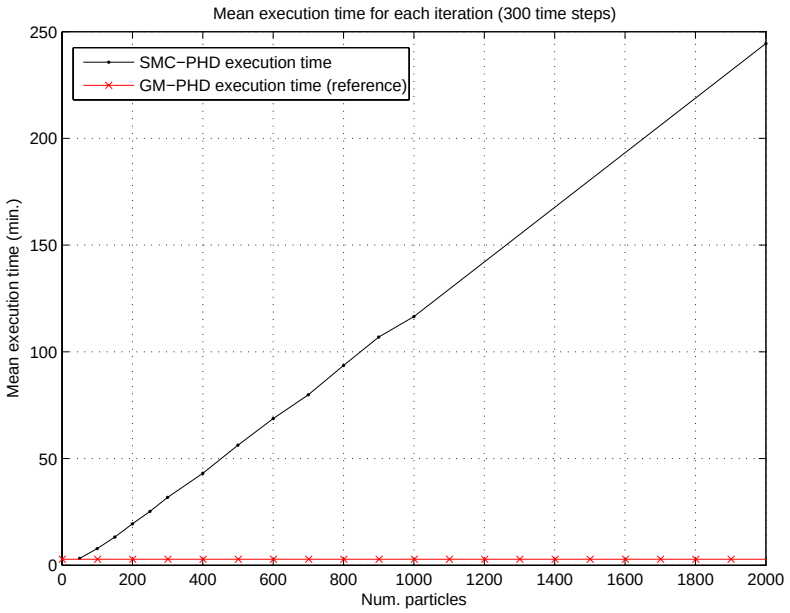


Figure 5.15: Processing time required with an increasing number of particles. The SMCPHD filter is implemented in Matlab and the simulation is run on a Intel 2.4 Ghz processor. As a reference the time required by the GM-PHD filter to process the same scenario (2.8 min) is reported.

Conclusions

In the simulations, the GM-PHD filter outperformed the SMC-PHD in both estimate quality and computational cost. The time required by the SMC-PHD to filter a scenario was always substantially larger than the time required by the GM-PHD. The reason is mainly because the GM-PHD applies pruning and merging techniques to reduce the number of Gaussian terms in the posterior intensity and does not require clustering to extract target estimates.
The computational cost is a fundamental aspect in practical applications, where new measurements may be available every few seconds and filters should provide a result as soon as possible. Even if the GM-PHD constitutes a closed form for the PHD recursion only for linear Gaussian models, its extension using the Unscented Transform makes it applicable to targets having a more complex and non-linear dynamic. In the tests, when a considerable number of targets is used, the number of particles required by the SMC-PHD to obtain an acceptable filtering quality rapidly becomes prohibitive. It is also important to note that both filters need a very careful choice of the parameters and it is generally easy to have one filter working better than one other if a better choice of the parameters is done; furthermore, a good knowledge of the clutter process and target dynamic is important to obtain good results. Due to the high number of configuration parameters, the process of adapting the filter may require a careful fine-tuning.

5.2 Stochastic pruning strategy for the GM-PHD

Introduction

In the GM-PHD filter [130] discussed in section 4.8, the PHD intensity is represented by mixture of m_t Gaussian terms. The number of terms increases according to the relation:

$$m_t = (m_{t-1} + m_{\gamma,t})(1 + |Y_t|) \tag{5.2.1}$$

where $|Y_t|$ denotes the number of measurements at time t and $m_{\gamma,t}$ the number of gaussians modelling the birth intensity. It is therefore necessary to make approximations.
The general approach consists in keeping the number of Gaussian terms bounded by using two steps commonly called *pruning* and *merging*. Merging consists in fusing the Gaussians whose Mahalanobis distance is below a given threshold, while the pruning step eliminates the Gaussians with low weights. The error made by using this strategy has been studied in [24]. Clearly, the pruning strategy is biased.
In the following sections we study an unbiased resampling scheme that leaves a chance to low-weighted Gaussians to be selected; the algorithm has been proposed by P. Clifford and P. Fearnhead and originally applied to the problem of estimating break points in well-log data [41]. Unfortunately however, the application of this algorithm in the pruning step of the GM-PHD does not provide satisfactory results and generally underperforms biased pruning strategies.

A practical track-management issue related to the pruning and merging step of the GM-PHD that has been noticed during numerical tests is the disactivation of tracks caused by repeated misdetection. In this case repeated misdetections cause certain Gaussian terms to be incorrectly eliminated. Section 5.2.3 proposes a multitarget-tracker adapted pruning strategy developed to address this problem.

5.2.1 Optimal resampling GM-PHD

The algorithm studied in this section is based on the determination of a threshold that is used to discriminate the Gaussian terms to accept during the pruning step. We assume in the following that we want to keep the overall number of elements at time t equal to a fixed number N.
Assume that in the updated PHD intensity we have M Gaussian terms with weights $w_{t,k}$, $k = 1, \ldots, M$ and we want to obtain M Gaussian terms with weights $\widetilde{w}_{t,k}$ such that only N among them have non-zero weights. As introduced before, the deterministic approach consists in fusing the Gaussian terms whose Mahalanobis distance is less than a given threshold U, and in keeping the N Gaussians with highest weights. Assume that we rank the weights after the Gaussians are fused,

then we have:

$$\begin{cases} \widetilde{w}_{t,k} = w_{t,k} & \text{If rank}(w_{t,k}) \leq N \\ \widetilde{w}_{t,k} = 0 & \text{Otherwise} \end{cases}$$

This strategy is biased as $\mathbb{E}[\widetilde{w}_{t,k}] \neq w_{t,k}$ for low weights.

The Fearnhead-Clifford resampling scheme is instead unbiased since it leaves a chance to some low-weight Gaussians to survive. Basically, it automatically sets a threshold calculated by satisfying to the following properties [41]:

(a) $\mathbb{E}[\widetilde{w}_{t,k}] = w_{t,k}$ (unbiasedness)

(b) The support of $\widetilde{w}_t$ has no more than N points

(c) $\mathbb{E}\left[\sum_{k=1}^{M}(\widetilde{w}_{t,k} - w_{k,t})^2\right]$ is minimized

Algorithm 4 describes its main steps; the running time is $O(N)$.

Algorithm 4 Fearnhead-Clifford resampling

- Let $w^{\star}_{t,k}$ be the normalized weights at time t. Calculate the unique solution c of $N = \sum_{k=1}^{M} \min(w^{\star}_{t,k}, 1)$
- For $k = 1, \ldots, M$, if $w^{\star}_{t,k} > 1/c$ then place it in set 1; otherwise place it in set 2. Assume there are L elements in set 1.
- Use the stratified sampling of Carpenter et al. [14] to resample $N-L$ Gaussians from set 2. The expected number of times that each particle is resampled is proportional to its weight.
- The new set of Gaussians consists of the L elements in set 1, each given its original weight, and the $N - L$ elements in set 2, each assigned a weight $1/c$.

5.2.2 Monodimensional multitarget model

For illustration, consider a mono-dimensional scenario with an unknown and time varying number of targets observed in clutter over the surveillance region [-100,100]. The state of each target $x_t = [p_t, \dot{p}_t]$ consists of position and velocity, while the measurement is a noisy observation of the position component. Each target has a survival probability $p_{S,t} = 0.995$, a probability of detection $p_{D,t} = 0.95$ and follows a linear Gaussian dynamics. The process model is as follows:

$$x_{t+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ T \end{bmatrix} u_t \tag{5.2.2}$$

with $u_t \sim \mathcal{N}(0, \sigma_u^2)$ and $\sigma_u = 0.7$. Each target, if detected, generates an observation according to:

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + v_t \tag{5.2.3}$$

The sampling period T = 1 and $v_t \sim \mathcal{N}(0, \sigma_v^2)$ with $\sigma_v = 1$. The birth intensity is defined as $\gamma_b = 0.35\mathcal{N}(.; x_b, Q_b)$ where $x_b = [0, 0]^T$ and

$$Q_b = \begin{bmatrix} 20 & 0 \\ 0 & 1 \end{bmatrix}$$

The clutter is Poisson with intensity λ_c. If a target exits from the surveillance region it is eliminated from the simulation. Each simulation lasts 100 time steps and to avoid short-living targets the latest birth-time is set to the 30th time step. While the structure of this model is simple, the clutter intensity and the number of misdetections make the filtering of the scenarios non trivial. Figure 5.16 shows a realization of a scenario.

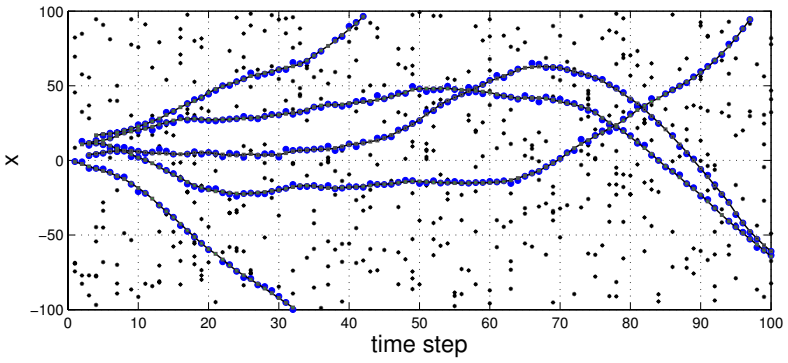


Figure 5.16: Test scenario generated with the model described in Section 5.2.2.

Reference function

In order to assess the effect of different resampling strategies we consider the L2 norm of the error function between the posterior PHD intensity $\gamma_{t|t}$ provided by the GM-PHD filter, and a reference function γ_{ref}:

$$E_{(\gamma_{t|t}, \gamma_{ref})} = \left(\int_S |\gamma_{t|t} - \gamma_{ref}|^2 d\gamma \right)^{\frac{1}{2}} \tag{5.2.4}$$

The integral is computed numerically on a bidimensional grid where S covers a sufficiently large subset of the domain. One natural choice for the reference function would be the posterior intensity obtained without pruning and merging, but the computational burden of this approach leads to the necessity of finding good substitutes. In order to compare possible reference functions, it is possible to consider

the complete posterior on a limited number of time steps. Table 5.3 reports the number of Gaussian terms and the mass of the PHD intensity on a test scenario with two targets and clutter intensity $\lambda_c = 5$ when no merging and no pruning is used. Despite their rapid explosion, most of the terms give no contribution at all, as their weights fall under machine precision rapidly. The elimination of these terms by using a small pruning threshold (for example 10^{-10} as in Table 5.4), determines a sharp reduction in the number of modes and an error under machine precision. When merging is used, their number decreases even more, but errors begin to become noticeable. Table 5.5 reports the errors for the first 6 time steps by using the complete intensity as reference function.

Time step	**N.Gaussian Terms**	**I.**
T=1	8	0.07498
T=2	36	0.45724
T=3	407	1.0614
T=4	2448	1.5245
T=5	12245	1.4247
T=6	73476	1.8215

Table 5.3: Number of Gaussian terms and integral of the intensity on a test scenario with 2 targets and $\lambda_c = 5$. No pruning or merging are used.

Time	**PT** 10^{-10}	**I.**	**PT** 10^{-10} **mer. 5**	**I.**
T=1	8	0.07498	4	0.07498
T=2	20	0.45724	5	0.45697
T=3	79	1.0614	14	1.3956
T=4	171	1.5245	19	1.5632
T=5	311	1.4247	19	1.3206
T=6	612	1.8215	26	1.793
T=7	1227	2.5848	28	2.3956
T=8	2333	2.9292	25	2.9556
T=9	5622	2.9972	35	3.031
T=10	8327	2.1249	36	2.1434

Table 5.4: Number of Gaussian terms (column 2 and 4) and integral of the posterior PHD with pruning threshold (PT) and merging threshold (MT).

The computational cost of building reference functions by using only pruning is still too high, especially for Monte Carlo validations. Results using different merging thresholds suggest that a good compromise between the computational cost and the error is obtained by using a pruning threshold $PT = 10^{-10}$ and a merging threshold $MT = 1$.

Time	**err PT** 10^{-10}	**err. PT** 10^{-10}**, MT=5**
T=1	1.1065e-038	2.2738e-008
T=2	2.456e-034	0.001125
T=3	7.9324e-021	0.0062382
T=4	1.3978e-019	0.02316
T=5	4.652e-019	0.04974
T=6	5.311e-018	0.060487

Table 5.5: Errors w.r.t the reference function of Table 5.3.

5.2.3 Comparison of pruning and merging strategies

The resampling methods which are evaluated are:

- *N-Best*:
  After the GM-PHD update, the N terms with greater weight are kept. All the other terms are eliminated. The resulting Gaussian mixture is used in the next step.

- *Threshold*:
  After the GM-PHD update, all the terms whose weight is above a pre-defined threshold are kept.

- *Fearnhead-Clifford resampling*:
  The Fearnhead-Clifford resampling is used to resample N terms from the updated mixture. $N = \{5, 10, 20\}$. If the mixture has less than N terms, all of them are kept.

The results are obtained by processing randomly-generated scenarii with a number of targets $T \in \{2, \ldots, 5\}$, and clutter intensity parameter $\lambda_c \in \{2, \ldots, 5\}$. The reference function is computed for each scenario by using a pruning threshold of $PT = 10^{-10}$ and a merging threshold of $MT = 1$. The average number of terms in the PHD approximation and the average error for the different resampling methods are reported in Figures 5.17 and 5.18. Results show that deterministic methods always outperforms the stochastic resampling in terms of computational cost and quality of the results. The use of a pruning threshold generally provides better results, but the threshold has to be chosen carefully, by considering the clutter level and the misdetection probability. The performance of the threshold strategy is also influenced by the birth intensity (Figure 5.19 and 5.22). Similar conclusions has been reported in [130] and [100] where it is observed that selecting the highest peaks (corresponding to the N-Best resampling) may generate unreliable state estimates due to terms with very small weights.
As we measure the error w.r.t a reference function the better result obtained by setting a threshold is due to the contribution of all the additional terms that are

sufficiently great to be above the pruning threshold but not enough to be selected by the N-Best. Figures 5.17 and 5.18 report the average number of terms which are selected by the different methods and the average errors on a Monte Carlo simulation.
The Fearnhead-Clifford resampling and the N-Best strategy maintain the computational cost bounded as they produce a posterior PHD with a fixed number of terms. As the targets may disappear and are eliminated when they exit from the surveillance zone, their number tends to decrease over time, and so does the average error.
The stratified resampling used in the Fearnhead-Clifford algorithm reduces the performance in a way that the error introduced by using a parameter N is comparable with the error of N-Best with parameter $N/2$.

A disappointing result is indeed the poor performance of the Fearnhead-Clifford resampling compared to the simple deterministic method. The unbiasedeness of the choice of the Gaussian terms at a greater computational cost does not provide any advantage in terms of the approximation error. Fig. 5.20 plots the weights of the updated Gaussian terms at different time steps and the Fearnhead-Clifford threshold. Once again, after the update step most of the terms have very low weights, which cause a waste of computational time if not correctly pruned.

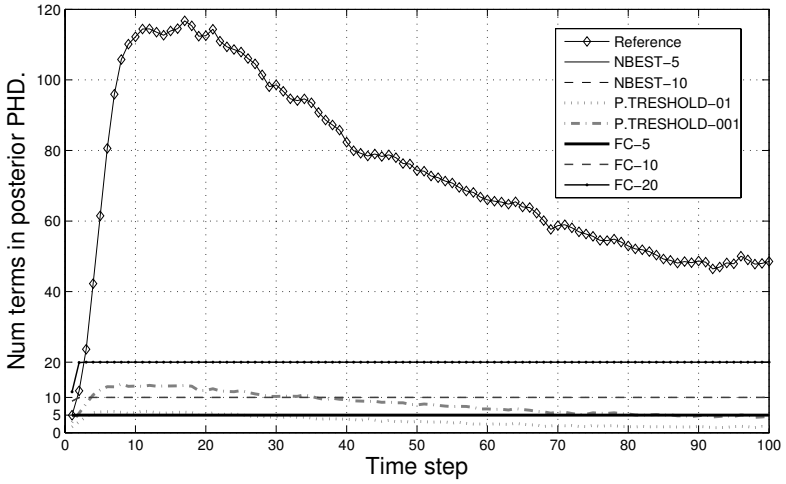


Figure 5.17: Number of gaussian terms in the posterior PHD with different resamplig strategies (100 iterations, birth intensity 0.35)

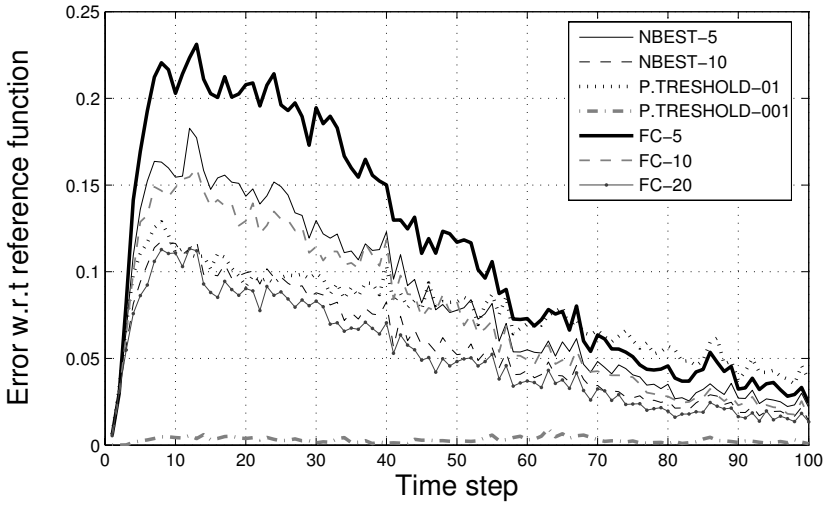


Figure 5.18: Error in the posterior PHD with different resampling strategies w.r.t the reference function. (100 iterations, birth intensity 0.35)

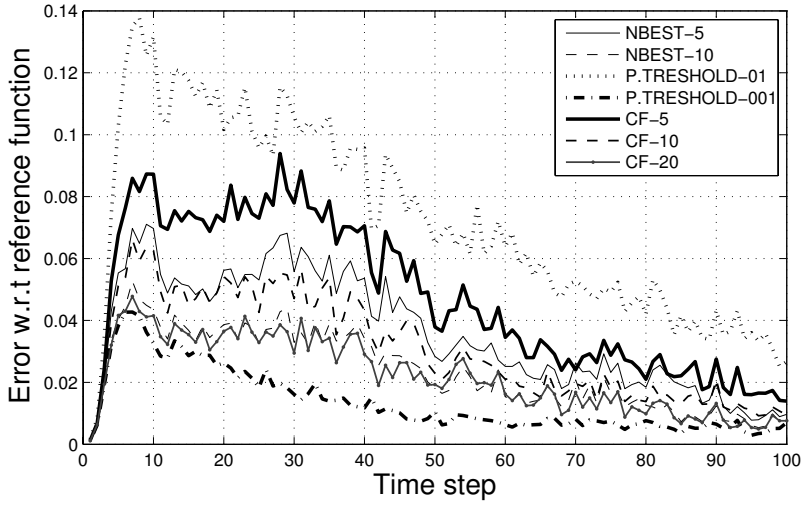


Figure 5.19: Error in the posterior PHD using different resamplig strategies w.r.t the reference function. (100 iterations, birth intensity 0.15)

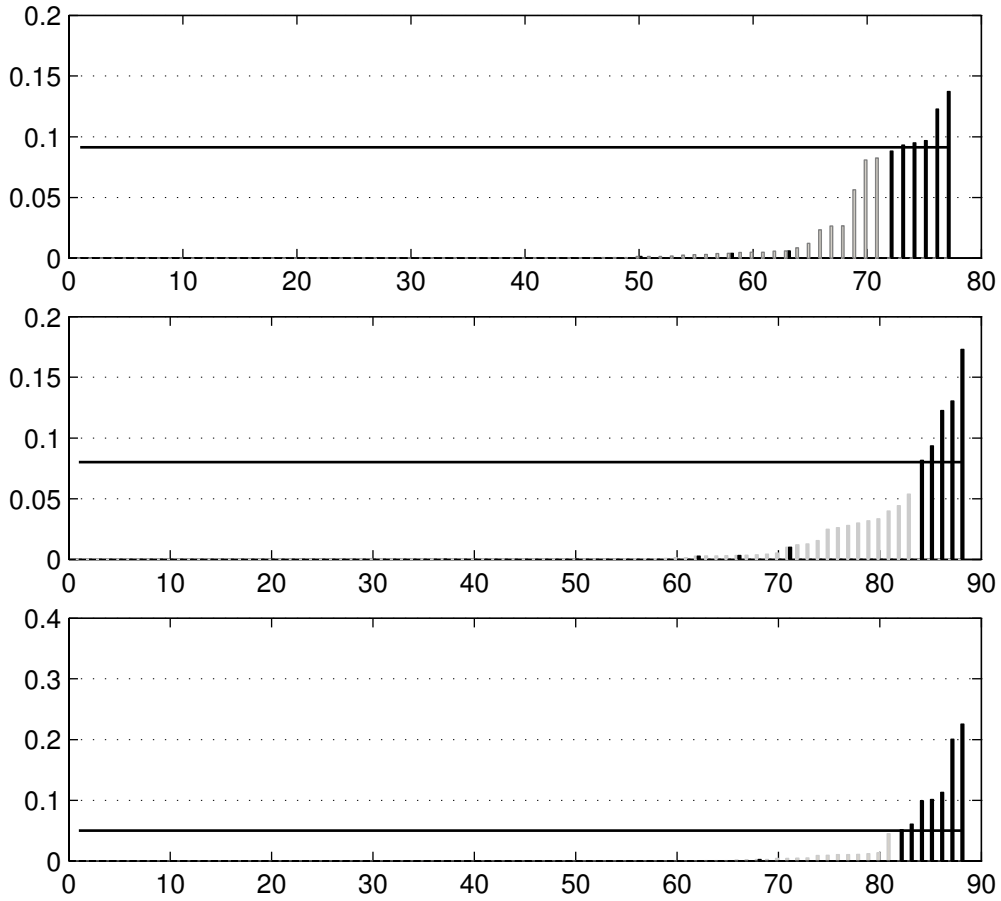


Figure 5.20: Fearnhead-Clifford threshold and weights of the terms in the posterior Gaussian mixture at different time steps.

Resampling efficiency

As the error is an inverse measure of the quality of the posterior and the cost is proportional to the number terms, we measure the resampling efficiency with $\eta = 1/(e_t N_t)$, where e_t is the error w.r.t the reference function and N_t the number of Gaussian terms in the posterior PHD. If a method produces no terms in the posterior the efficiency is not computed. Figure 5.21 reports the efficiencies in a logarithmic scale for the methods discussed. The deterministic acceptance of all the terms above a threshold provides the best cost-quality ratio. The Fearnhead-Clifford resampling has, on the contrary, a poor performance.

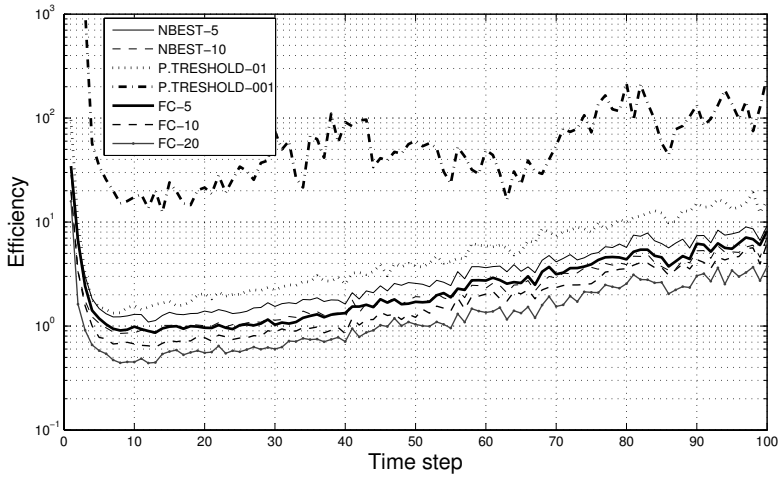


Figure 5.21: Efficiency of different resamplig strategies w.r.t the reference function. (100 iterations, birth intensity 0.35)

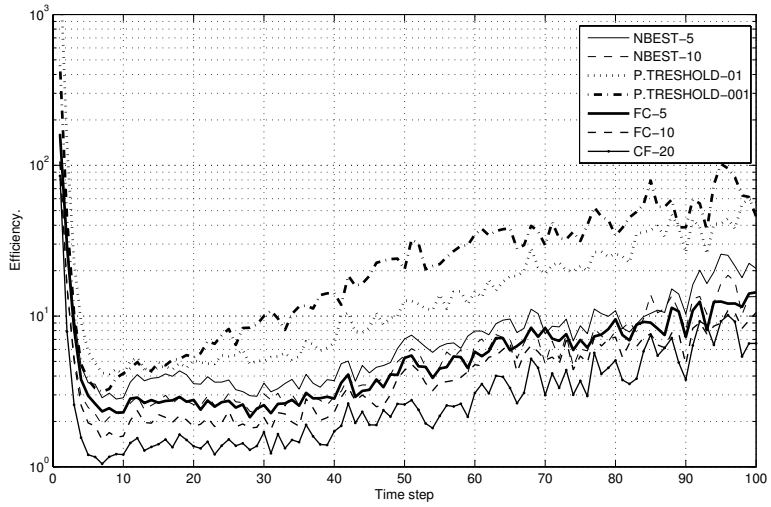


Figure 5.22: Efficiency of different resamplig strategies w.r.t the reference function. (100 iterations, birth intensity 0.15)

Multitarget-tracker adapted pruning

As the GM-PHD filter does not provide identities of individual state estimates, techniques to construct the tracks from the posterior mixture are required.
These techniques as well as different implementations of multi-target trackers have been investigated in [102, 48, 20].
Multi-target trackers maintain at each time step a list of confirmed targets as well as the labels of the Gaussian terms that have generated the estimates. In order to perform the peak-to-track association as a post processing operation the GM-PHD recursion is modified to maintain and propagate the labels associated to the Gaussian terms. The information about confirmed tracks can be exploited during the GM-PHD resampling to alleviate the loss of targets in case of misdetection. The rapid fall of the weights below the pruning threshold, however, may cause the elimination of an active target and a delay of several time steps before its reactivation.
These errors can be avoided by preventing the Gaussian terms to be pruned if they correspond to confirmed tracks until they reach a more conservative deactivation threshold. In order to evaluate this target-tracker-adapted pruning criterion we consider the Optimal Sub-Pattern Assignment (OSPA) Metric with cutoff parameter c = 20. Figure 5.24 reports the average OSPA distance and the average cardinality error in a Monte Carlo evaluation of 500 scenarios. The target detection probability has been lowered to $p_{d,t} = 0.9$, the probability of target survival is $p_{s,t} = 0.995$ and the birth intensity is a Gaussian mixture composed by three terms centered

in $[-50, 0]^T, [0, 0]^T$ and $[50, 0]^T$ with $\sigma_p = 20$ and $\sigma_v = 1$. The weights of birth intensity's terms are lowered to 0.15 to make it more difficult for the GM-PHD filter to catch new targets or to recover from misdetection.
Each simulation contains a random number of targets $N \in \{2 \dots 5\}$, and a Poisson number of false observations with mean $\lambda_c \in \{2 \dots 5\}$ uniformly distributed over the surveillance zone. Pruning threshold is set to 0.1; targets are extracted from the posterior Gaussian mixture if their weights are above 0.5.
By using the information about the active targets provided by the multi-target tracker, it is possible to avoid the pruning of Gaussian terms corresponding to confirmed tracks until their weights reach the elimination threshold (10^{-3} in the simulation).

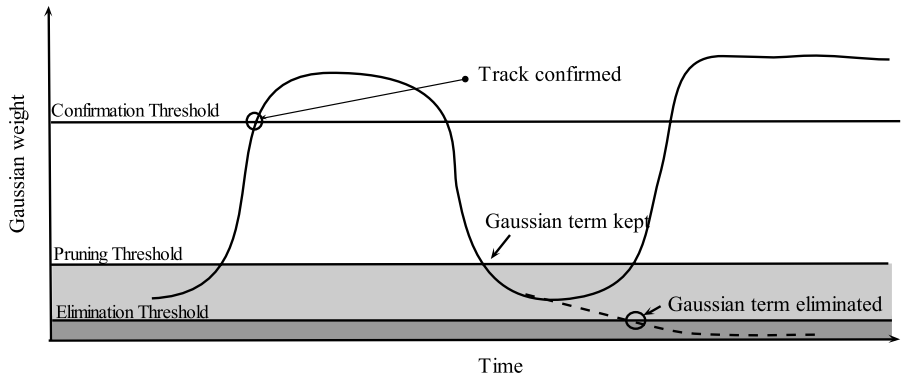


Figure 5.23: Schematic representation of the target confirmation and elimination events triggered by the growth of the Gaussian weight. An elimination threshold is added in order to avoid the pruning of confirmed terms until they reach a more conservative level.

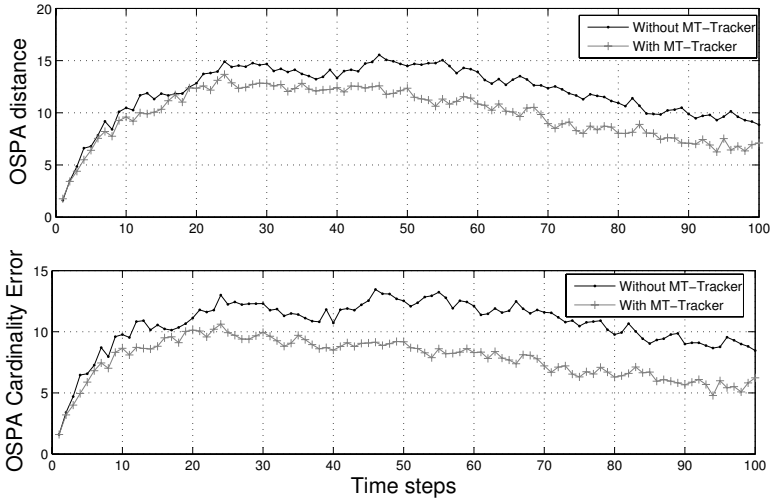


Figure 5.24: OSPA error and cardinality error in 500 simulated scenarios using a target-tracker-adapted pruning criterion.

Conclusions

This section evaluated a stochastic resampling method proposed by Fearnhead and Clifford [41] for the determination of the pruning threshold in the GM-PHD filter. Monte Carlo validations on different scenarios and with different parameters have demonstrated that deterministic strategies always outperform the Fearnhead-Clifford resampling in approximating the posterior PHD Gaussian mixture.
One problem of the deterministic algorithms is that when a target is undetected the weight of the corresponding term decreases rapidly and once the target is deleted it may take several time steps to be reactivated, especially if the clutter process is intense and the detection probability relatively low. In order to mitigate the problem we proposed a method which exploits the information on confirmed tracks in order to postpone the elimination of Gaussian terms until their weights reach a value corresponding to the almost sure target disappearance. Monte Carlo validations confirmed an improvement of the results when this strategy is adopted.

5.3 PHD quantization by using the Fast Fourier Transform

This section considers a method to calculate the Probability Hypothesis Density function over a grid by using the convolution method and the Fast Fourier Transform. Unlike the Sequential Monte Carlo PHD filter and the Gaussian Mixture PHD Filter this method provides an exact representation of the PHD function over the discretized domain. Moreover, it doesn't require Gaussian assumptions on the target dynamics and on the observation model. Section 5.3.1 outlines the technique of convolution proposed in [145] and discusses the application of the Fast Fourier Transform to filtering problems. The algorithm is presented in section 5.3.2; comparisons with the GM-PHD filter are presented in section 5.3.3 and conclusions discussed in section 5.3.3.

5.3.1 Single target filtering via convolution

We begin by describing the approach on a single target filtering problem defined by the following system, similar to (3.2.13):

$$x_{t+1} = F_t x_t + C_t u_t + v_t \tag{5.3.1}$$

$$y_t \quad = g(x_t) + w_t \tag{5.3.2}$$

where t is the discrete time index, x_t represents the target state-vector at time t, y_t the observation vector, u_t the input vector in a controlled environment, F_t and C_t the system and control matrices respectively and v_t and w_t the process and observation noises. The densities of the noise components are denoted by $\phi(x)$ and $\psi(x)$ respectively and are assumed to be time-independent. Moreover, the general assumptions required to derive the Bayes filter are considered verified: the target dynamics is described by a Markov process $p(x_{t+1}|x_{0:t}) = p(x_{t+1}|x_t)$, and the observations are mutually independent. We also assume that the matrix F_t is invertible.

The first step towards the construction of the numerical algorithm is to replace the continuous domain by a discrete domain. Let Ω_t^d be the uniform d-dimensional regular grid at time t and $(\Delta x_1, \ldots, \Delta x_d)$ the discretization steps along the $(1, \cdots, d)$ dimensions. A point $\bar{\mathtt{x}}^{\mathtt{j}}$ of the grid Ω_t^d is defined by $\bar{\mathtt{x}}^{\mathtt{j}} = (x_1^j \Delta x_1, \ldots, x_d^j \Delta x_d)$, where the coordinates $(x_1^j, \ldots, x_d^j)$ are integers. A function $f : \mathbb{R}^d \to \mathbb{R}$ can then be discretized over the points of the grid Ω_t^d by posing:

$$\bar{f}^j \triangleq f(\bar{\mathtt{x}}^{\mathtt{j}}), \quad \forall j \in \Omega_t^d$$

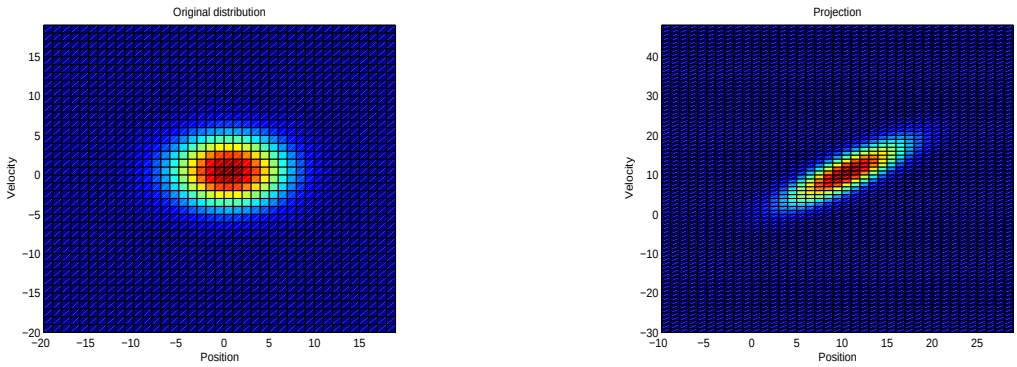
With this notation let $\varphi_{t|t}(x)$ denote the conditional density of the random variable x_t given the observations up to time t and $\bar{\varphi}_{t|t}^j = \varphi_{t|t}(\bar{\mathtt{x}}^{\mathtt{j}})$ its discretization over Ω_t^d,

such that $\sum_{j\in\Omega_t^d} \bar{\varphi}_{t|t}^j = 1$. The second step consists in calculating the distribution of the r.v. $x_{t+1}^- \triangleq F_t x_t + C_t u_t$, obtained by equation (5.3.1) without the noise component. Its pdf is:

$$\begin{aligned} \varphi_{t+1|t}^-(dx) &= \mathbb{P}(x_{t+1}^- \in dx) && (5.3.3)\\ &= \mathbb{P}(F_t x_t + C_t u_t \in dx) && (5.3.4)\\ &= \mathbb{P}(x_t \in F_t^{-1}(dx - C_t u_t)) && (5.3.5) \end{aligned}$$

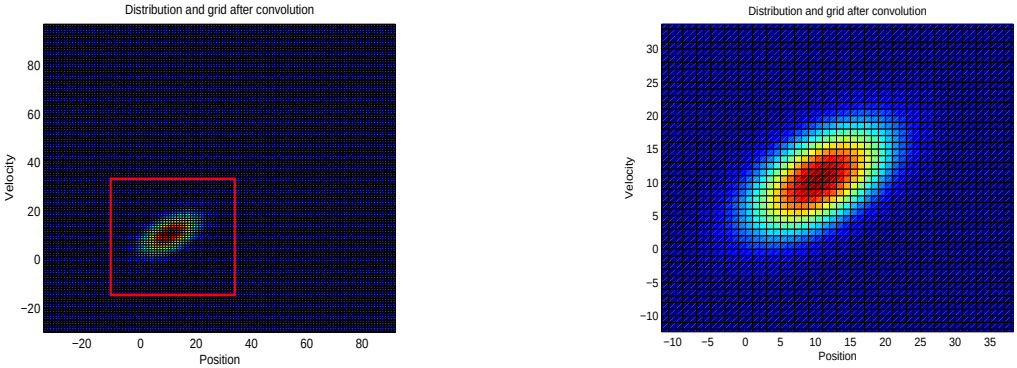
The computation can be done by applying the affine linear transformation $F_t(\bar{\mathtt{x}}^{\mathtt{j}}) + C_t u_t$ to each point of the grid. The resulting grid is in the general case no longer Cartesian, however it can be easily transformed to a Cartesian grid by interpolation. See [145] for details. Finally, as the predicted state is given by the sum of two independent random variables: $x_{t+1} = x_{t+1}^- + v_t$, the corresponding density $\varphi_{t+1|t}(\cdot)$ is the convolution of $\varphi_{t+1|t}^-(\cdot)$ and $\phi(\cdot)$.

Practically, the convolution increases the size of the grid so care has to be taken in order to prevent it from growing indefinitely at each time step. Most of the time, however, the intensity is concentrated in a limited region whose size depends on the predicted distribution and on the covariance matrix of the system noise. For this reason, the last step of the algorithm performs a reframing by eliminating the grid-nodes with negligible value and generating a grid that contains all the relevant information with a generally smaller number of points. Figure 5.25 and 5.26 illustrate the convolution of a Gaussian prior density with a Gaussian centered noise as well as the reframing operation on the resulting grid. The steps of the algorithm are detailed in box 5. In common computer implementations the execution time for FFT is faster for powers of two and it is generally convenient to resize the grids to appropriate dimensions by padding zeros.
Once the grid representation of the predicted probability density has been obtained, the full posterior can be calculated by multiplying each node of the grid by the likelihood with respect to the current observation and normalizing.



(a) (b)

Figure 5.25: Prior density (a) and density after the projection (b)



(c) (d)

Figure 5.26: Reframing operation (c) and final density (d)

Algorithm 5 Single target filtering via FFT

Initialization.

Let $\bar{\varphi}^j_{t|t}$ be the discretization of $\varphi_{t|t}(\cdot)$ over the grid Ω^d_t, and $\bar{\phi}^j_t$ the discretization of the system noise distribution.

for $t = 1, 2, \ldots$ **do**

Projection:

$\varphi^-_{t+1|t}(F_t \bar{\mathtt{x}}^j + C_t u_t) = \bar{\varphi}^j_{t|t}(\bar{\mathtt{x}}^j) \quad \forall j \in \Omega^d_t$

Interpolate $\varphi^-_{t+1|t}$ on a Cartesian grid $\Omega^d_{t+1|t}$.

Prediction:

Apply the convolution theorem

$\bar{\varphi}^j_{t+1|t} = IFFT(FFT(\varphi^-_{t+1|t}) \cdot FFT(\bar{\phi}^j_t))$

Update

$\bar{\varphi}^j_{t+1|t+1}(\bar{\mathtt{x}}^{\mathtt{j}}) = C_t^{-1} g_t(\bar{\mathtt{x}}^{\mathtt{j}}|y_t) \bar{\varphi}^j_{t+1|t}(\bar{\mathtt{x}}^{\mathtt{j}}) \quad \forall j \in \Omega^d_{t+1}$

where C_t^{-1} is the normalization constant.

Reframing

ϵ: error tolerance

Let $\varphi_{d'}(\bar{\mathtt{x}}_{d'})$ be the marginal of $\bar{\varphi}^j_{t+1|t+1}(\cdot)$ along the dimension d' and $\widehat{\varphi}_{d'}(\bar{\mathtt{x}}_{d'})$ the cdf of $\varphi_{d'}(\bar{\mathtt{x}}_{d'})$

for each dimension **do**

$a_{d'} = argmin \widehat{\varphi}_{d'}(\bar{\mathtt{x}}_{d'}) \leq \epsilon$

$b_{d'} = argmax \widehat{\varphi}_{d'}(\bar{\mathtt{x}}_{d'}) \geq 1 - \epsilon$

end for

Reframe the grid to a new grid Ω^d_{t+1} made by all the points $\bar{\mathtt{x}}^j$ of $\Omega^d_{t+1|t}$ such that $a_{d'} \leq \bar{\mathtt{x}}^j_{d'} \leq b_{d'}$, $\forall d'$

end for

5.3.2 PHD Filtering by convolution

The technique described in section 5.3.1 can be extended to the multitarget framework by combining the convolution method to the PHD recursion. The objective is to obtain a discretization of the propagated intensity on a Cartesian grid and then apply the PHD update equation. However, unlike the single target case, in order to obtain the representation of the propagated intensity it is necessary to take into account not only the target dynamics but also the birth intensity, the survival probability and the spawn intensity.
The algorithm initially computes the PHD intensity on a numerical grid by using the convolution technique; the convolution of the original intensity with the Markov kernel modeling the dynamical model of the targets does not change the intensity mass. After the projection, every node of the grid is multiplied by the target's survival probability before adding the birth and spawning intensity. The update step consists in applying the PHD update equation to each node of the grid. The algorithm is described in box 6. To simplify the algorithm we will assume a zero spawning intensity; the presence of spawning intensity does not pose conceptual problems.

Figure 5.27 shows the comparison between the marginal PHD intensity as computed by the FFT-PHD and GM-PHD respectively on the mono-dimensional scenario described in 5.2.2. With a suitable grid size and discretization steps the FFT-PHD produces a better approximation of the PHD posterior as it doesn't use pruning and merging thresholds which eliminate Gaussian components.

Extraction of target state estimates

As for the GM-PHD and for the SMC-PHD a procedure to extract target estimates from the grid-approximated PHD intensity is required. The procedure we describe is not as straightforward as for the GM-PHD but neither as computationally intensive as the general, cluster-based strategies for the SMC-PHD. The peak extraction for the FFTPHD begins by first finding a list of grid nodes with the highest intensity value. These maxima are then validated as true peaks if the intensity in the surrounding region (defined by a pre-determined window) is above an acceptance threshold. Figure 5.28 illustrates intuitively the procedure. The algorithm is analogous to the peak extraction of the GM-PHD for what concerns the presence of an acceptance threshold. The dimension of the validation window has to be chosen by taking into account the system noise covariance matrix. The pruning threshold by taking into account the clutter intensity. Figure 5.29 shows the target locations estimated on a test scenario by the GM-PHD and by the FFT-PHD configured with the same acceptance threshold.

Algorithm 6 FFT-PHD Filter (no spawning intensity)

Initialization.

Let $\mu_t^j(\cdot)$ be the discretization of the birth intensity at time t over the grid Ω_t^d, and $\bar{\phi}_t^j$ the discretization of the system noise distribution.

Initial intensity $\gamma_0^j(\bar{\mathbf{x}}^j) = \mu_0^j(\bar{\mathbf{x}}^j) \quad \forall j \in \Omega_t^d$

Denote by $Y_{t+1} = y_{t,1}, ..., y_{t,|Y_t|}$ the measurements received at time $t+1$

for $t = 1, 2, \ldots$ **do**

Step1: Prediction.

Compute the projection of the PHD

$\gamma'_{t+1|t}(F_t\bar{\mathbf{x}}^j) = \gamma_{t|t}(\bar{\mathbf{x}}^j) \quad \forall j \in \Omega_t^d$

Interpolate $\gamma'_{t+1|t}$ on a Cartesian grid $\Omega_{t+1|t}^d$.

$\bar{\gamma}_{t+1|t} = IFFT(FFT(\gamma'_{t+1|t}) \cdot FFT(\phi_t))$

Add the birth intensity

$\gamma_{t+1|t}^j(\bar{\mathbf{s}}^j) = \bar{\gamma}_{t+1|t}(\bar{\mathbf{s}}^j) + \mu_t(\bar{\mathbf{s}}^j) \quad \forall j \in \Omega_{t+1|t}^d$

Step2: Update

$\tilde{\gamma}_{t+1|t}(\bar{\mathbf{s}}^j) = (1 - p_{d,t+1})\gamma_{t+1|t}(\bar{\mathbf{s}}^j)$

$$\gamma_{t+1|t+1}(\bar{\mathbf{s}}^j) = \tilde{\gamma}_{t+1|t}(\bar{\mathbf{s}}^j) + \sum_{y \in Y_{t+1}} \frac{p_{d,t+1}\gamma_{t+1|t}(\bar{\mathbf{s}}^j) g_t(\bar{\mathbf{s}}^j | y_{t,j})}{h_t(y) + \sum_{\Omega_{t+1|t}} p_{d,t+1}\gamma_{t+1|t}(\bar{\mathbf{s}}^j) g_t(\bar{\mathbf{s}}^j | y_{t,j})}$$

Reframing

ϵ: error tolerance

$\eta_{d'}(\cdot)$: marginal of $\frac{\gamma_{t+1|t+1}(\cdot)}{\int \gamma_{t+1|t+1}(\cdot)}$ along the dimension d'

$\widehat{\eta}_{d'}(\cdot)$: cdf of $\eta_{d'}(\cdot)$

for each dimension d' **do**

$a_{d'} = \arg\min \widehat{\eta}_{d'}(\cdot) \leq \epsilon$

$b_{d'} = \arg\max \widehat{\eta}_{d'}(\cdot) \geq 1 - \epsilon$

end for

Reframe $\Omega_{t+1|t}^d$ into a new grid Ω_{t+1}^d made by all the points $\bar{\mathbf{s}}^j$ of $\Omega_{t+1|t}^d$ such that $a_{d'} \leq \bar{\mathbf{s}}_{d'}^j \leq b_{d'}$, $\forall d'$

end for

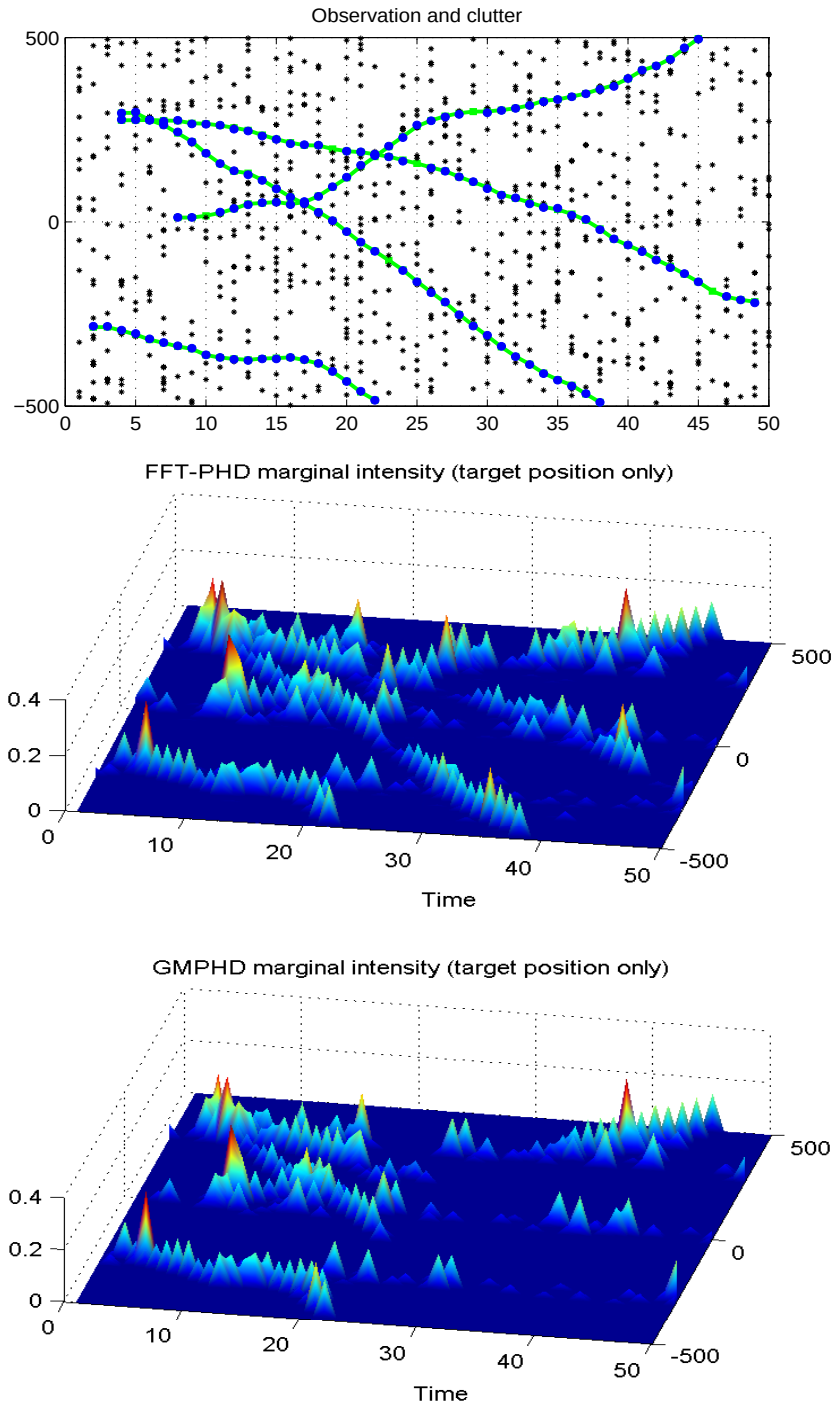


Figure 5.27: Test scenario and marginal PHD intensity as computed by the FFT-PHD and GM-PHD respectively. $\lambda_c = 15$, $p_d = 0.95$, GM-PHD pruning threshold 0.1.

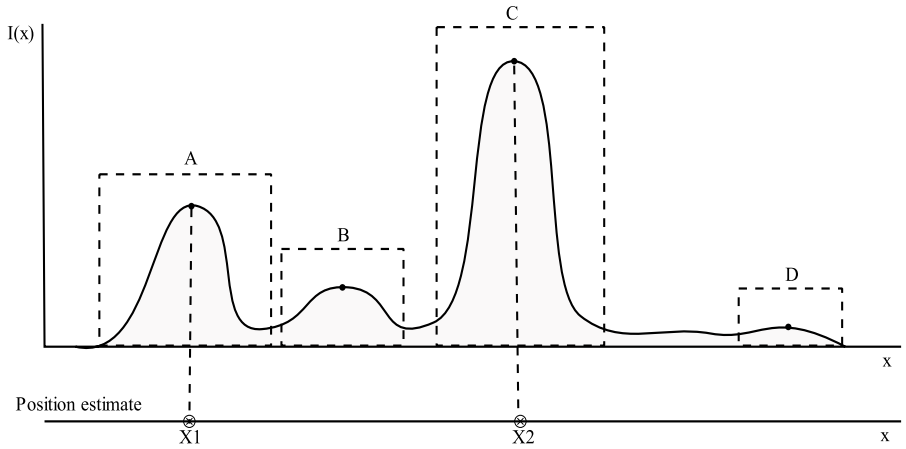


Figure 5.28: Maxima over the grid and validation windows. An estimate is generated at the coordinates of a local maxima if the integral of the intensity over the validation window is greater than a pre-defined threshold. In the example only the points A and C pass the test and generate the estimates X_1 and X_2.

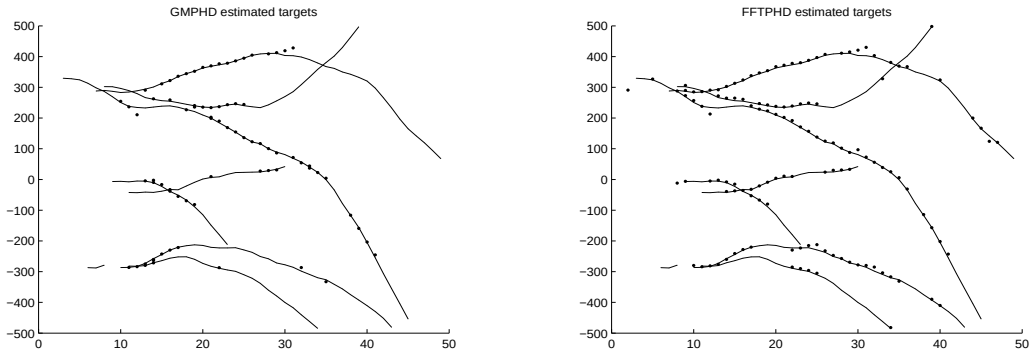


Figure 5.29: Estimated positions generated by the peak extraction algorithm of the GM-PHD and FFT-PHD on a test scenario. The continuous lines show the ground truth trajectories of the targets.

5.3.3 Numerical results

We consider the monodimensional case described in section 5.2.2. The surveillance region is set to [-500,500], the sampling period $T = 1$s, the survival probability is $p_{s,t} = 0.9$, and the detection probability is $p_{d,t} = 0.9$. Moreover $v_t \sim \mathcal{N}(0, \sigma_v^2)$ with $\sigma_v = 2$ and $u_t \sim \mathcal{N}(0, \sigma_u^2)$ and $\sigma_u = 5$. The birth intensity is defined as $\gamma_b = 0.33\mathcal{N}(.; x_b^1, Q_b) + 0.34\mathcal{N}(.; x_b^2, Q_b) + 0.33\mathcal{N}(.; x_b^3, Q_b)$ where $x_b^1 = [-300, 0]^T$, $x_b^2 = [0, 0]^T$, $x_b^3 = [300, 0]^T$.

The clutter is Poisson with intensity λ_c. If a target exits from the surveillance region it is eliminated from the simulation.
A total of 1000 scenarios are generated with different clutter intensities and a varying number of targets. Each scenario is filtered with the GM-PHD and the FFT-PHD filter. The results are compared by using the OSPA metric and both the average cardinality errors and the average positional error are measured. The parameters used to generate the scenario are reported in tables below. Each scenario contains between 4 and 10 targets. The clutter is uniform on the surveillance region. Its intensity λ is an integer randomly chosen from three sets, simulating a moderate $\lambda \in \{1, \ldots, 4\}$, average: $\lambda \in \{5, \ldots, 8\}$ and high: $\lambda \in \{9, \ldots, 12\}$ level of clutter.

GM-PHD	
Pruning threshold	0.1
Merging threshold	5
Max num. Gaussian	200
Target extraction th.	0.7

FFT-PHD	
Grid size	$[1001 \times 21]$
Grid dx (pos.)	2
Grid dv (vel.)	1
Target extraction th.	0.7

Monte Carlo cardinality errors and OSPA distance

On average, the FFT-PHD filter is able to provide a much more precise estimate on the cardinality of the targets. When the number of active targets in the surveillance zone is relatively high, a slight overestimation is registered, basically because of clutter observations occurring close to a zone where active targets are present. When the simulation is run with an intense level of clutter the FFT-PHD always greatly outperforms the GM-PHD For the GM-PHD, the target estimates are extracted by taking the mean of the Gaussian terms whose weight is above a threshold (called target extraction threshold) and in the case of the FF-PHD filter by using the procedure discussed in section 5.3.2. The same extraction threshold of 0.7 is used. Intuitively, this generates an estimate if the expected value of the umber of targets in the considered region of the space is greater than 0.7.
Three plots are reported for each clutter configuration; the OSPA distance averaged over the 1000 iterations is in Fig: 5.30(I), 5.31(I) and 5.32 (I); the cardinality errors Fig: 5.30(III), 5.31(III) 5.32 (III) and the localization errors in Fig: 5.30(II),

5.31(II) and 5.32 (II). The average number of targets estimated in the three clutter configuration is reported in Figure 5.33.
The results show that the FFT-PHD filter is much more precise in estimating the number of ground truth targets and, on average, the OSPA distance is smaller compared to the GM-PHD. The procedure of peak extraction, however, is less precise compared to the GM-PHD. The greater error relies on the fact that the estimates are extracted from the points of the grid, as opposed to the better estimates which are directly available in the GM-PHD. This error can be clearly reduced by increasing the discretization step of the grid, at the expense of a higher computational cost.

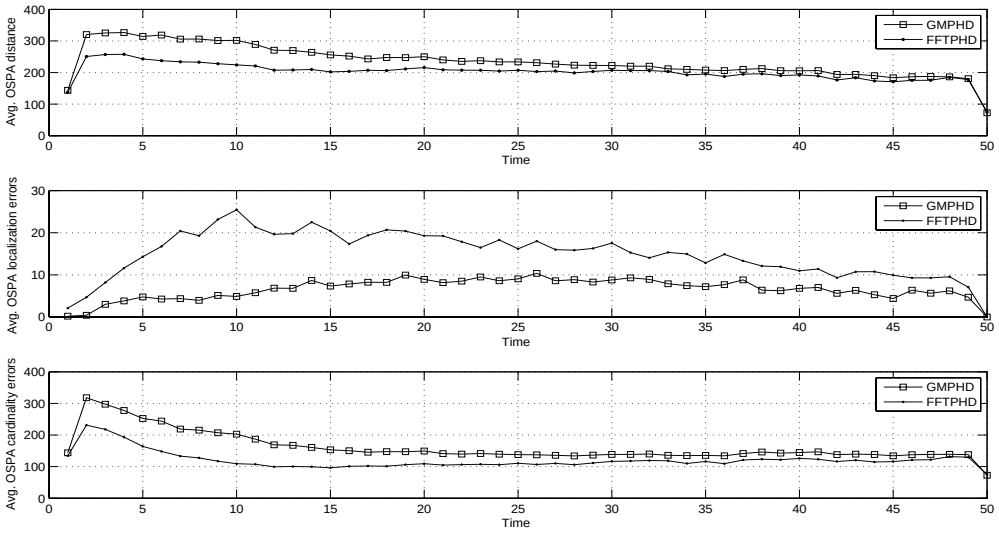


Figure 5.30: GM-PHD and FFT-PHD filtering results. Moderate clutter intensity, 1000 Monte Carlo runs.

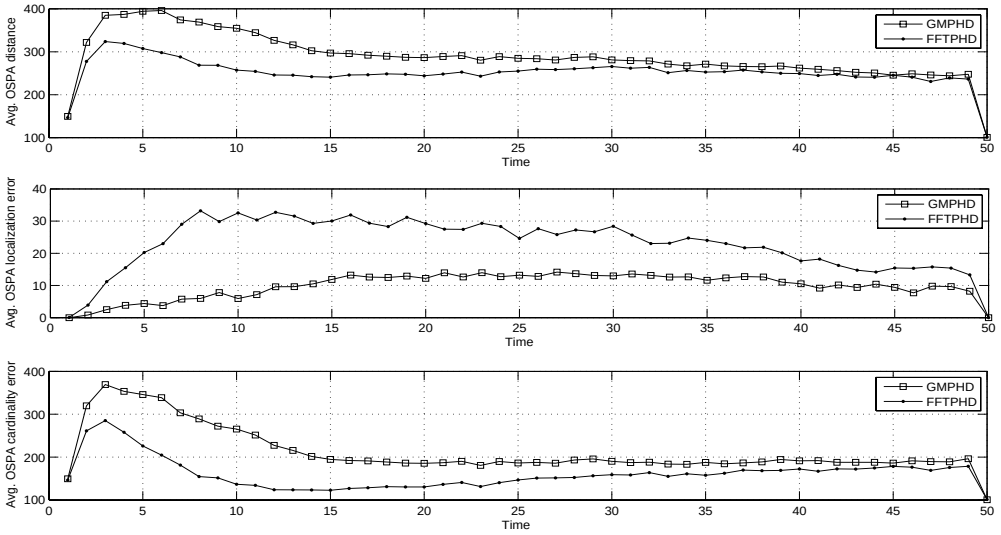


Figure 5.31: GM-PHD and FFT-PHD filtering results. Average clutter intensity, 1000 Monte Carlo runs.

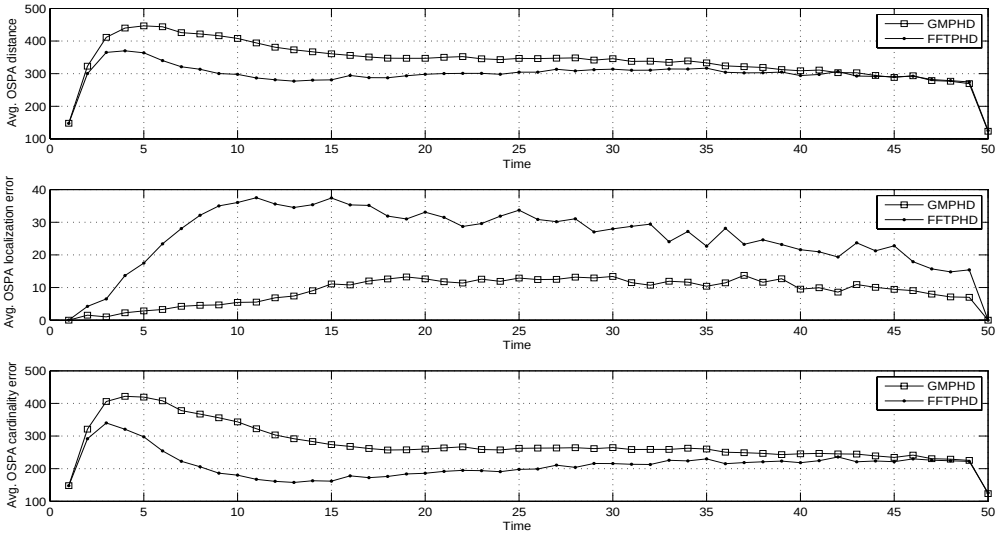
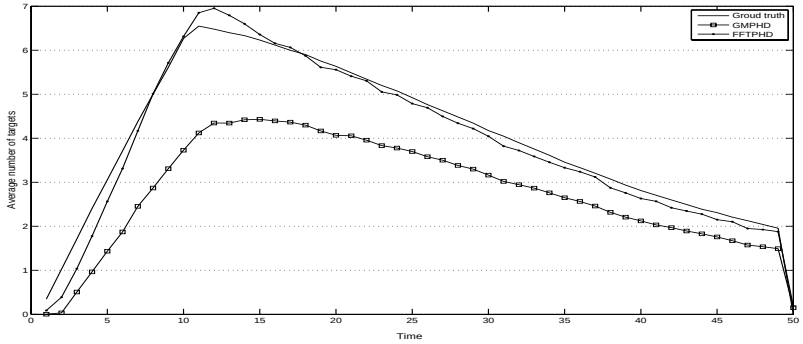
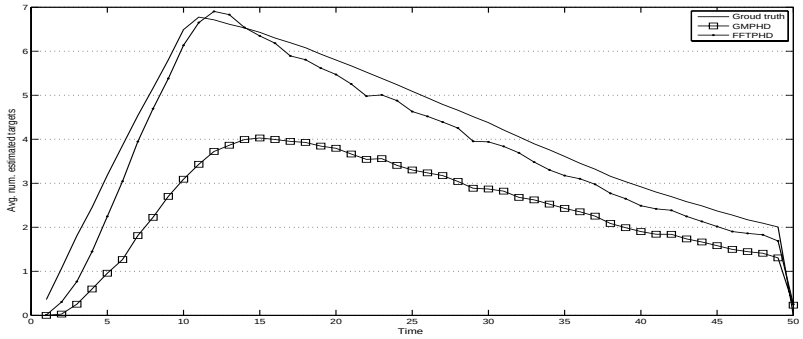


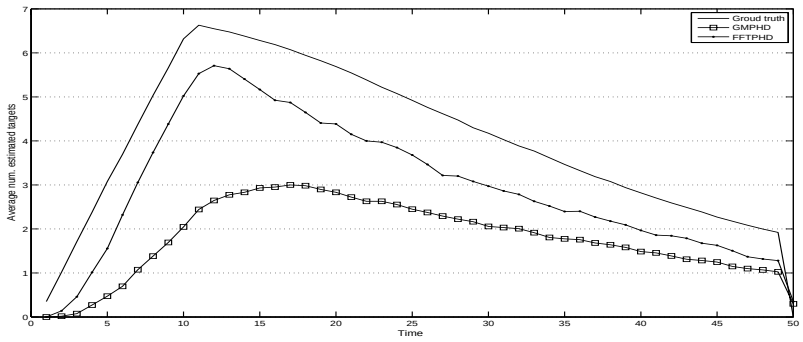
Figure 5.32: GM-PHD and FFT-PHD filtering results. High clutter intensity, 1000 Monte Carlo runs.



(a) Moderate clutter intensity ($\lambda \in [1, \ldots, 4]$)



(b) Average clutter intensity ($\lambda \in [5, \ldots, 8]$)



(c) High clutter intensity ($\lambda \in [9, \ldots, 12]$)

Figure 5.33: Average number of targets as estimated by the GM-PHD filter and by the FFT-PHD filter on 1000 simulated scenarios with different clutter intensities.

Monte Carlo validation

In order to better illustrate the performances of the FFT-PHD in handling critical difficulties, such as a high number of targets in a relatively small surveillance zone with repeated misdetections, this section reports the results obtained in filtering a fixed scenario (Figure 5.3.3) where the probability of target detection is lowered to 0.9 and the average clutter intensity at is increased to 5. Figure 5.3.3 reports the average number of target estimated by the GM-PHD, the FFT-PHD and the number of ground truth targets. The overall filtering performance as well as the localization and cardinality error are reported in Figure 5.36. Results show that the FFT-PHD filter is able to provide a much more accurate estimate of the number of targets than the GM-PHD. Even if the position estimates provided by the FFT-PHD are fairly accurate, the GM-PHD is able to obtain a better result, mainly because of the Kalman filter equations involved in the GM-PHD filter and because of the more straightforward peak extraction procedure.

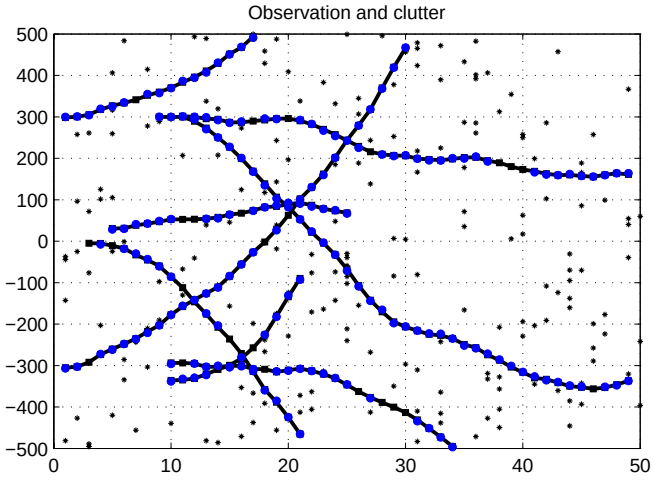


Figure 5.34: Scenario used to compare the GM-PHD and the FFT-PHD.

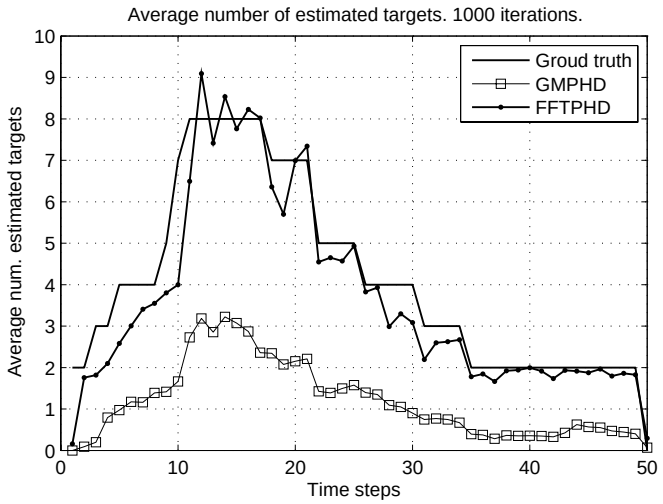


Figure 5.35: Average number of estimated targets over 1000 iterations.

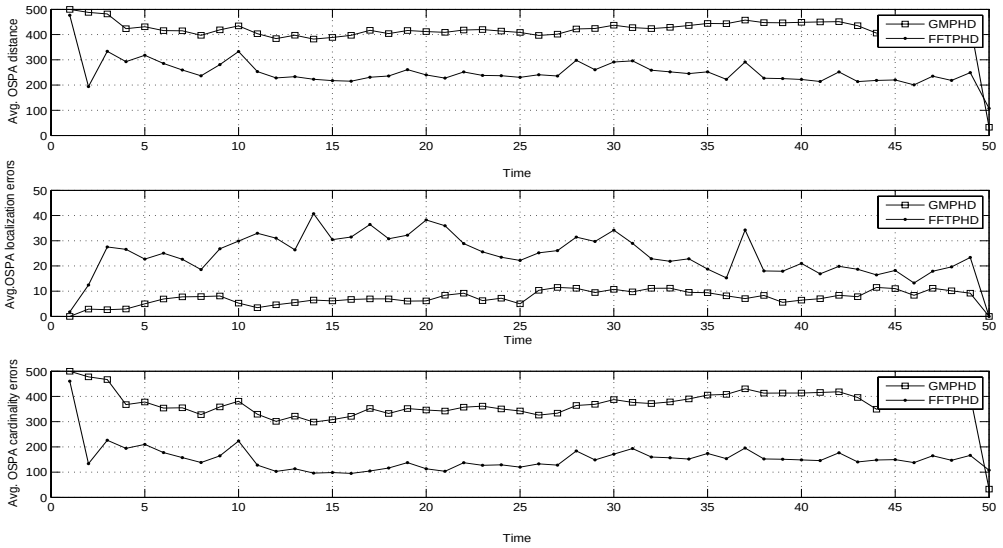


Figure 5.36: OSPA distance, cardinality error and localization error for the test scenario.

Computational efficiency

Figure 5.37 shows the amount of time required to process a scenario in the configurations of clutter previously described. In the case of the FFT-PHD the grid covers the region of the state space comprised between $[-500, 500]$ for the positional coordinate and $[-10, 10]$ for the velocity component. The grid is build by using the discretization steps $dx = 1$ and $dv = 1$. This generates a domain approximation with 1001×21 points. As a result the time required by the FFT-PHD is sensibly

higher than the time required by the GM-PHD. The main reason is accounted to the pruning techniques used in the Gaussian Mixture PHD filter to keep the number of terms bounded and to the computational cost of the Discrete Fourier Transform required in the projection step of the FFT-PHD filter. However, in the current implementation of the filter, the computation of the likelihood functions are done on the whole grid; a straightforward and obvious optimization would be to restrict this computation only to the regions of the space with a non negligible value, if this is possible. Figure 5.38 shows the results of Monte Carlo filterings of the same scenario when the resolution of the grid is changed. In the example, because of the characteristics of the system dynamics, only the velocity dimension is critical for the convolution. In order to test the convolution operation which is the most time-consuming part of the algorithm, the discretization step for the velocity coordinate is incremented from $dv = 0.1$ to $dv = 10$ with a discretization step $\Delta v = 0.1$. The discretization step over the position component is kept at $dx = 1$.
Obviously, by decreasing the resolution of the grid, the computational cost decreases rapidly, at the expense of the precision.

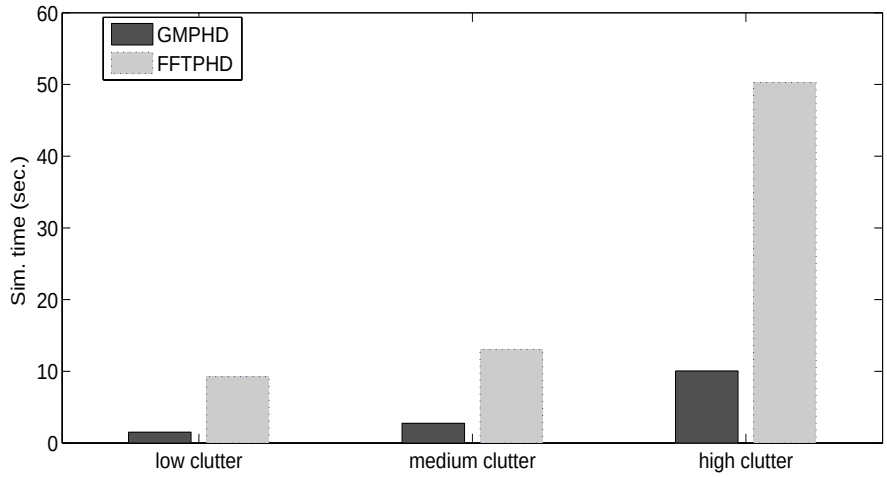


Figure 5.37: Processing time for a complete scenario in case of moderate, medium and high clutter. The discretization grid in the case of FFT-PHD is $dx = 1$, $dv = 1$

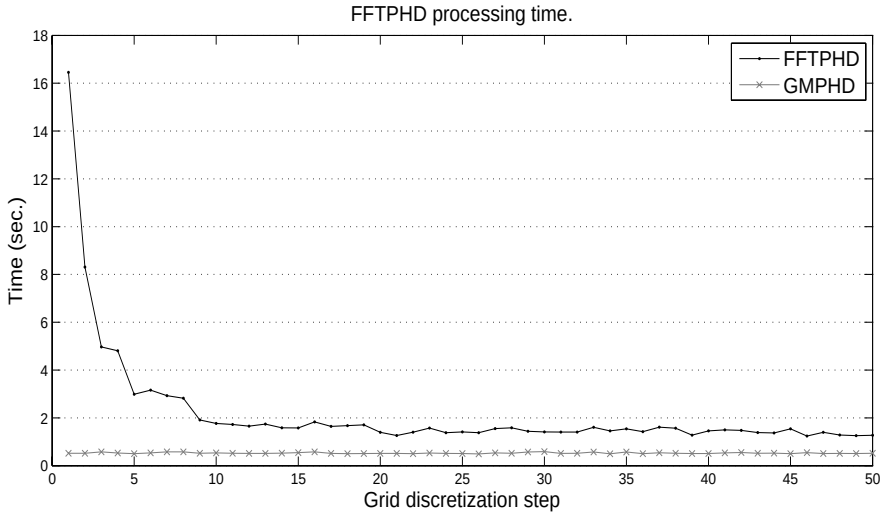


Figure 5.38: Average processing time of the FFT-PHD and GM-PHD obtained when the discretization step is widened.

Conclusions

This section proposed an approach for the computation of the PHD intensity based on numerical grids paired with a technique to compute the predicted PHD intensity according to Mahler's equations by using the Fast Fourier Transform.
One point that the FFT-PHD has in common with the SMC-PHD is the necessity to use post processing to extract target estimates from the peaks of the PHD function. The peak extraction for the FFTPHD is less expensive and produces more accurate estimates. Unfortunately, as the peak extraction relies on the grid-approximated PHD it is not possible to achieve the degree of precision of the GM-PHD. Nevertheless, the overall quality of the multi-target tracking as measured by the OSPA metric is generally good compared to existing solutions and sensibly better than the GM-PHD when the birth intensity is weak and misdetections frequent. Although the computational cost is greater compared to the GM-PHD, the approach is not limited to Gaussian models. The number of calculations depends on the cardinality of the observation sets and on the dimension of the grid, not on the number of targets as in the GM-PHD; moreover, the use of an adaptable grid and the computation of the likelihood function only in the regions of interest may greatly improve the basic approach. The algorithm has been initially tested on low dimensional state space models and it has shown promising results. The FFT-PHD may be a viable approach to the computation of the PHD recursion, especially when the surveillance zone is fixed and when an accurate estimation is preferred to the execution speed.

5.4 Association-sampling particle filters

The problem of characterizing a hidden Poisson point process given a realization of an observed Poisson point process is addressed in the second part of the thesis. By using the results presented in Chapter 6 it is possible to study the measure-valued processes arising in multitarget tracking as a generalization of Feynman-Kac measures.
This alternative formulation leads to the particle approximation algorithm discussed in this section. The central idea is to resample according to the association probabilities between particles and observations. This section details the implementation of the particle approximation algorithm; the derivation and the mathematical details are discussed in Chapter 7.
Before discussing this alternative definition, let's review some example of transport equations for simple branching-type systems.
Consider a sequence of state spaces E_t indexed by the parameter t. Assume that a point $x_t \in E_t$ at time t survives with probability $p_{s,t}(x_t)$ and evolves according to a Markov kernel $K_{t+1} : E_t \mapsto E_{t+1}$. The transport kernel associated its dynamics is:

$$R_{t+1}(x_t, dx_{t+1}) = p_{s,t}(x_t)K_{t+1}(x_t, dx_{t+1}) \tag{5.4.1}$$

Clearly, equation 5.4.1 represents the degenerate case of a branching process in which particles may only survive/evolve or die. In case of spawning the transport kernel must be modified to take into account the additional mass introduced at each time step. In this case it may be written as:

$$R_{t+1}(x_t, dx_{t+1}) = G_t(x_t)K_{t+1}(x_t, dx_{t+1}) \tag{5.4.2}$$

where $G_t(x_t) : E_t \mapsto (0, \infty)$ represents a positive function governing the variation of the mass caused by the disappearence, birth or spawning of points.

In the PHD recursion the intensity of the Poisson point process associated to the targets is propagated by the prediction equation which is similar to the transport equation (5.4.2) if we exclude the additional term corresponding to the birth intensity.
The predicted intensity is then updated according to the parameters of the observation model and to the observations RFS in order obtain the posterior intensity.

To establish the PHD recursion, let $\mathcal{X}_t$ and $\mathcal{Y}_t$ on E_s and E_o denote the two random measures associated to the targets and to the observations at time t. Assume that the initial measure $\mathcal{X}_0$ is a Poisson point process with intensity $\gamma_0(\cdot)$ equal to the birth intensity $\mu_0(\cdot)$. At each time step, $\mu_t(1)$ represents the number of targets which are expected to appear in the surveillance region. For any function $f \in \mathcal{B}(E_s)$

the posterior intensity is given by:

$$\widehat{\gamma}_0(f) = \mathbb{E}(\mathcal{X}_0(f)|\mathcal{Y}_0) = \gamma_0((1-p_{d,t})f)) + \int \mathcal{Y}_0(dy)(1-\beta_{\gamma_0}(y))\Psi_{p_{d,t}g(\bullet,y)}(\gamma_0)(f) \tag{5.4.3}$$

where $g(\bullet, y)$ is the single-target likelihood function and the probability $\beta_{\gamma_0}(y)$, defined in (6.3.6) is related to the chances that the point y on the observation space is generated by clutter. We remind once again that the details as well as the complete derivation of equation (5.4.3) will be given in Chapter 6.

The recursion is obtained by defining the pair of random sequences $(\mathcal{X}_{t+1}, \mathcal{Y}_{t+1})$ such that $\mathcal{X}_{t+1}$ is a Poisson point process with intensity γ_{t+1} defined by:

$$\begin{cases} \widehat{\gamma}_t &= \gamma_t(1-p_{d,t}) + \int \mathcal{Y}_t(dy)\,(1-\beta_{\gamma_t}(y))\ \ \Psi_{p_{d,t}g_t(\bullet,y)}(\gamma_t) \\ \gamma_{t+1} &= \widehat{\gamma}_t R_{t+1} + \mu_t \end{cases} \tag{5.4.4}$$

The first equation of (5.4.4) is the functional representation of the updated PHD intensity, while the second equation combines the updated PHD intensity with the transport kernel (defined later) and with the birth intensity μ_t in order to obtain the predicted intensity γ_{t+1}.
Figure 5.39 provides an intuitive representation of the PHD evolution after the operations of propagation and update: (1) the predicted intensity γ_t is updated by using the observation RFS (2). The resulting intensity $\widehat{\gamma}_t$ is then propagated by the transport equation (5.4.4) (3). The resulting intensity γ_{t+1} is then recursively updated and propagated in time (4-6).

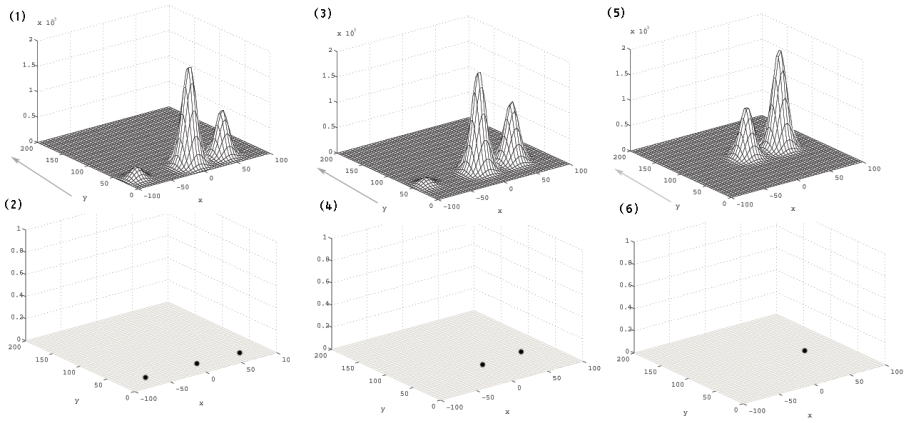


Figure 5.39: Schematic representation of the intensity evolution and corresponding observation RFSs. The maxima of the intensity function correspond to the points where the highest local concentration of targets is expected and are therefore used to estimate the target states.

Consider the PHD recursion in (5.4.4) and assume for simplicity no branching intensity (i.e. the intensity of the point process associated to the objects that

are created by existing objects). By substituting (6.3.6) in (5.4.4) we obtain the following integral expression of the updated posterior intensity on a test function $f \in \mathcal{B}(E_s)$:

$$\widehat{\gamma}_t(f) = \gamma_t((1-p_{d,t})f) + \int \mathcal{Y}_t(dy) \left(1 - \frac{h(y)}{h(y) + \gamma_t(p_{d,t}g_t(y,\,.))}\right) \Psi_{p_{d,t}g_t(y,\,.)}(\gamma_t)(f)$$

and since by definition $\Psi_{p_{d,t}g_t(y,\,.)}(\gamma_t)(f) = \frac{\gamma_t(p_{d,t}g_t(y,\,.)f)}{\gamma_t(p_{d,t}g_t(y,\,.))}$:

$$\widehat{\gamma}_t(f) = \gamma_t((1-p_{d,t})f) + \int \mathcal{Y}_t(dy) \frac{\gamma_t(p_{d,t}g(y,\,.)f)}{h(y) + \gamma_t(p_{d,t}g(y,\,.))} \tag{5.4.5}$$

In case of no spawning, the transport kernel reduces to the single target Markov kernel weighted by the survival probability: $R_{t+1}(x_t, dx_{t+1}) = p_{s,t}(x_t)M_{t+1}(x_t, dx_{t+1})$. By defining the function g_{t,γ_t} as

$$g_{t,\gamma_t}(x_t) = (1 - p_{d,t}(x_t))p_{s,t}(x_t) + \int \mathcal{Y}_t(dy_t) \frac{g_t(x_t, y_t)p_{d,t}(x_t)p_{s,t}(x_t)}{h_t(y_t) + \gamma_t(p_{d,t}g_t(\,.\,, y_t))} \tag{5.4.6}$$

the recursion can be written in a single equation as:

$$\gamma_{t+1}(f) = \gamma_t(g_{t,\gamma_t}M_{t+1}(f)) + \mu_{t+1}(f) \quad = \underbrace{\gamma_t\left(g_{t,\gamma_t}M_{t+1}(f) + \frac{\mu_{t+1}(f)}{\gamma_t(1)}\right)}_{Q_{t+1,\gamma_t}} \tag{5.4.7}$$

The previous manipulation shows that by combining the prediction and update equations of system (5.4.4), the PHD recursion can be written in the form of a general measure-valued dynamical system as follows:

$$\gamma_{t+1}(dx_{t+1}) = (\gamma_t Q_{t+1,\gamma_t})(dx_{t+1}) := \int \gamma_t(dx_t)Q_{t+1,\gamma_t}(x_t, dx_{t+1}) \tag{5.4.8}$$

where the operator Q_{t+1,γ_t} governs the evolution of the intensities. The subscript γ_t is used to stress the dependence of the transport operator on the intensity at time t.

When the branching intensity is not null, the PHD recursion is obtained from equation (5.4.8) with the following choice of operators:

$$Q_{t+1,\gamma_t}(x_t, dx_{t+1}) = \widehat{g}_{t,\gamma_t}(x_t)M_{t+1}(x_t, dx_{t+1}) + \gamma_t(1)^{-1}\ \mu_{t+1}(dx_{t+1}) \tag{5.4.9}$$

with $\widehat{g}_{t,\gamma_t}(x_t) = b_t(x_t)g_{t,\gamma_t}(x_t)$ and $b_t(x_t) = B_{t+1}(1)(x_t)$ denoting the intensity of the spawning process at the point x_t. It can be shown [32] that the Bernoulli filter also satisfies equation (5.4.8) for a different choice of integral operators.

The evolution of the intensity measures can be equivalently described by decoupling the process in the pair $(\gamma_t(1), \eta_t) \in (\mathbb{R}_+ \times \mathcal{P}(E))$ corresponding to the mass and

normalized intensity at each time step respectively. The intensity at time $t+1$ is then described by the pair mass-distribution given by a non-linear transformation:

$$(\gamma_{t+1}(1), \eta_{t+1}) = \Gamma_{t+1}(\gamma_t(1), \eta_t) \tag{5.4.10}$$

One natural way to solve the non-linear integral equation (5.4.8) is to find a judicious probabilistic interpretation of the normalized distributions $\eta_t(dx_t) = \gamma_t(dx_t)/\gamma_t(1)$ and to ensure that the total mass of the process at time $(t+1)$ can be computed in terms of the particle approximations of η_t. In the following, the first and the second components of the mapping Γ_{t+1} will be respectively denoted by:

$$\Gamma^1_{t+1} : \mathbb{R}_+ \times \mathcal{P}(E) \to \mathbb{R}_+ \tag{5.4.11}$$

$$\Gamma^2_{t+1} : \mathbb{R}_+ \times \mathcal{P}(E) \to \mathcal{P}(E) \tag{5.4.12}$$

where (5.4.11), represents the non-linear transformation of the process mass and (5.4.12), the transformation of the normalized intensities. In the next sections we study a mean field particle system and two interacting particle systems for the approximation of the measures η_t. The interacting particle systems are based on the sampling of the associations between the observations and the cluster of particles (or Kalman filters) approximating the posterior intensity.

5.4.1 Mean field and association-based PHD filters

As previously introduced, the mean field-type interpretation of the system (5.4.4) is based on the fact that the distribution η_t can be thought as the laws of a non linear Markov chain whose elementary transitions depend on the distributions η_{t-1} as well as on the mass process $\gamma_{t-1}(1)$.
The non linear transformation Γ^2_{t+1} can be rewritten in the form of a Markov transport:

$$\Gamma^2_{t+1}(\gamma_t(1), \eta_t) = \eta_t K_{t+1,\gamma_t} \quad \text{with} \quad \gamma_t = \gamma_t(1)\eta_t \tag{5.4.13}$$

The subscript γ_t is used to stress the dependency of the Markov kernel on the intensity measure at time t. The updated distribution (5.4.13) can be further decomposed into a product of a selection operator and an update operator [28], both depending on the measure γ_t:

$$\eta_t K_{t+1,\gamma_t} = \eta_t S_{t,\gamma_t} M_{t+1,\gamma_t} \tag{5.4.14}$$

and $\eta_t S_{\eta_t}$ can be written accordingly to (2.0.6) in term of the Boltzmann-Gibbs transformation:

$$\eta_t S_{\eta_t} = \Psi_{g_{t,\gamma_t}}(\eta_t) \tag{5.4.15}$$

where g_{t,γ_t} denotes a positive *potential* function governing the update process. The problem behind the design of a mean field interacting particle system consists in designing the right potential function and Markov transition such that:

$$\Gamma^2_{t+1}(\gamma_t(1), \eta_t) = \Psi_{g_{t,\gamma_t}}(\eta_t) M_{t+1,\gamma_t} \tag{5.4.16}$$

and then approximate η_t by a population of particles that evolve according to M_{t+1,γ_t} and survive/reproduce according to the branching intensities dictated by the potential function g_{t,γ_t}; the design of the updating transformation $\Psi_{g_{t,\gamma_t}}(\eta_t)$ in terms of a Markov transport equation is not unique.

In the rest of the section a pair Markov transitions and potential functions for the PHD recursion are discussed. Their complete derivation and convergence properties can be found in [32]. By construction:

$$\Gamma^1_{t+1}(\gamma_t(1),\eta_t) = \gamma_t(g_{t,\gamma_t}) + \mu_{t+1}(1) \quad \text{and} \quad \Gamma^2_{t+1}(\gamma_t(1),\eta_t) = \Psi_{g_{t,\gamma_t}}(\eta_t) M_{t+1,\gamma_t}$$

M_{t+1,γ_t} is the collection of Markov transitions defined as:

$$M_{t+1,\gamma_t}(x,\boldsymbol{.}) = \alpha_t(\gamma_t)\, M_{t+1}(x,\boldsymbol{.}) + (1-\alpha_t(\gamma_t))\ \overline{\mu}_{t+1}$$

with $\overline{\mu}_{t+1} = \mu_{t+1}/\mu_{t+1}(1)$ and

$$\alpha_t(\gamma_t) = \frac{\gamma_t(g_{t,\gamma_t})}{\gamma_t(g_{t,\gamma_t}) + \mu_{t+1}(1)}$$

For convenience in the design of the algorithm the potential function g_{t,γ_t} is rewritten by introducing a virtual observation point c' (corresponding to undetectable targets) to the observation RFS and by setting: $\mathcal{Y}_t^{c'} = \mathcal{Y}_t + \delta_{c'}$. Let $r_t(x) = p_{s,t}(x) + b_t(x)$ and denote by $g^{c'}_{t,\gamma_t}(x,y)$ the function defined below

$$g^{c'}_{t,\gamma_t}(x,y) = \begin{cases} r_t(x)(1-p_{d,t}(x)) & \text{if} \quad y = c' \\ r_t(x)\ \dfrac{p_{d,t}(x) g_t(x,y_t)}{h_t(y) + \gamma_t(p_{d,t} g_t(x,y))} & \text{if} \quad y \neq c' \end{cases} \tag{5.4.17}$$

The definition of this function may look unnecessary but it will clarify the way undetected targets are dealt in the algorithm. By using (5.4.17) the updating transformations $\Psi_{g_{t,\gamma_t}}(\eta_t)$ can be rewritten equivalently as:

$$\Psi_{g_{t,\gamma_t}}(\eta_t) = \Psi_{\overline{g}_{t,\gamma_t}}(\eta_t) \quad \text{with} \quad \overline{g}_{t,\gamma_t} = \int \mathcal{Y}_t^{c'}(dy)\ g^{c'}_{t,\gamma_t}(\cdot,y)$$

If the functions $\Psi_{g_{t,\gamma_t}}$ and $\Psi_{g_{t,\gamma_t}}(\eta_t) M_{t+1,\gamma_t}$ have a closed form, the evolution of the intensity can be computed with no error, otherwise the mean field interpretation consists in the approximation of the normalized intensity measures and process mass by the discrete measures:

$$\eta_t^N = \frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_t^{(N,j)}} \quad \text{and} \quad \gamma_t^N(dx) = \gamma_t^N(1)\ \eta_t^N(dx)$$

where N denotes the number of particles in the approximation. These particles approximation evolve by using selection type mechanisms dictated by $\Psi_{\overline{g}_{t,\gamma_t}}(\eta_t)$ and mutation mechanisms given by M_{t+1,γ_t}:

$$\begin{pmatrix} \gamma_t^N(1) \\ \eta_t^N \end{pmatrix} \xrightarrow{\text{updating } (\Psi_{g_{t,\gamma_t}}(\eta_t^N))} \begin{pmatrix} \widehat{\gamma}_t^N(1) \\ \widehat{\eta}_t^N \end{pmatrix} \xrightarrow{\text{prediction } (M_{t+1,\gamma_t})} \begin{pmatrix} \gamma_{t+1}^N(1) \\ \eta_{t+1}^N \end{pmatrix}$$

Unfortunately, as in the Sequential Monte Carlo PHD filter, the particle approximation of the normalized intensity doesn't provide the target estimates directly and a post-processing operation is still required. With the explicit expression of the potential function, the mean field version of the algorithm can be summarized with the steps resumed by the Algorithm 7.

The mean-field interpretation of the flow of measures defined by the PHD recursion and the abstract framework described in this section will be useful to develop two sampling-based algorithms for the approximation of the PHD filter. From the practical point of view however, this simple mean-field algorithm does not provide competitive advantages compared to the SMC-PHD implementation.

Algorithm 7 Mean field particle approximation of the PHD recursion. For simplicity the survival and detection probabilities are assumed constant and the spawn intensity is considered null.

Initialize parameters
N: number of particles
μ_t: birth intensity at time t with mass $\mu_t(1)$
$p_{d,t}$: probability of detection at time t
$p_{s,t}$: survival probability at time t
$\mathcal{Y}_t = \{y_1^t, \ldots, y_n^t\}$: observation RFS at time t
c^*: auxiliary observations corresponding to new targets event.
c': auxiliary observations corresponding to undetected targets.
Generate the initial population: Sample N particles $\xi_0^{(N,p)} \quad p = 1, \ldots, N$ from the birth distribution $\mu_0(\cdot)/\mu_0(1)$
Normalized PHD: $\eta_0^N = \frac{1}{N}\sum_{p=1}^N \delta_{\xi_0^{(N,p)}}$
PHD mass: $\gamma_0^N(1) = \mu_0(1)$
Propagate the particles according to the dynamical model
for time $t \geq 1$ **do**
  for $j = 1$ to $|\mathcal{Y}_t|$ **do**
    $L^j = \sum_{p=1}^N g(y_t^j|\xi_t^{(N,p)})$;
    $m^j = \frac{p_{s,t}\cdot p_{d,t}\cdot L^j \cdot \gamma_{t-1}(1)}{h(y_t^j)+p_{d,t}\cdot L^j\cdot\gamma_{t-1}(1)}$
  end for
  Compute the updated mass
  $\gamma_t(1) = \gamma_{t-1}(1)\cdot(1-p_{d,t})\cdot p_{s,t} + \mu_t(1) + \sum_j^{|\mathcal{Y}_t|} m^j$
  for $p = 1 \ldots N$ **do**
    for $y \in \mathcal{Y}_t \cup \{c^*, c'\}$ **do**
      if $y = c^*$ **then**
        $w^{(p,y)} = \mu_t(1)/(N\cdot\gamma_{t-1}(1))$
      else if $y = c'$ **then**
        $w^{(p,y)} = p_{s,t}\cdot(1-p_{d,t})/N$
      else
        $w^{(p,y)} = (p_{s,t}p_{d,t}g(y|\xi_t^{(N,p)}))/(h(y) + p_{d,t}\gamma_{t-1}(1)\cdot L^y)$
      end if
    end for
  end for
  Compute the weight of each particle: $w^p = \sum_y^{|\mathcal{Y}_t|} w^{(p,y)}$
  Resample N particles $\hat{\xi}^{(N,i)} \sim \frac{w^p}{\sum w^p}$
  Propagate the resampled particles according to the dynamical model to obtain $\eta_t^N = \frac{1}{N}\sum_{p=1}^N \delta_{\xi_t^{(N,p)}}$
end for

5.4.2 Association-sampling particle filters

This section describes an approximation technique for the PHD recursion based on the sampling of the associations between the terms of the PHD and the observations at each time step.
Similarly to the GM-PHD and SMC-PHD filter, the posterior PHD is approximated by a Gaussian mixture or by using sets of particles; the main difference relies on the fact that the terms approximating the posterior intensity are propagated in time by using an observations sampled from the RFS according to a specific probability measure. Usual pruning and merging or clustering algorithms are thus not required.
In the case of non-linear, non-Gaussian models, another significant difference with respect to the SMC-PHD is that the approximation of the posterior is done by a set of particle *clusters*, instead of a set of weighted particles.
The discrete probability measure on the space of all the possible associations between measurements and Gaussians or particle clusters is built from equation (5.4.3) and then, once an association has been sampled, the corresponding term is propagated and updated by using the Kalman filter's equations or Sequential Monte Carlo methods.
With linear-Gaussian models this generates a set of interacting Kalman filters while in non-linear models, the relevant modes of the PHD are approximated by clusters of particles propagated and updated in time by using the observation that has been sampled.
The idea towards the construction of the association measure is similar to the idea discussed in Section 5.4.1: starting from equation (5.4.8) the goal is the definition of a discrete probability measure by normalization of the transport kernel Q_{t+1,γ_t} which gives the association probabilities and the Markov transitions corresponding to the events of existing target, misdetection or new targets.
In both cases, the usual thresholding or clustering algorithms are not necessary since the number of Gaussian terms or particle clusters is chosen at the beginning and doesn't change.
In order to handle targets entering in the scene and potential misdetections, the RFS of observations has to be enlarged with two virtual measurements, denoted by c^* and c'. These two auxiliary points are added to the observation space in order to build the potential functions and the Markov transport kernels associated to the events of misdetection and to the appearence of new targets. With the introduction of these two auxiliary points, in fact, the full transport kernel Q_{t+1,γ_t} can be written as:

$$\begin{aligned} Q_{t+1,\gamma_t}(x_t, dx_{t+1}) &= G^{c^*}_{t,\gamma_t}(x_t) M^{c^*}_{t+1}(x_t, dx_{t+1}) + \\ & \quad G^{y}_{t,\gamma_t}(x_t) M^{y}_{t+1}(x_t, dx_{t+1}) + \\ & \quad G^{c'}_{t,\gamma_t}(x_t) M^{c'}_{t+1}(x_t, dx_{t+1}) \end{aligned} \tag{5.4.18}$$

where:

$$\begin{array}{ll} G^{c'}_{t,\gamma_t}(x_t) & := p_{s,t}(x_t)(1-p_{d,t}(x_t)) \\ G^{y}_{t,\gamma_t}(x_t) & := \int \mathcal{Y}_t(dy)\left(\frac{p_{s,t}(p_{d,t}g_t(y,x_t))}{h(y)+\gamma_t(p_{d,t}g_t(y,\bullet))}\right) \\ G^{c^*}_{t,\gamma_t}(x_t) & := \frac{\mu_{t+1}(1)}{\gamma_t(1)} \end{array} \qquad \begin{array}{ll} M^{c'}_{t+1}(x_t,dx_{t+1}) & := M_{t+1}(x_t,dx_{t+1}) \\ M^{y}_{t+1}(x_t,dx_{t+1}) & := M_{t+1}(x_t,dx_{t+1}) \\ M^{c^*}_{t+1}(x_t,dx_{t+1}) & := \bar{\mu}_{t+1}(dx_{t+1}) \end{array}$$

As equation (5.4.18) shows, the full transport kernel Q_{t+1,γ_t} is a combination of different terms expressed in the form of potential functions and Markov transitions. The terms $G^{c^*}_{t,\gamma_t}(x)$ $G^{c'}_{t,\gamma_t}(x)$ and $G^{y}_{t,\gamma_t}(x)$ are the potential functions associated to the birth intensity, misdetection and measurements respectively. The terms $M^{c^*}_{t+1}$ $M^{c'}_{t+1}$ and M^{y}_{t+1} are the Markov transitions corresponding to the events of new target, misdetection, and existing target.
Let $\nu_{t+1}(dy) := \mathcal{Y}_t(dy) + \delta_{c^*}(dy) + \delta_{c'}(dy)$ be a discrete measure on the observation space; by substitution we obtain the formula:

$$\begin{aligned} Q_{t+1,\gamma_t}(x_t,dx_{t+1}) &= \int \left(\mathcal{Y}_t(dy)+\delta_c(dy)+\delta_{c'}(dy)\right) G^{y}_{t,\gamma_t}(x_t) M^{y}_{t+1}(x_t,dx_{t+1}) \\ &= \int \nu_{t+1}(dy) Q^{y}_{t+1,\gamma_t}(x_t,dx_{t+1}) \end{aligned}$$

where the three different dynamics are determined by the transport kernels Q^{y}_{t+1,γ_t} indexed by the observations of the augmented RFS.
In case of linear Gaussian models, the potential functions $G^{c^*}_{t,\gamma_t}(x), G^{y}_{t,\gamma_t}(x), G^{c'}_{t,\gamma_t}(x)$ the distribution η_t and the Markov transitions $M_{t+1}, M^{c^*}_{t+1}, M^{c'}_{t+1}$ are constant functions or Gaussian mixtures, as well as the functions $\eta_t(G^{y}_{t,\gamma_t}(x)M^{y}_{t+1})$ which can be computed exactly. In the general case an approximation is required.
The discrete distribution is obtained by normalization:

$$\begin{aligned} \eta_{t+1}(f) = \frac{\gamma_{t+1}(f)}{\gamma_{t+1}(1)} &= \frac{\gamma_t Q_{t+1,\gamma_t}(f)}{\gamma_t Q_{t+1,\gamma_t}(1)} \\ &= \frac{\gamma_t(1)\eta_t Q_{t+1,\gamma_t}(f)}{\gamma_t(1)\eta_t Q_{t+1,\gamma_t}(1)} \\ &= \int \nu_{t+1}(dy) \frac{\eta_t(G^{y}_{t,\gamma_t} M^{y}_{t+1}(f))}{\int \nu_{t+1}(d\xi)\eta_t(G^{\xi}_{t,\gamma_t})} \end{aligned} \tag{5.4.19}$$

Equation (5.4.19) provides the expression of the normalized PHD at time $t+1$. In order to obtain an expression in the form of a sum of weighted functions, equation (5.4.19) is multiplied and divided by $\eta_t(G^{y}_{t,\gamma_t}(f))$ and rewritten by using the operators:

$$\frac{\eta_t G^{y}_{n,\gamma_t}(f)}{\eta_t(G^{y}_{t,\gamma_t})} = \Psi_{G^{y}_{t,\gamma_t}}(\eta_t)(f) \quad \text{and} \quad \Psi_{G^{y}_{t,\gamma_t}}(\eta_t) M^{y}_{t+1}(f) = \Phi^{y}_{t+1}(\eta_t)(f)$$

where $\Psi_{G^y_{t,\gamma_t}}(\eta_t)$ represents the change of probability given by equation (2.0.5) according to the potential function G^y_{t,γ_t} and $\Phi^y_{t+1}(\eta_t)$ denotes the full update-transport operator.
The normalized distribution of the targets at time t is a sum of weighted terms by the integral expression:

$$\eta_{t+1}(f) = \int \nu_{t+1}(dy) \underbrace{\frac{\eta_t(G^y_{t,\gamma_t}(f))}{\int \nu_{t+1}(d\xi)\eta_t(G^\xi_{t,\gamma_t})}}_{\Theta(y)} \Phi^y_{t+1}(\eta_t)(f) \tag{5.4.20}$$

In case of linear-Gaussian models, the distributions η_t have the form of a Gaussian mixture and the normalized posterior intensity is given by their update and propagation $\Phi^y_{t+1}(\eta_t)$ operator, weighted by the quantity $\Theta(y)$.
Once we have the expression of the normalized posterior PHD as a sum of weighted functions, the approximation is done by sampling a fixed number of terms according to their weights.
The linear-Gaussian case is examined first.
While the usual implementation of the GM-PHD filter works by propagating and updating all the terms of the mixture and then by pruning the Gaussians with a small weight, the proposed algorithm works by approximating the normalized posterior at time t as:

$$\eta_t \approx \frac{1}{N}\sum_{i=1}^{N} \eta_t^{\{i\}} \tag{5.4.21}$$

where $\eta_t^{\{i\}}$ are Gaussian terms. The normalized posterior PHD is obtained by substituing (5.4.21) into (5.4.20):

$$\eta_{t+1}(f) \approx \int \nu_{t+1}(dy) \underbrace{\frac{\sum_{i=1}^N \eta_t^{\{i\}}(G^y_t(f))}{\int \mathcal{Y}_t(d\xi)\sum_{j=1}^N \eta_t^{\{j\}}(G^\xi_t)}}_{\Theta(i,y)} \Phi^y_{t+1}(\eta_t^{\{i\}})(f) \tag{5.4.22}$$

Equation (5.4.22) provides a way to compute recursively in time the normalized intensity of the process by using a set of Kalman filters which get resampled according to the association weights $\Theta(i, y_t)$. Note that the association weight $\Theta(i, y)$ is computed for all the pairs Gaussian/measurement, including the two virtual observations c^* and c' discussed before. Once the terms have been resampled, the operator $\Phi^y_{t+1}(\eta_t^{\{i\}})$ corresponds to the usual Kalman update-prediction of the Gaussian $\eta_t^{\{i\}}$ with respect to the observation y. Algorithm 8 provides a detailed description of the implementation for the so-called Interacting Kalman PHD filter (IKF-PHD).
When linear Gaussian assumptions don't hold it is not possible to have an analytic solution for the integrals $\eta_t^{\{i\}}(G^y_t)$ of equation (5.4.22).

It is however possible to rely on their particle approximation:

$$\eta_t^{\{i\}} \approx \frac{1}{M} \sum_{j=1}^{M} \eta_{t,i}^{\{j\}} \tag{5.4.23}$$

In this case the terms $\eta_t^{\{i\}}$ that corresponded to Gaussian distributions in the IKF-PHD are approximated by clusters of M particles that are resampled with the same technique as in the linear Gaussian case.
We call this approach Interacting Group Particle PHD filter (IGP-PHD). When the number of particles in each cluster is 1, the algorithm reduces to the mean field particle filter.
An intuitive visualization of one step of the algorithm is given in Figure 5.40.

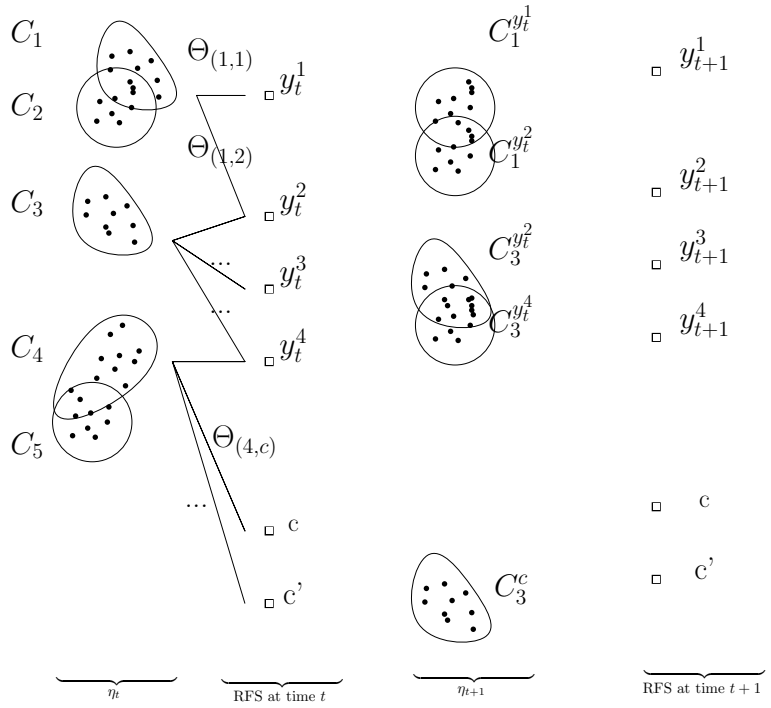


Figure 5.40: The normalized intensity at time t is approximated by a set of particle clusters $\eta_t^{\{i\}}$. The weights $\Theta(i, y_t)$ form a discrete probability measure on the space of the associations between the clusters and the observations, obtained from the PHD recursion. The sampling according to the weights correspond to the choice of a pair cluster-observation. Once the association has been chosen the particles are propagated and updated to obtain a mode of the normalized PHD $\eta_{t+1}^{\{i\}}$

With the same reasoning used to obtain equation (5.4.20) it is possible to derive the expression of the discrete probability measure on the association between the terms and the observations over multiple time steps. In this case the resampling would be done over all the possible associations between the terms of η_t and the

observations in the RFSs received at time $t+1$, $t+2$, etc. However, the exponential complexity of the approach would probably allow the use of no more than three or four time steps.

5.4.3 Numerical results

To test the algorithms, we consider two-dimensional scenarios with an unknown and time-varying number of targets observed in clutter. Two different models are considered: in the first the target dynamics is described by a jump Markov process and the observation model is linear Gaussian; in the latter both the target dynamics and the observation models are non-linear.
The OSPA metric is used to measure the cardinality and estimation errors of the IKF-PHD and IGP-PHD compared to the traditional SMC-PHD and GM-PHD filters. Different clutter rates, target survival probabilities and detection probabilities are used.

Jump Markov multi-target model

The state of each target consists of its position and velocity $x_t = [p_{x,t}, \dot{p}_{x,t}, p_{y,t}, \dot{p}_{y,t}]^T$. An additional state variable r denotes the active dynamical model. The constant velocity (CV) model ($r = 1$) is described by the transition and covariance matrices:

$$F_{t-1}(r=1) = \begin{bmatrix} A_1 & \mathbf{0}_2 \\ \mathbf{0}_2 & A_1 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \tag{5.4.24}$$

$$Q_t(r=1) = \sigma_{1,v}^2 \begin{bmatrix} \Sigma_1 & \mathbf{0}_2 \\ \mathbf{0}_2 & \Sigma_1 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} \frac{T^4}{4} & \frac{T^3}{3} \\ \frac{T^3}{3} & T^2 \end{bmatrix} \tag{5.4.25}$$

where $\mathbf{0}_n$ denotes a $n \times n$ zero matrix, and $\sigma_{1,v}$ denotes the standard deviation of process noise. Otherwise targets may follow a constant turn (CT) model with a counterclockwise turn rate of $0.015°s^{-1}$ ($r = 2$) or with a clockwise turn rate of $0.015°s^{-1}$ ($r = 3$):

$$F_{t-1}(r=2) = \begin{bmatrix} A_2 & -\tilde{A}_2 \\ \tilde{A}_2 & A_2 \end{bmatrix} \tag{5.4.26}$$

$$A_2 = \begin{bmatrix} 1 & \frac{1-cos\omega T}{\omega} \\ 0 & cos\omega T \end{bmatrix}, \tilde{A}_2 = \begin{bmatrix} 0 & \frac{1-cos\omega T}{\omega} \\ 0 & sin\omega T \end{bmatrix}, \tag{5.4.27}$$

$$Q_t(r=\{2,3\}) = \sigma_{1,v}^2 \begin{bmatrix} \Sigma_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \tilde{\Sigma}_2 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} \frac{2(\omega T - sin\omega T)}{\omega^3} & \frac{1-cos\omega T}{\omega^2} \\ \frac{1-cos\omega T}{\omega^2} & T \end{bmatrix}, \tilde{\Sigma}_2 = \begin{bmatrix} 0 & -\frac{\omega T - sin\omega T}{\omega^2} \\ \frac{\omega T - sin\omega T}{\omega^2} & 0 \end{bmatrix} \tag{5.4.28}$$

Each target, if detected, generates a measurement according to a linear Gaussian model with observation and noise covariance matrix:

$$H_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, R_t = \sigma_\epsilon^2 I^2 \tag{5.4.29}$$

Targets appearence is modeled by the birth intensity $\mu_t(\cdot)$ which is a weighted Gaussian mixture of the form:

$$\mu_t(\cdot) = \mu_t(1)[\mathcal{N}(\cdot; m^{(1)}, \Sigma_b^{(1)}) + \mathcal{N}(\cdot; m^{(2)}, \Sigma_b^{(2)}) + \mathcal{N}(\cdot; m^{(3)}, \Sigma_b^{(3)})] \tag{5.4.30}$$

$$m^{(1)} = [4e3,\ 1,\ 0,\ 10]^T,\ m^{(2)} = [-4e3,\ 1,\ 0,\ 10]^T,\ m^{(3)} = [0,\ 1,\ 0,\ 10]^T,$$

$$\Sigma_b^{(1)} = \Sigma_b^{(2)} = \Sigma_b^{(3)} = \begin{bmatrix} 2500 & 0 & 0 & 0 \\ 0 & 400 & 0 & 0 \\ 0 & 0 & 2500 & 0 \\ 0 & 0 & 0 & 400 \end{bmatrix}$$

The survival probability $p_{s,t}$, detection probability $p_{d,t}$ and expected number of new targets at each time step $\mu_t(1)$ are reported for each simulation. The switching between motion models is given by the Markovian transition probability matrix:

$$\tau(r_t|r_{t-1}) = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$

Linear Gaussian tests are performed by using targets with a fixed dynamical model $r = 1$.

Non-linear multi-target model

Consider a non-linear bearings and range example with a time varying number of objects observed in clutter: the state variable $x_t = [\widehat{x}_t^T, \omega_t^T]$ consists in the position and velocity components: $\widehat{x}_t^T = [p_{x,t}, \dot{p}_{x,t}, p_{y,t}, \dot{p}_{y,t}]$ and the turn rate ω_t^T. The dynamical model is:

$$\widehat{x}_t^T = F(\omega_{t-1})\widehat{x}_{t-1} + Gw_{t-1} \tag{5.4.31}$$

$$\omega_t = \omega_{t-1} + \Delta u_{t-1} \tag{5.4.32}$$

where $w_{t-1} \sim \mathcal{N}(\cdot; 0, \sigma_w^2 I)$,$u_{t-1} \sim \mathcal{N}(\cdot; 0, \sigma_u^2 I)$, $\Delta = 1s$, $\sigma_w = 5m/s^2$ and $\sigma_u = \pi/180$ rad/s and:

$$F(\omega) = \begin{bmatrix} 1 & \frac{\sin\omega\Delta}{\omega} & 0 & -\frac{1-\cos\omega\Delta}{\omega} \\ 0 & \cos\omega\Delta & 0 & -\sin\omega\Delta \\ 0 & \frac{1-\cos\omega\Delta}{\omega} & 1 & \frac{\sin\omega\Delta}{\omega} \\ 0 & \sin\omega\Delta & 0 & \cos\omega\Delta \end{bmatrix} G = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ \Delta & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \end{bmatrix} \tag{5.4.33}$$

If the object is detected, the observation is a noisy bearing and range vector given by:

$$y_t = \begin{bmatrix} \arctan(p_{x,t}/p_{y,t}) \\ \sqrt{p_{x,t}^2 + p_{y,t}^2} \end{bmatrix} + \epsilon_t \tag{5.4.34}$$

where $\epsilon_t \sim \mathcal{N}(\cdot; 0, R_t)$, with

$$R_t = \text{diag}([\sigma_\theta^2, \sigma_r^2]^T)$$

and $\sigma_\theta = \pi/180$ rad, $\sigma_r = 5$m. The birth process follows a Poisson RFS with intensity:

$$\mu_t(x) = \sum_{i=1}^{3} \mu_t(1)\mathcal{N}(x; m_b^{(i)}, P_b)$$

where $\mu_t(1) = 0.1$ and $m_b^{(1)} = [500, 15, 0, 15]^T$, $m_b^{(2)} = [-500, 15, 0, 15]^T$,$m_b^{(3)} = [0, 15, 0, 15]^T$. The covariance matrix $P_b = diag([10, 10, 10, 10]$.

State estimation and track management

In the IKF-PHD and IGP-PHD each Gaussian term or each particle cluster is associated at each time step to one of the observations in the augmented measurement RFS. Table 5.6 shows an example of the association history for various clusters during a simulation (the virtual observations associated to new targets and to misdetection events are denoted by the indices -1 and -2 respectively). The numbers in each row indicate the index of the observation assigned to a cluster of particles or to a Gaussian term. Different approaches to build tracks and estimates by exploiting this information are possible. In the examples that follow we use Algorithm 9.
Generally it is difficult to perform a robust and correct track management in the SMC-PHD since the errors introduced by clustering algorithms tend to interfere with the correct association. By reducing the clustering errors the IKF-PHD and IGP-PHD filters are able to provide much more coherent and consistent results. An example of track reconstruction with the IGP-PHD for the test scenarios of Figures 5.41 and 5.43 is shown in Figures 5.42 and 5.44 respectively.

ξ_1 :	-1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
ξ_2 :	-1	2	2	2	2	2	2	-1	-1	2	2	2	2	2	2	2	2	2	2
ξ_3 :	-1	1	1	1	1	1	1	1	-2	1	1	1	1	1	1	1	1	1	1
ξ_4 :	-1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
ξ_5 :	-1	1	1	1	1	1	1	1	-2	1	1	1	1	1	1	1	1	1	1
ξ_6 :	-1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3
ξ_7 :	-1	1	1	1	1	1	1	1	-2	1	1	1	1	1	1	1	1	1	1
ξ_8 :	-1	1	1	1	1	1	1	1	-2	1	1	1	1	1	1	1	1	1	1
ξ_9 :	-1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3

Time step $\rightarrow$

Table 5.6: Example of the association history. Particle clusters used to approximate the posterior intensity are denoted by (ξ_i). Each column represent a time step. At each time step the clusters are associated to a number corresponding to an observation in the measurement RFS. A negative index indicates that the particle cluster is associated to a birth event (-1) or to a misdetection event (-2). Tracks are built by exploiting the associations between measurements and particle clusters.

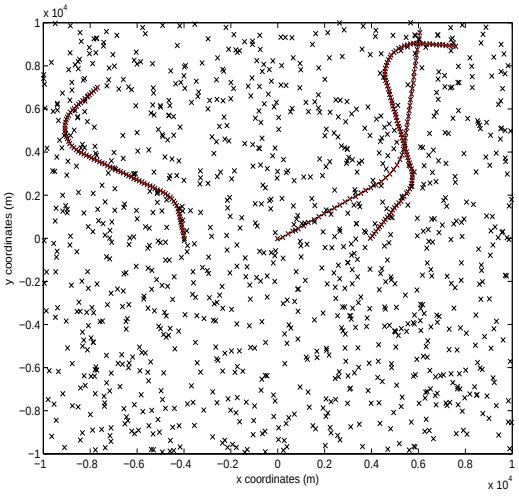


Figure 5.41: Track manager test scenario. 10 clutter points per scan, 100 time steps.

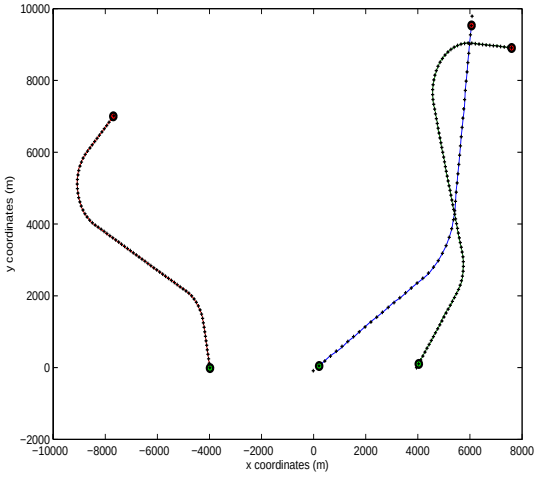


Figure 5.42: Tracks built by exploiting the association information between particle clusters and observations.

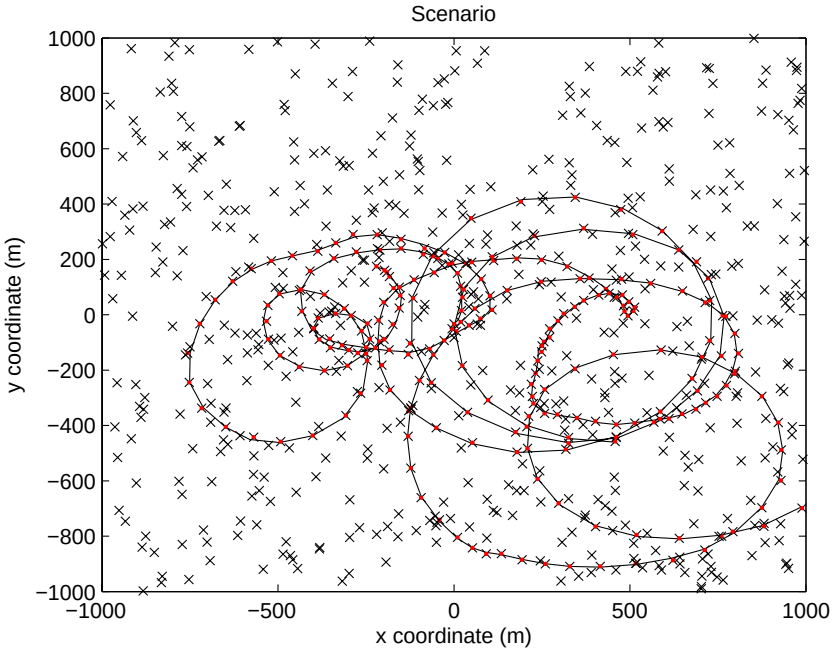


Figure 5.43: Scenario used to illustrate the association-based track manager. 10 clutter points per scan, 100 time steps. Three targets following the non-linear model described in section 5.4.3 are active during the simulation.

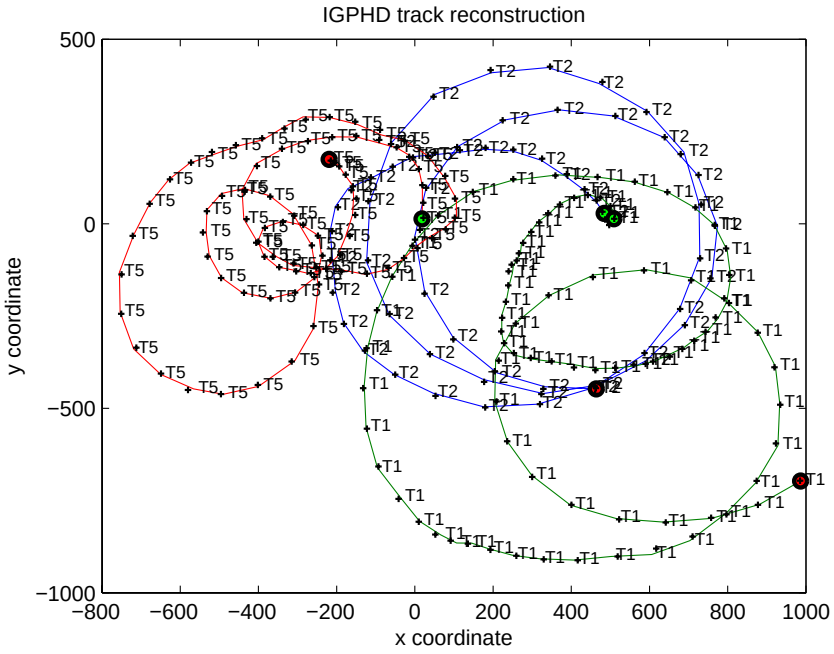


Figure 5.44: Labelled tracks built by exploiting the associations between particle clusters and observations. The targets are tracked as `T1´, `T2´and `T5´. Labels `T3´and `T4´ have been temptatively associated to tracks that have been discarded as false tracks.

Linear-Gaussian models

In linear-Gaussian examples a varying number of targets move in the surveillance zone following a constant velocity model. Figure 5.45 show an instance of the scenarios generated by this model. The Interacting Kalman PHD filter is configured with 1000 particles, the SMC-PHD filter uses an equal number of particles and the K-Means algorithm to extract estimates. The probability of a target detection is set to 0.9 while 5 false observations are expected, on average, at each time step. Figure 5.46 reports the OSPA distance (with parameter 2 and cut-off value 500) as well as the cardinality and assignment errors for the IKF-PHD filter compared to the SMC-PHD and GM-PHD over 50 scenarios; Figure 5.47 shows the OSPA errors of the IGP-PHD filter over 50 iterations. For the IGP-PHD 100 clusters of 50 particles are used.

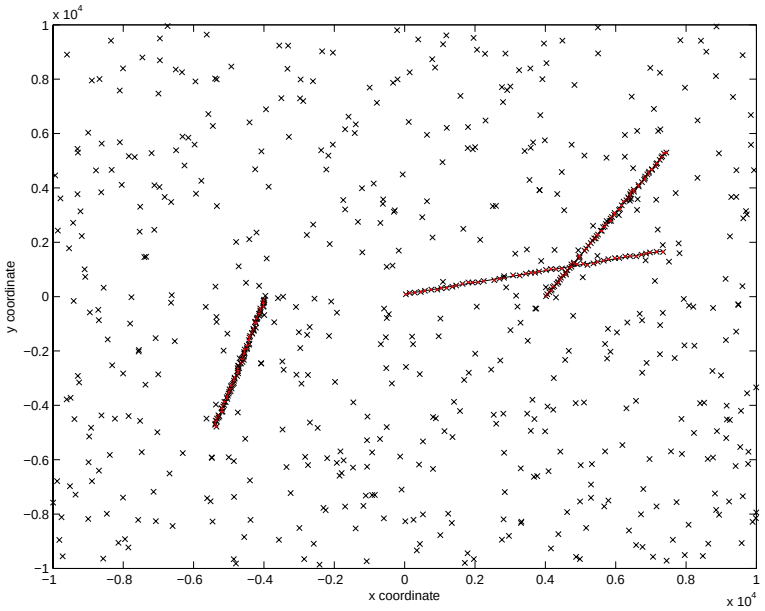


Figure 5.45: Test scenario for the linear model. Three targets following the linear model described in section 5.4.3 are active during the simulation. Detection probability $p_d = 0.9$, $\lambda_c = 5$, 100 time steps.

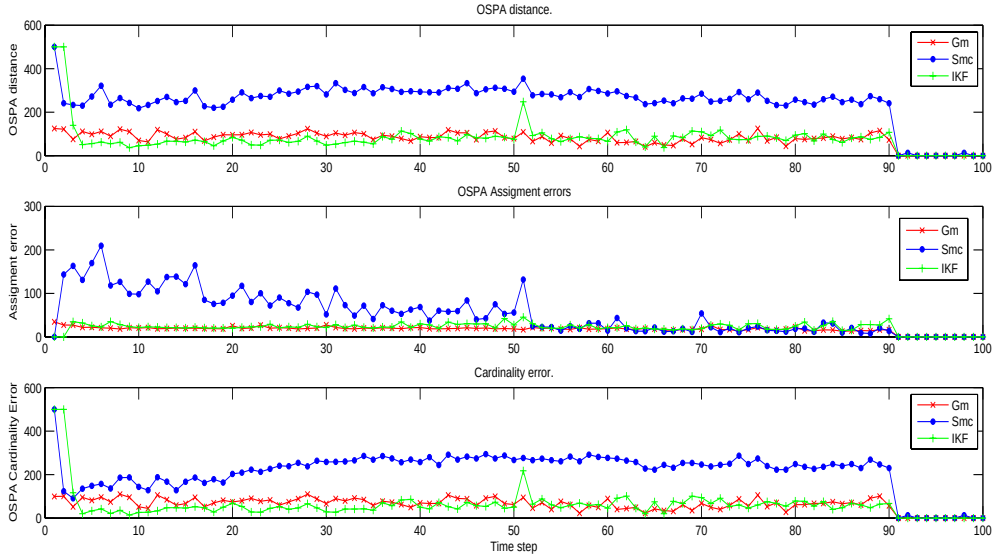


Figure 5.46: OSPA distance, cardinality and assignment errors for the GM-PHD, SMC-PHD and Interacting Kalman PHD filters over 50 scenarii generated according to the linear model described in section 5.4.3, $\lambda_c = 5$ $p_d = 0.9$

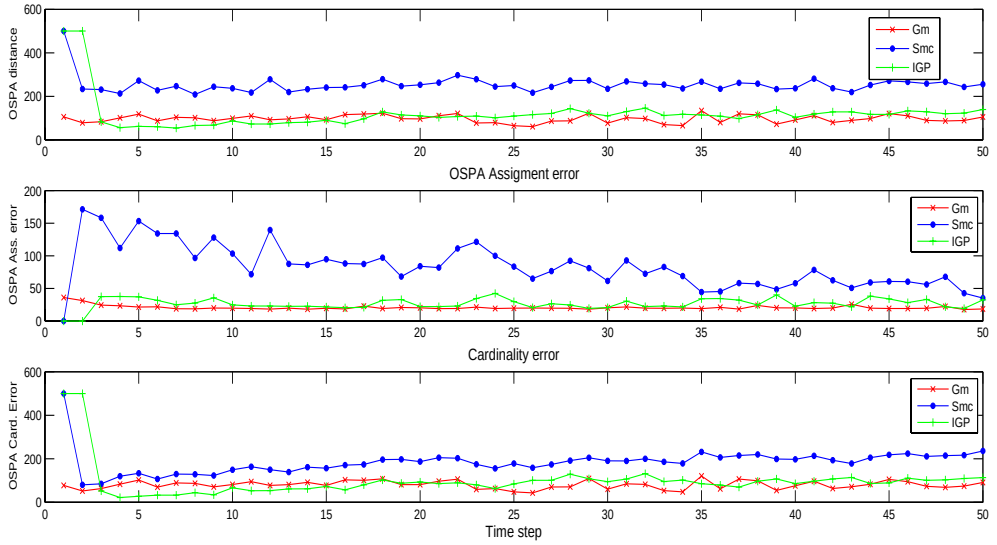


Figure 5.47: OSPA distance, cardinality and assignment errors for the GM-PHD, SMC-PHD and Interacting Group Particle PHD filters over 50 scenarii generated according to the linear model described in section 5.4.3, $\lambda_c = 5$ $p_d = 0.9$

Linear Gaussian Jump Markov models

More complex scenarios are generated by using the interacting jump Markov system described in Section 5.4.3. Figure 5.48 shows a scenario generated by the model. Figures 5.49 and 5.50 report the OSPA distance as well as the cardinality and assignment errors over 10 simulated scenarios. The SMC-PHD filter uses the K-Means algorithm to extract target estimates.

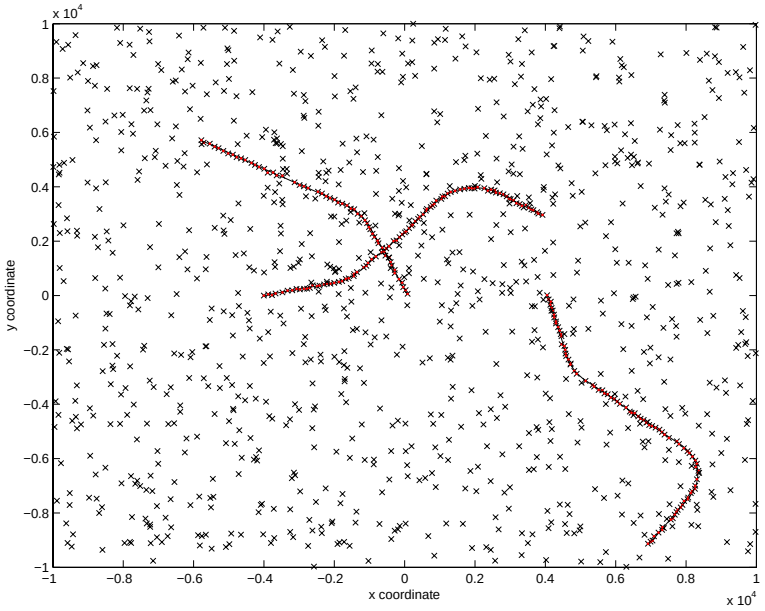


Figure 5.48: Test scenario of the jump Markov model. Three targets having the interacting jump Markov dynamics described in Section 5.4.3 are active during the simulation. Detection probability $p_d = 0.9$, $\lambda_c = 5$, 100 time steps.$p_d = 0.9$, $\lambda_c = 10$

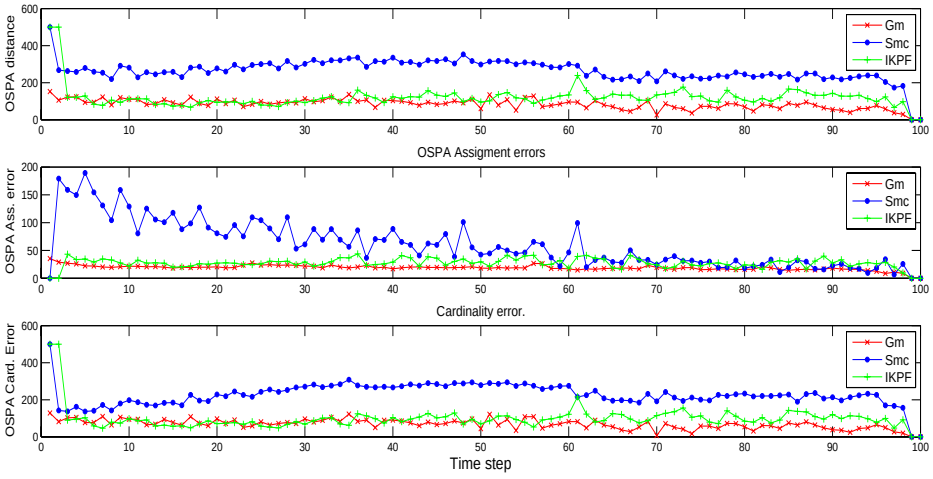


Figure 5.49: OSPA distance, cardinality and assignment errors for the GM-PHD, SMC-PHD and IKF-PHD filters over 50 iterations. Clutter intensity 10, misdetection probability $p_d = 0.9$. The IKF-PHD filter is configured with 1000 particles. The SMC-PHD has an equal number of particles and it uses the K-Means to build estimates.

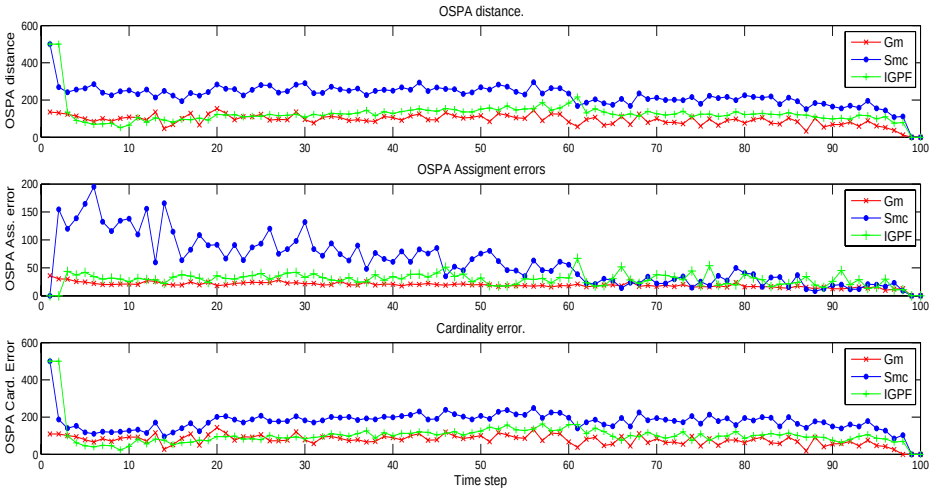


Figure 5.50: OSPA distance, cardinality and assignment errors for the GM-PHD, SMC-PHD and IGP-PHD filters over 50 iterations. Clutter intensity $\lambda_c = 10$, misdetection probability $p_d = 0.9$. The IGP-PHD filter is configured with 200 groups of 50 particles. The SMC-PHD is configured with an equal number of particles and it uses the K-Means to build estimates.

Nonlinear models

Figures 5.51 and 5.52 show an instance of the scenarios generated by the non-linear model described in Section 5.4.3 . SMC-PHD and IGP-PHD filter estimates are

shown in Figures 5.53 and 5.54 while the OSPA errors are shown in Figures 5.55.

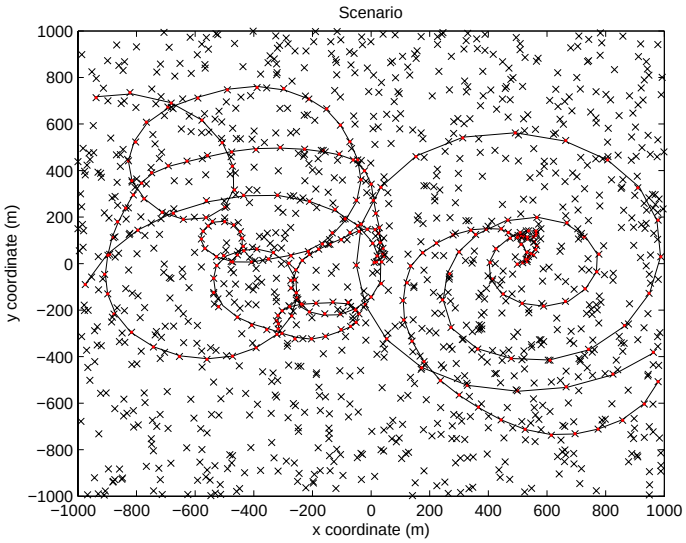


Figure 5.51: Test scenario for the non-linear model described in section 5.4.3. Five targets are active during the simulation. Detection probability $p_d = 0.9$, clutter intensity $\lambda_c = 10$. 100 time steps.

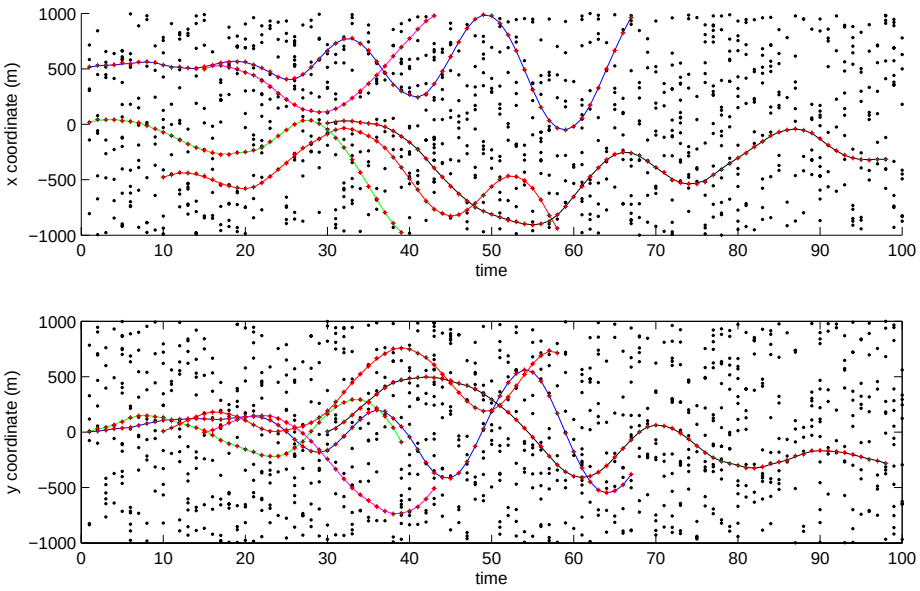


Figure 5.52: Position coordinates of the targets of scenario 5.51.

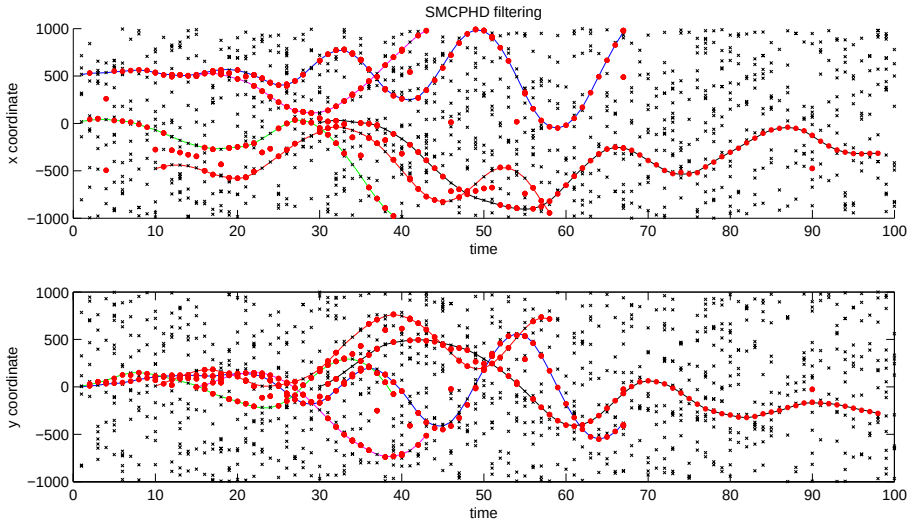


Figure 5.53: SMC-PHD filtering of scenario 5.51. Target estimates in red.

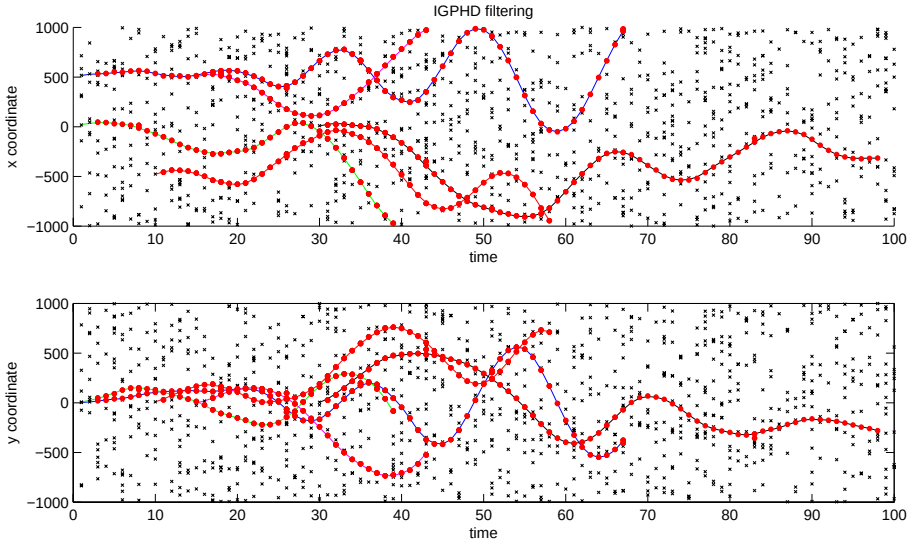


Figure 5.54: IGP-PHD filtering of scenario 5.51. Target estimates in red.

Monte Carlo validation

Since clustering errors are reduced by exploiting the association information, the IKF-PHD and IGP-PHD are able to provide more reliable target estimates and constitute a viable alternative to the SMC-PHD especially in models characterized by strong non-linearities. Figure 5.56 reports the average OSPA errors over 50 filtering iterations on scenarios generated according to the nonlinear model discussed

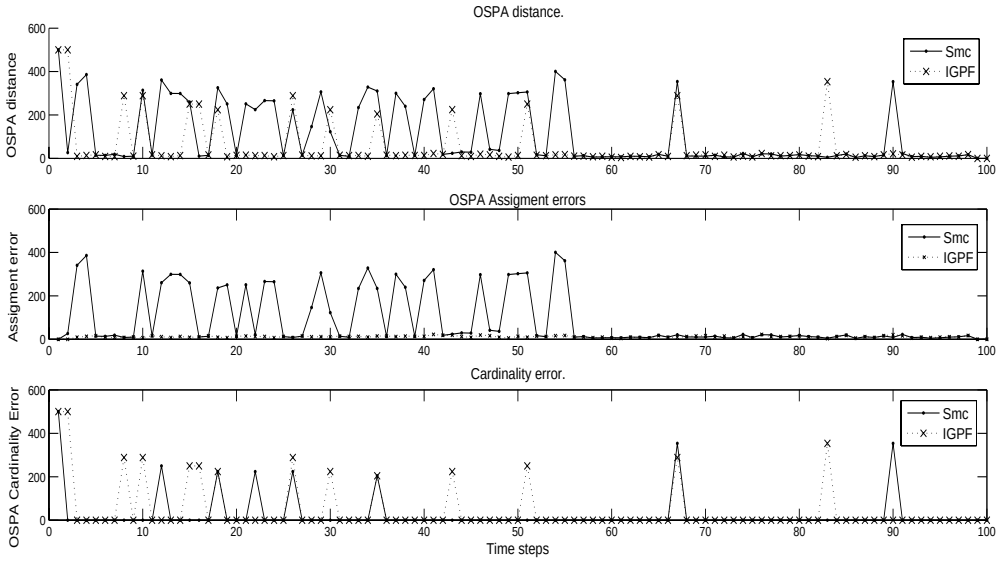


Figure 5.55: OSPA distance, cardinality and assignment errors for the SMC-PHD and IGP-PHD filtering of scenario 5.51.

in section 5.4.3. The scenarios are generated by using a clutter level of $\lambda_c = 10$, with 5 targets appearing at time $t = [1, 10, 1, 15, 30]$ sec. respectively.

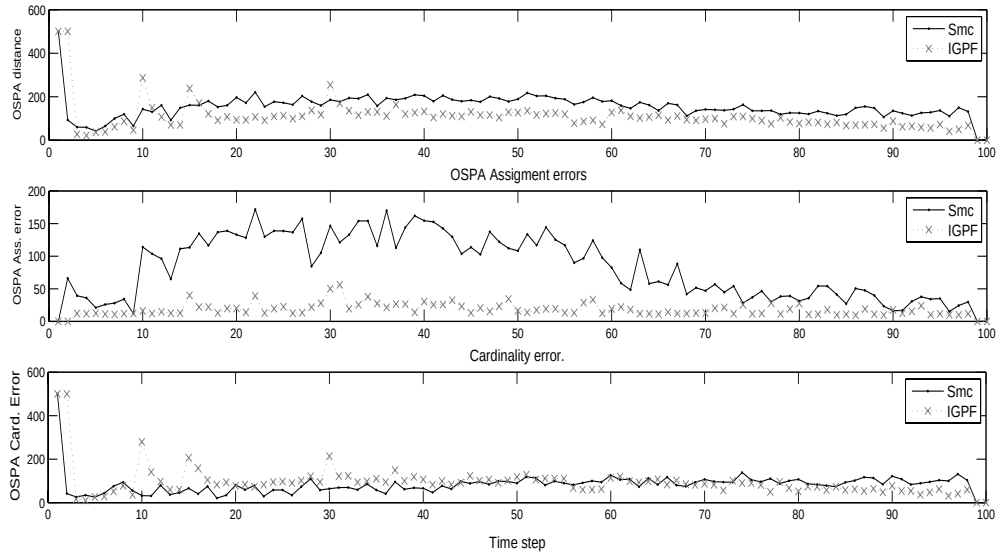


Figure 5.56: Monte Carlo simulation. OSPA distance, cardinality and assignment errors for the SMC-PHD and IGP-PHD filters over 50 iterations. Clutter intensity $\lambda_c = 10$, misdetection probability $p_d = 0.95$

Conclusion

Results show that the estimates built from the associations between particles and observations are more reliable than the estimates obtained with the K-Means in the SMC-PHD. In models where the dynamics of the targets is modelled as a jump Markov process, the estimation obtained with the IGP-PHD is slightly worse than the solution obtained with the Jump Markov GM-PHD (implemented via UKF) but better than that obtained with the SMC-PHD, even when the misdetection probability is relatively low.

In nonlinear models, Monte Carlo validations show that the estimates obtained by the IGP-PHD are more precise than those obtained by the SMC-PHD especially in what concern the assignment errors. Once again the reason is that clustering errors are not present. However, as Figure 5.56 shows, the IGP-PHD needs at least 3 time steps to validate a target. This delay appears in the peaks at time $t = 1, 10, 15, 30$ in the OSPA cardinality error plot.

The IKF-PHD and IGP-PHD filters are thus a viable alternative to the classical algorithms, especially because the clustering errors are sensibly reduced and robust trackers exploiting the association information can be built. The real advantage of these formulations, however, is in dealing with models characterized by strong non-linearities, where the GM-PHD cannot be used, and in models where the posterior PHD is hardly approximated by Gaussian mixtures.

Algorithm 8 Interacting Kalman filter implementation of the PHD recursion

N number of interacting Kalman Filters, S number of gaussian birth modes with mean and covariances $\{\mu_B^{(1,\dots,S)}, P_B^{(1,\dots,S)}\}$, μ_t birth intensity at time t with mass $\mu_t(1)$
$p_{d,t}$,$p_{s,t}$: detection and survival probabilities at time t
Initial normalized intensity $\eta_0 = \frac{1}{N}\sum_{i=0}^{N}\mathcal{N}(m_0^i, P_0^i)$
Observations set at time t $\mathcal{Y}_t := \{y_1^t, \dots, y_n^t\}$; c', c^* = auxiliary observations corresponding to undetected target and to birth events.
Initialization of $\eta_0 = \{m_0^{(i)}, P_0^{(i)}\}_{i=1}^N$:

for $i = 1 \dots N$ **do**
  $j \sim U(\{1, \dots, S\})$
  $m_0^{(i)} = \mu_B^{(j)}, P_0^{(i)} = P_B^{(j)}$
end for
for $t = 1, 2, \dots$ **do**
  for $y \in \mathcal{Y}_t$ **do**
    $L^y = 0$;
    for $i = 1 \dots N$ **do**
      $\bar{h} = H \cdot m_t^{(i)}$; $\Sigma = R + HP_t^{(i)}H'$
      $L^y = L^y + \mathcal{N}(y, \bar{h}, \Sigma)$
    end for
    $\widehat{m}^y = (p_{s,t} \cdot p_{d,t} \cdot L^y \cdot \gamma_{t-1}(1))/(h(y) + p_{d,t} \cdot L^y \cdot \gamma_{t-1}(1))$
  end for
  $\gamma_t(1) = \gamma_{t-1}(1) \cdot (1 - p_{d,t}) \cdot p_{s,t} + \mu_t(1) + \sum_y \widehat{m}^y$
  for $k = 1 \dots N$ **do**
    for $y \in \mathcal{Y}_t \cup \{c^*, c'\}$ **do**
      if $y = c^*$ **then**
        $w^{(k,y)} = \mu_t/(N \cdot \gamma_{t-1}(1))$
      else if $y = c'$ **then**
        $w^{(k,y)} = p_{s,t} \cdot (1 - p_{d,t})/N$
      else
        $H_m = H \cdot m_t^{(k)}$; $\Sigma = R + HP_t^{(k)}H'$; $l = \mathcal{N}(y, H_m, \Sigma)$
        $w(k,y) = (p_{s,t}p_{d,t}l)/(h(y) + p_{d,t}\gamma_{t-1}(1) \cdot L^y)$
      end if
    end for
  end for
  $W = \sum_k \sum_y w(k,y)$
  $\widehat{w}(k,y) = w(k,y)/W$
  Sample N associations: $\{a_i^{(\widehat{k},\widehat{y})}\}_{i=1}^N \sim \widehat{w}(k,y)$
  for $z = 1 \dots N$ **do**
    if $\hat{y} = c^*$ **then**
      $v \sim U(\{1, \dots, L\})$
      $\eta_t^z = \mathcal{N}(\mu_B^v, \Sigma_B^v)$
    else if $\hat{y} = c'$ **then**
      $\eta_t^z = \eta_{t-1}^{\widehat{k}}$
    else
      $\eta_t^z = \mathrm{KFupdate}(\eta_{t-1}^{\widehat{k}}, \widehat{y})$
    end if
  end for
  $\eta_t = \{\mathrm{KFpredict}(\eta_t^z)\}_{z=1}^N$
end for

Algorithm 9 Estimate extraction algorithm

Denote by $\mathcal{Y}_t = \{y_t^1 \cdots y_t^{N_1}\}, \mathcal{Y}_{t-1} = \{y_{t-1}^1 \cdots y_{t-1}^{N_2}\}, \mathcal{Y}_{t-2} = \{y_{t-2}^1 \cdots y_{t-2}^{N_3}\}$ the sets of observations at time t, $t-1$ and $t-2$ respectively.
Denote by ξ_t^i the i-th particle at time t and by $\xi_t^i.y$ and $\xi_t^i.x$ the associated observation and its state.
Estimates set $J = \emptyset$
for each $y_t^k > 0 \in \mathcal{Y}_t$ **do**
  $I_1 = \{i : \xi_t^i.y = y_t^k, \xi_{t-1}^i.y = -2, \xi_{t-2}^i.y > 0\}$
  $I_2 = \{l : \xi_t^l.y = y_t^k, \xi_{t-1}^l.y > 0\}$
  $J = J \cup (1/|I_1|) \sum_{i \in I_1} (\xi_t^i.x) \cup (1/|I_2|) \sum_{l \in I_2"} (\xi_t^l.x)$
end for
Take care of the possible undetected states by adding the following estimates:
for each $y_{t-1}^k > 0 \in \mathcal{Y}_{t-1}$ **do**
  B = { r: ξ_t^r having $\xi_{t-1}^r.y = y_{t-1}^k$}
  if $\xi_t^r.x = 2\ \forall r$ **then**
    $J = J \cup (1/|B|) \sum_{i \in B} (\xi_t^i.x)$
  end if
end for

Part II

Theoretical aspects and Stochastic Analysis

Overview

The first part of the thesis introduced the problem of multi-object filtering in the random finite sets framework and presented the PHD recursion and its common implementations. The objective was to introduce the conceptual framework of multi-object filtering as well as to analyze the numerical performances of PHD filters.
The second part is focused on the theoretical aspects of the PHD recursion considered as a measure-valued system.
Each chapter is introduced by a short paragraph and can be read independently. The organization of this part is as follows:

- Chapter 6 considers the problem of the estimation of a latent point process, given the realization of another point process on abstract measurable state spaces. By establishing the expression of the conditional distribution of a latent Poisson point process given an observation point process, the PHD recursion is derived without the use of probability generating functionals or other advanced FISST concepts.

- Chapter 7 analyzes the sequence of intensity measures associated to a class of nonlinear branching processes and studies their stability properties and long time behavior. A particle scheme to approximate numerically these intensity measures is also proposed.

- Chapter 8 presents an analysis of the exponential stability properties of a class of measure-valued equations arising in nonlinear multi-target filtering and proves the uniform convergence properties of a general class of stochastic filtering algorithms, including sequential Monte Carlo type models and mean field particle interpretation models.

Chapter 6

Conditional Distributions of Spatial Point Processes

6.1 Chapter overview

The approach developed in this chapter consists in the characterization of a hidden Poisson point process given a realization of an observed Poisson point process without the use of probability generating functionals or symmetrization techniques.
By using the connection between the Random Finite Set formulation and the theory of spatial branching processes, a measure theoretic formulation of the Probability Hypothesis Density Filter is derived, complementing [120].
Moreover, the measure-valued processes arising in multi-object filtering are described as a generalization of Feynman-Kac measures where the flow of measures depends on the total mass of the process.
The chapter is organized as follows: section 6.2 introduces the objective, the mathematical point of view and the main contributions as well as the definition of the mathematical concepts. Section 6.3 describes a static model associated to a pair of signal-observation Poisson point processes and section 6.4 establishes the connection between spatial filtering models and the probability hypothesis density equations and provides an alternative, measure-theoretic derivation of the PHD recursion.

This chapter has been accepted and is due to appear as a journal article in *Advances in Applied Probability* (June 2011).

On the Conditional Distributions of Spatial Point Processes

*François Caron, †Pierre Del Moral, ‡Arnaud Doucet, §Michele Pace

Keywords : filtering; multitarget tracking; spatial point processes; probability hypothesis density filter
Mathematics Subject Classification : 62M30, 93E11, 60D05

Abstract

We consider the problem of estimating a latent point process, given the realization of another point process on abstract measurable state spaces. We establish an expression of the conditional distribution of a latent Poisson point process given the observation process when the transformation from the latent process to the observed process includes displacement, thinning and augmentation with extra points. Our original analysis is based on an elementary and self-contained random measure theoretic approach. This simplifies and complements previous derivations given in [77], [120].

6.2 Introduction

Spatial point processes occur in a wide variety of scientific disciplines including environmetrics, epidemiology and seismology; see [26] and [122] for recent books on the subject. In this paper, we are interested in scenarios where the spatial point process of interest is unobserved and we only have access to another spatial point process which is obtained from the original process through displacement, thinning and augmentation with extra points. Such problems arise in forestry [69], [70] but our motivation for this work stems from target tracking applications [77], [120], [136]. In this context, we want to infer the number of targets and their locations; this number can vary as targets enter and exit the surveillance area. We only have access to measurements from a sensor. Some targets may not be detected by the sensor and additionally this sensor also provides us with a random number of false measurements.

*Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Francois.Caron@inria.fr

†Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Pierre.Del-Moral@inria.fr

‡Department of Statistics & Department of Computer Science, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, V6T 1Z2, Canada and The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan, Arnaud@stat.ubc.ca

§Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Michele.Pace@inria.fr

From a mathematical point of view, we are interested in the computation of the conditional distributions of a sequence of random measures with respect to a sequence of noisy and partial observations given by spatial point processes. Recently a few articles have addressed this problem. In a seminal paper [77], R. Mahler has proposed an original and elegant multi-object filtering algorithm known as the PHD (Probability Hypothesis Density) filter which relies on a first order moment approximation of the posterior. The mathematical techniques used by R. Mahler are essentially based on random finite sets techniques including set derivatives and probability generating functionals. In a more recent article [120], S.S. Singh, B.N. Vo, A. Baddeley and S. Zuyev have clarified some important technicalities concerning the use of the derivatives of the joint probability generating functionals to characterize conditional distributions. They have proposed a simplified derivation of the PHD filter and have extended this algorithm to include second moment information. An alternative way to obtain such conditional distributions appeared in [68] and, using Janossy densities, in [66].

The main contribution of this article is to propose an original analysis based on a self-contained random measure theoretic approach. The elementary techniques developed in this paper complement the more traditional random finite sets analysis involving symmetrization techniques or related to other technicalities associated with the computation of moment generating functions derivatives.

The rest of this article is organized as follows. In section 6.3 we first present a static model associated to a pair of signal-observation Poisson point processes. We establish a functional representation of the conditional distribution of a Poisson signal process w.r.t. noisy and partial observations. The proof is elementary. It is extended in section 6.4 to dynamic models in order to establish the PHD equations [77], [120]. We end this introductory section with some standard notations used in the paper.

We denote respectively by $\mathcal{M}(E)$, $\mathcal{P}(E)$ and $\mathcal{B}(E)$, the set of all finite positive measures on some measurable space $(E, \mathcal{E})$, the set of all probability measures, and the Banach space of all bounded and measurable real-valued functions. For $\mu \in \mathcal{M}(E)$ and $f \in \mathcal{B}(E)$, we let $\mu(f) = \int \mu(dx)\ f(x)$ be the Lebesgue integral. The Dirac measure at $a \in E$ is denoted δ_a. We also denote by $\mu^{\otimes p}$ the product measure of $\mu \in \mathcal{M}(E)$ on the product space E^p.
Let $G : x \in E \mapsto G(x) \in [0, \infty)$ be a bounded non-negative potential function. Define $\Psi_G(\eta) \in \mathcal{P}(E)$ by its density $G(x)/\eta(G)$ with respect to a measure η.
In various places in this article, we shall add an auxiliary "death" state d to the original state space E. The functions $f \in \mathcal{B}(E)$ are extended to the augmented space $E \cup \{d\}$ by setting $f(d) = 0$.

For every sequence of points $x = (x^i)_{1\le i\le k}$ in E and every $0 \le p \le k$, we denote by $m_p(x)$ the occupation measure of the first p coordinates $m_p(x) = \sum_{1\le i\le p} \delta_{x^i}$. For $p = 0$, we use the convention $m_0(x) = 0$, the null measure on E.
We recall that a bounded and positive integral operator $f \mapsto L(f)$ from $\mathcal{B}(E_2)$ into $\mathcal{B}(E_1)$ is such that the functions

$$x \mapsto L(f)(x) = \int_{E_2} L(x, dy) f(y)$$

are $\mathcal{E}_1$-measurable and bounded for some measures $L(x, .) \in \mathcal{M}(E_2)$. These operators also generate a dual operator $\mu \mapsto \mu L$ from $\mathcal{M}(E_1)$ into $\mathcal{M}(E_2)$ defined by $(\mu L)(f) = \mu(L(f))$. A Markov kernel is obtained when $L(x, \cdot) \in \mathcal{P}(E)$ for any x.

6.3 Conditional distributions for Poisson processes

Assume the unobserved point process is a finite Poisson point process $\mathcal{X} = \sum_{1\le i\le N} \delta_{X^i}$ with intensity measure γ on some measurable state space $(E_1, \mathcal{E}_1)$. We set $\eta(dx) = \gamma(dx)/\gamma(1)$. The observed point process consists of a collection of random observations directly generated by a random number of points of $\mathcal{X}$ plus some random observations unrelated to $\mathcal{X}$.

To describe more precisely this observed point process, we let α be a measurable function from E_1 into $[0, 1]$ and we consider a Markov transition $L(x, dy)$ from E_1 to E_2. Given a realization of $\mathcal{X}$, every random point $X^i = x$ generates with probability $\alpha(x)$ an observation Y'^i on E_2 with distribution $L(x, dy)$; otherwise it goes into a death state d. Hence $\alpha\,(x)$ measures the "detectability" degree of x. In other words, a given point x generates a random observation in $E_2' = E_2 \cup \{d\}$ with distribution

$$L_d(x, dy) = \alpha(x)\ L(x, dy) + (1 - \alpha(x))\ \ \delta_d(dy). \tag{6.3.1}$$

The resulting point process is the random measure $\sum_{1\le i\le N} \delta_{Y'^i}$ on the augmented state space E_2'.
In addition to this point process we also observe an additional, and independent of $\mathcal{X}$, Poisson point process $\sum_{1\le i\le N_c} \delta_{Y_c'^i}$ with intensity measure ν on E_2; this is known as the clutter noise in multitarget tracking.
In other words, we obtain a process on E_2' given by the random measure

$$\mathcal{Y}' = \sum_{1\le i\le N} \delta_{Y'^i} + \sum_{1\le i\le N_c} \delta_{Y_c'^i}.$$

The state d being unobservable, the observed point process is the random measure $\mathcal{Y}$ on E_2 given by

$$\mathcal{Y} = \sum_{1\le i\le N} 1_{E_2}(Y'^i)\ \delta_{Y'^i} + \sum_{1\le i\le N_c} \delta_{Y_c'^i} = \mathcal{Y}' - N_d\ \delta_d = \sum_{1\le i\le M} \delta_{Y^i}$$

where $N_d = \left(\sum_{1\leq i\leq N} 1_d(Y'^i)\right)$ corresponds to the number of undetected/dead points, and $M = N - N_d + N_c$ is the number of observed points.

Let $\mathcal{X}' = \mathcal{X} + N_c\delta_c$ be defined on $E_1' = E_1 \cup \{c\}$ where c is some cemetery state associated with clutter observations. We present in the following proposition an explicit integral representation of a version of the conditional distributions of $\mathcal{Y}'$ given $\mathcal{X}$ and $\mathcal{X}'$ given $\mathcal{Y}$.

Proposition 6.3.1. *A version of the conditional distribution of $\mathcal{Y}'$ given $\mathcal{X}$ is given for any function $F \in \mathcal{B}(\mathcal{M}(E_2'))$ by*

$$\mathbb{E}\left(F\left(\mathcal{Y}'\right)|\mathcal{X}\right) = e^{-\nu(1)}\sum_{k\geq 0}\frac{1}{k!}\int_{\left(E_2'\right)^{k+N}} F\left(m_k(y_c') + m_N(y')\right)\ \nu^{\otimes k}(dy_c')\ \prod_{i=1}^{N} L_d(X^i, dy'^i) \tag{6.3.2}$$

We further assume that $\nu \ll \lambda$ and $L(x,\,\boldsymbol{.}) \ll \lambda$, for any $x \in E_1$, for some reference measure $\lambda \in \mathcal{M}(E_2)$, with Radon Nikodym derivatives given by

$$g(x,y) = \frac{dL(x,\,\boldsymbol{.}\,)}{d\lambda}(y) \quad \textit{and} \quad h(y) = \frac{d\nu}{d\lambda}(y) \tag{6.3.3}$$

and such that $h(y) + \gamma(\alpha g(\,\boldsymbol{.}\,, y)) > 0$, for any $y \in E_2$.

In this situation, a version of the conditional distribution of $\mathcal{X}'$ given the observation point process $\mathcal{Y}$ is given for any function $F \in \mathcal{B}(\mathcal{M}(E_1'))$ by

$$\mathbb{E}\left(F\left(\mathcal{X}'\right)|\mathcal{Y}\right)$$

$$= e^{-\gamma(1-\alpha)}\sum_{k\geq 0}\frac{\gamma(1-\alpha)^k}{k!}\int_{\left(E_1'\right)^{k+M}} F\left(m_k(x') + m_M(x)\right)\ \Psi_{(1-\alpha)}(\eta)^{\otimes k}\left(dx'\right)\ \prod_{i=1}^{M} Q\left(Y^i, dx^i\right) \tag{6.3.4}$$

where Q is a Markov transition from E_2 into E_1' defined by the following formula

$$Q(y,dx) = (1-\beta(y))\ \Psi_{\alpha g(\,\boldsymbol{.}\,,y)}(\eta)(dx) + \beta(y)\ \delta_c(dx) \tag{6.3.5}$$

with

$$\beta(y) = \frac{h(y)}{h(y) + \gamma(\alpha g(\,\boldsymbol{.}\,, y))}\ . \tag{6.3.6}$$

Proof:
The proof of the first assertion in Eq. (6.3.2) is elementary, thus it is skipped. We provide here a proof of the second result given in Eq. (6.3.4). First, we observe that the random measure

$$\mathcal{Z} = \sum_{1\leq i\leq N}\delta_{(X^i,Y'^i)} + \sum_{1\leq i\leq N_c}\delta_{(c,Y_c'^i)} = \sum_{1\leq i\leq N+N_c}\delta_{(Z_1^i,Z_2^i)} \tag{6.3.7}$$

is a Poisson point process in $E' = E'_1 \times E'_2$. More precisely, the random variable $N+N_c$ is a Poisson random variable with parameter $\kappa = \gamma(1)+\nu(1)$, and $(Z_1^i, Z_2^i)_{i\geq 0}$ is a sequence of independent random variables with common distribution

$$\Gamma(d(z_1,z_2)) = \eta'(dz_1)L'(z_1,dz_2) \qquad \text{with} \qquad \kappa\eta' = \gamma(1)\ \eta + \nu(1)\ \delta_c\ ,$$
$$L'(z_1,dz_2) = 1_{E_1}(z_1)\ L_d(z_1,dz_2) + 1_c(z_1)\ \overline{\nu}(dz_2) \quad \text{with} \quad \overline{\nu}(dz_2) = \nu(dz_2)/\nu(1)\ .$$

From the joint distribution $\Gamma(d(z_1,z_2))$, we can obtain the conditional distribution $L'_{\eta'}(z_2,dz_1)$ of Z_1 given $Z_2 = z_2$ using the easily checked reversal formula, i.e. Bayes'rule

$$\eta'(dz_1)L'(z_1,dz_2) = (\eta' L')\,(dz_2)\ L'_{\eta'}(z_2,dz_1).$$

This yields

$$L'_{\eta'}(z_2,dz_1) = 1_d(z_2)\ \Psi_{(1-\alpha)}(\eta)(dz_1) + 1_{E_2}(z_2)\ Q(z_2,dz_1).$$

Hence we can conclude that for any function $F \in \mathcal{B}(\mathcal{M}(E'_1))$

$$\mathbb{E}\left(F(\mathcal{Z}_1)\,|\mathcal{Z}_2\right) = \int_{\left(E'_1\right)^{N+N_c}} F\left(m_{N+N_c}(z_1)\right)\ \prod_{i=1}^{N+N_c} L'_{\eta'}\left(Z_2^i, dz_1^i\right)$$

where $\mathcal{Z}_j$ stands for the j-th marginal of $\mathcal{Z}$, with $j \in \{1,2\}$. The end of the proof is now a direct consequence of the fact that $(\mathcal{Z}_1,\mathcal{Z}_2) = (\mathcal{X}',\mathcal{Y}')$, $\mathbb{E}\left(F\left(\mathcal{X}'\right)|\mathcal{Y}\right) = \mathbb{E}\left(\mathbb{E}\left(F(\mathcal{X}')\,|\mathcal{Y}'\right)|\mathcal{Y}\right)$ and

$$\mathbb{E}\left(F(\mathcal{Y}')\mid \mathcal{Y}\right) = e^{-\gamma(1-\alpha)} \sum_{k\geq 0}\ \frac{\gamma(1-\alpha)^k}{k!}\ F\left(k\delta_d + \mathcal{Y}\right)$$

for any function $F \in \mathcal{B}(\mathcal{M}(E'_2))$ as N_d follows a Poisson distribution of parameter $\gamma(1-\alpha)$. This ends the proof of the proposition. ■

The expressions of the conditional expectations of linear functionals of the random point processes $\mathcal{X}'$ and $\mathcal{X}$ given the point process $\mathcal{Y}$ follow straightforwardly from the previous proposition. Recall that $f\,(c) = 0$ by convention.

Corollary 6.3.2. *For any function $f \in \mathcal{B}(E'_1)$ we have*

$$\begin{aligned}\mathbb{E}\left(\mathcal{X}'(f)\mid\mathcal{Y}\right) &= \mathbb{E}\left(\mathcal{X}(f)\mid\mathcal{Y}\right)\\ &= e^{-\gamma(1-\alpha)}\sum_{k\geq 0}\frac{\gamma(1-\alpha)^k}{k!}\left(k\ \Psi_{(1-\alpha)}(\eta)\,(f) + \int \mathcal{Y}(dy)Q\,(f)\,(y)\right)\\ &= \gamma((1-\alpha)f) + \int\ \mathcal{Y}(dy)\,(1-\beta(y))\ \ \Psi_{\alpha g(\bullet,y)}(\eta)(f)\ . \qquad (6.3.8)\end{aligned}$$

In particular, the conditional expectation of the number of points N in $\mathcal{X}$ given the observations is given by

$$\mathbb{E}\left(N|\mathcal{Y}\right) = \mathbb{E}\left(\mathcal{X}(1)\mid\mathcal{Y}\right) = \ \gamma(1-\alpha) + \mathcal{Y}\,(1-\beta)\,. \qquad (6.3.9)$$

6.4 Spatial filtering models and probability hypothesis density equations

We show here how the results obtained in proposition 6.3.1 and corollary 6.3.2 allows us to establish directly the PHD filter equations [77], [120].
In what follows the parameter n is interpreted as a discrete time index. We consider a collection of measures $\mu_n \in \mathcal{M}(E_1)$ and a collection of positive operators R_{n+1} from E_1 into E_1.

We then define recursively a sequence of random measures $\mathcal{X}_n$ and $\mathcal{Y}_n$ on E_1 and E_2 as follows. The initial measure $\mathcal{X}_0$ is a Poisson point process with intensity measure $\gamma_0 = \mu_0$ on E_1. Given a realization of $\mathcal{X}_0$, the corresponding observation process $\mathcal{Y}_0$ on E_2 is defined as in section 6.3 with a detection function α_0 on E_1, a clutter intensity measure ν_0, and some Markov transitions $L_{d,0}$ and L_0 defined as in (6.3.1) and satisfying (6.3.3) for some reference measure λ_0 and some functions h_0 and g_0. From corollary 6.3.2, we have for any function $f \in \mathcal{B}(E_1)$

$$\begin{aligned}\widehat{\gamma}_0(f) =& \mathbb{E}\left(\mathcal{X}_0(f) \mid \mathcal{Y}_0\right) \\ &= \gamma_0((1-\alpha_0)f) + \int \mathcal{Y}_0(dy)\,(1-\beta_0(y))\;\Psi_{\alpha_0 g_0(\bullet,y)}(\gamma_0)(f)\end{aligned}$$

with a function β_0 defined as in Eq. (6.3.6) by substituting (α_0, h_0, g_0) to (α, h, g). Given a realization of the pair random sequences $(\mathcal{X}_p, \mathcal{Y}_p)$, with $0 \leq p \leq n$, the pair of random measures $(\mathcal{X}_{n+1}, \mathcal{Y}_{n+1})$ is *defined* as follows. We set $\mathcal{X}_{n+1}$ to be a Poisson point process with intensity measure γ_{n+1} defined by the following recursions for any function $f \in \mathcal{B}(E_1)$

$$\begin{aligned}\widehat{\gamma}_n(f) =& \gamma_n((1-\alpha_n)f) + \int \mathcal{Y}_n(dy)\,(1-\beta_n(y))\;\Psi_{\alpha_n g_n(\bullet,y)}(\gamma_n)(f) \\ \gamma_{n+1} =& \widehat{\gamma}_n R_{n+1} + \mu_{n+1}\end{aligned}$$

In the context of spatial branching processes, μ_n stands for the intensity measure of a spontaneous birth model while R_{n+1} represents the first moment transport kernel associated with a spatial branching type mechanism. For example, assume that each point $X_n^i = x$ at time n dies with probability $\rho(x)$ or survives and evolves according to a Markov kernel K_{n+1} from E_1 into E_1 then R_{n+1} corresponds to

$$R_{n+1}\left(x, dx'\right) = \left(1-\rho(x)\right) K_{n+1}\left(x, dx'\right).$$

It is also possible to modify R_{n+1} to include some spawning points [77], [120], [136]. In addition, given a realization of $\mathcal{X}_{n+1}$, the corresponding observation process $\mathcal{Y}_{n+1}$ is defined as in section 6.3 with a detection function α_{n+1} on E_1, a clutter intensity measure ν_{n+1}, and some Markov transitions $L_{d,(n+1)}$ and L_{n+1} defined as in (6.3.1) and satisfying (6.3.3) for some reference measure λ_{n+1} and some functions h_{n+1}

and g_{n+1}. We let $N_{c,n}$ be the number of death states c associated with clutter observations at time n and M_n be the number of observations at time n.

The following elementary corollary proves that the PHD filter propagates the first moment of the multi-target posterior distribution of the filtering model defined above. This is a direct consequence of proposition 6.3.1 and corollary 6.3.2.

Corollary 6.4.1. *An integral version of the conditional distribution of $\mathcal{X}'_n = \mathcal{X}_n + N_{c,n}\delta_c$ given the filtration $\mathcal{F}_n^Y = \sigma\left(\mathcal{Y}_p,\ 0 \leq p \leq n\right)$ generated by the observation point processes $\mathcal{Y}_p = \sum_{1\leq i\leq M_p} \delta_{Y_p^i}$, from the origin $p = 0$ up to the current time $p = n$, is given for any function $F \in \mathcal{B}\left(\mathcal{M}(E_1')\right)$ by the following formula*

$$\mathbb{E}\left(F\left(\mathcal{X}'_n\right)\middle|\mathcal{F}_n^Y\right) = e^{-\gamma_n(1-\alpha_n)} \sum_{k\geq 0} \frac{\gamma_n(1-\alpha_n)^k}{k!}$$
$$\int_{\left(E_1'\right)^{k+M_n}} F\left(m_k(x') + m_{M_n}(x)\right)\ \Psi_{(1-\alpha_n)}(\gamma_n)^{\otimes k}\left(dx'\right)\ \prod_{i=1}^{M_n} Q_n\left(Y_n^i, dx^i\right)$$

with the Markov transitions

$$Q_n(y, dx) = (1-\beta_n(y))\ \Psi_{\alpha_n g_n(\bullet, y)}(\gamma_n)(dx) + \beta_n(y)\ \delta_c(dx)\ .$$

In particular, the random measures γ_n and $\widehat{\gamma}_n$ defined below coincide with the first moment of the random measures $\mathcal{X}^n$ given the sigma-fields $\mathcal{F}_{n-1}^Y$ and $\mathcal{F}_n^Y$; that is, for any function $f \in \mathcal{B}(E_1)$, we have

$$\gamma_n(f) = \mathbb{E}\left(\mathcal{X}_n(f)\ |\ \mathcal{F}_{n-1}^Y\right) \quad \text{and} \quad \widehat{\gamma}_n(f) = \mathbb{E}\left(\mathcal{X}_n(f)\ |\ \mathcal{F}_n^Y\right).$$

Chapter 7

Particle approximations of a class of branching distribution flows arising in multi-target tracking

7.1 Chapter overview

In this chapter we analyze the sequence of intensity measures associated to the spatial branching point processes arising in the context of multiple object filtering and study their stability properties and long time behavior. Under appropriate regularity conditions, it is possible to demonstrate uniform and non asymptotic estimates and a functional Central Limit Theorem.
In the second part of the article we propose a technique to approximate numerically the flow of intensity measures.

This result has been accepted and is due to appear as a journal article in *SIAM Journal on Control and Optimization*.

Particle approximation of the intensity measures of a spatial branching point process arising in multi-target tracking

*François Caron, †Pierre Del Moral, ‡Arnaud Doucet, §Michele Pace

Abstract

The aim of this paper is two-fold. First we analyze the sequence of intensity measures of a spatial branching point process arising in a multiple target tracking context. We study its stability properties, characterizes its long time behavior and provide a series of weak Lipschitz type functional contraction inequalities. Second we design and analyze an original particle scheme to approximate numerically these intensity measures. Under appropriate regularity conditions, we obtain uniform and non asymptotic estimates and a functional central limit theorem. To the best of our knowledge, these are the first sharp theoretical results available for this class of spatial branching point processes.

Keywords : Spatial branching processes, multi-target tracking problems, mean field and interacting particle systems, Feynman-Kac semigroups, uniform estimates w.r.t. time, functional central limit theorems.

7.2 Introduction

Multi-target tracking problems deal with tracking several targets simultaneously given noisy sensor measurements. Over recent years, point processes approaches to address these problems have become very popular. The use of point processes in a multiple-target tracking context was first proposed in S. Mori et al. [89] as early as in 1986. Using a random sets formalism, a formalism essentially equivalent to the point process formalism [37, 90], R. Malher and his co-authors proposed in two books [86, 42] a systematic treatment of multi-sensor multi-target filtering problems. However, as mentioned in [90], "... although the random sets formalism (or the point process formalism) for multitarget tracking has provided a unified view

*Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Francois.Caron@inria.fr

†Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Pierre.Del-Moral@inria.fr

‡Department of Statistics & Department of Computer Science, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, V6T 1Z2, Canada and The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan, Arnaud@stat.ubc.ca

§Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux , Universit Bordeaux, 351 cours de la Libration 33405 Talence cedex, France, Michele.Pace@inria.fr

on the subject of multiple target tracking, it has failed to produce any significant practical tracking algorithms...".

This situation has recently changed following the introduction of the PHD (probability hypothesis density) filter by R. Malher [73, 76]. The PHD filter is a powerful multi-target tracking algorithm which is essentially a Poisson type approximation to the optimal multi-target filter [73, 86, 120]. It has found numerous applications since its introduction. The PHD filter cannot be computed analytically but it can be approximated by a mixture of Gaussians for linear Gaussian target models [130] and by non-standard particle methods for nonlinear non-Gaussian target models [133, 53].

Despite their increasing popularity, the theoretical performance of these multi-target particle methods remain poorly understood. Indeed their mathematical structure is significantly different from standard particle filters so the detailed theoretical results for particle filters provided in [28] are not applicable. Some convergence results have been already established in [133, 53] but remain quite limited. Reference [133] presents a basic convergence result for the PHD filter but does not establish any rate of convergence. In [53] the authors provide some quantitative bounds and a central limit theorem. However these quantitative bounds are not sharp and no stability result is provided.

The aim of this work is to initiate a thorough theoretical study of these non-standard particle methods by first characterizing the stability properties of the "signal" process and establishing uniform w.r.t the time index convergence results for its particle approximation. This "signal" process is a spatial branching point process whose intensity measure always satisfies a closed recursive equation in the space of bounded positive measures. We will not consider any observation process in this article. The analysis of the particle approximations of PHD filters is presented in [32]. It builds heavily upon the present work but it is even more complex as it additionally involves at each time step a nonlinear update of the intensity measure.

The rest is organized as follows: in section 7.3, we present a spatial branching point process which is general enough to model a wide variety of multiple target problems. We establish the linear evolution equation associated to the intensity measures of this process and introduces an original particle scheme to approximate numerically these measures. Section 7.4 summarizes the main results of this paper. In Section 7.5, we provide a detailed analysis of the stability properties and the long time behavior of these sequence of intensity measures, including the asymptotic behavior of the total mass process, i.e. the integral of the intensity measure over the state space, and the convergence to equilibrium of the corresponding sequence of normalized intensity measures. For time-homogeneous models, we exhibit three different types of asymptotic behavior. The analysis of these stability properties is essential in order to guarantee the robustness of the model and to obtain reli-

able numerical approximation schemes. Section 7.6 is devoted to the theoretical study of the non-standard particle scheme introduced to approximate the intensity measures. Our main result in this section is an non-asymptotic convergence for this scheme. Under some appropriate stability conditions, we additionaly obtain uniform estimates w.r.t. the time parameter.

7.3 Spatial branching point process and its particle approximation

7.3.1 Spatial branching point process for multi-target tracking

Assume that at a given time n there are N_n target states $(X_n^i)_{1\leq i\leq N_n}$ taking values in some measurable state space E_n enlarged with an auxiliary cemetery point c. The state space E_n depends on the problem at hand. It may vary with the time parameter and can include all the characteristics of a target such as its type, its kinetic parameters as well as its complete path from the origin. As usual, we extend the measures γ_n and the bounded measurable functions f_n on E_n by setting $\gamma_n(c) = 0$ and $f_n(c) = 0$.

Each target has a survival probability $e_n(X_n^i) \in [0, 1]$. When a target dies, it goes to the cemetery point c. We also use the convention $e_n(c) = 0$ so that a dead target can only stay in the cemetery. Survival targets give birth to a random strictly positive number of individuals $h_n^i(X_n^i)$ where $(h_n^i(X_n^i))_{1\leq i\leq N_n}$ is a collection of independent random variables such that $\mathbb{E}\left(h_n^i(x_n)\right) = H_n(x_n)$ for any $x_n \in E_n$ where H_n is a given collection of bounded functions H_n. We have $H_n\left(x_n\right) \geq 1$ for any $x_n \in E_n$ as $h_n^i(x_n) \geq 1$. This branching transition is called spawning in the multi-target tracking literature. We define $G_n = e_n H_n$.

After this branching transition, the system consists of a random number $\widehat{N}_n$ of individuals $(\widehat{X}_n^i)_{1\leq i\leq \widehat{N}_n}$. Each of them evolves randomly $\widehat{X}_n^i = x_n \rightsquigarrow X_{n+1}^i$ according to a Markov transition $M_{n+1}(x_n, dx_{n+1})$ from E_n into E_{n+1}. We use the convention $M_{n+1}(c, c) = 1$, so that any dead target remains in the cemetery state.
At the same time, an independent collection of new targets is added to the current configuration. This additional and spontaneous branching process is often modeled by a spatial Poisson process with a prescribed intensity measure μ_{n+1} on E_{n+1}. This spontaneous branching scheme is used to model new targets entering the state space.

At the end of this transition, we obtain $N_{n+1} = \widehat{N}_n + N'_{n+1}$ targets $(X_{n+1}^i)_{1\leq i\leq N_{n+1}}$, where N'_{n+1} is a Poisson random variable with parameter given by the total mass $\mu_{n+1}(1)$ of the positive measure μ_{n+1}, and $(X_{n+1}^{\widehat{N}_n+i})_{1\leq i\leq N'_{n+1}}$ are independent and identically distributed random variables with common distribution $\overline{\mu}_{n+1} = \mu_{n+1}/\mu_{n+1}(1)$

where $\mu_{n+1}(1) := \int_{E_{n+1}} \mu_{n+1}(dx)$.

Example. To illustrate the model, we present here a simple yet standard example [130]. In this case, we set $E_n = E = \mathbb{R}^4$ corresponding to the area monitored by a radar. All the targets are assumed to be of the same type. The state of a target $X_n = [p_n^x, p_n^y, v_n^x, v_n^y]$ consists of its position $[p_n^x, p_n^y]$ and velocity $[v_n^x, v_n^y]$ and is assumed to evolve according to a linear Gaussian model

$$X_n = AX_{n-1} + V_n \tag{7.3.1}$$

where $V_n \sim \mathcal{N}(0, \Sigma)$ is a sequence of i.i.d zero-mean normal random variables of covariance Σ; i.e. $M_n(x_{n-1}, dx_n) = M(x_{n-1}, x_n)dx_n$ with

$$M(x_{n-1}, x_n) = \frac{1}{(2\pi |\Sigma|)^{1/2}} \exp\left(-\frac{1}{2}(x_n - Ax_{n-1})^{\mathrm{T}} \Sigma^{-1} (x_n - Ax_{n-1})\right).$$

We assume that $\mu_n(x) = \mu(x)$, $e_n(x) = s > 0$ and $h_n(x_n) = h \in \{1, 2\}$ with $\mathbb{P}(h = 1) = 1 - \mathbb{P}(h = 2) = \alpha$. Hence for this model, each target X_{n-1} survives at time $n-1$ with a probability s. Each survival target has one offspring with probability α which evolves according to (7.3.1) or two offspring with probability $1 - \alpha$ which, conditional upon X_{n-1}, independently evolve according to (7.3.1). Additionally, a random number of targets distributed according to a Poisson distribution of parameter λ appear. These targets are independent and distributed in E as $\overline{\mu}(x) = \mu(x)/\mu(1)$.

7.3.2 Sequence of intensity distributions

At every time n, the intensity measure of the point process $\mathcal{X}_n := \sum_{i=1}^{N_n} \delta_{X_n^i}$ associated to the targets is given for any bounded measurable function f on $E_n \cup \{c\}$ by the following formula:

$$\gamma_n(f) := \mathbb{E}(\mathcal{X}_n(f)) \quad \text{with} \quad \mathcal{X}_n(f) := \int f(x)\ \mathcal{X}_n(dx)$$

To simplify the presentation, we suppose that the initial configuration of the targets is a spatial Poisson process with intensity measure μ_0 on the state space E_0.

Given the construction defined in section 7.3.1, it follows straightforwardly that the intensity measures γ_n on E_n satisfy the following recursive equation.

Lemma 7.3.1. *For any $n \geq 0$, we have*

$$\gamma_{n+1}(dx') = \int \gamma_n(dx)\ Q_{n+1}(x, dx') + \mu_{n+1}(dx') \tag{7.3.2}$$

with the initial condition $\gamma_0 = \mu_0$ where μ_{n+1} is the intensity measure of the spatial point process associated to the birth of new targets at time $n+1$ while the integral operator Q_{n+1} from E_n into E_{n+1} is defined by

$$Q_{n+1}(x_n, dx_{n+1}) := G_n(x_n)\ M_{n+1}(x_n, dx_{n+1}). \tag{7.3.3}$$

Proof:
For any bounded measurable function f on $E_{n+1} \cup \{c\}$, we have using the notation introduced at the end of section 7.3.1

$$\begin{aligned}\gamma_{n+1}(f) &= \mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) + \mathbb{E}\left(\sum_{i=\widehat{N}_n}^{\widehat{N}_n+N'_{n+1}} f\left(X_{n+1}^i\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) + \mu_{n+1}(1)\,\overline{\mu}_{n+1}(f).\end{aligned}$$

Let $\mathcal{G}_n$ denote the σ-field generated by $(X_n^i)_{1\leq i\leq N_n}$ then

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\middle|\,\mathcal{G}_n\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{N_n} e_n\left(X_n^i\right) h_n^i(X_n^i) M_{n+1}(f)\left(X_n^i\right)\right) \\ &= \gamma_n\left(e_n H_n M_{n+1}(f)\right)\end{aligned}$$

and the result follows. ∎

These intensity measures typically do not admit any closed-form expression. A natural way to approximate them numerically is to use a particle interpretation of the associated sequence of probability distributions given by

$$\eta_n(dx) := \gamma_n(dx_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(1) := \int_{E_n} \gamma_n(dx)$$

To avoid unnecessary technical details, we further assume that the potential functions G_n are chosen so that for any $x \in E_n$

$$0 < g_{n,-} \leq G_n(x) \leq g_{n,+} < \infty \tag{7.3.4}$$

for any time parameter $n \geq 0$. Note that this assumption is satisfied in most realistic multiple target scenarios such as the example discussed at the end of section 7.3.1. Indeed the condition $g_{n,-} \leq G_n(x)$ essentially states that there exists $e_{n,-} > 0$ such that $e_n(x) \geq e_{n,-}$ for any $x \in E_n$ as $H_n(x) \geq 1$. The condition $G_n(x) \leq g_{n,+}$ states

that there exists $H_{n,+} < \infty$ such that $H_n(x) \leq H_{n,+}$ for any $x \in E_n$ as $e_n(x) \leq 1$. In the unlikely scenario where (7.3.4) is not satisfied then the forthcoming analysis can be extended to more general models using the techniques developed in section 4.4 in [28]; see also [16]. We denote by $\mathcal{P}(E_n)$ the set of probability measures on the state space E_n.

To describe these particle approximations, it is important to observe that the pair process $(\gamma_n(1), \eta_n) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))$ satisfies an evolution equation of the following form

$$(\gamma_n(1), \eta_n) = \Gamma_n(\gamma_{n-1}(1), \eta_{n-1}) \tag{7.3.5}$$

We let Γ_n^1 and Γ_n^2 be the first and the second component mappings from $(\mathbb{R}_+ \times \mathcal{P}(E_n))$ into $\mathbb{R}_+$, and from $(\mathbb{R}_+ \times \mathcal{P}(E_n))$ into $\mathcal{P}(E_n)$. The mean field particle approximation associated with the equation (7.3.5) relies on the fact that it is possible to rewrite the mapping Γ_{n+1}^2 in the following form

$$\Gamma_{n+1}^2(\gamma_n(1), \eta_n) = \eta_n K_{n+1,(\gamma_n(1),\eta_n)} \tag{7.3.6}$$

where $K_{n+1,(m,\eta)}$ is a Markov kernel indexed by the time parameter n, a mass parameter $m \in \mathbb{R}_+$ and a probability measure η on the space E_n. In the literature on mean field particle systems, $K_{n,(m,\eta)}$ is called a McKean transition. The choice of such Markov transitions $K_{n,(m,\eta)}$ is not unique and will be discussed in section 7.6.1.

Before concluding this section, we note that

$$\gamma_{n+1}(dx') = (\gamma_n Q_{n+1})(dx') := \int \gamma_n(dx)\ Q_{n+1}(x, dx') \tag{7.3.7}$$

when $\mu_n = 0$. In this particular situation, the solution of the equation (7.3.2) is given by the following Feynman-Kac path integral formulae

$$\gamma_n(f) = \gamma_0(1)\ \mathbb{E}\left(f(X_n) \prod_{0 \leq p < n} G_p(X_p)\right) \tag{7.3.8}$$

where X_n stands for a Markov chain taking values in the state spaces E_n with initial distribution $\eta_0 = \gamma_0/\gamma_0(1)$ and Markov transitions M_n (see for instance section 1.4.4.in [28]). These measure-valued equations have been studied at length in [28].

7.3.3 Mean field particle interpretation

The transport formula presented in (7.3.6) provides a natural interpretation of the probability distributions η_n as the laws of a nonlinear Markov chain $\overline{X}_n$ whose elementary transitions $\overline{X}_n \rightsquigarrow \overline{X}_{n+1}$ depends on the distribution $\eta_n = \text{Law}(\overline{X}_n)$ as well as on the current mass $\gamma_n(1)$. In contrast to the more traditional McKean type nonlinear Markov chains presented in [28], the dependency on the mass process

induces a dependency of the whole sequence of measures η_p, from the origin $p = 0$ up to the current time $p = n$.

From now on, we will always assume that the mappings

$$\left(m, \left(x^i\right)_{1\leq i\leq N}\right) \in \left(\mathbb{R}_+ \times E_n^N\right) \mapsto K_{n+1,\left(m,\frac{1}{N}\sum_{i=1}^N \delta_{x^i}\right)}\left(x, A_{n+1}\right)$$

are measurable w.r.t the product sigma fields on $(\mathbb{R}_+ \times E_n^N)$, for any $n \geq 0$, $N \geq 1$, and $1 \leq i \leq N$, and any measurable subset $A_{n+1} \subset E_{n+1}$. In this situation, the mean field particle interpretation of (7.3.6) is an E_n^N-valued nonlinear Markov chain $\xi_n^{(N)} = \left(\xi_n^{(N,i)}\right)_{1\leq i\leq N}$ with transitions defined as

$$\left\{\begin{array}{rcl} \gamma_{n+1}^N(1) & = & \gamma_n^N(1)\ \eta_n^N(G_n) + \mu_{n+1}(1) \\ \\ \mathbb{P}\left(\xi_{n+1}^{(N)} \in dx \ \middle|\ \mathcal{F}_n^{(N)}\right) & = & \prod_{i=1}^N\ K_{n+1,(\gamma_n^N(1),\eta_n^N)}(\xi_n^{(N,i)}, dx^i) \end{array}\right. \tag{7.3.9}$$

with the pair of occupation measures $\left(\gamma_n^N, \eta_n^N\right)$ defined below

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^{(N,i)}} \quad \text{and} \quad \gamma_n^N(dx) := \gamma_n^N(1)\ \eta_n^N(dx)$$

In the above displayed formula, $\mathcal{F}_n^N$ stands for the σ-field generated by the random sequence $(\xi_p^{(N)})_{0\leq p\leq n}$, and $dx = dx^1 \times \ldots \times dx^N$ stands for an infinitesimal neighborhood of a point $x = (x^1, \ldots, x^N) \in E_n^N$. The initial system $\xi_0^{(N)}$ consists of N independent and identically distributed random variables with common law η_0. As usual, to simplify the presentation, we will suppress the parameter N when there is no possible confusion, so that we write ξ_n and ξ_n^i instead of $\xi_n^{(N)}$ and $\xi_n^{(N,i)}$.

In the above discussion, we have implicitly assumed that the quantities $\mu_n(1)$ are known and that it is easy to sample from the probability distribution $\overline{\mu}_n(dx) := \mu_n(dx)/\mu_n(1)$. In practice, we often need to resort to an additional approximation scheme to approximate $\mu_n(1)$ and $\overline{\mu}_n$. This situation is discussed in section 7.7. This additional level of approximation has essentially a minimal impact on the properties of the particle approximation scheme which can be analyzed using the same tools.

7.3.4 Notation

For the convenience of the reader, we end this introduction with some notation used in the present article. We denote by $\mathcal{M}(E)$ the set of measures on some measurable state space $(E, \mathcal{E})$ and we recall that $\mathcal{P}(E)$ is the set of probability measures. We also denote $\mathcal{B}(E)$ the Banach space of all bounded and measurable functions f equipped with the uniform norm $\|f\|$ and $\mathrm{Osc}_1(E)$ the convex set of $\mathcal{E}$-measurable functions f with oscillations $\mathrm{osc}(f) \leq 1$ where $\mathrm{osc}(f) = \sup_{(x,y)\in E^2} |f(x) - f(y)|$.

We let $\mu(f) = \int \mu(dx)\ f(x)$ be the Lebesgue integral of a function $f \in \mathcal{B}(E)$ with respect to a measure $\mu \in \mathcal{M}(E)$. We recall that a bounded integral kernel $M(x, dy)$ from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(E', \mathcal{E}')$ is an operator $f \mapsto M(f)$ from $\mathcal{B}(E')$ into $\mathcal{B}(E)$ such that the functions $x \mapsto M(f)(x) := \int_{E'} M(x, dy) f(y)$ are $\mathcal{E}$-measurable and bounded for any $f \in \mathcal{B}(E')$. The kernel M also generates a dual operator $\mu \mapsto \mu M$ from $\mathcal{M}(E)$ into $\mathcal{M}(E')$ defined by $(\mu M)(f) := \mu(M(f))$. A Markov kernel is a positive and bounded integral operator M with $M(1)(x) = 1$ for any $x \in E$. Given a pair of bounded integral operators (M_1, M_2), we let $(M_1 M_2)$ be the composition operator defined by $(M_1 M_2)(f) = M_1(M_2(f))$. For time-homogenous state spaces, we denote by $M^k = M^{k-1} M = M M^{k-1}$ the k-th composition of a given bounded integral operator M, with $k \geq 0$, with the convention $M^0 = Id$ the identity operator. We also use the notation

$$M\left([f_1 - M(f_1)]\,[f_2 - M(f_2)]\right)(x) := M\left([f_1 - M(f_1)(x)]\,[f_2 - M(f_2)(x)]\right)(x)$$

for some bounded functions f_1, f_2.

We also denote the total variation norm on $\mathcal{M}(E)$ by $\|\mu\|_{\mathrm{tv}} = \sup_{f \in \mathrm{Osc}_1(E)} |\mu(f)|$. When the bounded integral operator M has a constant mass, that is $M(1)(x) = M(1)(y)$ for any $(x, y) \in E^2$, the operator $\mu \mapsto \mu M$ maps $\mathcal{M}(E)$ into $\mathcal{M}(E')$. In this situation, we let $\beta(M)$ be the Dobrushin coefficient of a bounded integral operator M defined by the following formula

$$\beta(M) := \sup\ \{\mathrm{osc}(M(f)); f \in \mathrm{Osc}_1(E)\}$$

Given a positive function G on E, we let $\Psi_G\ :\ \eta \in \mathcal{P}(E) \mapsto \Psi_G(\eta) \in \mathcal{P}(E)$ be the Boltzmann-Gibbs transformation defined by

$$\Psi_G(\eta)(dx) := \frac{1}{\eta(G)}\ G(x)\ \eta(dx)$$

We recall that $\Psi_G(\eta)$ can be expressed in terms of a Markov transport equation

$$\eta S_\eta = \Psi_G(\eta) \tag{7.3.10}$$

for some selection type transition $S_\eta(x, dy)$. For instance, by noticing that, for any $\epsilon \geq 0$ s.t. $G(x) \geq \epsilon$

$$\Psi_{(G-\epsilon)}(\eta) = \frac{\eta(G)}{\eta(G) - \epsilon} \left(\Psi_{(G)}(\eta) - \frac{\epsilon \eta}{\eta(G)} \right)$$

we can take

$$S_\eta(x, dy) := \frac{\epsilon}{\eta(G)}\ \delta_x(dy) + \left(1 - \frac{\epsilon}{\eta(G)}\right)\ \Psi_{(G-\epsilon)}(\eta)(dy) \tag{7.3.11}$$

for any $\epsilon \geq 0$ s.t. $G(x) \geq \epsilon$. Notice that for $\epsilon = 0$, we have $S_\eta(x, dy) = \Psi_G(\eta)(dy)$. We can also choose

$$S_\eta(x, dy) := \epsilon G(x)\ \delta_x(dy) + (1 - \epsilon G(x))\ \Psi_G(\eta)(dy) \tag{7.3.12}$$

for any $\epsilon \geq 0$ that may depend on the current measure η, and s.t. $\epsilon G(x) \leq 1$. For instance, we can choose $1/\epsilon$ to be the η-essential supremum of G.

7.4 Statement of the main results

At the end of section 7.3.2, we have seen the evolution equation (7.3.2) coincides with that of a Feynman-Kac model (7.3.8) for $\mu_n = 0$. In this specific situation, the distributions γ_n are simply given by the recursive equation

$$\gamma_n = \gamma_{n-1} Q_n \implies \forall 0 \leq p \leq n \qquad \gamma_n = \gamma_p Q_{p,n} \quad \text{with} \quad Q_{p,n} = Q_{p+1} \ldots Q_{n-1} Q_n \tag{7.4.1}$$

For $p = n$, we use the convention $Q_{n,n} = Id$. In addition, the nonlinear semigroup associated to this sequence of distributions is given by

$$\eta_n(f) = \Phi_{p,n}(\eta_p)(f) := \eta_p Q_{p,n}(f)/\eta_p Q_{p,n}(1) = \eta_p \left(Q_{p,n}(1) P_{p,n}(f)\right)/\eta_p Q_{p,n}(1) \tag{7.4.2}$$

with the Markov kernel $P_{p,n}(x_p, dx_n) = Q_{p,n}(x_p, dx_n)/Q_{p,n}(x_p, E_n)$. The analysis of the mean field particle interpretations of such models have been studied in [28]. Various properties including contraction inequalities, fluctuations, large deviations and concentration properties have been developed for this class of models. In this context, the fluctuations properties as well as $\mathbb{L}_r$-mean error estimates, including uniform estimates w.r.t. the time parameter are often expressed in terms of two central parameters:

$$q_{p,n} = \sup_{x,y} \frac{Q_{p,n}(1)(x)}{Q_{p,n}(1)(y)} \quad \text{and} \quad \beta(P_{p,n}) = \sup_{x,y \in E_p} \|P_{p,n}(x, .) - P_{p,n}(y, .)\|_{\text{tv}} \tag{7.4.3}$$

with the pair of Feynman-Kac semigroups $(P_{p,n}, Q_{p,n})$ introduced in (7.4.1) and (7.4.2).

We also consider the pair of parameters $(g_-(n), g_+(n))$ defined below

$$g_-(n) = \inf_{0 \leq p < n} \inf_{E_p} G_p \leq \sup_{0 \leq p < n} \sup_{E_p} G_p = g_+(n)$$

The first main objective is to extend some of these properties to models where μ_n is non necessarily null. We illustrate our estimates in three typical scenarios

$$1) \quad G = g_{-/+} = 1 \qquad 2) \quad g_+ < 1 \quad \text{and} \quad 3) \quad g_- > 1 \tag{7.4.4}$$

arising in time homogeneous models

$$(E_n, G_n, M_n, \mu_n, g_-(n), g_+(n)) = (E, G, M, \mu, g_-, g_+) \tag{7.4.5}$$

These three scenarios correspond to the case where, *independently from the additional spontaneous births*, the existing targets die or survive and spawn in such a way that either their number remains constant ($G = g_{-/+} = 1$), decreases ($g_+ < 1$) or increases ($g_- > 1$).

Our first main result concerns three different types of long time behavior for these three types of models. This result can basically be stated as follows.

Theorem 7.4.1. *For time homogeneous models (7.4.5), the limiting behavior of* $(\gamma_n(1), \eta_n)$ *in the three scenarios (7.4.4) is as follows:*

1. *When* $G(x) = 1$ *for any* $x \in E$*, we have*

$$\gamma_n(1) = \gamma_0(1) + \mu(1)\ n \quad \textit{and} \quad \|\eta_n - \eta_\infty\|_{\mathrm{tv}} = O\left(\frac{1}{n}\right)$$

when M *is chosen so that*

$$\sum_{n\geq 0} \sup_{x\in E} \|M^n(x, \bullet) - \eta_\infty\|_{\mathrm{tv}} < \infty \quad \textit{for some invariant measure } \eta_\infty = \eta_\infty M. \tag{7.4.6}$$

2. *When* $g_+ < 1$*, there exists a constant* $c < \infty$ *such that*

$$\forall f \in \mathcal{B}(E), \qquad |\gamma_n(f) - \gamma_\infty(f)| \vee |\eta_n(f) - \eta_\infty(f)| \leq c\ g_+^n\ \|f\|$$

with the limiting measures

$$\gamma_\infty(f) := \sum_{n\geq 0} \mu Q^n(f) \ \textit{and}\ \eta_\infty(f) := \gamma_\infty(f)/\gamma_\infty(1) \tag{7.4.7}$$

3. *When* $g_- > 1$ *and there exist* $k \geq 1$ *and* $\epsilon > 0$ *such that* $M^k(x, \bullet) \geq \epsilon\ M^k(y, \bullet)$ *for any* $x, y \in E$ *then the mapping* $\Phi = \Phi_{n-1,n}$ *introduced in (7.4.2) has a unique fixed point* $\eta_\infty = \Phi(\eta_\infty)$ *and*

$$\lim_{n\to\infty} \frac{1}{n} \log \gamma_n(1) = \log \eta_\infty(G) \quad \textit{and} \quad \|\eta_n - \eta_\infty\|_{\mathrm{tv}} \leq c\ e^{-\lambda n}$$

for some finite constant $c < \infty$ *and some* $\lambda > 0$*.*

A more precise statement and a detailed proof of the above theorem can be found in section 7.5.2.

Our second main result concerns the convergence of the mean field particle approximations presented in (7.3.9). We provide rather sharp non asymptotic estimates including uniform convergence results w.r.t. the time parameter. Our results can be basically stated as follows.

Theorem 7.4.2. *For any $n \geq 0$, and any $N \geq 1$, we have $\gamma_n(1)$ and $\gamma_n^N(1) \in I_n$ with the compact interval I_n defined below*

$$I_n := [m_-(n), m_+(n)] \quad \text{where} \quad m_{-/+}(n) := \sum_{p=0}^{n} \mu_p(1) g_{-/+}(n)^{(n-p)} \tag{7.4.8}$$

In addition, for any $r \geq 1$, $f \in Osc_1(E_n)$, and any $N \geq 1$, we have

$$\sqrt{N}\ \mathbb{E}\left(\left|\left[\eta_n^N - \eta_n\right](f)\right|^r\right)^{\frac{1}{r}} \leq a_r\ b_n \quad \text{with} \quad b_n \leq \sum_{p=0}^{n} b_{p,n} \tag{7.4.9}$$

where $a_r < \infty$ stands for a constant whose value only depends on the parameter r and $b_{p,n}$ is the collection of constants given by

$$b_{p,n} := 2\ (1 \wedge m_{p,n})\ q_{p,n} \left[q_{p,n}\ \beta(P_{p,n}) + \sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\ \beta(P_{q,n}) \right] \tag{7.4.10}$$

with the pair of parameters

$$m_{p,n} = m_+(p)\|Q_{p,n}(1)\| / \sum_{p<q\leq n} c_{q,n} \quad \text{and} \quad c_{p,n} := \mu_p Q_{p,n}(1)$$

Furthermore, the particle measures γ_n^N are unbiased, and for the three scenarios (7.4.4) with time homogenous models s.t. $M^k(x, \boldsymbol{.}) \geq \epsilon\ M^k(y, \boldsymbol{.})$, for any $x, y \in E$ and some pair of parameters $k \geq 1$ and $\epsilon > 0$, the constant b_n in (7.4.9) can be chosen so that $\sup_{n\geq 0} b_n < \infty$; in addition, we have the non asymptotic variance estimates for some $d < \infty$, any $n \geq 1$ and for any $N > 1$

$$\mathbb{E}\left(\left[\frac{\gamma_n^N(1)}{\gamma_n(1)} - 1\right]^2\right) \leq d\ \frac{n+1}{N-1}\ \left(1 + \frac{d}{N-1}\right)^{n-1} \tag{7.4.11}$$

The non asymptotic estimates stated in the above theorem extend the one presented in [16, 28] for Feynman-Kac type models (7.3.8) where $\mu_n = 0$. For such models, the $\mathbb{L}_r$-mean error estimates (7.4.9) are satisfied with the collection of parameters $b_{p,n} := 2q_{p,n}^2\ \beta(P_{p,n})$, with $p \leq n$. The extra terms in (7.4.10) are intimately related to μ_n whose effects in the semigroup stability depend on the nature of G_n. We refer to theorem 7.4.1, section 7.5.2 and section 7.5.3, for a discussion on three different behaviors in the three cases presented in (7.4.4).

A direct consequence of this theorem is that it implies the almost sure convergence results:

$$\lim_{N\to\infty} \eta_n^N(f) = \eta_n(f) \quad \text{and} \quad \lim_{N\to\infty} \gamma_n^N(f) = \gamma_n(f)$$

for any bounded function $f \in \mathcal{B}(E_n)$.

Our last main result, is a functional central limit theorem. We let W_n^N be the centered random fields defined by the following formulae

$$\eta_n^N = \eta_{n-1}^N K_{n,(\gamma_n^N(1),\eta_{n-1}^N)} + \frac{1}{\sqrt{N}}\, W_n^N \,. \tag{7.4.12}$$

We also consider the pair of random fields

$$V_n^{\eta,N} := \sqrt{N}[\eta_n^N - \eta_n] \quad \text{and} \quad V_n^{\gamma,N} := \sqrt{N}[\gamma_n^N - \gamma_n]$$

For $n = 0$, we use the convention $W_0^N = V_0^{\eta,N}$.

Theorem 7.4.3. *The sequence of random fields $(W_n^N)_{n\geq 0}$ converges in law, as N tends to infinity, to the sequence of n independent, Gaussian and centered random fields $(W_n)_{n\geq 0}$ with a covariance function given for any $f, g \in \mathcal{B}(E_n)$ and $n \geq 0$ by*

$$\begin{aligned}&\mathbb{E}(W_n(f)W_n(g))\\ &= \eta_{n-1}K_{n,(\gamma_{n-1}(1),\eta_{n-1})}\left([f - K_{n,(\gamma_{n-1}(1),\eta_{n-1})}(f)][g - K_{n,(\gamma_{n-1}(1),\eta_{n-1})}(g)]\right).\end{aligned} \tag{7.4.13}$$

In addition, the pair of random fields $V_n^{\gamma,N}$ and $V_n^{\eta,N}$ converge in law as $N \to \infty$ to a pair of centered Gaussian fields V_n^γ and V_n^η defined by

$$V_n^\gamma(f) := \sum_{p=0}^{n} \gamma_p(1)\; W_p(Q_{p,n}(f)) \quad \textit{and} \quad V_n^\eta(f) := V_n^\gamma\left(\frac{1}{\gamma_n(1)}(f - \eta_n(f))\right)$$

The details of the proof of theorem 7.4.2 and theorem 7.4.3 can be found in section 7.6.2 dedicated to the convergence of the unnormalized particle measures γ_n^N. The proof of the non-asymptotic variance estimate (7.4.11) is given in section 7.6.2 as well as the $\mathbb{L}_r$-mean error estimates (7.4.9) and the fluctuation theorem 7.4.3. Under additional regularity conditions, we conjecture that it is possible to obtain uniform estimates for theorem 7.4.3 but have not established it here.
The rest of the chapter is organized as follows: in section 7.5, we analyze the semigroup properties of the total mass process $\gamma_n(1)$ and the sequence of probability distributions η_n. This section is mainly concerned with the proof of theorem 7.4.1. The long time behavior of the total mass process is discussed in section 7.5.1, while the asymptotic behavior of the probability distributions is discussed in section 7.5.2. In section 7.5.3, we develop a series of Lipschitz type functional inequalities for uniform estimates w.r.t. the time parameter for the particle approximation. In section 7.6, we present the McKean models associated to the sequence $(\gamma_n(1), \eta_n)$ and their mean field particle interpretations. Section 7.6.2 is concerned with the convergence analysis of these particle approximations. We discuss the convergence of the approximations of $\gamma_n(1)$, including their unbiasedness property and the non asymptotic variance estimates presented in (7.4.11). The proof of the $\mathbb{L}_r$-mean error

estimates (7.4.9) is presented in section 7.6.3. The proof of the functional central limit theorem 7.4.3 is a more or less direct consequence of the decomposition formulae presented in section 7.6.2 and is just sketched at the end of this very section.

7.5 Semigroup analysis

The purpose of this section is to analyze the semigroup properties of the intensity measure recursion (7.3.2). We establish a framework for the analysis of the long time behavior of these measures and their particle approximations (7.3.9). First, we briefly recall some estimate of the quantities $(q_{p,n}, \beta(P_{p,n}))$ in terms of the potential functions G_n and the Markov transitions M_n. Further details on this subject can be found in [28], and in references therein.

We assume here that the following condition is satisfied for some $k \geq 1$, some collection of numbers $\epsilon_p \in (0,1)$

$$M_{p,p+k}(x_p, .) \geq \epsilon_p \; M_{p,p+k}(y_p, .) \;\; \text{with} \;\; M_{p,p+k} = M_{p+1} M_{p+2} \ldots M_{p+k} \tag{7.5.1}$$

for any time parameter p and any pair of states $(x_p, y_p) \in E_p^2$. It is well known that the mixing type condition $(M)_k$ is satisfied for any aperiodic and irreducible Markov chains on finite spaces, as well as for bi-Laplace exponential transitions associated with a bounded drift function and for Gaussian transitions with a mean drift function that is constant outside some compact domain. We introduce the following quantities

$$\delta_{p,n} := \sup \prod_{p \leq q < n} (G_q(x_q)/G_q(y_q)) \quad \text{and} \quad \delta_p^{(k)} := \delta_{p+1,p+k} \tag{7.5.2}$$

where the supremum is taken over all admissible pair of paths with transitions M_q where an admissible path $(x_{p-1}, x_{p+1}, ..., x_{n-1})$ is such that $\prod_{p \leq q < n} M_{\,q}\,(x_{q-1}, dx_q) > 0$. Under the above conditions, we have [28, p. 140]

$$\beta(P_{p,p+n}) \leq \prod_{l=0}^{\lfloor n/k \rfloor - 1} \left(1 - \epsilon_{p+lk}^2 / \delta_{p+lk}^{(k)}\right) \quad \text{and} \quad q_{p,p+n} \leq \delta_{p,p+k}/\epsilon_p \tag{7.5.3}$$

For time-homogeneous Feynman-Kac models we set $\epsilon := \epsilon_k$ and $\delta_k := \delta_{0,k}$, for any $k \geq 0$. Using this notation, the above estimates reduce to [28, p. 142]

$$q_{p,p+n} \leq \; \delta_k/\epsilon \quad \text{and} \quad \beta(P_{p,p+n}) \leq \left(1 - \epsilon^2/\delta_{k-1}\right)^{\lfloor n/k \rfloor} \tag{7.5.4}$$

7.5.1 Description of the models

The next proposition gives a Markov transport formulation of Γ_n introduced in (7.3.5).

Proposition 7.5.1. *For any $n \geq 0$, we have the recursive formula*

$$\left\{ \begin{array}{rcl} \gamma_{n+1}(1) & = & \gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1) \\ \\ \eta_{n+1} & = & \Psi_{G_n}(\eta_n) M_{n+1,(\gamma_n(1),\eta_n)} \end{array} \right. \tag{7.5.5}$$

with the collection of Markov transitions $M_{n+1,(m,\eta)}$ indexed by the parameters $m \in \mathbb{R}_+$ and the probability measures $\eta \in \mathcal{P}(E_n)$ given below

$$M_{n+1,(m,\eta)}(x, dy) := \alpha_n(m, \eta)\, M_{n+1}(x, dy) + (1 - \alpha_n(m, \eta))\ \overline{\mu}_{n+1}(dy) \tag{7.5.6}$$

with the collection of $[0, 1]$-parameters $\alpha_n(m, \eta)$ defined below

$$\alpha_n(m, \eta) = \frac{m\eta(G_n)}{m\eta(G_n) + \mu_{n+1}(1)}$$

Proof:
Observe that for any function $f \in \mathcal{B}(E_{n+1})$, we have that

$$\eta_{n+1}(f) = \frac{\gamma_n(G_n M_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n(G_n) + \mu_{n+1}(1)} = \frac{\gamma_n(1)\ \eta_n(G_n M_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1)}$$

from which we find that

$$\eta_{n+1} = \alpha_n(\gamma_n(1), \eta_n)\ \ \Phi_{n+1}(\eta_n) + (1 - \alpha_n(\gamma_n(1), \eta_n))\ \ \overline{\mu}_{n+1}$$

From these observations, we prove (7.5.5). This ends the proof of the proposition. ■

We let Γ_{n+1} be the mapping from $\mathbb{R}_+ \times \mathcal{P}(E_n)$ into $\mathbb{R}_+ \times \mathcal{P}(E_{n+1})$ given by

$$\Gamma_{n+1}(m, \eta) = \left(\Gamma^1_{n+1}(m, \eta), \Gamma^2_{n+1}(m, \eta)\right) \tag{7.5.7}$$

with the pair of transformations:

$$\Gamma^1_{n+1}(m, \eta) = m\ \eta(G_n) + \mu_{n+1}(1) \quad \text{and} \quad \Gamma^2_{n+1}(m, \eta) = \Psi_{G_n}(\eta) M_{n+1,(m,\eta)}$$

We also denote by $(\Gamma_{p,n})_{0 \leq p \leq n}$ the corresponding semigroup defined by

$$\forall 0 \leq p \leq n \qquad \Gamma_{p,n} = \Gamma_{p+1,n}\Gamma_{p+1} = \Gamma_n \Gamma_{n-1} \ldots \Gamma_{p+1}$$

with the convention $\Gamma_{n,n} = Id$.

The following lemma collects some important properties of the sequence of intensity measures γ_n.

Lemma 7.5.2. *For any $0 \leq p \leq n$, we have the semigroup decomposition*

$$\gamma_n = \gamma_p Q_{p,n} + \sum_{p<q\leq n} \mu_q Q_{q,n} \quad \text{and} \quad \gamma_n = \sum_{0\leq p\leq n} \mu_p Q_{p,n} \tag{7.5.8}$$

In addition, we also have the following formula

$$\gamma_n(1) = \sum_{p=0}^{n} \mu_p(1) \prod_{p\leq q<n} \eta_q(G_q) \tag{7.5.9}$$

Proof:

The first pair of formulae are easily proved using a simple induction, and recalling that $\gamma_0 = \mu_0$. To prove the last assertion, we use an induction on the parameter $n \geq 0$. The result is obvious for $n = 0$. We also have by (7.3.2)

$$\gamma_{n+1}(1) = \gamma_n Q_{n+1}(1) + \mu_{n+1}(1) = \gamma_n(G_n) + \mu_{n+1}(1)$$

This implies

$$\begin{aligned}
\gamma_{n+1}(1) &= \gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1) \\
&= \gamma_{n-1}(1)\ \eta_{n-1}(G_{n-1})\ \eta_n(G_n) + \mu_n(1)\ \ \eta_n(G_n) + \mu_{n+1}(1) \\
&= \ldots \\
&= \gamma_0(1) \prod_{p=0}^{n} \eta_p(G_p)\ + \sum_{p=1}^{n+1} \mu_p(1) \prod_{p\leq q\leq n} \eta_q(G_q)
\end{aligned}$$

Recalling that $\gamma_0(dx_0) = \mu_0(dx_0)$, we prove (7.5.9). This ends the proof of the lemma. ∎

Using lemma 7.5.2, one proves that the semigroup $\Gamma_{p,n}$ satisfies the pair of formulae described below

Proposition 7.5.3. *For any $0 \leq p \leq n$, we have*

$$\Gamma^1_{p,n}(m,\eta) = m\ \eta Q_{p,n}(1) + \sum_{p<q\leq n} \mu_q Q_{q,n}(1) \tag{7.5.10}$$

$$\Gamma^2_{p,n}(m,\eta) = \alpha_{p,n}(m,\eta)\ \ \Phi_{p,n}(\eta) + (1 - \alpha_{p,n}(m,\eta)) \sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\ \Phi_{q,n}(\overline{\mu}_q) \tag{7.5.11}$$

with the collection of parameters $c_{p,n} := \mu_p Q_{p,n}(1)$ and the $[0,1]$-valued parameters $\alpha_{p,n}(m,\eta)$ defined below

$$\alpha_{p,n}(m,\eta) = \frac{m\eta Q_{p,n}(1)}{m\eta Q_{p,n}(1) + \sum_{p<q\leq n} c_{q,n}} \leq \alpha^{\star}_{p,n}(m) := 1 \wedge \left[m \left| \frac{Q_{p,n}(1)}{\sum_{p<q\leq n} c_{q,n}} \right| \right] \tag{7.5.12}$$

One central question in the theory of spatial branching point processes is the long time behavior of the total mass process $\gamma_n(1)$. Notice that $\gamma_n(1) = \mathbb{E}(\mathcal{X}_n(1))$ is the expected size of the n-th generation. For time homogeneous models with null spontaneous branching $\mu_n = \mu = 0$, the exponential growth of these quantities are related to the logarithmic Lyapunov exponents of the semigroup $Q_{p,n}$. The prototype of these models is the Galton-Watson branching process. In this context three typical situations may occur: 1) $\gamma_n(1)$ remains constant and equals to the initial mean number of individuals. 2) $\gamma_n(1)$ goes exponentially fast to 0, 3) $\gamma_n(1)$ grows exponentially fast to infinity,

The analysis of spatial branching point processes with $\mu_n = \mu \neq 0$ considered here is more involved. Loosely speaking, in the first situation discussed above the total mass process is generally strictly increasing; while in the second situation the additional mass injected in the system stabilizes the total mass process. Before giving further details, by lemma 7.5.2 we observe $\gamma_n(1) \in I_n$, for any $n \geq 0$, with the compact interval I_n defined in 7.4.8.
We end this section with a more precise analysis of the effect of μ in the three scenarios (7.4.4).

In the further developments of this section, we illustrate the stability properties of the sequence of probability distributions η_n in these three scenarios.

1. When $G(x) = 1$ for any $x \in E$, the total mass process $\gamma_n(1)$ grows linearly w.r.t. the time parameter and we have

$$\gamma_n(1) = m_-(n) = m_+(n) = \gamma_0(1) + \mu(1)\ n \tag{7.5.13}$$

   Note that the estimates in (7.5.12) take the following form

$$\alpha_{p,n}\left(\gamma_p(1), \eta_p\right) \leq \alpha^{\star}_{p,n}(\gamma_p(1)) := 1 \wedge \frac{\gamma_0(1) + \mu(1)\ p}{\mu(1)\ (n-p)} \to_{(n-p)\to\infty} 0$$

2. When $g_+ < 1$, the total mass process $\gamma_n(1)$ is uniformly bounded w.r.t the time parameter. More precisely, we have that

$$m_{-/+}(n) = g^n_{-/+}\ \gamma_0(1) + \left(1 - g^n_{-/+}\right)\ \frac{\mu(1)}{1 - g_{-/+}}$$

   This yields the rather crude estimates

$$\gamma_0(1) \wedge \frac{\mu(1)}{1-g_-} \leq \gamma_n(1) \leq \gamma_0(1) \vee \frac{\mu(1)}{1-g_+} \tag{7.5.14}$$

   We end this discussion with an estimate of the parameter $\alpha_{p,n}(m)$ given in (7.5.12). When the mixing condition $(M)_k$ stated in (7.5.1) is satisfied for some k and some fixed parameters $\epsilon_p = \epsilon$, using (7.5.4) we prove that

$$\sum_{p<r\leq n} \frac{\mu Q_{r,n}(1)}{Q_{p,r}(Q_{r,n}(1))} \geq \frac{\epsilon\mu(1)}{\delta_k} \sum_{p<r\leq n} \frac{1}{Q_{p,r}(1)} \geq \frac{\epsilon\mu(1)}{\delta_k}\ \frac{g_+^{-(n-p)} - 1}{1 - g_+}$$

from which we conclude that for any $n > p$ and any $m \in I_p$

$$\begin{aligned}\alpha^{\star}_{p,n}(m) &\leq 1 \wedge \left[m\ g_+^{(n-p)}\ \frac{\delta_k\ (1-g_+)}{\epsilon\mu(1)(1-g_+^{(n-p)})} \right] \\ &\leq 1 \wedge \left[m\ g_+^{(n-p)}\ \delta_k/(\epsilon\mu(1)) \right] \\ &\leq 1 \wedge \left[\left(\gamma_0(1) \vee \frac{\mu(1)}{1-g_+} \right)\ g_+^{(n-p)}\ \delta_k/(\epsilon\mu(1)) \right] \to_{(n-p)\to\infty} 0 \quad (7.5.15)\end{aligned}$$

3. When $g_- > 1$, the total mass process $\gamma_n(1)$ grows exponentially fast w.r.t the time parameter and we can easily show that

$$g_- > 1 \Longrightarrow \gamma_n(1) \geq m_-(n) = \gamma_0(1)\ g_-^n + \mu(1)\ \frac{g_-^n - 1}{g_- - 1} \quad (7.5.16)$$

7.5.2 Asymptotic properties

This section is concerned with the long time behavior of the semigroups $\Gamma_{p,n}$ in the three scenarios discussed in (7.5.13), (7.5.14), and (7.5.16). Our results are summarized in theorem 7.4.1. We consider time-homogeneous models $(E_n, G_n, M_n, \mu_n) = (E, G, M, \mu)$.

1. When $G(x) = 1$ for any $x \in E$, we have seen in (7.5.13) that $\gamma_n(1) = \gamma_0(1) + \mu(1)\ n$. In this particular situation, the time-inhomogeneous Markov transitions $M_{n,(\gamma_{n-1}(1),\eta_{n-1})} := \overline{M}_n$ introduced in (7.5.5) are given by

$$\overline{M}_n(x, dy) = \left(1 - \frac{\mu(1)}{\gamma_0(1) + n\mu(1)} \right)\ M(x, dy) + \frac{\mu(1)}{\gamma_0(1) + n\mu(1)}\ \overline{\mu}(dy)$$

This shows that $\eta_n = \mathrm{Law}(\overline{X}_n)$ can be interpreted as the distribution of the states $\overline{X}_n$ of a time inhomogeneous Markov chain with transitions $\overline{M}_n$ and initial distribution η_0. If we choose in (7.3.6) $K_{n+1,(\gamma_n(1),\eta_n)} = \overline{M}_{n+1}$, the N-particle model (7.3.9) reduces to a series of N independent copies of $\overline{X}_n$. In this situation, the mapping $\Gamma^2_{0,n}$ is given by

$$\Gamma^2_{0,n}(\gamma_0(1), \eta_0) := \frac{\gamma_0(1)}{\gamma_0(1) + n\mu(1)}\ \ \eta_0 M^n + \frac{n\mu(1)}{\gamma_0(1) + n\mu(1)}\ \ \frac{1}{n} \sum_{0 \leq p < n} \overline{\mu} M^p$$

The above formula shows that for a large time horizon n, the normalized distribution flow η_n is almost equal to $\frac{1}{n}\sum_{0\leq p<n} \overline{\mu}M^p$. Let us assume that the Markov kernel M is chosen so that (7.4.6) is satisfied for some invariant measure $\eta_\infty = \eta_\infty M$. In this case, for any starting measure γ_0, we have

$$\|\eta_n - \eta_\infty\|_{\mathrm{tv}} \leq \frac{\gamma_0(1)}{\gamma_0(1) + n\mu(1)}\ \tau_n + \frac{n\mu(1)}{\gamma_0(1) + n\mu(1)}\ \ \frac{1}{n} \sum_{0 \leq p < n} \tau_p = O\left(\frac{1}{n} \right)$$

with $\tau_n = \sup_{x\in E} \|M^n(x,\,.\,) - \eta_\infty\|_{\rm tv}$. For instance, suppose the mixing condition $(M)_k$ presented in (7.5.1) is met for some $k \geq 1$ and $\epsilon > 0$. In this case, the above upper bound is satisfied with $\tau_n = (1-\epsilon)^{\lfloor n/k \rfloor}$.

2. Consider the case where $g_+ < 1$. In this situation, the pair of measures (7.4.7) are well defined. Furthermore, for any $f \in \mathcal{B}(E)$ with $\|f\| \leq 1$, we have the estimates

$$\begin{aligned} |\gamma_n(f) - \gamma_\infty(f)| &\leq \gamma_0(1)\ \eta_0 Q^n(1) + \sum_{p\geq n} \mu Q^p(1) \\ &\leq g_+^n\ [\gamma_0(1) + \mu(1)/(1-g_+)] \longrightarrow_{n\to\infty} 0 \end{aligned}$$

In addition, using the fact that $\gamma_n(1) \geq \mu(1)$, we find that for any $f \in {\rm Osc}_1(E)$

$$\begin{aligned} |\eta_n(f) - \eta_\infty(f)| &\leq \frac{1}{\gamma_n(1)} \, |\gamma_n[f - \eta_\infty(f)] - \gamma_\infty[f - \eta_\infty(f)]| \\ &\leq g_+^n\ [\gamma_0(1)/\mu(1) + 1/(1-g_+)] \longrightarrow_{n\to\infty} 0 \end{aligned}$$

3. Consider the case where $g_- > 1$. We further assume that the mixing condition $(M)_k$ presented in (7.5.1) is met for some $k \geq 1$ and some fixed parameters $\epsilon_p = \epsilon > 0$. In this situation, it is well known that the mapping $\Phi = \Phi_{n-1,n}$ introduced in (7.4.2) has a unique fixed point $\eta_\infty = \Phi(\eta_\infty)$, and for any initial distribution η_0, we have

$$\|\Phi_{0,n}(\eta_0) - \eta_\infty\|_{\rm tv} \leq a\ e^{-\lambda\ n} \tag{7.5.17}$$

with

$$\lambda = -\frac{1}{k} \log\left(1 - \epsilon^2/\delta_{0,k-1}\right) \quad \text{and} \quad a = 1/\left(1 - \epsilon^2/\delta_{0,k-1}\right)$$

as well as

$$\sup_{\eta\in\mathcal{P}(E)} \left| \frac{1}{n} \log \eta Q^n(1) - \log \eta_\infty(G) \right| \leq b/n \tag{7.5.18}$$

for some finite constant $b < \infty$. For a more thorough discussion on the stability properties of the semigroup $\Phi_{0,n}$ and the limiting measures η_∞, we refer the reader to [28]. Our next objective is to transfer these stability properties to the one of the sequence η_n. First, using (7.5.18), we readily prove that

$$\lim_{n\to\infty} \frac{1}{n} \log \gamma_n(1) = \log \eta_\infty(G)$$

Next, we simplify the notation and we set $\alpha_n := \alpha_{0,n}\,(\gamma_0(1), \eta_0)$ and $c_n := c_{0,n}$. Using (7.5.11), we find that for any $n > 1$

$$a^{-1}\ \|\eta_n - \eta_\infty\|_{\rm tv} \leq \alpha_n\ e^{-\lambda n} + (1-\alpha_n) \sum_{0\leq p<n} \frac{c_p}{\sum_{0\leq q<n} c_q}\ e^{-\lambda p}$$

Recalling that

$$\mu(1)\ g_-^p \leq c_p = \mu Q^p(1) \leq \mu(1)\ g_+^p$$

we also obtain that

$$\begin{aligned}\sum_{0\leq p<n} \frac{c_p}{\sum_{1\leq q<n} c_q}\ e^{-\lambda p} &\leq \frac{1}{\left[\sum_{0\leq q<n} c_q\right]^{1/r}}\ \left[\sum_{0\leq p<n} c_p e^{-\lambda p r}\right]^{1/r} \\ &\leq \frac{1}{\left[\sum_{0\leq q<n} g_-^q\right]^{1/r}}\ \left[\sum_{0\leq p<n} (e^{-\lambda r} g_+)^p\right]^{1/r} \qquad (7.5.19)\end{aligned}$$

for any $r \geq 1$. We conclude that

$$r > \frac{1}{\lambda}\ \log g_+ \Longrightarrow \sum_{0\leq p<n} \frac{c_p}{\sum_{0\leq q<n} c_q}\ e^{-\lambda p} \leq g_-^{-(n-1)/r}/(1-e^{-\lambda r} g_+)^{1/r}$$

and therefore

$$a^{-1}\ \|\eta_n - \eta_\infty\|_{\mathrm{tv}} \leq e^{-\lambda n} + g_-^{-(n-1)/r}\ /(1-e^{-\lambda r} g_+)^{1/r} \to_{n\to\infty} 0$$

7.5.3 Stability and Lipschitz regularity properties

We describe in this section a framework that allows to transfer the regularity properties of the Feynman-Kac semigroups $\Phi_{p,n}$ introduced in (7.4.2) to the ones of the semigroup $\Gamma_{p,n}$ of the sequence $(\gamma_n(1), \eta_n)$. Before proceeding we recall a lemma that provides some weak Lipschitz type inequalities for the Feynman-Kac semigroup $\Phi_{p,n}$ in terms of the Dobrushin contraction coefficient associated with the Markov transitions $P_{p,n}$ introduced in (7.4.2). The details of the proof of this result can be found in [28] or in [31] (see Lemma 4.4. in [31], or proposition 4.3.7 on page 146 in [28]).

Lemma 7.5.4 ([31]). *For any $0 \leq p \leq n$, any $\eta, \mu \in \mathcal{P}(E_p)$ and any $f \in Osc_1(E_n)$, we have*

$$|[\Phi_{p,n}(\mu) - \Phi_{p,n}(\eta)](f)| \leq 2\ q_{p,n}^2\ \beta(P_{p,n})\ |(\mu-\eta)\mathcal{D}_{p,n,\eta}(f)| \qquad (7.5.20)$$

for a collection of functions $\mathcal{D}_{p,n,\eta}(f) \in Osc_1(E_p)$ whose values only depend on the parameters (p, n, η).

Proposition 7.5.5. *For any $0 \leq p \leq n$, any $\eta, \eta' \in \mathcal{P}(E_p)$ and any $f \in Osc_1(E_n)$, there exits a collection of functions $\mathcal{D}_{p,n,\eta'}(f) \in Osc_1(E_p)$ whose values only depend on the parameters (p, n, η) and such that, for any $m \in I_p$, we have*

$$\begin{aligned}&\left|\left[\Gamma_{p,n}^2(m,\eta) - \Gamma_{p,n}^2(m,\eta')\right](f)\right| \\ &\leq 2\ \alpha_{p,n}^\star\ q_{p,n}\left[q_{p,n}\ \beta(P_{p,n})\ |(\eta-\eta')\mathcal{D}_{p,n,\eta'}(f)| + \beta_{p,n}\,|(\eta-\eta')h_{p,n,\eta'}|\right]\end{aligned} \qquad (7.5.21)$$

with the collection of functions $h_{p,n,\eta'} = \frac{1}{2q_{p,n}} \frac{Q_{p,n}(1)}{\eta' Q_{p,n}(1)} \in Osc_1(E_p)$ *and the sequence of parameters* $\epsilon_{p,n}$ *and* $\beta_{p,n}$ *defined below*

$$\alpha^\star_{p,n} := \alpha^\star_{p,n}(m_+(p)) \quad \textit{and} \quad \beta_{p,n} := \sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}} \, \beta(P_{q,n}) \tag{7.5.22}$$

Before getting into the details of the proof of proposition 7.5.5, we illustrate some consequences of these weak functional inequalities for time-homogeneous models $(E_n, G_n, M_n, \mu_n) = (E, G, M, \mu)$ in the three scenarios discussed in (7.5.13), (7.5.14), and (7.5.16).

1. When $G(x) = 1$ for any $x \in E$, we have

   $$\Phi_{p,n}(\eta) = \eta M^{(n-p)}, \quad h_{p,n,\eta'} = 1/2 \quad c_{p,n} = \mu(1) \quad q_{p,n} = 1 \quad \alpha^\star_{p,n} \leq 1$$

   Let us assume that there exist $a < \infty$ and $0 < \lambda < \infty$ such that $\beta(M^n) \leq a e^{-\lambda n}$ for any $n \geq 0$. In this situation, we prove using (7.5.21) that

   $$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right| \leq 2a e^{-\lambda(n-p)} \; |(\mu-\eta)\mathcal{D}_{p,n,\eta'}(f)|$$

2. When $g_+ < 1$ and when the mixing condition $(M)_k$ stated in (7.5.1) is satisfied for some k and some fixed parameters $\epsilon_p = \epsilon$, we have seen in (7.5.15) that

   $$\sup_{m\in I_p} \alpha^\star_{p,n}(m) \leq 1 \wedge \left(d \; g_+^{(n-p)}\right) \quad \text{with} \quad d = \left((\gamma_0(1)/\mu(1)) \vee (1-g_+)^{-1}\right) \delta_{0,k} \epsilon^{-1}$$

   Furthermore, using the estimates given in (7.5.3) and (7.5.4), we also have that

   $$q_{p,n} \leq \delta_k/\epsilon \qquad \beta_{p,n} \leq 1 \quad \text{and} \quad \beta(P_{p,n}) \leq a \, e^{-\lambda \; (n-p)} \quad \text{with } (a,\lambda) \text{ given in (7.5.17)}$$

   In this situation, we prove using (7.5.21) that

   $$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right|$$

   $$\leq 2 \; \left[1 \wedge \left(d \; g_+^{(n-p)}\right)\right] \; (\delta_k/\epsilon) \left[(\delta_k/\epsilon) \; a \; e^{-\lambda(n-p)} \; |(\mu-\eta)\mathcal{D}_{p,n,\eta'}(f)| + |(\mu-\eta)h_{p,n,\eta'}|\right]$$

   Notice that for $(n-p) \geq \log(d)/\log(1/g_+)$, this yields

   $$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right|$$

   $$\leq a_0 \; e^{-\lambda_0(n-p)} \; |(\mu-\eta)\mathcal{D}_{p,n,\eta'}(f)| + a_1 \; e^{-\lambda_1(n-p)} \, |(\mu-\eta)h_{p,n,\eta'}|$$

   with

   $$a_0 = 2ad(\delta_k/\epsilon)^2 \quad a_1 = 2d(\delta_k/\epsilon) \quad \lambda_0 = \lambda + \log(1/g_+) \quad \text{and} \quad \lambda_1 = \log(1/g_+)$$

3. When $g_- > 1$ and when the mixing condition $(M)_k$ presented in (7.5.1) is met for some k and some fixed parameters $\epsilon_p = \epsilon > 0$, then we use the fact that

$$\alpha^\star_{p,n} \leq 1 \quad q_{p,n} \leq \delta_k/\epsilon \quad \text{and} \quad \beta(P_{p,n}) \leq a\, e^{-\lambda(n-p)} \quad \text{with } (a,\lambda) \text{ given in (7.5.17)}$$

Arguing as in (7.5.19), we prove that for any $r > \frac{1}{\lambda}\, \log g_+$

$$\beta_{p,n} \leq g_-^{-(n-p-1)/r}/(1-e^{-\lambda r} g_+)^{1/r}$$

from which we conclude that

$$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right|$$

$$\leq a_0\; e^{-\lambda_0(n-p)}\; |(\mu-\eta)\mathcal{D}_{p,n,\eta'}(f)| + a_1\; e^{-\lambda_1(n-p)}\, |(\mu-\eta)h_{p,n,\eta'}|$$

with

$$a_0 = 2a(\delta_k/\epsilon)^2 \quad a_1 = 2g_-^r(\delta_k/\epsilon)/(1-e^{-\lambda r}g_+)^{1/r} \quad \lambda_0 = \lambda \quad \text{and} \quad \lambda_1 = \log(g_-)$$

Now, we come to the proof of proposition 7.5.5.

Proof of proposition 7.5.5:

First, we observe that

$$\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m',\eta')$$

$$= \alpha_{p,n}(m,\eta)\;\left[\Phi_{p,n}(\eta) - \textstyle\sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\; \Phi_{q,n}(\overline{\mu}_q)\right]$$

$$-\alpha_{p,n}(m',\eta')\;\left[\Phi_{p,n}(\eta') - \textstyle\sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\; \Phi_{q,n}(\overline{\mu}_q)\right]$$

Using the following decomposition

$$ab - a'b' = a'(b-b') + (a-a')b' + (a-a')(b-b') \tag{7.5.23}$$

which is valid for any $a, a', b, b' \in \mathbb{R}$, we prove that

$$\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m',\eta')$$

$$= \alpha_{p,n}(m',\eta')\,[\Phi_{p,n}(\eta) - \Phi_{p,n}(\eta')]$$

$$+\left[\Phi_{p,n}(\eta') - \textstyle\sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\; \Phi_{q,n}(\overline{\mu}_q)\right][\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m',\eta')]$$

$$+[\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m',\eta')]\,[\Phi_{p,n}(\eta) - \Phi_{p,n}(\eta')] \tag{7.5.24}$$

For $m = m'$, using (7.5.24) we find that

$$\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')$$

$$= \alpha_{p,n}(m,\eta)\ \left[\Phi_{p,n}(\eta) - \Phi_{p,n}(\eta')\right]$$

$$+ \left[\Phi_{p,n}(\eta') - \textstyle\sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\ \Phi_{q,n}(\overline{\mu}_q)\right] \left[\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m,\eta')\right]$$

We also notice that

$$\alpha_{p,n}(m,\eta) = \frac{1}{1+\mu_{p,n}/[m\eta Q_{p,n}(1)]}$$

from which we easily prove that

$$\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m',\eta')$$

$$= \tfrac{\mu_{p,n}}{\mu_{p,n}+m\eta Q_{p,n}(1)} \tfrac{1}{\mu_{p,n}+m'\eta' Q_{p,n}(1)} \left[m\eta Q_{p,n}(1) - m'\eta' Q_{p,n}(1)\right]$$

and therefore

$$\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m,\eta') = \left(\alpha_{p,n}(m,\eta')\left(1-\alpha_{p,n}(m,\eta)\right)\right)\ \left[\eta-\eta'\right]\left(\frac{Q_{p,n}(1)}{\eta' Q_{p,n}(1)}\right)$$

The proof of $\alpha_{p,n}(m,\eta) \leq \alpha^\star_{p,n}(m)$ is elementary. From the above decomposition, we prove the following upper bounds

$$\left|\alpha_{p,n}(m,\eta) - \alpha_{p,n}(m,\eta')\right| \leq \alpha^\star_{p,n}(m)\ \left|\left[\eta-\eta'\right]\left(\frac{Q_{p,n}(1)}{\eta' Q_{p,n}(1)}\right)\right|$$

and

$$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right|$$

$$\leq \alpha^\star_{p,n}(m)\left[\left|\left[\Phi_{p,n}(\eta) - \Phi_{p,n}(\eta')\right](f)\right|\right.$$

$$\left. + \left|\left[\eta-\eta'\right]\left(\tfrac{Q_{p,n}(1)}{\eta' Q_{p,n}(1)}\right)\right|\ \left|\textstyle\sum_{p<q\leq n} \frac{c_{q,n}}{\sum_{p<r\leq n} c_{r,n}}\ \left[\Phi_{q,n}(\overline{\mu}_q) - \Phi_{q,n}\left(\Phi_{p,q}(\eta')\right)\right](f)\right|\right]$$

This yields

$$\left|\left[\Gamma^2_{p,n}(m,\eta) - \Gamma^2_{p,n}(m,\eta')\right](f)\right|$$

$$\leq \alpha^\star_{p,n}(m)\left[\left|\left[\Phi_{p,n}(\eta) - \Phi_{p,n}(\eta')\right](f)\right| + \beta_{p,n}\ \left|\left[\eta-\eta'\right]\left(\tfrac{Q_{p,n}(1)}{\eta' Q_{p,n}(1)}\right)\right|\right]$$

The last formula comes from the fact that

$$\beta(P_{q,n}) := \sup_{\nu,\nu'\in\mathcal{P}(E_q)} \left\|\Phi_{q,n}(\nu) - \Phi_{q,n}(\nu')\right\|_{\mathrm{tv}}$$

The proof of this result can be found in [28] (proposition 4.3.1 on page 134). The end of the proof is now a direct consequence of lemma 7.5.4. This ends the proof of the proposition. ∎

7.6 Mean field particle approximations

7.6.1 McKean particle interpretations

In proposition 7.5.1, the evolution equation (7.5.5) of the sequence of probability measures $\eta_n \rightsquigarrow \eta_{n+1}$ is a combination of an updating type transition $\eta_n \rightsquigarrow \Psi_{G_n}(\eta_n)$ and an integral transformation w.r.t. a Markov transition $M_{n+1,(\gamma_n(1),\eta_n)}$ that depends on the current total mass $\gamma_n(1)$ and the current probability distribution η_n. The operator $M_{n+1,(\gamma_n(1),\eta_n)}$ defined in (7.5.6) is a mixture of the Markov transition M_{n+1} and the spontaneous birth normalized measure $\overline{\mu}_{n+1}$. We let S_{n,η_n} be any Markov transition from E_n into itself satisfying

$$\Psi_{G_n}(\eta_n) = \eta_n S_{n,\eta_n}$$

The choice of these transitions is not unique. We can choose for instance one of the collection of transitions presented in (7.3.10), (7.3.12) and (7.3.12). Further examples of McKean acceptance-rejection type transitions can also be found in section 2.5.3 in [28]. By construction, we have the recursive formula

$$\eta_{n+1} = \eta_n K_{n+1,(\gamma_n(1),\eta_n)} \quad \text{with} \quad K_{n+1,(\gamma_n(1),\eta_n)} = S_{n,\eta_n} M_{n+1,(\gamma_n(1),\eta_n)} \tag{7.6.1}$$

with the auxiliary total mass evolution equation

$$\gamma_{n+1}(1) = \gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1) \tag{7.6.2}$$

As already mentioned in section 7.3, the sequence of probability distributions η_n can be interpreted as the distributions of the states $\overline{X}_n$ of a nonlinear Markov chain defined by the elementary transitions

$$\mathbb{P}\left(\overline{X}_{n+1} \in dx \mid \overline{X}_n\right) = K_{n,(\gamma_n(1),\eta_n)}\left(\overline{X}_n, dx\right) \quad \text{with} \quad \eta_n = \text{Law}(\overline{X}_n)$$

Next, we define the mean field particle interpretations of the sequence $(\gamma_n(1), \eta_n)$ given in (7.6.1) and (7.6.2). First, mimicking formula (7.6.2) we set

$$\gamma^N_{n+1}(1) := \gamma^N_n(1)\ \eta^N_n(G_n) + \mu_{n+1}(1) \quad \text{and} \quad \gamma^N_n(f) = \gamma^N_n(1)\ \times\ \eta^N_n(f)$$

for any $f \in \mathcal{B}(E_n)$, with the initial measure $\gamma^N_0 = \gamma_0$. It is important to notice that

$$\gamma^N_n(1) = \gamma_0(1) \prod_{0\leq q<n} \eta^N_q(G_q) + \sum_{p=1}^{n} \mu_p(1) \prod_{p\leq q<n} \eta^N_q(G_q) \Longrightarrow \gamma^N_n(1) \in I_n$$

The mean field particle interpretation of the nonlinear measure valued model (7.6.1) is an E^N_n-valued Markov chain ξ_n with elementary transitions defined in (7.3.9) and (7.6.1). By construction, the particle evolution is a simple combination of a selection and a mutation genetic type transition

$$\xi_n \rightsquigarrow \widehat{\xi}_n = (\widehat{\xi}^i_n)_{1\leq i\leq N} \rightsquigarrow \xi_{n+1}$$

During the selection transitions $\xi_n \rightsquigarrow \widehat{\xi}_n$, each particle $\xi_n^i \rightsquigarrow \widehat{\xi}_n^i$ evolves according to the selection type transition $S_{n,\eta_n^N}(\xi_n^i, dx)$. During the mutation stage, each of the selected particles $\widehat{\xi}_n^i \rightsquigarrow \xi_{n+1}^i$ evolves according to the Markov transition

$$M_{n+1,(\gamma_n^N(1),\eta_n^N)}(x,dy) := \alpha_n\left(\gamma_n^N(1),\eta_n^N\right) M_{n+1}(x,dy) + \left(1-\alpha_n\left(\gamma_n^N(1),\eta_n^N\right)\right)\ \overline{\mu}_{n+1}(dy)$$

7.6.2 Asymptotic behavior

This section is mainly concerned with the proof of theorem 7.4.2. We discuss the unbiasedness property of the particle measures γ_n^N and their convergence properties towards γ_n, as the number of particles N tends to infinity. We mention that the proof of the non asymptotic variance estimates (7.4.11) is simpler than the one provided in a recent article by the second author with F. Cerou and A. Guyader [16]. Section 7.6.3 is concerned with the convergence and the fluctuations of the occupation measures η_n^N around their limiting measures η_n.

Intensity measures

We start this section with a simple unbiasedness property.

Proposition 7.6.1. *For any $0 \leq p \leq n$, and any $f \in \mathcal{B}(E_n)$, we have*

$$\mathbb{E}\left(\gamma_{n+1}^N(f) \,\middle|\, \mathcal{F}_p^{(N)}\right) = \gamma_p^N Q_{p,n+1}(f) + \sum_{p<q\leq n+1} \mu_q Q_{q,n+1}(f) \tag{7.6.3}$$

In particular, we have the unbiasedness property: $\mathbb{E}\left(\gamma_n^N(f)\right) = \gamma_n(f)$.

Proof:
By construction of the particle model, for any $f \in \mathcal{B}(E_n)$ we have

$$\mathbb{E}\left(\eta_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) = \eta_n^N K_{n+1,(\gamma_n^N(1),\eta_n^N)}(f) = \Gamma_{n+1}^2\left(\gamma_n^N(1),\eta_n^N\right)(f)$$

with the second component Γ_{n+1}^2 of the transformation Γ_{n+1} introduced in 7.5.7. Using the fact that

$$\Gamma_{n+1}^2\left(\gamma_n^N(1),\eta_n^N\right)(f) = \frac{\gamma_n^N(1)\ \eta_n^N(Q_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n^N(1)\ \eta_n^N(Q_{n+1}(1)) + \mu_{n+1}(1)} = \frac{\gamma_n^N(Q_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n^N(Q_{n+1}(1)) + \mu_{n+1}(1)}$$

and

$$\gamma_{n+1}^N(1) = \gamma_n^N(1)\ \eta_n^N(G_n) + \mu_{n+1}(1) = \gamma_n^N(Q_{n+1}(1)) + \mu_{n+1}(1)$$

we prove that

$$\begin{aligned}\mathbb{E}\left(\gamma_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) &= \mathbb{E}\left(\gamma_{n+1}^N(1)\ \eta_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) = \gamma_{n+1}^N(1)\ \mathbb{E}\left(\eta_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) \\ &= \gamma_n^N(Q_{n+1}(f)) + \mu_{n+1}(f)\end{aligned}$$

This also implies that

$$\begin{aligned}\mathbb{E}\left(\gamma_{n+1}^N(f)\,\middle|\,\mathcal{F}_{n-1}^{(N)}\right) &= \mathbb{E}\left(\gamma_n^N(Q_{n+1}(f))\,\middle|\,\mathcal{F}_{n-1}^{(N)}\right)+\mu_{n+1}(f)\\ &= \gamma_{n-1}^N(Q_nQ_{n+1}(f))+\mu_n(Q_{n+1}(f))+\mu_{n+1}(f)\end{aligned}$$

Iterating the argument one proves (7.6.3). The end of the proof is now clear. ∎

The next theorem provides a key martingale decomposition and a rather crude non asymptotic variance estimate.

Theorem 7.6.2. *For any $n \geq 0$ and any function $f \in \mathcal{B}(E_n)$, we have the decomposition*

$$\sqrt{N}\ \left[\gamma_n^N-\gamma_n\right](f)=\sum_{p=0}^{n}\gamma_p^N(1)\ W_p^N(Q_{p,n}(f)) \tag{7.6.4}$$

In addition, if the mixing condition $(M)_k$ presented in (7.5.1) is met for some $k \geq 1$ and some constant parameters $\epsilon_p = \epsilon > 0$, then we have for any $N > 1$ and any $n \geq 1$

$$\mathbb{E}\left(\left[\frac{\gamma_n^N(1)}{\gamma_n(1)}-1\right]^2\right)\leq\frac{n+1}{N-1}\ \frac{\delta_k^2}{\epsilon^2}\left(1+\frac{\delta_k^2}{\epsilon^2(N-1)}\right)^{n-1} \tag{7.6.5}$$

Before presenting the proof of this theorem, we would like to make a couple of comments. On the one hand, we observe that the unbiasedness property follows directly from the decomposition (7.6.4). On the other hand, using Kintchine's inequality, for any $r \geq 1$, $p \geq 1$, and any $f \in \mathrm{Osc}_1(E_n)$ we have the almost sure estimates

$$\sqrt{N}\ \mathbb{E}\left(\left|W_p^N(f)\right|^r\ \middle|\mathcal{F}_{p-1}^{(N)}\right)^{\frac{1}{r}}\leq a_r$$

A detailed proof of these estimates can be found in [28], see also lemma 7.2 in [7] for a simpler proof by induction on the parameter N. From this elementary observation, and recalling that $\gamma_n^N(1) \in I_n$ for any $n \geq 0$, we find that

$$\sqrt{N}\ \mathbb{E}\left(\left|\left[\gamma_n^N-\gamma_n\right](f)\right|^r\right)^{\frac{1}{r}}\leq a_r\ b_n$$

for some finite constant b_n whose values only depend on the time parameter n.
Now, we present the proof of theorem 7.6.2.

Proof of theorem 7.6.2:
We use the decomposition:

$$\gamma_{n+1}^N(f)-\gamma_{n+1}(f)=\left[\gamma_{n+1}^N(f)-\mathbb{E}\left(\gamma_{n+1}^N(f)\,\middle|\,\mathcal{F}_n^{(N)}\right)\right]+\left[\mathbb{E}\left(\gamma_{n+1}^N(f)\,\middle|\,\mathcal{F}_n^{(N)}\right)-\gamma_{n+1}(f)\right]$$

By (7.6.3), we find that

$$\gamma_{n+1}^N(f)-\mathbb{E}\left(\gamma_{n+1}^N(f)\,\middle|\,\mathcal{F}_n^{(N)}\right)=\gamma_{n+1}^N(f)-\left[\gamma_n^N(Q_{n+1}(f))+\mu_{n+1}(f)\right]$$

Since we have

$$\begin{aligned}\gamma_n^N(Q_{n+1}(1)) + \mu_{n+1}(1) &= \gamma_n^N(G_n) + \mu_{n+1}(1) \\ &= \gamma_n^N(1)\ \eta_n^N(G_n) + \mu_{n+1}(1) = \gamma_{n+1}^N(1)\end{aligned}$$

this implies that

$$\begin{aligned}\gamma_{n+1}^N(f) - \left[\gamma_n^N(Q_{n+1}(f)) + \mu_{n+1}(f)\right] &= \gamma_{n+1}^N(1)\ \left[\eta_{n+1}^N(f) - \frac{\left[\gamma_n^N(Q_{n+1}(f)) + \mu_{n+1}(f)\right]}{\left[\gamma_n^N(Q_{n+1}(1)) + \mu_{n+1}(1)\right]}\right] \\ &= \gamma_{n+1}^N(1)\ \left[\eta_{n+1}^N(f) - \eta_n^N K_{n+1,(\gamma_n^N(1),\eta_n^N)}(f)\right]\end{aligned}$$

and therefore

$$\gamma_{n+1}^N(f) - \mathbb{E}\left(\gamma_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) = \gamma_{n+1}^N(1)\ \left[\eta_{n+1}^N(f) - \eta_n^N K_{n+1,(\gamma_n^N(1),\eta_n^N)}(f)\right]$$

Finally, we observe that

$$\mathbb{E}\left(\gamma_{n+1}^N(f) \,\middle|\, \mathcal{F}_n^{(N)}\right) - \gamma_{n+1}(f) = \gamma_n^N(Q_{n+1}(f)) - \gamma_n(Q_{n+1}(f))$$

from which we find the recursive formula

$$\left[\gamma_{n+1}^N - \gamma_{n+1}\right](f) = \gamma_{n+1}^N(1)\ \left[\eta_{n+1}^N - \eta_n^N K_{n+1,(\gamma_n^N(1),\eta_n^N)}\right](f) + \left[\gamma_n^N - \gamma_n\right](Q_{n+1}(f))$$

The end of the proof of (7.6.4) is now obtained by a simple induction on the parameter n.

Now, we come to the proof of (7.6.5). Using the fact that

$$\begin{aligned}\mathbb{E}\left(\gamma_p^N(1)W_p^N(f^{(1)})\ \gamma_q^N(1)W_q^N(f^{(2)})\right) &= \mathbb{E}\left(\gamma_p^N(1)\gamma_q^N(1)W_p^N(f^{(1)})\ \mathbb{E}\left(W_q^N(f^{(2)}) \mid \mathcal{F}_{q-1}^N\right)\right) \\ &= 0\end{aligned}$$

for any $0 \leq p < q \leq n$, and any $f^{(1)} \in \mathcal{B}(E_p)$, and $f^{(2)} \in \mathcal{B}(E_q)$, we prove that

$$N\ \mathbb{E}\left(\left[\gamma_n^N(1) - \gamma_n(1)\right]^2\right) = \sum_{p=0}^{n} \mathbb{E}\left(\gamma_p^N(1)^2\ \mathbb{E}\left(W_p^N(Q_{p,n}(1))^2 | \mathcal{F}_{p-1}^N\right)\right)$$

Notice that

$$\frac{1}{\gamma_n(1)^2} = \frac{1}{\gamma_p(1)^2}\ \frac{1}{\eta_p(Q_{p,n}(1))^2}\ \left(\frac{\gamma_p(Q_{p,n}(1))}{\gamma_n(1)}\right)^2 \leq \alpha_{p,n}^\star(\gamma_p(1))^2 \frac{1}{\gamma_p(1)^2}\ \frac{1}{\eta_p(Q_{p,n}(1))^2} \tag{7.6.6}$$

The r.h.s. estimate comes from the fact that

$$\frac{\gamma_p(Q_{p,n}(1))}{\gamma_n(1)} = \frac{\gamma_p(1)\ \eta_p(Q_{p,n}(1))}{\gamma_p(1)\ \eta_p(Q_{p,n}(1)) + \sum_{p<q\leq n} \mu_q Q_{q,n}(1)} = \alpha_{p,n}\left(\gamma_p(1), \eta_p\right) \leq \alpha_{p,n}^\star(\gamma_p(1))$$

Using the above decompositions, we readily prove that

$$N\ \mathbb{E}\left(\left[\frac{\gamma_n^N(1)}{\gamma_n(1)}-1\right]^2\right)\leq\sum_{p=0}^{n}\alpha^\star_{p,n}(\gamma_p(1))^2\ \mathbb{E}\left(\left(\frac{\gamma_p^N(1)}{\gamma_p(1)}\right)^2\ \mathbb{E}\left(W_p^N(\overline{Q}_{p,n}(1))^2|\mathcal{F}_{p-1}^N\right)\right)$$

with

$$\overline{Q}_{p,n}(1)=\overline{Q}_{p,n}(1)/\eta_p(Q_{p,n}(1))\leq q_{p,n}$$

We set

$$U_n^N:=\mathbb{E}\left(\left[\frac{\gamma_n^N(1)}{\gamma_n(1)}-1\right]^2\right)\quad\text{then we find that}\quad N\ U_n^N\leq a_n+\sum_{p=0}^{n}b_{p,n}\ U_p^N$$

with the parameters

$$a_n:=\sum_{p=0}^{n}\left(q_{p,n}\alpha^\star_{p,n}(\gamma_p(1)\right)^2\quad\text{and}\quad b_{p,n}:=\left(q_{p,n}\alpha^\star_{p,n}(\gamma_p(1)\right)^2$$

Using the fact that $b_{n,n}\leq 1$, we prove the following recursive equation

$$U_n^N\leq a_n^N+\sum_{0\leq p<n}b_{p,n}^N\ U_p^N\quad\text{with}\quad a_n^N:=\frac{a_n}{N-1}\quad\text{and}\quad b_{p,n}^N:=\frac{b_{p,n}}{N-1}$$

Using an elementary proof by induction on the time horizon n, we prove the following inequality:

$$U_n^N\leq\left[\sum_{p=1}^{n}a_p^N\sum_{e\in\langle p,n\rangle}b^N(e)\right]+\left[\sum_{e\in\langle 0,n\rangle}b^N(e)\right]\ U_0^N$$

In the above display, $\langle p,n\rangle$ stands for the set of all integer valued paths $e=(e(l))_{0\leq l\leq k}$ of a given length k from p to n

$$e_0=p<e_1<\ldots<e_{k-1}<e_k=n\quad\text{and}\quad b^N(e)=\prod_{1\leq l\leq k}b^N_{e(l-1),e(l)}$$

We have also used the convention $b^N(\emptyset)=\prod_\emptyset=1$ and $\langle n,n\rangle=\{\emptyset\}$, for $p=n$. Recalling that $\gamma_0^N=\gamma_0$, we conclude that

$$U_n^N\leq\sum_{p=1}^{n}a_p^N\sum_{e\in\langle p,n\rangle}b^N(e)$$

We further assume that the mixing condition $(M)_k$ presented in (7.5.1) is met for some parameters $k\geq 1$, and some constant parameters $\epsilon_p=\epsilon>0$. In this case, we use the fact that

$$\alpha^\star_{p,n}\leq 1\quad\text{and}\quad q_{p,n}\leq\delta_k/\epsilon$$

to prove that

$$\sup_{0\leq p\leq n} a_p^N \leq (n+1)\ (\delta_k/\epsilon)^2/(N-1) \quad \text{and} \quad \sup_{0\leq p\leq n} b_{p,n}^N \leq (\delta_k/\epsilon)^2/(N-1)$$

Using these rather crude estimates, we find that

$$U_n^N \leq a_n^N + \sum_{0<p<n} a_p^N \sum_{l=1}^{(n-p)} \left(\begin{array}{c} n-p-1 \\ l-1 \end{array} \right) \left(\frac{\delta_k^2}{\epsilon^2(N-1)} \right)^l$$

and therefore

$$\begin{aligned} U_n^N &\leq \frac{(n+1)}{(N-1)}\ \frac{\delta_k^2}{\epsilon^2} \left(1 + \frac{\delta_k^2}{\epsilon^2(N-1)} \sum_{0<p<n} \left(1 + \left(\frac{\delta_k^2}{\epsilon^2(N-1)} \right) \right)^{n-p-1} \right) \\ &= \frac{(n+1)}{(N-1)}\ \frac{\delta_k^2}{\epsilon^2} \left(1 + \frac{\delta_k^2}{\epsilon^2(N-1)} \right)^{n-1} \end{aligned}$$

This ends the proof of the theorem. ∎

7.6.3 Probability distributions

This section is mainly concerned with the proof of the $\mathbb{L}_r$-mean error estimates stated in (7.4.9). We use the decomposition

$$\begin{aligned} \left(\gamma_n^N(1), \eta_n^N\right) - (\gamma_n(1), \eta_n) &= \left[\Gamma_{0,n}\left(\gamma_0^N(1), \eta_0^N\right) - \Gamma_{0,n}\left(\gamma_0(1), \eta_0\right)\right] \\ &\quad + \sum_{p=1}^{n} \left[\Gamma_{p,n}\left(\gamma_p^N(1), \eta_p^N\right) - \Gamma_{p-1,n}\left(\gamma_{p-1}^N(1), \eta_{p-1}^N\right)\right] \quad (7.6.7) \end{aligned}$$

to prove that

$$\eta_n^N - \eta_n$$

$$= \left[\Gamma_{0,n}^2\left(\gamma_0^N(1), \eta_0^N\right) - \Gamma_{0,n}^2\left(\gamma_0(1), \eta_0\right)\right] + \textstyle\sum_{p=1}^n \left[\Gamma_{p,n}^2\left(\gamma_p^N(1), \eta_p^N\right) - \Gamma_{p-1,n}^2\left(\gamma_{p-1}^N(1), \eta_{p-1}^N\right)\right]$$

Using the fact that

$$\Gamma_{p-1,n}(m,\eta) = \Gamma_{p,n}\left(\Gamma_p(m,\eta)\right) \Rightarrow \Gamma_{p-1,n}^2(m,\eta) = \Gamma_{p,n}^2\left(\Gamma_p(m,\eta)\right)$$

we readily check that

$$\begin{aligned} \Gamma_p\left(\gamma_{p-1}^N(1), \eta_{p-1}^N\right) &= \left(\gamma_{p-1}^N(1)\eta_{p-1}^N(G_{p-1}) + \mu_p(1), \Psi_{G_{p-1}}\left(\eta_{p-1}^N\right) M_{p,(\gamma_{p-1}^N(1),\eta_{p-1}^N)}\right) \\ &= \left(\gamma_p^N(1), \eta_{p-1}^N K_{p,(\gamma_{p-1}^N(1),\eta_{p-1}^N)}\right) \end{aligned}$$

Since we have $\gamma_0^N(1) = \mu_0(1) = \gamma_0(1)$, one concludes that

$$\eta_n^N - \eta_n = \left[\Gamma_{0,n}^2\left(\gamma_0(1), \eta_0^N\right) - \Gamma_{0,n}^2\left(\gamma_0(1), \eta_0\right)\right] + \sum_{p=1}^{n}\left[\Gamma_{p,n}^2\left(\gamma_p^N(1), \eta_p^N\right) - \Gamma_{p,n}^2\left(\gamma_p^N(1), \eta_{p-1}^N K_{p,(\gamma_{p-1}^N(1),\eta_{p-1}^N)}\right)\right]$$

Using the fact that $\gamma_p^N(1) \in I_p$, for any $p \geq 0$, the end of the proof is a direct consequence of lemma 7.5.5 and Kintchine inequality. The proof of the uniform convergence estimates stated in the end of theorem 7.4.2 are a more or less direct consequence of the functional inequalities derived at the end of section 7.5.3. The end of the proof of the theorem 7.4.2 is now completed.

We end this section with the fluctuations properties of the N-particle approximation measures γ_n^N and η_n^N around their limiting values. Using the same line of arguments as those we use in the proof of the functional central limit theorem, theorem 3.3 in [34], we can prove that the sequence $(W_n^N)_{n\geq 0}$ defined in (7.4.12) converges in law, as N tends to infinity, to the sequence of n independent, Gaussian and centered random fields $(W_n)_{n\geq 0}$ with a covariance function given in (7.4.13). Using the decompositions (7.6.4) and

$$\eta_n^N(f) - \eta_n(f) = \frac{\gamma_n(1)}{\gamma_n^N(1)}\ \left([\gamma_n^N - \gamma_n]\left(\frac{1}{\gamma_n(1)}(f - \eta_n(f))\right)\right)$$

by the continuous mapping theorem, we deduce the functional central limit theorem 7.4.3.

7.7 Particle approximations of spontaneous birth measures

Assume that the spontaneous birth measures μ_n are chosen so that $\mu_n \ll \lambda_n$ for some reference probability measures λ_n and that the Radon Nikodim derivatives $H_n = d\mu_n/d\lambda_n$ are bounded. For any $n \geq 0$, we let $\lambda_n^{N'} := \frac{1}{N'}\sum_{i=1}^{N'} \delta_{\zeta_n^i}$ be the empirical measure associated with N' independent and identically distributed random variables $(\zeta_n^i)_{1\leq i\leq N}$ with common distribution λ_n. We also denote by $\mu_n^{N'}$ the particle spontaneous birth measures defined below

$$\forall n \geq 0 \qquad \mu_n^{N'}(dx_n) := H_n(x_n)\ \lambda_n^{N'}(dx_n)$$

In this notation, the initial distribution η_0 and the initial mass γ_0 are approximated by the weighted occupation measure $\eta_0^{N'} := \Psi_{H_0}(\lambda_0^{N'})$ and $\gamma_0^{N'}(1) := \lambda_0^{N'}(H_0)$.

We let $\widetilde{\gamma}_n^{N'}$ and $\widetilde{\eta}_n^{N'}$ the random measures defined as γ_n and η_n by replacing in (7.3.2) the measures μ_n by the random measures $\mu_n^{N'}$, for any $n \geq 0$; that is, we have that

$$\widetilde{\gamma}_n^{N'} = \widetilde{\gamma}_{n-1}^{N'} Q_n + \mu_n^{N'} \quad \text{and} \quad \widetilde{\eta}_n^{N'}(f_n) = \widetilde{\gamma}_n^{N'}(f_n)/\widetilde{\gamma}_n^{N'}(1)$$

for any $f_n \in \mathcal{B}(E_n)$. By construction, using the same arguments as the ones we used in the proof of (7.5.8), we have

$$\widetilde{\gamma}_n^{N'} = \sum_{0 \leq p \leq n} \mu_p^{N'} Q_{p,n}$$

This yields for any $f \in \mathcal{B}(E_n)$ the decomposition

$$\left[\widetilde{\gamma}_n^{N'} - \gamma_n\right](f) = \sum_{0 \leq p \leq n} \left[\mu_p^{N'} - \mu_p\right] Q_{p,n}(f) = \sum_{0 \leq p \leq n} \left[\lambda_p^{N'} - \lambda_p\right] (H_p \; Q_{p,n}(f))$$

Several estimates can be derived from these formulae, including $\mathbb{L}_p$-mean error bounds, functional central limit theorems, empirical process convergence, as well as sharp exponential concentration inequalities. For instance, we have the unbiasedness property

$$\mathbb{E}\left(\widetilde{\gamma}_n^{N'}(f)\right) = \gamma_n(f)$$

and the variance estimate

$$N \; \mathbb{E}\left(\left[\widetilde{\gamma}_n^{N'}(f) - \gamma_n(f)\right]^2\right) = \sum_{0 \leq p \leq n} \lambda_p \left[(H_p Q_{p,n}(f) - \lambda_p(H_p Q_{p,n}(f))]^2\right)$$

Using the same arguments as the ones we used in (7.6.6), we prove the following rather crude upper bound

$$\begin{aligned} N \; \mathbb{E}\left(\left[\frac{\widetilde{\gamma}_n^{N'}(f)}{\gamma_n(1)} - \eta_n(f)\right]^2\right) &\leq \sum_{0 \leq p \leq n} \alpha_{p,n}^{\star}(\gamma_p(1))^2 \frac{1}{\gamma_p(1)^2} \; \frac{\mu_p\left(H_p Q_{p,n}(f)^2\right)}{\eta_p(Q_{p,n}(1))^2} \\ &\leq \sum_{0 \leq p \leq n} \alpha_{p,n}^{\star}(\gamma_p(1))^2 \frac{1}{\gamma_p(1)^2} \; \|H_p\| \; \mu_p(1) \; q_{p,n}^2 \end{aligned}$$

We illustrate these variance estimates for time homogeneous models $(E_n, G_n, H_n, M_n, \mu_n) = (E, G, H, M, \mu)$, in the three situations discussed in (7.5.13), (7.5.14), and (7.5.16). We further assume that the mixing condition $(M)_k$ presented in (7.5.1) is met for some parameters $k \geq 1$, and some $\epsilon > 0$. In this case, we use the fact that $q_{p,n} \leq \delta_k/\epsilon$, to prove that

$$N \; \mathbb{E}\left(\left[\frac{\widetilde{\gamma}_n^{N'}(f)}{\gamma_n(1)} - \eta_n(f)\right]^2\right) \leq c \sum_{0 \leq p \leq n} \left[\alpha_{p,n}^{\star}(\gamma_p(1))/\gamma_p(1)\right]^2$$

with some constant $c := (\|H\| \; \mu(1) \; (\delta_k/\epsilon)^2)$.

1. When $G(x) = 1$ for any $x \in E$, we have $\gamma_p(1) = \gamma_0(1) + \mu(1) \; p$. Recalling that $\alpha_{p,n}^{\star}(\gamma_p(1)) \leq 1$, we prove the uniform estimates

$$N \; \sup_{n \geq 0} \mathbb{E}\left(\left[\frac{\widetilde{\gamma}_n^{N'}(f)}{\gamma_n(1)} - \eta_n(f)\right]^2\right) \leq c \sum_{p \geq 0} (\gamma_0(1) + \mu(1) \; p)^{-2}$$

2. When $g_+ < 1$ and when the mixing condition $(M)_k$ stated in (7.5.1) is satisfied, we have seen in (7.5.15) that

$$\alpha^{\star}_{p,n}(\gamma_p(1)) \leq 1 \wedge \left(d_1 \; g_+^{(n-p)} \right) \quad \text{and} \quad \inf_n \gamma_n(1) \geq d_2$$

for some finite constants $d_1 < \infty$ and $d_2 > 0$. From previous calculations, we prove the following uniform variance estimates

$$N \; \sup_{n \geq 0} \mathbb{E} \left(\left[\frac{\widetilde{\gamma}_n^{N'}(f)}{\gamma_n(1)} - \eta_n(f) \right]^2 \right) \leq (c/d_2^2) \sum_{p \geq 0} \left[1 \wedge \left(d_1^2 \; g_+^{2p} \right) \right]$$

3. When $g_- > 1$ we have seen in (7.5.16) that $\gamma_n(1) \geq d \; g_-^n$ for any $n \geq n_0$, for some finite constant $d < \infty$ and some $n_0 \geq 1$ so

$$N \; \sup_{n \geq 0} \mathbb{E} \left(\left[\frac{\widetilde{\gamma}_n^{N'}(f)}{\gamma_n(1)} - \eta_n(f) \right]^2 \right) \leq c \left(\sum_{0 \leq p \leq n_0} \gamma_p(1)^{-2} + d \sum_{n \geq n_0} g_-^{-2n} \right)$$

Chapter 8

On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering

8.1 Chapter overview

In this chapter we analyze the exponential stability properties of a class of measure-valued equations arising in nonlinear multi-object filtering and prove the uniform convergence properties of a general class of stochastic filtering algorithms. We illustrate these results in the context of the Bernoulli and the Probability Hypothesis Density filter.
The results can be resumed as follows: the Bernoulli filter with a sufficiently mixing prediction and almost equal survival and spontaneous births rates is exponentially stable and the PHD filter is exponentially stable for small clutter intensities and sufficiently high detection probability and spontaneous birth rates. In both situations, the estimation errors of any N-approximation model satisfying certain conditions do not accumulate over time.

In the second part we propose three different classes of stochastic particle approximation models. The first is a mean field particle interpretation of the flow of normalized intensity measures associated to a general filtering model, the second is an interacting particle association model while the third is a combination of the previous two approaches. The second approximation strategy was the object of sec-

tion 5.4.

This result has been accepted and is due to appear as a journal article in *Journal of Stochastic Analysis and Applications* (2011).

On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering

*François Caron, †Pierre Del Moral, ‡Michele Pace, §Ba-Ngu Vo

Abstract

We analyse the exponential stability properties of a class of measure-valued equations arising in nonlinear multi-target filtering problems. We also prove the uniform convergence properties of a rather general class of stochastic filtering algorithms, including sequential Monte Carlo type models and mean field particle interpretation models. We illustrate these results in the context of the Bernoulli and the Probability Hypothesis Density filter, yielding what seems to be the first results of this kind in this subject.

Keywords: Measure-valued equations, nonlinear multi-target filtering, Bernoulli filter, Probability hypothesis density filter, interacting particle systems, particle filters, sequential Monte Carlo methods, exponential concentration inequalities, semigroup stability, functional contraction inequalities.

Mathematics Subject Classification: Primary: 93E11, 65C35, 37L15; Secondary: 60J85, 65C05, 65C35.

8.2 Introduction

Let $(E_n)_{n\geq 0}$ be a sequence of measurable spaces equipped with the σ-fields $(\mathcal{E}_n)_{n\geq 0}$, and for each with $n \geq 0$, denote $\mathcal{M}(E_n)$, $\mathcal{M}_+(E_n)$ and $\mathcal{P}(E_n)$ the set of all finite signed measures, the subset of positive measures and the subset of probability measures, respectively, over the space E_n. The aim of this work is to present a stochastic interacting particle interpretation for numerical solutions of the general measure-valued dynamical systems $\gamma_n \in \mathcal{M}_+(E_n)$ defined by the following non-linear equation

$$\gamma_n(dx_n) = \left(\gamma_{n-1}Q_{n,\gamma_{n-1}}\right)(dx_n) := \int_{E_{n-1}} \gamma_{n-1}(dx_{n-1})Q_{n,\gamma_{n-1}}(x_{n-1}, dx_n) \qquad (8.2.1)$$

*Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux. Francois.Caron@inria.fr

†Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux. Pierre.Del-Moral@inria.fr

‡Centre INRIA Bordeaux et Sud-Ouest & Institut de Mathmatiques de Bordeaux. Michele.Pace@inria.fr

§School of Electrical Electronic & Computer Engineering University of Western Australia, bangu.vo@uwa.edu.au. The work of B.-N. Vo is supported by Australian Research Council under the Future Fellowship FT0991854

with initial measure $\gamma_0 \in \mathcal{M}_+(E_0)$, and positive and bounded integral operators $Q_{n,\gamma}$ from E_{n-1} into E_n, indexed by the time parameter $n \geq 1$ and the set of measures $\gamma \in \mathcal{M}_+(E_n)$.

This class of measure-valued equations arises in a natural way in the analysis of the first moments evolution of nonlinear branching processes, as well as in signal processing and more particularly in multiple targets tracking models. A pair of filtering models is discussed in some details in section 8.2.1 and in section 8.2.1. In the context of multiple targets tracking problems these measure-valued equations represents the first-order statistical moments of the conditional distributions of the target occupation measures given observation random measures obscured by clutter, detection uncertainty and data association uncertainty.

As in most of the filtering problems encountered in practice, the initial distribution of the targets is usually unknown. It is therefore essential to check wether or not the filtering equation "forgets" any erroneous initial distribution. For a thorough discussion on the stability properties of traditional nonlinear filtering problems with a detailed overview of theoretical developments on this subject, we refer to the book [28] and to the more recent article by M. L. Kleptsyna and A. Y. Veretennikov [60]. Besides the fact that significant progress has been made in the recent years in the rigorous derivation of multiple target tracking nonlinear equations (see for instance Chapter 6 and [13, 76, 135, 120]), up to our knowledge the stability and the robustness properties of these measure-valued models have never been addressed so far in the literature on the subject. One aim of this paper is to study one such important property: the exponential stability properties of multiple target filtering models. We present an original and general perturbation type technique combining the continuity property and the stability analysis of nonlinear semigroups of the form (8.2.1). A more thorough presentation of these results is provided in section 8.2.2 dedicated to the statement of the main results of the present article. The detailed presentation of this perturbation technique can be found in section 8.4.

On the other hand, while the integral equation (8.2.1) appears to be simple at first glance, numerical solutions are computationally intensive, often requiring integrations in high dimensional spaces. One natural way to solve the non-linear integral equation (8.2.1) is to find a judicious probabilistic interpretation of the normalized distributions flow given below

$$\eta_n(dx_n) := \gamma_n(dx_n)/\gamma_n(1)$$

To describe with some conciseness these stochastic models, it is important to observe that the pair process $(\gamma_n(1), \eta_n) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))$ satisfies an evolution equation of the following form

$$(\gamma_n(1), \eta_n) = \Gamma_n(\gamma_{n-1}(1), \eta_{n-1}) \tag{8.2.2}$$

Let the mappings $\Gamma_n^1 : \mathbb{R}_+ \times \mathcal{P}(E_n) \to \mathbb{R}_+$ and $\Gamma_n^2 : \mathbb{R}_+ \times \mathcal{P}(E_n) \to \mathcal{P}(E_n)$, denote the first and the second components of Γ_n respectively. By construction, we notice

that the total mass process can be computed using the recursive formula

$$\gamma_{n+1}(1) = \gamma_n(G_{n,\gamma_n}) = \eta_n(G_{n,\gamma_n})\ \gamma_n(1) \quad \text{with} \quad G_{n,\gamma_n} := Q_{n+1,\gamma_n}(1) \tag{8.2.3}$$

Suppose that we are given an approximation $\left(\gamma_n^N(1), \eta_n^N\right)$ of the pair $(\gamma_n(1), \eta_n)$ at some time horizon n, where N stands for some precision parameter; that is $\left(\gamma_n^N(1), \eta_n^N\right)$ converges (in some sense) to $(\gamma_n(1), \eta_n)$, as $N \to \infty$. Then, the N-approximation of the measure γ_n is given by $\gamma_n^N = \gamma_n^N(1) \times \eta_n^N$. The central idea behind any approximation model is to ensure that the total mass process at time $(n+1)$ defined by

$$\gamma_{n+1}^N(1) = \eta_n^N(G_{n,\gamma_n^N})\ \gamma_n^N(1) \tag{8.2.4}$$

can be "easily" computed in terms of the N-approximation measures γ_n^N. Assuming that the initial mass $\gamma_0(1) = \gamma_0^N(1)$ is known, the next step is to find some strategy to approximate the quantities $\Gamma_{n+1}^2(\gamma_n^N(1), \eta_n^N)$ by some N-approximation measures η_{n+1}^N, and to set $\gamma_{n+1}^N = \gamma_{n+1}^N(1) \times \eta_{n+1}^N$.

The local fluctuations of η_n^N around the measures $\Gamma_n^2(\gamma_{n-1}^N(1), \eta_{n-1}^N)$ is defined in terms of a collection of random fields W_n^N :

$$W_n^N := \sqrt{N}\ \left[\eta_n^N - \Gamma_n^2(\gamma_{n-1}^N(1), \eta_{n-1}^N)\right] \Longleftrightarrow \eta_n^N = \Gamma_n^2\left(\gamma_{n-1}^N(1), \eta_{n-1}^N\right) + \frac{1}{\sqrt{N}}\ W_n^N \tag{8.2.5}$$

which satisfies for any $r \geq 1$ and any test function f with uniform norm $\|f\| \leq 1$,

$$\mathbb{E}\left(W_n^N(f) \mid \mathcal{F}_{n-1}^N\right) = 0 \quad \text{and} \quad \mathbb{E}\left(\left|W_n^N(f)\right|^r \mid \mathcal{F}_{n-1}^N\right)^{\frac{1}{r}} \leq a_r \tag{8.2.6}$$

where $\mathcal{F}_{n-1}^N = \sigma\left(\eta_p^N, 0 \leq p < n\right)$ is the σ-field generated by the random measures η_p^N, $0 \leq p < n$, while b and a_r are universal constants whose values do not depend on the precision parameter N. The stochastic analysis of the resulting particle approximation model relies on the analysis of the propagation of the local sampling errors defined in (8.2.5). The main objective is to control, at any time horizon n, the fluctuations of the random measures (γ_n^N, η_n^N) around their limiting values (γ_n, η_n) defined by the following random fields:

$$V_n^{\gamma,N} := \sqrt{N}\ \left[\gamma_n^N - \gamma_n\right] \quad \text{with} \quad V_n^{\eta,N} := \sqrt{N}\ \left[\eta_n^N - \eta_n\right]. \tag{8.2.7}$$

The construction of the N-approximation measures η_n^N is far from being unique. In the present article, we devise three different classes of stochastic particle approximation models. These stochastic algorithms are discussed in section 8.5. The first one is a mean field particle interpretation of the flow of probability measures η_n, and it is presented in section 8.5.1. The second model is an interacting particle association model while the third one is a combination of these two approximation algorithms. These pair of approximation models are respectively discussed in

section 8.5.2 and in section 8.5.3. In the context of multi-target tracking models, the first two approximation models are closely related to the sequential Monte Carlo technique presented in the series of articles [117, 133, 96, 94, 95, 144], and respectively, the Gaussian mixture Probability Hypothesis Density filter discussed in the article by B.-N. Vo, and W.-K. Ma [129, 130], and the the Rao-Blackwellized Particle multi-target filters presented by S. Sarkka, A. Vehtari, and J. Lampinen in [114, 112]. These modern stochastic algorithms are rather simple to implement and computationally tractable, and they exhibit excellent performance.

Nevertheless, despite advances in recent years [53, 133], these Monte Carlo particle type multi-target filters remain poorly understood theoretically. One aim of this article is to present a novel class of stochastic algorithms with a refined analysis including uniform convergence results w.r.t. the time parameter. We also illustrate these results in the context of multi-target tracking models, yielding what seems to be the first uniform results of this type in this subject.

The rest of the article is organized as follows: In section 8.2.1 we illustrate the abstract measure-valued equations (8.2.1) with connection to two recent multi-target filters, namely the Bernoulli filter and the Probability Hypothesis Density filter (*abbreviate PHD filter*). Section 8.2.2 is devoted to the statement of our main results. In section 8.3, we describe the semigroups and the continuity properties of the nonlinear equation 8.2.1. We show that this semigroup analysis can be applied to analyse the convergence of the Bernoulli and the PHD approximation filters. Section 8.4 is devoted to the stability properties of nonlinear measure-valued processes of the form (8.2.2). We present a perturbation technique and a series of functional contraction inequalities. In the next three sections, we illustrate these results in the context of Feynman-Kac models, as well as Bernoulli and PHD models. Section 8.5 is concerned with the detailed presentation and the convergence analysis of three different classes of particle type approximation models, including mean field type particle approximations and particle association stochastic algorithms. Finally, the appendix of the article contains most of technical proofs in the text.

8.2.1 Measure-valued systems in Multi-target tracking

The measure-valued process given by (8.2.1) is a generalisation of Feynman-Kac measures. Its continuous time version naturally arise in the modeling and analysis of the first moments of spatial branching process [28, 40].

Our major motivation for studying this class of measure-valued system stems from advanced signal processing, more specifically, multiple target tracking. Driven primarily in the early 1970's by aerospace applications such as radar, sonar, guidance, navigation, and air traffic control, today multi-target filtering has found applications in many diverse disciplines, see for example the texts [4], [8] [80] and references therein. These nonlinear filtering problems deal with jointly estimating

the number and states of several interacting targets given a sequence of partial observations corrupted by noise, false measurements as well as miss-detection. This rapidly developing subject is, arguably, one of the most interesting contact points between the theory of spatial branching processes, mean field particle systems and advanced signal processing.

The first connections between stochastic branching processes and multi-target tracking seem to go back to the article by S. Mori, et. al. [89] published in 1986. However it was Mahler's systematic treatment of multi-sensor multi-target filtering using random finite sets theory [86, 42, 73, 76] that lead to the development novel multi-target filters and sparked world wide interests. To motivate the article, we briefly outline two recent multi-target filters that do not fit the standard Feynman-Kac's framework, but fall under the umbrella of the measure-valued equation (8.2.1). The first is the Bernoulli filter for joint detection and tracking of a single target while the second is the Probability Hypothesis Density filter.

Bernoulli filtering

A basic problem in target tracking is that the target of interest may not always be present and exact knowledge of target existence/presence cannot be determined from observations due to clutter and detection uncertainty [80]. The *Bernoulli* filter is a generalisation of the standard Bayes filter, which accommodates presence and absence of the target [135]. In a Bernoulli model, the birth of the target at time $n+1$ is modelled by a measure μ_{n+1} on E_{n+1}. The target enters the scene with a probability $\mu_{n+1}(1) < 1$ and its state is distributed according to the normalised measure $\mu_{n+1}/\mu_{n+1}(1)$. At time n, a target X_n has a probability $s_n(X_n)$ of surviving to the next time and evolve to a new state according to a given elementary Markov transition M_{n+1} from E_n into E_{n+1}. At time $n+1$, the target (if it exists) generates with probability $d_{n+1}(X_{n+1})$ an observation Y_{n+1} on some auxiliary state space, say E^Y_{n+1} with likelihood function $l_{n+1}(X_{n+1}, y)$. This so-called Bernoulli observation point process is superimposed with an additional and independent Poisson point process with intensity function $h_n > 0$ to form the occupation (or counting) measure observation process $\mathcal{Y}_{n+1} = \sum_{1\le i\le N^Y_{n+1}} \delta_{Y^i_{n+1}}$.

In its original form, the Bernoulli filter jointly propagates the probability existence of the target and the distribution of the target state [135]. Combining the probability of existence and the state distribution into a single measure, it can be shown that the Bernoulli filter satisfies the integral equation (8.2.1), with the probability of existence of the target given by the mass $\gamma_n(1)$ and the distribution of the target state given by the normalised measure $\eta_n = \gamma_n/\gamma_n(1)$. The integral operator

for the Bernoulli filter takes the following form

$$Q_{n+1,\gamma_n}(x_n, dx_{n+1}) := \frac{s_n(x_n)g_n(x_n)M_{n+1}(x_n, dx_{n+1}) + (\gamma_n(1)^{-1} - 1)\mu_{n+1}(dx_{n+1})}{(1 - \gamma_n(1)) + \gamma_n(g_n)} \tag{8.2.8}$$

where g_n is a likelihood function given by

$$g_n(x_n) \;\; : \;\; = (1 - d_n(x_n)) + d_n(x_n)\mathcal{Y}_n\left(l_n(x_n, \cdot)/h_n\right) \tag{8.2.9}$$

PHD filtering

A more challenging problem arises when the number of targets varies randomly in time, obscured by clutter, detection uncertainty and data association uncertainty. Suppose that at a given time n there are N_n^X targets $(X_n^i)_{1\leq i\leq N_n^X}$ each taking values in some measurable state space E_n. A target X_n^i, at time n, survives to the next time step with probability $s_n(X_n^i)$ and evolves to a new state according to a given elementary Markov transition M'_{n+1} from E_n into E_{n+1}. In addition X_n^i can spawn new targets at the next time, usually modelled by a spatial Poisson process with intensity measure $B_{n+1}(X_n^i, \cdot)$ on E_{n+1}. At the same time, an independent collection of new targets is added to the current configuration. This additional and spontaneous branching process is often modeled by a spatial Poisson process with a prescribed intensity measure μ_{n+1} on E_{n+1}. Each target X_{n+1}^i generates with probability $d_{n+1}(X_{n+1}^i)$ an observation Y_{n+1}^i on some auxiliary state space, say E_{n+1}^Y, with probability density function $g_{n+1}(X_{n+1}^i, y)$. In addition to this partial observation point process we also observe an additional and independent Poisson point process with intensity function h_n. Multi-target tracking concerns the estimation of the random measures $\mathcal{X}_{n+1} = \sum_{1\leq i\leq N_n^X} \delta_{X_n^i}$, given the observation occupation measures $\mathcal{Y}_p = \sum_{1\leq i\leq N_p^Y} \delta_{Y_p^i}$.

The multi-target tracking problem is computationally intractable in general and the Probability Hypothesis Density PHD (filter), is an approximation that propagates the first-order statistical moment, or intensity, of the multi-target state forward in time [76]. The PHD filter satisfies the integral equation (8.2.1), with the integral operator given below

$$Q_{n+1,\gamma_n}(x_n, dx_{n+1}) = g_{n,\gamma_n}(x_n)M_{n+1}(x_n, dx_{n+1}) + \gamma_n(1)^{-1}\ \mu_{n+1}(dx_{n+1}) \tag{8.2.10}$$

where M_{n+1} is a Markov kernel defined by

$$M_{n+1}(x_n, dx_{n+1}) := \frac{s_n(x_n)M'_{n+1}(x_n, dx_{n+1}) + B_{n+1}(x_n, dx_{n+1})}{s_n(x_n) + b_n(x_n)} \tag{8.2.11}$$

with the branching rate $b_n(x_n) = B_{n+1}(1)(x_n)$. The likelihood function g_{n,γ_n} is given by

$$g_{n,\gamma_n} := r_n \times \widehat{g}_{n,\gamma_n} \quad \text{with} \quad r_n := (s_n + b_n) \tag{8.2.12}$$

and

$$\widehat{g}_{n,\gamma_n}(x_n) := (1 - d_n(x_n)) + d_n(x_n) \int \mathcal{Y}_n(dy) \frac{g_n(x_n, y_n)}{h_n(y_n) + \gamma_n(d_n g_n(\cdot, y_n))} \quad (8.2.13)$$

Since its inception by Mahler [76] in 2003, the PHD filter has attracted substantial interest to date. The development of numerical solutions for the PHD filter [133], [130] have opened the door to numerous novel extensions and applications. More details on the derivation of the PHD filter using random finite sets, Poisson techniques or random measures theoretic approaches can be found in the series of articles [13, 76, 120].

8.2.2 Statement of the main results

To describe with some conciseness the main result of this article, we need to introduce some notation. We let $\mathrm{Osc}_1(E_n)$, be the set of $\mathcal{E}_n$-measurable functions f on E_n with oscillations $\mathrm{osc}(f) = \sup_{x,x'} |f(x) - f(x')| \leq 1$. We denote by $\mu(f) = \int \mu(dx)\ f(x)$ the Lebesgue integral of f w.r.t. some measure $\mu \in \mathcal{M}(E_n)$, and we let $|\mu - \nu|_{\mathrm{tv}}$ be the total variation distance between two probability measures ν and μ on E_n.

We assume that the following pair of regularity conditions are satisfied.

(H_1) : There exists a series of compact sets $I_n \subset (0, \infty)$ such that the initial mass value $\gamma_0(1) \in I_0$, and for any $m \in I_n$ $\eta \in \mathcal{P}(E_n)$, we have

$$\theta_{-,n}(m) \leq \eta\,(G_{n,m\eta}) \leq \theta_{+,n}(m) \quad \textit{for some pair of positive functions } \theta_{+/-,n}.$$

The main implication of condition (H_1) comes from the fact that the total mass processes $\gamma_n(1)$ and their N-approximation models $\gamma_n^N(1)$ are finite and they evolves at every time n in a series of compact sets

$$I_n \subset [m_n^-, m_n^+] \subset (0, \infty)$$

with the sequence of parameters $m_n^{+/-}$ defined by the recursive equations $m_{n+1}^- = m_n^- \theta_{-,n}(m_n^-)$ and $m_{n+1}^+ = m_n^+ \theta_{+,n}(m_n^+)$, with the initial conditions $m_0^- = m_0^+ = \gamma_0(1)$.

(H_2) : For any $n \geq 1$, $f \in Osc_1(E_n)$, and any $(m, \eta), (m', \eta') \in (I_n \times \mathcal{P}(E_n))$, the one step mappings $\Gamma_n = (\Gamma_n^1, \Gamma_n^2)$ defined in (8.2.2) satisfy the following Lipschitz type inequalities:

$$\left|\Gamma_n^1(m, \eta) - \Gamma_n^1(m', \eta')\right| \leq c(n)\ |m - m'| + \int |[\eta - \eta'](\varphi)|\ \Sigma_{n,(m',\eta')}^1(d\varphi) \quad (8.2.14)$$

$$\left|\left[\Gamma_n^2(m, \eta) - \Gamma_n^2(m', \eta')\right](f)\right| \leq c(n)\ |m - m'| + \int |[\eta - \eta'](\varphi)|\ \Sigma_{n,(m',\eta')}^2(f, d\varphi) \quad (8.2.15)$$

for some finite constants $c(n) < \infty$, and some collection of bounded measures $\Sigma^1_{n,(m',\eta')}$ and $\Sigma^2_{n,(m',\eta')}(f, \bullet)$ on $\mathcal{B}(E_n)$ such that

$$\int osc(\varphi)\ \Sigma^1_{n,(m,\eta)}(d\varphi) \leq \delta\left(\Sigma^1_n\right) \quad \text{and} \quad \int osc(\varphi)\ \Sigma^2_{n,(m,\eta)}(f, d\varphi) \leq \delta\left(\Sigma^2_n\right)$$

for some finite constant $\delta\left(\Sigma^i_n\right) < \infty$, $i = 1, 2$, whose values do not depend on the parameters $(m, \eta) \in (I_n \times \mathcal{P}(E_n)$ and $f \in Osc_1(E_n)$.

Condition (H_2) is a rather basic and weak continuity type property. It states that the one step transformations of the flow of measures (8.2.2) are weakly Lipschitz, in the sense that the mass variations and the integral differences w.r.t. some test function f can be controlled by the different initial masses and measures w.r.t. a collection of integrals of a possibly infinite number of test functions. It is satisfied for a large class of one step transformations Γ_n. In section 8.3.3, we will verify that it is satisfied for the general class of Bernoulli and the PHD filters discussed in section 8.2.1 and section 8.2.1.

We are now in position to state the main results of this article. The first one is concerned with the exponential stability properties of the semigroup $\Gamma_{p,n} = \left(\Gamma^1_{p,n}, \Gamma^2_{p,n}\right)$, with $0 \leq p \leq n$ associated with the one step transformations of the flow (8.2.2). A more precise description and the complete proof of the next theorem is provided in section 8.4.

Theorem 8.2.1. *We let $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ be the semigroups associated with the one step transformations of the flow of total masses $\Phi^1_{n,\nu_{n-1}} := \Gamma^1_n(\bullet, \nu_{n-1})$ and measures $\Phi^2_{n,m_{n-1}} := \Gamma^2_n(m_{n-1}, \bullet)$, with a fixed collection of measures $\nu := (\nu_n)_{n\geq 0} \in \prod_{n\geq 0} \mathcal{P}(E_n)$ and masses $m := (m_n)_{n\geq 0} \in \prod_{n\geq 0} I_n$. When these semigroups are exponentially stable (in the sense that they forget exponentially fast their initial conditions) and when the pair of mappings $\nu_{n-1} \mapsto \Phi^1_{n,\nu_{n-1}}$ and $m_{n-1} \mapsto \Phi^2_{n,m_{n-1}}$ are sufficiently regular then we have the following contraction inequalities*

$$\left|\Gamma^1_{p,n}(u', \eta') - \Gamma^1_{p,n}(u, \eta)\right| \vee \left|\Gamma^2_{p,n}(u', \eta') - \Gamma^2_{p,n}(u, \eta)\right|_{tv} \leq c\ e^{-\lambda(n-p)}$$

for any $p \leq n$, $u, u' \in I_p$, $\eta, \eta' \in \mathcal{P}(E_p)$, and some finite constants $c < \infty$ and $\lambda > 0$ whose values do not depend on the time parameters $p \leq n$.

The second theorem is concerned with estimating the approximation error associated with a N-approximation model satisfying condition (8.2.6). The first part of the theorem is proved in section 8.2.2. The proof of the uniform estimates is discussed in section 8.4.1 (see for instance lemma 8.4.4).

Theorem 8.2.2. *Under the assumptions (H_1) and (H_2), the semigroup $\Gamma_{p,n}$ satisfies the same Lipschitz type inequalities as those stated in (8.2.14) and (8.2.15) for*

some collection of measures $\Sigma^1_{p,n}$ *and* $\Sigma^2_{p,n}(f,\cdot)$ *on* $\mathcal{B}(E_p)$*. In addition, for any* N*-approximation model satisfying condition (8.2.6) we have the estimates:*

$$\mathbb{E}\left(\left|V_n^{\gamma,N}(1)\right|^r\right)^{\frac{1}{r}} \leq a_r \sum_{p=0}^{n} \delta\left(\Sigma^1_{p,n}\right) \quad \text{and} \quad \mathbb{E}\left(\left|V_n^{\eta,N}(f)\right|^r\right)^{\frac{1}{r}} \leq a_r \sum_{p=0}^{n} \delta\left(\Sigma^2_{p,n}\right) \tag{8.2.16}$$

for any $r \geq 1$*, and* $N \geq 1$*, with some constants* $a_r < \infty$ *whose values only depend on* r*. Furthermore, under the regularity conditions of theorem 8.2.1 the couple of estimates stated above are uniform w.r.t. the time horizon; that is, we have that* $\sup_{n\geq 0} \sum_{p=0}^{n} \delta\left(\Sigma^i_{p,n}\right) < \infty$*, for any* $i = 1, 2$*.*

These rather abstract theorems apply to a general class of discrete generation measure-valued equations of the form (8.2.1). We illustrate the application of this pair of theorems in the analysis of the stability properties and the approximation convergence of the pair of multiple target filters presented in this introductory section. These results can basically be stated as follows:

- The Bernoulli filter presented in section 8.2.1 with a sufficiently mixing prediction and almost equal survival and spontaneous births rates $s_n \sim \mu_n(1)$ is exponentially stable.

- The PHD filter presented in section 8.2.1 is exponentially stable for small clutter intensities and sufficiently high detection probability and spontaneous birth rates.

- In both situations, the estimation error of any N-approximation model satisfying condition (8.2.6) does not accumulate over time. Furthermore, the uniform rates of convergence provided in theorem 8.2.2 allows to design stochastic algorithms with prescribed performance index at any time horizon.

We end this section with some direct consequences of theorem 8.2.2:

Firstly, we observe that the mean error estimates stated in the above theorem clearly implies the almost sure convergence results

$$\lim_{N\to\infty} \eta_n^N(f) = \eta_n(f) \quad \text{and} \quad \lim_{N\to\infty} \gamma_n^N(f) = \gamma_n(f)$$

for any bounded function f on E_n. Furthermore, with some information on the constants a_r, these $\mathbb{L}_r$-mean error bounds can be turned to exponential concentration inequalities. To be more precise, by lemma 7.3.3 in [28], the collection of constants a_r in theorem 8.2.2, can be chosen so that

$$a_{2r}^{2r} \leq b^{2r}\ (2r)!\ 2^{-r}/r! \quad \text{and} \quad a_{2r+1}^{2r+1} \leq b^{2r+1}(2r+1)!\ 2^{-r}/r! \tag{8.2.17}$$

for some $b < \infty$, whose values do not depend on r. Using the above $\mathbb{L}_r$-mean error bounds we can establish the following non asymptotic Gaussian tail estimates:

$$\mathbb{P}\left(\left|\left[\eta_n^N - \eta_n\right](f)\right| \geq \frac{b_n}{\sqrt{N}} + \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{2b_n^2}\right) \quad \text{with} \quad b_n \leq b \sum_{p=0}^{n} \delta\left(\Sigma_{p,n}^2\right)$$

The above result is a direct consequence of the following observation

$$\forall r \geq 1 \qquad \mathbb{E}\left(U^r\right)^{\frac{1}{r}} \leq a_r\ b \Rightarrow \mathbb{P}\left(U \geq b + \epsilon\right) \leq \exp\left(-\epsilon^2/(2b)\right)$$

for any non negative random variable U. To check this claim, we use the following Laplace estimate

$$\forall t \geq 0 \quad \mathbb{E}\left(e^{tU}\right) \ \leq \ \exp\left(\frac{(bt)^2}{2} + bt\right) \Rightarrow \mathbb{P}\left(U \geq b + \epsilon\right) \leq \exp\left(-\sup_{t \geq 0}\left(\epsilon t - \frac{(bt)^2}{2}\right)\right)$$

It is worth noting that the above constructions allows us to consider with further work branching particle models in path spaces. These path space models arise in the analysis of the historical process associated with a branching models as well as the analysis of a filtering problem of the whole signal path given a series of observations. For instance, let us suppose that the Markov transitions M_n defined in (8.2.10) are the elementary transition of a Markov chain of the following form

$$X_n := \left(X'_p\right)_{0 \leq p \leq n} \in E_n := \prod_{0 \leq p \leq n} E'_p$$

In other words X_n represents the paths from the origin up to the current time of an auxiliary Markov chain X'_n taking values in some measurable state spaces E'_n, with Markov transitions M'_n. We assume that the potential functions g_{n,γ_n} only depend on the terminal state of the path, in the sense that $g_{n,\gamma_n}(X_n) = g'_{n,\gamma_n}(X'_n)$, for some potential function g'_{n,γ_n} on E'_n. In multiple target tracking problems, these path space models provide a way to estimate the conditional intensity of the path of a given target in a multi-target environment related to some likelihood function that only depends on the terminal state of the signal path.

In practice, it is essential to observe that the mean field particle interpretations of these path space models simply consist of keeping track of the whole history of each particle. It can be shown that the resulting particle model can be interpreted as the genealogical tree model associated with a genetic type model (see for instance [28]). In this situation, η_n^N is the occupation measure of a random genealogical tree, each particle represents the ancestral lines of the current individuals.

We end this section with some standard notation used in the paper:

We denote respectively by $\mathcal{M}(E)$, $\mathcal{P}(E)$, and $\mathcal{B}(E)$, the set of all finite positive measures μ on some measurable space $(E, \mathcal{E})$, the convex subset of all probability measures, and the Banach space of all bounded and measurable functions f

equipped with the uniform norm $\|f\|$. We denote by f^- and f^+ the infimum and the supremum of a function f. For measurable subsets $A \in \mathcal{E}$, in various instances we slightly abuse notation and we denote $\mu(A)$ instead of $\mu(1_A)$; and we set δ_a the Dirac measure at $a \in E$. We recall that a bounded and positive integral operator Q from a measurable space $(E_1, \mathcal{E}_1)$ into an auxiliary measurable space $(E_2, \mathcal{E}_2)$ is an operator $f \mapsto Q(f)$ from $\mathcal{B}(E_2)$ into $\mathcal{B}(E_1)$ such that the functions

$$x \mapsto Q(f)(x) := \int_{E_2} Q(x, dy) f(y)$$

are $\mathcal{E}_1$-measurable and bounded for some measures $Q(x, .) \in \mathcal{M}(E_2)$. These operators also generate a dual operator $\mu \mapsto \mu Q$ from $\mathcal{M}(E_1)$ into $\mathcal{M}(E_2)$ defined by $(\mu Q)(f) := \mu(Q(f))$. A Markov kernel is a positive and bounded integral operator M with $M(1) = 1$. We denote by $Q_{p,n} = Q_{p+1} Q_{p+2} \ldots Q_n$, with $p \leq n$ the semigroup associated with a given sequence of bounded and positive integral operator Q_n from some measurable spaces $(E_{n-1}, \mathcal{E}_{n-1})$ into $(E_n, \mathcal{E}_n)$. For $p = n$, we use the convention $Q_{n,n} = Id$, the identity operator.

We associate with a bounded positive potential function $G : x \in E \mapsto G(x) \in [0, \infty)$, the Bayes-Boltzmann-Gibbs transformations

$$\Psi_G \; : \; \eta \in \mathcal{M}(E) \mapsto \Psi_G(\eta) \in \mathcal{P}(E) \quad \text{with} \quad \Psi_G(\eta)(dx) := \frac{1}{\eta(G)} \; G(x) \; \eta(dx)$$

provided $\eta(G) > 0$. We recall that $\Psi_G(\eta)$ can be expressed in terms of a Markov transport equation

$$\eta S_\eta = \Psi_G(\eta) \tag{8.2.18}$$

for some selection type transition $S_\eta(x, dy)$. For instance, we can take

$$S_\eta(x, dy) := \frac{\epsilon}{\eta(G)} \; \delta_x(dy) + \left(1 - \frac{\epsilon}{\eta(G)}\right) \; \Psi_{(G-\epsilon)}(\eta)(dy) \tag{8.2.19}$$

for any $\epsilon \geq 0$ s.t. $G(x) \geq \epsilon$. Notice that for $\epsilon = 0$, we have $S_\eta(x, dy) = \Psi_G(\eta)(dy)$. We can also choose

$$S_\eta(x, dy) := \epsilon G(x) \; \delta_x(dy) + (1 - \epsilon G(x)) \; \Psi_G(\eta)(dy) \tag{8.2.20}$$

for any $\epsilon \geq 0$ that may depend on the current measure η, and s.t. $\epsilon G(x) \leq 1$. For instance, we can choose $1/\epsilon$ to be the η-essential maximum of the potential function G. Finally, in the context of Bernoulli and PHD filtering we set $\overline{\mu}_{n+1} = \mu_{n+1}/\mu_{n+1}(1)$, for any $n \geq 0$, the normalized spontaneous birth measures.

8.3 Semigroup description

8.3.1 The Bernoulli filter semigroup

By construction, we notice that the mass process and the normalized measures are given by the rather simple recursive formulae

$$\gamma_{n+1}(1) = \frac{\gamma_n(1)\eta_n(g_n)}{(1-\gamma_n(1)) + \gamma_n(1)\eta_n(g_n)}\ \Psi_{g_n}(\eta_n)(s_n) + \frac{(1-\gamma_n(1))}{(1-\gamma_n(1)) + \gamma_n(1)\eta_n(g_n)}\ \mu_{n+1}(1) \tag{8.3.1}$$

and

$$\eta_{n+1} := \alpha_n(\gamma_n)\ \Psi_{g_n s_n}(\eta_n) M_{n+1} + (1-\alpha_n(\gamma_n))\ \overline{\mu}_{n+1}$$

with the mappings $\alpha_n\ :\ \gamma \in \mathcal{M}(E_n) \mapsto \alpha_n(\gamma) \in [0,1]$ defined by

$$\alpha_n(\gamma) = \frac{\gamma(g_n s_n)}{\gamma(s_n g_n) + (1-\gamma(1))\mu_{n+1}(1)}$$

By construction, if we set $\gamma = m \times \eta$ then

$$\begin{aligned} \Gamma^1_{n+1}(m,\eta) &= \frac{\gamma(g_n)}{(1-m)+\gamma(g_n)}\ \Psi_{g_n}(\eta)(s_n) + \frac{(1-m)}{(1-m)+\gamma(g_n)}\ \mu_{n+1}(1) \\ \Gamma^2_{n+1}(m,\eta) &= \Psi_{g_n s_n}(\eta) M_{n+1,\gamma} \end{aligned}$$

with the collection of Markov transitions $M_{n+1,\gamma}$ defined below

$$M_{n+1,\gamma}(x,\boldsymbol{.}) := \alpha_n(\gamma)\, M_{n+1}(x,\boldsymbol{.}) + (1-\alpha_n(\gamma))\ \overline{\mu}_{n+1} \tag{8.3.2}$$

Next we provide an alternative interpretation of the mapping Γ^2_{n+1}. Firstly, observe that

$$\Psi_{g_n s_n}(\eta) M_{n+1,\gamma}(f) = \frac{\eta\left(Q_{n+1,m}(f)\right)}{\eta\left(Q_{n+1,m}(1)\right)} \tag{8.3.3}$$

with the integral operator

$$Q_{n+1,m}(f)(x) := m g_n(x) s_n(x) M_{n+1}(f)(x) + (1-m)\ \mu_{n+1}(f)$$

This implies that

$$\Gamma^2_{n+1}(m,\eta) = \Psi_{\widehat{G}_{n,m}}(\eta)\widehat{M}_{n+1,m}$$

with the potential function

$$\widehat{G}_{n,m} = m g_n s_n + (1-m)\ \mu_{n+1}(1) \tag{8.3.4}$$

and the Markov transitions

$$\widehat{M}_{n+1,m}(f) := \frac{m g_n s_n}{m g_n s_n + (1-m)\ \mu_{n+1}(1)}\ M_{n+1}(f) + \frac{(1-m)\mu_{n+1}(1)}{m g_n s_n + (1-m)\ \mu_{n+1}(1)}\ \overline{\mu}_{n+1}(f) \tag{8.3.5}$$

The condition (H_1) is clearly not met for the Bernoulli filter (8.2.8) when $s_n = 0$ and $\mu_{n+1}(1) = 0$, since in this situation $\gamma_n = 0$ for any $n \geq 1$. Nevertheless, this condition is met with $I_n \subset (0, 1]$ and $m\theta_{+,n}(m) = 1$, as long as s_n and $\mu_{n+1}(1)$ are uniformly bounded from below. It is also met for $s_n = 0$, as long as $0 < \mu_{n+1}(1) < 1$ and the likelihood function given in (8.2.9) is uniformly bounded. The condition is also met for $\mu_{n+1}(1) = 0$, as long as $\gamma_0(1) > 0$, and the likelihood function given in (8.2.9) and the function s_n are uniformly lower bounded.

We prove these assertions using the fact that

$$\gamma_{n+1}(1) = \widehat{\gamma}_n(1)\ \Psi_{g_n}(\eta_n)(s_n) + (1 - \widehat{\gamma}_n(1))\ \mu_{n+1}(1) \tag{8.3.6}$$

with the updated mass parameters $\widehat{\gamma}_n(1) \in [0, 1]$ given below

$$\widehat{\gamma}_n(1) := \frac{\gamma_n(1)\eta_n(g_n)}{(1 - \gamma_n(1)) + \gamma_n(1)\eta_n(g_n)}$$

If we set $s_n^- := \inf_{E_n} s_n$ and $s_n^+ = \sup_{E_n} s_n$ then

$$\forall n \geq 1 \qquad \gamma_n(1) \in \left[m_n^-, m_n^+\right]$$

with parameters

$$m_n^- = \mu_n(1) \wedge s_{n-1}^- \quad \text{and} \quad m_n^+ = \mu_n(1) \vee s_{n-1}^+ \ (\leq 1)$$

If s_n and $\mu_{n+1}(1)$ are uniformly bounded from below then we have $m_n^- > 0$. In addition, for the constant mapping $s_n = \mu_{n+1}(1)$, the total mass process is constant

$$\gamma_{n+1}(1) = m_{n+1}^+ = m_{n+1}^- = \mu_{n+1}(1)$$

for any $n \geq 0$. Furthermore, in this situation the flow of normalized measures is given by the updating-prediction transformation defined by

$$\forall n \geq 0 \qquad \eta_{n+1} = \Psi_{g_n^{(s)}}(\eta_n)\, M_{n+1}^{(s)}$$

with the likelihood function $g_n^{(s)}$ and the Markov transitions $M_{n+1}^{(s)}$ defined by

$$g_n^{(s)} := s_n g_n + (1 - s_n) \quad \text{and} \quad M_{n+1}^{(s)}(f) := \frac{s_n g_n M_{n+1}(f) + (1 - s_n)\ \overline{\mu}_{n+1}(f)}{s_n g_n + (1 - s_n)} \tag{8.3.7}$$

When $\mu_{n+1}(1) = 0$, the flow of normalized measures is again given by a simple updating-prediction equation

$$\eta_{n+1} = \Psi_{g_n s_n}(\eta_n) M_{n+1} \quad \text{and} \quad \gamma_{n+1}(1) = \Psi_{g_n}(\eta_n)(s_n) \times \theta_{\eta_n(g_n)}(\gamma_n(1)) \tag{8.3.8}$$

with the increasing mappings θ_a defined below

$$x \in [0, 1] \mapsto \theta_a(x) := ax/[ax + (1 - x)] \tag{8.3.9}$$

In addition, if $s_n^- > 0$ then

$$m_{n+1}^- \geq s_n^- \times \frac{g_n^- m_n^-}{g_n^- m_n^- + (1 - m_n^-)} > 0$$

as long as $g_n^- := \inf_{E_n} g_n > 0$, and $\gamma_0(1) > 0$. We prove this inequality using the fact that the mapping $(a, x) \in [0, \infty[\times[0, 1] \mapsto \theta_a(x)$ is increasing in both coordinates. In the case where $s_n = 1$, using the fact that and $\theta_a \circ \theta_b = \theta_{ab}$, we prove that

$$\gamma_{n+1}(1) = \theta_{\eta_n(g_n)}(\gamma_n(1)) = \theta_{\prod_{p=0}^n \eta_p(g_p)}(\gamma_0(1))$$

Conversely, when $\gamma_0(1) < 1$ and $0 < \mu_{n+1}(1) < 1$ and $s_n = 0$, for any $n \geq 0$, then we have a constant flow of normalized measures

$$\forall n \geq 1 \qquad \eta_n = \overline{\mu}_n$$

and the total mass process is such that

$$\gamma_n(1) \in]0, 1[\implies \gamma_{n+1}(1) = \mu_{n+1}(1) \times \left[1 - \theta_{\overline{\mu}_n(g_n)}(\gamma_n(1))\right] \in]0, 1[$$

with the convention $\overline{\mu}_0 = \eta_0$, for $n = 0$. In addition, if $\mu_{n+1}(1) = 1$ then we have

$$\gamma_{2(n+1)}(1) = \theta_{\prod_{p=0}^n (b_{2p}/b_{2p+1})}(\gamma_0(1)) \quad \text{and} \quad \gamma_{2n+1}(1) = \theta_{b_{2n}^{-1} \prod_{p=0}^{n-1} (b_{2p+1}/b_{2p})}(\gamma_0(1))$$

for any $n \geq 0$, with the parameters $b_n := \overline{\mu}_n(g_n)$. We prove these formuae using the the fact that $1 - \theta_a(x) = \theta_{1/a}(1 - x)$, and $\theta_a \circ \theta_b = \theta_{ab}$. This again implies that $m_n^- > 0$ as long as $\gamma_0(1) > 0$ and the likelihood function are uniformly lower bounded.

8.3.2 The PHD filter semigroup

By construction, if we set $\gamma = m \times \eta$ then we find that

$$\Gamma_{n+1}^1(m, \eta) = \gamma(g_{n,\gamma}) + \mu_{n+1}(1) \quad \text{and} \quad \Gamma_{n+1}^2(m, \eta) = \Psi_{g_{n,\gamma}}(\eta) M_{n+1,\gamma}$$

In the above display, $M_{n+1,\gamma}$ is the collection of Markov transitions defined below

$$M_{n+1,\gamma}(x, .) := \alpha_n(\gamma)\, M_{n+1}(x, .) + (1 - \alpha_n(\gamma))\, \overline{\mu}_{n+1} \quad \text{with} \quad \alpha_n(\gamma) = \frac{\gamma(g_{n,\gamma})}{\gamma(g_{n,\gamma}) + \mu_{n+1}(1)}$$

The interpretation of the updating transformation $\Psi_{g_{n,\gamma}}(\eta)$ in terms of a Markov transport equation is non unique. For instance, using (8.2.12) this Bolzmann-Gibbs transformation can be decomposed into two parts. The first one relates to the undetectable targets and the second is associated with non clutter observations. An

alternative description is provided below. We consider a virtual auxiliary observation point c (corresponding to undetectable targets) and set $\mathcal{Y}_n^c = \mathcal{Y}_n + \delta_c$. We also denote by $g_{n,\gamma}^c(\bullet, y)$ the function defined below

$$g_n^\gamma(\bullet, y) = \begin{cases} r_n(1-d_n) & \text{if } \ y = c \\ r_n \ \dfrac{d_n g_n(\bullet, y_n)}{h_n(y) + \gamma(d_n g_n(\bullet, y))} & \text{if } \ y \neq c \end{cases}$$

In this notation, the updating transformation $\Psi_{g_{n,\gamma}}(\eta)$ can be rewritten in the following form

$$\Psi_{g_{n,\gamma}}(\eta) = \Psi_{\overline{g}_{n,\gamma}}(\eta) \quad \text{with} \quad \overline{g}_{n,\gamma} = \int \mathcal{Y}_n^c(dy)\ g_n^\gamma(\bullet, y)$$

The averaged potential function $\overline{g}_{n,\gamma}$ allows us to measure the likelihood of signal states w.r.t. the current observation measure $\mathcal{Y}_n^c$. Using (8.2.18), the Bolzmann-Gibbs transformation $\Psi_{\overline{g}_{n,\gamma}}(\eta)$ can be interpreted as non linear Markov transport equation of the following form

$$\Psi_{\overline{g}_{n,\gamma}}(\eta) = \eta S_{n,\gamma} \quad \text{and} \quad \Gamma^2(m,\eta) = \eta K_{n+1,\gamma} \quad \text{with} \quad K_{n+1,\gamma} = S_{n,\gamma} M_{n+1,\gamma} \tag{8.3.10}$$

for some Markov transitions $S_{n,\gamma}$ from E_n into itself.

We also notice that condition (H_1) holds as long as the functions s_n, b_n, and $g_n(\bullet, y_n)$ are uniformly bounded and $\mu_n(1) > 0$. It is also met when $\mu_n(1) = 0$, as long as $r_n = (s_n + b_n)$ is uniformly lower bounded and $\mathcal{Y}_n \neq 0$ or $d_n < 1$.

8.3.3 Lipschitz regularity properties

Firstly, we mention that condition (H_2) can be replaced by the following regularity condition:

(H_2') : For any $n \geq 1$, $f \in Osc_1(E_n)$, and any $(m,\eta), (m',\eta') \in (I_n \times \mathcal{P}(E_n))$, the integral operators $Q_{n,m\eta}$ satisfy the following Lipschitz type inequalities:

$$|Q_{n,m\eta}(f) - Q_{n,m'\eta'}(f)| \leq c(n)\ |m - m'| + \int \ |[\eta - \eta'](\varphi)|\ \Sigma_{n,(m',\eta')}(f, d\varphi) \tag{8.3.11}$$

for some collection of bounded measures $\Sigma_{n,(m',\eta')}(f, \bullet)$ on $\mathcal{B}(E_n)$ such that

$$\int \ osc(\varphi)\ \Sigma_{n,(m,\eta)}(f, d\varphi) \leq \delta\,(\Sigma_n)$$

for some finite constant $\delta\,(\Sigma_n) < \infty$, whose values do dot depend on the parameters $(m,\eta) \in (I_n \times \mathcal{P}(E_n))$ and $f \in Osc_1(E_n)$.

We prove $(H_2') \Rightarrow$ (8.2.14) using the decompositions

$$m\eta Q_{n,m\eta} - m'\eta' Q_{n,m'\eta'} = m\eta\,[Q_{n,m\eta} - Q_{n,m'\eta'}] + [m\eta - m'\eta']\,Q_{n,m'\eta'}$$

and of course $[m\eta - m'\eta'] = [m - m']\,\eta + m'\,[\eta - \eta']$. To prove $(H_2') \Rightarrow$ (8.2.15), we let $\gamma = m\eta$ and $\gamma' = m'\eta'$ and we use the decomposition

$$\left[\Gamma_n^2(m,\eta) - \Gamma_n^2(m',\eta')\right](f) = \frac{1}{\gamma Q_{n,\gamma}(1)}\ \left[\gamma Q_{n,\gamma} - \gamma' Q_{n,\gamma'}\right]\left(f - \Gamma_n^2(m',\eta')(f)\right)$$

The Bernoulli filter (8.2.8) satisfies (H_2'), as long as the likelihood functions g_n given in (8.2.9) are uniformly bounded above. In this situation, (8.3.11) is met with

$$|Q_{n,m\eta}(f) - Q_{n,m'\eta'}(f)| \leq c(n)\ |m - m'| + c'(n)\,|[\eta - \eta'](g_n)|$$

for some finite constant $c'(n) < \infty$.

The PHD equation satisfies (H_2'), as long as the functions $h_n(y) + g'_{n,y}$ with $g'_{n,y} := d_n g_n(\bullet, y)$ are uniformly bounded above and below. To prove this claim, we simply use the fact that

$$\|\widehat{g}_{n,\gamma} - \widehat{g}_{n,\gamma'}\| \leq c_n \left[|m' - m| + \int \mathcal{Y}_n(dy)\ \left|[\eta' - \eta](g'_{n,y})\right|\right]$$

This estimate is a direct consequence of the following one

$$\widehat{g}_{n,\gamma}(x) - \widehat{g}_{n,\gamma'}(x) = \int \mathcal{Y}_n(dy)\ \frac{g'_{n,y}(x)}{h_n(y) + \gamma(g'_{n,y})} \frac{[\gamma' - \gamma]\,(g'_{n,y})}{h_n(y) + \gamma'(g'_{n,y})}$$

Next, we provide a pivotal regularity property of the semigroup $(\Gamma_{p,n})_{0\leq p\leq n}$ associated with the one step transformations of the flow (8.2.2).

Proposition 8.3.1. *We assume that conditions (H_1) and (H_2) are satisfied. Then, for any $0 \leq p \leq n$, $f \in Osc_1(E_n)$, and any $(m,\eta),(m',\eta') \in (I_p \times \mathcal{P}(E_p))$, we have the following Lipschitz type inequalities:*

$$\begin{aligned}
\left|\Gamma_{p,n}^1(m,\eta) - \Gamma_{p,n}^1(m',\eta')\right| &\leq c_p(n)\ |m - m'| + \int\ |[\eta - \eta'](\varphi)|\ \Sigma_{p,n,(m',\eta')}^1(d\varphi) \\
\left|\left[\Gamma_{p,n}^2(m,\eta) - \Gamma_{p,n}^2(m',\eta')\right](f)\right| &\leq c_p(n)\ |m - m'| + \int\ |[\eta - \eta'](\varphi)|\ \Sigma_{p,n,(m',\eta')}^2(f, d\varphi)
\end{aligned}$$

for some finite constants $c_p(n) < \infty$, and some collection of bounded measures $\Sigma_{p,n,(m',\eta')}^1$ and $\Sigma_{p,n,(m',\eta')}^2(f,\bullet)$ on $\mathcal{B}(E_p)$ such that

$$\int\ osc(\varphi)\ \Sigma_{p,n,(m,\eta)}^1(d\varphi) \leq \delta\left(\Sigma_{p,n}^1\right) \quad \text{and} \quad \int\ osc(\varphi)\ \Sigma_{p,n,(m,\eta)}^2(f, d\varphi) \leq \delta\left(\Sigma_{p,n}^2\right) \tag{8.3.12}$$

for some finite constant $\delta\left(\Sigma_{p,n}^i\right) < \infty$, $i = 1, 2$, whose values do dot depend on the parameters $(m,\eta) \in (I_p \times \mathcal{P}(E_p)$ and $f \in Osc_1(E_n)$.

Proof:
To prove this proposition, we use a backward induction on the parameter $1 \leq p \leq n$. For $p = (n-1)$, we have $\Gamma^i_{n-1,n} = \Gamma^i_n$, with $i = 1, 2$, so that the desired result is satisfied for $p = (n-1)$. We further assume that the estimates hold at a given rank $p < n$. To prove the estimates at rank $(p-1)$, we recall that

$$\Gamma_{p-1,n}(m,\eta) = \Gamma_{p,n}\left(\Gamma_p(m,\eta)\right) \Rightarrow \forall i = 1,2 \qquad \Gamma^i_{p-1,n}(m,\eta) = \Gamma^i_{p,n}\left(\Gamma_p(m,\eta)\right)$$

Under the induction hypothesis

$$\begin{aligned}\left|\Gamma^1_{p-1,n}(m,\eta) - \Gamma^1_{p-1,n}(m',\eta')\right| &= \left|\Gamma^1_{p,n}\left(\Gamma_p(m,\eta)\right) - \Gamma^1_{p,n}\left(\Gamma_p(m',\eta')\right)\right| \\ &\leq c_p(n)\ \left|\Gamma^1_p(m,\eta) - \Gamma^1_p(m',\eta')\right| \\ &\quad + \int \left|\left[\Gamma^2_p(m,\eta) - \Gamma^2_p(m',\eta')\right](\varphi)\right|\ \Sigma^1_{p,n,\Gamma_p(m',\eta')}(d\varphi)\end{aligned}$$

On the other hand

$$\left|\Gamma^1_p(m,\eta) - \Gamma^1_p(m',\eta')\right| \leq c(p)\ |m - m'| + \int\ |[\eta - \eta'](\varphi)|\ \Sigma^1_{p,(m',\eta')}(d\varphi)$$

and

$$\left|\left[\Gamma^2_p(m,\eta) - \Gamma^2_p(m',\eta')\right](\varphi)\right| \leq c(p)\ |m - m'| + \int\ |[\eta - \eta'](\psi)|\ \Sigma^2_{p,(m',\eta')}(\varphi, d\psi)$$

The end of the proof is now clear. The analysis of $\Gamma^2_{p-1,n}$ follows the same line of arguments and is omitted. This ends the proof of the proposition. ∎

8.3.4 Proof of theorem 8.2.2

This section is mainly concerned with the proof of the couple of estimates (8.2.16) stated in theorem 8.2.2.

We use the decomposition

$$\begin{aligned}\left(\gamma^N_n(1), \eta^N_n\right) - \left(\gamma_n(1), \eta_n\right) &= \left[\Gamma_{0,n}\left(\gamma^N_0(1), \eta^N_0\right) - \Gamma_{0,n}\left(\gamma_0(1), \eta_0\right)\right] \\ &\quad + \sum_{p=1}^{n} \left[\Gamma_{p,n}\left(\gamma^N_p(1), \eta^N_p\right) - \Gamma_{p-1,n}\left(\gamma^N_{p-1}(1), \eta^N_{p-1}\right)\right] \qquad (8.3.13)\end{aligned}$$

and the fact that

$$\Gamma_{p-1,p}\left(\gamma^N_{p-1}(1), \eta^N_{p-1}\right) = \left(\gamma^N_p(1), \Gamma^2_{p-1,p}\left(\gamma^N_{p-1}(1), \eta^N_{p-1}\right)\right)$$

to show that

$$\begin{aligned}\gamma^N_n(1) - \gamma_n(1) &= \left[\Gamma^1_{0,n}\left(\gamma^N_0(1), \eta^N_0\right) - \Gamma^1_{0,n}\left(\gamma_0(1), \eta_0\right)\right] \\ &\quad + \sum_{p=1}^{n} \left[\Gamma^1_{p,n}\left(\gamma^N_p(1), \eta^N_p\right) - \Gamma^1_{p,n}\left(\gamma^N_p(1), \Gamma^2_{p-1,p}\left(\gamma^N_{p-1}(1), \eta^N_{p-1}\right)\right)\right]\end{aligned}$$

Recalling that $\gamma_0^N(1) = \gamma_0(1)$, using proposition 8.3.1, we find that

$$\sqrt{N}\left|\gamma_n^N(1) - \gamma_n(1)\right| \leq \sum_{p=0}^{n} c_p(n) \int \left|[W_p^N(\varphi)\right| \Sigma_{p,n}^{(N,1)}(d\varphi)$$

with the predictable measure $\Sigma_{p,n}^{(N,1)} = \Sigma_{p,n,(m,\eta)}^1$ associated with the parameters $(m,\eta) = (\gamma_p^N(1), \Gamma_{p-1,p}^2(\gamma_{p-1}^N(1), \eta_{p-1}^N))$, with $0 < p \leq n$; and for $p = 0$, we set $\Sigma_{0,n}^{(N,1)} = \Sigma_{0,n,(\gamma_0(1),\eta_0)}$. Combing the generalized Minkowski's inequality with (8.2.6) we have

$$\mathbb{E}\left(\left|\int \left|W_p^N(\varphi)\right| \Sigma_{p,n}^{(N,1)}(d\varphi)\right|^r \left|\mathcal{F}_{p-1}^{(N)}\right)^{\frac{1}{r}} \leq a_r\ \delta\left(\Sigma_{p,n}^1\right)$$

for some constants a_r whose values only depend on the time parameter. This clearly implies that

$$\mathbb{E}\left(\left|\gamma_n^N(1) - \gamma_n(1)\right|^r\right)^{\frac{1}{r}} \leq a_r \sum_{p=0}^{n} \delta\left(\Sigma_{p,n}^1\right)$$

The normalized occupation measures can be analyzed in the same way using the decomposition given below:

$$\begin{aligned}\eta_n^N - \eta_n &= \left[\Gamma_{0,n}^2\left(\gamma_0^N(1), \eta_0^N\right) - \Gamma_{0,n}^2\left(\gamma_0(1), \eta_0\right)\right] \\ &\quad + \sum_{p=1}^{n}\left[\Gamma_{p,n}^2\left(\gamma_p^N(1), \eta_p^N\right) - \Gamma_{p,n}^2\left(\gamma_p^N(1), \eta_{p-1}^N K_{p,(\gamma_{p-1}^N(1),\eta_{p-1}^N)}\right)\right]\end{aligned}$$

This ends the proof of the theorem 8.2.2. ∎

8.4 Functional contraction inequalities

8.4.1 Stability properties

This section is concerned with the long time behavior of nonlinear measure-valued processes of the form (8.2.2). The complexity of these models depend in part on the interaction function between the flow of masses $\gamma_n(1)$ and the flow of probability measures $\eta_n = \gamma_n/\gamma_n(1)$. One natural way to start the analysis of these models is to study the stability properties of the measure-valued semigroup associated with a fixed flow of masses, and vice versa. These two mathematical objects are defined below.

Definition 8.4.1. *We associate with a flow of masses $m = (m_n)_{n\geq 0} \in \prod_{n\geq 0} I_n$ and probability measures $\nu := (\nu_n)_{n\geq 0} \in \prod_{n\geq 0} \mathcal{P}(E_n)$ the pair of semigroups*

$$\Phi_{p,n,\nu}^1 := \Phi_{n,\nu_{n-1}}^1 \circ \ldots \circ \Phi_{1,\nu_0}^1 \quad \textit{and} \quad \Phi_{p,n,m}^2 := \Phi_{n,m_{n-1}}^2 \circ \ldots \circ \Phi_{1,m_0}^2 \tag{8.4.1}$$

with $0 \leq p \leq n$, and the one step transformations

$$\begin{aligned} \Phi^1_{n,\nu_{n-1}} &: u \in I_{n-1} \mapsto \Phi^1_{n,\nu_{n-1}}(u) := \Gamma^1_n(u, \nu_{n-1}) \in I_n \\ \Phi^2_{n,m_{n-1}} &: \eta \in \mathcal{P}(E_{n-1}) \mapsto \Phi^2_{n,m_{n-1}}(\eta) := \Gamma^2_n(m_{n-1}, \eta) \in \mathcal{P}(E_n) \end{aligned}$$

By construction, using a simple induction on the time parameter n, we find that

$$(m_0, \nu_0) = (\gamma_0(1), \eta_0) \quad \text{and} \quad \forall n \geq 1 \quad m_n = \Phi^1_{n,\nu_{n-1}}(m_{n-1}) \quad \text{and} \quad \nu_n = \Phi^2_{n,m_{n-1}}(\nu_{n-1})$$

$$\Updownarrow$$

$$\forall n \geq 0 \quad (m_n, \nu_n) = (\gamma_n(1), \eta_n)$$

In the cases that are of particular interest, the semigroups $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ will have a Feynman-Kac representation. These models are rather well understood. A brief review on their contraction properties is provided in section 8.4.2. Further details can be found in the monograph [28]. The first basic regularity property of these models which are needed is the following weak Lipschitz type property :

(*Lip*(Φ)) *For any $p \leq n$, $u, u' \in I_p$, $\eta, \eta' \in \mathcal{P}(E_p)$ and $f \in Osc_1(E_n)$ the following Lipschitz inequalities*

$$\left|\Phi^1_{p,n,\nu}(u) - \Phi^1_{p,n,\nu}(u')\right| \leq a^1_{p,n} \, |u - u'| \tag{8.4.2}$$

$$\left|\left[\Phi^2_{p,n,m}(\eta) - \Phi^2_{p,n,m}(\eta')\right](f)\right| \leq a^2_{p,n} \int |[\eta - \eta'](\varphi)| \; \Omega^2_{p,n,\eta'}(f, d\varphi) \tag{8.4.3}$$

for some finite constants $a^i_{p,n} < \infty$, with $i = 1, 2$, and some collection of Markov transitions $\Omega^2_{p,n,\eta'}$ from $Osc_1(E_n)$ into $Osc_1(E_p)$, with $p \leq n$, whose values only depend on the parameters p, n, resp. p, n and η'.

The semigroups $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ may or may not be asymptotically stable depending on whether $a^i_{p,n}$ tends to 0, as $(n - p) \to \infty$. In section 8.4.3 we provide a set of easily checked regularity conditions under which the semigroups associated with the Bernoulli models discussed in 8.3.1 are asymptotically stable.

The second step in the study of the stability properties of the semigroups associated with the flow (8.2.2) is the following continuity property:

(*Cont*(Φ)) *For any $n \geq 1$, $u, u' \in I_{n-1}$, $\eta, \eta' \in \mathcal{P}(E_{n-1})$ and any $f \in Osc_1(E_n)$*

$$\left|\Phi^1_{n,\eta}(u) - \Phi^1_{n,\eta'}(u)\right| \leq \tau^1_n \int |[\eta - \eta'](\varphi)| \; \Omega^1_{n,\eta'}(d\varphi) \tag{8.4.4}$$

$$\left|\left[\Phi^2_{n,u}(\eta) - \Phi^2_{n,u'}(\eta)\right](f)\right| \leq \tau^2_n \, |u - u'| \tag{8.4.5}$$

for some finite constants $\tau^i_n < \infty$, with $i = 1, 2$, and some collection probability measures $\Omega^1_{n,\nu'}$ on $Osc_1(E_{n-1})$, whose values only depend on the parameters n, resp. n and ν'.

This elementary continuity condition allows us to enter the contraction properties of the semigroups $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ in the stability analysis of the flow of measures (8.2.2). The resulting functional contraction inequalities will be described in terms of the following collection of parameters.

Definition 8.4.2. *When the couple of conditions* $(Lip(\Phi))$ *and* $(Cont(\Phi))$ *stated above are satisfied, for any* $i = 1, 2$ *and* $p \leq n$ *we set*

$$\overline{a}^i_{p,n} = \tau^i_{p+1}\ a^i_{p+1,n} \qquad b_{p,n} = \sum_{p<q<n} \overline{a}^1_{p,q}\ \overline{a}^2_{q,n} \quad \text{and} \quad b'_{p,n} = \sum_{p\leq q<n} a^1_{p,q}\ \overline{a}^2_{q,n} \tag{8.4.6}$$

The main result of this section is the following proposition.

Proposition 8.4.3. *If conditions* $(Lip(\Phi))$ *and* $(Cont(\Phi))$ *are satisfied, then for any* $p \leq n$, $u, u' \in I_p$, $\eta, \eta' \in \mathcal{P}(E_p)$ *and* $f \in Osc_1(E_n)$ *we have the following Lipschitz inequalities*

$$\begin{aligned}
\left|\Gamma^1_{p,n}(u',\eta') - \Gamma^1_{p,n}(u,\eta)\right| &\leq c^{1,1}_{p,n}\ |u-u'| + c^{1,2}_{p,n} \int |[\eta-\eta'](\varphi)|\ \Sigma^1_{p,n,u',\eta'}(d\varphi) \\
\left|\Gamma^2_{p,n}(u',\eta')(f) - \Gamma^2_{p,n}(u,\eta)(f)\right] &\leq c^{2,1}_{p,n}\ |u-u'| + c^{2,2}_{p,n} \int |[\eta-\eta'](\varphi)|\ \Sigma^2_{p,n,u'\eta'}(f,d\varphi)
\end{aligned}$$

for some probability measures $\Sigma^1_{p,n,u',\eta'}(d\varphi)$ *and Markov transitions* $\Sigma^2_{p,n,m'\eta'}$, *with the collection of parameters*

$$\begin{aligned}
c^{1,1}_{p,n} &= a^1_{p,n} + \sum_{p\leq q<n} c^{2,1}_{p,q}\ \overline{a}^1_{q,n} \qquad \text{and} \qquad c^{1,2}_{p,n} = \sum_{p\leq q<n} c^{2,2}_{p,q}\ \overline{a}^1_{q,n} \\
c^{2,1}_{p,n} &= b'_{p,n} + \sum_{l=1}^{n-p} \sum_{p\leq r_1<\ldots r_l<n} b'_{p,r_1} \prod_{1\leq k\leq l} b_{r_k,r_{k+1}} \\
c^{2,2}_{p,n} &= a^2_{p,n} + \sum_{l=1}^{n-p} \sum_{p\leq r_1<\ldots r_l<n} a^2_{p,r_1} \prod_{1\leq k\leq l} b_{r_k,r_{k+1}}\ , \quad \text{with the convention } r_{l+1} = n.
\end{aligned}$$

In particular, the collection of parameters $\delta\left(\Sigma^i_{p,n}\right)_{i=1,2}$, $p \leq n$ *introduced in (8.2.16) and (8.3.12) are such that*

$$\delta\left(\Sigma^1_{p,n}\right) \leq c^{1,2}_{p,n} \quad \text{and} \quad \delta\left(\Sigma^2_{p,n}\right) \leq c^{2,2}_{p,n}$$

The proof of this proposition is rather technical and it is postponed to section 8.6 in the appendix. Now we conclude this section with a direct application of the above estimates. The proof of the theorem 8.2.1 stated in the introduction and the uniform estimates discussed in theorem 8.2.2 are a direct consequence of the following lemma.

Lemma 8.4.4. *Suppose that* $\tau^i = \sup_{n\geq 1} \tau^i_n < \infty$, *and* $a^i_{p,n} \leq c_i\ e^{-\lambda_i(n-p)}$, *for any* $p \leq n$, *and some finite parameters* $c_i < \infty$ *and* $\lambda_i > 0$, *with* $i = 1, 2$, *satisfying the following condition*

$$\lambda_1 \neq \lambda_2 \quad \text{and} \quad c_1 c_2\ \tau^1\tau^2 \leq \left(1 - e^{-(\lambda_1\wedge\lambda_2)}\right)\ \left(e^{-(\lambda_1\wedge\lambda_2)} - e^{-(\lambda_1\vee\lambda_2)}\right)$$

Then, for any $i,j \in \{1,2\}$ we have

$$c^{i,j}_{p,n} \leq c^{i,j}\ e^{-\lambda(n-p)} \quad \textit{with} \quad \lambda = (\lambda_1 \wedge \lambda_2) - \log\left(1 + c\tau^1\tau^2 \frac{e^{(\lambda_1\wedge\lambda_2)}}{e^{-(\lambda_1\wedge\lambda_2)} - e^{-(\lambda_1\vee\lambda_2)}}\right) > 0$$

and the parameters $c^{i,j}$ defined below

$$\begin{array}{rclrcl} c^{2,2} &=& c_2 & c^{2,1} &=& c_1c_2\tau^2/\left(e^{-(\lambda_1\wedge\lambda_2)} - e^{-(\lambda_1\vee\lambda_2)}\right) \\ c^{1,1} &=& c_1\ \left(1 + c^{2,1}\tau^1/(e^{-\lambda} - e^{-\lambda_1})\right) & c^{1,2} &=& c_1c_2\tau^1/(e^{-\lambda} - e^{-\lambda_1}) \end{array}$$

In particular, for any N-approximation models $(\gamma^N_n(1), \eta^N_n)$ of the flow $(\gamma_n(1), \eta_n)$ satisfying condition (8.2.6), the $\mathbb{L}_r$-mean error estimates presented in (8.2.16) are uniform w.r.t. the time parameter

$$\sup_{n\geq 0} \mathbb{E}\left(\left|V^{\gamma,N}_n(1)\right|^r\right)^{\frac{1}{r}} \leq a_r\ c^{1,2}/(1-e^{-\lambda}) \quad \textit{and} \quad \sup_{n\geq 0} \mathbb{E}\left(\left|V^{\eta,N}_n(f)\right|^r\right)^{\frac{1}{r}} \leq a_r\ c^{2,2}/(1-e^{-\lambda})$$

with some constants $a_r < \infty$ whose values only depend on r.

Proof. Under the premise of the lemma

$$b_{p,n} \leq c\tau \sum_{p<q<n} e^{-\lambda_1(q-(p+1))}\ e^{-\lambda_2(n-(q+1))} \quad \text{and} \quad b'_{p,n} \leq c\tau^2 \sum_{p\leq q<n} e^{-\lambda_1(q-p)}\ e^{-\lambda_2(n-(q+1))}$$

with $c = c_1c_2$ and $\tau = \tau^1\tau^2$. We further assume that $\lambda_1 > \lambda_2$ and we set $\Delta = |\lambda_1 - \lambda_2|$.

$$b_{p,n} \leq c\tau e^{-\lambda_2((n-1)-(p+1))} \sum_{p<q<n} e^{-\Delta(q-(p+1))} \leq c\tau e^{-\lambda_2((n-1)-(p+1))}/(1-e^{-\Delta})$$

In the same way, if $\lambda_2 > \lambda_1$ we have

$$b_{p,n} \leq c\tau e^{-\lambda_1((n-1)-(p+1))} \sum_{p<q<n} e^{-\Delta(n-(q+1))} \leq c\tau e^{-\lambda_1((n-1)-(p+1))}/(1-e^{-\Delta})$$

This implies that

$$b_{p,n} \leq c\tau e^{-(\lambda_1\wedge\lambda_2)((n-1)-(p+1))}/(1-e^{-\Delta})$$

In much the same way, it can be shown that

$$b'_{p,n} =\leq c\tau^2 e^{-(\lambda_1\wedge\lambda_2)((n-1)-p)}/(1-e^{-\Delta}) \tag{8.4.7}$$

We are now in a position to estimate the parameters $c^{i,j}_{p,n}$. Firstly, we observe that

$$c^{2,2}_{p,n} \leq c_2\ e^{-\lambda_2(n-p)} + c_2 \sum_{l=1}^{n-p} \left(\frac{c\tau^1\tau^2 e^{2(\lambda_1\wedge\lambda_2)}}{1-e^{-\Delta}}\right)^l \sum_{p\leq r_1<\ldots r_l<n} e^{-\lambda_2(r_1-p)} e^{-(\lambda_1\wedge\lambda_2)(n-r_1)}$$

When $\lambda_1 > \lambda_2$, we find that

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda_2(n-p)} \sum_{l=0}^{n-p} \ \left(\frac{c\tau e^{2\lambda_2}}{1-e^{-\Delta}} \right)^l \left(\begin{array}{c} n-p \\ l \end{array} \right)$$

and therefore

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda_2(n-p)} \ \left(1 + c\tau \ \frac{e^{2\lambda_2}}{1-e^{-\Delta}} \right)^{n-p} \Rightarrow c_{p,n}^{2,2} = c_2 \ e^{-\lambda(n-p)}$$

with

$$\lambda = \lambda_2 - \log \left(1 + c\tau \frac{e^{\lambda_2}}{e^{-\lambda_2} - e^{-\lambda_1}} \right) > 0$$

as long as

$$c\tau \leq \left(1 - e^{-\lambda_2} \right) \ \left(e^{-\lambda_2} - e^{-\lambda_1} \right)$$

When $\lambda_2 > \lambda_1$ we have $\lambda_2 = \lambda_1 + \Delta$, we find that

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda_2(n-p)} + c_2 e^{-\lambda_1(n-p)} \sum_{l=1}^{n-p} \ \left(\frac{c\tau e^{2\lambda_1}}{1-e^{-\Delta}} \right)^l \sum_{p \leq r_1 < \ldots r_l < n} e^{-\Delta(r_1 - p)}$$

from which it follows that

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda_1(n-p)} \left(1 + c\tau \frac{e^{2\lambda_1}}{1-e^{-\Delta}} \right)^{n-p}$$

Using a similar line of argument as above, we have

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda(n-p)}$$

with

$$\lambda = \lambda_1 - \log \left(1 + c\tau \frac{e^{\lambda_1}}{e^{-\lambda_1} - e^{-\lambda_2}} \right) > 0$$

as long as

$$c\tau \leq \left(1 - e^{-\lambda_1} \right) \ \left(e^{-\lambda_1} - e^{-\lambda_2} \right)$$

We conclude that

$$c_{p,n}^{2,2} \leq c_2 \ e^{-\lambda(n-p)}$$

with

$$\lambda = (\lambda_1 \wedge \lambda_2) - \log \left(1 + c\tau \frac{e^{(\lambda_1 \wedge \lambda_2)}}{e^{-(\lambda_1 \wedge \lambda_2)} - e^{-(\lambda_1 \vee \lambda_2)}} \right) > 0$$

as long as

$$c\tau \leq \left(1 - e^{-(\lambda_1 \wedge \lambda_2)} \right) \ \left(e^{-(\lambda_1 \wedge \lambda_2)} - e^{-(\lambda_1 \vee \lambda_2)} \right)$$

Using (8.4.7) we also show that

$$c_{p,n}^{2,1} \leq c^{2,1} \ e^{-\lambda(n-p)} \quad \text{with} \quad c^{2,1} = c\tau^2 \ \frac{1}{e^{-(\lambda_1 \wedge \lambda_2)} - e^{-(\lambda_1 \vee \lambda_2)}}$$

Using these estimates

$$c_{p,n}^{1,1} = c_1\ e^{-\lambda_1(n-p)} + \sum_{p\leq q<n} c_{p,q}^{2,1}\ c_1\tau^1\ e^{-\lambda_1(n-(q+1))}$$

and

$$c_{p,n}^{1,1} = c_1\ e^{-\lambda_1(n-p)} + c^{2,1}c_1\tau^1 \sum_{p\leq q<n} e^{-\lambda(q-p)}\ e^{-\lambda_1(n-(q+1))}$$

Since $\lambda_1 > \lambda$ we find that

$$c_{p,n}^{1,1} \leq c_1\ e^{-\lambda_1(n-p)} + c^{2,1}c_1\tau^1\ e^{-\lambda((n-1)-p)}/(1-e^{-\Delta'}) \quad \text{with} \quad \Delta' = \lambda_1 - \lambda > 0$$

This yields

$$c_{p,n}^{1,1} \leq c^{1,1}\ e^{-\lambda(n-p)} \quad \text{with} \quad c^{1,1} := c_1\ \left(1 + c^{2,1}\tau^1/(e^{-\lambda} - e^{-\lambda_1})\right)$$

Finally, we observe that

$$c_{p,n}^{1,2} = c\tau^1 \sum_{p\leq q<n} e^{-\lambda(q-p)}\ e^{-\lambda_1(n-(q+1))} \leq c\tau^1\ e^{-\lambda((n-1)-p)}/(1-e^{-\Delta'})$$

which implies that

$$c_{p,n}^{1,2} \leq c^{1,2}\ e^{-\lambda(n-p)} \quad \text{with} \quad c^{1,2} := c\tau^1/(e^{-\lambda} - e^{-\lambda_1})$$

This ends the proof of the lemma. ∎

8.4.2 Feynman-Kac models

We let $Q_{p,n}$, with $0 \leq p \leq n$, be the Feynman-Kac semi-group associated with a sequence of bounded and positive integral operator Q_n from some measurable spaces $(E_{n-1}, \mathcal{E}_{n-1})$ into $(E_n, \mathcal{E}_n)$. For any $n \geq 1$, we denote by G_{n-1} and M_n the potential function on E_{n-1} and the Markov transition from E_{n-1} into E_n defined below

$$G_{n-1}(x) = Q_n(1)(x) \quad \text{and} \quad M_n(f)(x) = \frac{Q_n(f)(x)}{Q_n(1)(x)}$$

We also denote by $\Phi_{p,n}$, $0 \leq p \leq n$, the nonlinear semigroup from $\mathcal{P}(E_p)$ into $\mathcal{P}(E_n)$ defined below

$$\forall \eta \in \mathcal{P}(E_p),\ \forall f \in \mathcal{B}(E_n) \quad \Phi_{p,n}(\eta)(f) = \eta Q_{p,n}(f)/\eta Q_{p,n}(1) \tag{8.4.8}$$

As usual we use the convention $\Phi_{n,n} = Id$, for $p = n$. It is important to observe that this semigroup is alternatively defined by the formulae

$$\Phi_{p,n}(\eta)(f) = \frac{\eta(G_{p,n}\ P_{p,n}(f))}{\eta(G_{p,n})} \quad \text{with} \quad G_{p,n} = Q_{p,n}(1) \quad \text{and} \quad P_{p,n}(f_n) = Q_{p,n}(f_n)/Q_{p,n}(1)$$

The next two parameters

$$r_{p,n} = \sup_{x,x' \in E_p} \frac{G_{p,n}(x)}{G_{p,n}(x')} \quad \text{and} \quad \beta(P_{p,n}) = \sup_{x_p, y_p \in E_p} |P_{p,n}(x_p, .) - P_{p,n}(y_p, .)|_{tv} \quad (8.4.9)$$

measure respectively the relative oscillations of the potential functions $G_{p,n}$ and the contraction properties of the Markov transition $P_{p,n}$. Various estimates in the forthcoming sections will be expressed in terms of these parameters. For instance and for further use in several places in this article, we have the following Lipschitz regularity property.

Proposition 8.4.5 ([33]). *For any $f_n \in Osc_1(E_n)$ we have*

$$|[\Phi_{p,n}(\eta_p) - \Phi_{p,n}(\mu_p)](f_n)| \leq 2\ r_{p,n}\ \beta(P_{p,n}) \quad \left|[\eta_p - \mu_p]\, \overline{P}^{\mu_p}_{p,n}(f_n)\right| \quad (8.4.10)$$

for some function $\overline{P}^{\mu_p}_{p,n}(f_n) \in Osc_1(E_p)$ that doesn't depends on the measure η_p.

Our next objective is to estimate the the contraction coefficients $r_{p,n}$ and $\beta(P_{p,n})$ in terms of the mixing type properties of the semigroup

$$M_{p,n}(x_p, dx_n) := M_{p+1} M_{p+2} \ldots M_n(x_p, dx_n)$$

associated with the Markov operators M_n. We introduce the following regularity condition.

$(MG)_m$ *There exists an integer $m \geq 1$ and a sequence $(\epsilon_p(M))_{p \geq 0} \in (0,1)^{\mathbb{N}}$ and some finite constant r_p such that for any $p \geq 0$ and any $(x, x') \in E_p^2$ we have*

$$M_{p,p+m}(x_p, .) \geq \epsilon_p(m) \quad M_{p,p+m}(x'_p, .) \quad \text{and} \quad G_p(x) \leq r_p\ G_n(x') \quad (8.4.11)$$

It is well known that the above condition is satisfied for any aperiodic and irreducible Markov chains on finite spaces. Loosely speaking, for non compact spaces this condition is related to the tails of the transition distributions on the boundaries of the state space. For instance, let us suppose that $E_n = \mathbb{R}$ and M_n is the bi-Laplace transition given by

$$M_n(x, dy) = \frac{c(n)}{2}\ e^{-c(n)\,|y - A_n(x)|}\ dy$$

for some $c(n) > 0$ and some drift function A_n with bounded oscillations $\mathrm{osc}(A_n) < \infty$. In this case, it is readily checked that condition $(M)_m$ holds true for $m = 1$ with the parameter $\epsilon_{n-1}(1) = \exp(-c(n)\ \mathrm{osc}(A_n))$.

Under the mixing type condition $(M)_m$ we have for any $n \geq m \geq 1$, and $p \geq 1$

$$r_{p,p+n} \leq \epsilon_p(m)^{-1} \prod_{0 \leq k < m} r_{p+k} \quad (8.4.12)$$

and

$$\beta(P_{p,p+n}) \leq \prod_{k=0}^{\lfloor n/m \rfloor - 1} \left(1 - \epsilon_{p+km}^{(m)}\right) \quad \text{with} \quad \epsilon_p^{(m)} := \epsilon_p^2(m) \prod_{0<k<m} r_{p+k}^{-1} \tag{8.4.13}$$

Notice that these estimates are also valid for any $n \geq 0$. Several contraction inequalities can be deduced from these estimates (see for instance chapter 4 of the book [28]). To give a flavor of these results, we further assume that $(M)_m$ is satisfied with $m = 1$, and we have $\epsilon = \inf_n \epsilon_n(1) > 0$. In this case, we can show that

$$r_{p,p+n} \leq r_p/\epsilon \quad \text{and} \quad \beta(P_{p,p+n}) \leq \left(1 - \epsilon^2\right)^n$$

We end this short section with a direct consequence of proposition 8.4.5.

Corollary 8.4.6. *Consider the Bernoulli semigroup presented in section 8.3.1. For constant mappings $s_n = \mu_{n+1}(1)$, the first component mapping is constant $\Phi^1_{n+1,\nu_n}(u) = s_n$ and the second component mapping $\Phi^2_{n+1,m_n}(\eta) = \Psi_{g_n^{(s)}}(\eta) M^{(s)}_{n+1}$ induces a Feynman-Kac semigroup with the likelihood function $g_n^{(s)}$ and the Markov transitions $M^{(s)}_{n+1}$ defined in (8.3.7). In this situation, the condition (8.4.2) is clearly met with $a^1_{p,n} = 0$, for any $p < n$,. We further assume that the semigroup of associated with the Markov transitions M_n satisfies the mixing property stated in the l.h.s. of (8.4.11) for some integer $m \geq 1$ and some parameter $\epsilon_p(m) \in]0, 1]$. In this situation, the condition (8.4.3) is also met with the collection of parameters $a^2_{p,n}$ given below*

$$a^2_{p,n} \leq 2\ \rho_p(m) \prod_{k=0}^{\lfloor (n-p)/m \rfloor - 1} \left(1 - \epsilon_{p+km}^{(m,s)}\right)$$

with

$$\rho_p(m) := \epsilon_p^{-1}(m) \prod_{p \leq k < p+m} r_k^2(s_k) r_k(1) \ \text{ and } \ \epsilon_p^{(s,m)} = \epsilon_p^2(m) r_p(s_p) / \prod_{p \leq k < p+m} r_k(s_k)^3 r_k(1)^2$$

and the collection of parameters $r_n(s_n)$ defined below

$$r_n(s_n) := \frac{s_n g_n^+ + (1 - s_n)}{s_n g_n^- + (1 - s_n)} \, (\leq r_n(1))$$

8.4.3 Bernoulli models

This section is concerned with the contraction properties of the semigroups $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ associated with the Bernoulli filter discussed in section 8.3.1. Before proceeding, we provide a brief discussion on the oscillations of the likelihood functions g_n given below

$$g_n(x_n) = (1 - d_n(x_n)) + d_n(x_n) \mathcal{Y}_n \left(l_n(x_n, \cdot)/h_n\right)$$

in terms of some $[0,1]$-valued detection probability functions d_n, some local likelihood functions l_n, and some positive clutter intensity function h_n. The oscillations of these likelihood functions strongly depend on the nature of the functions (d_n, h_n, l_n).

Assuming that $h_n^- > 0$ we have

$$(1 - d_n^{\circ,-}) + d_n^{\circ,-} \frac{l_n^-}{h_n^+} \mathcal{Y}_n(1) \leq g_n^- \leq g_n^+ \leq (1 - d_n^{\circ,+}) + d_n^{\circ,+} \frac{l_n^+}{h_n^-} \mathcal{Y}_n(1) \tag{8.4.14}$$

with the parameters

$$d_n^{\circ,+} = d_n^+ \, 1_{l_n^+ \mathcal{Y}_n(1) \geq h_n^-} + d_n^- \, 1_{l_n^+ \mathcal{Y}_n(1) < h_n^-}$$

$$d_n^{\circ,-} = d_n^- \, 1_{l_n^- \mathcal{Y}_n(1) \geq h_n^+} + d_n^+ \, 1_{l_n^- \mathcal{Y}_n(1) < h_n^+}$$

The semigroup contraction inequalities developed in this section will be expressed in terms of the following parameters

$$\delta_n(sg) := \frac{g_n^+ s_n^+}{g_n^- s_n^-}, \quad \delta_n(g) := \frac{g_n^+}{g_n^-} \quad \text{and} \quad \delta_n'(g) := \frac{1}{g_n^-} \wedge g_n^+$$

For time homogeneous models $(d_n, h_n, l_n) = (d, h, l)$, with constant detection probability $d_n(x) = d$ and uniformly bounded number of observations $\sup_n \mathcal{Y}_n(1) \leq \mathcal{Y}^+(1) < \infty$ we have the following estimates

$$(1 - d) \leq g_n^- \leq g_n^+ \leq (1 - d) + d \frac{l^+}{h^-} \mathcal{Y}^+(1)$$

In this situation, we have

$$\delta_n(g) \leq 1 + \frac{d}{1 - d} \frac{l^+}{h^-} \mathcal{Y}^+(1)$$

For small clutter intensity function with $h^- > 0$ and $l^- > 0$ we also have the observation free estimates $\frac{g_n^+}{g_n^-} \leq \frac{l^+ h^+}{l^- h^-}$, from which we find that the upper bound

$$\delta(g) := \sup_{n \geq 0} \delta_n(g) \leq \inf \left\{ 1 + \frac{d}{1 - d} \frac{l^+}{h^-} \mathcal{Y}(1), \frac{l^+ h^+}{l^- h^-} \right\} \tag{8.4.15}$$

and for $d < 1$

$$\delta'(g) := \sup_{n \geq 0} \delta_n'(g) \leq \sup \left\{ (1 - d) + d \frac{l^+}{h^-} \mathcal{Y}(1), \frac{1}{1 - d} \right\} \tag{8.4.16}$$

To be more precise, if we set $\inf_n \mathcal{Y}_n(1) = \mathcal{Y}^-(1)$ then

$$1 \leq \frac{l^-}{h^+} \mathcal{Y}(1)^- \Rightarrow \delta'(g) \leq (1 - d) + d \frac{l^+}{h^-} \mathcal{Y}^+(1)$$

In addition, if we have $d(1-d)\mathcal{Y}(1) \leq h^-/l^+$ and $d<1$ then we find the observation free estimates

$$d\mathcal{Y}(1)\ l^+/h^- \leq 1/(1-d) \Rightarrow \delta'(g) \leq (1-d) + \frac{1}{1-d}$$

Conversely, we have the observation free estimates

$$\frac{l^+}{h^-}\mathcal{Y}(1)^+ \leq 1 \Rightarrow \delta'(g) \leq \frac{1}{(1-d) + d\ \frac{l^-}{h^+}\ \mathcal{Y}^-(1)} \leq \frac{1}{1-d}$$

We are now in position to state the main result of this section.

Theorem 8.4.7. *If $\mu_{n+1}(1) \in]0,1[$, $0 < s_n^- \leq s_n^+ < 1$, and the semigroup $M_{p,n}$ satisfies the condition stated in the l.h.s. of (8.4.11) for some integer $m \geq 1$ and some positive constant $\epsilon_p(m)$, then the condition $(Lip(\Phi))$ is met with*

$$a^1_{p,n} \leq 2\ \epsilon_p^{-1}\delta'_p(g) \prod_{p\leq k<p+n}\left(1-\epsilon_k^2\right) \quad \textit{and} \quad a^2_{p,n} \leq 2\ \rho_p(m) \prod_{k=0}^{\lfloor n/m\rfloor-1}\left(1-\epsilon^{(m)}_{p+km}\right)$$

with some parameters

$$\epsilon_n \geq \inf\left\{\frac{s_n^-}{\mu_{n+1}(1)}, \frac{\mu_{n+1}(1)}{s_n^+}, \frac{1-s_n^+}{1-\mu_{n+1}(1)}, \frac{1-\mu_{n+1}(1)}{1-s_n^-}\right\}$$

and

$$\rho_p(m) \leq \epsilon_p(m)^{-1} \prod_{0\leq k<m}\delta_{p+k}(sg)^3 \quad \textit{and} \quad \epsilon_p^{(m)} \geq \epsilon_p(m)^2\ \delta_p(sg)^{-4} \prod_{0<k<m}\delta_{p+k}(sg)^{-5}$$

In addition condition $(Cont(\Phi))$ is met with

$$\tau^1_{n+1} \leq \delta_n(g)\ \left[(s_n^+ - s_n^-) + |s_n - \mu_{n+1}(1)|\right] \quad \textit{and} \quad \tau^2_{n+1} \leq \delta'_n(g)\ \sup\left\{\frac{\mu_{n+1}}{s_n^-}, \frac{s_n^+}{\mu_{n+1}(1)}\right\}$$

The proof of the theorem is postponed to section 8.6. To give a flavour of these estimates we examine time homogeneous models

$$(d_n, h_n, l_n, s_n, \mu_n) = (d, h, l, s, \mu)$$

with constant detection and survival probabilities $d_n(x) = d$, $s_n(x) = s$, and uniformly bounded number of observations $\sup_n \mathcal{Y}_n(1) \leq \mathcal{Y}(1) < \infty$. In this situation, we have $(\epsilon_p(m), \epsilon_p^{(s)}(m)) = (\epsilon(m), \epsilon^{(s)}(m))$ and using the estimates (8.4.15) we prove the following bounds

$$\tau^1_{n+1} \leq \delta(g)\ |s-\mu(1)| \quad \text{and} \quad \tau^2_{n+1} \leq \delta'(g)\ \frac{\mu(1)\vee s}{\mu(1)\wedge s}$$

and

$$a^1_{0,n} \leq 2\epsilon^{-1}\delta'(g)\left(1-\epsilon^2\right)^n \quad \text{and} \quad a^2_{0,n} \leq 2\epsilon(m)^{-1}\delta(g)^{3m}\ \left(1-\epsilon(m)^2\delta(g)^{-5m+1}\right)^{\lfloor n/m \rfloor}$$

with some parameter ϵ such that

$$\inf\left\{\frac{s}{\mu(1)}, \frac{\mu(1)}{s}, \frac{1-s}{1-\mu(1)}, \frac{1-\mu(1)}{1-s}\right\} \leq \epsilon \leq 1$$

It is also readily verified that the assumptions of lemma 8.4.4 are satisfied with the parameters

$$\begin{array}{rclcrcl} \tau^1 & \leq & \delta(g)\ |s-\mu(1)| & & \tau^2 & \leq & \delta'(g)\ ((\mu(1)\vee s)/(\mu(1)\wedge s)) \\ c_1 & = & 2\epsilon^{-1}\delta'(g) & & c_2 & = & 2\epsilon(m)^{-1}\left(1-\epsilon(m)^2\delta(g)^{-5m+1}\right)^{-1}\delta(g)^{3m} \end{array}$$

and the Lyapunov constants

$$\lambda_1 = -\log\left(1-\epsilon^2\right) \quad \text{and} \quad \lambda_2 = -\frac{1}{m}\log\left(1-\epsilon(m)^2\delta(g)^{-5m+1}\right)$$

We notice that ϵ tends to 1 and τ^1 tends to 0, as $|s-\mu(1)|$ tends to 0. Thus, there exists some $\varsigma \geq 0$ such that

$$\lambda_1 > \lambda_2 \quad \text{and} \quad c_1c_2\tau^1\tau^2 < \left(1-e^{-\lambda_2}\right)\ \left(e^{-\lambda_2}-e^{-\lambda_1}\right)$$

as long as $|s-\mu(1)| \leq \varsigma$. We summarize this discussion with the following corollary.

Corollary 8.4.8. *Consider the time homogeneous model discussed above. Under the assumptions of theorem 8.4.7, for any N-approximation models $(\gamma^N_n(1), \eta^N_n)$ of the Bernoulli model $(\gamma_n(1), \eta_n)$ satisfying condition (8.2.6), the $\mathbb{L}_r$-mean error estimates presented in (8.2.16) are uniform w.r.t. the time parameter*

$$\sup_{n\geq 0}\mathbb{E}\left(\left|V^{\gamma,N}_n(1)\right|^r\right)^{\frac{1}{r}} \leq a_r\ c^{1,2}/(1-e^{-\lambda}) \quad \textit{and} \quad \sup_{n\geq 0}\mathbb{E}\left(\left|V^{\eta,N}_n(f)\right|^r\right)^{\frac{1}{r}} \leq a_r\ c^{2,2}/(1-e^{-\lambda})$$

with the parameters $(c^{1,2}, c^{2,2}, \lambda)$ defined in lemma 8.4.4, and some finite constants $a_r < \infty$ whose values only depend on r.

Remark 8.4.9. *When $\mu_{n+1}(1) = 0$ we have seen in (8.3.8) that*

$$\Phi^1_{n+1,\nu_n}(u) = \Psi_{g_n}(\nu_n)(s_n) \times \theta_{\nu_n(g_n)}(u) \quad \textit{and} \quad \Phi^2_{n+1,m_n}(\eta) = \Psi_{g_ns_n}(\eta)M_{n+1}$$

with the collection of mappings θ_a, with $a \in [0,\infty[$, defined in (8.3.9). Using the fact that

$$\left|\Phi^1_{n+1,\nu_n}(u) - \Phi^1_{n+1,\nu_n}(u')\right| \ = \ \frac{\Psi_{g_n}(\nu_n)(s_n)\ \nu_n(g_n)}{\left[\nu(g_n)u+(1-u)\right]\left[\nu(g_n)u'+(1-u')\right]}\ |u-u'|$$

one proves that (8.4.2) is met with the rather crude upper bound

$$a^1_{p,n} \leq \prod_{p\leq k<n} a_{k,k+1} \quad \textit{and} \quad a^1_{k,k+1} \leq (s^+_k g^+_k)/(1 \wedge g^-_k)^2$$

We also notice that the second component mapping Φ^2_{n+1,m_n} doesn't depends on the parameter m_n, and it induces a Feynman-Kac semigroup of the same form as the one studied in section 8.4.2. Assuming that the mixing condition stated in the l.h.s. of (8.4.11) is satisfied some integer $m \geq 1$ and some parameter $\epsilon_p(m) > 0$, one can prove that (8.4.3) is met with the collection of parameters $a^2_{p,n}$ given below

$$a^2_{p,n} \leq 2\ \rho_p(m) \prod_{k=0}^{\lfloor (n-p)/m \rfloor - 1} \left(1 - \epsilon^{(m)}_{p+km}\right) \quad \textit{with} \quad \rho_p(m) = \epsilon^{-1}_p(m) \prod_{p\leq q<p+m} \delta_q(sg)$$

and the collection of parameters $\epsilon^{(m)}_p = \epsilon^{(m)}_p = \epsilon^2_p(m)/\prod_{p<q<p+m} \delta_q(sg)$.

8.4.4 PHD Models

This section is concerned with the contraction properties of the semigroups $\Phi^1_{p,n,\nu}$ and $\Phi^2_{p,n,m}$ associated with the PHD filter discussed in section 8.2.1 and in section 8.3.2.

The analysis of these nonlinear models is much more involved than the one of the Bernoulli models. We simplify the analysis and we further assume that the clutter intensity function, the detectability rate as well as the survival and the spawning rates introduced in section 8.3.2 are time homogeneous and constants functions, and we set

$$(b_n(x), h_n(x), s_n(x), r_n(x)) = (b, h, s, r)$$

To simplify the presentation, we also assume that the state spaces, the Markov transitions of the targets, the likelihood functions and the spontaneous birth measures are time homogeneous, that is we have that $E_n = E$, $E^Y_n = E^Y$, $M_n = M$, $g_n(x, y) = g(x, y)$ and $\mu_{n+1} = \mu$. Without further mention, we suppose that $r(1 - d) < 1$, $\mu(1) > 0$, $r > 0$, and for any $y \in E^Y$ we have

$$0 \leq g^-(y) := \inf_{x\in E} g(x, y) \leq g^+(y) := \sup_{x\in E} g(x, y) < \infty$$

Given a mapping θ from E^Y into $\mathbb{R}$, we set $\mathcal{Y}^-(\theta) := \inf_n \mathcal{Y}_n(\theta)$ and $\mathcal{Y}^+(\theta) := \sup_n \mathcal{Y}_n(\theta)$.

We recall from (8.2.10) that the PHD filter is defined by the measure-valued equation

$$\gamma_{n+1} = \gamma_n Q_{n+1,\gamma_n}$$

with the integral operator

$$Q_{n+1,\gamma_n}(x_n, dx_{n+1}) = g_{n,\gamma_n}(x_n) M_{n+1}(x_n, dx_{n+1}) + \gamma_n(1)^{-1}\ \mu_{n+1}(dx_{n+1})$$

with the function g_{n,γ_n} defined below

$$g_{n,\gamma_n}(x) = r(1-d) + rd \int \mathcal{Y}_n(dy) \frac{g(x,y)}{h + d\gamma_n(g(\bullet, y))}$$

We also notice that the total mass process and the normalized distribution flow are given by the following equations

$$\begin{aligned}
\gamma_{n+1}(1) &= \Phi^1_{n+1,\eta_n}(\gamma_n(1)) \\
&= \gamma_n(1)\ r(1-d) + \int \mathcal{Y}_n(dy)\ w_{\gamma_n(1)}(\eta_n, y) + \mu(1) \\
\eta_{n+1}(1) &= \Phi^2_{n+1,\gamma_n(1)}(\eta_n) \\
&\propto \gamma_n(1)\ r(1-d)\ \eta_n M + \int \mathcal{Y}_n(dy)\ w_{\gamma_n(1)}(\eta_n, y)\ \Psi_{g(\bullet,y)}(\eta_n) M + \mu(1)\ \overline{\mu}
\end{aligned}$$

with the probability measure $\overline{\mu}$ and weight functions w defined below

$$\overline{\mu}(dx) = \mu(dx)/\mu(1) \quad \text{and} \quad w_u(\eta, y) := r\left(1 - \frac{h}{h + du\eta(g(\bullet, y))}\right)$$

For null clutter parameter $h = 0$, we already observe that the total mass transformation Φ^1_{n+1,η_n} doesn't depend on the flow of probability measures η_n and it is simply given by

$$\Phi^1_{n+1,\eta_n}(\gamma_n(1)) = \gamma_n(1)\ r(1-d) + r\ \mathcal{Y}_n(1) + \mu(1)$$

In this particular situation, we have

$$\gamma^N_n(1) = \gamma_n(1) = (r(1-d))^n \gamma_0(1) + \sum_{0 \le k < n} (r(1-d))^{n-1-k} (r\ \mathcal{Y}_k(1) + \mu(1))$$

Now, we easily show that the pair of conditions (8.4.2) and (8.4.4) are satisfied with the parameters $a^1_{p,n} = (r(1-d))^{n-p}$ and $\tau^1_n = 0$. In more general situations, the total mass process is not explicitly known. Some useful estimates are provided by the following lemma.

Lemma 8.4.10. *We assume that the number of observations is uniformly bounded; that is, we have that $\mathcal{Y}^+(1) < \infty$. In this situation, the total mass process $\gamma_n(1)$ and any approximation model $\gamma^N_n(1)$ given by the recursion (8.2.4) (with the initial condition $\gamma^N_0(1) = \gamma_0(1)$) take values in a sequence of compact sets $I_n \subset [m^-, m^+]$ with*

$$m^- := \frac{\mu(1)}{1 - r(1-d)} \left(1 + rd\ \mathcal{Y}^- \left(\frac{g^-}{h + d\mu(1)g^-}\right)\right) \quad \text{and} \quad m^+ := \gamma_0(1) + \frac{r\mathcal{Y}^+(1) + \mu(1)}{1 - r(1-d)}$$

Proof. Using the fact that $\gamma_n(1) \geq \mu(1)$ we prove that

$$r\left(1 - \frac{h}{h + d\mu(1)\ g^-(y)}\right) \leq w_{\gamma_n(1)}(\eta_n, y) \leq r$$

from which we conclude that

$$\gamma_n(1)\ r(1-d) + r\ \mathcal{Y}_{h,n}(1) + \mu(1) \leq \Phi^1_{n+1,\eta_n}(\gamma_n(1)) \leq \gamma_n(1)\ r(1-d) + r\ \mathcal{Y}_n(1) + \mu(1)$$

with the random measures

$$\mathcal{Y}_{h,n}(dy) := \mathcal{Y}_n(dy)\ \frac{d\mu(1)\ g^-(y)}{h + d\mu(1)\ g^-(y)}$$

For any sequence of probability measures $\nu := (\nu_n)_{n\geq 0} \in \mathcal{P}(E)^{\mathbb{N}}$, and any starting mass $u \in [0, \infty[$ one conclude that

$$(r(1-d))^n\ u + \frac{r\mathcal{Y}^-_h(1) + \mu(1)}{1 - r(1-d)} \leq \Phi^1_{0,n,\nu}(u) \leq (r(1-d))^n\ u + \frac{r\mathcal{Y}^+(1) + \mu(1)}{1 - r(1-d)}$$

This implies that $\gamma_n(1), \gamma^N_n(1) \in I_n \subset [m^-, m^+]$ with

$$m^- := \frac{r\mathcal{Y}^-_h(1) + \mu(1)}{1 - r(1-d)} = \frac{\mu(1)}{1 - r(1-d)}\ \left(1 + rd\ \mathcal{Y}^-\left(\frac{g^-}{h + d\mu(1)g^-}\right)\right)$$

The end of the proof of the lemma is now completed. ∎

We are now in position to state the main result of this section.

Theorem 8.4.11. *We assume that the number of observations is uniformly bounded; that is, we have that $\mathcal{Y}^+(1) < \infty$. In this situation, the condition (Lip(Φ)) is met with the Lipschitz constants $a^i_{p,n} \leq \prod_{p\leq k<n} a^i_{k,k+1}$, with $i = 1, 2$, and the sequence of parameters $\left(a^i_{n,n+1}\right)_{n\geq 0}$, $i = 1, 2$, defined below*

$$a^1_{n,n+1} \leq r(1-d) + rdh\ \mathcal{Y}_n\left(\frac{g^+}{[h + dm^- g^-]^2}\right)$$

and

$$a^2_{n,n+1} \leq m^+ \frac{\beta(M)\left[(1-d) + d\ \mathcal{Y}_n\left(\frac{g^+}{h+dm^+g^+}\ \frac{g^+}{g^-}\right)\right] + hd\mathcal{Y}_n\left(\frac{g^+ - g^-}{(h+dm^-g^-)^2}\right)}{(1-d)\ m^- + dm^-\mathcal{Y}_n\left(\frac{g^-}{h+dm^-g^-}\right) + \mu(1)/r}$$

In addition, condition (Cont(Φ)) is met with the sequence of parameters

$$\tau^1_{n+1} \leq rdhm^+\ \mathcal{Y}_n\left(\frac{g^+ - g^-}{[h + dm^- g^-]^2}\right) \quad \tau^2_{n+1} \leq \frac{(1-d) + hd\ \mathcal{Y}_n\left(\frac{g^+}{(h+dm^-g^-)^2}\right)}{(1-d)\ m^- + dm^-\mathcal{Y}_n\left(\frac{g^-}{h+dm^-g^-}\right) + \mu(1)/r}$$

The proof of theorem 8.4.11 is postponed to section 8.6.

Corollary 8.4.12. *We assume that* $\mathcal{Y}^+(g^+/g^-)$ *and* $\mathcal{Y}^+(g^+/(g^-)^2) < \infty$. *In this situation, there exists some parameters* $0 < \kappa_0 \leq 1$, $\kappa_1 < \infty$, *and* $\kappa_2 > 0$ *such that for any* $d \geq \kappa_0$, $\mu(1) \geq \kappa_1$, *and* $h \leq \kappa_2$, *the semigroups* $\Phi^1_{p,n,\nu}$ *and* $\Phi^2_{p,n,m}$ *satisfy the pair of conditions* $(Lip(\Phi))$ *and* $(Cont(\Phi))$ *with some parameters* $(a^i_{p,n}, \tau^i_n)_{i=1,2,p\leq n}$, *satisfying the assumptions of lemma 8.4.4. In particular, for any* N*-approximation models* $(\gamma^N_n(1), \eta^N_n)$ *of the PHD equation* $(\gamma_n(1), \eta_n)$ *satisfying condition (8.2.6), the* $\mathbb{L}_r$*-mean error estimates presented in (8.2.16) are uniform w.r.t. the time parameter*

$$\sup_{n\geq 0} \mathbb{E}\left(\left|V^{\gamma,N}_n(1)\right|^r\right)^{\frac{1}{r}} \leq a_r\, c^{1,2}/(1-e^{-\lambda}) \quad \textit{and} \quad \sup_{n\geq 0} \mathbb{E}\left(\left|V^{\eta,N}_n(f)\right|^r\right)^{\frac{1}{r}} \leq a_r\, c^{2,2}/(1-e^{-\lambda})$$

with the parameters $(c^{1,2}, c^{2,2}, \lambda)$ *defined in lemma 8.4.4, and some finite constants* $a_r < \infty$ *whose values only depend on* r.

Proof. There is no loss of generality to assume that $r(1-d) < 1/2 \leq d$ and $\mu(1) \geq 1 \geq h$. Recalling that $m^- \geq \mu(1)$, one readily proves that

$$\frac{m^+}{\mu(1)} = \frac{\gamma_0(1)}{\mu(1)} + \frac{1}{1-r(1-d)}\left(1 + \frac{r}{\mu(1)}\,\mathcal{Y}^+(1)\right) \leq 2 + \gamma_0(1) + 2r\mathcal{Y}^+(1) := \rho$$

If we set $\delta(g) := \rho \vee \mathcal{Y}^+\left(\frac{g^+}{g^-}\right) \vee \mathcal{Y}^+\left(\frac{g^+}{(g^-)^2}\right)$, then we find the rather crude estimates

$$a^1_{n,n+1}/r \leq (1-d) + \frac{2h}{\mu(1)^2}\,\delta(g) \quad \text{and} \quad a^2_{n,n+1}/r \leq \left[\beta(M)(1-d) + \frac{2h+\beta(M)}{\mu(1)}\right]\delta(g)$$

as well as

$$\tau^1_{n+1}/r \leq \frac{2h}{\mu(1)}\,\delta(g)^2 \quad \text{and} \quad \tau^2_{n+1}/r \leq \frac{1}{\mu(1)}\left[(1-d) + \frac{2h}{\mu(1)^2}\,\delta(g)\right]$$

from which we find that

$$\tau^1\tau^2 \leq \frac{2hr^2}{\mu(1)^2}\left[(1-d) + \frac{2h}{\mu(1)^2}\,\delta(g)\right]\,\delta(g)^2 \tag{8.4.17}$$

Thus, there exists some $0 < \kappa_0 \leq 1$ and some $\kappa_1 < \infty$ so that for any $d \geq \kappa_0$ and any $\mu(1) \geq \kappa_1$ we have

$$\begin{aligned} a^1_{n,n+1} &\leq r\left[(1-d) + \frac{2}{\mu(1)^2}\right]\,\delta(g) := e^{-\lambda_1} < 1 \\ a^2_{n,n+1} &\leq r\left[(1-d) + \frac{3}{\mu(1)}\right]\delta(g) := e^{-\lambda_2} < 1 \quad \text{with} \quad 0 < \lambda_2 < \lambda_1 \end{aligned}$$

Finally, using (8.4.17) we find some $\kappa_2 > 0$ such that for any $h \leq \kappa_2$, we have that $\tau^1\tau^2 \leq \left(1-e^{-\lambda_2}\right)\left(e^{-\lambda_2} - e^{-\lambda_1}\right)$. The end of the proof is now a direct consequence of lemma 8.4.4. This ends the proof of the corollary. ∎

8.5 Stochastic particle approximations

8.5.1 Mean field interacting particle systems

Description of the models

The mean field type interacting particle system associated with the equation (8.2.2) relies on the fact that the one step mappings Γ_{n+1}^2 can be rewritten in the following form

$$\Gamma_{n+1}^2(\gamma_n(1),\eta_n) = \eta_n K_{n+1,\gamma_n} \quad \text{with} \quad \gamma_n = \gamma_n(1) \times \eta_n \tag{8.5.1}$$

for some collection of Markov kernels $K_{n+1,\gamma}$ indexed by the time parameter n and the set of measures $\gamma \in \mathcal{M}_+(E_n)$. We mention that the choice of the Markov transitions $K_{n,\gamma}$ is not unique. In the literature on mean field particle models, $K_{n,\gamma}$ are called a choice of McKean transitions. Some McKean interpretation models of the Bernoulli and the PHD filter models (8.2.8) and (8.2.10) are discussed in section 8.3.2 (see for instance (8.3.10)) and in section 8.3.1 (see for instance 8.3.2)

These models provide a natural interpretation of the distribution laws η_n as the laws of a non linear Markov chain $\overline{X}_n$ whose elementary transitions $\overline{X}_n \rightsquigarrow \overline{X}_{n+1}$ depends on the distribution $\eta_n = \mathrm{Law}(\overline{X}_n)$, as well as on the current mass process $\gamma_n(1)$. In contrast to traditional McKean model, the dependency on the mass process induce a dependency of all the flow of measures η_p, for $0 \leq p \leq n$. For a thorough description of these discrete generation and non linear McKean type models, we refer the reader to [28].

In further developments of the article, we always assume that the mappings

$$\left(m, x_n, (x^i)_{1\leq i\leq N}\right) \mapsto K_{n+1,m\sum_{j=1}^N \delta_{x^j}}(x_n, A_{n+1}) \text{ and } G_{n+1,m\sum_{j=1}^N \delta_{x^j}}(x_n)$$

are pointwise known, and of course measurable w.r.t. the corresponding product sigma fields, for any $n \geq 0$, $N \geq 1$, $A_{n+1} \in \mathcal{E}_{n+1}$, and any $x_n \in E_n$. In this situation, the mean field particle interpretation of this nonlinear measure-valued model is an E_n^N-valued Markov chain $\xi_n^{(N)} = \left(\xi_n^{(N,i)}\right)_{1\leq i\leq N}$, with elementary transitions defined as

$$\gamma_{n+1}^N(1) = \gamma_n^N(1)\ \eta_n^N(G_{n,\gamma_n^N}) \tag{8.5.2}$$

$$\mathbb{P}\left(\xi_{n+1}^{(N)} \in dx \ \middle|\ \mathcal{F}_n^{(N)}\right) = \prod_{i=1}^N K_{n+1,\gamma_n^N}(\xi_n^{(N,i)}, dx^i) \tag{8.5.3}$$

with the pair of occupation measures $\left(\gamma_n^N, \eta_n^N\right)$ defined below

$$\eta_n^N := \frac{1}{N}\sum_{j=1}^N \delta_{\xi_n^{(N,j)}} \quad \text{and} \quad \gamma_n^N(dx) := \gamma_n^N(1)\ \eta_n^N(dx)$$

In the above displayed formula, $\mathcal{F}_n^N$ stands for the σ-field generated by the random sequence $(\xi_p^{(N)})_{0\leq p\leq n}$, and $dx = dx^1 \times \ldots \times dx^N$ stands for an infinitesimal neighborhood of a point $x = (x^1, \ldots, x^N) \in E_n^N$. The initial system $\xi_0^{(N)}$ consists of N independent and identically distributed random variables with common law η_0. As usual, to simplify the presentation, when there is no possible confusion we suppress the parameter N, so that we write ξ_n and ξ_n^i instead of $\xi_n^{(N)}$ and $\xi_n^{(N,i)}$.

Convergence analysis

The rationale behind the mean field particle model described in (8.5.3) is that η_{n+1}^N is the empirical measure associated with N independent variables with distributions $K_{n+1,\gamma_n^N}(\xi_n^i, dx)$, so as long as γ_n^N is a good approximation of γ_n then η_{n+1}^N should be a good approximation of η_{n+1}. Roughly speaking, this induction argument shows that η_n^N tends to η_n, as the population size N tends to infinity.

These stochastic particle algorithms can be thought of in various ways: From the physical view point, they can be seen as microscopic particle interpretations of physical nonlinear measure-valued equations. From the pure mathematical point of view, they can also be interpreted as natural stochastic linearizations of nonlinear evolution semigroups. From the probabilistic point of view, they can be interpreted as a interacting recycling acceptance-rejection sampling techniques. In this case, they can be seen as a sequential and interacting importance sampling technique.

By construction, the local fluctuation random fields $(W_n^N)_{n\geq 0}$ defined in (8.2.5) can be rewritten as follows

$$\eta_n^N = \eta_{n-1}^N K_{n,\gamma_{n-1}^N} + \frac{1}{\sqrt{N}}\, W_n^N$$

Using Khintchine's inequality, we can check that (8.2.6) is met for any $r \geq 1$ and any $f_n \in \mathrm{Osc}_1(E_n)$, with the collection of universal constants given below

$$a_{2r}^{2r} \leq (2r)!\; 2^{-r}/r! \quad \text{and} \quad a_{2r+1}^{2r+1} \leq (2r+1)!\; 2^{-r}/r!$$

We end this section with a brief discussion on the PHD equation presented in (8.2.10). This model combines in a single step the traditional updating and a prediction filtering transition. This combination allows us to reduce the fluctuations of the local sampling errors and their propagations w.r.t. the time parameter. Since these updating-prediction models are often used in the literature of multiple target tracking, we provide below a short summary. If we set

$$\widehat{g}_{n,\gamma}^c(\cdot, y) = \begin{cases} (1-d_n) & \text{if} \quad y = c \\ \dfrac{d_n g_n(\cdot, y_n)}{h_n(y) + \gamma(d_n g_n(\cdot, y))} & \text{if} \quad y \neq c \end{cases}$$

then

$$\gamma_{n+1} = \widehat{\gamma}_n Q_{n+1} + \mu_{n+1} \quad \text{with} \quad Q_{n+1}(f) := r_n\; M_{n+1}(f)$$

with the updated measures defined below

$$\widehat{\gamma}_n(f) := \gamma_n(\widetilde{g}^c_{n,\gamma_n} f) \quad \text{with} \quad \widetilde{g}^c_{n,\gamma_n} = \int \mathcal{Y}^c_n(dy)\ \widehat{g}^c_{n,\gamma_n}(\cdot, y)$$

Notice that

$$\widehat{\gamma}_n(1) = \gamma_n(\widetilde{g}^c_{n,\gamma_n} f) \quad \text{and} \quad \widehat{\eta}_n(dx) := \widehat{\gamma}_n(dx)/\widehat{\gamma}_n(1) = \Psi_{\widetilde{g}^c_{\widehat{\gamma}_n,n}}(\eta_n)(dx)$$

from which we find the recursive formulae:

$$\begin{pmatrix} \gamma_n(1) \\ \eta_n \end{pmatrix} \xrightarrow{\text{updating}} \begin{pmatrix} \widehat{\gamma}_n(1) \\ \widehat{\eta}_n \end{pmatrix} \xrightarrow{\text{prediction}} \begin{pmatrix} \gamma_{n+1}(1) \\ \eta_{n+1} \end{pmatrix}$$

with the prediction transition described below

$$\gamma_{n+1}(1) = \widehat{\gamma}_n(r_n) + \mu_{n+1}(1) \quad \text{and} \quad \eta_{n+1} = \Psi_{r_n}(\widehat{\eta}_n)\, M'_{n+1,\widehat{\gamma}_n}$$

In the above displayed formula, $M'_{n+1,\widehat{\gamma}_n}$ is the Markov transition defined by

$$M'_{n+1,\widehat{\gamma}_n}(x, \cdot) = \alpha'_n(\widehat{\gamma}_n)\ M_{n+1}(x, \cdot) + (1 - \alpha'_n(\widehat{\gamma}_n))\ \overline{\mu}_{n+1}$$

with the collection of $[0,1]$-valued parameters $\alpha'_n(\widehat{\gamma}_n) = \widehat{\gamma}_n(r_n)/(\widehat{\gamma}_n(r_n) + \mu_{n+1}(1))$. It should be clear that the updating and the prediction transitions can be approximated using a genetic type selection and mutation transition. Each of these sampling transitions introduces a separate local sampling fluctuation error. The stochastic analysis of the corresponding mean field particle interpretations can be developed using the same line of arguments as those used for the particle model discussed above.

8.5.2 Interacting particle association systems

Description of the models

We let $(\mathcal{A}_n)_{n\geq 0}$ be a sequence of finite sets equipped with some finite positive measures $(\nu_n)_{n\geq 0}$. We further assume that the initial distribution γ_0 and the integral operators Q_{n+1,γ_n} in (8.2.1) have the following form

$$\gamma_0 = \int \nu_0(da)\ \eta_0^{(a)} \quad \text{and} \quad Q_{n+1,\gamma_n} = \int \nu_{n+1}(da)\ Q^{(a)}_{n+1,\gamma_n}$$

In the above display $\eta_0^{(a)}$ stands for a collection of measures on E_0, indexed by the parameter $a \in \mathcal{A}_0$, and $Q^{(a)}_{n+1,\gamma_n}$ is a collection of integral operators indexed by the parameter $a \in \mathcal{A}_{n+1}$. In this situation, we observe that

$$\gamma_0(1) = \nu_0(1) \quad \text{and} \quad \eta_0 = \int A_0(da)\ \eta_0^{(a)} \quad \text{with} \quad A_0(da) := \nu_0(da)/\nu_0(1)$$

We also assume that the following property is met

$$G_{n,\gamma}^{(a)} := Q_{n+1,\gamma}^{(a)}(1) \propto G_n^{(a)} \quad \text{and} \quad Q_{n+1,\gamma}^{(a)}(f)/Q_{n+1,\gamma}^{(a)}(1) := M_{n+1}^{(a)}(f) \tag{8.5.4}$$

for some function $G_n^{(a)}$ on E_n, and some Markov transitions $M_{n+1}^{(a)}$ from E_n into E_{n+1} whose values do not depend on the measures γ. For clarity of presentation, sometimes we write $\Psi_{G_n}^{(a)}$ instead of $\Psi_{G_n^{(a)}}$.

Definition 8.5.1. *We consider the collection of probability measures $\eta_n^{(\mathrm{a}_n)} \in \mathcal{P}(E_n)$, indexed by sequences of parameters*

$$\mathrm{a}_n = (a_0, \ldots, a_n) \in \mathcal{A}_{[0,n]} := (\mathcal{A}_0 \times \ldots \times \mathcal{A}_n)$$

and defined by the following equations

$$\eta_n^{(\mathrm{a}_n)} \quad = \quad \left(\Phi_n^{(a_n)} \circ \ldots \circ \Phi_1^{(a_1)}\right)\left(\eta_0^{(a_0)}\right) \tag{8.5.5}$$

with the mappings $\Phi_n^{(a)} : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$ indexed by $a \in \mathcal{A}_n$ and defined by the updating-prediction transformation

$$\Phi_n^{(a)}(\eta) = \Psi_{G_{n-1}}^{(a)}(\eta)\, M_n^{(a)}$$

We illustrate these abstract conditions in the context of the multiple target tracking equation presented in (8.2.10). In this situation, it is convenient to add a pair of virtual observation states c, c' to E_n^Y. Using this notation, the above conditions are satisfied with the finite sets $\mathcal{A}_{n+1}$ and their counting measures ν_{n+1} defined below

$$\mathcal{A}_{n+1} = \left\{Y_n^i, 1 \leq i \leq N_n^Y\right\} \cup \{c, c'\} \qquad \nu_{n+1} = \mathcal{Y}_n + \delta_c + \delta_{c'} \in \mathcal{M}(\mathcal{A}_{n+1})$$

Using (8.2.10) and (8.2.12), we check that (8.5.4) is met with the couple of potential functions and Markov transitions defined by

$$(G_n^{(y_n)}, M_{n+1}^{(y_n)}) = \begin{cases} (r_n d_n g_n(\cdot, y_n), M_{n+1}) & \text{for} \quad y_n \notin \{c, c'\} \\ (r_n(1-d_n), M_{n+1}) & \text{for} \quad y_n = c \\ \left(1, \overline{\mu}_{n+1}\right) & \text{for} \quad y_n = c' \end{cases}$$

In this case, we observe that

$$Q_{n+1,\gamma_n}^{(y_n)}(x_n, \cdot) = G_{n,\gamma_n}^{(y_n)}(x_n)\; M_{n+1}^{(y_n)}(x_n, \cdot)$$

with the potential function $G_{n,\gamma_n}^{(y_n)}$ defined below

$$G_{n,\gamma_n}^{(y_n)}/G_n^{(y_n)} = \begin{cases} [h_n(y_n) + \gamma_n(d_n g_n(\cdot, y_n))]^{-1} & \text{for} \quad y_n \notin \{c, c'\} \\ 1 & \text{for} \quad y_n = c \\ \mu_{n+1}(1)/\gamma_n(1) & \text{for} \quad y_n = c' \end{cases} \tag{8.5.6}$$

Under our assumptions, using (8.2.2), we have the following result.

Proposition 8.5.2. *The solution the equation (8.2.2) has the following form*

$$\eta_n = \int A_n(da)\ \eta_n^{(a)}$$

with a total mass process $\gamma_n(1)$ *and the association measures* $A_n \in \mathcal{P}(\mathcal{A}_{[0,n]})$ *defined by the following recursive equations*

$$\gamma_{n+1}(1) = \gamma_n(1)\ \eta_n(G_{n,\gamma_n}) \quad \textit{and} \quad A_{n+1} = \Omega_{n+1}\left(\gamma_n(1), A_n\right)$$

With the mapping

$$\Omega_{n+1}\ :\ (m,A) \in \left(]0,\infty[\times\mathcal{P}(\mathcal{A}_{[0,n]})\right) \mapsto \Omega_{n+1}(m,A) \in \mathcal{P}(\mathcal{A}_{[0,n+1]})$$

defined by the following formula

$$\Omega_{n+1}\left(m,A\right)\left(d(a,b)\right) \propto A(da)\ \nu_{n+1}(db)\ \eta_n^{(a)}\left(G^{(b)}_{n,m\int A(da)\ \eta_n^{(a)}}\right) \tag{8.5.7}$$

Proof:
The proof of the above assertion is simply based on the fact that

$$\begin{aligned}\eta_{n+1} \propto \int \nu_{n+1}(db)\ \eta_n Q^{(b)}_{n+1,\gamma_n} &= \int A_n(da)\ \nu_{n+1}(db)\ \eta_n^{(a)} Q^{(b)}_{n+1,\gamma_n} \\ &= \int A_n(da)\ \nu_{n+1}(db)\ \eta_n^{(a)}\left(G^{(b)}_{n,\gamma_n}\right)\ \eta_{n+1}^{(a,b)}\end{aligned}$$

This clearly implies that

$$\Gamma_n^2\left(m, \int A(da)\ \eta_{n-1}^{(a)}\right) = \int \Omega_n\left(m,A\right)\left(d(a,b)\right)\ \eta_n^{(a,b)}$$

This ends the proof of the proposition. ∎

By construction, we notice that for any discrete measure $A \in \mathcal{P}(\mathcal{A}_{[0,n-1]})$, and any collection of measures $\eta^{(a)} \in \mathcal{P}(E_{n-1})$, with $a \in \mathcal{A}_{[0,n-1]}$ we have the formula

$$\Gamma_n^2\left(m, \int A(da)\ \eta^{(a)}\right) = \int \Omega_n\left(m,A\right)\left(d(a,b)\right)\ \ \Phi_n^{(b)}\left(\eta^{(a)}\right)$$

Particle approximation models

To get some feasible solution, we further assume that $\eta_n^{(a)}\left(G^{(b)}_{n,\gamma_n}\right)$ are explicitly known for any sequence of parameters $(a,b) \in \left(\mathcal{A}_{[0,n]} \times \mathcal{A}_{n+1}\right)$. This rather strong condition is satisfied for the multiple target tracking model discussed above as long as the quantities

$$\eta_n^{(a_0,y_0,\ldots,y_{n-1})}(r_n d_n g_n(\boldsymbol{.}, y_n)) \quad \eta_n^{(a_0,y_0,\ldots,y_{n-1})}\left(r_n(1-d_n)\right) \quad \eta_n^{(a_0,y_0,\ldots,y_{n-1})}(d_n g_n(\boldsymbol{.}, y_n))$$

are explicitly known. This condition is clearly met for linear gaussian target evolution and observation sensors as long as the survival and detection probabilities s_n and d_n are state independent, and spontaneous birth $\overline{\mu}_n$ and spawned targets branching rates b_n are Gaussian mixtures. In this situation, the collection of measures $\eta_n^{(a_0,y_0,\ldots,y_{n-1})}$ are gaussian distributions and the equation (8.5.5) coincides with the traditional updating-prediction transitions of the discrete generation Kalman-Bucy filter.

We let $A_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_0^i}$, be the empirical measure associated with N independent and identically distributed random variables $(\mathrm{a}_0^i)_{1\leq i\leq N}$ with common distribution A_0. By construction, we have

$$\eta_0^N := \int A_0^N(da)\ \eta_0^{(a)} = \eta_0 + \frac{1}{\sqrt{N}}\ W_0^N$$

with some local sampling random fields satisfying (8.2.6). We further assume that $\gamma_0(1)$ is known and we set $\gamma_0^N = \gamma_0(1)\ \eta_0^N$.

$$\gamma_1^N(1) = \gamma_0^N(1)\ \eta_0^N(G_{0,\gamma_0^N}) \quad \text{and} \quad \eta_1^N := \int A_1^N(da)\ \eta_1^{(a)}$$

with the occupation measure $A_1^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_1^i}$ associated with N conditionally independent and identically distributed random variables $\mathrm{a}_1^i := (a_{0,1}^i, a_{1,1}^i)$ with common law $\Omega_1\left(\gamma_0^N(1), A_0^N\right)$. By construction, we also have

$$\eta_1^N := \int \Omega_1\left(\gamma_0^N(1), A_0^N\right)(da)\ \eta_1^{(a)} + \frac{1}{\sqrt{N}}\ W_1^N = \Gamma_1^2\left(\gamma_0^N(1), \eta_0^N\right) + \frac{1}{\sqrt{N}}\ W_1^N$$

with some local sampling random fields satisfying (8.2.6). Iterating this procedure, we define by induction a sequence of N-particle approximation measures

$$\gamma_n^N(1) = \gamma_{n-1}^N(1)\ \eta_{n-1}^N(G_{n-1,\gamma_{n-1}^N}) \quad \text{and} \quad \eta_n^N := \int A_n^N(da)\ \eta_n^{(a)}$$

with the occupation measure $A_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_n^i}$ associated with N conditionally independent and identically distributed random variables $\mathrm{a}_n^i := (a_{0,n}^i, a_{1,n}^i, \ldots, a_{n,n}^i)$ with common law $\Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N\right)$. Arguing as above, we find that

$$\eta_n^N = \int \Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N\right)(da)\ \eta_n^{(a)} + \frac{1}{\sqrt{N}}\ W_n^N = \Gamma_n^2\left(\gamma_{n-1}^N(1), \eta_{n-1}^N\right) + \frac{1}{\sqrt{N}}\ W_n^N$$

with some local sampling random fields satisfying (8.2.6).

Convergence analysis

The main objective of this section is to show that N-particle occupation measures A_n^N converge in a sense to be given, as N tends to ∞, to the association probability

measures A_n. To this end we observe that the one step mapping Ω_{n+1} introduced in (8.5.7) can be rewritten in the following form

$$\Omega_{n+1}\,(m,A)\,(F) = \frac{A\mathcal{Q}_{n+1,mA}(F)}{A\mathcal{Q}_{n+1,mA}(1)}$$

with the collection of integral operators $\mathcal{Q}_{n+1,mA}$ from $\mathcal{A}_{[0,n]}$ into $\mathcal{A}_{[0,n+1]}$ defined below

$$\mathcal{Q}_{n+1,B}(a,d(a',b)) := \delta_a(da')\ \nu_{n+1}(db)\ \eta_n^{(a')}\left(\mathcal{G}_{n,B}^{(b)}\right) \quad \text{where} \quad \mathcal{G}_{n,B}^{(b)} := G^{(b)}_{n,\int B(da)\ \eta_n^{(a)}}$$

with $B = mA$. In the above display $d(a',b) = da' \times db$ stands for an infinitesimal neighborhood of the point $(a',b) \in \mathcal{A}_{[0,n+1]}$, with $a = (a'_0,\ldots,a'_n) \in \mathcal{A}_{[0,n]}$ and $b \in \mathcal{A}_{n+1}$, and $a = (a_0,\ldots,a_n) \in \mathcal{A}_{[0,n]}$. It is important to point out that

$$B_n := \gamma_n(1) \times A_n \Longrightarrow B_{n+1} = B_n\mathcal{Q}_{n+1,B_n}$$

Notice that the flow of measures $(B_n)_{n\geq 0}$ satisfies the same type of equation as in (8.2.1), with the a total mass evolution of the same form as (8.2.3):

$$B_{n+1}(1) = B_n(1)\ A_n\left(\mathcal{G}_{n,B_n}\right) \quad \text{with} \quad \mathcal{G}_{n,mA} := \int \nu_{n+1}(db)\ \mathcal{G}_{n,mA}^{(b)}$$

$$\mathcal{Q}_{n+1,B_n}(F)(a) = \int \nu_{n+1}(db)\ \eta_n^{(a)}\left(\mathcal{G}_{n,B_n}^{(b)}\right)\ F(a,b)$$

$$\left[\mathcal{Q}_{n+1,B}(F) - \mathcal{Q}_{n+1,B'}(F)\right](a) = \int \nu_{n+1}(db) \quad \left[\eta_n^{(a)}\left(\mathcal{G}_{n,B}^{(b)}\right) - \eta_n^{(a)}\left(\mathcal{G}_{n,B'}^{(b)}\right)\right]\ F(a,b)$$

If we set $B = mA$ and $B' = m'A'$ then condition (H'_2) is met as long as

$$\left|\eta_n^{(a)}\left(\mathcal{G}_{n,B}^{(b)}\right) - \eta_n^{(a)}\left(\mathcal{G}_{n,B'}^{(b)}\right)\right| \leq c(n)\ |m-m'| + \int |[A-A'](\varphi)|\ \Sigma_{n,B'}^{(b)}(d\varphi)$$

for some collection of bounded measures $\Sigma_{n,B'}^{(b)}$ on $\mathcal{B}(\mathcal{A}_n)$ such that $\int\ \mathrm{osc}(\varphi)\ \Sigma_{n,B'}^{(b)} \leq \delta\left(\Sigma_n^{(b)}\right)$, for some finite constant $\delta\left(\Sigma_n^{(b)}\right) < \infty$, whose values do dot depend on the parameters $(m,A) \in (I_n \times \mathcal{P}(\mathcal{A}_n))$. Under the assumptions (8.5.4), we have

$$\mathcal{G}_{n,B}^{(b)}(x) = \alpha_n^{(b)}(B)\ G_n^{(b)}(x)$$

for some collection of parameters $\alpha_n^{(b)}(B)$ satisfying

$$\left|\alpha_n^{(b)}(B) - \alpha_n^{(b)}(B')\right| \leq c(n)\ |m-m'| + \int |[A-A'](\varphi)|\ \Sigma_{n,B'}^{(b)}(d\varphi)$$

This condition is clearly satisfied for the PHD model discussed in (8.5.6), as long as the functions $h_n(y_n) + d_n g_n(\boldsymbol{.}, y_n)$ are uniformly bounded from above and below.

For instance, for $b = y_n \notin \{c, c'\}$ we have

$$\alpha_n^{(b)}(B) = \left[h_n(b) + \int B(da)\ \eta_n^{(a)}(d_n g_n(\bullet, b)) \right]^{-1}$$

In this case, we can check that

$$\left| \alpha_n^{(b)}(B) - \alpha_n^{(b)}(B') \right| \leq c(n) \left| [B - B'](\varphi_n^{(b)}) \right| \quad \text{with} \quad \varphi_n^{(b)}(a) := \eta_n^{(a)}(d_n g_n(\bullet, b))$$

In the same way, we show that the condition (H_1) is also met for the PHD model. This, by construction of A_n^N we find that

$$A_n^N = \Omega_n \left(\gamma_{n-1}^N(1), A_{n-1}^N \right) + \frac{1}{\sqrt{N}}\ \mathcal{W}_n^N$$

with some local sampling random fields satisfying (8.2.6). Notice that

$$\Omega_{n+1}(m, A) = \Psi_{\mathcal{H}_{n,mA}}(A)\, \mathcal{M}_{n+1,mA}(a, d(a', b))$$

with the collection of potential functions

$$\mathcal{H}_{n,mA}(a) \ := \ \mathcal{Q}_{n+1,mA}(1)(a) = \eta_n^{(a)}(\mathcal{G}_{n,mA})$$

and the Markov transitions

$$\mathcal{M}_{n+1,mA}(a, d(a', b)) \ := \ \frac{\mathcal{Q}_{n+1,mA}(a, d(a', b))}{\mathcal{Q}_{n+1,mA}(1)(a)} = \delta_a(da')\ \frac{\nu_{n+1}(db)\ \eta_n^{(a')}\left(\mathcal{G}_{n,mA}^{(b)}\right)}{\int \nu_{n+1}(db')\ \eta_n^{(a')}\left(\mathcal{G}_{n,mA}^{(b')}\right)}$$

8.5.3 Mixed particle association models

We consider the association mapping

$$\Omega_{n+1} \ : \ (m, A, \eta) \in \left( ]0, \infty[\times \mathcal{A}_{[0,n]} \times \mathcal{P}(E_n)^{\mathcal{A}_{[0,n]}} \right) \mapsto \Omega_{n+1}(m, A, \eta) \in \mathcal{P}(\mathcal{A}_{[0,n+1]})$$

defined for any $(m, A) \in \left( ]0, \infty[\times \mathcal{A}_{[0,n]} \right)$ and any mapping $\eta : a \in \mathrm{Supp}(A) \mapsto \eta^{(a)} \in Pa(E_n)$ by

$$\Omega_{n+1}(m, A, \eta)(d(a, b)) \propto A(da)\ \nu_{n+1}(db)\ \eta^{(a)}\left(G_{n,m \int A(da)\ \eta^{(a)}}^{(b)} \right)$$

By construction, for any discrete measure $A \in \mathcal{P}(\mathcal{A}_{[0,n-1]})$, and any mapping $a \in \mathrm{Supp}(A) \mapsto \eta^{(a)} \in \mathcal{P}(E_{n-1})$, we have the formula

$$\Gamma_n^2\left(m, \int A(da)\ \eta^{(a)} \right) = \int \Omega_n\left(m, A, \eta^{(\bullet)} \right)(d(a, b))\ \ \Phi_n^{(b)}\left(\eta^{(a)} \right)$$

We also mention that the updating-prediction transformation defined in (8.5.5)

$$\Phi_n^{(a)}(\eta) = \Psi_{G_{n-1}}^{(a)}(\eta)\, M_n^{(a)} = \eta \mathcal{K}_{n,\eta}^{(a)} \quad \text{with} \quad \mathcal{K}_{n,\eta}^{(a)} = \mathcal{S}_{n-1,\eta}^{(a)} M_n^{(a)} \tag{8.5.8}$$

In the above displayed formula $\mathcal{S}_{n,\eta}^{(a)}$ stands for some updating Markov transition from E_{n-1} into itself satisfying the compatibility condition $\eta\mathcal{S}_{n-1,\eta}^{(a)} = \Psi_{G_{n-1}}^{(a)}(\eta)$.

We let $A_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_0^i}$, be the empirical measure associated with N independent and identically distributed random variables $(\mathrm{a}_0^i)_{1\leq i\leq N}$ with common distribution A_0. For any $\mathrm{a}\in\mathcal{A}_0$, we let

$$\eta_0^N := \int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a},N')} \quad \text{and} \quad \eta_0^{(\mathrm{a},N')} = \frac{1}{N'}\sum_{i=1}^{N'} \delta_{\xi_0^{[\mathrm{a},j]}}$$

with the empirical measure $\eta_0^{(\mathrm{a},N')}$ associated with N' random variables $\xi_0^{[\mathrm{a}]} = \left(\xi_0^{[\mathrm{a},j]}\right)_{1\leq j\leq N'}$ with common law $\eta_0^{(\mathrm{a})}$. We further assume that $\gamma_0(1)$ is known and set

$$\gamma_0^N := \gamma_0(1)\ \eta_0^N \quad \text{and} \quad \gamma_1^N(1) := \gamma_0^N(1)\ \eta_0^N(G_{0,\gamma_0^N})$$

It is readily checked that the fluctuation random fields given below

$$\mathcal{W}_0^{(a,N')} = \sqrt{N'}\ \left(\eta_0^{(a,N')} - \eta_0^{(a)}\right)$$

satisfies (8.2.6), with $N = N'$, for any given $a\in\mathcal{A}_0$. Using the fact that

$$\int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a},N')} = \int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a})} + \frac{1}{\sqrt{N'}}\ \int A_0^N(d\mathrm{a})\ \mathcal{W}_0^{(a,N')}$$

we conclude that

$$\eta_0^N := \eta_0 + \frac{1}{\sqrt{N}}\ W_0^N$$

with some local sampling random fields W_0^N satisfying the same estimates as in (8.2.6) by replacing $1/\sqrt{N}$ by the sum $\left(1/\sqrt{N} + 1/\sqrt{N'}\right)$.

Using (8.5.8), for any $\mathrm{a}_1 = (a_0, a_1)$ we find that

$$\Phi_1^{(a_1)}\left(\eta_0^{(a_0,N')}\right) = \eta_0^{(a_0,N')}\mathcal{K}_{n,\eta_0^{(a_0,N')}}^{(a_1)}$$

We let $A_1^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_1^i}$ be the occupation measure associated with N conditionally independent and identically distributed random variables $\mathrm{a}_1^i := (a_{0,1}^i, a_{1,1}^i)$ with common law

$$\Omega_1\left(\gamma_0^N(1), A_0^N, \eta_0^{(\cdot,N')}\right)$$

In the above displayed formula $\eta_0^{(\cdot,N')}$ stands for the mapping $a_0\in\mathcal{A}_0 \mapsto \eta_0^{(a_0,N')}\in\mathcal{P}(E_0)$.

We consider a sequence of conditionally independent random variables $\xi_1^{[a_0,a_1,j]}$ with distribution $\mathcal{K}_{n,\eta_0^{(a_0,N')}}^{(a_1)}\left(\xi_0^{[a_0,j]}, \cdot\right)$, with $1\leq j\leq N'$, and we set

$$\eta_1^{((a_0,a_1),N')} = \frac{1}{N'}\sum_{i=1}^{N'} \delta_{\xi_1^{[(a_0,a_1),j]}} \quad \text{and} \quad \eta_1^N := \int A_1^N(d\mathrm{a})\ \eta_1^{(\mathrm{a},N')}$$

Arguing as before, for any given $\mathrm{a}_1 := (a_0, a_1) \in \mathrm{Supp}(A_1^N)$, the sequence of random fields

$$\mathcal{W}_1^{(\mathrm{a}_1 N')} := \sqrt{N}\ \left(\eta_1^{((a_0,a_1),N')} - \Phi_1^{(a_1)}\left(\eta_0^{(a_0,N')}\right)\right)$$

satisfies (8.2.6), with $N = N'$. Thus, we conclude that

$$\begin{aligned}\eta_1^N &= \int \Omega_1\left(\gamma_0^N(1), A_0^N, \eta_0^{(\bullet,N')}\right)(d(a_0,a_1))\ \Phi_1^{(a_1)}\left(\eta_0^{(a_0,N')}\right) + \frac{1}{\sqrt{N}}\ W_1^N \\ &= \Gamma_1^2\left(\gamma_0^N(1), \eta_0^N\right) + \frac{1}{\sqrt{N}}\ W_1^N\end{aligned}$$

with some local sampling random fields W_1^N satisfying the same estimates as in (8.2.6) by replacing $1/\sqrt{N}$ by the sum $\left(1/\sqrt{N} + 1/\sqrt{N'}\right)$. Iterating this procedure, we define by induction a sequence of N-particle approximation measures

$$\gamma_n^N(1) = \gamma_{n-1}^N(1)\ \eta_{n-1}^N(G_{n-1,\gamma_{n-1}^N}) \quad \text{and} \quad \eta_n^N := \int A_n^N(da)\ \eta_n^{(a,N')}$$

with the occupation measure $A_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_n^i}$ associated with N conditionally independent and identically distributed random variables $\mathrm{a}_n^i := (a_{0,n}^i, a_{1,n}^i, \ldots, a_{n,n}^i)$ with common law $\Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N, \eta_{n-1}^{(\bullet,N')}\right)$. Arguing as above, we find that

$$\eta_n^N = \int \Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N, \eta_{n-1}^{(\bullet,N')}\right)(d(a,b))\ \Phi_n^{(b)}\left(\eta_{n-1}^{(a,N')}\right) = \Gamma_n^2\left(\gamma_{n-1}^N(1), \eta_{n-1}^N\right) + \frac{1}{\sqrt{N}}\ W_n^N$$

with some local sampling random fields satisfying the same estimates as in (8.2.6) by replacing $1/\sqrt{N}$ by the sum $\left(1/\sqrt{N} + 1/\sqrt{N'}\right)$. As before, the N-particle occupation measures A_n^N converge as N tends to ∞ to the association probability measures A_n.

8.6 Appendix

Proof of corollary 8.4.6

For constant mappings $s_n = \mu_{n+1}(1)$, the mappings Φ_{n+1,ν_n}^1 and Φ_{n+1,m_n}^2 are given by

$$\Phi_{n+1,\nu_n}^1(u) = s_n \quad \text{and} \quad \Phi_{n+1,m_n}^2(\eta) = \Psi_{g_n^{(s)}}(\eta) M_{n+1}^{(s)}$$

with the likelihood function $g_n^{(s)}$ and the Markov transitions $M_{n+1}^{(s)}$ defined in (8.3.7). Firstly, we observe that $r_n(s_n) := \sup_{x,x' \in E_n} g_n^{(s)}(x)/g_n^{(s)}(x')$. We also notice that the second component mapping Φ_{n+1,m_n}^2 does not depends on the parameter m_n, and it induces a Feynman-Kac semigroup of the same form as the one discussed in section 8.4.2.

Under the premise of the proposition, the semigroup of associated with the Markov transitions M_n satisfies the mixing property stated in the l.h.s. of (8.4.11) for some integer $m \geq 1$ and some parameter $\epsilon_p(m) \in]0, 1]$. In this situation, we also have that

$$M^{(s)}_{p,p+m}(x, .) \geq \epsilon^{(s)}_p(m)\ M^{(s)}_{p,p+m}(x', .)$$

with some positive parameter

$$\epsilon^{(s)}_p(m) \geq \epsilon_p(m) / \prod_{p \leq k < p+m} r_k(s_k) r_k(1) \quad \text{and} \quad r_n(s_n) := \frac{s_n g^+_n + (1 - s_n)}{s_n g^-_n + (1 - s_n)} \, (\leq r_n(1))$$

To prove this claim, firstly we observe that $M^{(s)}_{p,p+m}(x, .) \ll M^{(s)-}_{p,p+m}(x, .)$ and

$$\prod_{p \leq k < p+m} r_k(s_k)^{-1} \leq dM^{(s)}_{p,p+m}(x, .)/dM^{(s)-}_{p,p+m}(x, .) \leq \prod_{p \leq k < p+m} r_k(1)$$

with the semigroup $M^{(s)-}_{p,n}$ associated with the Markov transition

$$M^{(s)-}_{p,p+1}(x, .) = \alpha_{p+1}\ M_{p+1}(x, .) + (1 - \alpha_{k+1})\ \overline{\mu}_{k+1} \quad \text{with} \quad \alpha_{p+1} := \frac{s_k g^-_k}{s_k g^-_k + (1 - s_k)}$$

Using the geometric representation

$$M^{(s)-}_{p,n}(x, .) = \left(\prod_{p < k \leq n} \alpha_k \right) M_{p,n}(x, .) + \sum_{p < k \leq n} (1 - \alpha_k) \ \left(\prod_{k < l \leq n} \alpha_l \right) \ \overline{\mu}_k M_{k,n}$$

it can be verified that

$$M^{(s)-}_{p,p+m}(x, .) \geq \epsilon_p(m)\ M^{(s)-}_{p,p+m}(x', .) \geq \epsilon_p(m) \ \left(\prod_{p \leq k < p+m} g^-_k / g^+_k \right) \ M^{(s)}_{p,p+m}(x', .)$$

from which we conclude that

$$M^{(s)}_{p,p+m}(x, .) \geq \epsilon^{(s)}_p(m)\ M^{(s)}_{p,p+m}(x', .) \quad \text{with} \quad \epsilon^{(s)}_p(m) \geq \epsilon_p(m) / \prod_{p \leq k < p+m} r_k(s_k) r_k(1)$$

We end the proof of the proposition combing the proposition 8.4.5 with the couple of estimates presented in (8.4.12) and (8.4.13). This ends the proof of the corollary. ■

Proof of theorem 8.4.7

The formulae presented in (8.3.6) can be rewritten in terms of matrix operations as follows

$$[\gamma_{n+1}(1)\ ,\ 1 - \gamma_{n+1}(1)] = [\widehat{\gamma}_n(1)\ ,\ 1 - \widehat{\gamma}_n(1)] \ \begin{bmatrix} \Psi_{g_n}(\eta_n)(s_n) & 1 - \Psi_{g_n}(\eta_n)(s_n) \\ \mu_{n+1}(1) & 1 - \mu_{n+1}(1) \end{bmatrix}$$

and

$$[\widehat{\gamma}_n(1)\ ,\ 1-\widehat{\gamma}_n(1)] = \frac{[\gamma_n(1)\ ,\ 1-\gamma_n(1)] \begin{bmatrix} \eta_n(g_n) & 0 \\ 0 & 1 \end{bmatrix}}{[\gamma_n(1)\ ,\ 1-\gamma_n(1)] \begin{bmatrix} \eta_n(g_n) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

With a slight abuse of notation, we set

$$\vartheta_n := [\gamma_n(1)\ ,\ 1-\gamma_n(1)] \quad \widehat{\vartheta}_n := [\widehat{\gamma}_n(1)\ ,\ 1-\widehat{\gamma}_n(1)] \quad \text{and} \quad 1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

We also denote by $\mathcal{M}_{n+1,\eta_n}$ and $\mathcal{D}_{n,\eta_n}$ the stochastic and the diagonal matrices defined by

$$\mathcal{M}_{n+1,\eta_n} := \begin{bmatrix} \Psi_{g_n}(\eta_n)(s_n) & 1-\Psi_{g_n}(\eta_n)(s_n) \\ \mu_{n+1}(1) & 1-\mu_{n+1}(1) \end{bmatrix} \quad \text{and} \quad \mathcal{D}_{n,\eta_n} := \begin{bmatrix} \eta_n(g_n) & 0 \\ 0 & 1 \end{bmatrix} \tag{8.6.1}$$

In this notation, the above recursion can be rewritten in a more compact form

$$\vartheta_{n+1} = \widehat{\vartheta}_n\ \mathcal{M}_{n+1,\eta_n} \quad \text{and} \quad \widehat{\vartheta}_n = \frac{\vartheta_n\ \mathcal{D}_{n,\eta_n}}{\vartheta_n \mathcal{D}_{n,\eta_n} 1} \implies \vartheta_{n+1} = \frac{\vartheta_n\ \mathcal{Q}_{n+1,\eta_n}}{\vartheta_n \mathcal{Q}_{n+1,\eta_n} 1}$$

with the product of matrices $\mathcal{Q}_{n+1,\eta_n} = \mathcal{D}_{n,\eta_n}\mathcal{M}_{n+1,\eta_n}$.

$$\forall u \in I_p(\subset [0,1]) \qquad \left[\Phi^1_{p,n,\nu}(u), 1-\Phi^1_{p,n,\nu}(u)\right] = \frac{[u,1-u]\ \mathcal{Q}_{p,n,\nu}}{[u,1-u]\ \mathcal{Q}_{p,n,\nu}(1)}$$

with the matrix semigroup

$$\mathcal{Q}_{p,n,\nu} = \mathcal{Q}_{p+1,\nu_p}\mathcal{Q}_{p+2,\nu_{p+1}} \cdots \mathcal{Q}_{n,\nu_{n-1}}$$

These semigroups are again of the same form as the Feynman-Kac models discussed in section 8.4.2 with a two point state space. When $\mu_{n+1}(1) \in]0,1[$ and $0 < s_n^- \leq s_n^+ < 1$, we have for any $n \geq 0$ and any $i, i', j \in \{1,2\}$

$$\mathcal{M}_{n+1,\nu_n}(i,j) \geq \epsilon_n\ \mathcal{M}_{n+1,\nu_n}(i',j) \quad \text{and} \quad \sup_{i,i' \in \{1,2\}} \frac{\mathcal{Q}_{n+1,\nu_n}(1)(i)}{\mathcal{Q}_{n+1,\nu_n}(1)(i')} \leq \delta'_n(g)$$

The first assertion is a direct consequence of the proposition 8.4.5 with the couple of estimates presented in (8.4.12) and (8.4.13).

Using (8.3.3), we find that Φ^2_{n+1,m_n} induces a Feynman-Kac models of the same form as the one discussed in section 8.4.2. More precisely, we have that

$$\Phi^2_{n+1,m_n}(\eta) = \Psi_{\widehat{G}_{n,m_n}}(\eta)\widehat{M}_{n+1,m_n}$$

with the potential functions G_{n,m_n} and the Markov transitions $\widehat{M}_{n+1,m_n}$ defined in (8.3.4) and (8.3.5). Notice that

$$\sup_{x,x' \in E_n} \frac{\widehat{G}_{n,m_n}(x)}{\widehat{G}_{n,m_n}(x')} \leq \delta_n(sg)$$

and for any $x \in E_n$ and any $n \geq 0$

$$\delta_n(sg)^{-1}\ \widehat{M}^-_{n+1,m_n}(x,.) \leq \widehat{M}_{n+1,m_n}(x,.) \leq \delta_n(sg)\ \widehat{M}^-_{n+1,m_n}(x,.)$$

with the Markov transitions $\widehat{M}^-_{n+1,m_n}$ defined as $\widehat{M}^-_{n+1,m_n}$ by replacing the functions (s_n, g_n) by their lower bounds (s_n^-, g_n^-). To prove this claim, we use the fact that for any positive function f we have

$$\frac{d\widehat{M}_{n+1,m_n}(f)}{d\widehat{M}^-_{n+1,m_n}(f)} = \frac{m_n g_n^- s_n^- + (1-m_n)\mu_{n+1}(1)}{m_n g_n s_n + (1-m_n)\mu_{n+1}(1)} \times \frac{m_n g_n s_n M_{n+1}(f) + (1-m_n)\mu_{n+1}(1)\overline{\mu}_{n+1}(f)}{m_n g_n^- s_n^- + (1-m_n)\mu_{n+1}(1)\overline{\mu}_{n+1}(f)}$$

and the two series of inequalities

$$\delta_n(sg)^{-1} \leq \frac{m_n g_n^- s_n^- + (1-m_n)\mu_{n+1}(1)}{m_n g_n s_n + (1-m_n)\mu_{n+1}(1)} \leq 1$$

and

$$1 \leq \frac{m_n g_n s_n M_{n+1}(f) + (1-m_n)\mu_{n+1}(1)\overline{\mu}_{n+1}(f)}{m_n g_n^- s_n^- + (1-m_n)\mu_{n+1}(1)\overline{\mu}_{n+1}(f)} \leq \delta_n(sg)$$

With a slight abuse of notation, we write $\widehat{M}_{p,n}$, and respectively $\widehat{M}^-_{p,n}$, the semi-group associated with the Markov transitions $\widehat{M}_{n+1,m_n}$, and resp. $\widehat{M}^-_{n+1,m_n}$. Using the same argument as in the proof of corollary 8.4.6 it follows that

$$\widehat{M}^-_{p,p+m}(x,.) \geq \epsilon_p(m)\ \widehat{M}^-_{p,p+m}(x',.)$$

from which we conclude that

$$\widehat{M}_{p,p+m}(x,.) \geq \widehat{\epsilon}_p(m)\ \widehat{M}_{p,p+m}(x',.) \quad \text{with} \quad \widehat{\epsilon}_p(m) \geq \epsilon_p(m) \prod_{0\leq k<m} \delta_{p+k}(sg)^{-2}$$

using proposition 8.4.5 with the couple of estimates presented in (8.4.12) and (8.4.13), we check that (8.4.3) is satisfied with

$$a^2_{p,n} \leq 2\ \rho_p(m) \prod_{k=0}^{\lfloor n/m \rfloor - 1} \left(1 - \epsilon^{(m)}_{p+km}\right)$$

and some parameters

$$\epsilon_p^{(m)} := \widehat{\epsilon}_p(m)^2 \prod_{0<k<m} \delta_{p+k}(sg)^{-1} \geq \epsilon_p(m)^2\ \delta_p(sg)^{-4} \prod_{0<k<m} \delta_{p+k}(sg)^{-5}$$

and

$$\rho_p(m) := \widehat{\epsilon}_p(m)^{-1} \prod_{0\leq k<m} \delta_{p+k}(sg) \leq \epsilon_p(m)^{-1} \prod_{0\leq k<m} \delta_{p+k}(sg)^3$$

This ends the proof of the first assertion of the theorem. Next, we discuss condition (Cont(Φ)). We observe that

$$\Phi^1_{n+1,\nu}(u) = \frac{u\ \nu(g_n s_n) + (1-u)\mu_{n+1}(1)}{u\ \nu(g_n) + (1-u)}$$

After some manipulations

$$\begin{aligned}
&\Phi^1_{n+1,\nu}(u) - \Phi^1_{n+1,\nu'}(u) \\
&= \tfrac{u\nu'(g_n)}{u\nu'(g_n)+(1-u)}\ \left[\Psi_{g_n}(\nu) - \Psi_{g_n}(\nu')\right](s_n) \\
&\quad + \tfrac{u}{u\nu(g_n)+(1-u)} \tfrac{(1-u)}{u\nu'(g_n)+(1-u)}\ \left[\Psi_{g_n}(\nu)(s_n) - \mu_{n+1}(1)\right]\ \left[\nu - \nu'\right](g_n)
\end{aligned}$$

Recalling that the mapping $\theta_a(x) = ax/(ax+(1-x))$ in increasing on $[0,1]$ and using the fact that

$$\Psi_{g_n}(\nu) = \nu S_{n,\nu} \Longrightarrow \Psi_{g_n}(\nu) - \Psi_{g_n}(\nu') = \frac{g_n^+}{\nu(g_n)}\ (\nu - \nu')S_{n,\nu'}$$

with the Markov transition

$$S_{n,\nu'}(x, dx') = \frac{g_n(x)}{g_n^+(x)}\ \delta_x(dx') + \left(1 - \frac{g_n(x)}{g_n^+(x)}\right)\ \Psi_{g_n}(\nu')(dx')$$

we prove

$$|\Psi_{g_n}(\nu)(s_n) - \Psi_{g_n}(\nu')(s_n)| \leq \frac{g_n^+}{g_n^-}\ |(\nu - \nu')S_{n,\nu'}(s_n)| \tag{8.6.2}$$

and for any $u \in I_n = [m_n^-, m_n^+]$

$$\begin{aligned}
&\left|\Phi^1_{n+1,\nu}(u) - \Phi^1_{n+1,\nu'}(u)\right| \\
&= \tfrac{m_n^+ g_n^+}{m_n^+ g_n^+ + (1-m_n^+)}\ \tfrac{g_n^+}{g_n^-}\ |(\nu - \nu')S_{n,\nu'}(s_n)| \\
&\quad + \tfrac{m_n^+ g_n^+}{m_n^+ g_n^+ + (1-m_n^+)} \tfrac{(1-m_n^-)}{m_n^- g_n^- + (1-m_n^-)}\ |s_n - \mu_{n+1}(1)|\ \ |[\nu - \nu'](g_n/g_n^-)|
\end{aligned}$$

This implies that

$$\begin{aligned}
\tau^1_{n+1} &\leq \frac{m_n^+ g_n^+}{m_n^+ g_n^+ + (1-m_n^+)}\ \frac{g_n^+}{g_n^-}\ (s_n^+ - s_n^-) \\
&\qquad + \frac{m_n^+ g_n^+}{m_n^+ g_n^+ + (1-m_n^+)}\ \frac{(1-m_n^-)}{m_n^- g_n^- + (1-m_n^-)}\ |s_n - \mu_{n+1}|\ \left(\frac{g_n^+}{g_n^-} - 1\right) \\
&\leq \frac{g_n^+}{g_n^-}\ \left[(s_n^+ - s_n^-) + |s_n - \mu_{n+1}(1)|\right]
\end{aligned}$$

Using (8.3.3) we also find that

$$\Phi^2_{n+1,m}(\eta)(f) = \frac{m\eta(s_n g_n M_{n+1}(f)) + (1-m)\ \mu_{n+1}(f)}{m\eta(s_n g_n) + (1-m)\mu_{n+1}(1)}$$

It is also readily check that

$$\left[\Phi^2_{n+1,m}(\eta) - \Phi^2_{n+1,m'}(\eta)\right](f) = \frac{\mu_{n+1}(1)\ \eta(g_n s_n)\ \left[\Psi_{g_n s_n}(\eta) M_{n+1} - \overline{\mu}_{n+1}\right](f)\ (m-m')}{[m\eta(s_n g_n) + (1-m)\mu_{n+1}(1)]\,[m'\eta(s_n g_n) + (1-m')\mu_{n+1}(1)]}$$

from which we conclude that

$$\tau^2_{n+1} \leq \sup\left\{\frac{\mu_{n+1}}{s^-_n g^-_n}, \frac{s^+_n g^+_n}{\mu_{n+1}(1)}\right\} \leq \delta'_n(g)\ \sup\left\{\frac{\mu_{n+1}}{s^-_n}, \frac{s^+_n}{\mu_{n+1}(1)}\right\}$$

This ends the proof of the theorem. ∎

Proof of proposition 8.4.3

The proof of proposition 8.4.3 is based on the following technical lemma.

Lemma 8.6.1. *We assume that the regularity conditions (Lip(Φ)) and (Cont(Φ)) are satisfied. In this situation, for any $p \leq n$, $u, u' \in I_p$, $\eta, \eta' \in \mathcal{P}(E_p)$ and $f \in Osc_1(E_n)$ and any flow of masses and probability measures $m = (m_n)_{n\geq 0} \in \prod_{n\geq 0} I_n$ and $\nu := (\nu_n)_{n\geq 0} \in \prod_{n\geq 0} \mathcal{P}(E_n)$ we have the following estimates*

$$\left|\Phi^1_{p,n,\nu'}(u') - \Phi^1_{p,n,\nu}(u)\right| \leq a^1_{p,n}\ |u-u'| + \sum_{p\leq q<n} \overline{a}^1_{q,n} \int \left|[\nu_q - \nu'_q](\varphi)\right|\ \Omega^1_{q+1,\nu'_q}(d\varphi)$$

$$\left|\Phi^2_{p,n,m'}(\eta')(f) - \Phi^2_{p,n,m}(\eta)(f)\right| \leq a^2_{p,n} \int |[\eta-\eta'](\varphi)|\,\Omega^2_{p,n,\eta'}(f, d\varphi) + \sum_{p\leq q<n} \overline{a}^2_{q,n}\ |m_q - m'_q|$$

with the collection of parameters $\overline{a}^i_{p,n}$, $i = 1, 2$, defined in (8.4.6).

Proof. We use the decomposition

$$\begin{aligned}\Phi^1_{p,n,\nu'}(u') - \Phi^1_{p,n,\nu}(u) &= \Phi^1_{p,n,\nu}(u') - \Phi^1_{p,n,\nu}(u) \\ &\quad + \sum_{p<q\leq n} \left[\Phi^1_{q,n,\nu}(\Phi^1_{p,q,\nu'}(u')) - \Phi^1_{q-1,n,\nu}(\Phi^1_{p,q-1,\nu'}(u'))\right]\end{aligned}$$

and the fact that

$$\begin{aligned}\Phi^1_{q-1,n,\nu}(\Phi^1_{p,q-1,\nu'}(u')) &= \Phi^1_{q,n,\nu}\left(\Phi^1_{q-1,q,\nu}\left[\Phi^1_{p,q-1,\nu'}(u')\right]\right) \\ \Phi^1_{q,n,\nu}(\Phi^1_{p,q,\nu'}(u')) &= \Phi^1_{q,n,\nu}\left(\Phi^1_{q-1,q,\nu'}\left[\Phi^1_{p,q-1,\nu'}(u')\right]\right)\end{aligned}$$

and

$$\left|\Phi^1_{p,n,\nu}(u') - \Phi^1_{p,n,\nu}(u)\right| \leq a^1_{p,n}\ |u-u'|$$

and

$$\begin{aligned}&\left|\Phi^1_{q,n,\nu}(\Phi^1_{p,q,\nu'}(u')) - \Phi^1_{q-1,n,\nu}(\Phi^1_{p,q-1,\nu'}(u'))\right| \\ &\leq a^1_{q,n}\ \left|\Phi^1_{q,\nu_{q-1}}\left[\Phi^1_{p,q-1,\nu'}(u')\right] - \Phi^1_{q,\nu'_{q-1}}\left[\Phi^1_{p,q-1,\nu'}(u')\right]\right| \\ &\leq\ \overline{a}^1_{q-1,n}\ \int\ \left|[\nu_{q-1}-\nu'_{q-1}](\varphi)\right|\ \Omega_{q,\nu'_{q-1}}(d\varphi)\end{aligned}$$

to show that

$$\left|\Phi^1_{p,n,\nu'}(u') - \Phi^1_{p,n,\nu}(u)\right| \leq a^1_{p,n}\,|u-u'| + \sum_{p<q\leq n}\overline{a}^1_{q-1,n}\int\ \left|[\nu_{q-1}-\nu'_{q-1}](\varphi)\right|\ \Omega^1_{q,\nu'_{q-1}}(d\varphi)$$

In the same way, we use the decomposition

$$\begin{aligned}\left[\Phi^2_{p,n,m'}(\eta') - \Phi^2_{p,n,m}(\eta)\right] &= \left[\Phi^2_{p,n,m}(\eta') - \Phi^2_{p,n,m}(\eta)\right] \\ &\quad + \sum_{p<q\leq n}\left[\Phi^2_{q,n,m}(\Phi^2_{p,q,m'}(\eta')) - \Phi^2_{q-1,n,m}(\Phi^2_{p,q-1,m'}(\eta'))\right]\end{aligned}$$

and the fact that

$$\begin{aligned}\Phi^2_{q-1,n,m}(\Phi^2_{p,q-1,m'}(\eta')) &= \Phi^2_{q,n,m}\left(\Phi^2_{q-1,q,m}\left[\Phi^2_{p,q-1,m'}(\eta')\right]\right) \\ \Phi^2_{q,n,m}(\Phi^2_{p,q,m'}(\eta')) &= \Phi^2_{q,n,m}\left(\Phi^2_{q-1,q,m'}\left[\Phi^2_{p,q-1,m'}(\eta')\right]\right)\end{aligned}$$

and

$$\left|\Phi^2_{p,n,m}(\eta')(f) - \Phi^2_{p,n,m}(\eta)(f)\right| \leq a^2_{p,n}\int\ |[\eta-\eta'](\varphi)|\ \Omega^2_{p,n,\eta'}(f,d\varphi)$$

to show that

$$\begin{aligned}&\left|\Phi^2_{q,n,m}(\Phi^2_{p,q,m'}(\eta')) - \Phi^2_{q-1,n,m}(\Phi^2_{p,q-1,m'}(\eta'))\right| \\ &\leq a^2_{q,n}\int\ \left|[\Phi^2_{q,m_{q-1}}\left[\Phi^2_{p,q-1,m'}(\eta')\right] - \Phi^2_{q,m'_{q-1}}\left[\Phi^2_{p,q-1,m'}(\eta')\right]](\varphi)\right|\ \Omega^2_{q,n,\Phi^2_{p,q,m'}(\eta')}(f,d\varphi) \\ &\leq \overline{a}^2_{q-1,n}\,|m_{q-1}-m'_{q-1}|\end{aligned}$$

Using these estimates we conclude that

$$\left|\left[\Phi^2_{p,n,m'}(\eta') - \Phi^2_{p,n,m}(\eta)\right](f)\right| \leq a^2_{p,n}\int\ |[\eta-\eta'](\varphi)|\ \Omega^2_{p,n,\eta}(f,d\varphi) + \sum_{p<q\leq n}\overline{a}^2_{q-1,n}\,|m_{q-1}-m'_{q-1}|$$

This ends the proof of the lemma. ∎

Now we come to the proof of proposition 8.4.3.

Proof of proposition 8.4.3:

We fix a parameter $p \geq 0$, and we let $(m_n)_{n\geq p}, (m'_n)_{n\geq p} \in \prod_{n\geq p} I_n$ and $(\nu_n)_{n\geq p}$, and $(\nu'_n)_{n\geq p} \in \prod_{n\geq p}\mathcal{P}(E_n)$ be defined by the following recursive formulae

$$\forall q > p \qquad m'_q = \Phi^1_{q,\nu'_{q-1}}(m'_{q-1}) \quad \text{and} \quad \nu'_q = \Phi^2_{q,m'_{q-1}}(\nu'_{q-1})$$

$$\forall q > p \qquad m_q = \Phi^1_{q,\nu_{q-1}}(m_{q-1}) \quad \text{and} \quad \nu_q = \Phi^2_{q,m_{q-1}}(\nu_{q-1})$$

with the initial condition for $q = p$

$$(\nu_p, \nu'_p) = (\eta, \eta') \quad \text{and} \quad (m_p, m'_p) = (u, u')$$

By construction, we have

$$\nu'_q = \Phi^2_{p,q,m'}(\eta') \quad \text{and} \quad \nu_q = \Phi^2_{p,q,m}(\eta)$$

as well as

$$m'_q = \Phi^1_{p,q,\nu'}(u') \quad \text{and} \quad m_q = \Phi^1_{p,q,\nu}(u)$$

In this case, using lemma 8.6.1 it follows that

$$\begin{aligned}
&\left|\left[\Gamma^2_{p,n}(m', \eta') - \Gamma^2_{p,n}(m, \eta)\right](f)\right] \\
&\leq a^2_{p,n} \int |[\eta - \eta'](\varphi)| \ \Omega^2_{p,n,\eta'}(f, d\varphi) + \sum_{p \leq q < n} \overline{a}^2_{q,n} \ |\Gamma^1_{p,q}(m', \eta') - \Gamma^1_{p,q}(m, \eta)|
\end{aligned}$$

and

$$\begin{aligned}
&\left|\Gamma^1_{p,n}(m', \eta') - \Gamma^1_{p,n}(m, \eta)\right| \\
&\leq a^1_{p,n} \ |m - m'| + \sum_{p \leq q < n} \overline{a}^1_{q,n} \int \left|[\Gamma^2_{p,q}(m', \eta') - \Gamma^2_{p,q}(m, \eta)](\varphi)\right| \ \overline{\Omega}^1_{p,q,m',\eta'}(d\varphi)
\end{aligned}$$

with the probability measure $\overline{\Omega}^1_{p,q,m',\eta'} = \Omega^1_{q+1,\Gamma^2_{p,q}(m',\eta')}$.

Combining these two estimates, we arrive at the following inequality

$$\begin{aligned}
&\left|\left[\Gamma^2_{p,n}(m', \eta') - \Gamma^2_{p,n}(m, \eta)\right](f)\right] \\
&\leq a^2_{p,n} \int |[\eta - \eta'](\varphi)| \ \Omega^2_{p,n,\eta'}(f, d\varphi) + \left[\sum_{p \leq q < n} a^1_{p,q} \ \overline{a}^2_{q,n}\right] \ |m - m'| \\
&+ \sum_{p \leq r < q < n} \overline{a}^1_{r,q} \ \overline{a}^2_{q,n} \int \left|[\Gamma^2_{p,r}(m', \eta') - \Gamma^2_{p,r}(m, \eta)](\varphi)\right| \ \overline{\Omega}^1_{p,r,m',\eta'}(d\varphi)
\end{aligned}$$

This implies that

$$\begin{aligned}
&\left|\left[\Gamma^2_{p,n}(m', \eta') - \Gamma^2_{p,n}(m, \eta)\right](f)\right] \\
&\leq b'_{p,n} \ |m - m'| + a^2_{p,n} \int |[\eta - \eta'](\varphi)| \ \Omega^2_{p,n,\eta'}(f, d\varphi) \\
&+ \sum_{p \leq r_1 < n} b_{r_1,n} \int \left|[\Gamma^2_{p,r_1}(m', \eta') - \Gamma^2_{p,r_1}(m, \eta)](\varphi)\right| \ \overline{\Omega}^1_{p,r_1,m',\eta'}(d\varphi)
\end{aligned}$$

Our next objective is to show that

$$\left|\left[\Gamma^2_{p,n}(m',\eta') - \Gamma^2_{p,n}(m,\eta)\right](f)\right|$$

$$\leq \alpha^k_{p,n}\ |m-m'| + \beta^k_{p,n} \int\ |[\eta-\eta'](\varphi)|\ \Theta^k_{p,n,\eta'}(f,d\varphi)$$

$$+ \sum_{p\leq r_1<r_2<...<r_k<n} b_{r_1,r_2}\ \ldots\ b_{r_k,n} \int\ \left|[\Gamma^2_{p,r_1}(m',\eta') - \Gamma^2_{p,r_1}(m,\eta)](\varphi)\right|\ \overline{\Omega}^1_{p,r_1,m',\eta'}(d\varphi)$$

for any $k \leq (n-p)$ for some Markov transitions $\Theta^k_{p,n,m'\eta'}(f,d\varphi)$ and the parameters

$$\begin{aligned}\alpha^k_{p,n} &= b'_{p,n} + \sum_{l=1}^{k-1} \sum_{p\leq r_1<...r_l<n} b'_{p,r_1}\ b_{r_1,r_2}\ldots b_{r_l,n} \\ \beta^k_{p,n} &= a^2_{p,n} + \sum_{l=1}^{k-1} \sum_{p\leq r_1<...r_l<n} a^2_{p,r_1}\ b_{r_1,r_2}\ldots b_{r_l,n}\end{aligned}$$

We proceed by induction on the parameter k. Firstly, we observe that the result is satisfied for $k=1$ with

$$\left(\alpha^1_{p,n},\beta^1_{p,n}\right) = \left(b'_{p,n},a^2_{p,n}\right) \quad \text{and} \quad \Theta^1_{p,n,\eta'} = \Omega^2_{p,n,\eta'}$$

We further assume that the result is satisfied at rank k. In this situation, using the fact that

$$\left|\left[\Gamma^2_{p,r_1}(m',\eta') - \Gamma^2_{p,r_1}(m,\eta)\right](\varphi)\right|$$

$$\leq b'_{p,r_1}\ |m-m'| + a^2_{p,r_1} \int\ |[\eta-\eta'](\varphi')|\ \Omega^2_{p,r_1,\eta'}(\varphi,d\varphi')$$

$$+ \sum_{p\leq r_0<r_1} b_{r_0,r_1} \int\ \left|[\Gamma^2_{p,r_0}(m',\eta') - \Gamma^2_{p,r_0}(m,\eta)](\varphi)\right|\ \overline{\Omega}^1_{p,r_0,m',\eta'}(d\varphi)$$

we conclude that

$$\left|\left[\Gamma^2_{p,n}(m',\eta') - \Gamma^2_{p,n}(m,\eta)\right](f)\right|$$

$$\leq \alpha^{k+1}_{p,n}\ |m-m'| + \beta^{k+1}_{p,n} \int\ |[\eta-\eta'](\varphi)|\ \Theta^{k+1}_{p,n,m'\eta'}(f,d\varphi)$$

$$+ \sum_{p\leq r_0<r_1<r_2<...<r_k<n} b_{r_0,r_1}\ b_{r_1,r_2}\ \ldots\ b_{r_k,n} \int\ \left|[\Gamma^2_{p,r_0}(m',\eta') - \Gamma^2_{p,r_0}(m,\eta)](\varphi)\right|\ \overline{\Omega}^1_{p,r_0,m',\eta'}(d\varphi)$$

with

$$\begin{aligned}\alpha^{k+1}_{p,n} &= \alpha^k_{p,n} + \sum_{p\leq r_1<r_2<...<r_k<n} b'_{p,r_1}\ b_{r_1,r_2}\ \ldots\ b_{r_k,n} \\ \beta^{k+1}_{p,n} &= \beta^k_{p,n} + \sum_{p\leq r_1<r_2<...<r_k<n} a^2_{p,r_1}\ b_{r_1,r_2}\ \ldots\ b_{r_k,n}\end{aligned}$$

and the Markov transition

$$\begin{aligned}\beta_{p,n}^{k+1}\ \Theta_{p,n,m'\eta'}^{k+1}(f,d\varphi) &= \beta_{p,n}^{k}\ \Theta_{p,n,\eta'}^{k}(f,d\varphi)\\ &\quad+\sum_{p\leq r_1<r_2<\ldots<r_k<n} a_{p,r_1}^2\ b_{r_1,r_2}\ \ldots\ b_{r_k,n}\ \left(\overline{\Omega}_{p,r_1,m',\eta'}^1\Omega_{p,r_1,\eta'}^2\right)(d\varphi)\end{aligned}$$

We end the proof of the proposition using the fact that

$$\begin{aligned}\left|\Gamma_{p,n}^1(m',\eta')-\Gamma_{p,n}^1(m,\eta)\right| &\leq \left[a_{p,n}^1+\sum_{p\leq q<n} c_{p,q}^{2,1}\ \overline{a}_{q,n}^1\right]\ |m-m'|\\ &\quad+\sum_{p\leq q<n}\overline{a}_{q,n}^1\ c_{p,q}^{2,2}\int\ |[\eta-\eta'](\varphi')|\ \left[\overline{\Omega}_{p,q,m',\eta'}^1\Theta_{p,q,\eta'}\right](d\varphi')\end{aligned}$$

This proof of the proposition is now completed. ∎

Proof of theorem 8.4.11

For any $\eta\in\mathcal{P}(E)$ and any $u,u'\in I_n$, we have

$$\begin{aligned}&\left|\Phi_{n+1,\eta}^1(u)-\Phi_{n+1,\eta}^1(u')\right|\\ &=|u-u'|\left[r(1-d)+rdh\ \int\mathcal{Y}_n(dy)\ \tfrac{\eta(g(\bullet,y))}{[h+du\eta(g(\bullet,y))][h+du'\eta(g(\bullet,y))]}\right]\\ &\leq|u-u'|\left[r(1-d)+rdh\ \mathcal{Y}_n\left(\tfrac{g^+}{[h+dm^-g^-)]^2}\right)\right]\end{aligned}$$

This implies that condition (8.4.2) is satisfied with

$$a_{n,n+1}^1\leq r(1-d)+rdh\ \mathcal{Y}_n\left(\frac{g^+}{[h+dm^-g^-)]^2}\right)$$

In the same way, for any $\eta,\eta'\in\mathcal{P}(E)$ and any $u\in I_n$, we have

$$\Phi_{n+1,\eta}^1(u)-\Phi_{n+1,\eta'}^1(u)=rdhu\ \int\mathcal{Y}_n(dy)\ \tfrac{1}{[h+du\eta(g(\bullet,y))][h+du\eta'(g(\bullet,y))]}\ (\eta-\eta')\,(g(\bullet,y))$$

$$\tau_{n+1}^1\leq rdhm^+\ \mathcal{Y}_n\left(\frac{g^+-g^-}{[h+dm^-g^-]^2}\right)$$

and the probability measure

$$\Omega_{n,\eta'}^1(d\varphi)\propto\int\mathcal{Y}_n(dy)\ \frac{g^+(y)-g^-(y)}{[h+dm^-g^-(y)]^2}\ \delta_{\frac{g(\bullet,y)}{g^+(y)-g^-(y)}}(d\varphi)$$

Now, we come to the analysis of the mappings

$$\Phi_{n+1,u}^2(\eta)\propto r(1-d)u\ \eta M+\int\mathcal{Y}_n(dy)\ w_u(\eta,y)\ \Psi_{g(\bullet,y)}(\eta)M+\mu(1)\ \overline{\mu}$$

with the weight functions

$$w_u(\eta, y) := \frac{rdu\eta(g(\boldsymbol{.}, y))}{h + du\eta(g(\boldsymbol{.}, y))} = r\left(1 - \frac{h}{h + du\eta(g(\boldsymbol{.}, y))}\right)$$

Notice that

$$w^-(y) := \frac{rdm^- g^-(y)}{h + dm^- g^-(y)} \leq w_u(\eta, y) \leq w^+(y) := \frac{rdm^+ g^+(y)}{h + dm^+ g^+(y)}$$

To have a more synthetic formula, we extend the observation state space with two auxiliary points c_1, c_2 and we set

$$\mathcal{Y}_n^c = \mathcal{Y}_n + \delta_{c_1} + \delta_{c_2}$$

we extend the likelihood and the weight functions by setting

$$g(x, c_1) = g(x, c_2) = 1$$

and

$$\begin{array}{rcl} w^-(c_1) & := & r(1-d)m^- \leq w_u(\eta, c_1) := r(1-d)u \leq w^+(c_1) := r(1-d)m^+ \\ w_u(\eta, c_2) & = & w^+(c_2) = w^-(c_2) := \mu(1) \end{array}$$

In this notation, we find that

$$\Phi^2_{n+1,u}(\eta) \propto \int \mathcal{Y}_n^c(dy)\ w_u(\eta, y)\ \Psi_{g(\boldsymbol{.},y)}(\eta) M_y$$

with the collection of Markov transitions M_y defined below

$$\forall y \notin \{c_2\} \qquad M_y = M \quad \text{and} \quad M_{c_2} = \overline{\mu}$$

Notice that the normalizing constants $\mathcal{Y}_n^c(w_u(\eta, \boldsymbol{.}))$ satisfy the following lower bounds

$$\mathcal{Y}_n^c(w_u(\eta, \boldsymbol{.})) \ \geq \ \mathcal{Y}_n^c(w^-) = r(1-d)\ m^- + \mathcal{Y}_n\left(w^-\right) + \mu(1)$$

We analyze the Lipschitz properties of the mappings $\Phi^2_{n+1,u}$ using the following decomposition

$$\Phi^2_{n+1,u}(\eta) - \Phi^2_{n+1,u}(\eta') = \Delta_{n+1,u}(\eta, \eta') + \Delta'_{n+1,u}(\eta, \eta')$$

with the signed measures

$$\Delta_{n+1,u}(\eta, \eta') = \int \mathcal{Y}_n^c(dy)\ \frac{w_u(\eta, y)}{\mathcal{Y}_n^c(w_u(\eta, \boldsymbol{.}))}\ \left[\Psi_{g(\boldsymbol{.},y)}(\eta) M_y - \Psi_{g(\boldsymbol{.},y)}(\eta') M_y\right]$$

and

$$\Delta'_{n+1,u}(\eta, \eta') = \frac{1}{\mathcal{Y}_n^c(w_u(\eta, \boldsymbol{.}))} \int \mathcal{Y}_n^c(dy)\ [w_u(\eta, y) - w_u(\eta', y)]\ \left(\Psi_{g(\boldsymbol{.},y)}(\eta') M_y - \Phi^2_{n+1,u}(\eta')\right)$$

Arguing as in the proof of theorem 8.4.7 given in the appendix (see for instance (8.6.2)), one checks that

$$|\Delta_{n+1,u}(\eta,\eta')(f)|$$

$$\leq \tfrac{1}{\mathcal{Y}_n^c(w^-)} \left(r(1-d)m^+ \, |(\eta-\eta')(M(f))| + \int \mathcal{Y}_n(dy) \; w^+(y) \; \tfrac{g^+(y)}{g^-(y)} \; \left|(\eta-\eta')(S^y_{\eta'} M(f))\right| \right)$$

for some collection of Markov transitions $S^y_{\eta'}$ from E into itself. It is also readily checked that

$$\left|\Delta'_{n+1,u}(\eta,\eta')(f)\right| \leq \frac{hrdm^+}{\mathcal{Y}_n^c(w^-)} \int \mathcal{Y}_n(dy) \frac{1}{(h+m^-dg^-(y))^2} \left|(\eta-\eta')(g(\centerdot,y))\right|$$

This clearly implies that condition (8.4.3) is satisfied with

$$a^2_{n,n+1} \leq \frac{1}{\mathcal{Y}_n^c(w^-)} \left(\beta(M) \left[r(1-d)m^+ + \mathcal{Y}_n\left(\frac{w^+g^+}{g^-}\right)\right] + hrdm^+ \mathcal{Y}_n\left(\frac{g^+-g^-}{(h+m^-dg^-)^2}\right)\right)$$

We analyze the continuity properties of the mappings $u \mapsto \Phi^2_{n+1,u}(\eta)$ using the following decomposition

$$\Phi^2_{n+1,u}(\eta) - \Phi^2_{n+1,u'}(\eta)$$

$$= \tfrac{1}{\mathcal{Y}_n^c(w_u(\eta,\centerdot))} \int \mathcal{Y}_n^c(dy) \; \left[w_u(\eta,y) - w_{u'}(\eta,y)\right] \; \left(\Psi_{g(\centerdot,y)}(\eta) M_y - \Phi^2_{n+1,u'}(\eta)\right)$$

This implies that

$$\left|\left[\Phi^2_{n+1,u}(\eta) - \Phi^2_{n+1,u'}(\eta)\right](f)\right| \leq \tfrac{1}{\mathcal{Y}_n^c(w^-)} \left[r(1-d) + hrd \; \mathcal{Y}_n\left(\tfrac{g^+}{(h+dm^-g^-)^2}\right)\right] \; |u-u'|$$

This shows that condition (8.4.5) is satisfied with

$$\tau^2_{n+1} \leq \frac{1}{\mathcal{Y}_n^c(w^-)} \left[r(1-d) + hrd \; \mathcal{Y}_n\left(\frac{g^+}{(h+dm^-g^-)^2}\right)\right]$$

This ends the proof of the theorem. ∎

Chapter 9

Summary, Conclusions and Future Works

9.1 Summary

This thesis addressed the measure-valued processes arising in multi-object filtering and the stochastic particle models adopted to approximate their solutions.
The starting point is the random finite set approach, which has been proposed by Ronald Mahler and considerably studied and extended during the past ten years demonstrating theoretical soundness and practical validity.
In this context, we first focused on the performance of PHD filters with a set of numerical studies that opened the door to a series of improvements and to alternative implementations, then we analyzed their mathematical structure.

We started by considering three-dimensional aerial and naval scenarios provided by the French naval defence company DCNS and compared the performance of PHD filters in situations similar to those encountered in real-life. Results demonstrated that PHD filters provide good estimates when applied to realistic multi-target tracking, in particular the GM-PHD, which outperformed the SMC-PHD in both filtering quality and computational cost in all the scenarios we have examined.

A second contribution consists in the study of an unbiased stochastic resampling algorithm for the pruning step of the GM-PHD filter. The main goal was to assess the performance of the algorithm originally proposed by Fearnhead and Clifford [41] for the automatic determination of a pruning threshold for the GM-PHD, choice that is usually left to the user. Unfortunately, despite the property of unbiasedness, Monte Carlo validations demonstrated that deterministic strategies always outperform the Fearnhead-Clifford resampling, and thus they should be preferred.

The analysis of the strengths and weaknesses of PHD filters led to further studies in the direction of alternative implementations of the PHD recursion. One of the contribution in this respect is a method to calculate the Probability Hypothesis Density function over a grid by using the convolution method and the Fast Fourier Transform. This method provides an exact representation of the PHD function over the discretized domain and doesn't require Gaussian assumptions. However due to the use of numerical grids it is computationally more expensive than the GM-PHD filter. Despite the greater computational cost however, it shows remarkable properties and provides a very accurate representation of the intensity function.

The second part of the dissertation is focused on the analysis of a class of measure-valued equations of which the PHD recursion is an instance. Despite the increasing popularity of these multi-object filters, there are still open questions regarding their theoretical performance. One of the reasons is that their mathematical structure is significantly different from standard particle filters and theoretical results developed for these methods are not directly applicable. For this reason we aimed at initiating a thorough theoretical study of non-standard particle methods by first characterizing the stability properties of signal processes and then by establishing uniform convergence results for their particle approximation. In the case of the PHD, the signal is a spatial branching point process and particle methods are designed to approximate its first moment.

Another contribution is an alternative derivation of the PHD recursion based on the expression of the conditional distribution of a latent Poisson point process given an observed Poisson point process. This result opened the door to the characterization of the processes arising in multi-object filtering as a generalization of Feynman-Kac measures. Based on this alternative derivation we proposed a novel particle implementation of the PHD filter that avoids the computational burden related to clustering, and improves the overall quality of the estimates.

9.2 Future directions

Given the wide range of applicability of multi-target tracking and the great number of engineering problems related to the recursive estimation of the state of multiple objects, many questions remain open.
The number of works and publications appeared in the literature during the past ten years consolidate the generality and the validity of the random finite set approach as a proper basis for the study of multi-object systems. As it is natural, however, many pertinent questions can still be posed concerning both theoretical aspects and algorithmic developments.
One interesting direction, for example, has been proposed in [135] and concerns the

joint multi-object state and trajectory estimation: as the PHD filters provide only point estimates of the targets, it would be interesting to investigate the problem of joint multi-object state and trajectory estimation in order to obtain "connected" estimates of objects and trajectories.
Another interesting line of research is related to the *track-before-detect* (TBD) problem, a direction that has been suggested in [135] and initially investigated in [134]. In filtering applications, in fact, most of the time targets are modelled as points and this is often a quite valid assumption. However, there also exist situations in which this assumption leads to performance degradation or divergence of the algorithms. Theoretically, the general random set framework provides the mathematical tools to deal with this problem, the solution however is so expensive in terms of numerical computation that the search for principled approximations is necessary.
This line of research can be extended towards the application of the PHD, CPHD, Multi-Bernoulli filters to the tracking of complex objects (such as cars, people, animals etc.) from image sequences. In those cases the application of the techniques discussed in this dissertation is extremely challenging because the fundamental assumption of the observations as RFS is, in some way, invalidated.
In addition, even if the foundation of the PHD recursion is the general RFS framework, these filters remain inherently single-sensor in their original formulation. Recent works such as [83, 35] have contributed to shed some light onto multi-sensor PHD and CPHD recursion which could be potentially computationally tractable. Computational techniques to implement these generalizations are only marginally investigated and could be very important to the applicability of the filters to a wider set of problems.
From an engineering point of view, the investigation of numerical techniques to address practical application of the PHD filters and smoothers to the tracking of extended or unresolved targets (Sec. 4.9) would be important as well.
In terms of theoretical developments, a characterization of the performance limits for multi-object systems would be a fundamental contribution to the field. The study of the theoretical bounds established in this thesis as well as the generalization of existing bounds would be an interesting line of research towards the construction of algorithms with precise convergence rates and error bounds.
An unresolved question is whether the PHD filter can be generalized so that a Poisson approximation is no longer necessary. In particular an interesting question would be whether it is possible to utilize Feynman-Kac techniques to analyze general multi-object systems where the update step does not contain Poisson assumptions, or if it is possible to generalize some of the results presented in the second part of this thesis in order to put under a common measure-theoretic umbrella different multi-object filters based on the propagation of intensity measures.

Appendices

Appendix A

Finite Set Statistics

This section summarizes concepts of set-integrals and set-derivatives in the finite set statistics (FISST) framework. Detailed discussion about the mathematical foundations can be found in [133, 42]. The individual target motion in a multi-target problem is often modelled by a transition density on the single-target state space E_s while the measurement process is modelled as a likelihood on the single-target observation space E_o. Consequently, it is very hard to construct multi-target transition density and likelihood as Radon-Nikodým derivatives of probability measures on the Borel subsets of $\mathcal{F}(E_s)$ and $\mathcal{F}(E_o)$. For this reason FISST introduces a non-measure theoretic notion of density defined directly on the closed subsets of E_s and E_o through set-integrals and set derivatives [42].
Let $\mathcal{C}(E)$ denote the collection of closed subsets of E and $F : \mathcal{C}(E) \to [0, \infty)$. A simplified version of the set-derivative of F at the point $x \in E$ is the mapping $(dF)_x : \mathcal{C}(E) \to [0, \infty)$ defined as:

$$(dF)_x(S) \equiv \lim_{\lambda_K(\Delta_x)} \frac{F(S \cup \Delta_x) - F(S)}{\lambda_K(\Delta_x)} \tag{A.0.1}$$

where $\lambda_K(\Delta_x)$ is the volume (Lebesgue measure) of a neighbourhood Δ_x of x in units of K (note $\lambda_K = K\lambda$). The complete definition is in [42]. The set-derivative at a finite set $X = \{x_1, \cdots x_n\}$ is defined by recursion:

$$(dF)_{\{x_1, \cdots x_n\}}(S) \equiv (d(dF)_{\{x_1, \cdots x_{n-1}\}})_{x_n}(S) \tag{A.0.2}$$

where, by convention $(dF)_{\{\emptyset\}} \equiv F$. An important point to note is that $(dF)_X(S)$ has unit of $K^-|X|$, hence for a fixed $S \subseteq E$ the set-derivatives $(dF)_X(S)$ and $(dF)_Y(S)$ have different units if $|X| \neq |Y|$. Let f be a functional defined by $f(X) = (dF)_X(\emptyset)$ then the set-integral of f over a closed subset $S \subseteq E$ is defined as follows [42]:

$$\int_S f(X)\delta X \equiv \sum_{i=0}^{\infty} \frac{1}{i!} \int_{S^i} f(\{x_1, \ldots, x_i\}) \lambda_K^i(dx_1, \ldots, dx_i) \tag{A.0.3}$$

The set-integral and set-derivative are related by the following *generalised fundamental theorem of calculus*

$$f(X) = (dF)_X(\emptyset) \qquad F(S) = \int_S f(X)\delta X \tag{A.0.4}$$

The relationship between finite set-statistics (FISST) and conventional probability theory has been established in [133] and allows the construction of the conditional densities from the underlying physical model of the sensors, individual target dynamics, target births and deaths using the tools of FISST.

A.1 Belief functional

One of the central entities in finite set statistics (FISST) is the *belief functional* (or *belief-mass function*) which is a useful descriptor of a RFS [80]. The belief functional β_X of an RFS X is defined by [42]:

$$\beta_X(S) = \mathbb{P}(X \subseteq S) \tag{A.1.1}$$

for all closed $S \subseteq E$. As previously mentioned, the fact that FISST is based on belief mass functions defined directly on the closed subsets of E_s and E_o allows descriptions of multi-target motion and measurements models to be systematically constructed from the single-target motion and observation models respectively. The belief function, however, is not a measure and hence the standard measure theoretic notion of a density is not applicable. To circumvent this difficulty, the theory of FISST provides an alternative notion of density for belief functionals via the constructs of set-integrals and set-derivatives.

A.2 Measure theoretic formulation

As the object of interest in Bayesian estimation is the posterior probability density, the application of Bayesian reasoning to multitarget estimation is based on a suitable notion of probability density for random finite sets. The probability density p_X of a RFS X is given, if it exist, by the Radon-Nikodým derivative of the probability distribution P_X with respect to an appropriate dominating measure μ, i.e.

$$P_X(\mathcal{T}) = \int_{\mathcal{T}} p_X(\Xi)\mu(d\Xi) \tag{A.2.1}$$

for any Borel subset $\mathcal{T} \subseteq \mathcal{F}(E)$.
The conventional choice of reference measure in point process theory [50] is the

dimensionless measure given by the unnormalized distribution of a Poisson point process [133]:

$$\mu(\mathcal{T}) = \sum_{i=1}^{\infty} \frac{\lambda^i(\chi^{-1}(\mathcal{T}) \cap E^i)}{i!} \tag{A.2.2}$$

where E^i is i-th Cartesian product of E with the convention $E^0 = \{\emptyset\}$, λ^i the is the i-th product dimensionless Lebesgue measure and $\chi : \uplus_{i=0}^{\infty} E^i \to \mathcal{F}(E)$ is a mapping of vectors to sets defined by $\chi(x_1, \ldots, x_i) = \{x_j : j = 1 \ldots i\}$. Note that $p_{\Xi}(X)$, unlike the usual probability density on the Euclidean space which has the dimension of probability per unit hyper-volume, is dimensionless since the reference measure is dimensionless.
The integral of a non-negative function $f : \mathcal{F}(E) \to \mathbb{R}$ over a subset $\mathcal{T}$ of $\mathcal{F}(E)$ with respect to the measure μ is given by [133]:

$$\int_{\mathcal{T}} f(X)\mu(dX) = \sum_{i=0}^{\infty} \frac{1}{i!} \int 1_{\mathcal{T}}(\chi(x_1, \ldots, x_i)) f(\{x_1, \ldots, x_i\}) \lambda^i(dx_1, \ldots, dx_i) \tag{A.2.3}$$

where $1_{\mathcal{T}}$ is the indicator function for $\mathcal{T}$.

Proposition A.2.1. *Given a RFS X on E with probability distribution P_X and belief mass function β_X, if P_X is absolutely continuous with respect to μ, the unnormalised distribution of a Poisson point process with rate K^{-1} then:*

$$\frac{dP_X}{d\mu}(\Xi) = K^{|\Xi|}(d\beta_X)_{\Xi}(\emptyset) \tag{A.2.4}$$

In other words the set-derivative of the belief mass function β_X without its unit is the probability density p_X with respect to the dominating measure μ given in [133], or, the unitless set-derivative of the belief mass function of a RFS is its probability density. It is important to note that the probability density p_X is unit dependent, since the dominating measure μ depends on the choice of units.

A.3 Moments

The moments of an RFS are important characterizations of the process. The **first moment** of a RFS is the analogue of the expectation of a random vector, however, since there is no notion of addition for sets the expectation of a RFS has no meaning. Nevertheless, it can be indirectly constructed by representing the RFS as a random counting measure or random density function as described in 4.4. The first moment measure M of a RFS X on E (commonly called *intensity measure*) is defined for any subset of the space $S \subseteq E$ by:

$$M(S) = \mathbb{E}\left[N(S)\right] \tag{A.3.1}$$

and gives the expected number of points of X contained in S. The intensity measure plays a central role in online filtering as it constitutes a sufficient statistics for the posterior probability of the targets RFS. If the measure M admits a density $\gamma(x)$:

$$M(S) = \int_S \gamma(x)dx = \mathbb{E}\left[|X \cap S|\right] \tag{A.3.2}$$

for any $S \subseteq E$ and $\gamma : E \to [0, +\infty[$, then γ is called in point process theory *intensity function*. For convenience sometimes the intensity function is simply referred to as the intensity.
The intesity function is more commonly known in multi-target tracking as Probability Hypothesis Density (PHD) function. Intuitively, the value $\gamma(x)dx$ is the expected number of points in an infinitesimally small region dx of x, i.e the expected target density at x. Just as the density of a continuous random vector represents the zero-probability event of a particular realization of the random vector, the intensity $\gamma(x)$ represents the zero-probability event $\mathbb{P}(x \in X)$. Consequently, $\gamma(x)$ is usually multimodal and the peaks are the regions of high target intensity.

Higher order moment measures M^k of an RFS X on E are defined for any $S_1\times, \ldots, \times, S_k \subseteq E$ by:

$$M^k(S_1\times, \ldots, \times, S_k) = \mathbb{E}\left[\sum_{x_1\neq,\ldots,\neq x_k \in X} 1_{S_1\times,\ldots,\times,S_k}(x_1, \ldots, x_k)\right] \tag{A.3.3}$$

In analogy to the first moment case, if M^k admits a density:

$$\gamma^k(S_1\times, \ldots, \times, S_k) = \int_{S_1} \ldots \int_{S_k} (x_1, \ldots, x_k)dx_1, \ldots, dx_k \tag{A.3.4}$$

for any $S_1\times, \ldots, \times, S_k \subseteq E$ and $\gamma^k : E^k \Rightarrow [0, \infty)$ then E^k is the k-order (multitarget) moment density. It is useful to consider also the FISST definition [77] of the multitarget moment densities which is as follow. Let X be a RFS with probability density f_X its multitarget moment density is defined as:

$$D_X(\Xi) = \int f_X(\Xi \cup W)\delta W \tag{A.3.5}$$

where $D_\Xi(\emptyset) = 1$ and the integral is a set-integral. Note that $\int D_X(\Xi \cup W)\delta W$ has always the same units of measurement as Ξ and so there is no incommensurability of units.
By using this definition and $\delta_X(x) = \sum_{w\in X} \delta_w(x)$ the first multitarget moment is

$$D_X(\{x\}) = \int f_\Xi(\{x\} \cup W)\delta W = \int \delta_X(x) f_\Xi(X)\delta X \tag{A.3.6}$$

and higher order moment analogously:

$$D_\Xi(\{x_1, \ldots, x_n\}) = \int f_\Xi(\{x_1, \ldots, x_n\} \cup W)\delta W \tag{A.3.7}$$

It is easy to see that the first-order multitarget moment is the PHD as:

$$\int_S D_\Xi(\{x\}) = \mathbb{E}(|X \cap S|) \tag{A.3.8}$$

A.4 Probability Generating Functionals

The probability generating functional for a RFS is a fundamental descriptor analogous to the probability generating function of a discrete random variable.
The original derivation of the PHD filter as well as many other important results are stated by using probability generating functionals (PGFl).
The PGFl $G[h]$ of an RFS X on E [80] is defined for any real-valued function h on E such that $0 \leq h(x) \leq 1$ by :

$$G[h] \equiv \mathbb{E}[h^X] = \int h^\Xi f_X(\Xi)\delta\Xi \tag{A.4.1}$$

where

$$h^X \equiv \prod_{x \in X} h(X) \tag{A.4.2}$$

with $h^\emptyset = 1$ by convention. It can be easily seen that:

$$G_X[\mathbf{1}_S] = \beta_X(S) \tag{A.4.3}$$

The PGFL shares the following useful property with the belief-mass function. Let $X_1, \ldots, X_n$ be statistically independent RFS with PGFLs $G_1[h], \ldots, , G_n[h]$ and $X = X_1 \cup \ldots \cup X_n$. Then, for all h:

$$G_X[h] = G_1[h] \cdots G_n[h] \tag{A.4.4}$$

Following [62] an interpretation of the Probability Generating Functional is given by supposing $A_1, \ldots, A_r$ is a partition of the space E, and considering the function:

$$h(x) = \sum_{i=1}^{r} y_i 1_{A_i}(x)$$

where $|y_i| \leq 1$. It follows that:

$$G\left[\sum_{i=1}^{r} y_i 1_{A_i}(x)\right] = \mathbb{E}\left[\prod_{i=1}^{r} y_i^{|X \cap A_i|}\right]$$

which results in the joint probability generating function of the number of points in the sets of the partition. Intuitively, an arbitrary function h can be considered as a limiting case of this form where the partition is generated by the collection of all the infinitesimal regions dx. The PGFl captures the probability generating functions of

all possible families of counts of point occurrences.
In the context of multi-target tracking Mahler [76] provided the following alternative intuition of the PGFl: let Ξ be a random finite set and $0 \leq h(x) \leq 1$ a function representing the probability of detection or FOV of some sensor, then it can be shown that $G_{\Xi}[h]$ is the probability that Ξ is contained in the FOV.

A.5 Densities and units of measurement

The behavior of RFS density functions, as opposed to conventional densities is dependent on the units of measurements. For example, let $f_Z(\cdot)$ be the density of the random vector $Z \in \mathbb{R}^n$; then $\forall\ z \in \mathbb{R}^n$, $f_Z(z)$ has the units of a density, for example $\frac{1}{\text{metres}}$ if the coordinates of z are in meters. Now, let Σ be an absolutely continuous finite random set of $\mathbb{R}^n$. The units of the global density $f_{\Sigma}(\cdot)$ vary with the cardinality of the realization σ of the RFS: for example suppose that the units of $\mathbb{R}^n$ are meters and that $|\sigma| = k$. Then the units of $f_{\Sigma}(\sigma)$ are $1/\text{metres}^k$. Because of this behaviour one should use a great care when dealing with mathematical operations which may be undefined in the case of RFSs. As an example, consider the well defined expression:

$$\int f_Z(z)^2 d\lambda(z)$$

which has unit 1/metres. The analogous set-integral however is not well defined as:

$$\int f_{\Sigma}(\sigma)^2 \delta\sigma = f_{\Sigma}(\emptyset) + \int f_{\Sigma}(\{\sigma_0\})^2 d\lambda\sigma_0 + \frac{1}{2}\int f_{\Sigma}(\{\sigma_1, \sigma_2\})^2 d\lambda\sigma_1 d\lambda\sigma_2 \cdots$$

where the first term is unitless, the second 1/metres the third $1/\text{metres}^2$ and so on.

As reported in [42] one example may clarify the structure of the RFS densities and belief functions. Suppose that two targets T_1 and T_2 are located on the real line and observed by a sensor with density

$$f(a|x) = \mathbb{N}_{\sigma^2}(a - x)$$

and probability measure $p_f(S|x)$. Assume that the sensor reports indipendent observation, with no false alarms but with a detection probability q lower than 1. What the sensor sees at each time step is described by a random finite set which may be as follows:

$$\begin{aligned} Z &= \{a_1, a_2\} \\ Z &= \{a\} \\ Z &= \{\emptyset\} \end{aligned}$$

Where a $a, a_1, a_2 \in \mathbb{R}$. The statistics of this RFS can be described by the belief measure $\beta_\Sigma(S|X) = P(\Sigma \subseteq X)$ where X can take the values $X = \{\emptyset\}$, $X = \{x\}$,$X = \{x_1, x_2\}$. In those cases

$$\begin{aligned} \beta_\Sigma(S|x) &= 1 - q + qp_f(S|x) \\ \beta_\Sigma(S|\{x_1, x_2\}) &= [1 - q + qp_f(S|x_1)][1 - q + qp_f(S|x_2)] \end{aligned}$$

From the belief measure it is possible to compute the global density: for the one track case $X = \{x\}$ it is

$$f_\Sigma(\emptyset|X) = 1 - q, \qquad f(\{z\}|X) = qf(z|x)$$

In the two track case $X = \{x_1, x_2\}$ it is

$$f_\Sigma(\emptyset|X) = (1 - q)^2$$

$$f_\Sigma(\{z\}|X) = q(1 - q)f(z|x_1) + q(1 - q)f(z|x_2)$$

$$f_\Sigma(\{z_1, z_2\}|X) = q^2 f(z_1|x_1)f(z_2|x_1) + q^2 f(z_2|x_2)f(z_1|x_2)$$

In all other cases $f_\Sigma(Z|X) = 0$.

The belief measure $\beta_\Sigma(S|X)$ can be recovered from the global density via the set integral

$$\begin{aligned} \beta_\Sigma(S|X) &= \int_S f_\Sigma(Z|X)\delta X \\ &= f_\Sigma(\emptyset|X) + \int f_\Sigma(\{z\}|X)^2 dz + \frac{1}{2}\int_{S\times S} f_\Sigma(\{z_1, z_2\})^2 dz_1 dz_2 \end{aligned}$$

Appendix B

Additional Background Material

B.1 Probability and measure theory

The following is a summary of measure theory and measure theoretic probability concepts. Further details can be found in [62] and [27].

B.1.1 Classes of sets

Let Ω be a nonempty set and $\mathcal{P}(\Omega) \equiv \{A : A \subset \Omega\}$ be the power set of Ω, i.e., the class of all subsets of Ω.

Definition B.1.1. *A collection of sets $\mathcal{F} \subset \mathcal{P}(\Omega)$ is called an algebra if:*

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

In words, an algebra is a class of sets containing Ω that is closed under complementation and pairwise (and hence finite) unions.

Definition B.1.2. *A class $\mathcal{F} \subset \mathcal{P}(\Omega)$ is called a σ-algebra if it is an algebra and if it satisfies the following condition:*

- $A_n \in \mathcal{F}$ *for* $n \geq 1 \Rightarrow \bigcup_{n\geq 1} A_n \in \mathcal{F}$

Thus, a σ-algebra is a class of subsets of Ω that contains Ω and is closed under complementation and countable unions.

Example:
Let $\Omega = a, b, c, d$ and consider the classes:

$$\mathcal{F}_1 = \{\Omega, \emptyset, \{a\}\}$$

$$\mathcal{F}_2 = \{\Omega, \emptyset, \{a\}, \{b, c, d\}\}$$

$\mathcal{F}_2$ is a σ-algebra but not $\mathcal{F}_1$ since $\{a\}^c \notin \mathcal{F}_1$.

Definition B.1.3. *A topological space is a pair $(\mathbb{S}, \mathcal{T})$ where $\mathbb{S}$ is a nonempty set and $\mathcal{T}$ is a collection of subsets of $\mathbb{S}$ such that:*

- $\mathbb{S} \in \mathcal{T}$
- $A_1, A_2 \in \mathcal{T} \Rightarrow A_1 \cap A_2 \in \mathcal{T}$
- $\{A_i : i \in I\} \in \mathcal{T} \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{T}$

Elements of $\mathcal{T}$ are called open sets.

A particularly useful class of σ-algebras are those generated by open sets of a topological space. These are called Borel σ-algebras

Definition B.1.4. *The Borel σ-algebra on a topological space $\mathbb{S}$ is defined as the σ-algebra generated by the collection of open sets in $\mathbb{S}$.*

B.1.2 Measures

A set function is an extended real valued function defined on a class of subsets of a set Ω. Measures are *nonnegative set functions* that, intuitively speaking, assign a measure to the content of a subset of Ω. They have to satisfy certain requirements:

Definition B.1.5. *Let Ω be a nonempty set and $\mathcal{F}$ be an algebra on Ω. Then, a set function μ on $\mathcal{F}$ is called a measure if:*

- $\mu(A) \in [0, \infty]$ *for all* $A \in \mathcal{F}$
- $\mu(\emptyset) = 0$
- *for any disjoint collection of sets* $A_1, \ldots, A_n \in \mathcal{F}$ *with* $\bigcup_{n \geq 1} A_n \in \mathcal{F}$

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

Definition B.1.6. *A measure μ is called finite if $\mu(\Omega) \leq \infty$ or infinite if $\mu(\Omega) = \infty$. A finite measure with $\mu(\Omega) = 1$ is called a probability measure. A measure μ on a σ-algebra $\mathcal{F}$ is called σ-finite if there exist a countable collection of sets $A_1, \ldots, A_n \in \mathcal{F}$ not necessarily disjoint, such that*

- $\bigcup_{n\geq 1} A_n = \Omega$
- $\mu(A_n) \leq \infty \ \forall \ n \geq 1$

Definition B.1.7. ***(The counting measure)*** *Let Ω be a nonempty set and $\mathcal{F} = \mathcal{P}(\Omega)$ be the set of all subsets of Ω. Let*

$$\mu(A) = |A|, A \in \mathcal{F}$$

where $|A|$ denotes the number of elements in A. This measure is called counting measure on Ω.

B.1.3 Measurable transformations

Sometimes the interest is posed only on certain functions defined on Ω rather than on the full details of a measure space $(\Omega, \mathcal{F}, \mu)$. For example, if Ω represents the outcomes of 100 rolls of a dice, one may only be interested in knowing the number of "ones" in the 100 rolls. By assigning measures (probabilities) to sets (events) involving such functions, only certain functions (called measurable functions) that satisfy some natural restrictions, are allowed.

Definition B.1.8. *Let Ω be a nonempty set and let $\mathcal{F}$ be a σ-algebra on Ω. The pair $(\Omega, \mathcal{F})$ is called a measurable space. If μ is a measure on $(\Omega, \mathcal{F})$, then the triplet $(\Omega, \mathcal{F}, \mu)$ is called a measure space. If μ is a probability measure, then $(\Omega, \mathcal{F}, \mu)$ is called a probability space.*

Definition B.1.9. *Let $(\Omega, \mathcal{F})$ be a measurable space. Then a function $f : \Omega \to \mathbb{R}$ is called $\mathcal{F}$-measurable if for each a in $\mathbb{R}$*

$$f^{-1}((-\infty, a) =\equiv \{\omega : f(\omega) \leq a\} \in \mathcal{F}$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then a function $X : \Omega \to \mathbb{R}$ is called random variable if the event

$$X^{-1}((-\infty, a) =\equiv \{\omega : X(\omega) \leq a\} \in \mathcal{F}$$

for each $a \in \mathbb{R}$. In other words a random variable is a real valued $\mathcal{F}$-measurable function on a probability space $(\Omega, \mathcal{F}, P)$.

B.1.4 Integration

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be a measurable function. The integral of a measurable function is defined in stages by starting from the integral of a non-negative simple (or step) function:

Definition B.1.10. *A function $f : \Omega \to \mathbb{R} \equiv [-\infty, \infty]$ is a simple function if there exist a finite set (of distinct elements) $\{c_1, ..., c_k\} \in \mathbb{R}$ and sets $A_1, ..., A_k \in \mathcal{F}$ such that f can be written as:*

$$f = \sum_{i=1}^{k} c_i I_{A_i}$$

Definition B.1.11. ***(The integral of a simple nonnegative function)*** *Let $f : \Omega \to \mathbb{R} \equiv [-\infty, \infty]$ be a simple function on $(\Omega, \mathcal{F}, \mu)$. The integral of f w.r.t. μ, denoted by $\int f d\mu$ is defined as*

$$\int f d\mu \equiv \sum_{i=1}^{k} c_i \mu_{A_i}$$

Definition B.1.12. ***(The integral of a nonnegative measurable function)*** *Let $f : \Omega \to \mathbb{R}_+$ be a nonnegative measurable function on $(\Omega, \mathcal{F}, \mu)$. The integral of f w.r.t. μ, denoted by $\int f d\mu$ is defined as*

$$\int f d\mu \equiv \lim_{n\to\infty} \int f_n d\mu$$

where $\{f_n\}_{n\geq 1}$ is any sequence of nonnegative simple functions such that $f_n(\omega) \uparrow f(\omega)$ for all ω.

Definition B.1.13. ***(The integral of a measurable function)*** *Let f be a real valued measurable function on $(\Omega, \mathcal{F}, \mu)$ and $f^+ = fI_{f\geq 0}$ and $f^- = fI_{f\leq 0}$. The integral of f w.r.t. μ, denoted by $\int f d\mu$ is defined as*

$$\int f d\mu \equiv \int f^+ d\mu - \int f^- d\mu$$

provided that at least one of the integrals on the right side is finite.

Definition B.1.14. ***(Integrable functions)*** *A measurable function f on a measure space $(\Omega, \mathcal{F}, \mu)$ is said to be integrable with respect a measure μ if*

$$\int |f| d\mu < \infty$$

Remark on notation: $\int f d\mu$ can be equivalently written as

$$\int_\Omega f(\omega)\mu(d\omega)$$

or

$$\int_\Omega f(\omega) d\mu(\omega)$$

B.2 Gaussian Identities

A list of important results on Gaussian functions are reported below. Gaussian identities are implicitly used extensively in the dissertation, and used in the derivation of the Kalman recursion from a Bayesian perspective. More details can be found in [108].

Lemma B.2.1. *The product of two gaussians is another gaussian (unnormalized):*

$$\mathcal{N}_x(\mu_a, \Sigma_a) \cdot \mathcal{N}_x(\mu_b, \Sigma_b) = z_c \mathcal{N}_x(\mu_c, \Sigma_c)$$

where:

$$\Sigma_c = (\Sigma_a^{-1} + \Sigma_b^{-1})^{-1} \quad \textit{and} \quad \mu_c = \Sigma_c(\Sigma_a^{-1}\mu_a + \Sigma_b^{-1}\mu_b)$$

and the normalization factor z_c is

$$z_c = |2\pi(\Sigma_a + \Sigma_b)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_a - \mu_b)^T(\Sigma_a + \Sigma_b)^{-1}(\mu_a - \mu_b)\right)$$

The obvious generalization to K gaussians is:

$$\prod_{k=1}^{K} \mathcal{N}_x(\mu_k, \Sigma_k) = \tilde{z}_c \mathcal{N}_x(\tilde{\mu}, \tilde{\Sigma})$$

where:

$$\tilde{\Sigma} = \sum_{k=1}^{K} \Sigma_k^{-1} \quad \textit{and} \quad \tilde{\mu} = \left(\sum_{k=1}^{K} \Sigma_k^{-1}\right)^{-1} \left(\sum_{k=1}^{K} \Sigma_k^{-1}\mu_k\right)$$

The normalizing constant is given by:

$$\tilde{z} = \frac{|2\pi\Sigma_d|^{\frac{1}{2}}}{\prod_{k=1}^{K} |2\pi\Sigma_k|^{\frac{1}{2}}} \prod_{i<j} \exp\left(-\frac{1}{2}(\mu_i - \mu_j)^T B_{ij}(\mu_i - \mu_j\right)$$

where:

$$B_{ij} = \Sigma_i^{-1} \left(\sum_{k=1}^{K} \Sigma_k^{-1}\right)^{-1} \Sigma_j^{-1}$$

The fact that products of gaussian functions are again a gaussian function makes gaussian integrals easier to calculate:

Lemma B.2.2. *Given F, d, Q, m, and P of appropriate dimensions, and given that Q and P are positive definite,*

$$\int \mathcal{N}(x; F\xi + d, Q) \cdot \mathcal{N}(\xi; m, P) = \mathcal{N}(x; Fm + d, Q + FPF^T)$$

Lemma B.2.3. *Given H, b, R, m, and P of appropriate dimensions, and given that R and P are positive definite,*

$$\int \mathcal{N}(z; Hx + b, R) \cdot \mathcal{N}(x; m, P) = q(z)\mathcal{N}(x; \tilde{m}, \tilde{P})$$

where

$$q(z) = \mathcal{N}(z; Hm + b, R + HPH^T), \tag{B.2.1}$$
$$\tilde{m} = m + K(z - Hm - b), \tag{B.2.2}$$
$$\tilde{P} = (I - KH)P, \tag{B.2.3}$$
$$K = PH^T(R + HPH^T)^{-1} \tag{B.2.4}$$
$$\tag{B.2.5}$$

B.3 Convolution

Denote by $L^p(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{C}, \text{ Borel measurable, } \int_{\mathbb{R}^d} |f|^p dm < \infty\}$ where $m(\cdot)$ is the Lebesgue measure and by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-algebra on $\mathbb{R}^d$ and $d \geq 1$. We begin by reporting the definition of the convolution of measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$; from this, one can easily obtain the formula for the convolution of functions in $L^1(\mathbb{R}^d)$ and convolution of functions with measures [62].

Proposition B.3.1. *Let μ and λ be two σ-finite measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$; for any Borel set A in $\mathcal{B}(\mathbb{R}^d)$*

$$(\mu * \lambda)(A) := \int \int \mathbf{I}_A(x + y)\mu(dx)\lambda(dy) \tag{B.3.1}$$

Then $(\mu * \lambda)(\cdot)$ is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and it is called the convolution of μ and λ.

Moreover, if μ and ν are two probability measures, the following holds:

Proposition B.3.2. *Let X and Y be two independent R^d random variables with pdf $\phi(x)$ and $\varphi(y)$ respectively. The pdf $\psi(z)$ of the random variable $Z = X + Y$ is given by the convolution product of $\phi(x)$ and $\varphi(y)$.*

The convolution between two complex-valued functions f and $g \in L^1(\mathbb{R}^d) := L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, denoted by $(f * g)$, is defined as the integral transform:

$$(f * g)(x) := \int_{\mathbb{R}^d} f(y)g(x - y)dy = \int_{\mathbb{R}^d} f(x - y)g(y)dy \tag{B.3.2}$$

The following propositions are easy to verify: let f , $g \in L^1(\mathbb{R}^d)$:

$$\begin{aligned} (f * g) &= (g * f) && \text{(B.3.3)} \\ a(f * g) &= (af) * g = f * (ag) && \text{(B.3.4)} \\ \int (f * g)(x)dx &= \left(\int f(x)dx\right)\left(\int g(x)dx\right) && \text{(B.3.5)} \end{aligned}$$

B.4 Fourier Transform

For notation simplicity we briefly report here only the monodimensional discrete formulation of the Fourier transform. Details on the multidimensional Fourier transform can be found in [11].

Definition B.4.1. *For $f \in L^1(\mathbb{R}), t \in \mathbb{R}$*

$$\hat{f}(t) := \int f(x)e^{-itx}dx \tag{B.4.1}$$

is called the Fourier transform of f.

One important property of the Fourier transform is the following, called Convolution Theorem:

Proposition B.4.2. *Let f and $g \in L^1(\mathbb{R})$, and $(f * g)$ their convolution,*

$$\widehat{(f * g)} = \hat{f}\hat{g} \tag{B.4.2}$$

The discrete Fourier transform of a sequence of real numbers is a sequence of complex numbers of the same length computed as follows:

Proposition B.4.3. *The sequence on N complex numbers $s_0 \dots s_{n-1}$ is transformed into a different sequence of N complex numbers $\hat{s}_0 \dots \hat{s}_{n-1}$ by the discrete Fourier transform (DFT) according to the formula:*

$$\hat{s}_k = \sum_{n=0}^{N-1} s_n e^{-\frac{2\pi i}{N}kn},\ k = 0, \dots, N-1. \tag{B.4.3}$$

The inverse discrete Fourier transform (IDFT) is given by

$$s_n = \frac{1}{N}\sum_{k=0}^{N-1} \hat{s}_k e^{\frac{2\pi i}{N}kn},\ n = 0, \dots, N-1. \tag{B.4.4}$$

The connection between the operation of convolution and the Fourier transform is given by the discrete case the Convolution theorem:

Proposition B.4.4. *Let $\widehat{X}_k$ and $\widehat{Y}_k$ $0 \leq k \leq N-1$ the DFT of N-periodic functions $x(n)$ and $y(n)$ respectively. The convolution of x and y can be obtained as the product of the individual transforms:*

$$\begin{aligned} F(x * y)(k) &= \widehat{X}(k)\widehat{Y}(k) \qquad k = 0 \dots N-1 \\ (x * y)(n) &= F^{-1}(\widehat{X} \cdot \widehat{Y})(n) \quad n = 0 \dots N-1 \end{aligned}$$

where F denotes the operation of Fourier transform.
Computing the DFT on N points by using the definition takes $O(N^2)$ arithmetical operations and it is often too slow to be practical. The Fast Fourier Transform algorithm [25] can reduce the complexity to $O(NlogN)$ operations.

Bibliography

[1] B. Anderson and J. Moore. *Optimal filtering.* Prentice-Hall, New York, 1979. [cited at p. 29]

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002. [cited at p. 32, 35]

[3] Y. Bar-Shalom. Tracking in a Cluttered Environment with Probabilistic Data Association. *Automatica*, pages 451–460, 1975. [cited at p. 4]

[4] Y. Bar-Shalom and T. Fortmann. *Tracking and Data Association.* Academic Press, San Diego, 1988. [cited at p. 36, 37, 39, 192]

[5] Y. Bar-Shalom and X. R. Li. *Multitarget-multisensor tracking: principles and techniques.* YBS Publishing, Storrs, CT, 1995. [cited at p. 37, 39]

[6] Y. Bar-Shalom, X. R. Li, and T. Kirubajan. *Estimation with applications to tracking and navigation.* John Wiley & Sons, Inc., New York, NY, USA, 2002. [cited at p. 29]

[7] B. Bercu, P. Del Moral, and A. Doucet. A functional central limit theorem for a class of interacting Markov Chain Monte Carlo methods. *Electronic Journal of Probability*, 14:2130–2155, 2009. [cited at p. 180]

[8] S. Blackman. *Multiple Target Tracking with Radar Applications.* Artech House, Norwood, 1986. [cited at p. 39, 192]

[9] S. Blackman. *Design and Analysis of Modern Tracking Systems.* Artech House, Norwood, 1999. [cited at p. 40]

[10] S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5–18, Jan. 2009. [cited at p. 4, 40]

[11] E. O. Brigham. *The Fast Fourier Transform.* Prentice-Hall, New York, 1988. [cited at p. 263]

[12] F. Caron. *Inférence bayésienne pour la détermination et la sélection de modèles stochastiques*. PhD thesis, Université des Sciences et Technologies de Lille, 2006. [cited at p. 29, 36]

[13] F. Caron, P. Del Moral, A. Doucet, and M. Pace. Particle approximations of a class of branching distribution flows arising in multi-target tracking. *To appear in SIAM Journal on Control and Optimization*, 2011. [cited at p. 190, 195]

[14] J. Carpenter, P. Clifford, and P. Fearnhead. Improved particle filter for nonlinear problems. *IEE Proceedings Radar Sonar and Navigation*, 146:2–7, 1999. [cited at p. 89]

[15] G. Casella and C. Robert. Rao-Blackwellisation of Sampling Schemes. *Biometrika*, 83(1):81–94, 1996. [cited at p. 36]

[16] F. Cérou, P. Del Moral, and A. Guyader. A non asymptotic variance theorem for unnormalized Feynman-Kac particle models. *Annales de l'Institut Henri Poincaré*, 2011. [cited at p. 161, 166, 179]

[17] D. Clark. *Multiple Target Tracking with The Probability Hypothesis Density Filter*. PhD thesis, Heriot-Watt University, 2006. [cited at p. 67]

[18] D. Clark. First-Moment Multi-Object Forward-Backward Smoothing. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 53]

[19] D. Clark and J. Bell. Bayesian Multiple Target Tracking in Forward Scan Sonar Images Using the PHD Filter. *IEE Radar, Sonar and Navigation*, 152(5):327–334, 2005. [cited at p. 6, 52]

[20] D. Clark and J. Bell. Data Association for the PHD Filter. In *Proc. International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, volume 55, pages 217–222, 2005. [cited at p. 52, 97]

[21] D. Clark and J. Bell. Convergence results for the particle PHD filter. *IEEE Transactions on Signal Processing,*, 54(7):2652 – 2661, 2006. [cited at p. 6]

[22] D. Clark, J. Bell, Y. D. Saint-Pern, and Y. Petillot. PHD filter for multi-target tracking in 3-D sonar. In *Proc. of Oceans European Conference*, pages 265–270, 2005. [cited at p. 6, 67]

[23] D. Clark and S. Godsill. Group Target Tracking with the Gaussian Mixture Probability Hypothesis Density Filter. *IEE Radar, Sonar and Navigation*, 152(5):327–334, 2005. [cited at p. 6]

[24] D. Clark and B.-N. Vo. Convergence Analysis of the Gaussian Mixture PHD Filter. *IEEE Transactions On Signal Processing*, 55(4):1204–1212, 2007. [cited at p. 6, 52, 57, 88]

[25] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. [cited at p. 264]

[26] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Process.* Springer Verlag, Berlin, Germany, 2nd edition, 2003. [cited at p. 41, 42, 46, 47, 148]

[27] A. DasGupta. *Asymptotic Theory of Statistics and Probability.* Springer, 2008. [cited at p. 257]

[28] P. Del Moral. *Feynman-Kac formulae. Genealogical and interacting particle systems with applications.* Springer Verlag, New York, 2004. [cited at p. 3, 14, 16, 119, 157, 161, 164, 166, 168, 173, 174, 177, 178, 180, 190, 192, 197, 198, 207, 213, 221]

[29] P. Del Moral and A. Doucet. Particle methods: An introduction with applications. *INRIA Research Report, RR-6991*, 2009. [cited at p. 14]

[30] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal Of The Royal Statistical Society*, 68(3):411–436, 2006. [cited at p. 35]

[31] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli Society Journal*, 47, 2010. [cited at p. 174]

[32] P. Del Moral, M. Pace, B.-N. Vo, and F. Caron. On the Stability and the Approximation of Branching Distribution Flows , with Applications to Nonlinear Multiple Target Filtering. *to appear in Stochastic Analysis and Applications*, 2010. [cited at p. 118, 120, 157]

[33] P. Del Moral, F. Patras, and S. Rubenthaler. *A Mean Field Theory of Nonlinear Filtering*, pages 1–39. Oxford University Press, Oxford, 2011. [cited at p. 212]

[34] P. Del Moral and E. Rio. Concentration inequalities for mean field particle models. *HAL-INRIA RR-6901 to appear in Annals of Applied Probability*, 2011. [cited at p. 184]

[35] E. Delande, E. Duflos, D. Heurguier, and P. Vanheeghe. Multitarget phd filtering: proposition of extensions to the multi-sensor case. *INRIA Research Report RR-7337*, 2010. [cited at p. 53, 245]

[36] E. Delande, E. Duflos, P. Vanheeghe, and D. Heurguier. Multi-sensor PHD: Construction and implementation by space partitioning. *36th International Conference on Acoustics, Speech and Signal Processing*, 2011. [cited at p. 53]

[37] A. Doucet, N. D. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag New York, 2001. [cited at p. 24, 32, 35, 156]

[38] A. Doucet and S. Godsill. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000. [cited at p. 35, 36]

[39] O. Drummond and B. Fridling. Ambiguities in evaluating performance of multiple target tracking algorithms. In *Proc. SPIE Signal and Data Processing of Small Targets*, pages 326–337, 1992. [cited at p. 60]

[40] E. Dynkin. *An Introduction to Branching Measure-Valued Processes*. American Mathematical Society, CRM, 1994. [cited at p. 192]

[41] P. Fearnhead and P. Clifford. Online Inference for Hidden Markov models. *Journal of the Royal Statistical Society*, 65:887–899, 2003. [cited at p. 88, 89, 99, 243]

[42] I. Goodman, R. Mahler, and H. Nguyen. *Mathematics of Data Fusion*. Kluwer Academic Publishers, 1997. [cited at p. 5, 41, 44, 47, 49, 156, 193, 249, 250, 254]

[43] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *Proc. IEE Radar and Signal Processing*, pages 107–113, 2002. [cited at p. 3, 32]

[44] K. Granstr, C. Lundquist, and U. Orguner. A Gaussian Mixture PHD filter for Extended Target Tracking Partitioning the Measurement Set. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 60]

[45] M. Guerriero, L. Svensson, D. Svensson, and P. Willett. Shooting two birds with two bullets: how to find Minimum Mean OSPA estimates. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 61]

[46] J. Hoffman and R. Mahler. Multitarget miss distance via optimal assignment. *IEEE Trans. Sys, Man, and Cybernetics-Part A*, 34(3):327–336, 2004. [cited at p. 61]

[47] J. Houssineau and D. Laneuville. PHD filter with diffuse spatial prior on the birth process with applications to GM-PHD filter The PHD filter. In *Proc. 13th Int. Conf. Information Fusion*, 2010. [cited at p. 75]

[48] W. Huang, N.-T. Pham, and S. Ong. Maintaining track continuity in GMPHD filter. In *International Conference on Information, Communications and Signal Processing*, 2007. [cited at p. 97]

[49] N. Ikoma, T. Uchino, and H. Maeda. Tracking of feature points in image sequence by SMC implementation of PHD filter. In *Proc. of SICE Annual Conference*, pages 1696 – 1701, 2004. [cited at p. 6]

[50] J. Moller. *Markov chain Monte Carlo and spatial point processes*, pages 141–172. Chapman and Hall, monographs edition, 1999. [cited at p. 250]

[51] J. Vermaak, S. Godsill, and P. Perez. Monte Carlo filtering for multi-target tracking and data association. *IEEE Trans. Aerospace and Electronic Systems*, 41(1):390–332, 2005. [cited at p. 40]

[52] A. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970. [cited at p. 28]

[53] A. Johansen, S. S. Singh, A. Doucet, and B.-N. Vo. Convergence of the sequential Monte Carlo implementation of the PHD filter. *Methodology and Computing in Applied Probability*, 8(2):265–291, 2008. [cited at p. 6, 157, 192]

[54] S. Julier. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 3086(3):110–422, Mar. 2004. [cited at p. 30]

[55] S. Julier and J. K. Uhlmann. A General Method for Approximating Nonlinear Transformations of Probability Distributions. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, 1996. [cited at p. 28, 32]

[56] S. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A New Approach for the Nonlinear Transformation of Means and Covariances in Linear Filters. *IEEE Transactions on Automatic Control*, 1996. [cited at p. 32]

[57] S. J. Julier and J. K. Uhlmann. A New Extension of the Kalman Filter to Nonlinear Systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, Orlando, Florida, 1997. [cited at p. 3, 28, 30]

[58] R. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME Journal of Basic Engineering*, 82(Series D):35–45, 1960. [cited at p. 3, 27]

[59] D. Kendall. *Foundations of a theory of random sets*, pages 322–376. J. Wiley, New York, 1963. [cited at p. 41]

[60] M. L. Kleptsyna and A. Y. Veretennikov. On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields*, 141:411–444, 2008. [cited at p. 190]

[61] A. Kong, J. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994. [cited at p. 36]

[62] S. N. L. Krishna B. Athreya. *Measure Theory and Probability Theory*. Springer Texts in Statistics, 2006. [cited at p. 253, 257, 262]

[63] D. Laneuville and J. Houssineau. Passive Multi Target Tracking with GM-PHD Filter. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 52]

[64] K. W. Lee, B. Kalyan, S. Wijesoma, M. Adams, F. S. Hover, and N. M. Patrikalakis. Tracking random finite objects using 3D-LIDAR in marine environments. *Proc. of the 2010 ACM Symposium on Applied Computing - SAC '10*, page 1282, 2010. [cited at p. 6]

[65] X. Li and V. Jilkov. Survey of maneuvering target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1333–1364, 2003. [cited at p. 48, 72]

[66] M. V. Lieshout and A. Baddeley. Extrapolating and interpolating spatial patterns. *Probability, networks and algorithms*, 17:1–26, 2001. [cited at p. 149]

[67] Lin, T. Kirubarajan, and Y. Bar-Shalom. Data Association Combined with the Probability Hypothesis Density Filter for Multitarget Tracking. *Proceedings of SPIE Signal and Data Processing of Small Targets*, 5428, 2004. [cited at p. 52]

[68] J. Lund, A. Penttinen, and M. Rudemo. Bayesian analysis of spatial point patterns from noisy observations. *Department of Mathematics and Physics, The Royal Veterinary and Agricultural University*, 1999. [cited at p. 149]

[69] J. Lund and M. Rudemo. Models for point processes observed with noise. *Biometrika*, 87:235–249, 2000. [cited at p. 148]

[70] J. Lund and E. Thonnes. Perfect simulation and inference for point processes given noisy observations. *Computational Statistics*, 19:317–336, 2004. [cited at p. 148]

[71] J. B. MacQueen. Some Methods for classification and Analysis of Multivariate Observations. *Proc. of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967. [cited at p. 67, 75]

[72] E. Maggio, E. Piccardo, C. Regazzoni, and A. Cavallaro. Particle PHD Filtering for Multi-Target Visual Tracking. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 1101–1104, 2007. [cited at p. 6]

[73] R. Mahler. A theoretical foundation for the Stein-Winter Probability Hypothesis Density (PHD) multi-target tracking approach. In *Proc. MSS Nat. Symp. on Sensor and Data Fusion*, 2000. [cited at p. 157, 193]

[74] R. Mahler. An Introduction to Multisource-Multitarget Statistics and Its Applications. *Lockheed Martin. Technical Monograph.*, 2000. [cited at p. 5, 41]

[75] R. Mahler. Random set theory for target tracking and identification. *Handbook of Multisensor Data Fusion*, 2001. [cited at p. 5]

[76] R. Mahler. Multitarget Bayes Filtering via First-Order Multitarget Moments. *Ieee Transactions On Aerospace And Electronic Systems*, 39(4):1152–1178, 2003. [cited at p. 5, 42, 49, 51, 157, 190, 193, 194, 195, 254]

[77] R. Mahler. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152–1178, Oct. 2003. [cited at p. 148, 149, 153, 252]

[78] R. Mahler. "Statistics 101" for multisensor, multitarget data fusion. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):53–64, Jan. 2004. [cited at p. 66]

[79] R. Mahler. PHD Filters of Higher Order in Target Number. *Ieee Transactions On Aerospace And Electronic Systems*, 43(4):1523 – 1543, 2007. [cited at p. 59]

[80] R. Mahler. *Statistical Multisource-Multitarget Information Fusion.* Artech House, 2007. [cited at p. 5, 41, 42, 53, 59, 192, 193, 250, 253]

[81] R. Mahler. PHD filters for nonstandard targets, I: Extended targets. In *Proc. 12th International Conference on Information Fusion*, pages 448–452, Seattle, WA, 2009. [cited at p. 60]

[82] R. Mahler. PHD filters for nonstandard targets, II: Unresolved targets. In *Proc. 12th International Conference on Information Fusion*, pages 922–929, Seattle, WA, 2009. [cited at p. 60]

[83] R. Mahler. Approximate multisensor CPHD and PHD filters. In *Proc. 13th International Conference on Information Fusion*, 2010. [cited at p. 53, 245]

[84] R. Mahler. Linear-complexity CPHD filters. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 53]

[85] R. Mahler. The Forward-Backward Probability Hypothesis Density Smoother. In *Proc. 13th International Conference on Information Fusion*, 2010. [cited at p. 53, 57]

[86] R. Mahler, J. Goutsias, and H. Nguyen. *Random Sets Theory and Applications.* Springer-Verlag New York, 1997. [cited at p. 5, 156, 157, 193]

[87] G. Matheron. *Random Sets and Integral Geometry.* J. Wiley, New York, 1975. [cited at p. 41, 42]

[88] J. Moller and R. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall CRC, 2004. [cited at p. 41]

[89] S. Mori, C. Chong, E. Tse, and R. Wishner. Tracking and identifying multiple targets without apriori identifications. *IEEE Trans. Automatic Control*, 21:401–409, 1986. [cited at p. 156, 193]

[90] S. Mori and C.-Y. Chong. Point process formalism for multiple target tracking. In *Proc. 5th Int. Conf. Information Fusion*, pages 10–17, Arlington, VA, 2002. [cited at p. 156]

[91] D. Musicki and R. Evans. Joint Integrated Probabilistic Data Association - JIPDA. *IEEE Trans. Aerospace and Electronic Systems*, 40(3):1093 – 1099, 2004. [cited at p. 40]

[92] D. Musicki, R. Evans, and S. Stankovic. Integrated probabilistic data association. *IEEE Trans. Automatic Control*, AC-39(6):1237–1241, 1994. [cited at p. 37]

[93] D. Musicki, B. LaScala, and R. Evans. Integrated track splitting filter - efficient multi-scan single target tracking in clutter. *IEEE Trans. Aerospace and Electronic Systems*, 43(4):1409–1425, 2007. [cited at p. 40]

[94] W. Ng, S. Godsill, J. Li, and J. Vermaak. A review of recent results in multiple target tracking. In *Proc. of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 40 – 45, 2005. [cited at p. 192]

[95] W. Ng, J. Li, S. Godsill, and J. Vermaak. Tracking variable number of targets using Sequential Monte Carlo Method. In *Proceedings of the European Signal Processing Conference*, 2005. [cited at p. 192]

[96] W. Ng, J. Li, S. Godsill, and J. Vermaak. Tracking variable number of targets using Sequential Monte Carlo Methods. In *Proc. 13th IEEE Workshop on Statistical Signal Processing*, pages 1286 – 1291, 2005. [cited at p. 192]

[97] M. Pace. Comparison of PHD based filters for the Tracking of 3D Aerial and Naval Scenarios. In *Proc. IEEE Radar Conference 2010*, pages 479 – 484, 2010. [cited at p. 6, 8]

[98] M. Pace, P. Del Moral, and F. Caron. Comparison of Implementations of Gaussian Mixture PHD Filters. In *Proc. 13th Int. Conf. Information Fusion*, 2010. [cited at p. 8]

[99] M. Pace and H. Zhang. Grid Based PHD Filtering by Fast Fourier Transform. Technical report, Working Paper, 2011. [cited at p. 8]

[100] K. Panta. *Multi-Target Tracking Using 1st Moment of Random Finite Sets*. PhD thesis, Department of Electrical and Electronic Engineering of The University of Melbourne Australia, 2007. [cited at p. 55, 59, 67, 92]

[101] K. Panta, D. Clark, and B.-N. Vo. Data Association and Track Management for the Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Transactions on Aerospace and Electronic Systems*, 45(3):1003, 2009. [cited at p. 52]

[102] K. Panta and B.-N. Vo. Convolution Kernels based Sequential Monte Carlo Approximation of the Probability Hypothesis Density (PHD) Filter. *Information, Decision and Control Conference*, pages 336–341, 2007. [cited at p. 97]

[103] K. Panta, B.-N. Vo, S. S. Singh, and A. Doucet. Probability Hypothesis Density filter versus Multiple Hypothesis Tracking. In *Proceedings of SPIE Signal Processing, Sensor Fusion, and Target Recognition XIII, Vol. 5429*, 2004. [cited at p. 52]

[104] K. Panta, Vo, Ba-Ngu, and S. Singh. Novel Data Association Schemes for the Probability Hypothesis Density Filter. *Ieee Transactions On Aerospace And Electronic Systems*, 43(2), 2007. [cited at p. 52]

[105] E. Pollard, B. Pannetier, G. Le Besnerais, and F. Champagnat. GM-PHD filters for multi-object tracking in uncalibrated aerial videos. In *Proc. 12th Int. Conf. Information Fusion*, pages 1171–1178, Seattle, WA, 2009. [cited at p. 6]

[106] G. Pulford and R. Evans. A multipath data association tracker for over-the-horizon radar. *IEEE Trans. Aerospace and Electronic Systems*, 34(4):1165–1183, 1998. [cited at p. 37]

[107] D. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843–854, Dec. 1979. [cited at p. 40]

[108] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artec House, 2004. [cited at p. 24, 32, 261]

[109] B. Ristic and D. Clark. Improved SMC implementation of the PHD filter. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 52]

[110] B. Ristic, B.-N. Vo, and D. Clark. Performance evaluation of multi-target tracking using the OSPA metric. In *Proc. 13th Int. Conf. Information Fusion*, 2010. [cited at p. 61]

[111] R. Rothrock and O. Drummond. Performance metrics for multiple-sensor, multiple-target tracking. In *Proc. Signal and Data Processing of Small Targets*, pages 521–531, Orlando, FL, USA, 2000. [cited at p. 60]

[112] S. Sarkka, A. Vehtari, and J. Lampinen. Rao-Blackwellized Particle Filter for Multiple Target Tracking. *Journal Information Fusion*, 8(1):2–15, 2007. [cited at p. 192]

[113] D. Salmond. Mixture reduction algorithms for target tracking in clutter. In *Signal and Data Processing of Small Targets, Proc. SPIE*, 1990. [cited at p. 37]

[114] S. Sarkka, A. Vehtari, and J. Lampinen. Rao-Blackwellized Monte Carlo data association for multiple target tracking. In *Proc. 7th Int. Conf. Information Fusion*, pages 583–590, 2004. [cited at p. 192]

[115] D. Schuhmacher, B.-N. Vo, and B.-t. Vo. A Consistent Metric for Performance Evaluation of Multi-Object Filters. *Ieee Transactions On Signal Processing*, 56(8):3447 – 3457, 2008. [cited at p. 61, 76]

[116] J. Serra. The Boolean model and random sets. *Computer Graphics and Image Processing*, 12:99–126, 1980. [cited at p. 41]

[117] H. Sidenbladh. Multi-target particle filtering for the Probability Hypothesis Density. In *Proc. 6th Int. Conf. Information Fusion*, pages 800–806, Cairns, Australia, 2003. [cited at p. 52, 67, 192]

[118] H. Sidenbladh. Tracking Random Sets of Vehicles in Terrain. In *Proc 2nd IEEE Workshop on Multi-Object Tracking*, Madison, WI, USA 2003, 2003. [cited at p. 6]

[119] R. Singer and J. Stein. An Optimal Tracking Filter for Processing Sensor Data of Imprecisely Determined Origin in Surveillance Systems. In *Proc. IEEE CDC*, pages 171–175, 1971. [cited at p. 4]

[120] S. S. Singh, B.-N. Vo, A. Baddeley, and S. Zuyev. Filters for Spatial Point Processes. *SIAM Journal on Control and Optimization*, 48(4):2275–2295, 2009. [cited at p. 147, 148, 149, 153, 157, 190, 195]

[121] R. Sittle. An Optimal Data Association Problem in Surveillance Theory. In *Proc. IEEE Tran on Military Electronics*, pages 125–139, 1964. [cited at p. 4]

[122] D. Stoyan, D. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley & Sons, 1995. [cited at p. 42, 148]

[123] R. Streit and T. Luginbuhl. Probabilistic multiple hypothesis tracking. *Research Report for Naval Undersea Warefare Center Division, NUWC-NPT/10/428,*, 1995. [cited at p. 40]

[124] M. Tobias and a.D. Lanterman. Probability hypothesis density-based multitarget tracking with bistatic range and Doppler observations. *IEE Proceedings - Radar, Sonar and Navigation*, 152(3):195, 2005. [cited at p. 6]

[125] M. Tobias and A. D. Lanterman. Techniques for birth-particle placement in the probability hypothesis density particle filter applied to passive radar. *IET Radar, Sonar and Navigation*, 2(5):351 – 365, 2008. [cited at p. 52]

[126] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157):1124–1131, 1974. [cited at p. 1]

[127] A. Tversky and D. Kahneman. Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4):S251–S278, 1986. [cited at p. 1]

[128] M. Van Lieshout. *Markov Point Processes and their Applications*. Imperial College Press, 2000. [cited at p. 41]

[129] B.-N. Vo and W.-K. Ma. A closed-form solution for the probability hypothesis density filter. In *Proc. 7th Int. Conf. Information Fusion*, pages 856–863, Philadelphia, USA, 2005. [cited at p. 57, 58, 192]

[130] B.-N. Vo and W.-K. Ma. The Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Transactions on Signal Processing*, 54(11):4091–4104, 2006. [cited at p. 6, 52, 57, 58, 75, 88, 92, 157, 159, 192, 195]

[131] B.-N. Vo, A. Pasha, and H. D. Tuan. A Gaussian Mixture PHD Filter for Nonlinear Jump Markov Models. In *45th IEEE Conference on Decision and Control*, volume 1, pages 3162–3167, San Diego, CA, 2006. [cited at p. 52, 57, 59]

[132] B.-N. Vo, S. S. Singh, and A. Doucet. Sequential Monte Carlo Implementation of the PHD Filter for Multi-Target Tracking. In *Proc. 6th Int. Conf. Information Fusion*, pages 792–799, 2003. [cited at p. 6, 52, 54, 67, 75]

[133] B.-N. Vo, S. S. Singh, and A. Doucet. Sequential Monte Carlo methods for multitarget filtering with random finite sets. *IEEE Transactions On Aerospace And Electronic Systems*, 41(4):1224–1245, 2005. [cited at p. 42, 44, 49, 52, 54, 157, 192, 195, 249, 250, 251]

[134] B.-N. Vo and D. Suter. Bayesian Multi-Object Estimation from Image Observations. In *Proc. 12th Int. Conf. Information Fusion*, pages 890–898, Seattle, WA, USA, 2009. [cited at p. 245]

[135] B.-T. Vo. *Random Finite Sets in Multi-Object Filtering*. PhD thesis, PhD Thesis, The University of Western Australia, 2008. [cited at p. 47, 59, 60, 61, 66, 190, 193, 244, 245]

[136] B.-T. Vo, B.-N. Vo, and A. Cantoni. Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter. *IEEE Transactions on Signal Processing*, 55(7):3553–3567, July 2007. [cited at p. 59, 148, 153]

[137] B.-T. Vo, B.-N. Vo, and A. Cantoni. The Cardinality Balanced Multi-Target Multi-Bernoulli Filter and its implementations. *IEEE Trans. Signal Processing*, 55(7):3553 – 3567, 2007. [cited at p. 59]

[138] Y.-D. Wang, J.-K. Wu, A. A. Kassim, and W.-M. Huang. Tracking a Variable Number of Human Groups in Video Using Probability Hypothesis Density. *Proceedings of the 18th International Conference on Pattern Recognition*, 3, 2006. [cited at p. 52]

[139] N. Wax. Signal to Noise Improvement and the Statistics of Tracking Populations. *Journal of Applied Physics*, pages 586–595, 1955. [cited at p. 4]

[140] P. Willett, Y. Ruan, and R. Streit. PMHT: Some problems and solutions. *IEEE Trans. Aerospace and Electronic Systems*, 38(3):738–754, 2002. [cited at p. 40]

[141] J. Wu. PHD Filter for Multi-target Visual Tracking with Trajectory Recognition. In *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, 2010. [cited at p. 53]

[142] W. Yang, J. Zhongliang, and S. Hu. Data association for PHD filter based on MHT. In *Proc. 8th Int. Conf. Information Fusion*, 2008. [cited at p. 52]

[143] T. Zajic and R. Mahler. A particle-systems implementation of the PHD multitarget tracking filter. *Signal Processing, Sensor Fusion, and Target Recognition*, 5096:291–299, 2003. [cited at p. 6, 52]

[144] T. Zajic, R. Ravichandran, R. Mahler, M. R., and N. M. Joint tracking and identification with robustness against unmodeled targets. In *SPIE Proc. 12th Signal Processing, Sensor Fusion and Target Recognition*, pages 279–290, 2003. [cited at p. 192]

[145] H. Zhang. Non-Linear Bayesian filtering by convolution method using FFT. *Proc. 14th Int. Conf. Information Fusion*, 2011. [cited at p. 100, 101]