Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Laurea Magistrale in Ingegneria Elettrica e dell'Automazione
Curriculum: Automazione

Tracking of Multiple Extended Objects

Tracciamento di Oggetti Estesi Multipli

Candidato
Matteo Tesori

Relatori
Prof. Luigi Chisci

Prof. Giorgio Battistelli

Correlatore
Dr. Lin Gao

anno accademico 2020-2021

Abstract

The work of the present thesis deals with the *Multiple Extended Object Tracking* (MEOT) problem in data fusion. As suggested by the name, such a problem amounts to jointly estimating the state of multiple extended objects, where the attribute *extended* means that each object can generate an arbitrary number of different measurements.

More specifically, in the context of this work, the MEOT problem is studied by taking into account the fact that the great amount of available measurements allows one to estimate not only the kinematic state of an object, but also its shape. As a result, two major difficulties have been faced: the first one consists of estimating how many objects are present in the surveilled scene and where there are located; the second consists of estimating the shape of each present object.

A Bayesian viewpoint is adopted, according to which the set of the objects states is modelled as a *random finite set* (RFS). In order to get a feasible solution from the point of view of the computational burden, the RFS of the tracked object is not represented in terms of its probability density function, but rather on first-order approximation known in the literature as *Probability Hypothesis Density* (PHD). The resulting algorithm, called *PHD filter*, is derived in the first part of the thesis and resolves the first difficulty of the MEOT problem, i.e. the joint estimation of the kinematic states of an unknown number of extended objects.

The second part of the thesis deals with the problem of estimating the shape of a single extended object, whose solution is consequently recast into the PHD filter formalism. In this part two different algorithms are discussed, namely the *Gaussian inverse Wishart* (GIW) filter and the *Multiplicative Error Model - Extended Kalman Filter Star* (MEM-EKF*) filter. The GIW filter can be regarded as a first simple solution to the shape estimation problem, while the MEM-EKF* filter is an evolution of the GIW filter aiming to provide better performances. Based on the main ideas of the GIW and the MEM-EKF* filters, a new solution, called by the author as *Lambda Omicron - Multiplicative Error Model* (LO-MEM) filter, is devised in the final chapter of the thesis.

Sommario

Il presente lavoro di tesi riguarda un problema di data fusion noto col nome di *Tracciamento di Oggetti Estesi Multipli* (TOEM). Come suggerito dal nome, tale problema consiste nella stima congiunta degli stati di molteplici oggetti estesi, dove il termine *esteso* indica che ogni oggetto ha la possibilità di generare un numero arbitrario di differenti misure.

In particolare, in questo lavoro il problema TOEM è studiato tenendo conto del fatto che l'elevato numero di misure disponibili permette di stimare non solo lo stato cinematico di un oggetto, ma anche la forma. Conseguentemente, nel presente lavoro sono affrontate due difficoltà principali: la prima consiste nella stima del numero di oggetti presenti nella scena e nella loro localizzazione; la seconda consiste nella stima della forma di ogni oggetto presente nella scena.

La soluzione considerata adotta un punto di vista Bayesiano, secondo il quale l'insieme degli stati degli oggetti è modellato da un *insieme finito aleatorio* (RFS = Random Finite Set). Nell'ottica di ottenere una soluzione computazionalmente accettabile, il RFS degli stati degli oggetti non è rappresentato in termini della sua densità di probabilità, bensì in termini della rispettiva approssimazione al primo ordine nota come *Probability Hypothesis Density* (PHD). L'algoritmo risultante, chiamato *filtro PHD*, è derivato nella prima parte della tesi e risolve la prima grande difficoltà del problema TOEM, i.e. la stima congiunta degli stati di un numero non precisato di oggetti estesi.

La seconda parte della tesi affronta il problema della stima della forma di un singolo oggetto esteso, la cui soluzione è conseguentemente riformulata nel formalismo del filtro PHD. In questa parte sono discussi due algoritmi, che sono rispettivamente il *filtro Gaussian inverse Wishart* (GIW) ed il *filtro al Modello di Errore Moltiplicativo - filtro di Kalman Esteso Stella* (MEM-EKF*). Il filtro GIW può essere visto come una prima semplice soluzione al problema della stima di forma, mentre il filtro MEM-EKF* è un'evoluzione del filtro GIW ideata per ottenere prestazioni migliori. Sulla base delle idee principali dei filtri GIW e MEM-EKF*, una nuova soluzione, chiamata dall'autore *filtro Lambda Omicron - Modello di Errore Moltiplicativo* (LO-MEM), è sviluppata nel capitolo finale della tesi.

Acronyms

APB	Approximate Poisson Body (model)
CDF	Cumulative Distribution Function
EKF	Extended Kalman Filter
FISST	Finite Set Statistics
GIW	Gaussian Inverse Wishart
GM	Gaussian Mixture
LO	Lambda Omicron
MEM	Multiplicative Error Model
MEOT	Multiple Extended Object Tracking
MOT	Multiple Object Tracking
MPDF	Multiobject Probability Density Function
PDF	Probability Density Function
PGF	Probability Generating Function
PGFL	Probability Generating Functional
PHD	Probability Hypotesis Density (function)
PMF	Probability Mass Function
RFS	Random Finite Set
RHS	Right Hand Side
SMC	Sequential Monte Carlo
SOT	Single Object Tracking

Contents

Chapter 1

Introduction

1.1 Objective and thesis structure

One of the crucial and most interesting problems in control systems engineering is the estimation of the state of a dynamic system. In fact, the usual solution adopted to control a dynamic system is to define the controller as a suitable state feedback. However, usually the state is unknown, therefore one must provide a sufficiently robust estimate of the state to the controller.

A special instance of the state estimation problem is called *tracking*, which consists in the estimation of the state of a moving object, for example an aircraft, based on remote measurements, provided by one or more sensors, for example a radar station, typically in a discrete time fashion.

In 1960, Kalman published a paper describing its solution to the state estimation problem, the *Kalman filter*, which can be considered as the starting point of the modern tracking systems. Despite its undoubted success in the vast majority of the state estimation problems, today the Kalman filter does not provide a practically usable solution to the tracking problem because of its basic assumptions that, in this special estimation context, read as follows:

- (**single object assumption**) at each time step, there is one and only object present in the observed scene;
- (**point object assumption**) at each time step, the object generates one and only one measurement.

On the one hand, the first assumption is not valid due to the fact that, in general, the scene observed by the sensors is populated by an unknown and time varing number of objects due to the fact that such objects can enter or leave the scene. Moreover, the scene can contain also false-objects detectable by the sensors. For example, in air surveillance problem one can be interested in tracking airplanes (true-objects) but not helicopters (false-objects).

On the other hand, the second assumption is no longer valid nowdays since the advance in sensor technology which has led to an increment into resolution capabilities, that is the modern sensors, for example the LIDAR sensors, can observe simultaneously multiple different points of the tracked object. Thanks to the amount of measurements, it is possible to estimate both the position of the object, represented by a point regarded as the center, and the shape of the object, represented by an ellipsoid (or, more often as an ellipse if the tracking problem is defined in a bidimensional enviroment) that encodes the extensions and the orientation of the object.

The objective of this thesis is to present three different tracking algorithms that address the two limitations of the Kalman filter.

The first limitation is overtaken by considering the simplest solution provided by Mahler's FISST (Finite Set Statistics): the so-called PHD (Probability Hypotesis Density) filter. This argument is discussed in the first part of the thesis (chapters 2, 3, 4, 5). The second limitation is overtaken by discussing in the second part of the thesis (chapters 6, 7, 8, 9) three different approaches to the tracking problem for extended objects:

1. the GIW (Gaussian Inverse Wishart) filter, which is the first filter in the literature, introduced by Koch between 2006 and 2011, that handles the problem of estimating the kinematic state and the shape of the tracked object;

2. the MEM-EKF* (Multiplicative Error Model - Extended Kalman Filter Star) filter, an evolution of the GIW filter, developed by Baum et al. between 2012-2019, that permits to explicitly separate the estimation of the extensions from the estimation of the orientation of the tracked object;

3. the LO-MEM (Lambda Omicron - Multiplicativer Error Model) filter, that is a new filter based on the MEM-EKF* proposed by the author in order to estimate the state of objects showing high manouvering behaviours, i.e. that change rapidly their position and orientation during time.

The multi-object trackers discussed in this thesis are nothing but the PHD implementations of the former single object trackers, specifically:

1. the GIW-PHD filter (Granstorm et al., 2012);

2. the PHD MEM-EKF* (Baum et al., 2017);

3. the LO-MEM PHD (developed in this work).

Finally, the performance of the trackers are compared by means of numerical simulations in the final chapter 10.

1.2 Contributions

The main contribution is the LO-MEM PHD filter, a new multi-object tracker that tries to improve the estimation performance of the MEM-EKF* filter in complex scenarios where the multiple objects move in the scene by rapidly changing in time their positions and orientations.

The LO-MEM PHD filter differs from the MEM-EKF* PHD filter under the following aspects:

- assumes that every object follows the so-called *unicycle motion model*, that is one object can move only along the orientation direction or, in other words, cannot move along its lateral direction. Due to this feature, the LO-MEM filter loose of generality with respect of the MEM-EKF* filter;

- the state of an object is represented in polar coordinates rather than a Cartesian coordinates. This feature allows to encode explicitly in the estimation process the former property of the unicycle motion model;

- the prediction step is based on a new non-linear motion model, called *Lambda-Omicron motion model*, rather than a linear motion model. This feature leads to a more accurate estimation of the object orientation;

- the correction step is computed according to a new measurement vector, which allows to perform the correction in one operation (rather than multiple operations). Thanks to this feature, the new orrection step is less computational demanding and solves some minor problems of the MEM-EKF* corrector;

- the model of the measurement likelihood, which plays a central role in the PHD implementation.

1.3 Preliminary discussion on the PHD filter

The PHD filter, the central algorithm of this thesis, can be seen as an evolution of the so called Bayes's filter. The Bayes filter is a generalization of the

Kalman filter that allows to estimate the state of stochastic systems that are not linear in their dynamics or characterized by Gaussian uncertainties.

The PHD filter, conversly, is an advanced approximation of the so-called *multiobject Bayes filter*, which generalizes the Bayes filter (referred as *single object Bayes filter*) because the latter can handle the estimation of the state of single object (mathematically represented by a dynamic system[1]) while the former can handle the estimation of the states of multiple objects simultaneously.

In order to clarify the mathematics behind the PHD filter, in the next two subsections are briefly discussed the single object and multiobject Bayes filters.

1.3.1 Single object Bayes filter

In the *single object tracking* (SOT) problem, at the time step k one wants to estimate the value of the actual state $X_k \in \mathbb{R}^n$ of one object (which can generate only one measurement per time step) given the set $y_{1:k} \triangleq \{y_1, \ldots, y_k\}$ of the measurements provided by a sensor up to the actual time step k.

The conceptual solution of the SOT problem is represented by the (single object) Bayes filter, which is a recursive algorithm relying on the following basic concepts:

- measurement model $h_k(\cdot, \cdot)$ - a function that describes how the measurements are generated by the considered sensor;
- motion model $f_k(\cdot, \cdot)$ - a function that describes what the future value of the object state will be;
- likelihood function $\ell_k(\cdot)$ - a probabilistic representation of the measurement model;
- transition density $\varphi_{k+1|k}(\cdot|\cdot)$ - a probabilistic representation of the motion model;
- filtered density $p_{k|k}(\cdot)$ - a probabilistic representation of the actual value of the object state based on the set of measurement actually available;
- predicted density $p_{k+1|k}(\cdot)$ - a probabilistic representation of the future value of the object state based on the set of measurement actually available;

[1]for this reason the terms *object* and *dynamic system* are considered synonyms in this work

- optimal estimate $\hat{x}_{k|k}$ - a point in the state space that estimates the actual value of the object state;
- estimator covariance $P_{k|k}$ - a matrix that describes the accuracy of the optimal considered estimate.

Briefly speaking, the derivation of the (single object) Bayes filter consists essentially in the following procedure:

- **step 1**: define the measurement and motion models according to the sensor considered and the time-behaviour of the object under study

$$\begin{aligned} x_{k+1} &= f_k(x_k, w_k) \\ y_k &= h_k(x_k, v_k) \end{aligned} \tag{1.1}$$

  here x_k and x_{k+1} are the (real) states of the objects at time k and $k+1$, y_k is the measure observed at time k by the sensor and w_k, v_k are noise signals (typically stationary) characterized by the PDFs $p_W(\cdot)$ and $p_V(\cdot)$;

- **step 2**: find the likelihood function and the transition density from the measurement and motion models according to the standard formulas provided by the probability calculus, i.e.

$$\begin{aligned} \ell_k(y|x) \triangleq p(y_k|x_k) &= \frac{p_V(h_k^{-1}(x_k, y_k))}{[\det J_{h_k}]_{x_k, h^{-1}(x_k, y_k)}} \\ \varphi_{k+1|k}(x|w) \triangleq p(x_{k+1}|w_k) &= \frac{p_W(f_k^{-1}(w_k, x_{k+1}))}{[\det J_{f_k}]_{x_k, f^{-1}(w_k, x_{k+1})}} \end{aligned} \tag{1.2}$$

  here $h_k^{-1}(\cdot)$ and $f_k^{-1}(\cdot)$ are the inverse functions of $h_k(\cdot)$ and $f_k(\cdot)$, while $\det J_{h_k}$ and $\det J_{f_k}$ are the Jacobian determinants of $h_k(\cdot)$ and $f_k(\cdot)$;

- **step 3**: find the filtered and predicted densities from the likelihood function and the transition density according to the Bayes and the Chapman-Kolmogorov equations, i.e.

$$\begin{aligned} p_{k|k}(x) &\triangleq \frac{\ell_k(y|x) p_{k|k-1}(x)}{\int \ell_k(y|\xi) p_{k|k-1}(\xi)\, \mathrm{d}\xi} \\ p_{k+1|k}(x) &\triangleq \int \varphi_{k+1|k}(x|w) p_{k|k}(w)\, \mathrm{d}w \end{aligned} \tag{1.3}$$

  here $p_{k|k-1}(\cdot)$ is the predicted density at time $k-1$;

- **step 4**: extract from the filtered density the optimal estimate by minimizing the risk $R_{C_k}[\cdot]$, where $C_k : \mathbb{R}^{2n} \mapsto \mathbb{R}^+$ is the choosen cost function, i.e.
$$\hat{x}_k \triangleq \arg\min_x R_{C_k}[p_{k|k}] \tag{1.4}$$
with
$$R_{C_k}[p_{k|k}] \triangleq \int C_k(x, w) p_{k|k}(w) \mathrm{d}w \tag{1.5}$$

1.3.2 Multiobject Bayes filter

In the more general and challenging *multiobject tracking* (MOT) problem, the objective is to estimate the state of several objects given severals measurements provided by several sensors (where, in general, every object can generate more than one measurement per time step). In this case, if at the time step k the number of objects is N_k and the number of measurements is m_k, the MOT problem consists of estimate simultaneously the values of the states $X_{k,1}, \ldots, X_{k,N_k}$ given the set of measurements $y_{1:m_1,m_k} \triangleq \{\cup_{i=1}^{m_1}\{y_{1,i}\}, \ldots, \cup_{i=1}^{m_k}\{y_{1,k}\}\}$.

Note that in the MOT problem the numbers N_k is a random integer (eventually zero), while m_k is a realization of a second random integer M_k (eventually zero), in fact:

- N_k is random because it is not known with certaintly how many objects are actually present in the scene supervised by the sensors. Such uncertaintly is a consequence of the fact that objects actually present in the scene can leave that scene in the future, while objects actually non present in the scene can enter that scene in the future. In the Bayesian reasoning the uncertainties are represented in probabilistic terms, so N_k is random and, in principle, can assume all the possible values $n_k = 0, 1, 2, \ldots$;
- M_k is random because a sensor may not read any measurement (because the sensor can be occluted) or may read, in addition to the measurements generated by the objects in the scene, several false measurements (due to the presence of false objects in the scene). As a result, M_k can assume all the possible values $m_k = 0, 1, 2, \ldots$ and it is modelled as a random variable likewise N_k;

A possible, and convenient, approach to resolve the MOT problem is the following:

1. collect the actual single object states in the state set
$$\mathsf{X}_k \triangleq \{X_{k,1}, \ldots, X_{k,N_k}\} \tag{1.6}$$

and collect the actual measurements observed by the sensors in the measurement set

$$\mathsf{y}_k \triangleq \{y_{k,1}, \ldots, y_{k,m_k}\} \tag{1.7}$$

2. think the starting MOT problem as a non-ordinary SOT problem where at the time step k one wants to estimate the value of the actual state X_k of one meta-object given the set $\mathsf{y}_{1:k} \triangleq \{\mathsf{y}_1, \ldots, \mathsf{y}_k\}$ of meta-measurements readed by a meta-sensor up to the actual time step.

Unfortunately X_k and y_k are *sets* rather than *vectors*, and this means that there are several problem to resolve, for example

- the filtered and predicted densities are not defined, since it is not defined the concept of probability density of a set;
- the likelihood and the transition density are not defined, since, once again, these two terms are in this case probability densities of sets, a concept not defined;
- the optimal estimate is not defined, since the filtered density is not defined.

The *finite set statistics* (FISST) is a theory that fill up all theese theoretical gaps by introducing the concept of *random finite set*. The main ideas behind FISST are the following:

- the introduction of the concept of RFS allows to define properly a concept of PDF for sets, called *multiobject PDF* (MPDF) in what follows, and a concept of integration over the space of MPDFs, called *set integral* in what follows. As a result, in FISST the filtered and predicted densities are well defined and denoted as $p_{k|k}(\{\cdot\})$, $p_{k+1|k}(\{\cdot\})$;
- due to the definition of set integral, the Bayes equation and the Chapman-Kolmogorov equation are appliable in the new context of RFSs

$$\begin{aligned} p_{k|k}(\mathsf{x}) &\triangleq \frac{\ell_k(\mathsf{y}|\mathsf{x})p_{k|k-1}(\mathsf{x})}{\int \ell_k(\mathsf{y}|\mathsf{w})p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}} \\ p_{k+1|k}(\mathsf{x}) &\triangleq \int \varphi_{k+1|k}(\mathsf{x}|\mathsf{w})p_{k|k}(\mathsf{w})\ \mathrm{d}\mathsf{w} \end{aligned} \tag{1.8}$$

  where the usual definition of integration is replaced by the definition of set-integral;
- given a generic filtered density $p_{k|k}(\{\cdot\})$, FISST provides also different procedures to extract an optimal estimate $\hat{\mathsf{x}}_k$ (which is a finite set of vectors in $\mathbb{R}^n$ rather than a single vector in $\mathbb{R}^n$).

The multiobject Bayes filter is an algorithm that propagates in time the predicted and filtered densities $p_{k|k}(\{\cdot\})$, $p_{k+1|k}(\{\cdot\})$ by computing the Bayes and Chapman-Kolomogorov equations and produces the estimates by processing the filtered densities $p_{k|k}(\{\cdot\})$.

1.3.3 Motivation for the PHD filter

The multiobject Bayes filter is computational intractable even in the simple applications. The PHD filter is an algorithm that solves this problem by propagating in time not the predicted and filtered densities $p_{k|k}(\{\cdot\})$, $p_{k+1|k}(\{\cdot\})$ but rather the predicted and filtered *probability hypothesis densities* $D_{k|k}(\cdot)$, $D_{k+1|k}(\cdot)$, which are standard functions of the type $f : \mathbb{R}^n \mapsto \mathbb{R}^+$ that, despite this is not completely true, can be regarded as rough approximations of $p_{k|k}(\{\cdot\})$, $p_{k+1|k}(\{\cdot\})$.

However, the analitical computation of $D_{k|k}(\cdot)$, $D_{k+1|k}(\cdot)$ is not straightforward and requires advanced tools, the so-called *probability genereting functionals* (PGFLs). For this reason, the entire third chapter of the thesis is devoted only to the PGFLs and fourth chapter to the derivation of the standard PHD filter (i.e., the computation of $D_{k|k}(\cdot)$, $D_{k+1|k}(\cdot)$). The fifth chapter discusses the derivation of the PHD filter for extended objects.

Part I

PHD filters

Chapter 2

Standard multiobject calculus

2.1 Summary

The idea of FISST is to define the multiobject Bayes filter as a single object Bayes filter that deals with random finite sets, a new concept of random variable, rather than random vectors. Hence the objective of FISST is to recast in the random finite set enviroment the single object Bayes filtering theory. In order achieve this goal, FISST develops the so called multiobject calculus, which is the extension of the ordinary multivariate calculus to the random finite set enviroment.

The multiobject calculus consist in the following 4 steps:

- **step 1**: define a probabilistic model $(p_N, \mathcal{S})$, which is clarified by the following definition 1, for random finite sets;

- **step 2**: translate the probabilistic model $(p_N, \mathcal{S})$ that defines a random finite set into a more convenient function $\beta_{\mathsf{X}}(\cdot)$, called belief mass function. The belief mass function $\beta_{\mathsf{X}}(\cdot)$ is a compact probabilistic descriptor of X equivalent to the model $(p_N, \mathcal{S})$. Equivalent means that it is possible to recover $(p_N(\cdot), \mathcal{S})$ from $\beta_{\mathsf{X}}(\cdot)$ and viceversa. The belief mass function $\beta_{\mathsf{X}}(\cdot)$ is the random finite set counterpart of the ordinary probability mass function (PMF) $\mathbb{P}_X(\cdot)$ (which is the distribution function of a random vector X).

- **step 3**: introduce a new concept of integral, the so called set integral, that permits to integrate functions of random finite sets. Basing on the definition of set integral, the belief mass function can be expressed as the set integral of a special multiobject function $p_{\mathsf{X}}(\{\cdot\})$

$$\beta_{\mathsf{X}}(S) = \int_S p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) \tag{2.1}$$

equation (3)[1] is the random finite set counterpart of the much more familiar equation

$$\mathbb{P}_X(S) = \int_S p_X(x)\ \mathrm{d}x \qquad \forall\, S \in \mathcal{B}(\mathbb{R}^n) \tag{2.2}$$

thus the multiobject function $p_{\mathsf{X}}(\{\cdot\})$ is the random finite set counterpart of the ordinary PDF $p_X(\cdot)$, and for this reason is called multiobject PDF.

- **step 4**: introduce a new concept of differentation, the so called Lebesgue-differentation, that permits to extract, without loosing any information[2] about a random finite set, the multiobject PDF $p_{\mathsf{X}}(\{\cdot\})$ from a belief mass function $\beta_{\mathsf{X}}(\cdot)$

$$p_{\mathsf{X}}(\mathsf{x}) = \left.\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}\mathsf{x}}\right|_{S=\varnothing} \qquad \forall \mathsf{x} \in \mathcal{F}(\mathbb{R}^n) \tag{2.3}$$

equation (5) is the random finite set counterpart of the much more familiar equation

$$p_X(x) = \left.\frac{\mathrm{d}\mathbb{P}_X(S)}{\mathrm{d}x}\right|_{S=(-\infty,x]^n} = \frac{\mathrm{d}P_X(x)}{\mathrm{d}x} \qquad \forall\, x \in \mathbb{R}^n \tag{2.4}$$

where $P_X(\cdot)$ is the cumulative distribution function (CDF) of X, i.e. the probability mass function $\mathbb{P}_X(\cdot)$ restricted to the hyper-intervals[3] $S = (-\infty, x]^n$

$$P_X(x) \triangleq \mathbb{P}_X((-\infty, x]^n) \qquad \forall\, x \in \mathbb{R}^n \tag{2.5}$$

Due to (3) and (5), the multiobject PDF $p_{\mathsf{X}}(\{\cdot\})$ can be regarded as a probabilistic descriptor of X equivalent to the belief mass function $\beta_{\mathsf{X}}(\cdot)$. Moreover, turns out that the multiobject PDF $p_{\mathsf{X}}(\{\cdot\})$ is also

[1]equations (3), (4), (5) and (6) will be discussed in details in the next sections, for this reason it is not explained here the notation adopted.

[2]once again, "without loosing any information" means that it is possible to recover back the belief mass function from the multiobject PDF.

[3]the notation $(-\infty, x]^n$ is a shorthand for the cartesian product $\prod_{i=1}^n(-\infty, x_i]$.

equivalent to the model $(p_N, \mathcal{S})$, and this prove the equivalence between the model $(p_N, \mathcal{S})$ and the belief mass function $\beta_{\mathsf{X}}(\cdot)$.

In conclusion, there are 3 equivalent ways to characterize the statistical properties of a random finite set: the probabilistic model $(p_N, \mathcal{S})$, the belief mass function $\beta_{\mathsf{X}}(\cdot)$ and the multiobject PDF $p_{\mathsf{X}}(\{\cdot\})$.

- **step 5**: define new concept of statistics which summarize in a compact form the information contained in the multiobject PDF. Such statistics are the marginal multitarget estimate, the joint multitarget estimate and the probability hypothesis density. Likewisei n the ordinary domain of random vectors, such statistics are not complete probabilistic desriptors, meaning that it is not possible to recover back the multiobject PDFs from such statistics (it is only possible to extract such statistics given the multiobject PDFs). More importantly, this fact means that there is a loss of information in the extraction of the statistics from the multiobject PDFs, and this can be a problem in complex scenarios (for example, when the signal-to-noise ratio is low). This fact will be proved by the so called inversion formula.

2.2 Random finite sets

The ordinary probability theory defines different types of random variables, for example:

- **random integer** I: a random variable that draws its instantiations i from the set $\mathbb{Z}$ of all integers;
- **random number** A: a random variable that draws its instantiations a from the set $\mathbb{R}$ of all real numbers;
- **random vector** X: a random variable that draws its instantiations x from the Euclidean space $\mathbb{R}^n$ of all real-valued vectors;

In the MOT is involved a new and more sophisticated concept of random variable, that is:

- **random finite set** X: a random variable that draws its instantiations x from the set $\mathcal{F}(\mathbb{R}^n)$ of all finite subsets of the Euclidean space $\mathbb{R}^n$.

For example, possible instantiations of a random finite set X are the following: $\mathsf{x} = \varnothing$, $\mathsf{x} = \{x\}$, $\mathsf{x} = \{x_1, x_2\}$, $\mathsf{x} = \{x_1, x_2, x_3\}$ and so on, where the elements x_1,x_2,..., are ordinary random vectors belonging to $\mathbb{R}^n$. In short, a random finite set is essentially a set with random cardinality, which can be zero as well, composed by random vectors.

FISST assumes a simplifying hypothesis that prevents some mathematical issues: the molteplicity of the elements of a random finite set is always unitary, i.e. are not allowed repetead elements in random finite sets. This hypothesis is not restrictive because the elements of a random finite set are vectors drawn from the continuos space $\mathbb{R}^n$, so it is almost impossible that two elements in a random finite set can be identical.

To gain intuition about what is a random finite set, one can think about the following algorithm which explains how to sample a random finite set:

1. initialize $\mathsf{x} = \varnothing$;
2. generate an integer η according to some *finite* discrete density $p_N(\cdot)$;
3. if $\eta > 0$ then:
   (a) generate the real-valued vectors x_1, x_2, …, x_η in $\mathbb{R}^n$ according to some joint *symmetric* PDF $p_{X_1,X_2,\dots,X_\eta}(\cdot,\cdot,\dots,\cdot)$;
   (b) include the vectors x_1, x_2, …, x_η in x: $\mathsf{x} \leftarrow \mathsf{x} \cup \{x_1, x_2, \dots, x_\eta\}$.

Note that the attributes *finite* and *symmetric* describing the densities $p_N(\cdot)$ and $p_{X_1,X_2,\dots,X_\eta}(\cdot,\cdot,\dots,\cdot)$ have specific meanings:

- finite means that $p_N(\cdot)$ vanishes as the cardinality considered increase, i.e. $\lim_{\eta\to\infty} p_N(\eta) = 0$. This is because a random finite set is a set composed by a finite number of elements, so it is not allowed the case $|\mathsf{X}| = \infty$;
- symmetric means that $p_{X_1,X_2,\dots,X_\eta}(\cdot,\cdot,\dots,\cdot)$ is invariant with respect to permutations of it arguments, i.e.

$$p_{X_1,X_2,\dots,X_\eta}(x_1, x_2, \dots, x_\eta) = p_{X_1,X_2,\dots,X_\eta}(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(\eta)}) \quad \forall \text{ permutations } \sigma(\cdot) \,. \tag{2.6}$$

For example, if $\eta = 2$ then the symmetry states that it is true

$$p_{X_1,X_2}(x_1, x_2) = p_{X_1,X_2}(x_2, x_1). \tag{2.7}$$

This is because a random finite set is a collection of *unordered* elements, in the sense that, for example, $\{x_2, x_3, x_1\}$ and $\{x_1, x_2, x_3\}$ are different representations of the same set. In this case, since

$$\{x_2, x_3, x_1\} = \{x_1, x_2, x_3\} \tag{2.8}$$

it is necessary that

$$p_{X_1,X_2,X_3}(x_2, x_3, x_1) = p_{X_1,X_2,X_3}(x_1, x_2, x_3) \tag{2.9}$$

so that the PDF $p_{X_1,X_2,X_3}(\cdot,\cdot,\cdot)$ must be symmetric, i.e. can be written as

$$p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta) \triangleq \frac{1}{\eta!}\sum_\sigma \tilde{p}_{X_1,X_2,\ldots,X_\eta}(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(\eta)}) \tag{2.10}$$

where $\tilde{p}_{X_1,X_2,\ldots,X_\eta}(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(\eta)})$ is a generic PDF (potentially non-symmetric) and the summation is taken over all $\eta!$ possible permutations $\sigma(\cdot)$ of the arguments $x_1, \ldots, x_\eta$. Note that the factor $1/\eta!$ guarantees that $p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta)$ integrates to unity, while the summation over all the permutations guarantees that the joint PDF $p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta)$ is symmetric.

Although, as prescribed by the assiomatic theory of probability, a random finite set is rigorously defined as a measurable map $\mathsf{X} : \Omega \mapsto \mathcal{F}(\mathbb{R}^n)$, where

- Ω is some sample space in an underliyng probability space $(\Omega, \Sigma, \mathbb{P})$;
- $\mathcal{F}(\mathbb{R}^n)$ is the hyperspace of finite subsets of $\mathbb{R}^n$;

it is possible to give a more intuitive and equivalent definition based on the previous intuitive algorithm.

Definition 1. A finite set $\mathsf{X} = \{X_1, X_2, \ldots, X_N\} \subset \mathbb{R}^n$ is called *random finite set* (RFS) over $\mathbb{R}^n$ if and only if:

- the cardinality N is a random integer, characterized by a finite discrete density $p_N : \mathbb{N} \mapsto [0, 1]$ called *cardinality density*;
- for all η such that $p_N(\eta) > 0$, the elements $X_1, X_2, \ldots, X_\eta \in \mathbb{R}^n$ are random vectors, characterized by a symmetric joint PDF

$$p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta) \triangleq \frac{1}{\eta!}\sum_\sigma \tilde{p}_{X_1,X_2,\ldots,X_\eta}(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(\eta)}) \tag{2.11}$$

on $\mathbb{R}^{n\times\eta}$ called *spatial density*.

An RFS is therefores completely characterized by a finite discrete density $p_N(\cdot)$ and a *family* $\mathcal{S} \triangleq \{\tilde{p}_{X_1,X_2,\ldots,X_\eta}(\cdot,\cdot,\ldots,\cdot)\}_{\eta:p_N(\eta)>0}$ of PDFs or, more coincisely, by a couple $(p_N, \mathcal{S})$, referred in this document as the *model* of the RFS.

2.3 Belief mass functions

The central set function of the FISST is the-so called *belief mass function* (BMF)

$$\beta_{\mathsf{X}}(S) \triangleq \mathbb{P}(\mathsf{X} \subseteq S) \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) \tag{2.12}$$

where $\mathcal{O}(\mathbb{R}^n)$ is the class of open subsets of $\mathbb{R}^n$ and $\mathbb{P} : \Sigma \mapsto [0,1]$ is the underliyng probability measure. Note the difference between X and S: despite the fact both X and S are subsets of $\mathbb{R}^n$, the former is a finite subset of $\mathbb{R}^n$, while the second is a continuos (i.e. non-countably infinite) subset of $\mathbb{R}^n$.

The BMF is the probability of the event $\mathsf{X} \subseteq S$ or, in simpler terms, the probability that the outcome x of X is completely contained in a open region of interest $S \subseteq \mathbb{R}^n$.

It turns out immediately that the BMF is a direct generalization of the probability mass function (PMF) of a random vector. In fact, if the RFS X reduces to a random vector X then

$$\beta_{\mathsf{X}}(S) = \mathbb{P}(\{X\} \subseteq S) = \mathbb{P}(X \subseteq S) = \mathbb{P}(X \in S) \triangleq \mathbb{P}_X(S) \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) \tag{2.13}$$

It is possible express the BMF in a factorized formula. Definition 1 suggests the following two facts:

- with probability $p_N(\eta)$, the RFS X is composed by η vectors $X_1, \ldots, X_\eta$, i.e. its cardinality $|\mathsf{X}|$ is exactly equal to η, so

$$\mathbb{P}(|\mathsf{X}| = \eta) = p_N(\eta) \tag{2.14}$$

- assuming $|\mathsf{X}| = \eta$, the probability of the event $\mathsf{X} \subseteq S$, denoted concisely as $\beta_{\mathsf{X}|\eta}(S) \triangleq \mathbb{P}(\mathsf{X} \subseteq S || \mathsf{X}| = \eta)$, is the probability of the event 'X_1, X_2,...,X_η are simultaneously in the region S', which is $\mathbb{P}_{X_1,X_2,\ldots,X_\eta}(S)$. Consequently hold the following

$$\begin{aligned} \beta_{\mathsf{X}|\eta}(S) &= \mathbb{P}_{X_1,X_2,\ldots,X_\eta}(S) \\ &\triangleq \int_{S^\eta} p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots\, \mathrm{d}x_\eta \end{aligned} \tag{2.15}$$

  where $p_{X_1,X_2,\ldots,X_\eta}(\cdot,\cdot,\ldots,\cdot)$ is the joint PDF of X_1, X_2, ..., X_η and S^η denotes the Cartesian product $\prod_{i=1}^{\eta} S = S \times S \times \cdots \times S$. Note that, as long as the subset S is in the Borel σ-algebra of $\mathbb{R}^n$ (denoted as $\mathcal{B}(\mathbb{R}^n)$), the probability measure $\beta_{\mathsf{X}|\eta}(\cdot)$ is well defined since it is an ordinary joint absolute-continue distribution on $\mathbb{R}^n$. This fact is not in contrast with definition (16) (which considers an open subset S of $\mathbb{R}^n$) since $\mathcal{O}(\mathbb{R}^n) \subset \mathcal{B}(\mathbb{R}^n)$.

Recalling that $\mathbb{P}(\cdot)$ is a probability measure, the law of total probabilities can be applied in order to exploit the two aforementioned facts, leading to the expressions

$$
\begin{aligned}
\beta_{\mathsf{X}}(S) = \mathbb{P}(\mathsf{X} \subseteq S) = \sum_{\eta=0}^{\infty} \mathbb{P}(\mathsf{X} \subseteq S \,|\, |\mathsf{X}| = \eta)\, \mathbb{P}(|\mathsf{X}| = \eta) \\
= \sum_{\eta=0}^{\infty} \beta_{\mathsf{X}|\eta}(S)\, p_N(\eta) = p_N(0) + \\
\sum_{\eta=1}^{\infty} \left(\int_{S^\eta} p_{X_1, X_2, \ldots, X_\eta}(x_1, x_2, \ldots, x_\eta)\; \mathrm{d}x_1\; \mathrm{d}x_2 \cdots\; \mathrm{d}x_\eta \right) p_N(\eta) \\
\forall\, S \in \mathcal{O}(\mathbb{R}^n)
\end{aligned}
\tag{2.16}
$$

where it is assumed by convention that $\beta_{\mathsf{X}|0}(S) \triangleq 1$, which is reasonable because, since the empty set $\varnothing$ is by definition a subset of any set S, the probability $\beta_{\mathsf{X}|0}(S)$ of the event $\varnothing \subseteq S$ is unitary regardless of the value of S.

2.3.1 Independence for RFSs

Let X_1 and X_2 be two independent RFSs with BMFs $\beta_{\mathsf{X}_1}(\cdot)$ and $\beta_{\mathsf{X}_2}(\cdot)$ respectively. Let $\mathsf{X} \triangleq \mathsf{X}_1 \cup \mathsf{X}_2$ be the union between X_1 and X_2. It is easy to see that the BMF $\beta_{\mathsf{X}}(\cdot)$ is factorized into the product of the BMFs $\beta_{\mathsf{X}_1}(\cdot)$ and $\beta_{\mathsf{X}_2}(\cdot)$. In fact observe

$$
\beta_{\mathsf{X}}(S) = \mathbb{P}(\mathsf{X} \subseteq S) = \mathbb{P}(\mathsf{X}_1 \cup \mathsf{X}_2 \subseteq S) = \mathbb{P}(\mathsf{X}_1, \mathsf{X}_2 \subseteq S). \tag{2.17}
$$

Thus, due to the independence of X_1 and X_2, it follows that $\mathbb{P}(\mathsf{X}_1, \mathsf{X}_2 \subseteq S) = \mathbb{P}(\mathsf{X}_1 \subseteq S)\, \mathbb{P}(\mathsf{X}_1 \subseteq S)$ and

$$
\beta_{\mathsf{X}}(S) = \mathbb{P}(\mathsf{X}_1, \mathsf{X}_2 \subseteq S) = \mathbb{P}(\mathsf{X}_1 \subseteq S)\, \mathbb{P}(\mathsf{X}_1 \subseteq S) = \beta_{\mathsf{X}_1}(S)\, \beta_{\mathsf{X}_1}(S). \tag{2.18}
$$

2.4 Set integrals and multiobject PDFs

According to the expression (2.16), a BMF can be also written also in the following form

$$\begin{aligned}\beta_{\mathsf{X}}(S) &= p_N(0) + \sum_{\eta=1}^{\infty}\left(\int_{S^\eta} p_{X_1,\dots,X_\eta}(x_1,\dots,x_\eta)\ \mathrm{d}x_1\cdots\mathrm{d}x_\eta\right) p_N(\eta)\\ &= p_N(0) + \sum_{\eta=1}^{\infty}\frac{1}{\eta!}\left(\int_{S^\eta} \eta!\, p_{X_1,\dots,X_\eta}(x_1,\dots,x_\eta)\ \mathrm{d}x_1\cdots\mathrm{d}x_\eta\right) p_N(\eta)\\ &= p_N(0) + \sum_{\eta=1}^{\infty}\frac{1}{\eta!}\int_{S^\eta} \eta!\, p_{X_1,\dots,X_\eta}(x_1,\dots,x_\eta)\, p_N(\eta)\ \mathrm{d}x_1\cdots\mathrm{d}x_\eta\end{aligned} \tag{2.19}$$

where this final expression plays a central role in FISST since it leads to the following definitions.

Definition 2. Let $(p_N, \mathcal{S})$ be the model of an RFS X, then the finite set function $p_{\mathsf{X}} : \mathcal{F}(\mathbb{R}^n) \mapsto \mathbb{R}^+$ defined as

$$\begin{aligned}p_{\mathsf{X}}(\mathsf{x}) &\triangleq \begin{cases} p_N(0) & \text{if } \mathsf{x} = \varnothing\\ 1!\, p_{X_1}(x_1)\, p_N(1) & \text{if } \mathsf{x} = \{x_1\}\\ \vdots & \\ \eta!\, p_{X_1,\dots,X_\eta}(x_1,\dots,x_\eta)\, p_N(\eta) & \text{if } \mathsf{x} = \{x_1,\dots,x_\eta\}\end{cases}\\ &= \begin{cases} p_N(0) & \text{if } \mathsf{x} = \varnothing\\ p_N(1)\, \tilde{p}_{X_1}(x_1) & \text{if } \mathsf{x} = \{x_1\}\\ \vdots & \\ \sum_\sigma p_N(\eta)\, \tilde{p}_{X_1,\dots,X_\eta}(x_{\sigma(1)},\dots,x_{\sigma(\eta)}) & \text{if } \mathsf{x} = \{x_1,\dots,x_\eta\}\end{cases}\end{aligned} \tag{2.20}$$

is called *multiobject PDF* (MPDF) of the RFS X.

Definition 3. Let $f : \mathcal{F}(\mathbb{R}^n) \mapsto \mathbb{R}$ be a finite set function. The *set-integral* of $f(\cdot)$ concentrated over $S \subseteq \mathbb{R}^n$ is defined as

$$\int_S f(\mathsf{x})\ \mathrm{d}\mathsf{x} \triangleq f(\varnothing) + \sum_{\eta=1}^{\infty}\frac{1}{\eta!}\int_{S^\eta} f(\{x_1,x_2,\dots,x_\eta\})\ \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_\eta. \tag{2.21}$$

Often the RHS will be expressed in short as

$$\sum_{\eta=0}^{\infty}\frac{1}{\eta!}\int_{S^\eta} f(\{x_1,x_2,\dots,x_\eta\})\ \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_\eta. \tag{2.22}$$

According to definitions 2 and 3, equations (2.19) are equivalent to

$$\begin{aligned}\beta_{\mathsf{X}}(S) &= p_{\mathsf{X}}(\varnothing) + \sum_{\eta=1}^{\infty} \frac{1}{\eta!} \int_{S^\eta} p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})\ \mathrm{d}x_1 \cdots \mathrm{d}x_\eta \\ &\triangleq \sum_{\eta=0}^{\infty} \frac{1}{\eta!} \int_{S^\eta} p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})\ \mathrm{d}x_1 \cdots \mathrm{d}x_\eta \triangleq \int_S p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x}\end{aligned} \tag{2.23}$$

which expresses the fact that the BMF is nothing but the set-integral of a MPDF.

2.5 Janossy densities

Let $p_{\mathsf{X}}(\{\cdot\})$ be a MPDF. Given the cardinality $\eta = |\mathsf{X} = \mathsf{x}|$, the function $p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\}) : \mathbb{R}^{n\times\eta} \mapsto \mathbb{R}^+$

$$\begin{aligned}p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\}) &\triangleq \begin{cases} p_N(0) & \text{if } \eta = 0 \\ \eta!\, p_{X_1,\ldots,X_\eta}(x_1, \ldots, x_\eta)\, p_N(\eta) & \text{if } \eta > 0 \end{cases} \\ &= \begin{cases} p_N(0) & \text{if } \eta = 0 \\ \sum_\sigma p_N(\eta)\, \tilde{p}_{X_1,\ldots,X_\eta}(x_{\sigma(1)}, \ldots, x_{\sigma(\eta)}) & \text{if } \eta > 0 \end{cases}\end{aligned} \tag{2.24}$$

is called *Janossy density* of order η. Equations (2.24) show how to compute the MPDF $p_{\mathsf{X}}(\{\cdot\})$ given the model $(p_N, \mathcal{S})$. On the other hand, it is also possible to recover the model $(p_N, \mathcal{S})$ given the MPDF $p_{\mathsf{X}}(\{\cdot\})$ by the following two operations:

- **cardinality marginalization**: since

$$\int p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})\ \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_\eta = \eta!\, p_N(\eta) \tag{2.25}$$

  it follows that the generic value $p_N(\eta)$ of the cardinality density is given by

$$p_N(\eta) = \frac{1}{\eta!} \int p_{\mathsf{X}}(\{x_1, x_2, \ldots, x_\eta\})\ \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_\eta \tag{2.26}$$

- **spatial marginalization**: since

$$p_{X_1,X_2,\ldots,X_\eta}(x_1, x_2, \ldots, x_\eta) = \frac{p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})}{\eta!\, p_N(\eta)} \tag{2.27}$$

it follows from (2.26) that the spatial density $p_{X_1,X_2,\ldots,X_\eta}(\cdot)$ can be computed as

$$p_{X_1,X_2,\ldots,X_\eta}(x_1,x_2,\ldots,x_\eta) = \frac{p_{\mathsf{X}}(\{x_1,x_2,\ldots,x_\eta\})}{\int p_{\mathsf{X}}(\{x_1,x_2,\ldots,x_\eta\})\ \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_\eta} \tag{2.28}$$

Equations (2.26) and (2.28) state that there is a 1-to-1 corrispondance between the model $(p_N,\mathcal{S})$ and the MPDF $p_{\mathsf{X}}(\cdot)$. So in conclusion, one can switch between the characterizations $(p_N,\mathcal{S})$ and $p_{\mathsf{X}}(\cdot)$ without loosing any information about the considered RFS X. The BMF $\beta_{\mathsf{X}}(\cdot)$ is a third and equivalent characterization of an RFS X, in the sense that there is a 1-to-1 corrispondance with the MPDF $p_{\mathsf{X}}(\cdot)$ or the model $(p_N,\mathcal{S})$: the transformation $p_{\mathsf{X}}(\cdot)\mapsto\beta_{\mathsf{X}}(\cdot)$ is provided by the set integration (2.23), on the other hand the inverse transformation $\beta_{\mathsf{X}}(\cdot)\mapsto p_{\mathsf{X}}(\cdot)$ is given by a new operation, which will be defined in the following section, that is the so-called *set differentation*.

2.6 Set derivatives

2.6.1 Lebesgue differentation

Consider an ordinary PMF $\mathbb{P}_X(\cdot)$ on $\mathbb{R}^n$ which admits a PDF $p_X(\cdot)$, so

$$\mathbb{P}_X(S) = \int_S p_X(x)\ \mathrm{d}x \qquad \forall\, S\in\mathcal{B}(\mathbb{R}^n) \tag{2.29}$$

the ordinary vector calculus tells that, given $\mathbb{P}_X(S)$, it is possible to recover $p_X(\cdot)$ via the ordinary vectorial differentation of the CDF

$$p_X(x) = \left.\frac{\mathrm{d}\mathbb{P}_X(S)}{\mathrm{d}x}\right|_{S=(-\infty,x]^n} = \left.\frac{\partial^n\mathbb{P}_X(S)}{\partial x_1\cdots\partial x_n}\right|_{S=(-\infty,x]^n} \tag{2.30}$$

unfortunately such method is not directy applicable to the RFS case, because the result of the ordinary differentation is not a MPDF but an ordinary PDF. However there is a more powerful procedure to recover $p_X(\cdot)$, which can be easily extended to the RFS case, called *Lebesgue differentation*.

Consider a neighborhood E_x of the generic point $x\in\mathbb{R}^n$, for example the hypercube $C_{x,\epsilon}\triangleq\{x'\in\mathbb{R}^n : |x'_i-x_i|<\epsilon/2 \text{ for } i=1,2,\ldots,n\}$ with small edge $\epsilon>0$. Since $E_x = C_{x,\epsilon}$ is a simple open subset of $\mathbb{R}^n$, so $C_{x,\epsilon}\in\mathcal{B}(\mathbb{R}^n)$, it holds that

$$\mathbb{P}_X(C_{x,\epsilon}) = \int_{C_{x,\epsilon}} p_X(\xi)\ \mathrm{d}\xi. \tag{2.31}$$

Now observe that, since ϵ is small, its possible to approximate, in the entire neighborhood $B_{x,\epsilon}$, the density $p_X(\cdot)$ with a constant density with value $p_X(x)$[4]

$$\mathbb{P}_X(C_x) \approx \int_{C_{x,\epsilon}} p_X(x)\,\mathrm{d}\xi = p_X(x) \int_{C_{x,\epsilon}} \mathrm{d}\xi = p_X(x)\,\epsilon^n. \tag{2.32}$$

Hence it holds that

$$p_X(x) \approx \frac{\mathbb{P}_X(C_{x,\epsilon})}{\epsilon^n} \tag{2.33}$$

which, if the limit exists, takes the following strong form for $\epsilon \downarrow 0$

$$p_X(x) = \lim_{\epsilon\downarrow 0} \frac{\mathbb{P}_X(C_{x,\epsilon})}{\epsilon^n}. \tag{2.34}$$

In conclusion, the Lebesgue differentation of $\mathbb{P}_X(\cdot)$ consists of the following transformation $\mathbb{P}_X(\cdot) \mapsto p_X(\cdot)$

$$\frac{\mathrm{d}\mathbb{P}_X(x)}{\mathrm{d}x} \triangleq \lim_{\epsilon\downarrow 0} \frac{\mathbb{P}_X(C_{x,\epsilon})}{\epsilon^n}. \tag{2.35}$$

The function $\frac{\mathrm{d}\mathbb{P}_X(\cdot)}{\mathrm{d}x}$ is called *Radon-Nykodim derivative* of $\mathbb{P}_X(\cdot)$ in x and, by construction, is the PDF $p_X(\cdot)$. Note that the limit is performed only from above since the edge ϵ is a non-negative quantity.

The Lebesgue differentation can be expressed in a more general form which is suitable for the next extension to the RFS domain. Consider a generic open subset S of $\mathbb{R}^n$ such that $S \cap C_{x,\epsilon} = \varnothing$, i.e. a set S disjoint from the small hyper-cube $C_{x,\epsilon}$. Thanks to the additivity of the PMF $\mathbb{P}_X(\cdot)$ (it is a probability measure), it is possible to write

$$\mathbb{P}_X(S \cup C_{x,\epsilon}) = \mathbb{P}_X(S) + \mathbb{P}_X(C_{x,\epsilon}) \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) : S \cap C_{x,\epsilon} = \varnothing \tag{2.36}$$

from which it follows immediately that

$$\lim_{\epsilon\downarrow 0} \frac{\mathbb{P}_X(C_{x,\epsilon})}{\epsilon^n} = \lim_{\epsilon\downarrow 0} \frac{\mathbb{P}_X(S \cup C_{x,\epsilon}) - \mathbb{P}_X(S)}{\epsilon^n} \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) : S \cap C_{x,\epsilon} = \varnothing. \tag{2.37}$$

Thus the transformation over $\mathbb{P}_X(\cdot)$

$$\frac{\mathrm{d}\mathbb{P}_X(S)}{\mathrm{d}x} \triangleq \lim_{\epsilon\downarrow 0} \frac{\mathbb{P}_X(S \cup C_{x,\epsilon}) - \mathbb{P}_X(S)}{\epsilon^n} \qquad \forall\, S \in \mathcal{O}(\mathbb{R}^n) : S \cap C_{x,\epsilon} = \varnothing \tag{2.38}$$

[4] such approximation is valid if $\mathbb{P}_X(\cdot)$ is sufficiently smooth, for example if it is absolutely continuos

also returns the density $p_X(\cdot)$ like (125). In this case, the function $\frac{\mathrm{d}\mathbb{P}_X(\cdot)}{\mathrm{d}x}$ is called *generalized Radon-Nykodim derivative* of $\mathbb{P}_X(\cdot)$ in x.

The constraint $S \cap C_{x,\epsilon} = \varnothing$ can also be seen as a constraint on x given S: equation (2.38) is valid only when the point x moves outside S. In fact, if x is inside S then for ϵ sufficiently small it holds that $C_{x,\epsilon} \subset S$, so $S \cup C_{x,\epsilon} = S$ and equation (2.38) vanishes.

$$\frac{\mathrm{d}\mathbb{P}_X(S)}{\mathrm{d}x} \triangleq \lim_{\epsilon \downarrow 0} \frac{\mathbb{P}_X(S \cup C_{x,\epsilon}) - \mathbb{P}_X(S)}{\epsilon^n} \qquad \forall\, S \not\ni x \tag{2.39}$$

In the MOT this is limitation because there is a case (see formulas (2.69)) where the differentation needs to be evaluated on $S = \mathbb{R}^n$, so that equation (2.39) will be generalized furthermore.

The BMF, likewise the PMF, is an ordinary set function of the type $\Phi : \mathcal{O}(\mathbb{R}^n) \mapsto \mathbb{R}^+$, so the generalized Lebesgue differentation can also be applied to a BMF leading to the following generalized Radon-Nykodim derivative

$$\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}x} \triangleq \lim_{\epsilon \downarrow 0} \frac{\beta_{\mathsf{X}}(S \cup C_{x,\epsilon}) - \beta_{\mathsf{X}}(S)}{\epsilon^n} \qquad \forall\, S \not\ni x^n \tag{2.40}$$

2.6.2 Properties of the Lebesgue differentation

The generalized Lebesgue differentation obeys to the ordinary differentation rules, for example:

- **constant rule**: let $\Phi(S) = K$ be a constant set function. Then

$$\frac{\mathrm{d}K}{\mathrm{d}x} = 0 \tag{2.41}$$

- **linear rule**: let $\Phi(S) = \int_S f(x)\, \mathrm{d}x$ be a set function induced by the density $f(\cdot)$. Then

$$\frac{\mathrm{d}\Phi(S)}{\mathrm{d}x} = f(x) \tag{2.42}$$

  note that this fact is true because the generalized Lebesgue differentation is defined as the operation that extracts the density from a set function of the type $\Phi(S) = \int_S f(x)\, \mathrm{d}x$;

- **sum rule**: let $\Phi_1(S) = \int_S f_1(x)\, \mathrm{d}x$, $\Phi_2(S) = \int_S f_2(x)\, \mathrm{d}x$ be set functions induced by the densities $f_1(\cdot)$, $f_2(\cdot)$ and let a_1, a_2 be real numbers. Then

$$\frac{\mathrm{d}[a_1\, \Phi_1(S) + a_2\, \Phi_2(S)]}{\mathrm{d}x} = a_1\, \frac{\mathrm{d}\Phi_1(S)}{\mathrm{d}x} + a_2\, \frac{\mathrm{d}\Phi_2(S)}{\mathrm{d}x} = a_1\, f_1(x) + a_2\, f_2(x) \tag{2.43}$$

in other words, the generalized Lebesgue differentation is a linear transformation of set functions.

- **monomial rule**: let $\Phi(S) = \int_S f(x)\ \mathrm{d}x$ be a set function induced by the density $f(\cdot)$ and let k be an integer. Then

$$\frac{\mathrm{d}[\Phi(S)]^k}{\mathrm{d}x} = k\,[\Phi(S)]^{k-1}\frac{\mathrm{d}\Phi(S)}{\mathrm{d}x} = k\,[\Phi(S)]^{k-1}\,f(x) \tag{2.44}$$

- **product rule**: let $\Phi_1(S) = \int_S f_1(x)\ \mathrm{d}x$, $\Phi_2(S) = \int_S f_2(x)\ \mathrm{d}x$ be set functions inducted by the densities $f_1(\cdot)$, $f_2(\cdot)$. Then

$$\begin{aligned}\frac{\mathrm{d}[\Phi_1(S)\,\Phi_2(S)]}{\mathrm{d}x} &= \frac{\mathrm{d}\Phi_1(S)}{\mathrm{d}x}\,\Phi_2(S) + \Phi_1(S)\,\frac{\mathrm{d}\Phi_2(S)}{\mathrm{d}x} \\ &= f_1(x)\,\Phi_2(S) + \Phi_1(S)\,f_2(x)\end{aligned} \tag{2.45}$$

- **chain rule**: let $\Phi(S) = \int_S f(x)\ \mathrm{d}x$ be a set function inducted by the density $f_1(\cdot)$ and let $\phi(\cdot)$ be a real valued function. Then

$$\frac{\mathrm{d}[f(\Phi_1(S))]}{\mathrm{d}x} = \left.\frac{\mathrm{d}\varphi(y)}{\mathrm{d}y}\right|_{y=\Phi(S)}\frac{\mathrm{d}\Phi(S)}{\mathrm{d}x} = \left.\frac{\mathrm{d}\varphi(y)}{\mathrm{d}y}\right|_{y=\Phi(S)} f(x). \tag{2.46}$$

Note that $\frac{\mathrm{d}\varphi(\cdot)}{\mathrm{d}y}$ is the ordinary derivative of the ordinary function $\varphi(\cdot)$.

2.6.3 Set differentation

The objective of this section is to define an operation $\beta_{\mathsf{X}}(\cdot) \mapsto p_{\mathsf{X}}(\cdot)$ that allows to extract the relative MPDF $p_{\mathsf{X}}(\cdot)$ from a given BMF $\beta_{\mathsf{X}}(\cdot)$. In order to get the solution to this problem, it is useful to firstly understand how to recover from the BMF $\beta_{\mathsf{X}}(\cdot)$ the first 3 Janossy densities $p_{\mathsf{X}}(\varnothing)$, $p_{\mathsf{X}}(\{x_1\})$, $p_{\mathsf{X}}(\{x_1, x_2\})$.

- **Janossy density of order 0**: consider the expression

$$\beta_{\mathsf{X}}(S) = p_{\mathsf{X}}(\varnothing) + \sum_{\eta=1}^{\infty}\frac{1}{\eta!}\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})\ \mathrm{d}x_1\cdots\mathrm{d}x_\eta. \tag{2.47}$$

By setting $S = \varnothing$ turns out for any $\eta > 0$ that

$$\int_{\varnothing^\eta} p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})\ \mathrm{d}x_1\cdots\mathrm{d}x_\eta = 0 \tag{2.48}$$

thus the zero order Janossy density $p_{\mathsf{X}}(\varnothing)$ is directly given by the BMF $\beta_{\mathsf{X}}(\cdot)$ restricted to the empty set $\varnothing$

$$p_{\mathsf{X}}(\varnothing) = \beta_{\mathsf{X}}(\varnothing) \tag{2.49}$$

- **Janossy density of order 1**: the first order derivative (in the generalized Lebsegue sense) of the generic BMF $\beta_{\mathsf{X}}(\cdot)$ is

$$\begin{aligned}\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}x_1} &= \frac{\mathrm{d}}{\mathrm{d}x_1}\left[p_{\mathsf{X}}(\varnothing) + \sum_{\eta=1}^{\infty}\frac{1}{\eta!}\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\ldots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right] \\ &= \underbrace{\frac{\mathrm{d}p_{\mathsf{X}}(\varnothing)}{\mathrm{d}x_1}}_{=0} + \underbrace{\frac{\mathrm{d}}{\mathrm{d}x_1}\left[\int_S p_{\mathsf{X}}(\{x_1\})\,\mathrm{d}x_1\right]}_{=p_{\mathsf{X}}(\{x_1\})} \\ &\qquad + \sum_{\eta=2}^{\infty}\frac{1}{\eta!}\frac{\mathrm{d}}{\mathrm{d}x_1}\left[\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\ldots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right]\end{aligned} \tag{2.50}$$

as one can show, setting $S = \varnothing$ leads for any $\eta > 1$ to

$$\frac{\mathrm{d}}{\mathrm{d}x_1}\left[\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\ldots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right]_{S=\varnothing} = 0 \tag{2.51}$$

so that the following relation holds

$$p_{\mathsf{X}}(\{x_1\}) = \left.\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}x_1}\right|_{S=\varnothing} \tag{2.52}$$

which expresses the fact that the first order Janossy density $p_{\mathsf{X}}(\{x_1\})$ is given by the following couple of operations

1. differentation of the BMF $\beta_{\mathsf{X}}(\cdot)$ with respect x_1;
2. restriction of the derivative $\frac{\mathrm{d}\beta_{\mathsf{X}}(\cdot)}{\mathrm{d}x_1}$ to the empty set $\varnothing$.

- **Janossy density of order 2**: the second order derivative (in the generalized Lebsegue sense) of the generic BMF $\beta_{\mathsf{X}}(\cdot)$ is defined as

$$\frac{\mathrm{d}^2\beta_{\mathsf{X}}(S)}{\mathrm{d}x_2\,\mathrm{d}x_1} \triangleq \frac{\mathrm{d}}{\mathrm{d}x_2}\left[\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}x_1}\right] \tag{2.53}$$

thus, by recalling (2.23),

$$\begin{aligned}\frac{\mathrm{d}^2\beta_{\mathsf{X}}(S)}{\mathrm{d}x_2\,\mathrm{d}x_1} &= \frac{\mathrm{d}}{\mathrm{d}x_2}\left[p_{\mathsf{X}}(\{x_1\}) + \sum_{\eta=2}^{\infty}\frac{1}{\eta!}\frac{\mathrm{d}}{\mathrm{d}x_1}\left[\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right]\right]\\ &= \underbrace{\frac{\mathrm{d}p_{\mathsf{X}}(\{x_1\})}{\mathrm{d}x_2}}_{=0} + \frac{1}{2}\underbrace{\frac{\mathrm{d}^2}{\mathrm{d}x_2\,\mathrm{d}x_1}\left[\int_{S^2} p_{\mathsf{X}}(\{x_1,x_2\})\,\mathrm{d}x_1\,\mathrm{d}x_2\right]}_{=2p_{\mathsf{X}}(\{x_1,x_2\})} \\ &\quad + \sum_{\eta=3}^{\infty}\frac{1}{\eta!}\frac{\mathrm{d}^2}{\mathrm{d}x_2\,\mathrm{d}x_1}\left[\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right]\end{aligned} \quad . \tag{2.54}$$

Finally, setting $S=\varnothing$ implies for any $\eta>2$ that

$$\frac{\mathrm{d}}{\mathrm{d}x_2\,\mathrm{d}x_1}\left[\int_{S^\eta} p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})\,\mathrm{d}x_1\cdots\mathrm{d}x_\eta\right]_{S=\varnothing} = 0 \tag{2.55}$$

so that

$$p_{\mathsf{X}}(\{x_1,x_2\}) = \left.\frac{\mathrm{d}^2\beta_{\mathsf{X}}(S)}{\mathrm{d}x_2\,\mathrm{d}x_1}\right|_{S=\varnothing} \quad . \tag{2.56}$$

The second order Janossy density $p_{\mathsf{X}}(\{x_1,x_2\})$ is, once again, given by the following couple of operations

1. differentation of the BMF $\beta_{\mathsf{X}}(\cdot)$ in x_2 and x_1 or, more coincisely, in the finite set $\mathsf{x} \triangleq \{x_1,x_2\}$;
2. restriction of the derivative $\frac{\mathrm{d}\beta_{\mathsf{X}}(\cdot)}{\mathrm{d}\mathsf{x}} = \frac{\mathrm{d}^2\beta_{\mathsf{X}}(\cdot)}{\mathrm{d}x_2\,\mathrm{d}x_1}$ to the empty set $\varnothing$

In general, the generic Janossy density $p_{\mathsf{X}}(\{x_1,\dots,x_\eta\})$ is given by the 2-step procedure

1. differentation of the BMF $\beta_{\mathsf{X}}(\cdot)$ in x_η, $x_{\eta-1}$, ..., x_1 or, in short, in the finite set $\mathsf{x} \triangleq \{x_1,\dots,x_\eta\}$. Such derivative is defined recursively via the rule

$$\frac{\mathrm{d}^\eta\beta_{\mathsf{X}}(S)}{\mathrm{d}x_\eta\,\mathrm{d}x_{\eta-1}\dots\mathrm{d}x_1} \triangleq \frac{\mathrm{d}}{\mathrm{d}x_\eta}\left[\frac{\mathrm{d}^{\eta-1}\beta_{\mathsf{X}}(S)}{\mathrm{d}x_{\eta-1}\,\mathrm{d}x_{\eta-2}\dots\mathrm{d}x_1}\right] \tag{2.57}$$

2. restriction of the derivative $\frac{\mathrm{d}\beta_{\mathsf{X}}(\cdot)}{\mathrm{d}\mathsf{x}} = \frac{\mathrm{d}^\eta\beta_{\mathsf{X}}(\cdot)}{\mathrm{d}x_\eta\,\mathrm{d}x_{\eta-1}\dots\mathrm{d}x_1}$ to the empty set $\varnothing$

Definition 4. Let $\Phi : \mathcal{O}(\mathbb{R}^n) \mapsto \mathbb{R}$ be a set function. Then, if it exists, the *set derivative* of $\Phi(\cdot)$ is the set function $\frac{\mathrm{d}\Phi}{\mathrm{d}\mathsf{x}} : \mathcal{O}(\mathbb{R}^n) \mapsto \mathbb{R}$ defined as

$$\frac{\mathrm{d}\Phi(S)}{\mathrm{d}\mathsf{x}} \triangleq \begin{cases}\Phi(S) & \text{if } \mathsf{x}=\varnothing\\ \frac{\mathrm{d}^\eta\Phi(S)}{\mathrm{d}x_\eta\cdots\mathrm{d}x_1} & \text{if } \mathsf{x}=\{x_1,\dots,x_\eta\}\end{cases} \qquad \forall\, S \not\supseteq \mathsf{x} \tag{2.58}$$

where $\frac{\mathrm{d}^\eta \Phi(S)}{\mathrm{d}x_\eta \cdots \mathrm{d}x_1}$ is the η-th generalized Radon-Nykodim derivative of $\Phi(\cdot)$.

Thanks to the definition of set derivative, the inverse operation of set-integration is expressed coincisely as set-differentation, which consists in the computation of every generalized Radon-Nykodim derivatives restricted to the empty set.

This means that integrating an MPDF $p_{\mathsf{X}}(\{\cdot\})$ produces the relative BMF $\beta_{\mathsf{X}}(\cdot)$, while evaluating on the empty set $S = \varnothing$ the set-derivative of a BMF $\beta_{\mathsf{X}}(\cdot)$ provides the relative MPDF $p_{\mathsf{X}}(\{\cdot\})$

$$p_{\mathsf{X}}(\mathsf{x}) \xrightarrow{\int_S (\cdot)\, \mathrm{d}\mathsf{x}} \int_S p_{\mathsf{X}}(\mathsf{x})\, \mathrm{d}\mathsf{x} \triangleq \beta_{\mathsf{X}}(S) \xrightarrow{\left.\frac{\mathrm{d}(\cdot)}{\mathrm{d}x}\right|_{S=\varnothing}} p_{\mathsf{X}}(\mathsf{x}) \tag{2.59}$$

2.7 Probability Hypothesis Density

2.7.1 Definition

The so-called probability hypothesis density (PHD) can be regarded as the RFS counterpart of the concept of expected value. Its definition is not straightforward, in fact consider the naively

$$\mathbb{E}[\mathsf{X}] = \int \mathsf{x}\, p_{\mathsf{X}}(\mathsf{x})\, \mathrm{d}\mathsf{x} = \sum_{\eta=0}^{\infty} \frac{1}{\eta!} \int \{x_1, \ldots, x_\eta\}\, p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})\, \mathrm{d}x_1 \cdots \mathrm{d}x_\eta. \tag{2.60}$$

Clearly, the product $\{x_1, \ldots, x_\eta\}\, p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})$ is not defined because $\{x_1, \ldots, x_\eta\}$ is a set rather than a vector, so that such approach cannot be considered. On the other hand, a well-defined concept is the expected value of a multiobject function $f : \mathcal{F}(\mathbb{R}^n) \mapsto \mathbb{R}$:

$$\begin{aligned} \mathbb{E}[f(\mathsf{X})] &\triangleq \int f(\mathsf{x})\, p_{\mathsf{X}}(\mathsf{x})\, \mathrm{d}\mathsf{x} \\ &= \sum_{\eta=0}^{\infty} \frac{1}{\eta!} \int f(\{x_1, \ldots, x_\eta\})\, p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\})\, \mathrm{d}x_1 \cdots \mathrm{d}x_\eta \end{aligned} \tag{2.61}$$

since now the products, the sums and the integrations are all performed between two multiobject functions $f(\{\cdot\})$, $p_{\mathsf{X}}(\{\cdot\})$. Due to this fact, the standard approach adopted by FISST in order to define the expected value of an RFS is to replace the RFS X with a "look-like" identity function $f(\{\cdot\})$ for the considerd RFS X. The (quite natural) choice of FISST for $f(\{\cdot\})$ is to consider the indicator function $\delta_{(\{\cdot\})}(x)$, which is defined as follows

$$\mathsf{x} \to \delta_{\mathsf{X}}(x) \triangleq \begin{cases} \sum_{x_i \in \mathsf{X}} \delta_i(x) & \text{if } \mathsf{x} \neq \varnothing \\ 0 & \text{if } \mathsf{x} = \varnothing \end{cases} = \begin{cases} \sum_{x_i \in \mathsf{X}} \delta(x_i - x) & \text{if } \mathsf{x} \neq \varnothing \\ 0 & \text{if } \mathsf{x} = \varnothing \end{cases} \tag{2.62}$$

where $\delta_i(\cdot)$ is the Dirac delta concentrated at the point x_i. Note that $\delta_{\mathsf{X}}(x)$ is a function of the RFS X and not of the vector x: the latter has to be regarded as a fixed external parameter. As a result, the expected value of an RFS X, that is the PHD, is not a finite set but a standard function

Definition 5. Let X be an RFS characterized by the MPDF $p_{\mathsf{X}}(\{\cdot\})$. The *probability hypothesis density* (PHD) or *intensity* of the RFS X is the function $D_{\mathsf{X}} : \mathbb{R}^n \mapsto \mathbb{R}$ defined as follows

$$D_{\mathsf{X}}(x) \triangleq \mathbb{E}[\delta_{\mathsf{X}}(x)] = \int \delta_{\mathsf{X}}(x)\, p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x}. \tag{2.63}$$

Notice that in the computation of the set-integral that defines a PHD, the argument x has to be regarded as a fixed parameter, while the integration variable is rather the finite set x. Note that in order to know the PHD one has to compute a different set integral for every possible value of x.

2.7.2 Practical interpretation

According to definition 5, the integral of a PHD $D_{\mathsf{X}}(\cdot)$ over a region $S \in \mathcal{O}(\mathbb{R}^n)$ is

$$\begin{aligned} \int_S D_{\mathsf{X}}(x)\ \mathrm{d}x &= \int_S \int \delta_{\mathsf{x}}(x)\, p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x}\mathrm{d}x \\ &= \int \left(\int_S \delta_{\mathsf{x}}(x)\ \mathrm{d}x \right) p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} \\ &= \int \left(\int_S \sum_{x_i \in \mathsf{x}} \delta(x_i - x)\ \mathrm{d}x \right) p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} \\ &= \int \left(\sum_{x_i \in \mathsf{x}} \int_S \delta(x_i - x)\ \mathrm{d}x \right) p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} \end{aligned}. \tag{2.64}$$

Now, by observing that

$$\int_S \delta(x_i - x)\ \mathrm{d}x = \begin{cases} 1 & \text{if } x_i \in S \\ 0 & \text{otherwise} \end{cases} \tag{2.65}$$

it turns out that

$$\sum_{x_i \in \mathsf{x}} \int_S \delta(x_i - x) \, \mathrm{d}x = |\mathsf{x} \cap S| \tag{2.66}$$

so that the combination of (2.66) with (2.64) yields to

$$\int_S D_{\mathsf{X}}(x) \, \mathrm{d}x = \int |\mathsf{x} \cap S| \, p_{\mathsf{X}}(\mathsf{x}) \, \mathrm{d}\mathsf{x} = \mathbb{E}[|\mathsf{X} \cap S|] \quad . \tag{2.67}$$

Hence, in conclusion, the PHD $D_{\mathsf{X}}(\cdot)$ is a function that, if integrated over a region S, expresses how many elements of the considered RFS are expected in S.

In analogy to the ordinary concept of PDF, the PHD $D_{\mathsf{X}}(\cdot)$ is a function (more precisely, a density) which tends to take large values in regions S that most likely contain at least one element of X. Note however that $D_{\mathsf{X}}(\cdot)$ is not a PDF because in general does not integrate to the unity. In fact, by setting $S = \mathbb{R}^n$, it turns out that the integral of $D_{\mathsf{X}}(\cdot)$ is the expected cardinality of the considered RFS X, i.e.,

$$\int D_{\mathsf{X}}(x) \, \mathrm{d}x = \mathbb{E}[|\mathsf{X}|] \tag{2.68}$$

is a generic, not necessarely integer, non-negative number.

2.7.3 Relationship with the BMF

Given the BMF $\beta_{\mathsf{X}}(\cdot)$ of an RFS X, the PHD $D_{\mathsf{X}}(\cdot)$ can be found via the following set-differentation

$$D_{\mathsf{X}}(x) = \left. \frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}\{x\}} \right|_{S=\mathbb{R}^n} . \tag{2.69}$$

Note that, in principle,it is not well defined because the definition of generalized Radon-Nykodim derivative requires that the point x where is differentation is performed be outside the region S where the set function under differentation is evaluated (and this cannot occur since $S = \mathbb{R}^n$). In this situation, it is necessary to use a more general definition of derivative, which, in particular, is

$$\frac{\mathrm{d}\beta_{\mathsf{X}}(S)}{\mathrm{d}x} \triangleq \lim_{\epsilon_1, \epsilon_2 \downarrow 0} \frac{\beta_{\mathsf{X}}((S \setminus C^2_{x,\epsilon_2}) \cup C^1_{x,\epsilon_1}) - \beta_{\mathsf{X}}(S \setminus C^2_{x,\epsilon_2})}{\epsilon_1^n} \tag{2.70}$$

where $C^1_{x,\epsilon_1}, C^2_{x,\epsilon_2}$ are hypercubes with edges ϵ_1, $\epsilon_2 > \epsilon_1$ centered in x Such definition works for every possible choice of x and S.

Hence, equation (2.69) in combination with the newly defined Radon-Nykodim derivative tells how to compute the PHD $D_{\mathsf{X}}(\cdot)$ as a function of the BMF $\beta_{\mathsf{X}}(\cdot)$. Conversly, the BMF $\beta_{\mathsf{X}}(\cdot)$ cannot be recovered from the PHD $D_{\mathsf{X}}(\cdot)$, meaning that there is not a 1-to-1 corrispondance between $\beta_{\mathsf{X}}(\cdot)$ and $D_{\mathsf{X}}(\cdot)$.

The PHD $D_{\mathsf{X}}(\cdot)$ does not provide a full statistical characterization of an RFS X, but a synthetic representation of X consting only of the most important information. In this sense, the PHD $D_{\mathsf{X}}(\cdot)$ can be regarded as a first order statistical description of X, while the BMF $\beta_{\mathsf{X}}(\cdot)$ (or the MPDF $p_{\mathsf{X}}(\{\cdot\})$ or the model $(p_N, \mathcal{S})$) is a full statistical description of X.

2.8 RFS models

2.8.1 IID RFS

Let $p_N(\cdot)$ be a discrete density and let $p_X(\cdot)$ be a symmetric PDF over $\mathbb{R}^n$. An RFS X is called *independent and identically distributed* (IID) with cardinality $p_N(\cdot)$ and spatial density $p_X(\cdot)$ if and only if its MPDF takes the form

$$p_{\mathsf{X}}(\mathsf{x}) = |\mathsf{x}|!\, p_N(|\mathsf{x}|) \prod_{x \in \mathsf{x}} p_X(x) \tag{2.71}$$

by set-integrating the MPDF turns out that the BMF is

$$\beta_{\mathsf{X}}(S) = \sum_{\eta=0}^{\infty} \left(\int_S p_X(x) \, \mathrm{d}x \right)^{\eta} p_N(\eta) \tag{2.72}$$

2.8.2 Poisson RFS

Let X be an IID RFS. The RFS X is called *Poisson* if and only if its cardinality is Poisson for some parameter $\lambda > 0$, i.e.

$$p_N(\eta) = \mathrm{Po}_{\lambda}(\eta) \triangleq \exp(-\lambda) \frac{\lambda^{\eta}}{\eta!}. \tag{2.73}$$

It turns out immediately that the multiobject PDF of a Poisson RFS is

$$p_{\mathsf{X}}(\mathsf{x}) = \exp(-\lambda) \lambda^{|\mathsf{x}|} \prod_{x \in \mathsf{x}} p_X(x). \tag{2.74}$$

By introducing the so-called *intensity function* $I(\cdot) \triangleq \lambda p_X(\cdot)$, the expression of the multiobject PDF of a Poisson RFS simplifies to

$$p_{\mathsf{X}}(\mathsf{x}) = \exp(-\lambda) \prod_{x \in \mathsf{x}} I(x). \tag{2.75}$$

Note that the integral of the intensity function provides an estimate of the cardinality of the RFS X, indeed it holds that

$$\int I(x)\ \mathrm{d}x = \lambda = \mathbb{E}_{p_N}[N]. \tag{2.76}$$

This fact explains the reason why $I(\cdot)$ is called intensity of the RFS X: if $I(\cdot)$ has a large integral then the mean value of the number of vectors contained in X is large ($\equiv \mathsf{X}$ is 'intense').

By set-integrating the MPDF follows that the BMF is

$$\beta_{\mathsf{X}}(S) = \exp\left(-\lambda + \int_S I(x)\ \mathrm{d}x\right) \tag{2.77}$$

on the other hand, by set-differentiating the BMF follows that the PHD is

$$D(x) = I(x) \tag{2.78}$$

2.8.3 Bernoulli RFS

Let X be an IID RFS. The RFS X is called *Bernoulli* if and only if its cardinality is Bernoulli for some parameter $p \in [0,1]$ called *probability of existence*, i.e.

$$p_N(\eta) = \mathrm{Ber}_p(\eta) \triangleq (1-p)\delta_0(\eta) + p\delta_1(\eta). \tag{2.79}$$

It turns out immediately that the multiobject PDF of a Bernoulli RFS is

$$p_{\mathsf{X}}(\mathsf{x}) = \left[(1-p)\,\delta_0(|\mathsf{x}|) + p\,p_X(\mathsf{x})\,\delta_1(|\mathsf{x}|)\right]. \tag{2.80}$$

Note that $p_{\mathsf{X}}(\mathsf{x}) = 0$ if $|\mathsf{x}| > 2$. Set-integrating the BMF gives the following BMF

$$\beta_{\mathsf{X}}(S) = 1 - p + p\int_S p_X(x)\ \mathrm{d}x \tag{2.81}$$

moreover the PHD is

$$D(x) = p\,p_X(x) \tag{2.82}$$

2.8.4 Multi-Bernoulli RFS

Let be $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_m$ be a finite family of independent Bernoulli RFSs with probabilities of existence $p_1,\ p_2,\ \ldots,\ p_m$ respectively. An RFS X is called *multi-Bernoulli* with components $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_m$ if and only if

$$\mathsf{X} = \bigcup_{i=1}^{m} \mathsf{X}_i. \tag{2.83}$$

Due to independence, the BMF is

$$\beta_{\mathsf{X}}(S) = \prod_{i=1}^{m} \beta_{\mathsf{X}_i}(S) = \prod_{i=1}^{m} \left(1 - p_i + p_i \int_S p_{X_i}(x) \, \mathrm{d}x \right). \tag{2.84}$$

One can show via set-differentation that, for $\eta = 0, 1, \ldots, m$, the generic Janossy density is

$$p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\}) = \sum_{A \in F_\eta^m} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i) \left(\sum_{\sigma} p_{X_i}(x_{\sigma(i)}) \right) \tag{2.85}$$

where $F_\eta^m \triangleq \{A : A \subseteq \{1, 2, \ldots, m\}, |A| = \eta\}$, while $p_{\mathsf{X}}(\{x_1, \ldots, x_\eta\}) = 0$ for any $\eta > m$.

By set differentiating the BMF, turns out that the cardinality density of a multi-Bernoulli RFS is Poisson-Binomial with parameters $p_1, \ldots, p_m$, i.e.

$$\begin{aligned} p_N(\eta) &= \mathrm{PB}_{p_1, \ldots, p_m}(\eta) \\ &\triangleq \begin{cases} \sum_{A \in F_\eta^m} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i) & \text{if } \eta \in \{0, 1, \ldots, m\} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2.86}$$

the PHD is

$$D(x) = \sum_{i=1}^{m} p_i \, p_{X_i}(x) \tag{2.87}$$

Chapter 3

Generalized multiobject calculus

3.1 Summary

In order to clarify the derivation of the PHD filter, the main objective of this chapter is to provide a good understanding of the meaning of PGFL and a good understanding of the main tool used to manipulate them, i.e. the the functional derivative. The PGFLs are RFS descriptions equivalent to the BMFs or the MPDFs, but they have the advantage to greatly simplify some calculations in the derivation of the PHD filter. Without the PGFLs and the functional derivatives, such derivation is much harder even if not impossible.

The main concepts of the chapter are the following:

- the definitions of functional and PGFL. In the first part of the chapter such definitions are given and then the PGFLs of the most common RFSs are presented;

- the definition of functional derivative. This is the tool used to extract the PHD density from a PGFL. For this reason, the concept of functional derivative plays a central role in the derivation of the PHD filter;

- properties of functional derivatives. Instead to apply the definitions, the functional derivatives can be computed more easily according to some rules presented in the final part of the chapter.

3.2 Functionals

A *functional* $F[\cdot]$ is a function of the type $F : T \mapsto \mathbb{R}$, where T is the following set of functions

$$T \triangleq \{h : \mathbb{R}^n \mapsto \mathbb{R} : 0 \leq h(x) \leq 1 \quad \forall\, x \in \mathbb{R}^n\}. \tag{3.1}$$

In simple words, a functional is a transformation $F[\cdot]$ that associates a special function $h(\cdot)$, called *test function*, to a real number. For example, the following tranformations are functionals

$$f[h] \triangleq \int h(x)\, f(x)\; \mathrm{d}x \tag{3.2}$$

$$h^{\mathsf{X}} \triangleq \begin{cases} 1 & \text{if } \mathsf{X} = \varnothing \\ \prod_{x \in \mathsf{X}} h(x) & \text{if } \mathsf{X} \neq \varnothing \end{cases} \tag{3.3}$$

called respectively *linear functional* and *power functional.* If $f(\cdot)$ is a PDF then the linear functional assumes the meaning of expected value of the test function $h(\cdot)$. In particular, if X is a random vector with PDF $f(\cdot)$, then the probability of the event $X \in S$ can be written as the following linear functional

$$\mathbb{P}(X \in S) = \mathbb{P}_X(S) \triangleq \int_S f(x)\; \mathrm{d}x = \int 1_S(x)\, f(x)\; \mathrm{d}x = f[1_S] \tag{3.4}$$

where 1_S is the indicator function of the open set S ($1_S(x) \triangleq 1$ iff $x \in S$, 0 otherwise). It is easy to see that the linear functional, as suggested by the name, is linear, i.e. $f[a\, h_1 + b\, h_2] = a\, f[h_1] + b\, f[h_2]$ for any choice of scalars a, b and test functions h_1, h_2.

3.3 Probability Generating Functional

3.3.1 Definition and interpretations

The most important functional involved in FISST is the so-called *probability generating functional* (PGFL), defined as follows:

$$G[h] \triangleq \int h^{\mathsf{X}}\, p_{\mathsf{X}}(\mathsf{X})\; \mathrm{d}\mathsf{X} \tag{3.5}$$

where $h(\cdot)$ is a generic test function and $p_{\mathsf{X}}(\{\cdot\})$ is an MPDF. Essentially, the generic PGFL is a linear functional of the power functional (3.3), where the involved integral is a set integral and the considered density is the MPDF $p_{\mathsf{X}}(\{\cdot\})$, so that $G[h] = p_{\mathsf{X}}[h^{\mathsf{X}}]$. In this sense, a PGFL can be seen as the expected value of the power h^{X}, where X is a RFS whose MPDF is $p_{\mathsf{X}}(\{\cdot\})$, i.e.,

$$G[h] = \mathbb{E}[h^{\mathsf{X}}]. \tag{3.6}$$

Another interesting interpretation of the meaning of PGFL is provided by the following relationship

$$G[1_S] = \int 1_S^{\mathsf{x}}\, p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} = \int_S p_{\mathsf{X}}(\mathsf{x})\ \mathrm{d}\mathsf{x} = \beta(S) \tag{3.7}$$

where it is observed that the power of the indicator function $1_S(\cdot)$ does not vanish only when the dummy finite set $\mathsf{x} = \{x_1, \ldots, x_\eta\}$ is included in the region S. Equation (3.7) states that the PGFL is a generalized BMF, where the indicator function $1_S(\cdot)$ is replaced by a generic test function $h(\cdot)$. For this reason, the multiobject calculus involving the PFGLs is referred to as "generalized".

Note that the PGFL represents a full characterization of an RFS: given a PGFL, equation (3.7) expresses how to recover the BMF. On the other hand, given a BMF one can recover the MPDF (via set differentation) and then compute the PGFL according to the definition.

3.3.2 Properties

- **(Linearity)**: Let $G_{a\,\mathsf{X}_1+b\,\mathsf{X}_2}[\cdot]$ be the PGFL of the linear combination $a\,p_{\mathsf{X}_1}(\{\cdot\})+b\,p_{\mathsf{X}_2}(\{\cdot\})$ of the two MPDFs $p_{\mathsf{X}_1}(\{\cdot\})$, $p_{\mathsf{X}_2}(\{\cdot\})$. Let $G_{\mathsf{X}_1}[\cdot]$, $G_{\mathsf{X}_2}[\cdot]$ be the PGFLs of the MPDFs $p_{\mathsf{X}_1}(\{\cdot\})$, $p_{\mathsf{X}_2}(\{\cdot\})$. Then

$$G_{a\,\mathsf{X}_1+b\,\mathsf{X}_2}[h] = a\,G_{\mathsf{X}_1}[h] + b\,G_{\mathsf{X}_2}[h]. \tag{3.8}$$

  Note that this is a trivial consequence of the set integral linearity.

- **(Independent factorization)**: Let $\mathsf{X} = \cup_{i=1}^k \mathsf{X}_i$ be the union of k independent RFSs X_1, ..., X_k. Let $G_{\mathsf{X}}[\cdot]$ be the PGFL of X and let $G_{\mathsf{X}_i}[\cdot]$ be the PGFL of X_i. Then

$$G_{\mathsf{X}}[h] = \prod_{i=1}^{k} G_{\mathsf{X}_i}[h]. \tag{3.9}$$

3.4 Examples of PGFLs

3.4.1 Poisson PGFL

Let X be a Poisson RFS with intensity $I(\cdot) = \lambda\, p_X(\cdot)$. According to the definition, the PGFL of such RFS is

$$G[h] = \exp(I[h-1]) \tag{3.10}$$

where it is used the linear functional notation $I[h] \triangleq \int h(x)\, I(x)\, \mathrm{d}x$ and it is observed that

$$I[h] - \lambda = I[h-1] \tag{3.11}$$

.

3.4.2 IID PGFL

Let X be an IID RFS with cardinality $p_N(\cdot)$ and spatial density $p_X(\cdot)$. The PGFL of such RFS is

$$G[h] = [G_N(z)]_{z=p_X[h]} \quad . \tag{3.12}$$

Once again, $p_X[h]$ is the linear functional notation for $\int h(x)\, p_X(x)\, \mathrm{d}x$. Moreover it is introduced to so-called *probability generating function* (PGF) $G_N(\cdot)$ of the cardinality density $p_N(\cdot)$, which is defined as

$$G_N(z) \triangleq \sum_{\eta=0}^{\infty} z^{\eta}\, p_N(\eta) \tag{3.13}$$

and, essentially, it is the z-transform of $p_N(\cdot)$. In this sense, the concept of PGFL can be seen a RFS counterpart of the standard concept of PGF. Moreover, a PGFL can be also tought as a sort of z-tranform of a MPDF $p_{\mathsf{X}}(\{\cdot\})$.

3.4.3 Bernoulli PGFL

Let X be a Bernoulli RFS with probability of existence p and spatial density $p_X(\cdot)$. The PGFL of such RFS is given by

$$G[h] = 1 - p + p\, p_X[h] \tag{3.14}$$

3.4.4 Multi-Bernoulli PGFL

Let X be a multi-Bernoulli RFS with Bernoulli components $\{(p_i, p_{X_i}(\cdot))\}_{i=1}^m$. According to the independent factorization, the PGFL of such RFS is

$$G[h] = \prod_{i=1}^{m} (1 - p_i + p_i\, p_{X_i}[h]) \tag{3.15}$$

3.5 Functional derivatives

3.5.1 Heuristic definition

Likewise the set derivative can be used to extract the MPDF from a BMF, the functional derivative can be used to extract the MPDF from a PGFL. Moreover, the functional derivative can be used to extract the PHD from a PGFL.

Recall the definition of directional derivative (usually called *Frechet derivative*) for functions $F : \mathbb{R}^n \mapsto \mathbb{R}$

$$\frac{\partial F}{\partial w}(x) \triangleq \lim_{\epsilon \to 0} \frac{F(x + \epsilon w) - F(x)}{\epsilon} \tag{3.16}$$

which quanties the how much the function $F(\cdot)$ changes when onsidering a small perturbation of its argument x in the direction $w \in \mathbb{R}^n$.

The so-called *Gateaux derivative* is a direct generalization of the Frechet derivative for functionals $F : T \mapsto \mathbb{R}$

$$\frac{\partial F}{\partial g}[h] \triangleq \lim_{\epsilon \to 0} \frac{F[h + \epsilon g] - F[h]}{\epsilon} \tag{3.17}$$

which represents how much the functional $F[\cdot]$ changes when it is considering a small variation of its argument h in the direction $g \in T$. Finally, the functional derivative (aka *Volterra derivative*) is a special case of Gateaux derivative, which considers the special direction provided by the Dirac delta function concentrated at the point x

$$\frac{\partial F}{\partial \{x\}}[h] \triangleq \frac{\partial F}{\partial \delta_x}[h] \triangleq \lim_{\epsilon \to 0} \frac{F[h + \epsilon \delta_x] - F[h]}{\epsilon} \tag{3.18}$$

wich represents how the functional $F[\cdot]$ changes when $h(\cdot)$ is perturbated at the point x. More in general, the functional derivative of a functional $F[\cdot]$ with respect to a finite set $\mathsf{x} = \{x_1, \ldots, x_\eta\}$ is defined as iterated Gateaux derivative in the directions $\delta_{x_1}, \ldots, \delta_{x_\eta}$

$$\frac{\partial F}{\partial \mathsf{x}}[h] \triangleq \begin{cases} F[h] & \text{if } \mathsf{x} = \emptyset \\ \frac{\partial^\eta F}{\partial \delta_{x_\eta} \cdots \partial \delta_{x_1}}[h] & \text{if } \mathsf{x} = \{x_1, \ldots, x_\eta\} \end{cases} \tag{3.19}$$

- **Example 1**: consider the generic linear functional $F[h] \triangleq f[h] = \int h(w)\, f(w)\; \mathrm{d}w$. The first order functional derivative is

$$\begin{aligned}
\frac{\partial f[h]}{\partial\{x_1\}} &= \frac{\partial}{\partial \delta_{x_1}}\left[\frac{\partial f[h]}{\partial \varnothing}\right] = \frac{\partial}{\partial \delta_{x_1}}[f[h]] = \lim_{\epsilon\to 0}\frac{f[h+\epsilon\delta_{x_1}] - f[h]}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\int [h(w)+\epsilon\delta_{x_1}(w)]f(w)\; \mathrm{d}w - \int h(w)\, f(w)\; \mathrm{d}w}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\left(\int h(w)\, f(w)\; \mathrm{d}w + \epsilon \int \delta_{x_1}(w)\, f(w)\; \mathrm{d}w\right) - \int h(w)\, f(w)\; \mathrm{d}w}{\epsilon} \\
&= \lim_{\epsilon\to 0}\int \delta_{x_1}(w)\, f(w)\; \mathrm{d}w = \int \delta_{x_1}(w)\, f(w)\; \mathrm{d}w = f(x_1)
\end{aligned} \tag{3.20}$$

Since the result does not depend on the generic test functions $h(\cdot)$, the higher order functional derivative are all identically zero. In conclusion, the functional derivative of a linear functional is

$$\frac{\partial f[h]}{\partial \mathsf{x}} = \begin{cases} f[h] & \text{if } \mathsf{x} = \varnothing \\ f(x_1) & \text{if } \mathsf{x} = \{x_1\} \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

In simple words, the functional derivative has the effect of extracting the density function $f(\cdot)$ from the functional $f[\cdot]$.

3.5.2 Rigourous definition

The definition (3.18) is only a practical engineering heuristic. The rigorous definition of functional derivative is based on the definition of set derivative. First, consider the set function

$$\Phi(S) \triangleq \frac{\partial F}{\partial g}[h]\bigg|_{g=1_S}. \tag{3.22}$$

Such set function admits a density $\phi(\cdot)$, so it is possible to express $\Phi(\cdot)$ as follows:

$$\Phi(S) = \int_S \phi(x)\; \mathrm{d}x. \tag{3.23}$$

The density $\phi(\cdot)$, which is given by Lebesgue-differentation of $\Phi(\cdot)$, is the rigorous definition of functional derivative of $F[\cdot]$, i.e.

$$\frac{\partial F}{\partial \{x\}}[h] \triangleq \phi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\Phi(S)\right] = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial F}{\partial g}[h]\bigg|_{g=1_S}\right]. \tag{3.24}$$

In a more fancy way, it is possible to define the functional derivative in terms of set-differentatation of $\Phi(\cdot)$

$$\frac{\partial F}{\partial \{x\}}[h] \triangleq \frac{\mathrm{d}}{\mathrm{d}\{x\}} \left[\left. \frac{\partial F}{\partial g}[h] \right|_{g=1_S} \right]_{S=\varnothing} \tag{3.25}$$

This equation is the so-called *constructive definition* of functional derivative, since it provides an explicit method to compute the functional derivative of $F[\cdot]$. On the other hand, Equation (47) can be expressed with the more meaningful notation as follows

$$\left. \frac{\partial F}{\partial g}[h] \right|_{g=1_S} = \int_S \frac{\partial F}{\partial \{x\}}[h] \, \mathrm{d}x \tag{3.26}$$

this is the so-called *non-constructive definition* of functional derivative (it is only an implicit definition since the functional derivative is under the sign of integral). The non-constructive definition can be naturally extended to the generic Gateaux derivative as follows

$$\frac{\partial F}{\partial g}[h] = \int g(x) \, \frac{\partial F}{\partial \{x\}}[h] \, \mathrm{d}x \tag{3.27}$$

The idea about behind this formula is based on the fact that

$$\left. \frac{\partial F}{\partial g}[h] \right|_{g=1_S} = \int 1_S(x) \, \frac{\partial F}{\partial \{x\}}[h] \, \mathrm{d}x \tag{3.28}$$

so that replacing $1_S(\cdot)$ with the generic test function $g(\cdot)$ provides the general Gateaux derivative of the functional $F[\cdot]$. Equation (3.28) express the fact that it is possible to recover any Gateaux derivative by knowing the simpler singleton functional derivative.

3.6 Properties of the functional derivatives

3.6.1 Set derivatives as special functional derivatives

$$\left. \frac{\partial F[h]}{\partial \mathsf{x}} \right|_{h=1_S} = \frac{\mathrm{d}\phi_F(S)}{\mathrm{d}\mathsf{x}} \tag{3.29}$$

PROOF (SKETCH)

Let E_x be a neighborhood of x with hypervolume ϵ. If $\epsilon \to 0$ then

$$\delta_x(w) = \frac{1_{E_x}(w)}{\epsilon} \tag{3.30}$$

consequently $\epsilon\delta_x = 1_{E_x}$ and

$$\begin{aligned}\left.\frac{\partial F[h]}{\partial\{x\}}\right|_{h=1_S} &= \lim_{\epsilon\to 0}\left.\frac{F[h+\epsilon\delta_x]-F[h]}{\epsilon}\right|_{h=1_S} = \lim_{\epsilon\to 0}\left.\frac{F[h+1_{E_x}]-F[h]}{\epsilon}\right|_{h=1_S}. \\ &= \lim_{\epsilon\to 0}\frac{F[1_S+1_{E_x}]-F[1_S]}{\epsilon}\end{aligned} \tag{3.31}$$

Let $\phi_F(S) \triangleq F[1_S]$ the set function induced by the functional $F[\cdot]$ and assume for simplicity that S and E_x are disjoint, so that $1_S + 1_{E_x} = 1_{S\cup E_x}$ and

$$\left.\frac{\partial F[h]}{\partial\{x\}}\right|_{h=1_S} = \lim_{\epsilon\to 0}\frac{F[1_{S\cup E_x}]-F[1_S]}{\epsilon} = \lim_{\epsilon\to 0}\frac{\phi_F(S\cup E_x)-\phi_F(S)}{\epsilon} = \frac{\mathrm{d}\phi_F(S)}{\mathrm{d}\{x\}} \tag{3.32}$$

By iterating the procedure over the entire generic finite set $\mathsf{x} = \{x_1, \ldots, x_\eta\}$, equation (3.29) is obtained. □

In particular, the functional derivative of the PGFL restricted to the indicator function $h = 1_S$ is the set derivative of the BMF, i.e.,

$$\left.\frac{\partial G[h]}{\partial\mathsf{x}}\right|_{h=1_S} = \frac{\mathrm{d}\beta(S)}{\mathrm{d}\mathsf{x}}. \tag{3.33}$$

From the above fact, some interesting results arise:

- setting $S = \varnothing$ provides the MPDF, indeed

$$\left.\frac{\partial G[h]}{\partial\mathsf{x}}\right|_{h=1_\varnothing} = \left.\frac{\mathrm{d}\beta(S)}{\mathrm{d}\mathsf{x}}\right|_{S=\varnothing} = p_{\mathsf{X}}(\mathsf{x}). \tag{3.34}$$

  Since $1_\varnothing = 0$, the operation that extracts the MPDF from the PGFL is the functional derivative restricted to the null test function $h = 0$

$$\left.\frac{\partial G[h]}{\partial\mathsf{x}}\right|_{h=0} = p_{\mathsf{X}}(\mathsf{x}) \tag{3.35}$$

- setting $S = \mathbb{R}^n$ and by considering the singleton $\mathsf{x} = \{x\}$, provides the PHD density, in fact

$$\left.\frac{\partial G[h]}{\partial\{x\}}\right|_{h=1_{\mathbb{R}^n}} = \left.\frac{\mathrm{d}\beta(S)}{\mathrm{d}\{x\}}\right|_{S=\mathbb{R}^n} = D_{\mathsf{X}}(x). \tag{3.36}$$

  Since $1_{\mathbb{R}^n} = 1$, the operation that extracts the PHD density from the PGFL is the functional derivative restricted to unitary test function $h = 1$

$$\left.\frac{\partial G[h]}{\partial\{x\}}\right|_{h=1} = D_{\mathsf{X}}(x) \tag{3.37}$$

3.6.2 Fundamental theorem of multiobject calculus

$$
\begin{aligned}
p_{\mathsf{X}}(\mathsf{x}) &= \frac{\partial}{\partial \mathsf{x}} \left[\int h^{\mathsf{w}}\, p_{\mathsf{X}}(\mathsf{w})\, \mathrm{d}\mathsf{w} \right]_{h=0} && \text{(a)} \\
F[h] &= \int h^{\mathsf{w}} \left.\frac{\partial F[h]}{\partial \mathsf{w}}\right|_{h=0} \mathrm{d}\mathsf{w} && \text{(b)}
\end{aligned}
\tag{3.38}
$$

PROOF (SKETCH)

- **(a)**: equation (63a) is already proved. In fact, by observing that the set integral in the RHS is the PGFL $G[\cdot]$ of the MPDF $p_{\mathsf{X}}(\{\cdot\})$, follows trivially that equation (1.38a) is equation (1.34).

- **(b)**: suppose for simplicity that the considered functional $F[\cdot]$ is the PGFL $G[\cdot]$ of the MPDF $p_{\mathsf{X}}(\{\cdot\})$. In this case, equation (1.38a) tells that

$$
\left.\frac{\partial F[h]}{\partial \mathsf{w}}\right|_{h=0} = p_{\mathsf{X}}(\mathsf{w}). \tag{3.39}
$$

  Hence the RHS of (1.38b) is, as claimed, the PGLF of $p_{\mathsf{X}}(\{\cdot\})$ itself

$$
\int h^{\mathsf{w}}\, p_{\mathsf{X}}(\mathsf{w})\, \mathrm{d}\mathsf{w} = F[h]. \tag{3.40}
$$

  The result still holds in the general case where $F[\cdot]$ is as a generic functional.

□

The fundamental theorem proves that the PGFL of RFS is a descriptor equivalent to the BMF or MPDF, in fact:

- given a MPDF, the definition of PGFL provides the rule to compute the relative PGFL;

- given the PGFL, (1.38a) provides the rule to get back the MPDF.

3.6.3 Turn the crank rules for functional derivatives

- **(constant rule)** - Let $F[h] \triangleq K$ be a constant functional, i.e. a functional that does not depends on any test function $h(\cdot)$. Then

$$
\frac{\partial K}{\partial \{x\}} = 0 \tag{3.41}
$$

- **(linear rule)** - Let $F[h] \triangleq f[h] \triangleq \int h(x)\, f(x)\, \mathrm{d}x$ be a linear functional with density $f(\cdot)$. Then

$$\frac{\partial f[h]}{\partial \{x\}} = f(x) \tag{3.42}$$

- **(monomial rule)** - Let $F[\cdot]$ be a generic functional and let N be an integer. Then

$$\frac{\partial (F[h])^N}{\partial \{x\}} = N\,(F[h])^{N-1}\,\frac{\partial F[h]}{\partial \{x\}} \tag{3.43}$$

- **(sum rule)** - Let $F_1[\cdot]$, $F_2[\cdot]$ be two generic functionals and let a, b be two real numbers. Then

$$\frac{\partial}{\partial \{x\}}(a\, F_1[h] + b\, F_2[h]) = a\,\frac{\partial F_1[h]}{\partial \{x\}} + b\,\frac{\partial F_2[h]}{\partial \{x\}} \tag{3.44}$$

- **(product rule)** - Let $F_1[\cdot]$, $F_2[\cdot]$ be two generic functionals. Then

$$\frac{\partial}{\partial \{x\}}(F_1[h]\, F_2[h]) = \frac{\partial F_1[h]}{\partial \{x\}}\, F_2[h] + F_1[h]\,\frac{\partial F_2[h]}{\partial \{x\}} \tag{3.45}$$

- **(first chain rule)** - Let $F[\cdot]$ be a generic functional and let $\varphi(\cdot)$ be a generic scalar function. Then

$$\frac{\partial\, \varphi(F[h])}{\partial \{x\}} = \left.\frac{\mathrm{d}\varphi(y)}{\mathrm{d}y}\right|_{y=F[h]} \frac{\partial F[h]}{\partial \{x\}} \tag{3.46}$$

3.7 Second chain rule

3.7.1 Functional transformations

Let $\varphi[h](\cdot)$ be a *functional transformation*, i.e. a function that transforms a test function $h(\cdot)$ into another test function $\varphi[h](\cdot)$.

- **Example 2**: The functional transformations are defined pointwise on the domain of the starting test function $h(\cdot)$. An example of functional transformation is

$$\varphi[h](w) = 1 - p(w) + p(w)\, h(w) \tag{3.47}$$

where, necessarily, $p(\cdot)$ is a second test function (if this is not the case then $\varphi[h](\cdot)$ will not be a test function). Note that by fixing in

w_0 the argument w and by varing the test function $h(\cdot)$, the functional transformation $\varphi[\cdot](w_0)$ behaves exactly like an ordinary functional: $\varphi[\cdot](w_0)$ associates the actual test function $h(\cdot)$ to a scalar $\varphi[h](w_0) = 1 - p(w_0) + p(w_0)\,h(w_0)$. From this prospective, it make sense to compute the functional derivative of the functional $\varphi[\cdot](w_0)$

$$\frac{\partial\,\varphi[h](w_0)}{\partial\,\{x\}} \triangleq \lim_{\epsilon\to 0} \frac{\varphi[h+\epsilon\delta_x](w_0) - \varphi[h](w_0)}{\epsilon} \tag{3.48}$$

whatever it is the fixed point w_0.

In general a functional transformation can be seen as a continuos family of functionals parametrized by the argument w of the base test function $h(\cdot)$. Due to this interpretation, the definition of functional derivative for a functional transformation $\varphi[\cdot](\cdot)$ is

$$\frac{\partial\,\varphi[h](w)}{\partial\,\{x\}} \triangleq \lim_{\epsilon\to 0} \frac{\varphi[h+\epsilon\delta_x](w) - \varphi[h](w)}{\epsilon} \qquad \forall w \in \mathbb{R}^n \tag{3.49}$$

3.7.2 Theorem

Consider a functional $F[\cdot]$. Since $\varphi[h](\cdot)$ is a test function for any fixed base test function $h(\cdot)$, it is well defined the composition $F[\varphi[h]]$, wich is a functional as well. Thus, it make sense to compute the functional derivative of the composition $F[\varphi[h]]$.

Theorem 1. (second chain rule) - Let be $F[\cdot]$ be a generic functional and let $\varphi[h](\cdot)$ be a generic functional transformation. Then

$$\frac{\partial\,F[\varphi[h]]}{\partial\,\{x\}} = \int \frac{\partial\,\varphi[h](w)}{\partial\,\{x\}} \frac{\partial\,F[\tilde{h}]}{\partial\,\{w\}}\bigg|_{\tilde{h}=\varphi[h]} \mathrm{d}w \tag{3.50}$$

Proof

According to the practical definition of functional derivative, holds

$$\frac{\partial\,F[\varphi[h]]}{\partial\,\{x\}} = \lim_{\epsilon\to 0} \frac{F[\varphi[h+\epsilon\delta_x]] - F[\varphi[h]]}{\epsilon} \tag{3.51}$$

now, by considering the linear approximation

$$\varphi[h+\epsilon\delta_x] \approx \varphi[h] + \frac{\partial\,\varphi[h]}{\partial\,\{x\}}\,\epsilon \tag{3.52}$$

which can be written as

$$\begin{aligned}\frac{\partial F[\varphi[h]]}{\partial \{x\}} &= \lim_{\epsilon\to 0}\frac{F[\varphi[h]+\frac{\partial \varphi[h]}{\partial \{x\}}\,\epsilon]-F[\varphi[h]]}{\epsilon}\\ &= \lim_{\epsilon\to 0}\frac{F[\tilde{h}+\frac{\partial \varphi[h]}{\partial \{x\}}\,\epsilon]-F[\tilde{h}]}{\epsilon}\bigg|_{\tilde{h}=\varphi[h]}\end{aligned} \tag{3.53}$$

the third member is the Gateaux derivative of $F[\cdot]$ evaluated on the test function $\tilde{h}=\varphi[h]$ and in the direction $g=\frac{\partial \varphi[h]}{\partial \{x\}}$, so

$$\frac{\partial F[\varphi[h]]}{\partial \{x\}} = \frac{\partial F[\tilde{h}]}{\partial g}\bigg|_{g=\frac{\partial \varphi[h]}{\partial \{x\}},\tilde{h}=\varphi[h]} \tag{3.54}$$

thus, the conclusion is

$$\frac{\partial F[\varphi[h]]}{\partial \{x\}} = \int \frac{\partial \varphi[h](w)}{\partial \{x\}}\frac{\partial F[\tilde{h}]}{\partial \{w\}}\bigg|_{\tilde{h}=\varphi[h]}\,\mathrm{d}w \tag{3.55}$$

as claimed □

3.7.3 Observation

Note that the two notation

$$\frac{\partial F[\varphi[h]]}{\partial \{x\}} \qquad\qquad \frac{\partial F[\tilde{h}]}{\partial \{x\}}\bigg|_{\tilde{h}=\varphi[h]} \tag{3.56}$$

figuring in the previous proof are confusingly similar, but, as one naturally expects, in general doesn't represent the same quantity. In fact, the former is given by

$$\frac{\partial F[\varphi[h]]}{\partial \{x\}} = \lim_{\epsilon\to 0}\frac{F[\varphi[h+\epsilon\delta_x]]-F[\varphi[h]]}{\epsilon} \tag{3.57}$$

while the latter is given by the definition functional derivative restricted to $\tilde{h}=\varphi[h]$, i.e.

$$\frac{\partial F[\tilde{h}]}{\partial \{x\}}\bigg|_{\tilde{h}=\varphi[h]} = \lim_{\epsilon\to 0}\frac{F[\tilde{h}+\epsilon\delta_x]-F[\tilde{h}]}{\epsilon}\bigg|_{\tilde{h}=\varphi[h]} = \lim_{\epsilon\to 0}\frac{F[\varphi[h]+\epsilon\delta_x]-F[\varphi[h]]}{\epsilon}. \tag{3.58}$$

Now, by comparing the difference quotients, turns out clearly that

$$\begin{aligned}\frac{\partial F[\varphi[h]]}{\partial \{x\}} = \lim_{\epsilon\to 0}\frac{F[\varphi[h+\epsilon\delta_x]]-F[\varphi[h]]}{\epsilon} \neq\\ \lim_{\epsilon\to 0}\frac{F[\varphi[h]+\epsilon\delta_x]-F[\varphi[h]]}{\epsilon} = \frac{\partial F[\tilde{h}]}{\partial \{x\}}\bigg|_{\tilde{h}=\varphi[h]}\end{aligned} \tag{3.59}$$

rather, the correct equivalence is expressed by

$$\frac{\partial\, F[\varphi[h]]}{\partial\,\{x\}} = \lim_{\epsilon\to 0}\frac{F[\varphi[h+\epsilon\delta_x]] - F[\varphi[h]]}{\epsilon} \equiv \lim_{\epsilon\to 0}\frac{F\left[\varphi[h] + \epsilon\frac{\partial\,\varphi[h]}{\partial\,\{x\}}\right] - F[\varphi[h]]}{\epsilon} = \left.\frac{\partial F[\tilde{h}]}{\partial g}\right|_{g=\frac{\partial\,\varphi[h]}{\partial\,\{x\}},\tilde{h}=\varphi[h]} \tag{3.60}$$

so, in conclusion, in (1.50) the former functional derivative is performed in the direction $\frac{\partial\,\varphi[h]}{\partial\{x\}}$, while the latter derivative is performed in the direction δ_x, i.e.

$$\frac{\partial\, F[\varphi[h]]}{\partial\,\{x\}} = \left.\frac{\partial F[\tilde{h}]}{\partial g}\right|_{g=\frac{\partial\,\varphi[h]}{\partial\{x\}},\tilde{h}=\varphi[h]} \neq \left.\frac{\partial\, F[\tilde{h}]}{\partial\,\{x\}}\right|_{\tilde{h}=\varphi[h]} = \left.\frac{\partial F[\tilde{h}]}{\partial g}\right|_{g=\delta_x,\tilde{h}=\varphi[h]} \tag{3.61}$$

Chapter 4

Standard PHD filter

4.1 Summary

This chapter provides a detailed proof of the PHD filter equations, which express how to compute the predicted and corrected PHDs $D_{k|k-1}(\cdot)$, $D_{k|k}(\cdot)$.

In this chapter such filter is referred as *standard* in order to make a distinction with the second version designed to handle extended object. The difference between this two PHD filters is in the measurement model, where the standard PHD filter assumes that an object can generate no more than one measure per sampling step, while the PHD filter for extended object assumes that an object can generate an arbitrary number of measures per sampling step.

The chapter is organized as follows:

- firstly the equation of the multiobject Bayes filter, presented in the introduction of the thesis, are re-written in the language of PGFLs;

- then is defined the motion and measurement model of the standard PHD filter, in short called *standard model*, and the consequent multiobject Bayes filter, called *standard multiobject Bayes filter*, is derived in its PGFL form;

- in the final part the PHD filter equations are computed starting from the equations of the standard multiobject Bayes filter. In conclusion, the equations of the standard PHD filter are simplified by considering linear-Gaussian models, yielding to the so-called *Gaussian mixture PHD filter* (GM-PHD filter), which is an algorithm that can be easily implemented to the calculator.

4.2 General multi-object Bayes filter

4.2.1 Preliminary discussion

In the classical Kalman filtering theory, many real aspects involved in the detection process are not taken into account. For example, consider the situation where a ground-to-air radar is used to surveil a certain region of space called scene.

The output of the radar is a so-called *signature*, which is a continous signal $s : [0, 2\pi] \mapsto \mathbb{R}$ that maps an azimuthal angle α into a radio-frequency intensity $s(\alpha)$. If an object is located at angle α then the signature gets a value $s(\alpha)$ significantly stronger than the noise floor of the radar. Due to this fact, the azimuthal position of an object present in the scene is extracted from the signature by comparing the signature with a suitable treshold τ: the object is declared present in α if $s(\alpha) > \tau$, and in such case one says that the object is *detected* at azimuth α. The actual detection process is characterized by the following facts:

- due to measurement noise, if the treshold τ is set too low then it is very likely to get detections even when there is no object in the scene. Such detections are called *false detections*. The set of false detections is called *clutter*;

- if the treshold τ is set too high then it is very unlikely to detect objects present in the scene. In this case one says that the objects present in the scene that don't produce any detection are *miss detected*;

- if an object is too close to the radar then it is possible that it can generate more than one detection. In such case the object is called *extended object*;

- if a group of objects is too far from the radar then it is possible that the entire group generate only one global detection. In such case, the objects in the group are called *unresolved*;

- if an object is not too near and not too far from the radar then it is very likely that it generates only one detection. In such case the object is called *point object*;

- an object actually non present in the scene can enter the scene in the future. This event is called *object birth*;

- an object actually present in the scene can leave the scene in the future. This event is called *object death*.

These are the major aspects that are not considered in the classical single-object oriented Kalman filtering theory. It is worth to point out that the actual example is only one istance of many real applications where the classic Kalman filtering theory is not sufficient to address the state estimation problem.

On the other hand, the multiobject Bayes filter can handle all these issues by considering measurement and motion models based on the RFS representation. In particular, the standard multiobject Bayes filter considers:

- a standard measurement model which takes into account the clutter and miss-detections for point objects;
- a standard motion model which takes into account object birth and death.

Extended and unresolved objects are treated by the non standard multiobject Bayes filter.

4.2.2 General equations

The multiobject Bayes filter is the direct generalization of the single object Bayes filter, where the basic concepts of PDF, integration, likelihood, transition density are replaced by the FISST concepts of multiobject filtered and predicted densities, set integral, multiobject likelihood and transition density.

The two main steps of the multiobject Bayes filter are the following:

- **Correction step**: the Bayes equation for the multiobject Bayes filter gets the form

$$p_{k|k}(\mathsf{x}) = \frac{\ell_k(\mathsf{y}|\mathsf{x})\, p_{k|k-1}(\mathsf{x})}{\int \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\ \mathsf{dw}} \tag{4.1}$$

  where $p_{k|k}(\{\cdot\})$ is the filtered MPDF, $\ell_k(\{\cdot\}|\mathsf{x})$ is the multiobject likelihood and $p_{k|k-1}(\{\cdot\})$ is the predicted MPDF. Note that the evidence is computed as a set integral. The Bayes equation (104) defines the correction step performed by the multiobject Bayes filter.

- **Prediction step**: the Chapman-Kolmogorov equation for the multiobject Bayes filter gets the form

$$p_{k+1|k}(\mathsf{x}) = \int \varphi_{k+1|k}(\mathsf{x}|\mathsf{w})\, p_{k|k}(\mathsf{w})\ \mathsf{dw} \tag{4.2}$$

where $\varphi_{k+1|k}(\{\cdot\}|\mathsf{w})$ is the multiobject transition density. Once again, a set integration is involved in place of an ordinary integration. The Chapman-Kolmogorov equation (105) defines the prediction step of the multiobject Bayes filter.

4.2.3 PGFL form

The multiobject Bayes filter is expressed in terms of the predicted MPDF $p_{k|k-1}(\{\cdot\})$ as a function of the corrected MPDF $p_{k-1|k-1}(\{\cdot\})$ through the multiobject Chapman-Kolmogorov equation

$$p_{k|k-1}(\mathsf{x}) = \int \varphi_{k|k-1}(\mathsf{x}|\mathsf{w})\, p_{k-1|k-1}(\mathsf{w})\; \mathrm{d}\mathsf{w} \tag{4.3}$$

and in terms of the corrected MPDF $p_{k|k}(\{\cdot\})$ as function of the predicted MPDF $p_{k|k-1}(\{\cdot\})$ through the multiobject Bayes equation

$$p_{k|k}(\mathsf{x}) = \frac{\ell_k(\mathsf{y}|\mathsf{x})\, p_{k|k-1}(\mathsf{x})}{\int \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\; \mathrm{d}\mathsf{w}} \tag{4.4}$$

Since a PGFL, likewise an MPDF, provides a full characterization of an RFS, the multiobject Bayes filter can be expressed in terms of PGFLs as well. Consider the predicted and corrected PGFLs

$$G_{k|k-1}[h] \triangleq \int h^{\mathsf{x}}\, p_{k|k-1}(\mathsf{x})\; \mathrm{d}\mathsf{x} \qquad G_{k|k}[h] \triangleq \int h^{\mathsf{x}}\, p_{k|k}(\mathsf{x})\; \mathrm{d}\mathsf{x}. \tag{4.5}$$

These PGFLs are respectively given by a *transformed* Chapman-Kolmogorov and a *transformed* Bayes equations, which are derived as follows.

- **Chapman-Kolmogorov equation**: starting with the simpler predicted PGFL, it holds that

$$\begin{aligned} G_{k|k-1}[h] &= \int h^{\mathsf{x}} \left(\int \varphi_{k+1|k}(\mathsf{x}|\mathsf{w})\, p_{k-1|k-1}(\mathsf{w})\; \mathrm{d}\mathsf{w} \right) \mathrm{d}\mathsf{x} \\ &= \int \left(\int h^{\mathsf{x}}\, \varphi_{k|k-1}(\mathsf{x}|\mathsf{w})\; \mathrm{d}\mathsf{x} \right) p_{k-1|k-1}(\mathsf{w})\; \mathrm{d}\mathsf{w} \end{aligned}. \tag{4.6}$$

  By defining $\Phi_{k|k-1}[\cdot|\mathsf{w}]$ as the PGFL of the Markov transition MPDF $\varphi_{k|k-1}(\cdot|\mathsf{w})$

$$\Phi_{k|k-1}[h|\mathsf{w}] \triangleq \int h^{\mathsf{x}}\, \varphi_{k|k-1}(\mathsf{x}|\mathsf{w})\; \mathrm{d}\mathsf{x} \tag{4.7}$$

  from which it follows that the PGFL form of the Chapman-Kolmogorov equation is

$$G_{k|k-1}[h] = \int \Phi_{k|k-1}[h|\mathsf{w}]\, p_{k-1|k-1}(\mathsf{w})\; \mathrm{d}\mathsf{w}. \tag{4.8}$$

- **Bayes equation**: now consider the corrected PGFL, it holds that

$$\begin{aligned} G_{k|k}[h] &= \int h^{\mathsf{x}} \left(\frac{\ell_k(\mathsf{y}|\mathsf{x})\, p_{k|k-1}(\mathsf{x})}{\int \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}} \right) \mathrm{d}\mathsf{x} \\ &= \frac{\int h^{\mathsf{x}}\, \ell_k(\mathsf{y}|\mathsf{x})\, p_{k|k-1}(\mathsf{x})\ \mathrm{d}\mathsf{x}}{\int \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}} \\ &= \frac{\int h^{\mathsf{x}}\, \ell_k(\mathsf{y}|\mathsf{x})\, p_{k|k-1}(\mathsf{x})\ \mathrm{d}\mathsf{x}}{\left[\int h^{\mathsf{w}}\, \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}\right]_{h=1}} \end{aligned} . \qquad (4.9)$$

By defining $L_k[\cdot|\mathsf{w}]$ as the PGFL of the likelihood MPDF $\ell_{k|k-1}(\cdot|\mathsf{w})$

$$L_k[g|\mathsf{w}] \triangleq \int g^{\mathsf{y}}\, \ell_k(\mathsf{y}|\mathsf{w})\ \mathrm{d}\mathsf{y} \qquad (4.10)$$

it follows that

$$\left.\frac{\partial\, L_k[g|\mathsf{w}]}{\partial\, \mathsf{y}}\right|_{g=0} = \ell_k(\mathsf{y}|\mathsf{w}) \qquad (4.11)$$

from which it follows that the corrected PGFL can be written as

$$\begin{aligned} G_{k|k}[h] &= \frac{\int h^{\mathsf{x}} \left.\frac{\partial\, L_k[g|\mathsf{x}]}{\partial\, \mathsf{y}}\right|_{g=0} p_{k|k-1}(\mathsf{x})\ \mathrm{d}\mathsf{x}}{\left[\int h^{\mathsf{w}} \left.\frac{\partial\, L_k[g|\mathsf{w}]}{\partial\, \mathsf{y}}\right|_{g=0} p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}\right]_{h=1}} \\ &= \frac{\frac{\partial}{\partial_g\, \mathsf{y}} \left[\int h^{\mathsf{x}}\, L_k[g|\mathsf{x}]\, p_{k|k-1}(\mathsf{x})\ \mathrm{d}\mathsf{x}\right]_{g=0}}{\frac{\partial}{\partial_g\, \mathsf{y}} \left[\int h^{\mathsf{w}}\, L_k[g|\mathsf{w}]\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w}\right]_{h=1,g=0}} \end{aligned} \qquad (4.12)$$

where the symbol $\partial/\partial_g\, \mathsf{y}$ means that the differentation is performed (only) with respect to the test function $g(\cdot)$ (while $h(\cdot)$ is mantained fixed). In other words, $\partial/\partial_g\, \mathsf{y}$ is a partial functional derivative evaluated in y. By introducing the bivariate PGFL of the bivariate multi-object density $\ell_k(\cdot|\cdot)\, p_{k|k-1}(\cdot)$

$$\begin{aligned} F_k[h,g] &\triangleq \int h^{\mathsf{w}}\, L_k[g|\mathsf{w}]\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w} \\ &= \int h^{\mathsf{w}} \left(\int g^{\mathsf{y}}\, \ell_k(\mathsf{y}|\mathsf{w})\ \mathrm{d}\mathsf{y} \right) p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{w} \\ &= \int h^{\mathsf{w}}\, g^{\mathsf{y}}\, \ell_k(\mathsf{y}|\mathsf{w})\, p_{k|k-1}(\mathsf{w})\ \mathrm{d}\mathsf{y}\mathrm{d}\mathsf{w} \end{aligned} \qquad (4.13)$$

from which it turns out finally that the PGFL form of the Bayes equation (121) is

$$G_{k|k}[h] = \frac{\left.\frac{\partial\, F_k[h,g]}{\partial_g\, \mathsf{y}}\right|_{g=0}}{\left.\frac{\partial\, F_k[h,g]}{\partial_g\, \mathsf{y}}\right|_{h=1,g=0}} . \qquad (4.14)$$

4.3 Standard multi-object Bayes filter

4.3.1 Standard measurement model

The standard measurement model makes the following assumptions:

1. an object detection is produced according to the observation model $Y = h(X, V)$, where V is the measurement noise. Note that this assumption holds in two different situations:
   - the detection process is based on a single sensor, characterized by the model $h(\cdot, \cdot)$, that can produce simultaneously several measurements;
   - the detection process is based on several sensors, characterized by a common model $h(\cdot, \cdot)$, that can produce up to one measurement;
2. the clutter set is Poisson with intensity $I : \mathbb{R}^p \mapsto \mathbb{R}_{\geq 0}$, where $\mathbb{R}^p$ is the measurement space;
3. an object is detected with probability $p_\mathrm{D} : \mathbb{R}^n \mapsto [0, 1]$, where $\mathbb{R}^n$ is the state space;
4. an object can generate up to one measurement, i.e. all objects are considered pointwise. In other words, an object can be miss-detected or can generate one measurement;
5. a measurement concerns at most one object, i.e. all objects are resolved. In other words, a measurement can be a false detection or can be a detection of an object;
6. clutter and the set of object measurements are statistically independent.

Given the above assumptions, the standard measurement model states that the set of measurements $\mathsf{Y} \triangleq \{Y_1, \ldots, Y_{|\mathsf{Y}|}\}$, referred to as multiobject measurement (in short *multimeasurement*), consists of by two parts:

- the first part is the clutter C, which is simply the collection of the false detections produced by the detection process. In other words, the clutter C is the subset of the multimeasurement Y formed by the measurements wich are not produced by objects

$$\mathsf{C} = \{Y_i \in \mathsf{Y} : Y_i \text{ is a false detection}\} \tag{4.15}$$

- the second part is the set of detections $\mathsf{h}(\mathsf{X})$, referred as *multidetection*, which is simply the collection of the true detections produced by the detection process. In other words, the multidetection $\mathsf{h}(\mathsf{X})$ is the subset of the multimeasurement Y formed by the measurements wich are produced by objects

$$\mathsf{h}(\mathsf{X}) = \{Y_i \in \mathsf{Y} : Y_i \text{ is a detection}\} \tag{4.16}$$

.

At this high level of the discussion the standard measurement model is expressed by the following simple set equation

$$\mathsf{Y} = \mathsf{h}(\mathsf{X}) \cup \mathsf{C} \tag{4.17}$$

which, intrestingly, resembles the ordinary measurement model $Y = h(X) + V$ considered in the ordinary single object Bayes filter.

The clutter C is an RFS that does not require any further explanation, since assumption 2 states explicitly that it is Poisson. On the other hand the multidetection $\mathsf{h}(\mathsf{X})$ is an RFS that needs further explanation. Let $\mathsf{X} \triangleq \{X_1, \ldots, X_{|\mathsf{X}|}\}$ be the mulitobject state (in short *multistate*). Due to assumptions 4 and 5, the generic object with state $X_i \in \mathsf{X}$ generates one measure, which is $Y_i = h(X_i, V_i)$ according to assumption 1, if such object is detected, while doesn't generate any measurement if such object is not detected. As a consequence of this fact, the generic object with state $X_i \in \mathsf{X}$ is associated with an RFS $\mathsf{h}(X_i)$, referred to as *detection*, such that

$$\mathsf{h}(X_i) \triangleq \begin{cases} \{h(X_i, V_i)\} & \text{if object with state } X_i \text{ is detected} \\ \varnothing & \text{if object with state } X_i \text{ is not detected} \end{cases}. \tag{4.18}$$

Since the detection $\mathsf{h}(X_i)$ is an RFS with cardinality concetrated over {0,1}, it is Bernoulli. The actual definition considers a generic object, so it holds for every object in the scene and leads to the following natural definition for the multidetection $\mathsf{h}(\mathsf{X})$

$$\mathsf{h}(\mathsf{X}) \triangleq \bigcup_{X_i \in \mathsf{X}} \mathsf{h}(X_i). \tag{4.19}$$

Thus in conclusion, due to assumption 6, the multidetection $\mathsf{h}(X_i)$ is multi-Bernoulli and consequently the multimeasurement Y is Poisson-multi-Bernoulli.

4.3.2 Standard motion model

The standard motion model makes the following assumptions:

1. an object with actual state X can evolve to the future state X' according to the motion model $X' = f(X, W)$, where W is the process noise;

2. the birth set is Poisson with intensity $I_{\mathrm{B}} : \mathbb{R}^n \mapsto \mathbb{R}_{\geq 0}$, where $\mathbb{R}^n$ is the measurement space;

3. an object doesn't disappear from the scene, i.e. *survives*, with probability $p_{\mathrm{S}} : \mathbb{R}^n \mapsto [0, 1]$, where $\mathbb{R}^n$ is the state space;

4. object spawning is not allowed. i.e. an object can disappear or can survives and if survives then it generates only one new state;

5. a new state is generated up to one survived object, i.e. multiple actual states cannot fuse in one new state. This means that a future state is generated by the movement of one survived object or it is generated by the appearance of a new object;

6. the birth set and the set of survived object are statistically independent.

Given these assumptions, the standard measurement model states that the future multistate $\mathsf{X}' \triangleq \left\{X'_1, \ldots, X'_{|\mathsf{X}'|}\right\}$ consists of two parts:

- the first part is the set of new born objects B, wich is simply the collection of the objects that are just appeared in the scene. In other words, the birth set B is the subset of the future multistate X' formed by the states wich are not generated by the motion of the survived objects
$$\mathsf{B} = \{X'_i \in \mathsf{X}' : X'_i \text{ is a new born state}\} \tag{4.20}$$

- the second part is the set of survived objects $\mathsf{f}(\mathsf{X})$, which is simply the collection of the states generated by the motion of the survived objects
$$\mathsf{f}(\mathsf{X}) = \{X'_i \in \mathsf{X}' : X'_i \text{ is not a new born state}\} . \tag{4.21}$$

The standard motion model is expressed by the following simple set equation

$$\mathsf{X}' = \mathsf{f}(\mathsf{X}) \cup \mathsf{B} \tag{4.22}$$

wich resembles the ordinary motion model $X' = f(X) + W$ considered in the single object Bayes filter.

The birth set B is already complete defined by assumption 2 (it is Poisson with given intensity function). The set of survived objects $\mathsf{f}(\mathsf{X})$ is defined similarly to $\mathsf{h}(\mathsf{X})$: given $\mathsf{X} = \{X_1, \ldots, X_{|\mathsf{X}|}\}$, the generic object with state i, according to assumptions 4 and 5, can move in a new state $X_i' = f(X_i, W_i)$ or can disappear, so such generic object is associated to the RFS

$$\mathsf{f}(X_i) \triangleq \begin{cases} \{f(X_i, W_i)\} & \text{if object with state } X_i \text{ survives} \\ \varnothing & \text{if object with state } X_i \text{ disappears} \end{cases}. \tag{4.23}$$

In total there are $|\mathsf{X}|$ different instances of the previous RFS (one per object), which can be collected in the global RFS

$$\mathsf{f}(\mathsf{X}) \triangleq \bigcup_{X_i \in \mathsf{X}} \mathsf{f}(\mathsf{X}) \tag{4.24}$$

The birth set B is assumed Poisson, while the RFS $\mathsf{f}(X_i)$ is Bernoulli, so that $\mathsf{f}(\mathsf{X})$ turns out to be multi-Bernoulli and X' Poisson-multi-Bernoulli. The conclusion is that the standard motion model is analogous to the standard measurement model where the following substitutions are considered

$$\mathsf{C} \to \mathsf{B} \qquad \mathsf{h}(\mathsf{X}) \to \mathsf{f}(\mathsf{X}) \qquad \mathsf{Y} \to \mathsf{X}' \tag{4.25}$$

This means that all the previous results found for the standard measurement model still hold for the standard motion model (with the new conventions).

4.3.3 PGFL form

The standard motion and measurement models are respectively

$$\begin{aligned} \mathsf{X}_{k+1} &= \mathsf{f}(\mathsf{X}_k) \cup \mathsf{B}_k \\ \mathsf{Y}_k &= \mathsf{h}(\mathsf{X}_k) \cup \mathsf{C}_k \end{aligned} \tag{4.26}$$

where at any given time step k

- the RFS of survived objects $\mathsf{f}(\mathsf{X}_k)$ is multi-Bernoulli with parameters $\{p_\mathrm{S}(X), \varphi_{k+1|k}(\cdot|X)\}_{X \in \mathsf{X}_k}$;
- the RFS of new born objects B_k is Poisson with intensity $I_\mathrm{B}(\cdot)$;
- the RFS of detected objects $\mathsf{h}(\mathsf{X}_k)$ is multi-Bernoulli with parameters $\{p_\mathrm{D}(X), \ell_k(\cdot|X)\}_{X \in \mathsf{X}_k}$;
- the clutter C_k is Poisson with intensity $I_\mathrm{C}(\cdot)$.

Moreover, the standard model assumes that all involved RFS involved are statistically independent from each other. Due to this assumption, it is easy to derive the PGFLs of the Markov transition MPDF and the likelihood MPDF. Consequently, it is also easy to derive the PGFL form of the Chapman-Kolmogorov and Bayes equations.

- **Markov PGFL**: Let $G_k^{\mathrm{S}}[\cdot|\mathsf{X}_k]$, $G_k^{\mathrm{B}}[\cdot]$ be the PGFLs of $\mathsf{f}(\mathsf{X}_k)$, B_k respectively. It holds that

$$\begin{aligned} G_k^{\mathrm{S}}[h|\mathsf{X}_k] &= \prod_{X \in \mathsf{X}_k} (1 - p_{\mathrm{S}}(X) + p_{\mathrm{S}}(X)\, \varphi_{k+1|k}[h|X]) \\ &= (1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k+1|k}[h])^{\mathsf{X}_k} \end{aligned} \tag{4.27}$$

$$G_k^{\mathrm{B}}[h] = \exp(I_{\mathrm{B}}[h-1]) \tag{4.28}$$

where $\varphi_{k+1|k}[h|X] \triangleq \int h(x)\, \varphi_{k+1|k}(x|X)\, \mathrm{d}x$. Due to the independence, the PGFL $G_{k+1|k}[\cdot|\mathsf{X}_k]$ of the Markov transition MPDF $\varphi_{k+1|k}(\{\cdot\}|\mathsf{X}_k)$ is given by

$$\begin{aligned} \Phi_{k+1|k}[h|\mathsf{X}_k] &= G_k^{\mathrm{S}}[h|\mathsf{X}_k]\, G_k^{\mathrm{B}}[h] \\ &= (1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k+1|k}[h])^{\mathsf{X}_k}\, \exp(I_{\mathrm{B}}[h-1]) \end{aligned} \tag{4.29}$$

- **likelihood PGFL**: Let $G_k^{\mathrm{D}}[\cdot|\mathsf{X}_k]$, $G_k^{\mathrm{C}}[\cdot]$ be the PGFLs of $\mathsf{h}(\mathsf{X}_k)$, C_k respectively. With the same previous reasoning, it holds that

$$\begin{aligned} L_k[h|\mathsf{X}_k] &= G_k^{\mathrm{D}}[h|\mathsf{X}_k]\, G_k^{\mathrm{C}}[h] \\ &= (1 - p_{\mathrm{D}} + p_{\mathrm{D}}\, \ell_k[h])^{\mathsf{X}_k}\, \exp(I_{\mathrm{C}}[h-1]) \end{aligned} \tag{4.30}$$

where $\ell_k[h|X] \triangleq \int h(x)\, \ell_k(x|X)\, \mathrm{d}x$.

- **Chapman-Kolmogorov equation**: the PGFL form of the Chapman-Kolmogorov equation for the standard multiobject Bayes filter is given by

$$\begin{aligned} G_{k|k-1}[h] &= \int (1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k|k-1}[h])^{\mathsf{w}}\, \exp(I_{\mathrm{B}}[h-1])\, p_{k-1|k-1}(\mathsf{w})\, \mathrm{d}\mathsf{w} \\ &= \exp(I_{\mathrm{B}}[h-1]) \int (1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k|k-1}[h])^{\mathsf{w}}\, p_{k-1|k-1}(\mathsf{w})\, \mathrm{d}\mathsf{w} \\ &= \exp(I_{\mathrm{B}}[h-1])\, G_{k-1|k-1}[1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k|k-1}[h]] \end{aligned} \tag{4.31}$$

where, naturally, the corrected PGFL is defined as

$$G_{k-1|k-1}[h] \triangleq \int h^{\mathsf{w}}\, p_{k-1|k-1}(\mathsf{w})\, \mathrm{d}\mathsf{w}. \tag{4.32}$$

Note that the corrected PGFL $G_{k-1|k-1}[\cdot]$ operates over the functional transformation $h \mapsto 1 - p_{\mathrm{S}} + p_{\mathrm{S}}\, \varphi_{k|k-1}[h]$.

- **Bayes equation**: It turns out that

$$
\begin{aligned}
F_k[h,g] &= \int h^{\mathsf{w}} \left[(1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])^{\mathsf{w}} \exp(I_{\mathrm{C}}[g-1])\right] p_{k|k-1}(\mathsf{w})\ \mathsf{dw} \\
&= \exp(I_{\mathrm{C}}[g-1]) \int h^{\mathsf{w}} (1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])^{\mathsf{w}}\, p_{k|k-1}(\mathsf{w})\ \mathsf{dw} \\
&= \exp(I_{\mathrm{C}}[g-1]) \int [h\,(1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])]^{\mathsf{w}}\, p_{k|k-1}(\mathsf{w})\ \mathsf{dw} \\
&= \exp(I_{\mathrm{C}}[g-1])\, G_{k|k-1}[h\,(1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])]
\end{aligned}
\quad . \tag{4.33}
$$

Hence, the PGFL form of the Bayes equation for the standard multi-object Bayes filter is

$$
G_{k|k}[h] = \frac{\frac{\partial}{\partial_g\,\mathsf{y}} \left[\exp(I_{\mathrm{C}}[g-1])\, G_{k|k-1}[h\,(1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])]\right]_{g=0}}{\frac{\partial}{\partial_g\,\mathsf{y}} \left[\exp(I_{\mathrm{C}}[g-1])\, G_{k|k-1}[h\,(1-p_{\mathrm{D}}+p_{\mathrm{D}}\,\ell_k[g])]\right]_{h=1,g=0}} \tag{4.34}
$$

.

4.4 General PHD filter

4.4.1 Idea behind the PHD filter

The central idea of the PHD filter is to propagate the corrected PHD $D_{k|k}(x)$ of the estimand X_k rather than the full corrected MPDF $p_{k|k}(\{\cdot\})$.

Then, the estimate $\hat{\mathsf{x}}_k$ is extracted from $D_{k|k}(\cdot)$ by considering the $\mathbb{E}[|\mathsf{X}_k|] = D_{k|k}[1]$ largest peaks of $D_{k|k}(\cdot)$. Note that the PHD filter is optimal in the sense that it extracts the MAP estimate from $D_{k|k}(\cdot)$ (and not from $p_{k|k}(\cdot)$).

Due to this advanced approximation techinique, the multiobject Bayes filter, wich is combinatorially intractable, reduces to the PHD filter, wich is polynomial in complexity (more precisely, if the actual number of objects and measurements are n and m then the complexity is $O(n\,m)$). The PHD filter represents a rough approximation of the multiobject Bayes filter, resulting in a great amount of loss in information. Despite this fact, in some scenarios the PHD filter performs better than the conventional multihypothesis filters.

4.4.2 PHD filter vs Kalman filter

To understand why and when the PHD filter works correctly, one can think the PHD filter as the multiobject counterpart of the steady state Kalman filter. Consider a single object tracking problem where the corrected PDF

$p_{k|k}(\cdot)$ is unimodal, symmetric and with a time invariant covariance: in this situation the first-order approximation

$$p_{k|k}(x) \approx \mathcal{N}(x|\mu_{k|k}, P) \triangleq p_{k|k}(x|\mu_{k|k}) \tag{4.35}$$

where $\mathcal{N}(\cdot|\mu_{k|k}, P)$ denotes a Gaussian PDF with expected value $\mu_{k|k}$ and covariance P given by the steady state Kalman filter, is accurate. This means that one can propagate the first order moment $\mu_{k|k}$ alone (rather than the full PDF $p_{k|k}(\cdot)$) without loosing the most important information about the estimand x_k; the notation $p_{k|k}(\cdot|\mu_{k|k})$ reminds that $\mu_{k|k}$ contains the relevant information about x_k. In jargon, one says that $\mu_{k|k}$ is a sufficient statistics.

4.4.3 Limitations of the PHD filter

The PHD filter, likewise the steady state Kalman filter, assumes that the multiobject first order moment, the PHD $D_{k|k}(\cdot)$, is a sufficient statistics of the estimand X_k, i.e. the following approximation holds

$$p_{k|k}(\mathsf{x}) \approx p_{k|k}(\mathsf{x}|D_{k|k}). \tag{4.36}$$

Such approximation is reasonable under the following conditions:

1. sensors are unbiased and charactezied by small covariances, meaning that $\ell_k(\cdot|\cdot)$ is concentrated around the true values of the object states. Here *small* means that, somewhat, the covariances are small with respect to the distances between the real values of the object states;

2. clutter is not intense, meaning that the number of false measurements $I_{\mathrm{c}}[1]$ is small with respect the number of objects present in the scene.

To see why, consider the following consider the simple case where 2 objects are present in the scene. Suppose that the sensors are Gaussian and unbiased, so that

$$\ell_k(y|x) = \mathcal{N}(y|\, x, \sigma^2) \tag{4.37}$$

for some track variance σ^2. Moreover, suppose that the sensors are characterized by an ideal probability of detection $p_{\mathrm{D}} = 1$ and that there are no false measurements, so that the multimeasure $\mathsf{y}_k = \{y_1, y_2\}$ is collected knowing that there are no clutter measurements inside. Suppose for simplicity that the predicted MPDF $p_{k|k-1}(\{\cdot\})$ is not informative, hence the corrected MPDF $p_{k|k}(\{\cdot\})$ is essentially the multiobject likelihood $\ell_k(\{\cdot\}|\mathsf{y})$ (which is multi-Bernoulli with two deterministic Gaussian components). Thus

$$p_{k|k}(\mathsf{x}) = \begin{cases} 0 & \text{if } \mathsf{x} = \varnothing \\ 0 & \text{if } \mathsf{x} = \{x_1\} \\ \mathcal{N}(x_1|\, y_1, \sigma^2)\mathcal{N}(x_2|\, y_2, \sigma^2) & \\ \quad + \mathcal{N}(x_1|\, y_2, \sigma^2)\mathcal{N}(x_2|\, y_1, \sigma^2) & \text{if } \mathsf{x} = \{x_1, x_2\} \\ 0 & \text{otherwise} \end{cases}. \tag{4.38}$$

In this case, the corrected PHD is

$$D_{k|k}(x) = \mathcal{N}(x|\, y_1, \sigma^2) + \mathcal{N}(x|\, y_2, \sigma^2) \tag{4.39}$$

and the expected number of objects is

$$\mathbb{E}[|\mathsf{X}_k|] = \int D_{k|k}(x) \; \mathrm{d}x = 2 \tag{4.40}$$

so that the PHD filter produces the estimate $\hat{\mathsf{x}}_k = \{\hat{x}_1, \hat{x}_2\}$, where $\hat{x}_1$ and $\hat{x}_2$ are the locations of the two largest peaks of $D_{k|k}(\cdot)$. It is possible to distinguish three different cases:

- **case 1 -** $|y_1 - y_2| > 2\sigma$: in this case both $D_{k|k}(\cdot)$ and $p_{\mathsf{X}}(\{\cdot\})$ are bimodal, so that both the PHD filter and the multiobject Bayes filter recognize two objects. More precisely, the two filters produce the same estimate $\hat{\mathsf{x}}_k = \{y_1, y_2\}$;

- **case 2 -** $\sqrt{2}\sigma > |y_1 - y_2| > 2\sigma$: in this case $D_{k|k}(\cdot)$ is unimodal while $p_{\mathsf{X}}(\{\cdot\})$ is bimodal, so the PHD filter recognizes only one object while the multiobject Bayes filter recognizes two objects. More precisely, the PHD filter produces the estimate $\hat{\mathsf{x}}_k = \{1/2(y_1 + y_2)\}$, while the multiobject Bayes filter produces the estimate $\hat{\mathsf{x}}_k = \{y_1, y_2\}$;

- **case 3 -** $|y_1 - y_2| < \sqrt{2}\sigma$: in this case both $D_{k|k}(\cdot)$ and $p_{\mathsf{X}}(\{\cdot\})$ are unimodal, so both the PHD filter and the multiobject Bayes filter recognize only one object. More precisely, the PHD filter and the multiobject Bayes filter produce the same estimate $\hat{\mathsf{x}}_k = \{1/2(y_1 + y_2)\}$.

The conclusion is that, for this example, the PHD approximation is appropriate if and only if $|y_1 - y_2| > 2\sigma$. Moreover, the full multiobject Bayes filter outperforms the PHD filter if and only if $\sqrt{2}\sigma > |y_1 - y_2| > 2\sigma$. In the third case $|y_1 - y_2| < \sqrt{2}\sigma$ both filters cannot track the two objects present in the scene.

4.4.4 PHD predictor

Theorem 2. Consider the standard model and let $\tilde{\varphi}_{k|k-1}(\cdot|w) \triangleq p_{\mathrm{S}}(w)\,\varphi_{k|k-1}(\cdot|w)$ be the "pseudo transition density" at time $k-1$. Then, given the corrected PHD $D_{k-1|k-1}(\cdot)$, the predicted PHD is

$$D_{k|k-1}(x) = I_{\mathrm{B}}(x) + D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] \tag{4.41}$$

where the following linear functional is defined

$$D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] \triangleq \int \tilde{\varphi}_{k|k-1}(x|w)\, D_{k-1|k-1}(w)\; \mathrm{d}w \tag{4.42}$$

Proof

The predicted PHD is given by

$$\begin{aligned} D_{k|k-1}(x) &= \left.\frac{\partial\, G_{k|k-1}[h]}{\partial\,\{x\}}\right|_{h=1} \\ &= \frac{\partial}{\partial\,\{x\}}\left[\exp(I_{\mathrm{B}}[h-1])\, G_{k-1|k-1}[1-p_{\mathrm{S}}+\tilde{\varphi}_{k|k-1}[h]]\right]_{h=1} \end{aligned} \tag{4.43}$$

Then, due to the product rule, after some simple calculations it turns out that

$$\begin{aligned} D_{k|k-1}(x) &= \frac{\partial}{\partial\,\{x\}}\left[\exp(I_{\mathrm{B}}[h-1])\right]_{h=1} G_{k-1|k-1}[1-p_{\mathrm{S}}+\tilde{\varphi}_{k|k-1}[1]] + \\ &\qquad \exp(I_{\mathrm{B}}[1-1])\,\frac{\partial}{\partial\,\{x\}}\left[G_{k-1|k-1}[1-p_{\mathrm{S}}+\tilde{\varphi}_{k|k-1}[h]]\right]_{h=1} \\ &= \frac{\partial}{\partial\,\{x\}}\left[\exp(I_{\mathrm{B}}[h-1])\right]_{h=1} + \frac{\partial}{\partial\,\{x\}}\left[G_{k-1|k-1}[1-p_{\mathrm{S}}+\tilde{\varphi}_{k|k-1}[h]]\right]_{h=1} \end{aligned} \tag{4.44}$$

The first term is simply the PHD $I_{\mathrm{B}}(\cdot)$ of the birth set B_k: in fact $\exp(I_{\mathrm{B}}[h-1])$ is the PGFL of B_k, so the functional derivative restricted to $h=1$ of $\exp(I_{\mathrm{B}}[h-1])$ is the PHD of B_k, which is $I_{\mathrm{B}}(\cdot)$, i.e.

$$\frac{\partial}{\partial\,\{x\}}\left[\exp(I_{\mathrm{B}}[h-1])\right]_{h=1} = I_{\mathrm{B}}(x). \tag{4.45}$$

For the second term, the second chain rule can be applied: define the functional transformation

$$\varphi[h](w) \triangleq 1 - p_{\mathrm{S}}(w) + \tilde{\varphi}_{k|k-1}[h|w] \tag{4.46}$$

so that

$$G_{k-1|k-1}[1 - p_\mathrm{S} + \tilde{\varphi}_{k|k-1}[h]] = G_{k-1|k-1}[\varphi[h]]. \tag{4.47}$$

The functional derivative of the second term is

$$\left.\frac{\partial\, G_{k-1|k-1}[\varphi[h]]}{\partial \{x\}}\right|_{h=1} = \int \left.\frac{\partial\, \varphi[h](w)}{\partial \{x\}}\right|_{h=1} \left.\frac{\partial\, G_{k-1|k-1}[\tilde{h}]}{\partial \{w\}}\right|_{\tilde{h}=\varphi[1]} \mathrm{d}w \tag{4.48}$$

Thanks to the linearity of the functional derivative, as well as to the constant and linear rules, the first term under the integral is

$$\left.\frac{\partial\, \varphi[h](w)}{\partial \{x\}}\right|_{h=1} = \frac{\partial}{\partial \{x\}} \left[1 - p_\mathrm{S}(w) + \tilde{\varphi}_{k|k-1}[h|w]\right]_{h=1} = \tilde{\varphi}_{k|k-1}(x|w) \quad . \tag{4.49}$$

On the other hand, the second term under the integral is

$$\left.\frac{\partial\, G_{k-1|k-1}[\tilde{h}]}{\partial \{w\}}\right|_{\tilde{h}=\varphi[1]} = D_{k-1|k-1}(w) \tag{4.50}$$

where it is exploited the fact $\varphi[1](w) = 1$ identically for all $w \in \mathbb{R}^n$. Consequently,

$$\begin{aligned}\left.\frac{\partial\, G_{k-1|k-1}[\varphi[h]]}{\partial \{x\}}\right|_{h=1} &= \int \tilde{\varphi}_{k|k-1}(x|w)\, D_{k-1|k-1}(w)\ \mathrm{d}w \\ &\triangleq D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)]\end{aligned} \quad . \tag{4.51}$$

In conclusion, the predicted PHD is

$$D_{k|k-1}(x) = I_\mathrm{B}(x) + D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] \tag{4.52}$$

as claimed. □

The result can be interpreted as follows: the predicted PHD $D_{k|k-1}(\cdot)$ gets large values in those locations x such that it is likely that an object can appear x (i.e. $I_\mathrm{B}(x)$ is large) or such that it is likely that a survived object move towards (i.e. $D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)]$ is large). On the other hand, the predicted PHD $D_{k|k-1}(\cdot)$ gets low values in those locations x where it is unlikely that an object can appear and a survived object moves towards (i.e. both $I_\mathrm{B}(x)$ and $D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)]$ are small).

4.4.5 PHD corrector

Theorem 3. Consider the standard model (132) and let $\tilde{\ell}_k(\cdot|w) \triangleq p_\mathrm{D}(w)\, \ell_k(\cdot|w)$ be the pseudo likelihood at time k. Moreover, assume that the predicted

MPDF $p_{k|k-1}(\{\cdot\})$ is Poisson for some given intensity $D_{k|k-1}(\cdot)$. Then, the corrected PHD is

$$D_{k|k}(x) = \Lambda(x)\, D_{k|k-1}(x) \tag{4.53}$$

where the PHD likelihood $\Lambda(\cdot)$ is defined as follows

$$\Lambda(x) \triangleq (1 - p_{\mathrm{D}}(x)) + \bar{\Lambda}(\mathsf{y}|x) \tag{4.54}$$

with

$$\bar{\Lambda}(\mathsf{y}|x) \triangleq \sum_{y\in\mathsf{y}} \frac{\tilde{\ell}_k(y|x)}{I_{\mathrm{C}}(y) + D_{k|k-1}[\tilde{\ell}_k(y)]}. \tag{4.55}$$

Proof

Thanks to the assumption that the predicted MPDF is Poisson, the semplification $G_{k|k-1}[h] = \exp(D_{k|k-1}[h-1])$ holds, thus the bivariate PGFL used in the Bayes equation takes the following simple form

$$\begin{aligned} F_k[h,g] &= \exp(I_{\mathrm{C}}[g-1])\, G_{k|k-1}[\tilde{h}]_{\tilde{h}=h\,(1-p_{\mathrm{D}}+\tilde{\ell}_k[g])} \\ &= \exp(I_{\mathrm{C}}[g-1])\, [\exp(D_{k|k-1}[\tilde{h}-1])]_{\tilde{h}=h\,(1-p_{\mathrm{D}}+\tilde{\ell}_k[g])} \\ &= \exp(\underbrace{I_{\mathrm{C}}[g-1] + D_{k|k-1}[h\,(1-p_{\mathrm{D}}+\tilde{\ell}_k[g])-1]}_{\triangleq \iota[h,g]}) \end{aligned} \tag{4.56}$$

where the functional $\iota[\cdot,\cdot]$ is introduced for the sake of notation. Given this explicit expression for the bivariate PGFL, the corrected PHD is

$$D_{k|k}(x) = \left.\frac{\partial\, G_{k|k}[h]}{\partial\,\{x\}}\right|_{h=1} = \frac{\partial}{\partial\,\{x\}}\left[\frac{\left.\frac{\partial\, F_k[h,g]}{\partial_g\,\mathsf{y}}\right|_{g=0}}{\left.\frac{\partial\, F_k[h,g]}{\partial_g\,\mathsf{y}}\right|_{h=1,g=0}}\right]_{h=1}. \tag{4.57}$$

Notice that the denominator of the RHS does not depend on $h(\cdot)$ (since it is already fixed to $h = 1$ before taking the derivative in $\{x\}$), thus it is a constant functional. Consequently, thanks to the linearity of the functional derivative, the external derivative with respect to the singleton $\{x\}$ acts only on the numerator of the RHS, i.e.

$$D_{k|k}(x) = \frac{\frac{\partial}{\partial\,\{x\}}\left[\left.\frac{\partial\, F_k[h,g]}{\partial_g\,\mathsf{y}}\right|_{g=0}\right]_{h=1}}{\left.\frac{\partial\, F_k[h,g]}{\partial_g\,\mathsf{y}}\right|_{h=1,g=0}} = \frac{\left.\frac{\partial\, F_k[h,g]}{\partial\,(\mathsf{y}\cup\{x\})}\right|_{h=1,g=0}}{\left.\frac{\partial\, F_k[h,g]}{\partial_g\,\mathsf{y}}\right|_{h=1,g=0}} \tag{4.58}$$

with the convention that the differentation $\partial/\partial(\mathsf{y}\cup\{x\})$ is performed with respect to $h(\cdot)$ when evaluated in $\{x\}$ and performed with respect $g(\cdot)$ when evaluated in y. At this point the problem is to compute the numerator and the denominator of the RHS and then take their quotient.

step 1: Computation of the denominator

The objective is to compute

$$\left.\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}\right|_{h=1,g=0} = \frac{\partial}{\partial_g \mathsf{y}}\left[\exp(\iota_g[h])\right]_{h=1,g=0} \quad = \frac{\partial}{\partial_g \mathsf{y}}\left[\exp(\iota[g])\right]_{g=0} \quad (4.59)$$

where $\iota[g] \triangleq \iota[1,g]$. Now it is convenient to compute the derivative with respect y by increasing step by step the number of considered measurements. Taking in mind that the differentation in y is always performed with respect $g(\cdot)$, the subscript g will be omitted from the notation ∂/∂_g.

- **case 1**: $\mathsf{y} = \varnothing$ - trivially

$$\left.\frac{\partial F_k[h,g]}{\partial \varnothing}\right|_{h=1} = F_k[1,g] = \exp(\iota[g]) \quad (4.60)$$

- **case 2**: $\mathsf{y} = \{y_1\}$ - due to the first chain rule (82), holds

$$\left.\frac{\partial F_k[h,g]}{\partial \{y_1\}}\right|_{h=1,g=0} = \frac{\partial \exp(\iota[g])}{\partial \{y_1\}} = \left[\frac{\mathrm{d}\exp(\iota[g])}{\mathrm{d}\,\iota[g]}\,\frac{\partial\,\iota[g]}{\partial \{y_1\}}\right]_{g=0} \quad . \quad (4.61)$$

  The first factor is trivially

$$\frac{\mathrm{d}\exp(\iota[g])}{\mathrm{d}\,\iota[g]} = \exp(\iota[g]) \quad (4.62)$$

  while the second factor, according to the linearity of the functional derivative and to the linearity of the linear functional $D_{k|k-1}[\cdot]$, is

$$\frac{\partial\,\iota[g]}{\partial \{y_1\}} = I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)] \quad . \quad (4.63)$$

  Hence,

$$\left.\frac{\partial F_k[h,g]}{\partial \{y_1\}}\right|_{h=1} = \exp(\iota[g])\,\Big(I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)]\Big) \quad (4.64)$$

- **case 3**: $\mathsf{y} = \{y_1, y_2\}$ - According to the definition of iterated functional derivative and observing that the factor

$$\Big(I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)]\Big) \quad (4.65)$$

in $\frac{\partial F_k[h,\cdot]}{\partial\{y_1\}}$ does not depend on $g(\cdot)$, it holds that

$$\begin{aligned}\left.\frac{\partial F_k[h,g]}{\partial\{y_1,y_2\}}\right|_{h=1} &= \frac{\partial \exp(\iota[g])}{\partial\{y_2\}}\Big(I_{\mathrm{C}}(y_1)+D_{k|k-1}[\tilde{\ell}_k(y_1)]\Big)\\ &= \exp(\iota[g])\prod_{i=1}^{2}\Big(I_{\mathrm{C}}(y_i)+D_{k|k-1}[\tilde{\ell}_k(y_i)]\Big)\end{aligned}. \tag{4.66}$$

The previous result suggest the following general formula

$$\left.\frac{\partial F_k[h,g]}{\partial \mathsf{y}}\right|_{h=1,g=0} = \exp(\iota[g])\prod_{y\in\mathsf{y}}\Big(I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)]\Big) \tag{4.67}$$

which can be proved by the induction $\mathsf{y}=\{y_1,\dots,y_i\}\to\mathsf{y}=\{y_1,\dots,y_{i+1}\}$. Setting $g=0$ yields to the final expression of the denominator

$$\left.\frac{\partial F_k[h,g]}{\partial \mathsf{y}}\right|_{h=1,g=0} = \exp(\iota[0])\Big(I_{\mathrm{C}}+D_{k|k-1}[\tilde{\ell}_k]\Big)^{\mathsf{y}} \quad . \tag{4.68}$$

step 2: Computation of the numerator

The objective is to compute

$$\left.\frac{\partial F_k[h,g]}{\partial(\mathsf{y}\cup\{x\})}\right|_{h=1,g=0} = \left.\frac{\partial \exp(\iota_g[h])}{\partial(\mathsf{y}\cup\{x\})}\right|_{h=1,g=0} = \frac{\partial}{\partial\mathsf{y}}\left[\left.\frac{\partial\exp(\iota_g[h])}{\partial\{x\}}\right|_{h=1}\right]_{g=0}. \tag{4.69}$$

Hence, the functional derivative in (191) is splitted in two different functional derivatives: the former is in $\{x\}$ while the latter is in y. Note that, since the differentation is order invariant, it is also possible to perform the differentation by taking first the derivative in y and then in $\{x\}$, but this strategy leads to more complicated computations, hence is not considered in what follows.

Derivative with respect to $\{x\}$

Recall the fact that the differentation in $\{x\}$ acts only on the test function $h(\cdot)$, while the differentation in y acts only on the test function $g(\cdot)$. In other words, the two differentations behave like two partial differentations. Given this fact and according to the first chain rule, the differentation in $\{x\}$ yields to

$$\frac{\partial\exp(\iota_g[h])}{\partial\{x\}} = \frac{\mathrm{d}\exp(\iota_g[h])}{\mathrm{d}\,\iota_g[h]}\frac{\partial\,\iota_g[h]}{\partial\{x\}} = \exp(\iota_g[h])\frac{\partial\,\iota_g[h]}{\partial\{x\}}. \tag{4.70}$$

The second factor in the RHS, by linearity, is

$$\frac{\partial\, \iota_g[h]}{\partial\, \{x\}} = \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) D_{k|k-1}(x) \quad . \tag{4.71}$$

Consequently, it follows that

$$\frac{\partial\, \exp(\iota_g[h])}{\partial\, \{x\}} = \exp(\iota_g[h]) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) D_{k|k-1}(x). \tag{4.72}$$

Now, setting $h = 1$, leads to

$$\left.\frac{\partial\, \exp(\iota_g[h])}{\partial\, \{x\}}\right|_{h=1} = \exp(\iota[g]) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) D_{k|k-1}(x) \tag{4.73}$$

Derivative with respect to y

At this point it remains to evaluate the functional derivative in y of . To this end, repeat the procedure used to compute the denominator in (1.58).

- **case 1**: $\mathsf{y} = \varnothing$ - Trivially

$$\begin{aligned}\left.\frac{\partial\, F_k[h,g]}{\partial\, (\varnothing \cup \{x\})}\right|_{h=1} &= \frac{\partial}{\partial\, \varnothing}\left[\left.\frac{\partial\, \exp(\iota_g[h])}{\partial\, \{x\}}\right|_{h=1}\right] = \left.\frac{\partial\, \exp(\iota_g[h])}{\partial\, \{x\}}\right|_{h=1} \\ &= \exp\left(\iota[g]\right) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) D_{k|k-1}(x)\end{aligned} \tag{4.74}$$

- **case 2**: $\mathsf{y} = \{y_1\}$ - By linearity and according to the product rule, it holds that

$$\begin{aligned}\left.\frac{\partial\, F_k[h,g]}{\partial\, (\{y_1\} \cup \{x\})}\right|_{h=1} &= \frac{\partial}{\partial\, \{y_1\}}\left[\left.\frac{\partial\, \exp(\iota_g[h])}{\partial\, \{x\}}\right|_{h=1}\right] \\ &= \frac{\partial}{\partial\, \{y_1\}}\left[\exp\left(\iota[g]\right) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) D_{k|k-1}(x)\right] \\ &= \frac{\partial}{\partial\, \{y_1\}}\left[\exp\left(\iota[g]\right) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right)\right] D_{k|k-1}(x) \\ &= \exp(\iota[g])\, T[g]\, D_{k|k-1}(x)\end{aligned} \tag{4.75}$$

where, in order to simplify the notation, the following functional is defined

$$T[g] \triangleq \left(I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)]\right) \left(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]\right) + \tilde{\ell}_k(y_1|x) \tag{4.76}$$

- **case 3**: $\mathsf{y} = \{y_1, y_2\}$ - According to the definition of iterated functional derivative, holds

$$\begin{aligned}\left.\frac{\partial\, F_k[h,g]}{\partial\left(\{y_1,y_2\}\cup\{x\}\right)}\right|_{h=1} &= \frac{\partial}{\partial\,\{y_2\}}\left[\left.\frac{\partial\, F_k[h,g]}{\partial\left(\{y_1\}\cup\{x\}\right)}\right|_{h=1}\right] \\ &= \frac{\partial}{\partial\,\{y_2\}}\left[\exp(\iota[g])\,T[g]\,D_{k|k-1}(x)\right] \\ &= \frac{\partial\,\exp(\iota[g])\,T[g]}{\partial\,\{y_2\}}\,D_{k|k-1}(x) \\ &= \exp(\iota[g])\left(U[g] + \frac{\partial\,T[g]}{\partial\{y_2\}}\right)\,D_{k|k-1}(x)\end{aligned} \tag{4.77}$$

where the shorthand $U[\cdot]$ is defined as follows

$$\begin{aligned}U[g] = &\left(\prod_{i=1}^{2}(I_{\mathrm{C}}(y_i) + D_{k|k-1}[\tilde{\ell}_k(y_i)])\right)(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]) \\ &+ (I_{\mathrm{C}}(y_2) + D_{k|k-1}[\tilde{\ell}_k(y_2)])\tilde{\ell}_k(y_1|x)\end{aligned} \tag{4.78}$$

and

$$\frac{\partial\,T[g]}{\partial\{y_2\}} = (I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)])\,\tilde{\ell}_k(y_2|x) \quad . \tag{4.79}$$

Now focus on the sum $U[g] + \frac{\partial\,T[g]}{\partial\{y_2\}}$, which is

$$\begin{aligned}U[g] + \frac{\partial\,T[g]}{\partial\{y_2\}} = &\left(\prod_{i=1}^{2}(I_{\mathrm{C}}(y_i) + D_{k|k-1}[\tilde{\ell}_k(y_i)])\right)(1 - p_{\mathrm{D}}(x) + \tilde{\ell}_k[g|x]) \\ &+ (I_{\mathrm{C}}(y_2) + D_{k|k-1}[\tilde{\ell}_k(y_2)])\tilde{\ell}_k(y_1|x) \\ &\quad + (I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)])\,\tilde{\ell}_k(y_2|x)\end{aligned} \quad . \tag{4.80}$$

The sum, say S, of the second and third terms can be written in the following form

$$\begin{aligned}S = &\left(\prod_{i=1}^{2}(I_{\mathrm{C}}(y_i) + D_{k|k-1}[\tilde{\ell}_k(y_i)])\right)\left[\frac{\tilde{\ell}_k(y_1|x)}{I_{\mathrm{C}}(y_1) + D_{k|k-1}[\tilde{\ell}_k(y_1)]} \right. \\ &\left. + \frac{\tilde{\ell}_k(y_2|x)}{I_{\mathrm{C}}(y_2) + D_{k|k-1}[\tilde{\ell}_k(y_2)]}\right] \quad . \\ = &\left(\prod_{i=1}^{2}(I_{\mathrm{C}}(y_i) + D_{k|k-1}[\tilde{\ell}_k(y_i)])\right)\left(\sum_{i=1}^{2}\frac{\tilde{\ell}_k(y_i|x)}{I_{\mathrm{C}}(y_i) + D_{k|k-1}[\tilde{\ell}_k(y_i)]}\right)\end{aligned} \tag{4.81}$$

Hence it follows that

$$\left.\frac{\partial F_k[h,g]}{\partial\left(\{y_1,y_2\}\cup\{x\}\right)}\right|_{h=1} = \exp(\iota[g])\left(\prod_{i=1}^{2}(I_{\mathrm{C}}(y_i)+D_{k|k-1}[\tilde{\ell}_k(y_i)])\right) \times\left((1-p_{\mathrm{D}}(x)+\tilde{\ell}_k[g|x])+\sum_{i=1}^{2}\frac{\tilde{\ell}_k(y_i|x)}{I_{\mathrm{C}}(y_i)+D_{k|k-1}[\tilde{\ell}_k(y_i)])}\right) \times D_{k|k-1}(x) \tag{4.82}$$

The previous results suggest the following general formula

$$\left.\frac{\partial F_k[h,g]}{\partial\left(\mathsf{y}\cup\{x\}\right)}\right|_{h=1} = \exp(\iota[g])\left(\prod_{y\in\mathsf{y}}(I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)])\right) \times\left((1-p_{\mathrm{D}}(x)+\tilde{\ell}_k[g|x])+\sum_{y\in\mathsf{y}}\frac{\tilde{\ell}_k(y|x)}{I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)]}\right) \times D_{k|k-1}(x) \tag{4.83}$$

which can be proved by the induction $\mathsf{y}=\{y_1,\ldots,y_i\}\rightarrow\mathsf{y}=\{y_1,\ldots,y_{i+1}\}$. Setting $g=0$ yields to the final expression

$$\left.\frac{\partial F_k[h,g]}{\partial\left(\mathsf{y}\cup\{x\}\right)}\right|_{h=1} = \exp(\iota[g])\left(\prod_{y\in\mathsf{y}}(I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)])\right) \times\left((1-p_{\mathrm{D}}(x))+\sum_{y\in\mathsf{y}}\frac{\tilde{\ell}_k(y|x)}{I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)]}\right) \times D_{k|k-1}(x). \tag{4.84}$$

step 3: Final result

By dividing (1.68) with (1.84), it turns out, as claimed, that

$$D_{k|k}(x) = \left((1-p_{\mathrm{D}}(x))+\sum_{y\in\mathsf{y}}\frac{\tilde{\ell}_k(y|x)}{I_{\mathrm{C}}(y)+D_{k|k-1}[\tilde{\ell}_k(y)]}\right)D_{k|k-1}(x) = \Lambda(x)\,D_{k|k-1}(x) \tag{4.85}$$

where the PHD likelihood $\Lambda(\cdot)$ is defined as follows

$$\Lambda(x)\triangleq(1-p_{\mathrm{D}}(x))+\bar{\Lambda}(\mathsf{y}|x) \tag{4.86}$$

with

$$\bar{\Lambda}(\mathsf{y}|x) \triangleq \sum_{y \in \mathsf{y}} \frac{\tilde{\ell}_k(y|x)}{I_{\mathrm{C}}(y) + D_{k|k-1}[\tilde{\ell}_k(y)]}. \tag{4.87}$$

□

The result can be interpreted as follows: the corrected PHD $D_{k|k}(\cdot)$ gets high values in those location x where the sensors are blind (i.e. the probability of miss detection $1 - p_{\mathrm{D}}(x)$ is high) or where are near at least to one *reliable* measure y (i.e. $\bar{\Lambda}(x)$ is high). A measurement y is reliable in x if and only if:

- the pseudo likelihood $\tilde{\ell}_k(y|x) \triangleq p_{\mathrm{D}}(x)\, \ell_k(y|x)$ is large, i.e. if the spot x is well observed by the sensors ($p_{\mathrm{D}}(x)$ is large) and the measurement y is actually *near* (according to the intensity of the sensor noise) the spot x ($\ell_k(y|x)$ is large);
- it is likely that the measurement is generated by an object and not by clutter (i.e. $I_{\mathrm{C}}(y)$ is small with respect to

$$D_{k|k-1}[\tilde{\ell}_k(y)] \triangleq \int \tilde{\ell}_k(y|w)\, D_{k|k-1}(w)\; \mathrm{d}w \tag{4.88}$$

which is a pseudo expected value of $\tilde{\ell}_k(y|X)$ weighted by $D_{k|k-1}(x)$ and thus a sort of estimate of the event 'y is generated by an object').

On the other hand, the corrected PHD $D_{k|k}(\cdot)$ gets low values in those location x that are far from any reliable measurement.

4.5 Gaussian mixture PHD filter

4.5.1 PHD filter implementations

There are two ways to translate the standard PHD filter in an algorithm executable by a computer.

1. **Particle approximation**: the idea is to approximate the predicted and the corrected PHDs as linear combinations of delta densities, i.e.

$$D_{k|k-1}(x) \approx \sum_{i=1}^{\nu} w^i_{k|k-1}\, \delta_{x^i_{k|k-1}}(x) \qquad D_{k|k}(x) \approx \sum_{i=1}^{\nu} w^i_{k|k}\, \delta_{x^i_{k|k}}(x) \tag{4.89}$$

for suitable sets of predicted particles $\{w^i_{k|k-1}, x^i_{k|k-1}\}_{i=1}^{\nu}$ and corrected particles $\{w^i_{k|k}, x^i_{k|k}\}_{i=1}^{\nu}$, which are computed according to the prediction and correction steps of the PHD filter. Note that the PHDs are not normalized functions, so the importance weights $w^i_{k|k-1}$,$w^i_{k|k}$ does not sum to the unity but rather to $N_{k|k-1}$, $N_{k|k}$ (thus, the approximation is not a mixture of deltas but a linear combination, with non-negative coefficients, of deltas).

The resulting algorithm is called *Sequential Monte Carlo PHD filter* (SMC-PHD filter).

The SMC-PHD filter leads to good performance if the number of particles ν (which is usually fixed in time) is *sufficiently large*. Here the term *large* strongly depends on the SNR, the dimension of the object states and the number of tracked objects. Tipically the SMC-PHD filter is computationally demanding because the number of particles ν involved is large.

2. **Gaussian mixture approximation**: the idea is to approximate the predicted and corrected PHDs as linear combinations of Gaussian functions

$$\begin{aligned} D_{k|k-1}(x) &\approx \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1}\, \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \\ D_{k|k}(x) &\approx \sum_{i=1}^{\nu_{k|k}} w^i_{k|k}\, \mathcal{N}(x; x^i_{k|k}, P^i_{k|k}) \end{aligned} \tag{4.90}$$

for suitable sets of predicted Gaussian components

$$\{w^i_{k|k-1}, (x^i_{k|k-1}, P^i_{k|k-1})\}_{i=1}^{\nu_{k|k-1}} \tag{4.91}$$

and corrected Gaussian components

$$\{w^i_{k|k}, x^i_{k|k}, P^i_{k|k})\}_{i=1}^{\nu_{k|k}} \tag{4.92}$$

which are computed according the prediction and correction steps of the PHD filter. Once again, the importance weights sum to $N_{k|k-1}$ and $N_{k|k}$, thus the term *Gaussian mixture* is used improperly as a shorthand for *linear combination of Gaussians.*

The resulting algorithm is called *Gaussian mixture PHD filter* (GM-PHD filter).

For an individual element of the mixture, in principle, the GM-PHD is more computational demanding than the SMC-PHD filter since it requires the computation of the additional parameters $P^i_{k|k-1}$, $P^i_{k|k}$ (the predicted and corrected covariances of the Gaussian kernels). However in general the GM-PHD achieves good performance for relatively small number of Gaussian components $\nu_{k|k-1}$, $\nu_{k|k}$ (note that now these numbers are not fixed in time) with respect to the number of particles ν required by the SMC-PHD filter. Moreover, the GM-PHD filter makes some particular assumptions that allows to compute algebraically the predicted and corrected parameters of the Gaussian kernels by means of a standard Kalman filter (thus, with a very simple procedure). Hence, the GM-PHD filter tends to be less computationally demanding than the SMC-PHD filter, but, due to its additional assumptions, turns out to be more restrictive than the SMC-PHD filter.

The focus of this thesis is on the GM-PHD filter, which will be derived in what follows. Both the GM-PHD predictor and the GM-PHD corrector rely on the following well-known result about the product of Gaussian PDFs

Theorem 4. (Foundamental Gaussian identity) - Let C be a $p \times n$ matrix with $p \leq n$, let R and P be $p \times p$ and $n \times n$ covariance matrices, then

$$\mathcal{N}(y; Cx, R)\, \mathcal{N}(x; \hat{x}, P) = \mathcal{N}(y; \hat{y}; S)\, \mathcal{N}(x; \Omega^{-1} q, \Omega^{-1}) \tag{4.93}$$

where S, $\hat{y}$ are given by the Kalman predictor (standard form) while Ω, q are given by the Kalman corrector (information form)

$$\begin{aligned} S &\triangleq R + CPC' & \qquad \Omega &\triangleq P^{-1} + C' R^{-1} C \\ \hat{y} &\triangleq C\hat{x} & q &\triangleq P^{-1}\hat{x} + C' R^{-1} y \end{aligned} \tag{4.94}$$

4.5.2 GM-PHD predictor

Theorem 5. Consider the following assumptions:

- the corrected PHD is a Gaussian mixture of the following type

$$D_{k-1|k-1}(x) = \sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k-1|k-1}\, \mathcal{N}(x; x^i_{k-1|k-1}, P^i_{k-1|k-1}) \tag{4.95}$$

- the single-object motion model is linear-Gaussian, i.e. $X_{k+1} = A_k X_k + W_k$ with $W_k \sim \mathcal{N}(0, Q_k)$, thus the transition density considered is of the form

$$\varphi_{k|k-1}(x|w) = \mathcal{N}(x; A_k\, w, Q_k) \tag{4.96}$$

- the survival probability is constant all over the surveilled scene, i.e.

$$p_S(x) = p_S \qquad \forall\, x \in \mathbb{R}^n \tag{4.97}$$

- the PHD of the birth set is a "Gaussian mixture" of the following type

$$I_{\mathrm{B},k}(x) = \sum_{i=1}^{b_k} \beta_k^i \, \mathcal{N}(x; x_{\mathrm{B}}^i; B_k^i) \tag{4.98}$$

Then, the predicted PHD is still a "Gaussian mixture", more precisely

$$D_{k|k-1}(x) = I_{\mathrm{B},k}(x) + \sum_{i=1}^{\nu_{k-1|k-1}} w_{k|k-1}^i \, \mathcal{N}(x; x_{k|k-1}^i, P_{k|k-1}^i) \tag{4.99}$$

where the number of predicted Gaussian components is $\nu_{k|k-1} = b_k + \nu_{k-1|k-1}$ and for every survived Gaussian component, i.e. for $i = 1, 2, \ldots, \nu_{k-1|k-1}$, the following facts hold:

- the predicted weight is given by

$$w_{k|k-1}^i = p_{\mathrm{S}} \, w_{k-1|k-1}^i \tag{4.100}$$

- the parameters of the Gaussian kernel are given by the Kalman predictor which, in standard form, are the followings

$$\begin{aligned} x_{k|k-1}^i &= A_k \, x_{k-1|k-1} \\ P_{k|k-1}^i &= A_k \, P_{k-1|k-1}^i A_k' + Q_k \end{aligned} \tag{4.101}$$

Proof

The theorem is proved if and only if

$$D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] = \sum_{i=1}^{\nu_{k|k-1}} w_{k|k-1}^i \, \mathcal{N}(x; x_{k|k-1}^i, P_{k|k-1}^i). \tag{4.102}$$

In order to do that, observe that the pseudo-transition density is

$$\tilde{\varphi}_{k|k-1}(x|w) = p_{\mathrm{S}} \, \varphi_{k|k-1}(x|w) = p_{\mathrm{S}} \, \mathcal{N}(x; A_k \, w, Q_k). \tag{4.103}$$

As a consequence, it holds that

$$\begin{aligned} D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] &\triangleq \int \tilde{\varphi}_{k|k-1}(x|w) \, D_{k-1|k-1}(w) \, \mathrm{d}w \\ \triangleq \sum_{i=1}^{\nu_{k-1|k-1}} w_{k|k-1}^i & \int \mathcal{N}(x; A_k \, w, Q_k) \, \mathcal{N}(w; x_{k-1|k-1}^i, P_{k-1|k-1}^i) \, \mathrm{d}w \end{aligned} \tag{4.104}$$

where $w^i_{k|k-1} \triangleq p_S\, w^i_{k-1|k-1}$. Now, for the fundamental Gaussian identity, it general holds that

$$\mathcal{N}(x; A_k\, w, Q_k)\, \mathcal{N}(w; x^i_{k-1|k-1}, P^i_{k-1|k-1}) = \mathcal{N}(x; x^i_{k|k-1}; P^i_{k|k-1})\, \mathcal{N}(w; \mu; \Sigma) \tag{4.105}$$

for some suitables moments μ, Σ and for $x^i_{k|k-1}$, $P^i_{k|k-1}$ given by the Kalman predictor

$$\begin{aligned} x^i_{k|k-1} &\triangleq A_k x^i_{k-1|k-1} \\ P^i_{k|k-1} &\triangleq A_k\, P^i_{k-1|k-1}\, A'_k + Q_k \end{aligned} . \tag{4.106}$$

Thus, in conclusion, it follows that the PHD of survived objects $D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)]$ simplifies to

$$\begin{aligned} D_{k-1|k-1}[\tilde{\varphi}_{k|k-1}(x)] &= \sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k|k-1} \mathcal{N}(x; x^i_{k|k-1}; P^i_{k|k-1}) \underbrace{\int \mathcal{N}(w; \mu; \Sigma)\, \mathrm{d}w}_{=1} \\ &= \sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k|k-1} \mathcal{N}(x; x^i_{k|k-1}; P^i_{k|k-1}) \end{aligned} \tag{4.107}$$

as claimed. □

4.5.3 GM-PHD corrector

Theorem 6. Consider the following assumptions:

- the predicted PHD is a "Gaussian mixture" of the following form

$$D_{k|k-1}(x) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1}\, \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \tag{4.108}$$

- the single-object measurement model is linear-Gaussian, i.e. $Y_k = C_k X_k + V_k$ with $V_k \sim \mathcal{N}(0, R_k)$, thus the considered likelihood is of the form

$$\ell_k(y|w) = \mathcal{N}(y; C_k\, w, R_k) \tag{4.109}$$

- the detection probability is constant all over the surveilled scene, i.e.

$$p_D(x) = p_D \qquad \forall\, x \in \mathbb{R}^n. \tag{4.110}$$

Then, by denoting $\mathsf{y}_k = \{y_1, \ldots, y_{m_k}\}$, the corrected PHD is

$$D_{k|k}(x) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k} \mathcal{N}\left(x; x^i_{k|k}, P^i_{k|k}\right) + \sum_{i=1}^{\nu_{k|k-1}} \sum_{j=1}^{m_k} w^{i,j}_{k|k} \mathcal{N}\left(x; x^{i,j}_{k|k}, P^i_{k|k}\right) \tag{4.111}$$

where the number of corrected Gaussian components is $\nu_{k|k} = \nu_{k|k}(1 + m_k)$ and for every predicted component $i = 1, 2, \ldots, \nu_{k|k}$ and for every measurement $j = 1, 2, \ldots, m_k$, the following facts hold:

- the corrected weights are given
  - for the *undetected objects*, by

$$w^i_{k|k} \triangleq (1 - p_{\mathrm{D}})\, w^i_{k|k-1} \tag{4.112}$$

  - for the *detected objects*, by

$$w^{i,j}_{k|k} \triangleq \frac{p_{\mathrm{D}}\, w^i_{k|k-1}\, \mathcal{N}\left(y_j; \hat{y}^i_{k|k-1}, S^i_k\right)}{I_{\mathrm{C}}(y_j) + \sum_{\iota=1}^{\nu_{k|k-1}} p_{\mathrm{D}}\, w^{\iota}_{k|k-1} \mathcal{N}\left(y_j; y^{\iota}_{k|k-1}; S^{\iota}_k\right)} \tag{4.113}$$

    where $\hat{y}^i_{k|k-1}$, $S^i_{k|k-1}$ are given by the Kalman predictor, i.e. in standard form

$$\begin{aligned} S^i_k &\triangleq R_k + C_k P^i_{k|k-1} C'_k \\ \hat{y}^i_{k|k-1} &\triangleq C_k x^i_{k|k-1} \end{aligned} \tag{4.114}$$

- the corrected parameters of the Gaussian kernels are given
  - for the *undetected objects*, by the predicted parameters ($\equiv$ correction bypassed)

$$\begin{aligned} P^i_{k|k} &\triangleq P^i_{k|k-1} \\ x^i_{k|k} &\triangleq x^i_{k|k-1} \end{aligned} \tag{4.115}$$

  - for the *detected objects*, by the Kalman corrector, i.e. in standard form

$$\begin{aligned} L^i_k &\triangleq P^i_{k|k-1} C'_k \left(S^i_k\right)^{-1} \\ P^i_{k|k} &\triangleq (I - L^i_k C_k) P^i_{k|k-1} \\ x^{i,j}_{k|k} &\triangleq x^i_{k|k-1} + L^i_k (y_j - \hat{y}^j_{k|k-1}) \end{aligned} \tag{4.116}$$

PROOF

The objective is to compute the PHD correction step under the actual simplifying assumptions. First of all, trivially, since the probability of detection is constant and the notation considered to express the multimeasurement is $\mathsf{y}_k = \{y_1, \ldots, y_{m_k}\}$, it holds that

$$\begin{aligned} D_{k|k}(x) &= \left[(1 - p_{\mathrm{D}}) + \sum_{j=1}^{m_k} \frac{p_{\mathrm{D}}\, \ell_k(y_j|x)}{I_{\mathrm{C}}(y_j) + D_{k|k-1}[p_{\mathrm{D}}\, \ell_k(y_j)]} \right] D_{k|k-1}(x) \\ &= \underbrace{(1 - p_{\mathrm{D}})\, D_{k|k-1}(x)}_{\triangleq D^{\mathrm{D}}_{k|k-1}(x)} + \underbrace{\sum_{j=1}^{m_k} \frac{p_{\mathrm{D}}\, \ell_k(y_j|x)}{I_{\mathrm{C}}(y_j) + p_{\mathrm{D}}\, D_{k|k-1}[\ell_k(y_j)]}\, D_{k|k-1}(x)}_{\triangleq D^{\mathrm{D}}_{k|k-1}(x)} \end{aligned} \tag{4.117}$$

the corrected PHD $D_{k|k}(\cdot)$ thus is given by the sum of two distinct PHDs: the PHD of undetected objects $D^{\mathrm{ND}}_{k|k-1}(\cdot)$ and the PHD of detected objects $D^{\mathrm{D}}_{k|k-1}(\cdot)$.

Undetected objects PHD

Recalling the Gaussian form of the predicted PHD yields to

$$\begin{aligned} D^{\mathrm{ND}}_{k|k}(x) &= (1 - p_{\mathrm{D}}) \left(\sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1}\, \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \right) \\ &= \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k}\, \mathcal{N}(x; x^i_{k|k}, P^i_{k|k}) \end{aligned} \tag{4.118}$$

where are defined the undetected parameters as follows

$$\begin{aligned} w^i_{k|k} &\triangleq (1 - p_{\mathrm{D}})\, w^i_{k|k-1} \\ P^i_{k|k} &\triangleq P^i_{k|k-1} \\ x^i_{k|k} &\triangleq x^i_{k|k-1} \end{aligned} \quad . \tag{4.119}$$

Detected objects PHD

The Gaussian form of the predicted PHD implies

$$
\begin{aligned}
D^{\mathrm{D}}_{k|k}(x) &= \left(\sum_{j=1}^{m_k} \frac{p_{\mathrm{D}}\, \ell_k(y_j|x)}{I_{\mathrm{C}}(y_j) + p_{\mathrm{D}}\, D_{k|k-1}[\ell_k(y_j)]} \right) \left(\sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1} \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \right) \\
&= \sum_{i=1}^{\nu_{k|k-1}} \sum_{j=1}^{m_k} \frac{p_{\mathrm{D}}\, w^i_{k|k-1}}{I_{\mathrm{C}}(y_j) + p_{\mathrm{D}}\, D_{k|k-1}[\ell_k(y_j)]}\, \ell_k(y_j|x)\, \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1})
\end{aligned}
\tag{4.120}
$$

By exploiting the Gaussian form of the likelihood and according to the foundamental Gaussian identity, the kernel of the PHD can be written in the following equivalent form

$$
\begin{aligned}
\ell_k(y_j|x)\, \mathcal{N}\left(x; x^i_{k|k-1}, P^i_{k|k-1}\right) &= \mathcal{N}\left(y_j; C_k x, R_k\right)\, \mathcal{N}\left(x; x^i_{k|k-1}, P^i_{k|k-1}\right) \\
&= \mathcal{N}\left(y_j; C_k x, R_k\right)\, \mathcal{N}\left(x; x^{i,j}_{k|k-1}, P^i_{k|k-1}\right)
\end{aligned}
\tag{4.121}
$$

where $\hat{y}^i_{k|k-1}$, S^i_k are given by the Kalman predictor

$$
\begin{aligned}
S^i_k &\triangleq R_k + C_k P^i_{k|k-1} C'_k \\
\hat{y}^i_{k|k-1} &\triangleq C_k x^i_{k|k-1}
\end{aligned}
\tag{4.122}
$$

while $P^i_{k|k}$, $x^{i,j}_{k|k}$ are given by the Kalman corrector (in the standard correction gain form), i.e.

$$
\begin{aligned}
L^i_k &\triangleq P^i_{k|k-1} C'_k \left(S^i_k\right)^{-1} \\
P^i_{k|k} &\triangleq (I - L^i_k C_k) P^i_{k|k-1} \\
x^{i,j}_{k|k} &\triangleq x^i_{k|k-1} + L^i_k (y_j - \hat{y}^j_{k|k-1})
\end{aligned}
\tag{4.123}
$$

hence it turns out that

$$
D^{\mathrm{D}}_k(x) = \sum_{i=1}^{\nu_{k|k-1}} \sum_{j=1}^{m_k} \frac{p_{\mathrm{D}}\, w^i_{k|k-1} \mathcal{N}\left(y_j; \hat{y}^i_{k|k-1}, S^i_k\right)}{I_{\mathrm{C}}(y_j) + p_{\mathrm{D}}\, D_{k|k-1}[\ell_k(y_j)]}\, \mathcal{N}\left(x; x^{i,j}_{k|k-1}, P^i_{k|k-1}\right)
\tag{4.124}
$$

Now focus on the linear functional $D_{k|k-1}[\ell_k(y_j)]$. Recalling the Gaussian forms for the predicted PHD $D_{k|k-1}(\cdot)$ and for the likelihood $\ell_k(\cdot|w)$, it holds that

$$D_{k|k-1}[\ell_k(y_j)] \triangleq \int \ell_k(y_j|w)\, D_{k|k-1}(w)\ \mathrm{d}w \\ = \sum_{\iota=1}^{\nu_{k|k-1}} w^{\iota}_{k|k-1} \int \mathcal{N}\left(y; C_k w, R_k\right) \mathcal{N}\left(w; x^{\iota}_{k|k-1}, P^{\iota}_{k|k-1}\right)\ \mathrm{d}w \quad (4.125)$$

Once again, according to the fundamental Gaussian identity, the integrand can be written in the form

$$\mathcal{N}\left(y; C_k w, R_k\right) \mathcal{N}\left(w; x^{\iota}_{k|k-1}, P^{\iota}_{k|k-1}\right) = \mathcal{N}\left(y; \hat{y}^{\iota}_{k|k-1}, S^{\iota}_k\right) \mathcal{N}\left(w; \mu^{\iota}_W, P^{\iota}_W\right) \quad (4.126)$$

where μ^i_W, P^i_W are suitable moments, thus

$$D_{k|k-1}[\ell_k(y_j)] = \sum_{\iota=1}^{\nu_{k|k-1}} w^{\iota}_{k|k-1} \int \mathcal{N}\left(y; \hat{y}^{\iota}_{k|k-1}, S^{\iota}_k\right) \mathcal{N}\left(w; \mu^{\iota}_W, P^{\iota}_W\right)\ \mathrm{d}w \\ = \sum_{\iota=1}^{\nu_{k|k-1}} w^{\iota}_{k|k-1} \mathcal{N}\left(y; \hat{y}^{\iota}_{k|k-1}, S^{\iota}_k\right) \quad (4.127)$$

Consequently the detected PHD gets the form

$$D^{\mathrm{D}}_k(x) \triangleq \sum_{i=1}^{\nu_{k|k-1}} \sum_{j=1}^{m_k} w^{i,j}_{k|k} \mathcal{N}\left(x; x^{i,j}_{k|k-1}, P^i_{k|k-1}\right) \quad (4.128)$$

where the corrected weights are defined as

$$w^{i,j}_{k|k} \triangleq \frac{p_{\mathrm{D}}\, w^i_{k|k-1} \mathcal{N}\left(y_j; \hat{y}^i_{k|k-1}, S^i_k\right)}{I_{\mathrm{C}}(y_j) + \sum_{\iota=1}^{\nu_{k|k-1}} p_{\mathrm{D}}\, w^{\iota}_{k|k-1} \mathcal{N}\left(y_j; \hat{y}^{\iota}_{k|k-1}, S^{\iota}_k\right)} \quad (4.129)$$

Final result

Finally, by summing $D^{\mathrm{ND}}_k(\cdot)$ and $D^{\mathrm{D}}_k(\cdot)$, it turns out that

$$D_{k|k}(x) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k} \mathcal{N}\left(x; x^i_{k|k}, P^i_{k|k}\right) + \sum_{i=1}^{\nu_{k|k-1}} \sum_{j=1}^{m_k} w^{i,j}_{k|k} \mathcal{N}\left(x; x^{i,j}_{k|k}, P^i_{k|k}\right) \quad (4.130)$$

as claimed. □

4.5.4 Postprocessing

The number of predicted hypotheses is $\nu_{k|k-1} = b_k + \nu_{k-1|k-1}$, thus grows exponentially in time. In the same way, the number of corrected hypotheses is $\nu_{k|k} = \nu_{k|k-1}(1+m_k)$, wich grows exponentially in time (even more rapidly then $\nu_{k|k-1}$). In order to mantain the algorithm computationally feasible, particular techniques are employed after the prediction and correction steps of the GM-PHD filter to limit $\nu_{k|k-1}$ and $\nu_{k|k}$ below a desidered treshold $\nu_{\max}$. Assume that the corrected or predicted hypotheses are in the form

$$\{(w^i, x^i, P^i)\}_{i=1}^{\nu} \tag{4.131}$$

Such techniques are essentially the following three (performed in the same order as presented below):

- **Pruning**: The irrelevant hypothesis are discharged. Here *irrelevant* means that the relative weight of an hypothesis is smaller than a preset threshold.

  $$\text{if } w^i < \gamma_1 \text{ then } (w^i, x^i, P^i) \text{ is eliminated} \tag{4.132}$$

  Note that the weights, since they sum to a quantity that it is not necessarily the unity, don't need to be normalized to the unity after the pruning of the irrelevant hypotheses. Rather, if one doesn't want to loose the information about the volume of the PHD then the weights can be renormalized to the value of the volume of not pruned.

- **Merging**: if two or more hypotheses are similar then they can be approximated with a suitable unique hypothesis. More precisely, here *similar* means that if the merging distance[1] between two hypothesis i, j

  $$d(i,j) \triangleq ||x^i - x^j||^2_{P^i} \tag{4.133}$$

  is smaller than a preset threshold then the hypotheses i, j are merged into a new single hypothesis k

  $$\text{if } d(i,j) < \gamma_2 \text{ then } (w^i, x^i, P^i), (w^k, x^k, P^k) \text{ are replaced by } (w^k, x^k, P^k) \tag{4.134}$$

[1] This is not a distance because, due to the weight matrix P^i, is not symmetric: $d(i,j) \neq d(j,i)$. However, if i and j are corrected hypotheses for detected objects then the covariances P^i, P^j are the same and the term *merging distance* $d(i,j)$ assumes the real meaning of distance

where the merged hypothesis is defined as follows

$$\begin{aligned} w^k &\triangleq w^i + w^j \\ x^k &\triangleq \frac{w^i x^i + w^j x^j}{w^i + w^j} \\ P^k &\triangleq \frac{w^i P^i + w^j P^j}{w^i + w^j} + (x^i - x^j)(x^i - x^j)' \end{aligned} \tag{4.135}$$

- **Capping**: If after the pruning and the merging procedures the number of hypotheses is still too large, then only the $\nu_{\max}$ most relevant hypotheses are kept in the PHD. More precisely, suppose that the pruning and merging procedure reduces the number of hypotheses from ν to $\nu' < \nu$, but still is $\nu' > \nu_{\max}$. Suppose to have ordered the weights from the bigger, w^1, to the smallest, $w^{\nu'}$. Then, the capping procedure defines the post-processed PHD as

$$D(x) \triangleq \sum_{i=1}^{\nu_{\max}} w^i \mathcal{N}(x; x^i, P^i) \text{ if } \nu' > \nu_{\max}. \tag{4.136}$$

  On the other hand, if $\nu' \leq \nu_{\max}$ then, naturally, all hyphotheses are kept in the PHD, i.e.

$$D(x) \triangleq \sum_{i=1}^{\nu'} w^i \mathcal{N}(x; x^i, P^i) \text{ if } \nu' \leq \nu_{\max} \tag{4.137}$$

4.5.5 Estimate extraction

After the post-processing of the corrected PHD, one can extract the estimate $\hat{\mathsf{x}}_k$ of the actual object set X_k. The estimation procedure, which is an heuristic (i.e. is not Bayes-optimal) that resembles a sort of MAP estimation, operates as follows:

- **step 1**: Estimate the actual number of objects N_k present in the scene as

$$\begin{aligned} \hat{N}_k &\triangleq \mathbb{E}_{k|k}[|\mathsf{X}_k|] \triangleq \int D_{k|k}(x) \, \mathrm{d}x \\ &= \int \left(\sum_{i=1}^{\nu_{k|k}} w^i_{k|k} \mathcal{N}(x; x^i_{k|k}, P^i_{k|k}) \right) \mathrm{d}x \\ &= \sum_{i=1}^{\nu_{k|k}} w^i_{k|k} \underbrace{\int \mathcal{N}(x; x^i_{k|k}, P^i_{k|k}) \, \mathrm{d}x}_{=1} = \sum_{i=1}^{\nu_{k|k}} w^i_{k|k} \end{aligned} \tag{4.138}$$

- **step 2**: Given $\hat{N}_k$, define the estimate as the locations of the $\hat{N}_k$ largest peaks of the corrected PHD $D_{k|k}(\cdot)$. Due to the Gaussian representation, if the weights are ordered from the bigger to the smaller then such locations are $x^1_{k|k}, \ldots, x^{\hat{N}_k}_{k|k}$, then the estimate is

$$\hat{\mathsf{x}}_k = \{x^i_{k|k}\}_{i=1}^{\hat{N}_{k|k}}. \tag{4.139}$$

Note that the Gaussian representation provides a natural representation of the uncertantly affecting the single-object estimates $x^i_{k|k}$, which is the relative covariance matrix $P^i_{k|k}$.

Chapter 5

Extended object PHD filter theory

5.1 Summary

In this chapter the PHD filter for extended object is derived. As mentioned before, the unique difference between the standard PHD filter and the extended object PHD filter is in the measurement model, whereas the extended object PHD filter does not consider the single-measure semplification. Due to this fact, the corrected PHD computed by the extended object PHD gets an expression that is more complex than the expression of the corrected PHD computed by the standard object PHD. In particular, the new corrected PHD now depends on the partitions of the set of measures and this opens a new clustering problem. The chapter is structured as follows

- in the first part some essential concepts about the partitions of a finite sets are briefly discussed;

- then the measurement model for extended object is introduced in its exact definition and in its approximation, called *approximate Poisson body* (APB) model, that permits to obtain closed formulae;

- finally the new PHD corrector is derived and, after that, also the relative Gaussian mixture implementation is discussed. Moreover, a simple and effective clustering algorithm is given.

5.2 Partition of a finite set

5.2.1 Definition

Definition 6. Let $\mathsf{x} = \{x_1, \dots, x_\eta\}$ be a finite set on $\mathbb{R}^n$. A *partition* $\mathcal{P}$ of the finite set x is a set of the form

$$\mathcal{P} = \{\mathsf{w}_1, \mathsf{w}_2, \dots, \mathsf{w}_{|\mathcal{P}|}\} \tag{5.1}$$

where w_1, w_2, ..., $\mathsf{w}_{|\mathcal{P}|}$ are subsets of x, called *cells* of $\mathcal{P}$, such that:

1. every cell is not empty, i.e. $\mathsf{w}_p \neq \varnothing$ for all $p = 1, 2, \dots, |\mathcal{P}|$;
2. every cell is disjoint from the others, i.e. $\mathsf{w}_i \cap \mathsf{w}_j = \varnothing$ if $i \neq j$;
3. the union of the cells gives the starting set x, i.e. $\bigcup_{p=1}^{|\mathcal{P}|} \mathsf{w}_p = \mathsf{x}$.

Given a finite set $\mathsf{x} = \{x_1, \dots, x_\eta\}$, there are two particular partitions of x: the minimal partition, wich is composed by only 1 cell $\mathsf{w}_1 \triangleq \mathsf{x}$

$$\mathcal{P} \triangleq \{\mathsf{x}\} \tag{5.2}$$

and the maximal partition, which is composed by $|\mathsf{x}| = \eta$ cells $\mathsf{w}_1 \triangleq \{x_1\}$, ..., $\mathsf{w}_\eta \triangleq \{x_\eta\}$

$$\mathcal{P} \triangleq \{\{x_1\}, \dots, \{x_\eta\}\} \tag{5.3}$$

The minimal partition is the partition of x with smallest cardinality (minimum number of different cells), while the maximal partition is the partition of x with greatest cardinality (maximum number of different cells), thus in general hold for a generic partition $\mathcal{P}$

$$1 \leq |\mathcal{P}| \leq |\mathsf{x}| \tag{5.4}$$

the intuition suggests that there is only one minimal partition and only one maximal partition. On the other hand, it is possible to find different partitions with a cardinality that it is not maximal or minimal.

Two relevant problems regarding the partition of a finite set are: 1) given a finite set, count the number of its possible partitions; 2) given a finite set, find a systematic way to list its partitions. These two problems will be addressed in what follows.

5.2.2 Number of partitions of a finite set

The number of different partitions that is possible construct from a finite set x is given by the so called *Bell number* $B_{|\mathsf{x}|}$ of order $|\mathsf{x}|$. Such number is defined by the recursion

$$B_{i+1} \triangleq \sum_{j=0}^{i} \binom{i}{j} B_j \tag{5.5}$$

initialised by $B_1 \triangleq 1$.

Observation 1. The function $i \mapsto B_i$ grows rapidly, for example the first eight Bell numbers are

$$\begin{array}{llll} B_1 = 1 & B_2 = 2 & B_3 = 5 & B_4 = 15 \\ B_5 = 52 & B_6 = 203 & B_7 = 877 & B_8 = 4140 \end{array} . \tag{5.6}$$

In comparision, the first eight outcomes of the exponential function $i \mapsto 2^i$ are

$$\begin{array}{llll} 2^1 = 2 & 2^2 = 4 & 2^3 = 8 & 2^4 = 16 \\ 2^5 = 32 & 2^6 = 64 & 2^7 = 128 & 2^8 = 256 \end{array} \tag{5.7}$$

while the first eight outcome of the factorial function $i \mapsto i!$ are

$$\begin{array}{llll} 1! = 1 & 2! = 2 & 3! = 6 & 4! = 24 \\ 5! = 120 & 6! = 720 & 7! = 5040 & 8! = 40320 \end{array} . \tag{5.8}$$

A more convenient way to compute the Bell numbers is the following

$$B_i = \sum_{j=1}^{i} S_{i,j} \tag{5.9}$$

where $S_{i,j}$ is the so called *Stirling number of the second kind* of order i, j, and is given by

$$S_{i,j} \triangleq \frac{1}{j!} \sum_{\eta=0}^{j} (-1)^{\eta} \binom{j}{\eta} (j-\eta)^i. \tag{5.10}$$

The Stirling number $S_{i,j}$ counts the number of different partitions with $|\mathcal{P}| = j$ cells that is possible to construct from a finite set with $|\mathsf{x}| = i$ elements. Thus, equation (12) states simply that the total number of partitions is the sum of the number of partitions with 1 cell (which is one according to the intuition), the number of partitions with 2 cells, . . . , the number of partitions with i cells (which is one according to the intuition).

One can show that the Stirling numbers satisfy the recursion

$$S_{i+1,j} = S_{i,j-1} + j\, S_{i,j}. \tag{5.11}$$

As a consequence, the Bell numbers can be computed also with the following recursion

$$B_{i+1} = B_i + \sum_{j=1}^{i} j\, S_{i,j}. \tag{5.12}$$

To see why, consider

$$\begin{aligned} B_{i+1} &= \sum_{j=1}^{i+1} S_{i+1,j} = \sum_{j=1}^{i+1} \left(S_{i,j-1} + j\, S_{i,j}\right) \\ &= \sum_{j=1}^{i+1} S_{i,j-1} + \sum_{j=1}^{i+1} j\, S_{i,j} \end{aligned}. \tag{5.13}$$

Now, by defining $S_{i,0} \triangleq 0$ for all i ($\equiv$ the empty set is not a partition of a finite set) and by defining $S_{i,j} \triangleq 0$ for all $j > i$ ($\equiv$ the maximal cardinality of a partition is the cardinality of the considered finite set), it holds that

$$B_{i+1} = \sum_{j=2}^{i+1} S_{i,j-1} + \sum_{j=1}^{i} j\, S_{i,j} = \sum_{j=1}^{i} S_{i,j} + \sum_{j=1}^{i} j\, S_{i,j} = B_i + \sum_{j=1}^{i} j\, S_{i,j} \tag{5.14}$$

5.2.3 Listing partitions

There is a simple recursive procedure to list every partition of a finite set $\mathsf{x} = \{x_1, \ldots, x_\eta\}$. Before tackling the problem, consider the simpler sub-problem of finding all partitions of a generic subset $\mathsf{x}'' = \{x_1, \ldots, x_{|\mathsf{x}''|}\} \subset \mathsf{x}$ given the partitions of the "1 singleton predecessor" $\mathsf{x}' \subset \mathsf{x}''$ where $|\mathsf{x}'| = |\mathsf{x}''| - 1$.

Such sub-problem is resolved by the following two-step procedure:

- **step 1**: for each partition $\mathcal{P}'$ of x' append the singleton cell $\mathsf{w}'' \triangleq \{x_j\}$ to get the partitions of x''

$$\mathcal{P}_1'' \triangleq \mathcal{P}' \cup \mathsf{w}'' = \mathcal{P}' \cup \{x_j\} \tag{5.15}$$

- **step 2**: for each partition $\mathcal{P}'$ of x' and each cell $\mathsf{w}' \in \mathcal{P}'$ of the partition $\mathcal{P}'$ considered, replace the cell w' with the new cell $\mathsf{w}' \cup \mathsf{w}'' = \mathsf{w}' \cup \{x_j\}$ to get the remaing partitions of x''

$$\mathcal{P}_2'' \triangleq (\mathcal{P}' \setminus \mathsf{w}') \cup (\mathsf{w}' \cup \mathsf{w}'') = (\mathcal{P}' \setminus \mathsf{w}') \cup (\mathsf{w}' \cup \{x_j\}) \tag{5.16}$$

It is clear that the step 1 and step 2 in general produce different partitions of the new finite set x''. Moreover, step 1 and step 2 exhausts all of the partitions of the new finite set x'' because they result in exactly $B_{|\mathsf{x}''|}$ different partitions of x'', which is the total number of partition of x''.

Theorem 7. Step 1 and step 2 produces all the $B_{|\mathsf{x}''|}$ partitions of the finite set $\mathsf{x}'' \triangleq \mathsf{x}' \cup \{x_j\}$.

PROOF

To see why this fact holds, consider the number of partition generated by step 1 and step 2

$$B_{\text{step 1,step 2}} \triangleq B_{\text{step 1}} + B_{\text{step 2}} \tag{5.17}$$

where $B_{\text{step 1}}$ is the number of partitions generated by step 1 and $B_{\text{step 2}}$ is the number of partitions generated by step 2.

Trivially, the number of partitions generated by step 1 is equal to the number of the given partions, thus

$$B_{\text{step 1}} = B_{|\mathsf{x}'|} \tag{5.18}$$

On the other hand, finding the number of partitions generated by step 2 is more complicated.

Consider the minimal partition of the given set of partitions. Such partition contains only 1 cell (which is $\mathsf{w}' = \mathsf{x}'$) and step 2 produces for this partition 1 new partition. Thus let

$$B_{\text{step 2,1 cell}} \triangleq 1 \tag{5.19}$$

be the number of new partition generated by step 2 in this first case. Now consider the partitions with 2 cells of the given set of partitions. For each of theese partitions, step 2 produces 2 new partitions. Since there are $S_{|\mathsf{x}'|,2}$ different partitions of x' with 2 cells, the total number of new partitions generated by step 2 in this case is

$$B_{\text{step 2,2 cells}} \triangleq 2\, S_{|\mathsf{x}'|,2} \tag{5.20}$$

With the same reasoning, consider the partitions with 3 cells of the given set of partitions. For each of these partitions, step 2 produces 3 new partitions. There are $S_{|\mathsf{x}'|,3}$ different partitions of x' with 3 cells, so

$$B_{\text{step 2,3 cells}} \triangleq 3\, S_{|\mathsf{x}'|,3} \tag{5.21}$$

At this point it is clear that step 2, when considering partitions with j cells of the given set of partitions, generates

$$B_{\text{step 2},j \text{ cells}} \triangleq j\, S_{|\mathsf{x}'|,j} \tag{5.22}$$

new partitions. The total number of new partitions generated by step 2 is consequently

$$B_{\text{step 2}} = \sum_{j=1}^{|\mathsf{x}'|} B_{\text{step 2},j \text{ cells}} = \sum_{j=1}^{|\mathsf{x}'|} j\, S_{|\mathsf{x}'|,j} \tag{5.23}$$

where it is noted that $S_{|\mathsf{x}'|,1} = 1$ ($\equiv$ there is only one maximal partition), thus $B_{\text{step 2,1 cell}} = 1\, S_{|\mathsf{x}'|,1}$. As a consequence,

$$B_{\text{step 1,step 2}} = B_{|\mathsf{x}'|} + \sum_{j=1}^{|\mathsf{x}'|} j\, S_{|\mathsf{x}'|,j} \tag{5.24}$$

from which follows that

$$B_{\text{step 1,step 2}} = B_{|\mathsf{x}'|+1} = B_{|\mathsf{x}''|} \tag{5.25}$$

so, as claimed, step 1 and step 2 exhaust all the partitions of the new finite set x''. □

5.3 Measurement models for extended objects

5.3.1 Single extended object

The key concept used to generalize the standard model to extended objects is that an extended object, if detected, produces a finite set of measurements scattered around its surface. A single measurement can be seen as the result of the detection of a single reflecting point X_k^e, which behaves like a point-object in the standard measurement model, belonging to the edge of the extended object surface.

Thus, if X_k is the state of a single extended object (that is the centroid of the object) then the extended object is modelled as the collection of reflection points

$$X_k + X_k^1, \ldots, X_k + X_k^{E(X_k)} \tag{5.26}$$

disposed around the centroid X_k. Note that the number of reflection points $E(X)$ can vary during the time with X_k. Abbreviate the probability of detection of the reflection point $X + X^e$ as

$$p_{\mathrm{D}}^e(X_k) \triangleq p_{\mathrm{D}}(X_k + X_k^e) \qquad \forall\, e = 1, 2, \ldots, E(X_k) \tag{5.27}$$

and in the same way abbreviate the likelihood to observe a measurement Y_k if the reflection point considered is $X_k + X_k^e$ as

$$\ell_k^e(Y_k|X_k) \triangleq \ell_k(Y_k|X_k + X_k^e) \qquad \forall\, e = 1, 2, \ldots, E(X_k). \tag{5.28}$$

Then, the set of measurements generated by an extended object X_k is modelled as

$$\mathsf{h}(X_k) \triangleq \bigcup_{e=1}^{E(X_k)} \mathsf{h}^e(X_k) \tag{5.29}$$

where $\mathsf{h}^e(X_k)$ is a Bernoulli RFS with parameters $p_{\mathrm{D}}^e(X_k)$ and $\ell_k^e(\cdot|X_k)$. Assuming independence between Bernoulli components, the PGFL of the set of detections generated by the extended object X is

$$G_k^{\mathrm{D}}[h|X_k] \triangleq \prod_{e=1}^{E(X_k)} \Big(1 - p_{\mathrm{D}}^e(X_k) + \tilde{\ell}_k^e[h|X_k]\Big) \tag{5.30}$$

where $\tilde{\ell}_k^e \triangleq p_{\mathrm{D}}^e\, \ell_k^e$.

In brief, the result is that the considered model assumes that an extended object is the collection of multiple reflection points and that every reflection point gives rise to a Bernoulli RFS of detections (likewise the state of a point-object in the standard measurement model).

By merging the detection set with clutter (assumed to be Poisson), turns out the complete measurement model for a single extended object

$$\mathsf{Y}_k = \mathsf{h}(X_k) \cup \mathsf{C}_k = \left(\bigcup_{e=1}^{E(X)} \mathsf{h}^e(X_k)\right) \cup \mathsf{C}_k \tag{5.31}$$

in conclusion, the PGFL of the set of measurements in this case is

$$\begin{aligned} L_k[h|X_k] &= G_k^{\mathrm{D}}[h|X_k]\, G_k^{\mathrm{C}}[h] \\ &= \left[\prod_{e=1}^{E(X_k)} \Big(1 - p_{\mathrm{D}}^e(X_k) + \tilde{\ell}_k^e(h|X_k)\Big)\right] \exp(I_{\mathrm{C}}[h-1]) \end{aligned}. \tag{5.32}$$

5.3.2 Multiple extended objects

Now assume that multiple extended objects are present in the scene at the same time step k. In this case, for every extended object in X_k the single extended object model holds, so that

$$\mathsf{h}(\mathsf{X}_k) = \bigcup_{X \in \mathsf{X}_k} \mathsf{h}(X) = \bigcup_{X \in \mathsf{X}_k} \bigcup_{e=1}^{E(X)} \mathsf{h}^e(X) \tag{5.33}$$

Hence, the general measurement model for extended objects is

$$\mathsf{Y}_k = \mathsf{h}(\mathsf{X}_k) \cup \mathsf{C}_k = \left(\bigcup_{X \in \mathsf{X}_k} \bigcup_{e=1}^{E(X)} \mathsf{h}^e(X) \right) \cup \mathsf{C}_k \tag{5.34}$$

and, assuming that the clutter is Poisson, the relative PGFL is

$$L_k[h|\mathsf{X}_k] = G_k^{\mathrm{D}}[h|\mathsf{X}_k]\, G_k^{\mathrm{C}}[h] = \left[\prod_{e=1}^{E} \left(1 - p_{\mathrm{D}}^e + \tilde{\ell}_k^e[h] \right) \right]^{\mathsf{X}_k} \exp(I_{\mathrm{C}}[h-1]). \tag{5.35}$$

5.3.3 Poisson approximation

One can show that a multi-Bernoulli RFS composed by a great number of identical Bernoulli components with small probability of existence can be well approximated by a Poisson RFS.

Due to this fact, the single extended object multi-Bernoulli PGFL can be replaced by the simpler Poisson PGFL

$$G_k^{\mathrm{D}}[h|X_k] = \exp\left(I_{\mathrm{D}}[h-1|X_k] \right) \tag{5.36}$$

if the tracked extended object satisfies the following properties:

- every reflection point e is characterized by similar parameters p_{D}^e, $\tilde{\ell}_k$;
- the probability of detection p_{D}^e is small;
- the total number $E(X_k)$ of reflection points is big.

The Poisson model has one remarkable drawback: by considering the Poisson model (51), and assuming $I_{\mathrm{D}}(\cdot|X_k) > 0$, it holds that

$$\mathbb{P}(\mathsf{Y}_k \neq \varnothing|X_k) = 1 - \exp\left(-I_{\mathrm{D}}[1|X_k] \right) > 0 \tag{5.37}$$

which means that the probability of the event $\mathsf{Y}_k \neq \varnothing$ given X_k cannot be exactly zero. In other words, the Poisson approximation cannot represent the situation where an extended object does not produce any measure (i.e. the extended object is completely occluted). This limitation doesn't occour with the multi-Bernoulli model (45), in fact for the multi-Bernoulli it holds that

$$\mathbb{P}(\mathsf{Y}_k \neq \varnothing|X_k) = 1 - \prod_{e=1}^{E(X_k)} \left(1 - p_{\mathrm{D}}^e(X_k) \right) \tag{5.38}$$

so that $\mathsf{Y}_k \neq \varnothing$ given X_k can be a zero-probability event if every reflection point is not detected almost surely, that is $p_{\mathrm{D}}^{e}(X_k) = 0$ for all $e = 1, 2, \ldots, E(X_k)$.

In order to resolve this representation problem, define the probability of the event 'the extended object is in X_k' as $\mathring{p}_{\mathrm{D}}(X_k)$ and consequently define the 'corrected Poisson' PGFL of the set of detections as

$$\mathring{G}_k^{\mathrm{D}}[h|X_k] = 1 - \mathring{p}_{\mathrm{D}}(X_k) + \mathring{p}_{\mathrm{D}}(X_k) \exp\left(I_{\mathrm{D}}[h-1|X_k]\right) \tag{5.39}$$

now the correct model defines the probability of the event $\mathsf{Y}_k \neq \varnothing$ given X_k as

$$\mathbb{P}(\mathsf{Y}_k \neq \varnothing|X_k) = \mathring{p}_{\mathrm{D}}(X_k)\left(1 - \exp\left(-I_{\mathrm{D}}[1|X_k]\right)\right) \tag{5.40}$$

which can be exactly zero (if $\mathring{p}_{\mathrm{D}}(X_k) = 0$, i.e. if the extended object is not present in X_k).

The likelihood PGFL of Y_k given X_k takes the form

$$\begin{aligned} L_k[h|X_k] &= \mathring{G}_k^{\mathrm{D}}[h|X_k]\, G_k^{\mathrm{C}}[h] \\ &= [1 - \mathring{p}_{\mathrm{D}}(X_k) + \mathring{p}_{\mathrm{D}}(X_k)\ \exp(I_{\mathrm{D}}[h-1|X_k])] \exp(I_{\mathrm{C}}[h-1]) \end{aligned}^{\textstyle\cdot} \tag{5.41}$$

Hence by assuming that every extended object satisfies the approximation hypotheses, the multi extended object likelihood PGFL is given by

$$\begin{aligned} L_k[h|\mathsf{X}_k] &= (\mathring{G}_k^{\mathrm{D}}[h])^{\mathsf{X}_k}\, G_k^{\mathrm{C}}[h] \\ &= [1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\ \exp(I_{\mathrm{D}}[h-1])]^{\mathsf{X}_k} \exp(I_{\mathrm{C}}[h-1]) \end{aligned} \tag{5.42}$$

This equation represents the so-called *approximate Poisson-body model* (APB model) for extended objects. The simplest multiobject filter for extended objects is the APB-PHD filter, which is based on the simple APB model.

5.4 PHD filter for extended objects

5.4.1 Derivation workflow

The derivation of the PHD filter for extended objects follows the same procedure as the the standard PHD filter:

- **step 1**: according to the considered motion model, define the PGFL form of the multiobject Bayes predictor;
- **step 2**: via functional differentation, extract the predicted PHD from the predicted PGFL;

- **step 3**: according to the considered measurement model, and assuming that the predicted MPDF is Poisson, define the bivariate PGFL used to represent the corrected PGFL;
- **step 4**: via functional differentation, find the expression of the corrected PGFL from the bivariate PGFL;
- **step 5**: via functional differentation, extract the corrected PHD from the corrected PGFL.

Since the standard PHD filter and the PHD filter for extended objects share the same motion model, only the corrector of the PHD filter for extended objects will be derived. In other words, step 1 and step 2 have already addressed, so that only step 3, 4, 5 will be described in what follows (assuming the measurement model for extended objects rather than point objects).

5.4.2 Bivariate PGFL for extended objects

Recall that the bivariate PGFL $F_k[\cdot,\cdot]$ used to represent the multiobject Bayes corrector is defined as follows

$$F_k[h, g] \triangleq \int h^{\mathsf{w}}\, L_k[g|\mathsf{w}]\, p_{k|k-1}(\mathsf{w})\; \mathsf{dw}. \tag{5.43}$$

According to the APB measurement model, the bivariate PGFL reduces to

$$\begin{aligned} F_k[h, g] &= \int h^{\mathsf{w}}\, [1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\, \exp(I_{\mathrm{D}}[g-1])]^{\mathsf{w}} \exp(I_{\mathrm{C}}[g-1])\, p_{k|k-1}(\mathsf{w})\; \mathsf{dw} \\ &= \exp(I_{\mathrm{C}}[g-1]) \int \{\underbrace{h\,(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\, \exp(I_{\mathrm{D}}[g-1]))}_{\triangleq \tilde{h}}\}^{\mathsf{w}} p_{k|k-1}(\mathsf{w})\; \mathsf{dw}. \\ &= \exp(I_{\mathrm{C}}[g-1])\, G_{k|k-1}[h\,(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\, \exp(I_{\mathrm{D}}[g-1]))] \end{aligned} \tag{5.44}$$

Now, in order to simplify the differentation of such PGFL, assume that the predicted PGFL $G_{k|k-1}[\cdot]$ is Poisson for some predicted intesity function $D_{k|k-1}(\cdot)$, i.e.

$$G_{k|k-1}[\tilde{h}] = \exp\left(D_{k|k-1}[\tilde{h} - 1]\right). \tag{5.45}$$

Consequently, the bivariate PFGL assumes the following Poisson form

$$\begin{aligned} F_k[h, g] &= \exp(I_{\mathrm{C}}[g-1])\, [\exp\left(D_{k|k-1}[\tilde{h} - 1]\right)]_{\tilde{h} = h\,(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\, \exp(I_{\mathrm{D}}[g-1]))} \\ &= \exp\big(\underbrace{I_{\mathrm{C}}[g-1] + D_{k|k-1}[h\,(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}}\, \exp(I_{\mathrm{D}}[g-1])) - 1]}_{\triangleq \iota[h,g]}\big) \end{aligned} \tag{5.46}$$

5.4.3 Multiobject Bayes corrector for extended objects

Recall the general PGFL form of the multiobject Bayes corrector

$$G_{k|k}[h] = \frac{\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}\Big|_{g=0}}{\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}\Big|_{h=1,g=0}}. \tag{5.47}$$

According to this expression, the corrected PGFL $G_{k|k}[\cdot]$ is given by the following procedure:

- compute the general partial functional derivative $\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}$;
- given the general expression of the functional derivative $\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}$, find the corrected PGFL.

Theorem 8. Let $F_k[h,g]$ be the APB bivariate PGFL, then

$$\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}} = F_k[h,g]\, I_{\mathrm{C}}^{\mathsf{y}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h,g] \tag{5.48}$$

where

- the notation $\mathcal{P} \boxminus \mathsf{y}$ means that $\mathcal{P}$ is a partition of y (consequently, $\mathsf{w} \in \mathcal{P}$ means that w is a cell of the partition $\mathcal{P}$);
- the cell bivariate PGFL $d_{\mathsf{w}}[\cdot,\cdot]$ is defined as follows

$$d_{\mathsf{w}}[h,g] \triangleq \begin{cases} 1 + D_{k|k-1}[h \exp(I_{\mathrm{D}}[g-1])\, \mathring{p}_{\mathrm{D}}\, \ell_y] & \text{if } \mathsf{w} = \{y\} \\ D_{k|k-1}[h \exp(I_{\mathrm{D}}[g-1])\, \mathring{p}_{\mathrm{D}}\, \ell_{\mathsf{w}}] & \text{if } |\mathsf{w}| > 1 \end{cases} \tag{5.49}$$

- the cell likelihoods are defined as follows

$$\ell_y(x) \triangleq \frac{I_{\mathrm{D}}(y|x)}{I_{\mathrm{C}}(y)} \qquad \ell_{\mathsf{w}}(x) \triangleq \prod_{y \in \mathsf{w}} \ell_y(x) \tag{5.50}$$

  Note that here y is a generic measure, while x is the given state of an extended object.

PROOF

The proof is by induction on the number of considered measurements, starting from the simple case $\mathsf{y} = \{y_1\}$.

Induction base

If $\mathsf{y} = \{y_1\}$, then due to the first chain rule, it holds that

$$\frac{\partial F_k[h,g]}{\partial_g \{y_1\}} = \frac{\mathrm{d}\exp(\iota[g])}{\mathrm{d}\iota[g]} \frac{\partial \iota[g]}{\partial \{y_1\}} \tag{5.51}$$

The first factor is trivially

$$\frac{\mathrm{d}\exp(\iota[h,g])}{\mathrm{d}\iota[h,g]} = \exp(\iota[h,g]) = F_k[h,g] \tag{5.52}$$

while the second factor is

$$\begin{aligned} \frac{\partial \iota[h,g]}{\partial \{y_1\}} &= I_{\mathrm{C}}(y_1) + D_{k|k-1}\left[h\,\mathring{p}_{\mathrm{D}}\,\exp(I_{\mathrm{D}}[g-1])\,I_{\mathrm{D}}(y_1)\right] \\ &= I_{\mathrm{C}}(y_1)\left(1 + D_{k|k-1}\left[h\,\mathring{p}_{\mathrm{D}}\,\exp(I_{\mathrm{D}}[g-1])\,\frac{I_{\mathrm{D}}(y_1)}{I_{\mathrm{C}}(y_1)}\right]\right). \\ &\triangleq I_{\mathrm{C}}(y_1)\left(1 + D_{k|k-1}\left[h\,\mathring{p}_{\mathrm{D}}\,\exp(I_{\mathrm{D}}[g-1])\,\ell_{y_1}\right]\right) \\ &\triangleq I_{\mathrm{C}}(y_1)\,d_{\{y_1\}}[h,g] \end{aligned} \tag{5.53}$$

Now note that the singleton $\mathsf{y} = \{y_1\}$ admits only one partition, the trivial partition $\mathcal{P} = \{\{y_1\}\}$, which has only one cell, $\mathsf{w} = \{y_1\}$. Consequently, holds

$$\sum_{\mathcal{P} \boxminus \{y_1\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h,g] = \prod_{\mathsf{w} \in \{\{y_1\}\}} d_{\mathsf{w}}[h,g] = d_{\{y_1\}}[h,g] \tag{5.54}$$

consequently, it turns out the claimed formula for the special case $\mathsf{y} = \{y_1\}$

$$\frac{\partial F_k[h,g]}{\partial_g \{y_1\}} = F_k[h,g]\,I_{\mathrm{C}}^{\{y_1\}} \sum_{\mathcal{P} \boxminus \{y_1\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h,g] \tag{5.55}$$

so that the base of the induction is proved.

Induction step

assume that the claimed equation holds for $\mathsf{y} = \{y_1, \ldots, y_m\}$ with an arbitrary number $m > 1$ of measurements

$$\frac{\partial F_k[h,g]}{\partial_g \{y_1, \ldots, y_m\}} = F_k[h,g]\,I_{\mathrm{C}}^{\{y_1,\ldots,y_m\}} \sum_{\mathcal{P} \boxminus \{y_1,\ldots,y_m\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h,g]. \tag{5.56}$$

The objective is to show that this relation implies

$$\frac{\partial F_k[h,g]}{\partial_g \{y_1, \ldots, y_{m+1}\}} = F_k[h,g]\,I_{\mathrm{C}}^{\{y_1,\ldots,y_{m+1}\}} \sum_{\mathcal{P} \boxminus \{y_1,\ldots,y_{m+1}\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h,g]. \tag{5.57}$$

In order to do that, start by the trival observation

$$\{y_1, \dots, y_{m+1}\} = \{y_1, \dots, y_m\} \cup \{y_{m+1}\} \tag{5.58}$$

from which follows that

$$\begin{aligned} \frac{\partial F_k[h,g]}{\partial_g\{y_1,\dots,y_{m+1}\}} &= \frac{\partial F_k[h,g]}{\partial_g(\{y_1,\dots,y_m\}\cup\{y_{m+1}\})} = \frac{\partial}{\partial_g\,\{y_{m+1}\}}\frac{\partial F_k[h,g]}{\partial_g\{y_1,\dots,y_m\}} \\ &= \frac{\partial}{\partial_g\,\{y_{m+1}\}}\left[F_k[h,g]\,I_{\mathrm{C}}^{\{y_1,\dots,y_m\}}\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]\right]. \\ &= I_{\mathrm{C}}^{\{y_1,\dots,y_m\}}\frac{\partial}{\partial_g\,\{y_{m+1}\}}\left[F_k[h,g]\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]\right] \end{aligned} \tag{5.59}$$

For the product rule, the functional derivative splits into the sum of two different terms

$$\begin{aligned} \frac{\partial F_k[h,g]}{\partial_g\{y_1,\dots,y_{m+1}\}} = I_{\mathrm{C}}^{\{y_1,\dots,y_m\}}\Bigg[& \underbrace{\frac{\partial F_k[h,g]}{\partial_g\,\{y_{m+1}\}}\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]}_{\triangleq A[h,g]} \\ & \underbrace{+\,F_k[h,g]\frac{\partial}{\partial_g\,\{y_{m+1}\}}\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]}_{\triangleq B[h,g]}\Bigg]. \end{aligned} \tag{5.60}$$

The first term is

$$\begin{aligned} A[h,g] &= \big(F_k[h,g]I_{\mathrm{C}}(y_{m+1})d_{y_{m+1}}[h,g]\big)\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g] \\ &= F_k[h,g]\,I_{\mathrm{C}}(y_{m+1})\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\left(d_{y_{m+1}}[h,g]\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]\right) \\ &= F_k[h,g]\,I_{\mathrm{C}}(y_{m+1})\sum_{\mathcal{P}\boxminus\{y_1,\dots,y_m\}}\prod_{\mathsf{w}\in\mathcal{P}\cup\{y_{m+1}\}} d_{\mathsf{w}}[h,g] \end{aligned} \tag{5.61}$$

while the second one is

$$B[h,g] = F_k[h,g] \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \frac{\partial}{\partial_g \{y_{m+1}\}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]$$
$$= F_k[h,g] \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{\partial\, d_{\mathsf{w}}[h,g]}{\partial \{y_{m+1}\}} \prod_{\substack{\mathsf{w}'\in\mathcal{P}\\ \mathsf{w}'\neq\mathsf{w}}} d_{\mathsf{w}'}[h,g] \right) .$$
$$= F_k[h,g] \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \left(\prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h,g] \sum_{\mathsf{w}\in\mathcal{P}} \frac{1}{d_{\mathsf{w}}[h,g]} \frac{\partial\, d_{\mathsf{w}}[h,g]}{\partial \{y_{m+1}\}} \right) \tag{5.62}$$

After some simple functional manipulations, it turns out that

$$\frac{\partial\, d_{\mathsf{w}}[h,g]}{\partial \{y_{m+1}\}} = \frac{\partial \left(D_{k|k-1}[h \, \exp(I_{\mathrm{D}}[g-1])\, \mathring{p}_{\mathrm{D}}\, \ell_{\mathsf{w}}]\right)}{\partial \{y_{m+1}\}}$$
$$= I_{\mathrm{C}}(y_{m+1})\, d_{\mathsf{w}\cup\{y_{m+1}\}}[h,g] \tag{5.63}$$

which yields

$$B[h,g] = F_k[h,g]\, I_{\mathrm{C}}(y_{m+1}) \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \left(\prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h,g] \sum_{\mathsf{w}\in\mathcal{P}} \frac{d_{\mathsf{w}\cup\{y_{m+1}\}}[h,g]}{d_{\mathsf{w}}[h,g]} \right)$$
$$= F_k[h,g]\, I_{\mathrm{C}}(y_{m+1}) \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \left(\sum_{\mathsf{w}\in\mathcal{P}} \prod_{\mathsf{w}'\in(\mathcal{P}\setminus\mathsf{w})\cup(\mathsf{w}\cup\{y_{m+1}\})} d_{\mathsf{w}'}[h,g] \right) \tag{5.64}$$

Consequently,

$$\frac{\partial\, F_k[h,g]}{\partial_g\{y_1,\dots,y_{m+1}\}} = I_{\mathrm{C}}^{\{y_1,\dots,y_m\}} \left[A[h,g] + B[h,g]\right]$$
$$= F_k[h,g]\, I_{\mathrm{C}}^{\{y_1,\dots,y_m\, y_{m+1}\}} \sum_{\mathcal{P} \boxminus \{y_1,\dots,y_m\}} \left(\prod_{\mathsf{w}\in\mathcal{P}\cup\{y_{m+1}\}} d_{\mathsf{w}}[h,g] \right.$$
$$\left. + \sum_{\mathsf{w}\in\mathcal{P}} \prod_{\mathsf{w}'\in(\mathcal{P}\setminus\mathsf{w})\cup(\mathsf{w}\cup\{y_{m+1}\})} d_{\mathsf{w}'}[h,g] \right) \tag{5.65}$$

Finally, recalling that the set of partitions $\mathcal{P} \boxminus \{y_1,\dots,y_{m+1}\}$ is given by

$$\underbrace{\{\mathcal{P}\cup\{y_{m+1}\}\}}_{\text{step 1}} \quad \cup \quad \underbrace{\{(\mathcal{P}\setminus\mathsf{w})\cup(\mathsf{w}\cup\{y_{m+1}\})\}}_{\text{step 2}} \qquad \forall\,\mathcal{P}\boxminus\{y_1,\dots,y_m\},\ \forall\,\mathsf{w}\in\mathcal{P} \tag{5.66}$$

lead to the following formula

$$\frac{\partial\, F_k[h,g]}{\partial_g\{y_1,\dots,y_{m+1}\}} = F_k[h,g]\, I_{\mathrm{C}}^{\{y_1,\dots,y_m\, y_{m+1}\}} \sum_{\mathcal{P}\boxminus\{y_1,\dots,y_{m+1}\}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g] \tag{5.67}$$

which closes the induction step and concludes the proof. □

Theorem 9. The corrected PGFL is given by

$$G_{k|k}[h] \triangleq \tilde{F}_k[h] \frac{\sum_{\mathcal{P}\boxminus\mathsf{y}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h]}{\sum_{\mathcal{P}'\boxminus\mathsf{y}} \prod_{\mathsf{w}'\in\mathcal{P}'} d_{\mathsf{w}'}} \tag{5.68}$$

where

$$\begin{aligned}
\tilde{F}_k[h] &\triangleq \exp(D_{k|k-1}[(h-1)\,(1-\mathring{p}_{\mathrm{D}}+\mathring{p}_{\mathrm{D}}\exp(-I_{\mathrm{D}}[1]))])\\
d_{\mathsf{w}}[h] &\triangleq \begin{cases} 1 + D_{k|k-1}[h\ \exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_y] & \text{if } \mathsf{w}=\{y\} \\ D_{k|k-1}[h\ \exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_{\mathsf{w}}] & \text{if } |\mathsf{w}|>1 \end{cases}\\
d_{\mathsf{w}} &\triangleq \begin{cases} 1 + D_{k|k-1}[\exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_y] & \text{if } \mathsf{w}=\{y\} \\ D_{k|k-1}[\exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_{\mathsf{w}}] & \text{if } |\mathsf{w}|>1 \end{cases}
\end{aligned} \tag{5.69}$$

Proof

On one hand, the numerator of the corrected PGFL is

$$\begin{aligned}
\left.\frac{\partial\, F_k[h,g]}{\partial_g \mathsf{y}}\right|_{g=0} &= \left[F_k[h,g]\, I_{\mathrm{C}}^{\mathsf{y}} \sum_{\mathcal{P}\boxminus\mathsf{y}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,g]\right]_{g=0}. \\
&= F_k[h,0]\, I_{\mathrm{C}}^{\mathsf{y}} \sum_{\mathcal{P}\boxminus\mathsf{y}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h,0]
\end{aligned} \tag{5.70}$$

The first factor is given by

$$\begin{aligned}
F_k[h,0] &= \exp(\iota[0]) \\
&= \exp(-I_{\mathrm{C}}[1] + D_{k|k-1}[h\,(1-\mathring{p}_{\mathrm{D}}+\mathring{p}_{\mathrm{D}}\exp(-I_{\mathrm{D}}[1]))-1]) \triangleq F_k[h]
\end{aligned} \tag{5.71}$$

while

$$d_{\mathsf{w}}[h,0] = \begin{cases} 1 + D_{k|k-1}[h\ \exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_y] & \text{if } \mathsf{w}=\{y\} \\ D_{k|k-1}[h\ \exp(-I_{\mathrm{D}}[1])\,\mathring{p}_{\mathrm{D}}\,\ell_{\mathsf{w}}] & \text{if } |\mathsf{w}|>1 \end{cases} \triangleq d_{\mathsf{w}}[h] \quad . \tag{5.72}$$

Hence, the numerator is

$$\left.\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}\right|_{g=0} = F_k[h]\, I_{\mathrm{C}}^{\mathsf{y}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h] \quad . \tag{5.73}$$

On the other hand, the denominator of the corrected PGFL is

$$\left.\frac{\partial F_k[h,g]}{\partial_g \mathsf{y}}\right|_{h=1,g=0} \triangleq F_k[1]\, I_{\mathrm{C}}^{\mathsf{y}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}} \tag{5.74}$$

where

$$d_{\mathsf{w}} \triangleq d_{\mathsf{w}}[1] = \begin{cases} 1 + D_{k|k-1}[\exp(-I_{\mathrm{D}}[1])\, \mathring{p}_{\mathrm{D}}\, \ell_y] & \text{if } \mathsf{w} = \{y\} \\ D_{k|k-1}[\exp(-I_{\mathrm{D}}[1])\, \mathring{p}_{\mathrm{D}}\, \ell_{\mathsf{w}}] & \text{if } |\mathsf{w}| > 1 \end{cases} \quad . \tag{5.75}$$

In conclusion, the corrected PGFL is

$$G_{k|k}[h] \triangleq \tilde{F}_k[h] \frac{\sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h]}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \tag{5.76}$$

where

$$\tilde{F}_k[h] \triangleq \frac{F_k[h]}{F_k[1]} \quad = \exp(D_{k|k-1}[(h-1)\,(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}} \exp(-I_{\mathrm{D}}[1]))]). \tag{5.77}$$

□

5.4.4 APB-PHD corrector

Theorem 10. The APB-PHD corrector, which corresponds to the APB measurement model, is given by

$$D_{k|k}(x) = \Lambda(x)\, D_{k|k-1}(x) \tag{5.78}$$

where the likelihood $\Lambda(\cdot)$ is defined as follows

$$\Lambda(x) \triangleq (1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])) + \bar{\Lambda}(\mathsf{y}|x) \tag{5.79}$$

and

$$\bar{\Lambda}(\mathsf{y}|x) \triangleq \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x]) \sum_{\mathcal{P} \boxminus \mathsf{y}} \omega_{\mathcal{P}} \left(\sum_{\mathsf{w} \in \mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}} \right) \tag{5.80}$$

with

$$\omega_{\mathcal{P}} \triangleq \frac{\prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \tag{5.81}$$

PROOF

According to the generalized multiobject calculus, the corrected PHD $D_{k|k}(\cdot)$ can be computed as the following functional derivative of the corrected PGFL $G_{k|k}[\cdot]$

$$D_{k|k}(x) = \left. \frac{\partial\, G_{k|k}[h]}{\partial\, \{x\}} \right|_{h=1} \tag{5.82}$$

thus the objective is to compute the following differentation

$$\begin{aligned}
D_{k|k}(x) &= \frac{\partial}{\partial\, \{x\}} \left[\tilde{F}_k[h] \frac{\sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h]}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \right]_{h=1} \\
&= \frac{1}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \frac{\partial}{\partial\, \{x\}} \left[\tilde{F}_k[h] \sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h] \right]_{h=1} \\
&= \frac{1}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \frac{\partial}{\partial\, \{x\}} \left[\sum_{\mathcal{P} \boxminus \mathsf{y}} \tilde{F}_k[h] \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h] \right]_{h=1} \\
&= \frac{1}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \frac{\partial}{\partial\, \{x\}} \left[\tilde{F}_k[h] \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h] \right]_{h=1} \\
&= \frac{1}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \left[\underbrace{\frac{\partial\, \tilde{F}_k[h]}{\partial\, \{x\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h]}_{\triangleq A_{\mathcal{P}}[h]} + \underbrace{\tilde{F}_k[h] \frac{\partial}{\partial\, \{x\}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}[h]}_{\triangleq B_{\mathcal{P}}[h]} \right]_{h=1}
\end{aligned} \tag{5.83}$$

In order to simplify the notations, define the functional

$$\tilde{\imath}[h] \triangleq D_{k|k-1}[(h-1)(1 - \mathring{p}_{\mathrm{D}} + \mathring{p}_{\mathrm{D}} \exp(-I_{\mathrm{D}}[1]))] \tag{5.84}$$

so that $\tilde{F}_k[h] = \exp(\tilde{\imath}[h])$ and, after some standard computations,

$$\begin{aligned}
\frac{\partial\, \tilde{F}_k[h]}{\partial\, \{x\}} &= \frac{\partial \exp(\tilde{\imath}[h])}{\partial\, \{x\}} \\
&= \tilde{F}_k[h](1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x]))\, D_{k|k-1}(x)
\end{aligned} \tag{5.85}$$

Consequently, the first term takes the form

$$
\begin{aligned}
A_{\mathcal{P}}[h] &= \tilde{F}_k[h](1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x)\exp(-I_{\mathrm{D}}[1|x]))\, D_{k|k-1}(x) \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h] \\
&= \tilde{F}_k[h](1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x)\exp(-I_{\mathrm{D}}[1|x])) \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h]\, D_{k|k-1}(x)
\end{aligned}
\tag{5.86}
$$

Now, for the second term, it holds that

$$
\frac{\partial}{\partial\{x\}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h] = \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h] \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{1}{d_{\mathsf{w}}[h]} \frac{\partial\, d_{\mathsf{w}}[h]}{\partial\{x\}} \right) \tag{5.87}
$$

and

$$
\begin{aligned}
\frac{\partial\, d_{\mathsf{w}}[h]}{\partial\{x\}} &= \frac{\partial}{\partial\,\{x\}} \left[D_{k|k-1}[h\, \exp(-I_{\mathrm{D}}[1])\, \mathring{p}_{\mathrm{D}}\, \ell_{\mathsf{w}}] \right] \\
&= \exp(-I_{\mathrm{D}}[1|x])\, \mathring{p}_{\mathrm{D}}(x)\, \ell_{\mathsf{w}}(x)\, D_{k|k-1}(x)
\end{aligned}
\tag{5.88}
$$

meaning that

$$
\begin{aligned}
\frac{\partial}{\partial\{x\}} \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h] &= \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h] \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{1}{d_{\mathsf{w}}[h]} \exp(-I_{\mathrm{D}}[1|x])\, \mathring{p}_{\mathrm{D}}(x)\, \ell_{\mathsf{w}}(x)\, D_{k|k-1}(x) \right). \\
&= \mathring{p}_{\mathrm{D}}(x)\ \exp(-I_{\mathrm{D}}[1|x]) \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h] \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}[h]} \right) D_{k|k-1}(x)
\end{aligned}
\tag{5.89}
$$

Consequently the second term gets the form

$$
B_{\mathcal{P}}[h] = \tilde{F}_k[h]\mathring{p}_{\mathrm{D}}(x)\ \exp(-I_{\mathrm{D}}[1|x]) \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h] \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}[h]} \right) D_{k|k-1}(x) \tag{5.90}
$$

Hence, it turns out that

$$
\begin{aligned}
D_{k|k}(x) &= \frac{1}{\sum_{\mathcal{P}'\boxminus\mathsf{y}} \prod_{\mathsf{w}'\in\mathcal{P}'} d_{\mathsf{w}'}} \sum_{\mathcal{P}\boxminus\mathsf{y}} \Big[A_{\mathcal{P}}[h] + B_{\mathcal{P}}[h] \Big]_{h=1} \\
&= \frac{1}{\sum_{\mathcal{P}'\boxminus\mathsf{y}} \prod_{\mathsf{w}'\in\mathcal{P}'} d_{\mathsf{w}'}} \sum_{\mathcal{P}\boxminus\mathsf{y}} \Bigg[\tilde{F}_k[h] \Bigg((1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x)\exp(-I_{\mathrm{D}}[1|x])) \prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}[h] + \\
&\qquad \mathring{p}_{\mathrm{D}}(x)\ \exp(-I_{\mathrm{D}}[1|x]) \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}[h] \left(\sum_{\mathsf{w}\in\mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}[h]} \right) \Bigg) \Bigg]_{h=1} D_{k|k-1}(x)
\end{aligned}
\tag{5.91}
$$

Now, by observing that $d_{\mathsf{w}}[1] = d_{\mathsf{w}}$ and $\tilde{F}_k[1] = 1$, the claimed formula for the corrected PHD follows

$$
\begin{aligned}
D_{k|k}(x) &= \frac{1}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}} \sum_{\mathcal{P} \boxminus \mathsf{y}} \Bigg[(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])) \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}} + \\
&\quad \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x]) \prod_{\mathsf{w}' \in \mathcal{P}} d_{\mathsf{w}'} \left(\sum_{\mathsf{w} \in \mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}} \right) \Bigg] D_{k|k-1}(x) \\
&= \Bigg[(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])) \underbrace{\frac{\sum_{\mathcal{P} \boxminus \mathsf{y}} \prod_{\mathsf{w} \in \mathcal{P}} d_{\mathsf{w}}}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}}}_{=1} + \\
&\quad \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x]) \sum_{\mathcal{P} \boxminus \mathsf{y}} \underbrace{\frac{\prod_{\mathsf{w}'' \in \mathcal{P}} d_{\mathsf{w}''}}{\sum_{\mathcal{P}' \boxminus \mathsf{y}} \prod_{\mathsf{w}' \in \mathcal{P}'} d_{\mathsf{w}'}}}_{\triangleq \omega_{\mathcal{P}}} \left(\sum_{\mathsf{w} \in \mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}} \right) \Bigg] D_{k|k-1}(x) \\
&= \Bigg[(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])) \\
&\quad \underbrace{+ \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x]) \sum_{\mathcal{P} \boxminus \mathsf{y}} \omega_{\mathcal{P}} \left(\sum_{\mathsf{w} \in \mathcal{P}} \frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}} \right)}_{\triangleq \bar{\Lambda}(\mathsf{y}|x)} \Bigg] D_{k|k-1}(x) \\
&= \Bigg[\underbrace{(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])) + \bar{\Lambda}(\mathsf{y}|x)}_{\triangleq \Lambda(x)} \Bigg] D_{k|k-1}(x)
\end{aligned}
\tag{5.92}
$$

□

5.5 Gaussian mixture implementation

5.5.1 GM-APB-PHD corrector

Theorem 11. Consider the following assumptions:

- the predicted PHD is a Gaussian mixture of the following form

$$
D_{k|k-1}(x) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1} \mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \tag{5.93}
$$

- the detection intensity $I_{\mathrm{D}}[\cdot|x]$ is Gaussian, in the sense that

$$I_{\mathrm{D}}(y|x) \triangleq \lambda_D(x)\,\ell_k(y|x) \tag{5.94}$$

where the spatial distribution $\ell_k(\cdot|x) \triangleq I_{\mathrm{D}}(y|x)/I_{\mathrm{D}}[1|x]$, assuming that the single-point measurement model is linear-Gaussian, i.e. $Y_k = C_k\,X_k + V_k$ with $V_k \sim \mathcal{N}(0, R_k)$, is the following Gaussian single-point-object likelihood

$$\ell_k(y|w) = \mathcal{N}(y; C_k\,w, R_k) \tag{5.95}$$

- the detection probability $\mathring{p}_{\mathrm{D}}(\cdot)$ satisfies for all i and for all $x \in \mathbb{R}^n$ the relationship

$$\begin{aligned}\mathring{p}_{\mathrm{D}}(x)\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) &= \mathring{p}_{\mathrm{D}}(x^i_{k|k-1})\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \\ &\triangleq \mathring{p}^i_{\mathrm{D}}\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1})\end{aligned} \tag{5.96}$$

where $\mathring{p}^i_{\mathrm{D}} \triangleq \mathring{p}_{\mathrm{D}}(x^i_{k|k-1})$.

- the detection intensity $\lambda_{\mathrm{D}}(x) \triangleq I_{\mathrm{D}}[1|x]$ satisfies for all i and for all $x \in \mathbb{R}^n$ the relationship

$$\begin{aligned}\exp\left(-\lambda_{\mathrm{D}}(x)\right)\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) &= \exp\left(-\lambda_{\mathrm{D}}(x^i_{k|k-1})\right)\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1}) \\ &\triangleq \exp\left(-\lambda^i_{\mathrm{D}}\right)\,\mathcal{N}(x; x^i_{k|k-1}, P^i_{k|k-1})\end{aligned} \tag{5.97}$$

where $\lambda^i_{\mathrm{D}} \triangleq \lambda_{\mathrm{D}}(x^i_{k|k-1})$.

Then, the corrected PHD is a Gaussian mixture of the form

$$D_{k|k}(x) = D^{\mathrm{ND}}_{k|k}(x) + \sum_{\mathcal{P} \boxminus \mathsf{y}} \sum_{\mathsf{w} \in \mathcal{P}} D^{\mathrm{D}}_{k|k}(x; \mathcal{P}, \mathsf{w}) \tag{5.98}$$

where:

- **1)** the undetected object PHD $D^{\mathrm{ND}}_{k|k}(\cdot)$ is

$$D^{\mathrm{ND}}_{k|k}(x) \triangleq \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k}\,\mathcal{N}\left(x; x^i_{k|k}, P^i_{k|k}\right) \tag{5.99}$$

where for all predicted Gaussian components i, defined

$$\mathring{p}^i_{\mathrm{eff}} \triangleq \mathring{p}^i_{\mathrm{D}}\,\exp(-\lambda^i_{\mathrm{D}}) \tag{5.100}$$

the undetected object weights are given by

$$w^i_{k|k} \triangleq \left(1 - \mathring{p}^i_{\text{eff}}\right) \, w^i_{k|k-1} \tag{5.101}$$

while the parameters of the nondetection Gaussian components are given by

$$\begin{aligned} x^i_{k|k} &\triangleq x^i_{k|k-1} \\ P^i_{k|k} &\triangleq P^i_{k|k-1} \end{aligned} \tag{5.102}$$

- **2)** the detected object PHD $D^{\text{D}}_{k|k}(\cdot;\cdot,\cdot)$ is

$$D^{\text{D}}_{k|k}(x;\mathcal{P},\mathsf{w}) \triangleq \sum_{i=1}^{\nu_{k|k-1}} w^{i,\mathcal{P},\mathsf{w}}_{k|k} \, \mathcal{N}\left(x; x^{i,\mathsf{w}}_{k|k}, P^{i,\mathsf{w}}_{k|k}\right) \tag{5.103}$$

where for all predicted Gaussian component i, $\mathcal{P}$, w, defined (the symbol $\oplus$ denotes the vertical stacking operation)

$$y_{\mathsf{w}} \triangleq \bigoplus_{y \in \mathsf{w}} y \qquad C_{\mathsf{w}} \triangleq [\underbrace{C'_k, \ldots, C'_k}_{|\mathsf{w}| \text{ times}}]' \qquad R_{\mathsf{w}} \triangleq \text{diag}(\underbrace{R_k, \ldots, R_k}_{|\mathsf{w}| \text{ times}})'. \tag{5.104}$$

The detected object weights are given by

$$w^{i,\mathcal{P},\mathsf{w}}_{k|k} \triangleq \left(\frac{\mathring{p}^i_{\text{eff}} \, \omega_{\mathcal{P}} \, \ell^i_{\mathsf{w}}}{d_{\mathsf{w}}}\right) \, w^i_{k|k-1} \tag{5.105}$$

with

$$\begin{aligned} \ell^i_{\mathsf{w}} &\triangleq \frac{\left(\lambda^i_{\text{D}}\right)^{|\mathsf{w}|} \mathcal{N}\left(y_{\mathsf{w}}; \hat{y}^{i,\mathsf{w}}_{k|k-1}, S^{i,\mathsf{w}}_k\right)}{I^{\mathsf{w}}_{\text{C}}} \\ d_{\mathsf{w}} &= \delta_{|\mathsf{w}|,1} + \frac{1}{I^{\mathsf{w}}_{\text{C}}} \sum_{\iota=1}^{\nu_{k|k-1}} \mathring{p}^{\iota}_{\text{eff}} \, (\lambda^{\iota}_{\text{D}})^{|\mathsf{w}|} \, \mathcal{N}(y_{\mathsf{w}}; \hat{y}^{\iota,\mathsf{w}}_{k|k-1}, S^{\iota,\mathsf{w}}_k) \, w^{\iota}_{k|k-1} \end{aligned} \tag{5.106}$$

and $\hat{y}^{i,\mathsf{w}}_{k|k-1}, S^{i,\mathsf{w}}_k$ given by the following Kalman predictor

$$\begin{aligned} S^{i,\mathsf{w}}_k &\triangleq R_{\mathsf{w}} + C_{\mathsf{w}} P^i_{k|k-1} C'_{\mathsf{w}} \\ \hat{y}^{i,\mathsf{w}}_{k|k-1} &\triangleq C_{\mathsf{w}} \, x^i_{k|k-1} \end{aligned} \tag{5.107}$$

while the parameters of the detection Gaussian components are given by the following Kalman corrector

$$\begin{aligned} L^i_{\mathsf{w}} &\triangleq P^i_{k|k-1} \, C'_{\mathsf{w}} \left(S^{i,\mathsf{w}}_k\right)^{-1} \\ P^{i,\mathsf{w}}_{k|k} &\triangleq (I - L^i_{\mathsf{w}} \, C_{\mathsf{w}}) P^i_{k|k-1} \\ x^{i,\mathsf{w}}_{k|k} &\triangleq x^i_{k|k-1} + L^i_{\mathsf{w}} \left(y_{\mathsf{w}} - \hat{y}^{i,\mathsf{w}}_{k|k-1}\right) \end{aligned} \tag{5.108}$$

Proof

The corrected PHD is

$$\begin{aligned} D_{k|k}(x) &= \Lambda(x)\, D_{k|k-1}(x) \\ &= \underbrace{\left(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp(-I_{\mathrm{D}}[1|x])\right) D_{k|k-1}(x)}_{\triangleq D^{\mathrm{ND}}_{k|k}(x)} + \underbrace{\bar{\Lambda}(\mathsf{y}|x)\, D_{k|k-1}(x)}_{\triangleq D^{\mathrm{D}}_{k|k}(x)} \end{aligned} \tag{5.109}$$

so the objective is to show that the undetected object PHD $D^{\mathrm{ND}}_{k|k}(\cdot)$ is given by (1.99) and that the detected object PHD $D^{\mathrm{D}}_{k|k}(\cdot) = \sum_{\mathcal{P}} \sum_{\mathsf{w}} D^{\mathrm{D}}_{k|k}(\cdot; \mathcal{P}, \mathsf{w})$ where $D^{\mathrm{D}}_{k|k}(\cdot; \mathcal{P}, \mathsf{w})$ is given by (1.103).

Undetected object PHD

Recalling the Gaussian form of the predicted PHD and by defining $\lambda_{\mathrm{D}}(\cdot) \triangleq I_{\mathrm{D}}[1|\cdot]$, it follows that

$$D^{\mathrm{ND}}_{k|k}(x) = \sum_{i=1}^{\nu_{k|k-1}} \left(1 - \mathring{p}_{\mathrm{D}}(x) + \mathring{p}_{\mathrm{D}}(x) \exp\left(-\lambda_{\mathrm{D}}(x)\right)\right) w^i_{k|k-1} \mathcal{N}\left(x; x^i_{k|k-1}, P^i_{k|k-1}\right) \tag{5.110}$$

Now, thanks to the previous assumptions,

$$\begin{aligned} D^{\mathrm{ND}}_{k|k}(x) &\triangleq \sum_{i=1}^{\nu_{k|k-1}} \left(1 - \mathring{p}^i_{\mathrm{D}} + \mathring{p}^i_{\mathrm{D}} \exp\left(-\lambda^i_{\mathrm{D}}\right)\right) w^i_{k|k-1} \mathcal{N}\left(x; x^i_{k|k-1}, P^i_{k|k-1}\right) \\ &=\triangleq \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k} \mathcal{N}\left(x; x^i_{k|k}, P^i_{k|k}\right) \end{aligned} \tag{5.111}$$

where for all i

$$\begin{aligned} w^i_{k|k} &\triangleq (1 - \mathring{p}^i_{\mathrm{eff}})\, w^i_{k|k-1} \\ P^i_{k|k} &\triangleq P^i_{k|k-1} \\ x^i_{k|k} &\triangleq x^i_{k|k-1} \end{aligned} \quad . \tag{5.112}$$

Detected object PHD

From the Gaussian form of the predicted PHD. it holds that

$$\begin{aligned}D^{\mathrm{D}}_{k|k}(x) &= \mathring{p}_{\mathrm{D}}(x)\exp\left(-\lambda_{\mathrm{D}}(x)\right)\sum_{\mathcal{P}\boxminus \mathsf{y}}\omega_{\mathcal{P}}\left(\sum_{\mathsf{w}\in\mathcal{P}}\frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}}\right)\\ &\quad\times\left(\sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1}\mathcal{N}\left(x;x^i_{k|k-1};P^i_{k|k-1}\right)\right)\\ &= \sum_{\mathcal{P}\boxminus \mathsf{y}}\sum_{\mathsf{w}\in\mathcal{P}}\underbrace{\sum_{i=1}^{\nu_{k|k-1}}\mathring{p}_{\mathrm{D}}(x)\ \exp\left(-\lambda_{\mathrm{D}}(x)\right)\ \omega_{\mathcal{P}}\frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}}w^i_{k|k-1}\mathcal{N}(x;x^i_{k|k-1},P^i_{k|k-1})}_{\triangleq D^{\mathrm{D}}_{k|k}(x;\mathcal{P},\mathsf{w})}\end{aligned} \tag{5.113}$$

so the objective is to show that equivalence $D^{\mathrm{D}}_{k|k}(\cdot;\cdot,\cdot)$ is given by (1.103). Due to the previous assumptions,

$$\begin{aligned}D^{\mathrm{D}}_{k|k}(x;\mathcal{P},\mathsf{w}) &\triangleq \sum_{i=1}^{\nu_{k|k-1}}\mathring{p}^i_{\mathrm{D}}\ \exp\left(-\lambda^i_{\mathrm{D}}\right)\ \omega_{\mathcal{P}}\frac{\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}}w^i_{k|k-1}\mathcal{N}(x;x^i_{k|k-1},P^i_{k|k-1})\\ &= \sum_{i=1}^{\nu_{k|k-1}}\frac{\mathring{p}^i_{\mathrm{eff}}\,\omega_{\mathcal{P}}\ell_{\mathsf{w}}(x)}{d_{\mathsf{w}}}w^i_{k|k-1}\mathcal{N}(x;x^i_{k|k-1},P^i_{k|k-1})\end{aligned} \tag{5.114}$$

now focus on the cell likelihood $\ell_{\mathsf{w}}(\cdot)$. It holds that

$$\ell_{\mathsf{w}}(x) \triangleq \prod_{y\in\mathsf{w}}\ell_y(x) \triangleq \prod_{y\in\mathsf{w}}\frac{I_{\mathrm{D}}(y|x)}{I_{\mathrm{C}}(y)} = \frac{\left(\lambda_{\mathrm{D}}(x)\right)^{|\mathsf{w}|}\ \mathcal{N}(y_{\mathsf{w}};C_{\mathsf{w}}x,R_{\mathsf{w}})}{I^{\mathsf{w}}_{\mathrm{C}}} \tag{5.115}$$

where are defined

$$y_{\mathsf{w}} \triangleq \bigoplus_{y\in\mathsf{w}} y \qquad C_{\mathsf{w}} \triangleq [\underbrace{C'_k,\ldots,C'_k}_{|\mathsf{w}|\ \text{times}}]' \qquad R_{\mathsf{w}} \triangleq \mathrm{diag}(\underbrace{R_k,\ldots,R_k}_{|\mathsf{w}|\ \text{times}})' \tag{5.116}$$

Consequently,

$$D^{\mathrm{D}}_{k|k}(x;\mathcal{P},\mathsf{w}) = \sum_{i=1}^{\nu_{k|k-1}}\frac{\mathring{p}^i_{\mathrm{eff}}\,\omega_{\mathcal{P}}}{d_{\mathsf{w}}}\frac{\left(\lambda^i_{\mathrm{D}}\right)^{|\mathsf{w}|}}{I^{\mathsf{w}}_{\mathrm{C}}}w^i_{k|k-1}\mathcal{N}(y_{\mathsf{w}};C_{\mathsf{w}}x,R_{\mathsf{w}})\,\mathcal{N}(x;x^i_{k|k-1},P^i_{k|k-1}) \tag{5.117}$$

Now, for the foundamental Gaussian identity, it holds that

$$\mathcal{N}(y_{\mathsf{w}};C_{\mathsf{w}}x,R_{\mathsf{w}})\,\mathcal{N}(x;x^i_{k|k-1},P^i_{k|k-1}) = \mathcal{N}(y_{\mathsf{w}};\hat{y}^{i,\mathsf{w}}_{k|k-1},S^{i,\mathsf{w}}_k)\,\mathcal{N}(x;x^{i,\mathsf{w}}_{k|k},P^{i,\mathsf{w}}_{k|k}) \tag{5.118}$$

where $\hat{y}^{i,\mathsf{w}}_{k|k-1}, S^{i,\mathsf{w}}_k$ are given by the following Kalman predictor

$$\begin{aligned} S^{i,\mathsf{w}}_k &\triangleq R_\mathsf{w} + C_\mathsf{w} P^i_{k|k-1} C'_\mathsf{w} \\ \hat{y}^{i,\mathsf{w}}_{k|k-1} &\triangleq C_\mathsf{w}\, x^i_{k|k-1} \end{aligned} \tag{5.119}$$

while $x^{i,\mathsf{w}}_{k|k}, P^{i,\mathsf{w}}_{k|k}$ are given by the following Kalman corrector

$$\begin{aligned} L^i_\mathsf{w} &\triangleq P^i_{k|k-1}\, C'_\mathsf{w} \left(S^{i,\mathsf{w}}_k\right)^{-1} \\ P^{i,\mathsf{w}}_{k|k} &\triangleq (I - L^i_\mathsf{w}\, C_\mathsf{w}) P^i_{k|k-1} \\ x^{i,\mathsf{w}}_{k|k} &\triangleq x^i_{k|k-1} + L^i_\mathsf{w} \left(y_\mathsf{w} - \hat{y}^{i,\mathsf{w}}_{k|k-1}\right) \end{aligned} \quad . \tag{5.120}$$

Hence, it turns out that

$$\begin{aligned} D^{\mathrm{D}}_{k|k}(x; \mathcal{P}, \mathsf{w}) &\triangleq \sum_{i=1}^{\nu_{k|k-1}} \frac{\mathring{p}^i_{\mathrm{eff}}\, \omega_\mathcal{P}\, \ell^i_\mathsf{w}}{d_\mathsf{w}} w^i_{k|k-1}\, \mathcal{N}(x; x^{i,\mathsf{w}}_{k|k}, P^{i,\mathsf{w}}_{k|k}) \\ &\triangleq \sum_{i=1}^{\nu_{k|k-1}} w^{i,\mathcal{P},\mathsf{w}}_{k|k}\, \mathcal{N}(x; x^{i,\mathsf{w}}_{k|k}, P^{i,\mathsf{w}}_{k|k}) \end{aligned} \tag{5.121}$$

where

$$\ell^i_\mathsf{w} \triangleq \frac{\left(\lambda^i_{\mathrm{D}}\right)^{|\mathsf{w}|}\, \mathcal{N}(y_\mathsf{w}; \hat{y}^{i,\mathsf{w}}_{k|k-1}, S^{i,\mathsf{w}}_k)}{I^\mathsf{w}_{\mathrm{C}}} \qquad w^{i,\mathcal{P},\mathsf{w}}_{k|k} \triangleq \frac{\mathring{p}^i_{\mathrm{eff}}\, \omega_\mathcal{P}\, \ell^i_\mathsf{w}}{d_\mathsf{w}}. \tag{5.122}$$

Finally, the last step is show the explicit expression of d_w. In order to do that, recall the definition of d_w and apply the assumptions, so that

$$\begin{aligned} d_\mathsf{w} &= \begin{cases} 1 + D_{k|k-1}[\mathring{p}_{\mathrm{eff}}\, \ell_y] & \text{if } \mathsf{w} = \{y\} \\ D_{k|k-1}[\mathring{p}_{\mathrm{eff}}\, \ell_\mathsf{w}] & \text{if } |\mathsf{w}| > 1 \end{cases} = \delta_{|\mathsf{w}|,1} + D_{k|k-1}[\mathring{p}_{\mathrm{eff}}\, \ell_\mathsf{w}] \\ &= \delta_{|\mathsf{w}|,1} + \int \mathring{p}_{\mathrm{eff}}(x)\, \frac{(\lambda_{\mathrm{D}}(x))^{|\mathsf{w}|}\, \mathcal{N}\left(y_\mathsf{w}; C_\mathsf{w}\, x, R_\mathsf{w}\right)}{I^\mathsf{w}_{\mathrm{C}}} \\ &\quad \times \left(\sum_{\iota=1}^{\nu_{k|k-1}} w^\iota_{k|k-1}\, \mathcal{N}\left(x; x^\iota_{k|k-1}, P^\iota_{k|k-1}\right) \right) \mathrm{d}x \\ &= \delta_{|\mathsf{w}|,1} + \frac{1}{I^\mathsf{w}_{\mathrm{C}}} \sum_{\iota=1}^{\nu_{k|k-1}} \mathring{p}^\iota_{\mathrm{eff}}\, (\lambda^\iota_{\mathrm{D}})^{|\mathsf{w}|}\, \mathcal{N}\left(y_\mathsf{w}; \hat{y}^{\iota,\mathsf{w}}_{k|k-1}, S^{\iota,\mathsf{w}}_k\right) w^\iota_{k|k-1} \end{aligned} \tag{5.123}$$

□

5.5.2 Partitioning methods

Both the theorical corrector and its Gaussian mixture implementation are computationally intractable because they involve the computation of evry possible partition of the actual set of measure y gathered by the sensors. However, suppose that the extended objects present in the scene are well separated and that the measurements generated by every object are not mixed togheter but are tightly clustered around the corrisponding object. In this case exist one and only one natural partition $\mathcal{P}^*$ that haves a dominant weight with respect the others $B_{|\mathsf{y}|} - 1$ possible partitions.

In a more complicated situation where there is not a natural partition of the set of measures, multiple partitions can have large weights. In some sense, the number of relevant partitions grows with the entropy contained in the set of measures y.

The computational cost of the PHD corrector can be reduced to a tractable level by reducing drastically the number of partition considered. The idea is the following: instead to consider every possible partitions in the complete set $\mathcal{P} \boxminus \mathsf{y}$, consider only a small subset $\mathcal{S}_P(\mathsf{y}) \subset \mathcal{P} \boxminus \mathsf{y}$ composed by only $P(\mathsf{y}) \triangleq |\mathcal{S}_P(\mathsf{y})| \ll B_{|\mathsf{y}|} = |\mathcal{P} \boxminus \mathsf{y}|$ good partitions, where *good* means that hopefully every partition considered have large weight. The PHD corrector, then, can be approximated as follows

$$D_{k|k}(x) = D_{k|k}^{\mathrm{ND}}(x) + \sum_{\mathcal{P} \in \mathcal{S}_P(\mathsf{y})} \sum_{\mathsf{w} \in \mathcal{P}} D_{k|k}^{\mathrm{D}}(x; \mathcal{P}, \mathsf{w}). \tag{5.124}$$

There are several ways to define the set of relevant partitions $\mathcal{S}_P(\mathsf{y})$, for example common partitioning methods used in MOT are the following:

- distance partitioning;
- GLO partitioning;

5.5.3 Distance partitioning

The idea of this method is to collect in a same cell measures that are near from each other. This a idea seems reasonable if it is true that the extended objects produce their measures in the neighborhood of their centroids.

In order to understand how works distance partitioning, consider the following example.

- **example** - Consider the following set of measures (2-dimensional positions expressed in meters)

$$\mathsf{y} = \left\{ y_1 \triangleq \begin{bmatrix} 1 \\ 0 \end{bmatrix}, y_2 \triangleq \begin{bmatrix} 0 \\ 1 \end{bmatrix}, y_3 \triangleq \begin{bmatrix} 3 \\ 3 \end{bmatrix}, y_4 \triangleq \begin{bmatrix} 7 \\ 7 \end{bmatrix}, y_5 \triangleq \begin{bmatrix} 8 \\ 7 \end{bmatrix}, y_6 \triangleq \begin{bmatrix} 9 \\ 7 \end{bmatrix}, y_7 \triangleq \begin{bmatrix} 10 \\ 7 \end{bmatrix} \right\} \tag{5.125}$$

and consider the conventional Euclidean distance $\Delta_{i,j} \triangleq \sqrt{(y_i - y_j)'(y_i - y_j)}$, so the distance matrix $\Delta \triangleq [\Delta_{i,j}]$ is

$$\Delta \approx \begin{bmatrix} \cdot & 1.4 & 3.6 & 9.2 & 9.8 & 10.6 & 11.4 \\ \cdot & \cdot & 3.6 & 9.2 & 10 & 10.8 & 11.6 \\ \cdot & \cdot & \cdot & 5.6 & 6.4 & 7.2 & 8.1 \\ \cdot & \cdot & \cdot & \cdot & 1 & 2 & 3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{5.126}$$

in this case the measurements y_1, y_2 are near from each other ($\Delta_{1,2} \approx 1.4$ is small) and the measurements y_4, y_5, y_6, y_7 are near from each other ($\Delta_{4,5} = \Delta_{5,6} = \Delta_{6,7} = 1$ are "small"), so there is a natural partition

$$\mathcal{P} = \{\{y_1, y_2\}, \{y_3\}, \{y_4, y_5, y_6, y_7\}\} \tag{5.127}$$

Distance partitioning defines two measures as *near* if for a given treshold γ (which is a parameter of the algorithm) holds

$$\Delta_{i,j} \leq \gamma \tag{5.128}$$

and, based on this definition, choose the (unique, as one can show) partition that satisfies the following property: if $\Delta_{i,j} \leq \gamma$ ($\equiv y_i$ and y_j are near) then y_i and y_j are in the same cell. Note that, on the contrary, distance partitioning doesn't guarantee that if y_i and y_j are in the same cell then $\Delta_{i,j} \leq \gamma$.

As a consequence of this defining property, if $\gamma \triangleq 0$ then every measures are considered isolated and the partition generated is the maximal partition, which is in this example

$$\mathcal{P} = \{\{y_1\}, \{y_2\}, \{y_3\}, \{y_4\}, \{y_5\}, \{y_6\}, \{y_7\}\} \tag{5.129}$$

if $\gamma \triangleq \infty$ then every measures are considered near and the partition generated is the "minimal partition", which is in this example

$$\mathcal{P} = \{\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}\} \tag{5.130}$$

the specific partition is produced, for example, by the treshold $\gamma \triangleq 1.5$. Distance partitioning consider the following procedure to generate the partition.

- **step 1**: defines $\mathsf{w}_1 \triangleq \{y_1\}$ and searches the neighbohrs of y_1. If y_1 has not any neighbors then the partition w_1 is completed, otherwise includes in w_1 such neighborhs.

  In this example there is one neighbohr for y_1, that is y_2, so $\mathsf{w}_1 \triangleq \{y_1\}$ is augmented to $\mathsf{w}_1 \triangleq \{y_1, y_2\}$.
- **step 2**: if w_1 is not completed, repeat the operations of step 1 on the new measurement just included.

  In this example (only) the measurement y_2 was just added to w_1 so, firstly the set of neighborhs of y_2 is searched. Excluding y_1, the measure y_2 has not any neighbohrs so w_1 is declared complete, so

$$\mathsf{w}_1 \triangleq \{y_1, y_2\} \tag{5.131}$$

- **step 3**: the first measure that is not in w_1 is considered to initialize the new cell w_2. In this example such measure is y_3, so $\mathsf{w}_2 \triangleq \{y_3\}$. Then, step 1 and step 2 are applied to the new cell. In this example y_3 has not any neighbors, so the relative cell is declared complete

$$\mathsf{w}_2 \triangleq \{y_3\} \tag{5.132}$$

- **step 4**: the first measurement that is not in w_1 or w_2 is considered to initialize the new cell w_3. In this example such measurement is y_4, so $\mathsf{w}_3 \triangleq \{y_4\}$. Then, by repeating step 1 and step 2 on w_3 turns out for this example:
  * y_4 has one neighbor that is y_5. The cell is augmented to $\mathsf{w}_3 = \{y_4, y_5\}$.
  * the new measurement y_5 has a new neighbor that is y_6. The cell is augmented to $\mathsf{w}_3 = \{y_4, y_5, y_6\}$.
  * the new measurement y_6 has a new neighbor that is y_7. The cell is augmented to $\mathsf{w}_3 = \{y_4, y_5, y_6, y_7\}$.
  * the new measurement y_7 has not any new neighbors, so w_3 is declared complete.

$$\mathsf{w}_3 \triangleq \{y_4, y_5, y_6, y_7\} \tag{5.133}$$

- **step 5**: the first measurement that is not in w_1, w_2 or w_3 is considered to initialized the new cell w_4. In this example w_1, w_2 and w_3 exhaust the set of measurements y, so the partition composed by such cells is declared complete, thus it is returned

$$\mathcal{P} = \{\mathsf{w}_1, \mathsf{w}_2, \mathsf{w}_3\} \tag{5.134}$$

  which is the considered partition.

Note that the defining property is satisfied: the couples of near measurements are

- $\Delta_{1,2} = 1.4 < \gamma \triangleq 1.5$. The measures y_1 and y_2 are in the same cell w_1;
- $\Delta_{4,5} = 1 < \gamma \triangleq 1.5$. The measures y_4 and y_5 are in the same cell w_3;
- $\Delta_{5,6} = 1 < \gamma \triangleq 1.5$. The measures y_5 and y_6 are in the same cell w_3;
- $\Delta_{6,7} = 1 < \gamma \triangleq 1.5$. The measures y_6 and y_7 are in the same cell w_3;

On the other hand, if two measurements are in the same cell then in general is not true that are near. For example, consider y_4 and y_7, they belong to the same cell w_3 but their distance is $\Delta_{4,7} = 3 > \gamma \triangleq 1.5$, so they are not near.

5.5.4 GLO partitioning

In general, to achieve good tracking performance it is necessary to consider a partition that resembles accurately the true partition generated by the objects in the scene. Considering only one partition between the $B_{|\mathsf{y}|}$ possible partitions is not a good idea because it is unlikely that such partition is similar to the true one. In order to increase the chance to have included a good partition in the estimation problem, a set of partition $\mathcal{S}_P(\mathsf{y})$ containing $P > 1$ different partitions is required.

The GLO (Granstrom Lundquist Orguner - name of the authors) partitioning generates multiple partitions with the following procedure:

- **step 1**: Given the set of measurements y, every $\binom{|\mathsf{y}|}{2} = |\mathsf{y}| \cdot (\mathsf{y} - 1)/2$ possible distances $\Delta_{i,j}$are computed by considering the (unitless) Mahalanobis distance

$$\Delta_{i,j} \triangleq \sqrt{(y_i - y_j)'\, R^{-1} (y_i - y_j)} \tag{5.135}$$

- **step 2**: Since y is Gaussian and the distance considered is Mahalanobis, the variable $\Delta_{i,j}$ is a χ^2 with $p = \dim(y)$ degrees of freedom. The following confidence interval (based on the inverse CDF of the χ^2_p distribution) is thus computed

$$\delta_{\min}(1-\alpha) < \Delta_{i,j} < \delta_{\max}(1-\alpha) \tag{5.136}$$

where the significance level α is defined by $1 - \alpha = \mathbb{P}(\delta_{\min}(1-\alpha) < \Delta_{i,j} < \delta_{\min}(1-\alpha))$ and typically is chosen in $[0, 0.4]$.

- **step 3**: is defined the set of thresholds as follows

$$\Gamma \triangleq \{\Delta_{i,j} : \delta_{\min}(1-\alpha) < \Delta_{i,j} < \delta_{\max}(1-\alpha)\} \cup \{0\} \tag{5.137}$$

  where only the distances $\Delta_{i,j}$ with statistical relevance are considered.

- **step 4**: For every treshold $\gamma \in \Gamma$, distance partitioning is performed. Consequently, in this step $|\Gamma|$ partitions are generated.

- **step 5**: It is possible that distance partitioning produces the same partition when different tresholds are considered. In this final step, the set of partition $\mathcal{S}_P(\mathsf{y})$ is defined by considering only the different partitions generated in step 4.

Part II

Shape filters

Chapter 6

Fundamentals of random matrices

6.1 Summary

In this chapter are discussed the main mathematical tools, the Wishart and inverse Wishart distributions, used by the GIW, MEM-EKF* and LO-MEM filters to deal with the estimation of the shape of an extended object. Since theese two distributions are matrix-variate, new concepts such random matrices, matrix-variate PDFs are briefly introduced in the first part of the chapter. In the final part of the chapter are discussed the definitions and applications of the Wishart and inverse Wishart distributions.

6.2 Matrix vectorization

6.2.1 Full vectorization

Let A be a generic matrix in $\mathbb{R}^{m\times n}$. Let $A_1, \ldots, A_n$ be the column of A

$$A = \left[\begin{array}{c|c|c} A_1 & \cdots & A_n \end{array}\right] \tag{6.1}$$

the *vectorization* of A is an operation that maps the matrix A into a column vector $\text{vec}[A] \in \mathbb{R}^{m\cdot n}$ obtained by stacking vertically the columns $A_1, \ldots, A_n$ of A

$$\text{vec}[A] \triangleq \bigoplus_{i=1}^{n} A_i = \begin{bmatrix} A_1' & \cdots & A_n' \end{bmatrix}' \tag{6.2}$$

Note that given the vectorization $\text{vec}[A]$ it is possible to recover the orginal matrix A by partitioning $\text{vec}[A]$ in n consecutive column $m \times 1$ subvectors $A_1, \ldots, A_n$ and by organizing them in a $m \times n$ matrix, which is A itself. This means that $\text{vec}[A]$ provides an equivalent way to express the information encoded by matrix A. Thanks to the vectorial representation, familiar vectorial concepts, like the core concepts of the vectorial differential calculus or the concepts of the random vectors theory, can be easily extend to the space of matrices.

6.2.2 Half vectorization

Let A be a symmetric matrix in $\mathbb{R}^{m \times m}$, i.e. $A_{ji} = A_{ij}$ for all $i, j = 1, \ldots, m$. In this case the matrix A can be expressed as a "compressed" vectorization since there are not m^2 distinct elements but only $m(m+1)/2$, which are the diagonal elements $\{A_{ii}\}_{1 \leq i \leq m}$ and (by convention) the super-diagonal elements $\{A_{ij}\}_{1 \leq j \leq i \leq m}$ (rather than the sub-diagonal elements).

Exploiting this fact, define the *half vectorization* of $A = A' \in \mathbb{R}^{m \times m}$ as the operation that maps A into the column vector $\text{vech}[A] \in \mathbb{R}^{\frac{m(m+1)}{2}}$ given by

$$\text{vech}[A] \triangleq \begin{bmatrix} A_{11} & A_{12} & A_{22} & \cdots & A_{1m} & \cdots & A_{mm} \end{bmatrix}' \tag{6.3}$$

for example, if $m \triangleq 4$ then

$$\text{vech}[A] \triangleq \begin{bmatrix} A_{11} & A_{12} & A_{22} & A_{31} & A_{32} & A_{33} & A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}' \tag{6.4}$$

6.3 Kronecker product

The *Kronecker product* between a matrix $A \in \mathbb{R}^{m \times n}$ and a matrix $B \in \mathbb{R}^{p \times q}$ is the matrix $A \otimes B \in \mathbb{R}^{pm \times qn}$ defined as follows

$$A \otimes B \triangleq \begin{bmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix}$$

often the Kronecker product permits to express complicated expressions in simple and compact form. For example, consider a 2×2 sample covariance matrix obtained from a sample of m measures $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ characterized by the mean measure $\bar{\boldsymbol{y}} \triangleq m^{-1} \sum_{i=1}^{m} \boldsymbol{y}_i$

$$\boldsymbol{S} \triangleq \frac{1}{m-1} \sum_{i=1}^{m} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})' \tag{6.5}$$

if, for some reason, the vectorial representation $v_{\boldsymbol{S}} \triangleq \text{vech}[\boldsymbol{S}]$ is preferred to the matricial representation $\boldsymbol{S}$, then one can express in vectorial form the sample covariance only trough algebraic operations

$$v_{\boldsymbol{S}} = \frac{F}{m-1} \sum_{i=1}^{m} [(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \otimes (\boldsymbol{y}_i - \bar{\boldsymbol{y}})] \tag{6.6}$$

where

$$F \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.7}$$

6.4 Some useful properties

The vectorization and the Kronecker product give rise to some useful properties that are widely used in multivariate statistic. A brief list of such properties is the following:

- $\text{vec}[aA + bB] = a\text{vec}[A] + b\text{vec}[B]$ (linearity);
- $\text{vec}[AB] = (I \otimes A)\text{vec}[B] = (B' \otimes I)\text{vec}[A] = (B' \otimes A)\text{vec}[I]$;
- $\text{vec}[ABC] = (C' \otimes A)\text{vec}[B]$;
- $\text{vec}[vv'] = v \otimes v$ (with v column vector);
- $\text{tr}[A'B] = \text{vec}[A]'\text{vec}[B]$;
- $\text{tr}[ABC] = \text{vec}[A']'(I \otimes C)\text{vec}[C]$;
- $\text{tr}[D'ABC'] = \text{vec}[D]'(C \otimes A)\text{vec}[B]$;
- $\text{tr}[AB'CBD] = \text{vec}[B]'(DA \otimes C')\text{vec}[B] = \text{vec}[B]'(A'D' \otimes C)\text{vec}[B]$;
- $\text{tr}[AB'CB] = \text{vec}[B]'(A \otimes C')\text{vec}[B] = \text{vec}[B]'(A' \otimes C)\text{vec}[B]$;

6.5 Matrix integral

6.5.1 Full integration

Let $f : \mathbb{R}^{m \times n} \mapsto \mathbb{R}$ be a scalar function of the generic $m \times n$ matrix X. The integral of $f(\cdot)$ over a region $R \subseteq \mathbb{R}^{m \times n}$ is defined as the iterated integral of $f(\cdot)$ with respect each element of vec$[X]$, where such elements X_{ij} range according vec$[R]$

$$\begin{aligned}\int_R f(X)\ \mathrm{d}X &\triangleq \int_{\mathrm{vec}[R]} f(\mathrm{vec}[X])\ \mathrm{d}\,\mathrm{vec}[X] \\ &= \int_{\mathrm{vec}[R]} f(X_{11},\cdots,X_{mn})\ \mathrm{d}X_{11}\cdots\mathrm{d}X_{mn}\end{aligned} \tag{6.8}$$

in short, the measure considered on the space $\mathbb{R}^{m\times n}$ of $m\times n$ real-valued matrices is the Lebesgue measure of the $m\cdot n$-dimensional Euclidean space $\mathbb{R}^{m\cdot n}$

$$\mathrm{d}X \triangleq \mathrm{d}\,\mathrm{vec}[X] \triangleq \prod_{j=1}^{n}\prod_{i=1}^{m}\mathrm{d}X_{ij} \tag{6.9}$$

essentially, the matrix integral is a conventional multiple integral in a "high dimensional" Euclidean space.

6.5.2 Half integration

Let $f:\mathbb{R}^{m\times m}\mapsto\mathbb{R}$ be a scalar function of a symmetric $m\times m$ matrix X. The integral of $f(\cdot)$ over a region $R\subseteq\mathbb{R}^{m\times m}$ is defined as the the iterated integral of $f(\cdot)$ with respect each element vech$[X]$, where such elements X_{ij} range according vech$[R]$

$$\begin{aligned}\int_R f(X)\ \mathrm{d}X &\triangleq \int_{\mathrm{vech}[R]} f(\mathrm{vech}[X])\ \mathrm{d}\,\mathrm{vech}[X] \\ &= \int_{\mathrm{vech}[R]} f(X_{11},X_{12},X_{22},\cdots,X_{1m},\cdots X_{mm}) \\ &\quad\times\ \mathrm{d}X_{11}\,\mathrm{d}X_{12}\,\mathrm{d}X_{22}\cdots\mathrm{d}X_{1m}\ \cdots\mathrm{d}X_{mm}\end{aligned} \tag{6.10}$$

in short, the measure considered on the space $\mathbb{R}^{m\times m}$ of symmetric $m\times m$ real-valued matrices is the Lebesgue measure of the $m\cdot(m+1)/2$-dimensional Euclidean space $\mathbb{R}^{\frac{m(m+1)}{2}}$

$$\mathrm{d}X \triangleq \mathrm{d}\,\mathrm{vech}[X] = \prod_{j=1}^{m}\prod_{i\le j}^{m}\mathrm{d}X_{ij} \tag{6.11}$$

6.6 Random matrices

Roughly speaking, an $m\times n$ random matrix $\boldsymbol{X}$ is a matrix whose $m\cdot n$ entries $\boldsymbol{X}_{ij}$ are random variables. More formally, an $m\times n$ random matrix $\boldsymbol{X}$ is a measurable map $\boldsymbol{X} : (\Omega, \mathcal{E}, \mathbb{P}) \mapsto \mathbb{R}^{m \times n}$, where $(\Omega, \mathcal{E}, \mathbb{P})$ is a probability space. According to this fact, the basic concepts related to random matrices are direct generalizations of the basic concepts of the random variables. In what follows, such concepts, i.e. the definition of PDF and the definitions of the most common moment of a random matrix, are briefly discussed.

6.6.1 Generic random matrices

- **definition 1**: Let $p : \mathbb{R}^{m \times n} \mapsto \mathbb{R}$ be a scalar function of the generic $m \times n$ matrix X. The function $p(\cdot)$ is a *matrix-variate PDF* if and only if

  1. $p(X) \geq 0$ for all $X \in \mathbb{R}^{m \times n}$;
  2. $\int p(X)\ \mathrm{d}X \triangleq \int p(\mathrm{vec}[X]))\ \mathrm{d}\,\mathrm{vec}[X] = 1$.

- **definition 2**: Let $h : \mathbb{R}^{m \times n} \mapsto \mathbb{R}^{p \times q}$ be a matrix function and let $\boldsymbol{X}$ be a generic random $m \times n$ matrix with PDF $p(\cdot)$. The *expected value* of $h(\cdot)$, whos typical element is denotes as h_{ij}, is defined as

$$\mathbb{E}[h(\boldsymbol{X})] \triangleq \begin{bmatrix} \mathbb{E}[h_{11}(\boldsymbol{X})] & \cdots & \mathbb{E}[h_{1q}(\boldsymbol{X})] \\ \vdots & & \vdots \\ \mathbb{E}[h_{p1}(\boldsymbol{X})] & \cdots & \mathbb{E}[h_{pq}(\boldsymbol{X})] \end{bmatrix} \tag{6.12}$$

  where for all $i = 1, \ldots, p$ and for all $j = 1, \ldots, q$

$$\mathbb{E}[h_{ij}(\boldsymbol{X})] \triangleq \int h_{ij}(X)\, p(X)\ \mathrm{d}X \triangleq \int h_{ij}(X)\, p(\mathrm{vec}[X]))\ \mathrm{d}\,\mathrm{vec}[X] \tag{6.13}$$

- **definition 3**: The *expected value* of the $m \times n$ random matrix $\boldsymbol{X}$, whos typical element is $\boldsymbol{X}_{ij}$, with PDF $p(\cdot)$ is the expected value of the identical function $h(\boldsymbol{X}) = \boldsymbol{X}$, i.e.

$$\mathbb{E}[\boldsymbol{X}] \triangleq \begin{bmatrix} \mathbb{E}[\boldsymbol{X}_{11}] & \cdots & \mathbb{E}[\boldsymbol{X}_{1n}] \\ \vdots & & \vdots \\ \mathbb{E}[\boldsymbol{X}_{m1}] & \cdots & \mathbb{E}[\boldsymbol{X}_{mn}] \end{bmatrix} \tag{6.14}$$

  where for all $i = 1, \ldots, m$ and for all $j = 1, \ldots, n$

$$\mathbb{E}[\boldsymbol{X}_{ij}] \triangleq \int X_{ij}\, p(X)\ \mathrm{d}X \triangleq \int X_{ij}\, p(\mathrm{vec}[X]))\ \mathrm{d}\,\mathrm{vec}[X] \tag{6.15}$$

- **definition 4**: The *variance* of the $m \times n$ random matrix $\boldsymbol{X}$ is the expected value of the function $h(\boldsymbol{X}) = (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])'$, i.e.

$$\text{Var}[\boldsymbol{X}] \triangleq \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])'\right] \tag{6.16}$$

as one can show, the expectation operator (66) satisfies the familiar linear properties of the usual expectation operator:

- if A is a deterministic matrix then $\mathbb{E}[A] = A$;
- if A, B are deterministic matrices with suitables dimensions, then $\mathbb{E}[Ah(\boldsymbol{X})B] = A\mathbb{E}[h(\boldsymbol{X})]B$;
- if $h_1(\cdot)$ and $h_2(\cdot)$ are matrix functions of the same order then $\mathbb{E}[h_1(\boldsymbol{X}) + h_2(\boldsymbol{X})] = \mathbb{E}[h_1(\boldsymbol{X})] + \mathbb{E}[h_2(\boldsymbol{X})]$

from the previous properties of the expectation operator follows immediately the simplified formula for the variance

$$\text{Var}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}'] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]' \tag{6.17}$$

More commonly than the variance, the dispersion of a random matrix is quantified by the covariance. The extension to the matrix-variate case of the covariance definition is less straightforwad then the previous definitions and involves the vec[$\cdot$] operator. Since there is a bijection between $\boldsymbol{X}$ and vec[$\boldsymbol{X}$], the statistical properties of the random matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ are completely represented by the statistical properties of the random column vector $\text{vec}[\boldsymbol{X}] \in \mathbb{R}^{m \cdot n}$. This means that the distribution of a random matrix $\boldsymbol{X}$ is simply the distribution of the random vector vec[$\boldsymbol{X}$]. Naturally, one can think also in terms of the columns of $\boldsymbol{X}$, which are the columns of $\boldsymbol{X}'$, thus the same reasoning holds for the random column vector $\text{vec}[\boldsymbol{X}'] \in \mathbb{R}^{n \cdot m}$. According to this fact, from the definition of matrix-variate PDF and matrix integral follows the relation

$$p(X) = p(\text{vec}[X]) = p(\text{vec}[X']) = p(X_{11}, \cdots, X_{mn}) \tag{6.18}$$

which says that there are four equivalent ways to express the PDF of a random matrix:

1. (the most explicit) $p(X_{11}, \cdots, X_{mn})$, where every single random variable X_{ij} is denoted separately in the joint density $p(\cdot)$;
2. $p(\text{vec}[X]) = p(X_1, \cdots, X_n)$, where the random variables X_{ij} are organized and denoted in the joint density $p(\cdot)$ in n random m-dimensional column vectors, which are the columns of X;

3. $p(\mathrm{vec}[X']) = p(X_1, \cdots, X_m)$, where the random variables X_{ij} are organized and denoted in the joint density $p(\cdot)$ in m random n-dimensional column vectors, which are the transpose rows of X;
4. (the most coincise) $p(X)$, where the random variables X_{ij} are globally organized and denoted in the joint density $p(\cdot)$ in 1 random matrix.

As a consequence, one can define the density of a random matrix by using three different representations and then switch from one to another without loosing any information. Thus, now consider the vectorial representation given by $p(\mathrm{vec}[X'])$ instead of the global representation given by $p(X)$. Since $\mathrm{vec}[\boldsymbol{X}']$ is a random vector, it is well-defined its covariance.

- **definition 5**: The *covariance* of a random matrix $\boldsymbol{X}$ is the covariance of the the random vector $\mathrm{vec}[\boldsymbol{X}']$, i.e.

$$\begin{aligned}\mathrm{Cov}[\boldsymbol{X}] &\triangleq \mathrm{Cov}[\mathrm{vec}[\boldsymbol{X}']] \\ &\triangleq \mathbb{E}[(\mathrm{vec}[\boldsymbol{X}'] - \mathbb{E}[\mathrm{vec}[\boldsymbol{X}']])(\mathrm{vec}[\boldsymbol{X}'] - \mathbb{E}[\mathrm{vec}[\boldsymbol{X}']])']\end{aligned} \tag{6.19}$$

- **definition 6**: The *cross-covariance* between the random matrices $\boldsymbol{X}$, $\boldsymbol{Y}$ is the cross-covariance between the random vectors $\mathrm{vec}[\boldsymbol{X}']$, $\mathrm{vec}[\boldsymbol{Y}']$, i.e.

$$\begin{aligned}\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}] &\triangleq \mathrm{Cov}[\mathrm{vec}[\boldsymbol{X}'], \mathrm{vec}[\boldsymbol{Y}']] \\ &\triangleq \mathbb{E}[(\mathrm{vec}[\boldsymbol{X}'] - \mathbb{E}[\mathrm{vec}[\boldsymbol{X}']])(\mathrm{vec}[\boldsymbol{Y}'] - \mathbb{E}[\mathrm{vec}[\boldsymbol{Y}']])']\end{aligned} \tag{6.20}$$

Abbreviate $v_{\boldsymbol{X}} \triangleq \mathrm{vec}[\boldsymbol{X}']$, $\bar{v}_{\boldsymbol{X}} \triangleq \mathbb{E}[\mathrm{vec}[\boldsymbol{X}']]$, then the covariance of $\boldsymbol{X}$ can be expressed more clearly as

$$\mathrm{Cov}[\boldsymbol{X}] = \mathrm{Cov}[v_{\boldsymbol{X}}] = \mathbb{E}[(v_{\boldsymbol{X}} - \bar{v}_{\boldsymbol{X}})(v_{\boldsymbol{X}} - \bar{v}_{\boldsymbol{X}})'] \tag{6.21}$$

moreover, due to the usual elementary properties, holds the simplified formula

$$\mathrm{Cov}[\boldsymbol{X}] = \mathbb{E}[v_{\boldsymbol{X}} v'_{\boldsymbol{X}}] - \bar{v}_{\boldsymbol{X}} \bar{v}'_{\boldsymbol{X}} \tag{6.22}$$

and in the same way,

$$\begin{aligned}\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}] = \mathrm{Cov}[v_{\boldsymbol{X}}, v_{\boldsymbol{Y}}] &= \mathbb{E}[(v_{\boldsymbol{X}} - \bar{v}_{\boldsymbol{X}})(v_{\boldsymbol{Y}} - \bar{v}_{\boldsymbol{Y}})'] \\ &= \mathbb{E}[v_{\boldsymbol{X}} v'_{\boldsymbol{Y}}] - \bar{v}_{\boldsymbol{X}} \bar{v}'_{\boldsymbol{Y}}\end{aligned} \tag{6.23}$$

6.6.2 Symmetric random matrices

For the symmetric case one has to define the concepts of PDF and moments in terms of half integration rather than full integration. Thus, in every occurrance, the $\mathrm{vec}[\cdot]$ is replaced with the $\mathrm{vech}[\cdot]$ operator.

6.7 Wishart distribution

6.7.1 Definition

Definition 7. Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ be independent $\mathcal{N}(0, \Sigma)$ random vectors in $\mathbb{R}^p$. If $\Sigma > 0$ and $m > p - 1$, then the symmetric random matrix $\boldsymbol{A}$, called *scatter matrix*, defined as

$$\boldsymbol{A} \triangleq \sum_{i=1}^{m} \boldsymbol{y}_i \boldsymbol{y}_i' \tag{6.24}$$

is said to have the p-dimensional *Wishart distribution* with m degrees of freedom and scale matrix Σ. It will be used the notation $\boldsymbol{A} \sim \mathcal{W}_p(m, \Sigma)$

The Wishart distribution gets a familiar form if the attention is restricted to the simpler case where:

- the sample is univariate, i.e. $p = 1$ or $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ (with $m > 1$) are independent scalar random variables;
- the underlying distribution of the sample is a standard Gaussian, i.e.

$$\boldsymbol{y}_i \sim \mathcal{N}(0, 1) \qquad i = 1, 2, \ldots, m$$

then the Wishart distribution reduces to the familiar chi-squared distribution (with m degrees of freedom)

$$\boldsymbol{A} \sim \mathcal{W}_1(m, 1) = \chi^2_m$$

The Wishart distribution can be though as a multidimensional chi squared distribution and, likewise the χ^2_m distribution is used to estimate the variance of a Gaussian univariate sample, the $\mathcal{W}_p(m, \Sigma)$ distribution is used to estimate the covariance matrix of a Gaussian p-variate sample.

6.7.2 Density and moments

If $\boldsymbol{A} \sim \mathcal{W}_p(m, \Sigma)$ then the matrix-variate PDF of $\boldsymbol{A}$ is

$$p_{\boldsymbol{A}}(A; p, m, \Sigma) = \begin{cases} \frac{1}{K} (\det A)^{\frac{m-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1} A\right) & \text{if } A = A' > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.25}$$

where $\operatorname{etr}(\cdot)$ is the exponential trace operator[1] and the normalizer K is given by

$$K \triangleq 2^{m\frac{p}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{m+1-i}{2}\right) (\det \Sigma)^{\frac{m}{2}} \tag{6.26}$$

[1] if M is a square matrix, then $\operatorname{etr}(M) \triangleq \exp(\operatorname{tr}[M])$, where $\operatorname{tr}[\cdot]$ is the trace operator (sum of the diagonal elements of M)

with $\Gamma(\cdot)$ be the Euler's Gamma function[2], i.e.

$$\Gamma(x) \triangleq \int_{\mathbb{R}+} t^{x-1} \exp(-t) \, \mathrm{d}t \tag{6.27}$$

According to the PDF (???), the expected value and the covariance of $\boldsymbol{A}$ are given by

$$\begin{aligned} \mathbb{E}[\boldsymbol{A}] &= m\Sigma \\ \mathrm{Cov}[\mathrm{vec}[\boldsymbol{A}]] &= m(I_{p^2} + K_p)(\Sigma \otimes \Sigma) \end{aligned} \tag{6.28}$$

where I_{p^2} is the $p^2 \times p^2$ identity matrix and K_p, called *commutation matrix*[3], is defined as

$$K_p \triangleq \sum_{i,j=1}^{p} (H_{ij} \otimes H'_{ij}) \tag{6.29}$$

with $H_{ij} \in \{0,1\}^{p\times p}$ be a matrix having a unit in position ij and zero elsewhere. For example, if $p = 2$ then

$$H_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad H_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad H_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad H_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.30}$$

and

$$K_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.31}$$

Note that the covariance of $\boldsymbol{A}$ is expressed in a redundant form due to the use of the vec[$\cdot$] operator.

6.7.3 Application

The most important application of the Wishart distribution is expressed by the following famous result.

Theorem 12. (Wishart, 1920 circa) Let $\{\boldsymbol{y}_i\}_{i=1}^m$ be a sample of $m > p$ IID p-variate Gaussian vectors with expectation $\mu_y \in \mathbb{R}^p$ and non-singular covariance $\Sigma_y \in \mathbb{R}^{p\times p}$

$$\boldsymbol{y}_i \sim \mathcal{N}(\mu_y, \Sigma_y) \qquad i = 1, 2, \ldots, m$$

then

[2] the Gamma function is the extension of the factorial function $n \mapsto n!$ to non-integers

[3] the name arise from the properties $\mathrm{vec}[M'] = K_p \mathrm{vec}[M]$, $\mathrm{vec}[M] = K_p \mathrm{vec}[M']$, where M is a generic $p \times p$ matrix

1. the sample mean $\bar{\boldsymbol{y}} \in \mathbb{R}^p$ is Gaussian with expectation μ_y and covariance Σ_y/m
$$\bar{\boldsymbol{y}} \triangleq \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{y}_i \sim \mathcal{N}\left(\mu_y, \frac{\Sigma_y}{m}\right)$$
2. the centered scatter matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ is a p-variate Wishart with $m-1$ degrees of freedom and covariance matrix Σ_y
$$\boldsymbol{A} \triangleq \sum_{i=1}^{m} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})' \sim \mathcal{W}_p\left(m-1, \Sigma_y\right)$$
3. the sample mean $\bar{\boldsymbol{y}}$ and the scatter matrix $\boldsymbol{A}$ are independently distributed

due to this fact, turns out that $\bar{\boldsymbol{y}}$, $\boldsymbol{A}$ are sufficient statistics to infer the value of μ_y and Σ_y, indeed due to the facts

$$\begin{aligned} \mathbb{E}[\bar{\boldsymbol{y}}] &= \mu_y \\ \mathbb{E}[\boldsymbol{A}] &= (m-1)\Sigma_y \end{aligned} \tag{6.32}$$

follows that, given $\bar{\boldsymbol{y}} = \bar{y}$, $\boldsymbol{A} = A$, one can reasonably estimates for the parameters μ_y, Σ_y of the sample population as

$$\begin{aligned} \hat{\mu}_y &= \bar{y} \\ \hat{\Sigma}_y &= S = \frac{A}{m-1} \end{aligned} \tag{6.33}$$

the random matrix $\boldsymbol{S} \triangleq \boldsymbol{A}/(m-1)$ is nothing but more than the sample covariance matrix, and it turns out that it is Wishart as well

$$\boldsymbol{S} \sim \mathcal{W}_p\left(m-1, \frac{\Sigma_y}{m-1}\right) \tag{6.34}$$

this result plays a central role in the LO-MEM corrector.

6.8 Inverse Wishart distribution

6.8.1 Definition and expectation

Definition 8. A $p \times p$ random matrix $\boldsymbol{B}$ is said to have the p-dimensional *inverse Wishart distribution* with ν degrees of freedom and parameter matrix V if and only if $\nu > 2p$ and its density function is

$$p_{\boldsymbol{B}}(B; p, \nu, V) = \begin{cases} \frac{1}{K}\left(\det B\right)^{-\frac{\nu}{2}} \operatorname{etr}\left(-\frac{1}{2}B^{-1}V\right) & \text{if } B = B' > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.35}$$

where the normalizer K is given by

$$K \triangleq 2^{p\frac{\nu-p-1}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{\nu-p-i}{2}\right) (\det V)^{-\frac{\nu-p-1}{2}} \tag{6.36}$$

It will be used the notation $\boldsymbol{B} \sim \mathcal{IW}_p(\nu, V)$.

The inverse Wishart distribution is the distribution of the inverse of a Wishart random matrix, indeed hold the following facts:

- if $\boldsymbol{A} \sim \mathcal{W}_p(m, \Sigma)$ then $\boldsymbol{B} \triangleq \boldsymbol{A}^{-1} \sim \mathcal{IW}_p(\nu \triangleq m+p+1, V \triangleq \Sigma^{-1})$.
- if $\boldsymbol{B} \sim \mathcal{IW}_p(\nu, V)$ then $\boldsymbol{A} \triangleq \boldsymbol{B}^{-1} \sim \mathcal{W}_p(m \triangleq \nu-p-1, \Sigma \triangleq V^{-1})$.

If $\nu - 2p - 2 > 0$, the expected value of an inverse Wishart matrix $\boldsymbol{B} \sim \mathcal{IW}_p(\nu, V)$ is

$$\mathbb{E}[\boldsymbol{B}] = \frac{V}{\nu - 2p - 2} = \frac{\Sigma^{-1}}{m - p - 1} \tag{6.37}$$

6.8.2 Application

The central property of the inverse Wishart distribution is that it is the conjugate prior of the Wishart distribution: let $\Sigma = \Sigma' > 0$ be a unknown $p \times p$ matrix and $A = A' > 0$ be the observation of a random matrix $\boldsymbol{A}$, by considering the prior model

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_p(\nu, V) \tag{6.38}$$

and by considering the likelihood

$$\boldsymbol{A}|\Sigma \sim \mathcal{W}_p(m, \Sigma) \tag{6.39}$$

follows that, given $\boldsymbol{A} = A$, the posterior distribution of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}|A \sim \mathcal{IW}_p(\nu + m, V + A) \tag{6.40}$$

This important result is the core concept of the GIW filter. Such filter represents the shape of an extended object with a $d \times d$ (with $d = 2$ if the object moves on a plane or with $d = 3$ if the object moves in the space) *symmetric and positive definite* (SPD) random matrix $\boldsymbol{X} = \boldsymbol{X}' > 0$ and, in order to produce a Bayesian estimate of $\boldsymbol{X}$, assumes that $\boldsymbol{X} \sim \mathcal{IW}_d(\nu, V)$. Then consider as a measure the centered scatter matrix $\overline{\boldsymbol{Y}}$ generated by a sample $\{\boldsymbol{y}_i \in \mathbb{R}^d\}_{i=1}^n$ assumed to be Gaussian, so $\overline{\boldsymbol{Y}} \sim \mathcal{W}_d(n-1, \Sigma_y)$. Thanks to these positions, it turns out that the corrected density of $\boldsymbol{X}$ is

$\mathcal{IW}_2(\nu + n - 1, V + \overline{Y})$, so the Bayesian estimate $\boldsymbol{X}$ produced by GIW filter is

$$\hat{X} = \mathbb{E}[\boldsymbol{X}|\overline{Y}] = \frac{(V + \overline{Y})^{-1}}{(\nu + n - 1) - 2d - 2} \tag{6.41}$$

Notice that this is only the underlying idea of the GIW filter and some important aspects of the estimation process are neglected. For this reason, the above estimate does not represent the exact estimate produced by the GIW filter but only an approximation.

Chapter 7

GIW filter

7.1 Summary

This chapter provides the derivation of the first filter, the GIW filter, that is able to estimate both the position and the shape of a single extended object. In the first part of the chapter are discussed the main ideas involved to represent and estimate the jointly the position and the shape of the tracked object, then GIW predictor and corrector are derived in details. The chapter ends with the implementation of the GIW filter with the PHD filter for extended object, which results in the GIW-PHD filter, which is able to track simultaneously the positions and the shapes of multiple extended objects.

7.2 Elements of the Bayesian solution

A single extended object is modelled by two random variables:

- **kinematic state**: is a random vector $\boldsymbol{x}_k$ modelling the position, velocity, acceleration, etc, of the object. For example, in a 2-dimensional scenario, a possible choice of $\boldsymbol{x}_k$ is the following

$$\boldsymbol{x}_k \triangleq \begin{bmatrix} \boldsymbol{m}_k' & \dot{\boldsymbol{m}}_k' & \ddot{\boldsymbol{m}}_k' \end{bmatrix}' \in \mathbb{R}^{2\cdot 3} = \mathbb{R}^6 \tag{7.1}$$

  where:

  - $\boldsymbol{m}_k \triangleq [\boldsymbol{\xi}_k \ \boldsymbol{\eta}_k]' \in \mathbb{R}^2$ is the position of the object expressed in cartesian coordinates;
  - $\dot{\boldsymbol{m}}_k \triangleq [\dot{\boldsymbol{\xi}}_k \ \dot{\boldsymbol{\eta}}_k]' \in \mathbb{R}^2$ is the velocity of the object expressed in cartesian coordinates;

– $\ddot{\boldsymbol{m}}_k \triangleq [\ddot{\boldsymbol{\xi}}_k \; \ddot{\boldsymbol{\eta}}_k]' \in \mathbb{R}^2$ is the acceleration of the object expressed in cartesian coordinates.

naturally, one can generalize the expression of the kinematic state by considering a d-dimensional scenario (in a cartesian framework, the most common choices are $d \triangleq 2$, $d \triangleq 3$) and by including in the model the first $s-1$ derivatives of the position $\boldsymbol{m}_k$, i.e.

$$\boldsymbol{x}_k \triangleq \begin{bmatrix} \boldsymbol{m}_k' & \dot{\boldsymbol{m}}_k' & \ddot{\boldsymbol{m}}_k' & \dots & \boldsymbol{m}_k^{(s-1)} \end{bmatrix}' \in \mathbb{R}^{d \cdot s} \tag{7.2}$$

in other words, the kinematic state is a column vector $\boldsymbol{x}_k \in \mathbb{R}^n$ with dimension $n = s \cdot d$.

- **shape state**: is a $d \times d$ random matrix $\boldsymbol{X}_k$ modelling the shape of the object. It is assumed that the matrix $\boldsymbol{X}_k$ is SPD, likewise a non-singular covariance matrix, thus the locus of points $m \in \mathbb{R}^d$ satisfying the quadratic equation

$$m' \boldsymbol{X}_k m = 1 \tag{7.3}$$

consist in a d-dimensional hyper-ellipsoide representing the contour of the shape of the tracked object. For example, in the simple 2-dimensional cartesian case the shape of the object is represented by an ellipse in $\mathbb{R}^2$, while the extension matrix $\boldsymbol{X}_k$ assumes the simple form

$$\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{X}_{11,k} & \boldsymbol{X}_{12,k} \\ \cdot & \boldsymbol{X}_{22,k} \end{bmatrix} \tag{7.4}$$

note that the number of extension parameters is not d^2 but rather $d \cdot (d+1)/2 < d^2$ because the symmetry assumption on $\boldsymbol{X}_k$ (the parameters are only the diagonal components and the upper (or lower) triangle components of $\boldsymbol{X}_k$).

Given the representation $\boldsymbol{x}_k$, $\boldsymbol{X}_k$ of an extended object, the filtering problem is the following: estimate both $\boldsymbol{x}_k$ and $\boldsymbol{X}_k$ in function of the accumulated sensor measures $\mathsf{y}_{1:k} \triangleq \{\mathsf{y}_1, \dots, \mathsf{y}_k\}$ according the usual iterative scheme prescribed by the Bayesian approach. Note that y_k is not a proper RFS because the GIW filter assumes that the cardinality is not random but fixed to the value n_k, i.e.

$$\mathsf{Y}_k \triangleq \{\boldsymbol{y}_{1,k}, \dots, \boldsymbol{y}_{n_k,k}\} \qquad n_k \text{ known} \tag{7.5}$$

The predicted and corrected PDFs considered are the *joint* conditional PDFs

$$\begin{aligned} p_{k|k-1}(x, X) &\triangleq p(x_k, X_k | \mathsf{y}_{1:k-1}) && \text{predicted PDF} \\ p_{k|k}(x, X) &\triangleq p(x_k, X_k | \mathsf{y}_{1:k}) && \text{corrected PDF} \end{aligned} \tag{7.6}$$

7.2.1 Prediction

Given the corrected PDF $p_{k-1|k-1}(\cdot,\cdot)$ at the previous time step $k-1$, the predicted PDF is given by the Chapman-Kolmogorov equation

$$p_{k|k-1}(x,X) = \int \varphi_{k|k-1}(x,X|w,W)\, p_{k-1|k-1}(w,W)\ \mathrm{d}w\, \mathrm{d}W \tag{7.7}$$

this formula can be simplified by making the following assumptions:

1. the shape X is statistically independent from the kinematic x;
2. the shape X does not change rapidly in time.

as one can show, with this positions holds

$$\begin{aligned} p_{k|k-1}(x,X) \approx & \underbrace{\left(\int \varphi_{k|k-1}(x|X,w)\, p_{k-1|k-1}(w|X)\ \mathrm{d}w\right)}_{\triangleq p_{k|k-1}(x|X)} \\ & \times \underbrace{\left(\int \varphi_{k|k-1}(X|W)\, p_{k-1|k-1}(W)\ \mathrm{d}W\right)}_{\triangleq p_{k|k-1}(X)} \end{aligned} \tag{7.8}$$

which says that the predicted density is given by the product of the following two separate integrations

$$\begin{aligned} p_{k|k-1}(x|X) &\triangleq \int \varphi_{k|k-1}(x|X,w)\, p_{k-1|k-1}(w|X)\ \mathrm{d}w \\ p_{k|k-1}(X) &\triangleq \int \varphi_{k|k-1}(X|W)\, p_{k-1|k-1}(W)\ \mathrm{d}W \end{aligned} \tag{7.9}$$

therefore the kinematic state and the shape state can be predicted separately. Note that the kinematic state is conditioned on the shape state in order to take into account its dependence to the shape state.

7.2.2 Correction

Given the actual predicted PDF $p_{k|k}(\cdot,\cdot)$, the corrected PDF is given by the Bayes equation

$$p_{k|k}(x,X) = \frac{\ell_k(\mathsf{y}|x,X)\, p_{k|k-1}(x,X)}{\int \ell_k(\mathsf{y}|w,W)\, p_{k|k-1}(w,W)\ \mathrm{d}w\, \mathrm{d}W} \tag{7.10}$$

assuming that y_k is composed by n_k IID measures distributed according the conditional density $p(y|x, X)$, the measurement likelihood gets the following factorized form

$$\ell_k(\mathsf{y}|x, X) = \left(\prod_{i=1}^{n_k} \ell_k(y^i|x, X) \right) \tag{7.11}$$

where

$$\ell_k(y|x, X) \triangleq p(y_k|x_k, X_k) \tag{7.12}$$

is the single-measure likelihood. As mentioned before, in this model the number of measures n_k is not considered random.

7.3 GIW predictor

7.3.1 Kinematic state prediction

The motion model considered for the kinematic state is linear and Gaussian

$$\boldsymbol{x}_k = \Phi_{k|k-1}\, \boldsymbol{x}_{k-1} + \boldsymbol{w}_k \tag{7.13}$$

where, assuming the same dynamic in every dimension $i = 1, 2, \ldots, d$, the d-dimensional transition state matrix is given by

$$\Phi_{k|k-1} \triangleq F_{k|k-1} \otimes I_d \tag{7.14}$$

where F_k is a 1-dimensional transition state matrix. The term $\boldsymbol{w}_k$ is a white noise with a zero-mean Gaussian distribution

$$\boldsymbol{w}_k \sim \mathcal{N}(0, \Delta_{k|k-1}) \tag{7.15}$$

with covariance matrix $\Delta_{k|k-1}$ given by

$$\Delta_{k|k-1} \triangleq D_{k|k-1} \otimes \boldsymbol{X}_k \tag{7.16}$$

where $D_{k|k-1}$ is the 1-dimensional plant noise. This model states that the covariance matrix $\Delta_{k|k-1}$ is proportional to the shape $\boldsymbol{X}_k$ of the object. The reason why the actual model is chosen consists in the fact that permits to compute the easily the prediction and correction steps.

On the other hand this model does not represents accurately the dynamic of an extended object because it states that bigger (in its extension) is the object, more irregular is the motion of the object. Clearly, there is no a physical justification in support to this feature of the model.

In summary, the linear-Gaussian motion model considered is the following

$$\begin{aligned} \boldsymbol{x}_k &= (F_{k|k-1} \otimes I_d)\, \boldsymbol{x}_{k-1} + \boldsymbol{w}_k \\ \boldsymbol{w}_k &\sim \mathcal{N}(0, D_{k|k-1} \otimes \boldsymbol{X}_k) \end{aligned} \tag{7.17}$$

for suitables matrices $F_{k|k-1}$, $D_{k|k-1}$. As a consequence, the kinematic transition density is Gaussian, more precisely[1]

$$\varphi_{k|k-1}(x|X, w) = \mathcal{N}(x; (F_{k|k-1} \otimes I_d)w, D_{k|k-1} \otimes X) \tag{7.18}$$

Now, by assuming that the corrected PDF for the kinematic state is Gaussian with the following structure

$$p_{k-1|k-1}(x|X) \triangleq \mathcal{N}(x; x_{k-1|k-1}, P_{k-1|k-1} \otimes X) \tag{7.19}$$

for given $x_{k-1|k-1}$, $P_{k-1|k-1}$, turns out that the predicted PDF for the kinematic state is still Gaussian

$$\begin{aligned} p_{k|k-1}(x|X) &\triangleq \int \varphi_{k|k-1}(x|X, w)\, p_{k-1|k-1}(w|X) \;\mathrm{d}w \\ &= \mathcal{N}(x; x_{k|k-1}, P_{k|k-1} \otimes X) \end{aligned} \tag{7.20}$$

where the predicted parameters are given by the following Kalman predictor

$$\begin{aligned} x_{k|k-1} &\triangleq (F_{k|k-1} \otimes I_d)\, x_{k-1|k-1} \\ P_{k|k-1} &\triangleq D_{k|k-1} + F_{k|k-1}\, P_{k-1|k-1} F'_{k|k-1} \end{aligned} \tag{7.21}$$

7.3.2 Shape state prediction

The prediction for the extension state is a simple heuristic, where it is postulated that the shape state is inverse Wishart, which is the typical distribution used in the multivariate statistic to represent SPD random matrices. More precisely, it is assumed that the corrected PDF of the shape state is inverse Wishart with $\nu_{k-1|k-1}$ degrees of freedom and parameter matrix $X_{k-1|k-1}$, so

$$p_{k-1|k-1}(X) \triangleq \mathcal{IW}_d(X; \nu_{k-1|k-1}, X_{k-1|k-1}) \tag{7.22}$$

while the predicted PDF of the extension state is inverse Wishart with $\nu_{k|k-1}$ degrees of freedom and $X_{k|k-1}$ scale matrix

$$p_{k|k-1}(X) \triangleq \mathcal{IW}_d(X; \nu_{k|k-1}, X_{k|k-1}) \tag{7.23}$$

[1] w that figures in φ is not the process noise of the dynamic model, but a generic state at time $k-1$

the prediction consists only in the time-update of the parameters of the predicted inverse Wishart, starting from the parameters of the corrected parameters of the inverse Wishart. The models considered are the following:

$$\begin{aligned} \nu_{k|k-1} &\triangleq \exp\left(-T/\tau\right)\nu_{k-1|k-1} \\ X_{k|k-1} &\triangleq \frac{\nu_{k|k-1}-d-1}{\nu_{k-1|k-1}-d-1}X_{k-1|k-1} \end{aligned} \tag{7.24}$$

where τ is an hyperparameter. The reasons why the actual models are choosen stem from the following facts:

- **time-update of the degrees of freedom**: the degrees of freedom of an inverse Wishart are related to the precision of the corresponding expectation. Since the prediction step is the operation that increase the uncertainties, the precision shall decrease as the time T between two consecutive measurement updates increase. The exponential model considered takes into account this fact, and the decay parameter τ represents how much sensitive is the prediction with respect the absolute value of T.
- **time-update of the scale matrix**: the second equation of (47) simply states that the predicted expected value of the shape matrix is the same as the corrected expected value of the shape matrix, in fact (47) is equivalent to

$$\frac{X_{k|k-1}}{\nu_{k|k-1}-d-1} = \frac{X_{k-1|k-1}}{\nu_{k-1|k-1}-d-1} \quad \equiv \quad \mathbb{E}_{k|k-1}[\boldsymbol{X}] = \mathbb{E}_{k-1|k-1}[\boldsymbol{X}] \tag{7.25}$$

in other words, the time-update considered states that it is believed that the shape matrix doesn't change too much between a time step and another. This is reasonable if the sampling interval T is "small" with respect the dynamic of the object tracked. In fact, despite the fact that the width and the length of an extended object are reasonably fixed in time, the orientation can change a lot between a sampling interval and another, according to the dynamic behaviour of the object tracked.

7.3.3 Joint kinematic-shape prediction

According to the previous results for the prediction of the kinematic state and the shape state and the joint predicted density is given by

$$\begin{aligned} p_{k|k-1}(x,X) &= p_{k|k-1}(x|X)\,p_{k|k-1}(X) \\ &= \mathcal{N}(x;x_{k|k-1},P_{k|k-1}\otimes X)\,\mathcal{IW}(X;\nu_{k|k-1},X_{k|k-1}) \\ &\triangleq \mathcal{NIW}(x,X;x_{k|k-1},P_{k|k-1},\nu_{k|k-1},X_{k|k-1}) \end{aligned} \tag{7.26}$$

where $\mathcal{NIW}(\mu, \Lambda, \nu, \Psi)$ denotes the so-called *Gaussian-inverse-Wishart* probability density, defined over the product space $\mathbb{R}^{s \cdot d} \times \mathbb{S}_+^{d \times d}$, where $\mathbb{S}_+$ is the the space of the SPD matrices, and characterized by the four parameters $\mu \in \mathbb{R}^{s \cdot d}$, $\Lambda \in \mathbb{S}_+^{s \times s}$, $\nu \in \mathbb{R}$, $\Psi \in \mathbb{S}_+^{d \times d}$.

7.4 GIW corrector

7.4.1 Measument model

Let $\boldsymbol{y}_k \in \mathbb{R}^p$, where $p = \tilde{s} \cdot d$ for some $\tilde{s}$ representing how many variables are observed in one dimension. The measurement model considered is the standard linear and Gaussian model

$$\boldsymbol{y}_k = C_k \, \boldsymbol{x}_k + \boldsymbol{v}_k \tag{7.27}$$

where the d-dimensional observation matrix $C_k \in \mathbb{R}^{p \times s \cdot d}$ is given by

$$C_k \triangleq H_k \otimes I_d \tag{7.28}$$

for a 1-dimension observation matrix $H_k \in \mathbb{R}^{\tilde{s} \times s}$. The measurement noise $\boldsymbol{v}_k$ is assumed to be zero mean Gaussian with covariance equal to the shape matrix

$$\boldsymbol{v}_k \sim \mathcal{N}(0, \boldsymbol{X}_k) \tag{7.29}$$

this choice for the measurement noise arise from the following considerations.

1. in principle, every measures are scattered around the object because of the random measurement error (whose covariance is usually denoted as R_k) and because the reflection points of the objects are randomly illuminated by the sensors. In order to take into account this two phenomena one can define the simple model for the effective power of the measurement noise as

$$R_k + \boldsymbol{X}_k \tag{7.30}$$

   where the $\boldsymbol{X}_k$ takes into account the fact that bigger is the object and bigger is the distance between the reflection points, so larger is the "observed" distance between the measures.

2. due to equation (59), one can see the measures, which are generated by the individual reflection points with a measurement error R_k, as generated (only) by the centroid with an equivalent power $R_k + \boldsymbol{X}_k$.

3. assuming that the object extension is much bigger than the imprecision of the sensors, i.e. $\boldsymbol{X}_k \gg R_k$, the equivalent power of the measurement noise reduces to $\boldsymbol{X}_k$, so the actual model is taken in consideration.

7.4.2 Likelihood

Since the measurement model considered is linear and Gaussian, the single measure likelihood is Gaussian as well, in particular

$$\ell_k(y|x, X) = \mathcal{N}(y; (H_k \otimes I_d)x, X) \tag{7.31}$$

as a consequence, the joint likelihood of the set of measures $\mathsf{y}_k \triangleq \{y_k^1, \dots, y_k^{n_k}\}$ assumes the form

$$\ell_k(\mathsf{y}|x, X) = \prod_{i=1}^{n_k} \mathcal{N}(y^i; C_k x, X) \tag{7.32}$$

as one can show after some simple algebraic manipulations, such likelihood can be factorized as follows

$$\ell_k(\mathsf{y}|x, X) \propto \mathcal{N}\left(\bar{y}; (H_k \otimes I_d)x, \frac{X}{n_k}\right) \mathcal{W}_p\left(\overline{Y}; n_k - 1, X\right) \tag{7.33}$$

where are introduced the following statistics of the measurement set y_k

- **mean measure**: defined as

$$\bar{y} \triangleq \frac{1}{n_k} \sum_{i=1}^{n_k} y_k^i \tag{7.34}$$

  its likelihood is the Gaussian $\mathcal{N}\left(\bar{y}; (H_k \otimes I_d)x, \frac{X}{n_k}\right)$;

- **scatter matrix**: defined as

$$\overline{Y} \triangleq \sum_{i=1}^{n_k} (y_k^i - \bar{y})(y_k^i - \bar{y})' \tag{7.35}$$

  its likelihood is given by the Wishart $\mathcal{W}_p\left(\overline{Y}; n_k - 1, X\right)$.

7.4.3 Correction

Given the predicted joint density, the corrected density is given by the Bayes equation. For simplicity, only the numerator of the Bayes equation is discussed and the normalizing factor is considered absorbed by the proportionality sign. The corrected density can be factorized as follows

$$\begin{aligned} p_{k|k}(x, X) &\propto \ell_k(\mathsf{y}|x, X)\, p_{k|k-1}(x, X) \\ &\propto \left[\mathcal{N}\left(\bar{y}; C_k x, \frac{X}{n_k}\right) \mathcal{N}\left(x; x_{k|k-1}, P_{k|k-1} \otimes X\right)\right] \\ &\quad \times \left[\mathcal{LW}\left(\bar{Y}; n_k - 1, X\right) \mathcal{IW}(X; \nu_{k|k-1}, X_{k|k-1})\right] \end{aligned} \tag{7.36}$$

- **Gaussians product**: due to the fundamental Gaussian identity, the first two factors give

$$\mathcal{N}\left(\bar{y}; C_k x, \frac{X}{n_k}\right) \mathcal{N}\left(x; x_{k|k-1}, P_{k|k-1} \otimes X\right) = \mathcal{N}\left(\bar{y}; \bar{y}_{k|k-1}, S_k X\right) \mathcal{N}(x; x_{k|k}, P_{k|k} \otimes X) \tag{7.37}$$

where, as always, the $\bar{y}_{k|k-1}$, Σ_k are given by the Kalman predictor

$$\begin{aligned} \bar{y}_{k|k-1} &\triangleq (H_k \otimes I_d) x_{k|k-1} \\ S_k &\triangleq \frac{1}{n_k} + H_k\, P_{k|k-1} H_k' \end{aligned} \tag{7.38}$$

while $x_{k|k}$ and $\Pi_{k|k}$ are given by the Kalman corrector

$$\begin{aligned} L_k &\triangleq P_{k|k-1} H_k S_k^{-1} \\ P_{k|k} &\triangleq (I - L_k H_k) P_{k|k-1} \\ x_{k|k} &\triangleq x_{k|k-1} + (L_k \otimes I_d)(\bar{y} - \bar{y}_{k|k-1}) \end{aligned} \tag{7.39}$$

now note that the first factor on the RHS doesn't depends on the kinematic state $\boldsymbol{x}_k$ but only on the shape state $\boldsymbol{X}_k$. Due to this fact, it is convenient to express such factor in the following pseudo inverse Wishart form

$$\mathcal{N}\left(\bar{y}; \bar{y}_{k|k-1}, S_k X\right) \propto |X|^{-\frac{1}{2}} \operatorname{etr}\left(-\frac{1}{2} N_k X^{-1}\right) \tag{7.40}$$

where it is introduced the spread matrix

$$N_k \triangleq (\bar{y} - \bar{y}_{k|k-1})(\bar{y} - \bar{y}_{k|k-1})' S_k^{-1} \tag{7.41}$$

in conclusion,

$$\mathcal{N}\left(\bar{y}; C_k x, \frac{X}{n_k}\right) \mathcal{N}\left(x; x_{k|k-1}, P_{k|k-1} \otimes X\right) = |X|^{-\frac{1}{2}} \operatorname{etr}\left(-\frac{1}{2} N_k X^{-1}\right) \mathcal{N}(x; x_{k|k}, P_{k|k} \otimes X) \tag{7.42}$$

- **Complete product**: the corrected density gets the form

$$p_{k|k}(x, X) = \mathcal{N}(x; x_{k|k}, P_{k|k} \otimes X)\, F(X) \tag{7.43}$$

where

$$
\begin{aligned}
F(X) \triangleq |X_k|^{-\frac{1}{2}} \operatorname{etr}\left(-\frac{1}{2} N_k X^{-1}\right) \\
\times \mathcal{W}_p\left(\bar{Y}; n_k - 1, X\right) \mathcal{IW}_d(X; \nu_{k|k-1}, X_{k|k-1})
\end{aligned}
\tag{7.44}
$$

after some elementary calculations, turns out that

$$
F(X) \propto \mathcal{IW}_d\left(X; \nu_{k|k}, X_{k|k}\right) \tag{7.45}
$$

where are introduced the corrected parameters

$$
\begin{aligned}
\nu_{k|k} &\triangleq \nu_{k|k-1} + n_k \\
X_{k|k} &\triangleq X_{k|k-1} + N_k + \bar{Y}
\end{aligned}
\tag{7.46}
$$

in conclusion, the corrected density has the following final Gaussian inverse Wishart form

$$
p_{k|k}(x, X) \propto \mathcal{N}(x; x_{k|k}, P_{k|k} \otimes X)\, \mathcal{IW}_d\left(X; \nu_{k|k}, X_{k|k}\right) \tag{7.47}
$$

7.4.4 Joint kinematic-shape correction

The corrected density is proportional to a Gaussian inverse Wishart

$$
p_{k|k}(x, X) \propto \mathcal{NIW}\left(x; x_{k|k}, P_{k|k}, \nu_{k|k}, X_{k|k}\right) \tag{7.48}
$$

where the Gaussian parameters are given by the Kalman corrector (wich acts on the measures $\bar{y}$ of to the object centroid)

$$
\begin{aligned}
L_k &\triangleq P_{k|k-1} H_k S_k^{-1} \\
P_{k|k} &\triangleq (I_s - L_k H_k) P_{k|k-1} \\
x_{k|k} &\triangleq x_{k|k-1} + L_k[\bar{y} - (H_k \otimes I_d) x_{k|k-1}]
\end{aligned}
\tag{7.49}
$$

and the inverse Wishart parameters are given by

$$
\begin{aligned}
\nu_{k|k} &\triangleq \nu_{k|k-1} + n_k \\
X_{k|k} &\triangleq X_{k|k-1} + N_k + \bar{Y}
\end{aligned}
\tag{7.50}
$$

7.5 GIW estimates

Given the corrected density $\mathcal{NIW}\left(x_{k|k}, P_{k|k}, \nu_{k|k}, X_{k|k}\right)$, the kinematic state and the shape state are estimated as follows

$$
\begin{aligned}
\hat{x}_{k|k} &\triangleq x_{k|k} \\
\hat{X}_{k|k} &\triangleq \frac{X_{k|k}}{\nu_{k|k} - 2p - 2}
\end{aligned}
\tag{7.51}
$$

moreover the covariance of the kinematic state is

$$\Pi_{k|k} \triangleq P_{k|k} \otimes \hat{X}_{k|k} \tag{7.52}$$

7.6 PHD implementation

7.6.1 GIW-PHD model

By denoting the augmented true state of the generic object as the $(s \cdot d + d^2) \times 1$ column vector

$$\xi_k \triangleq \left[x'_k \quad (\mathrm{vech}[X_k])' \right]' \tag{7.53}$$

follows that the RFS model considered by the GIW-PHD is

$$\begin{aligned} \mathsf{X}_{k+1} &= \mathsf{f}(\mathsf{X}_k) \cup \mathsf{B}_k \\ \mathsf{Y}_k &= \mathsf{h}(\mathsf{X}_k) \cup \mathsf{C}_k \end{aligned} \tag{7.54}$$

where, as usual,

- the RFS of survived objects $\mathsf{f}(\mathsf{X}_k)$ is multi-Bernoulli with parameters $\{p_{\mathrm{S}}(\xi), \varphi_{k+1|k}(\cdot|\xi)\}_{\xi \in \mathsf{X}_k}$, where $p_{\mathrm{S}}(\xi)$ is the probability that the object ξ survives and $\varphi_{k+1|k}(\xi'|\xi)$ is the probability that the object ξ, knowing that survives, moves in ξ'. Due to the linear-Gaussian motion model considered, the transition density (relative only to the kinematic state) has the Gaussian form

$$\varphi_{k|k-1}(x'|X', x, X) \triangleq \mathcal{N}\left(x'; (F_s \otimes I_d)x, D_s \otimes X'\right) \tag{7.55}$$

- the RFS of birthed objects B_k is Poisson with intensity $I_{\mathrm{B}}(\cdot)$. The simple choice of the GIW-PHD filter here is

$$I_{\mathrm{B}}(\xi) \triangleq \sum_{i=1}^{\nu_{\mathrm{B}}} w_{\mathrm{B}}^i \mathcal{NIW}\left(x, X; \hat{x}_{\mathrm{B}}^i, P_{\mathrm{B}}^i, \nu_{\mathrm{B}}^i, V_{\mathrm{B}}^i\right) \tag{7.56}$$

  where the parameters are defined by the designer of the filter;

- the RFS of detections $\mathsf{h}(\mathsf{X}_k)$ is a mixed Bernoulli-Poisson RFS with parameters $\{p_{\mathrm{D}}(\xi), I_{\mathrm{D}}(\cdot|\xi)\}_{\xi \in \mathsf{X}_k}$, where $p_{\mathrm{D}}(\xi)$ is the probability that object ξ is detected and $I_{\mathrm{D}}(\cdot|\xi)$ is the intensity of the detections produced by ξ. Since the measurement model considered is linear-Gaussian, the detection intensity is

$$I_{\mathrm{D}}(\mathsf{y}|\xi) \triangleq \lambda_{\mathrm{D}}^{|\mathsf{y}|} \prod_{y \in \mathsf{y}} \mathcal{N}\left(y; (H_{\tilde{s},s} \otimes I_d)x, X\right) \tag{7.57}$$

- the RFS of clutter measures C_k is Poisson with intensity $I_{\mathrm{C}}(\cdot)$.

7.6.2 GIW-PHD predictor

Theorem 13. Consider the GIW model and suppose that the corrected PHD at time $k-1$ is the following non-normalized mixture of Gaussian Inverse Wishart densities

$$D_{k-1|k-1}(\xi) = \sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k-1|k-1} \mathcal{NIW}\left(\xi; \hat{x}^i_{k-1|k-1}, P^i_{k-1|k-1}, \nu^i_{k-1|k-1}, V^i_{k-1|k-1}\right) \tag{7.58}$$

then the predicted PHD is given by

$$D_{k|k-1}(\xi) = I_{\mathrm{B}}(\xi) + D^{\mathrm{S}}_{k|k-1}(\xi) \tag{7.59}$$

where the PHD of survived objects is the following non-normalized mixture of Gaussian Inverse Wishart densities

$$D^{\mathrm{S}}_{k|k-1}(\xi) = \sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k|k-1} \mathcal{NIW}\left(\xi; \hat{x}^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right) \tag{7.60}$$

where the predicted weights are

$$w^i_{k|k-1} \triangleq p_{\mathrm{S}}\, w^i_{k-1|k-1} \tag{7.61}$$

the predicted parameters for the kinematic state are

$$\begin{aligned} \hat{x}^i_{k|k-1} &\triangleq (F_s \otimes I_d) x^i_{k-1|k-1} \\ P^i_{k|k-1} &\triangleq F_s\, P^i_{k-1|k-1} F_s' + D_s \end{aligned} \tag{7.62}$$

while the predicted parameters for the shape state are

$$\begin{aligned} \nu^i_{k|k-1} &\triangleq \exp(-T/\tau)\, \nu^i_{k-1|k-1} \\ V^i_{k|k-1} &\triangleq \frac{\nu_{k|k-1} - d - 1}{\nu_{k-1|k-1} - d - 1} V_{k-1|k-1} \end{aligned} \tag{7.63}$$

Proof

The objective is to compute the following PHD

$$\begin{aligned} D^{\mathrm{S}}_{k|k-1}(\xi') &= D_{k-1|k-1}\left[\tilde{\varphi}_{k|k-1}(\xi')\right] \\ &= \sum_{i=1}^{\nu_{k-1|k-1}} \underbrace{p_{\mathrm{S}}\, w^i_{k-1|k-1}}_{\triangleq w^i_{k|k-1}} \Bigg[\int \varphi_{k|k-1}(\xi'|\xi) \\ &\qquad \times \mathcal{NIW}\left(\xi; \hat{x}^i_{k-1|k-1}, P^i_{k-1|k-1}, \nu^i_{k-1|k-1}, V^i_{k-1|k-1}\right) \mathrm{d}\xi \Bigg] \end{aligned} \tag{7.64}$$

now by factorizing the transition density as

$$\varphi_{k|k-1}(\xi'|\xi) = \varphi_{k|k-1}(x'|X', x, X)\, \varphi_{k|k-1}(X'|x, X) \tag{7.65}$$

and assuming that $\varphi_{k|k-1}(X'|x, X) = \varphi_{k|k-1}(X'|X)$, follows

$$\begin{aligned} &\int \varphi_{k|k-1}(\xi'|\xi)\, \mathcal{NIW}\left(\xi; \hat{x}^i_{k-1|k-1}, P^i_{k-1|k-1}, \nu^i_{k-1|k-1}, V^i_{k-1|k-1}\right)\, \mathrm{d}\xi \\ &= \underbrace{\int \mathcal{N}\left(x'; (F_s \otimes I_d)x, D_s \otimes X'\right)\, \mathcal{N}\left(x; \hat{x}^i_{k-1|k-1}, P^i_{k-1|k-1} \otimes X\right) \mathrm{d}x}_{\text{kinematic part}} \\ &\quad \times \underbrace{\int \varphi_{k|k-1}(X'|X) \mathcal{IW}\left(X; \nu^i_{k-1|k-1}, V^i_{k-1|k-1}\right)\, \mathrm{d}X}_{\text{shape part}} \end{aligned} \tag{7.66}$$

kinematic part

with the additional assumption that $X \approx X'$, the integral involving the kinematic part can be solved by applying the foundamental Gaussian identity, yielding after some basic computations to

$$\begin{aligned} &\int \mathcal{N}\left(x'; (F_s \otimes I_d)x, D_s \otimes X'\right)\, \mathcal{N}\left(x; \hat{x}^i_{k-1|k-1}, P^i_{k-1|k-1} \otimes X\right) \mathrm{d}x \\ &\qquad = \mathcal{N}\left(x'; \hat{x}^i_{k|k-1}, P^i_{k|k-1} \otimes X'\right) \end{aligned} \tag{7.67}$$

where

$$\begin{aligned} \hat{x}^i_{k|k-1} &\triangleq (F_s \otimes I_d)\hat{x}_{k-1|k-1} \\ P^i_{k|k-1} &\triangleq F_s\, P^i_{k-1|k-1}\, F_s' + D_s \end{aligned} \tag{7.68}$$

shape part

the integral involving the extension part is heuristically defined, whatever it is the form of the transition density $\varphi_{k|k-1}(X'|X)$, to be the following inverse Wishart

$$\int \varphi_{k|k-1}(X'|X)\, \mathcal{IW}\left(X; \nu^i_{k-1|k-1}, V^i_{k-1|k-1}\right)\, \mathrm{d}X \triangleq \mathcal{IW}_d\left(X'; \nu^i_{k|k-1}, V^i_{k|k-1}\right) \tag{7.69}$$

where

$$\begin{aligned} \nu^i_{k|k-1} &\triangleq \exp(-T/\tau)\nu^i_{k-1|k-1} \\ V^i_{k|k-1} &\triangleq \frac{\nu_{k|k-1} - d - 1}{\nu_{k|k} - d - 1} V^i_{k-1|k-1} \end{aligned} \tag{7.70}$$

joint kinematic-shape state

the combination of (38), (39) and (41) yields to the Gaussian inverse Wishart density

$$\int \varphi_{k|k-1}(\xi'|\xi)\,\mathcal{NIW}\left(\xi;\hat{x}^i_{k-1|k-1},P^i_{k-1|k-1},\nu^i_{k-1|k-1},V^i_{k-1|k-1}\right)\,\mathrm{d}\xi$$
$$=\mathcal{NIW}(\xi';\hat{x}^i_{k|k-1},P^i_{k|k-1},\nu^i_{k|k-1},V^i_{k|k-1}) \tag{7.71}$$

and so, according to (36),

$$D^{\mathrm{S}}_{k|k-1}(\xi')=\sum_{i=1}^{\nu_{k-1|k-1}} w^i_{k|k-1}\,\mathcal{NIW}(\xi';\hat{x}^i_{k|k-1},P^i_{k|k-1},\nu^i_{k|k-1},V^i_{k|k-1}) \tag{7.72}$$

which complete the proof. □

7.6.3 GIW-PHD corrector

In order to simplify the derivation of the GIW-PHD corrector, the cell likelihood is derived before the complete corrector. In what follows will be assumed $\tilde{s}=1$, so only the position is measured, and denoted in short $C \triangleq H_{1,s}\otimes I_d$. Moreover, the central factorization will be expressed in the terms of the mean measure $\bar{y}_{\mathsf{w}}$ and scatter matrix $\overline{Y}_{\mathsf{w}}$ of a generic cell of measures w, i.e.

$$\prod_{y\in\mathsf{w}}\mathcal{N}\left(y;Cx,X\right)=\underbrace{\mathcal{W}_d(\overline{Y}_{\mathsf{w}},|\mathsf{w}|-1,X)}_{\triangleq\mathcal{L}_{\mathrm{aux}}}\cdot\mathcal{N}\left(\bar{y}_{\mathsf{w}};Cx,\frac{X}{|\mathsf{w}|}\right) \tag{7.73}$$

where

$$\bar{y}_{\mathsf{w}}\triangleq\frac{1}{|\mathsf{w}|}\sum_{y\in\mathsf{w}}y \qquad \bar{Y}_{\mathsf{w}}\triangleq\sum_{y\in\mathsf{w}}(y-\bar{y}_{\mathsf{w}})(y-\bar{y}_{\mathsf{w}})' \tag{7.74}$$

Theorem 14. Assumes that y is a d-dimensional position-only measure, then

$$\left[\prod_{y\in\mathsf{w}}\mathcal{N}\left(y;Cx,X\right)\right]\cdot\mathcal{NIW}\left(x,X;x^i_{k|k-1},P^i_{k|k-1},\nu^i_{k|k-1},V^i_{k|k-1}\right)=$$
$$\mathcal{L}_{\mathsf{w},i}\cdot\mathcal{NIW}\left(x,X;x^{\mathsf{w},i}_{k|k},P^{\mathsf{w},i}_{k|k},\nu^{\mathsf{w},i}_{k|k},V^{\mathsf{w},i}_{k|k}\right) \tag{7.75}$$

where the cell likelihood $\mathcal{L}_{\mathsf{w},i}$ is given by

$$\mathcal{L}_{\mathsf{w},i} \triangleq (\pi^{d|\mathsf{w}|} S_{\mathsf{w},i}\,|\mathsf{w}|)^{-\frac{1}{2}} \frac{\left(\det[V^i_{k|k-1}]\right)^{\nu^i_{k|k-1}/2}}{\left(\det[V^{\mathsf{w},i}_{k|k}]\right)^{\nu^{\mathsf{w},i}_{k|k}/2}} \frac{\Gamma_d\left(\nu^{\mathsf{w},i}_{k|k}/2\right)}{\Gamma_d\left(\nu^i_{k|k-1}/2\right)} \tag{7.76}$$

with $\Gamma_d(\cdot)$ denoting d-variate Gamma function,

$$\Gamma_d(x) \triangleq \pi^{\frac{d(d-1)}{4}} \prod_{i=1}^{d} \Gamma\left(x - \frac{i-1}{2}\right) \tag{7.77}$$

and the parameters of the Gaussian inverse Wishart density are the following:

- **kinematic part**:

$$\begin{aligned}
S_{\mathsf{w},i} &\triangleq \frac{1}{|\mathsf{w}|} + H_{1,s} P^i_{k|k-1} H'_{1,s} \\
\hat{y}^i_{k|k-1} &\triangleq C x^i_{k|k-1} \\
L_{\mathsf{w},i} &\triangleq P^i_{k|k-1} H'_{1,s}\, S^{-1}_{\mathsf{w},i} \\
P^{\mathsf{w},i}_{k|k} &\triangleq (I_s - L_{\mathsf{w},i} H_{1,s}) P^i_{k|k-1} \\
x^{\mathsf{w},i}_{k|k} &\triangleq x^i_{k|k-1} + (L_{\mathsf{w},i} \otimes I_d)\left(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1}\right)
\end{aligned} \tag{7.78}$$

- **shape part**:

$$\begin{aligned}
N_{\mathsf{w},i} &\triangleq \frac{(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1})(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1})'}{S_{\mathsf{w},i}} \\
\nu^{\mathsf{w},i}_{k|k} &\triangleq \nu^i_{k|k-1} + |\mathsf{w}| \\
V^{\mathsf{w},i}_{k|k} &\triangleq V^i_{k|k-1} + \bar{Y}_{\mathsf{w}} + N_{\mathsf{w},i}
\end{aligned} \tag{7.79}$$

Proof

Start from the followin relationship,

$$\left[\prod_{y\in\mathsf{w}} \mathcal{N}(y; Cx, X)\right] \cdot \mathcal{NIW}\left(x, X; x^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right) =$$
$$\mathcal{L}_{\text{aux}} \cdot \mathcal{N}\left(\bar{y}_{\mathsf{w}}; Cx, \frac{X}{|\mathsf{w}|}\right) \cdot \mathcal{N}\left(x; x^i_{k|k-1}, P^i_{k|k-1} \otimes X\right) \cdot \mathcal{IW}\left(X; \nu^i_{k|k-1}, V^i_{k|k-1}\right) \tag{7.80}$$

the product between the two Gaussians give rise, thanks to the fundamental Gaussian identity and the basic properties of the Kronecker product, to the factorization

$$\begin{aligned}\mathcal{N}\left(\bar{y}_{\mathsf{w}}; Cx, \frac{X}{|\mathsf{w}|}\right) \cdot \mathcal{N}\left(x; x^i_{k|k-1}; P^i_{k|k-1} \otimes X\right) \\ = \mathcal{N}\left(x; x^{\mathsf{w},i}_{k|k}, P^{\mathsf{w},i}_{k|k} \otimes X\right) \cdot \mathcal{N}\left(\bar{y}_{\mathsf{w}}; \hat{y}^i_{k|k-1}, S_{\mathsf{w},i} X\right)\end{aligned} \tag{7.81}$$

where

$$\begin{aligned} S_{\mathsf{w},i} &\triangleq \frac{1}{|\mathsf{w}|} + H_{1,s} P^i_{k|k-1} H'_{1,s} \\ \hat{y}^i_{k|k-1} &\triangleq C x^i_{k|k-1} \\ L_{\mathsf{w},i} &\triangleq P^i_{k|k-1} H'_{1,s}\, S^{-1}_{\mathsf{w},i} \\ P^{\mathsf{w},i}_{k|k} &\triangleq (I_s - L_{\mathsf{w},i} H_{1,s}) P^i_{k|k-1} \\ x^{\mathsf{w},i}_{k|k} &\triangleq x^i_{k|k-1} + (L_{\mathsf{w},i} \otimes I_d)\left(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1}\right) \end{aligned} \tag{7.82}$$

so that

$$\begin{aligned} &\left[\prod_{y \in \mathsf{w}} \mathcal{N}\left(y; (H_{1,s} \otimes I_d) x, X\right)\right] \cdot \mathcal{NIW}\left(x, X; x^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right) = \\ &\mathcal{N}\left(x; x^{\mathsf{w},i}_{k|k}, P^{\mathsf{w},i}_{k|k} \otimes X\right) \cdot \underbrace{\mathcal{N}\left(\bar{y}_{\mathsf{w}}; \hat{y}^i_{k|k-1}, S_{\mathsf{w},i} X\right) \cdot \mathcal{L}_{\text{aux}} \cdot \mathcal{IW}_d\left(X; \nu^i_{k|k-1}, V^i_{k|k-1}\right)}_{\triangleq F} \end{aligned} \tag{7.83}$$

the factor F, by using the abbreviations $\nu \triangleq \nu^i_{k|k-1}$, $V \triangleq V^i_{k|k-1}$, can be written as follows

$$F = K_1 \left(\det[X]\right)^{-\frac{|\mathsf{w}|+\nu+d+1}{2}} \operatorname{etr}\left[-\frac{1}{2}\left(N_{\mathsf{w},i} + \bar{Y}_{\mathsf{w}} + V\right) X^{-1}\right] \tag{7.84}$$

where

$$\begin{aligned} K_1 &\triangleq (2\pi)^{-\frac{d|\mathsf{w}|}{2}} \left(S_{\mathsf{w},i}\, |\mathsf{w}|\right)^{-\frac{1}{2}} \frac{\left(\det[V]\right)^{\frac{\nu}{2}}}{2^{\frac{\nu d}{2}} \Gamma_d\left(\frac{\nu}{2}\right)} \\ N_{\mathsf{w},i} &\triangleq \frac{(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1})(\bar{y}_{\mathsf{w}} - \hat{y}^i_{k|k-1})'}{S_{\mathsf{w},i}} \end{aligned} \tag{7.85}$$

now the normalization constant K_1 can be adjusted in order to write F in terms of an inverse Wishart with parameters $\nu^{\mathsf{w},i}_{k|k} \triangleq |\mathsf{w}| + \nu^i_{k|k-1}$, $V^{\mathsf{w},i}_{k|k} \triangleq$

$N_{\mathsf{w},i} + \bar{Y}_{\mathsf{w}} + V^i_{k|k-1}$. The objective is to introduce the normalizer

$$K_2 \triangleq \frac{\left(\det[N_{\mathsf{w},i} + \bar{Y}_{\mathsf{w}} + V]\right)^{\frac{|\mathsf{w}|+\nu}{2}}}{2^{\frac{(|\mathsf{w}|+\nu)\,d}{2}}\Gamma_d\left(\frac{|\mathsf{w}|+\nu}{2}\right)} \tag{7.86}$$

in the expression of the factor F, so

$$F = \underbrace{\frac{K_1}{K_2}}_{\triangleq \mathcal{L}_{\mathsf{w},i}} \underbrace{K_2 \cdot (\det[X])^{-\frac{|\mathsf{w}|+\nu+d+1}{2}} \operatorname{etr}\left[-\frac{1}{2}\left(N_{\mathsf{w},i} + \bar{Y}_{\mathsf{w}} + V\right) X^{-1}\right]}_{=\mathcal{IW}_d(X;\nu^{\mathsf{w},i}_{k|k},V^{\mathsf{w},i}_{k|k})} \tag{7.87}$$

the cell likelihood $\mathcal{L}_{\mathsf{w},i}$, after some algebra, can be written in a more clear form as follows

$$\mathcal{L}_{\mathsf{w},i} \triangleq \frac{K_1}{K_2} \quad = (\pi^{d|\mathsf{w}|} S_{\mathsf{w},i}\,|\mathsf{w}|)^{-\frac{1}{2}} \frac{\left(\det[V^i_{k|k-1}]\right)^{\nu^i_{k|k-1}/2}}{\left(\det[V^{\mathsf{w},i}_{k|k}]\right)^{\nu^{\mathsf{w},i}_{k|k}/2}} \frac{\Gamma_d\left(\nu^{\mathsf{w},i}_{k|k}/2\right)}{\Gamma_d\left(\nu^i_{k|k-1}/2\right)} \tag{7.88}$$

□

Theorem 15. Consider the GIW-PHD model and suppose that the predicted PHD at time $k-1$ is the following non-normalized mixture of Gaussian inverse Wishart densities

$$D_{k|k-1}(\xi) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k-1} \mathcal{NIW}\left(\xi; \hat{x}^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right) \tag{7.89}$$

and assume that all the heuristic considerations of the GIW filter hold (including the hypothesis that y is a d-dimensional vector - so position-only measures are considered). Then the corrected PHD is given by

$$D_{k|k}(\xi) = D^{\mathrm{ND}}_{k|k-1}(\xi) + \sum_{\mathcal{P}\boxminus\mathsf{y}} \sum_{\mathsf{w}\in\mathcal{P}} D^{\mathrm{D}}_{k|k-1}(\xi, \mathsf{w}) \tag{7.90}$$

where the PHD of undetected objects is the following mixture of Gaussian inverse Wishart densities

$$D^{\mathrm{ND}}_{k|k-1}(\xi) = \sum_{i=1}^{\nu_{k|k-1}} w^i_{k|k} \mathcal{NIW}\left(\xi; \hat{x}^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right) \tag{7.91}$$

with

$$w^i_{k|k} \triangleq \left[1 - p_{\mathrm{D}} \cdot (1 - \exp(-\lambda_{\mathrm{D}}))\right] w^i_{k|k-1} \tag{7.92}$$

and the PHD of detected objects in w is

$$D^{\mathrm{D}}_{k|k-1}(\xi, \mathsf{w}) = \sum_{i=1}^{\nu_{k|k-1}} w^{i,\mathsf{w}}_{k|k} \mathcal{NIW}\left(\xi; \hat{x}^{i,\mathsf{w}}_{k|k}, P^{i,\mathsf{w}}_{k|k}, \nu^{i,\mathsf{w}}_{k|k}, V^{i,\mathsf{w}}_{k|k}\right) \tag{7.93}$$

with

$$\begin{aligned} w^{\mathsf{w},i}_{k|k} &\triangleq \omega_{\mathcal{P}} \frac{\tilde{\ell}_{\mathsf{w},i}}{d_{\mathsf{w}}} \\ \tilde{\ell}_{\mathsf{w},i} &\triangleq p_{\mathrm{D}} \cdot \exp(-\lambda_{\mathrm{D}}) \cdot \left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|} \cdot \mathcal{L}_{\mathsf{w},i} \cdot w^{i}_{k|k-1} \\ d_{\mathsf{w}} &\triangleq \delta_1(|\mathsf{w}|) + \sum_{i=1}^{\nu_{k|k-1}} \tilde{\ell}_{\mathsf{w},i} \\ \omega_{\mathcal{P}} &\triangleq \frac{\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}}{\sum_{\mathcal{P}\boxminus\mathsf{y}} \prod_{\mathsf{w}'\in\mathcal{P}} d_{\mathsf{w}'}} \end{aligned} \tag{7.94}$$

Proof

The PHD of non-detected object is trivivally given by

$$\begin{aligned} D^{\mathrm{ND}}_{k|k}(\xi) &= \Lambda^{\mathrm{ND}}(\mathsf{y}|\xi)\, D_{k|k-1}(\xi) \\ &= \sum_{i=1}^{\nu_{k|k-1}} \underbrace{\left[1 - p_{\mathrm{D}} \cdot (1 - \exp(-\lambda_{\mathrm{D}}))\right] w^{i}_{k|k-1}}_{\triangleq w^{i}_{k|k}} \\ &\quad \times \mathcal{NIW}\left(\xi; \hat{x}^{i}_{k|k-1}, P^{i}_{k|k-1}, \nu^{i}_{k|k-1}, V^{i}_{k|k-1}\right) \end{aligned} \tag{7.95}$$

Now the turn of the detected PHD,

$$\begin{aligned} D^{\mathrm{D}}_{k|k}(\xi) &= \Lambda^{\mathrm{D}}(\mathsf{y}|\xi)\, D_{k|k-1}(\xi) \\ &= \sum_{\mathcal{P}\boxminus\mathsf{y}} \sum_{\mathsf{w}\in\mathcal{P}} \underbrace{\sum_{i=1}^{\nu_{k|k-1}} \omega_{\mathcal{P}} \frac{\tilde{\ell}_{\mathsf{w}}\, w^{i}_{k|k-1} \mathcal{NIW}\left(\xi; \hat{x}^{i}_{k|k-1}, P^{i}_{k|k-1}, \nu_{k|k-1}, V_{k|k-1}\right)}{d_{\mathsf{w}}}}_{\triangleq D^{\mathrm{D}}_{k|k}(\xi,\mathsf{w})} \end{aligned} \tag{7.96}$$

focus on the product between the likelihood $\tilde{\ell}_{\mathsf{w}}$ and the Gaussian inverse Wishart density inside the summation in i. For the GIW model holds

$$\begin{aligned} \tilde{\ell}_{\mathsf{w}} &= p_{\mathrm{D}} \cdot \exp(-\lambda_{\mathrm{D}}) \cdot \prod_{y\in\mathsf{w}} \frac{\lambda_{\mathrm{D}} \cdot \mathcal{N}(y; Cx, X)}{I_{\mathrm{C}}} \\ &= p_{\mathrm{D}} \cdot \exp(-\lambda_{\mathrm{D}}) \left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|} \cdot \prod_{y\in\mathsf{w}} \mathcal{N}(y; Cx, X) \end{aligned} \tag{7.97}$$

consequently,

$$\begin{aligned}&\tilde{\ell}_{\mathsf{w}} \cdot \mathcal{NIW}\left(\xi; x^i_{k|k-1}, P^i_{k|k-1}, \nu^i_{k|k-1}, V^i_{k|k-1}\right)\\&= p_{\mathrm{D}} \cdot \exp\left(-\lambda_{\mathrm{D}}\right) \left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|} \cdot \mathcal{L}_{\mathsf{w},i} \cdot \mathcal{NIW}\left(\xi; x^{\mathsf{w},i}_{k|k}, P^{\mathsf{w},i}_{k|k-1}, \nu^{\mathsf{w},i}_{k|k-1}, V^{\mathsf{w},i}_{k|k-1}\right)\end{aligned} \tag{7.98}$$

which permits to write the detected PHD as

$$D^{\mathrm{D}}_{k|k}(\xi, \mathsf{w}) = \sum_{i=1}^{\nu_{k|k-1}} \omega_{\mathcal{P}} \frac{\tilde{\ell}_{\mathsf{w},i}}{d_{\mathsf{w}}} \cdot \mathcal{NIW}\left(\xi; x^{\mathsf{w},i}_{k|k}, P^{\mathsf{w},i}_{k|k-1}, \nu^{\mathsf{w},i}_{k|k-1}, V^{\mathsf{w},i}_{k|k-1}\right) \tag{7.99}$$

with

$$\tilde{\ell}_{\mathsf{w},i} \triangleq p_{\mathrm{D}} \cdot \exp\left(-\lambda_{\mathrm{D}}\right) \left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|} \cdot \mathcal{L}_{\mathsf{w},i} \cdot w^i_{k|k-1} \tag{7.100}$$

and

$$d_{\mathsf{w}} = \delta_1(|\mathsf{w}|) + \sum_{i=1}^{\nu_{k|k-1}} \tilde{\ell}_{\mathsf{w},i} \tag{7.101}$$

□

Chapter 8

MEM-EKF* filter

8.1 Summary

In this chapter is discussed the second type of filter for extended objects that is capable to estimate the shape of the tracked object, i.e. the MEM-EKF* filter. Unlike the GIW filter, which was born to track the flocks of multiple independent point objects[1] (the estimated ellipse, in fact, represents the shape of the flock), the MEM-EKF* is specifically designed to track extended object. This fact can be seen from the following properties, that are reflecting the dynamic of an extended object, of the MEM-EKF* filter:

- the variance of the position of the estimated ellipse is independent on its area. For the GIW filter such variance is proportional to the area of the estimated ellipse;
- the variance of the orientation angle of the estimate ellipse can be choose independently from the variance of its two radii. For the GIW filter such variances are equals;

As a consequence, the MEM-EKF* filter achieves better performance than the GIW filter when this two algorithms are compared in an estimation problem involing an extended object (rather than a cluster object).

The chapter is structured as follows

- in the first part are introduced preliminary concepts, such a how the MEM-EKF* represents the shape of an extended object and the what type of measurement vectors are employed to perform the correction step.

[1] in literature this type of estimand is usually called *cluster object*

- then are discussed the standard motion and measurement model and the single extended object predictor and corrector;
- in the final part is introduced an improved and more general version of the standard MEM-EKF* (which differs from the standard version in the motion model) deviced to track manuevering objects and after that the PHD extension of the general MEM-EKF* filter.

In the first part of the chapter is discussed how the MEM-EKF* represents mathematically the shape of the object and

8.2 Shape parametrization

One of the main difference between the GIW filter and the MEM-EKF* filter is that the shape of an object is not parametrized by the independent entries of the shape matrix but it is parametrized by the director angles and the diameters of the representing ellipsoide.

For example consider the planar case, here the shape parameters of the GIW filter are $\boldsymbol{X}_{11}$, $\boldsymbol{X}_{12}$, $\boldsymbol{X}_{22}$ forming the shape matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \cdot & \boldsymbol{X}_{22} \end{bmatrix} \tag{8.1}$$

while the shape parameters of the MEM-EKF* filter are:

- $\boldsymbol{l}_1 \in \mathbb{R}^+$, the biggest diameter of the representing ellipse, called *lenght* of the object. Since an extended object is assumed to be rigid, the real length is a quantity that it is considered fixed in time. From the point of view of the estimation, the length is well represented by a random variable with a small variance;
- $\boldsymbol{l}_2 \in \mathbb{R}^+$, the smallest diameter of the representing ellipse, called *width* of the object. Once again, due to rigidity of the objects, the real width is a quantity that it is considered fixed in time as well. Likewise the length, the width is well represented by a random variable with a small variance;
- $\boldsymbol{\theta} \in [0, 2\pi]$, the angle between the horizontal axis of the reference frame and the lenght of the representing ellipse, called *orientation* of the object. Since an extended object can moves in space, the real orientation, in the same manner as the real position of the object, is a time varying quantity. From the point of view of the estimation, the orientation is well represented by a random variable with large variance.

In what follows, it will be always assumed for simplicity that the tracking problem is planar. However, with some effort the MEM-EKF* filter can be adjusted to solve also tracking 3-dimensional problems.

The relationship between $\boldsymbol{X}$ and the MEM-EKF* parameters f $\boldsymbol{\theta}$, $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ is clear in the planar case: since $\boldsymbol{X}$ is symmetric, it is possible to compute its spectral decomposition.

$$\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' \tag{8.2}$$

where $\boldsymbol{V}$ is an orthonormal matrix, and thus, due to planar assumption, can be expressed as

$$\boldsymbol{V} = \begin{bmatrix} \cos\boldsymbol{\theta} & -\sin\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} \end{bmatrix} \tag{8.3}$$

while $\boldsymbol{\Lambda}$ is diagonal matrix whose diagonal elements are the eigenvalues $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$ of $\boldsymbol{X}$,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 & 0 \\ 0 & \boldsymbol{\lambda}_2 \end{bmatrix}. \tag{8.4}$$

Now, since $\boldsymbol{X} \geq 0$, the eigenvalues are $\lambda_1, \lambda_2 \geq 0$ and consequently the spectral decomposition of $\boldsymbol{X}$ can be also expressed in the following form

$$\boldsymbol{X} = \boldsymbol{S}\boldsymbol{S}' \tag{8.5}$$

where it is introduced the *profile matrix*

$$\boldsymbol{S} \triangleq \boldsymbol{V}\sqrt{\boldsymbol{\Lambda}} = \begin{bmatrix} \cos\boldsymbol{\theta} & -\sin\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} \sqrt{\boldsymbol{\lambda}_1} & 0 \\ 0 & \sqrt{\boldsymbol{\lambda}_2} \end{bmatrix} \tag{8.6}$$

the random matrix $\boldsymbol{S}$ depends on $\boldsymbol{\theta}$, which is the orientation angle, and $\sqrt{\boldsymbol{\lambda}_1}$, $\sqrt{\boldsymbol{\lambda}_2}$, which are the length and width $\boldsymbol{l}_1$, $\boldsymbol{l}_2$. In conclusion, the MEM-EKF* filter expresses $\boldsymbol{X}$ in terms of $\boldsymbol{\theta}$, $\boldsymbol{l}_1$, $\boldsymbol{l}_2$.

Thanks to the special parametrization choosen, the MEM-EKF* filter can track separately the orientation, the lenght and the width of an extended object. This feature, in concjunction to the fact that in a typical scenario $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ are fixed in time while $\boldsymbol{\theta}$ is time varying, permits the MEM-EKF* to achieve better performace than the GIW filter.

However there is a price to be paid, which consists in the introduction of a new and highly non-linear measurement model called *multiplicative error model* (MEM).

8.3 Multiplicative error model

In this section is derived the MEM model, which expresses how a measure vector $\boldsymbol{y}$ is generated by the tracked object .

The MEM model assumes that $\boldsymbol{y} \in \mathbb{R}^2$ is a position-only measure expressed in cartesian coordinates. Moreover, due to the fact that the object tracked is assumed to be extended, the MEM-EKF* assumes that at every time step are available multiple measures $\boldsymbol{y}_k^1, \dots \boldsymbol{y}_k^{n_k}$, where n_k is assumed to be known. Every measure is assumed to be originated by a random point $\boldsymbol{z}$ on the surface of the object and corrupted by an additive Gaussian noise $\boldsymbol{v}$.

$$\begin{aligned} \boldsymbol{y}_k &= \boldsymbol{z}_k + \boldsymbol{v}_k \\ \boldsymbol{v}_k &\sim \mathcal{N}(0, R_v) \end{aligned} \tag{8.7}$$

the problem is to relate $\boldsymbol{z}$ with the shape parameters $\boldsymbol{l}_1$, $\boldsymbol{l}_2$, $\boldsymbol{\theta}$. In the next subsections this task is achieved step-by-step by considering different shape models.

8.3.1 Unit disc

Assume that the object is positioned in the origin of the reference frame and its shape is a disc with unitary radius. In this case a random point $\boldsymbol{z} \in \mathbb{R}^2$ on the object surface can be expressed as follows

$$\underbrace{\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix}}_{\boldsymbol{h}} \tag{8.8}$$

where, if $C_{0,1} \triangleq \{h \in \mathbb{R}^2 : h_1^2 + h_2^2 \leq 1\}$, the random vector $\boldsymbol{h} \in \mathbb{R}^2$ is uniformly distributed on the disc $C_{0,1}$

$$\boldsymbol{h} \sim \mathcal{U}\left(C_{0,1}\right)$$

the reason why the uniform model is choosen arise from the fact that in general in a tracking problem the points of an extended object are equally visible. The random vector $\boldsymbol{h}$ is called *multiplicative error* and its moments are

$$\mathbb{E}[\boldsymbol{h}] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mathrm{Cov}[\boldsymbol{h}] = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \triangleq R_h \tag{8.9}$$

the origin of its name will be more clear in the next subsections.

In conclusion, trivially, $\boldsymbol{z}$ is uniformly distributed over the unit disc $\mathrm{C} \triangleq \{\xi, \eta \in \mathbb{R} : \xi^2 + \eta^2 \leq 1\}$ of measurement space

$$\boldsymbol{z} \sim \mathcal{U}(\mathrm{C}) \tag{8.10}$$

which means that represents a random point of the extended object.

8.3.2 Aligned ellipsoidal disc

Assume that the object is positioned in the origin of the reference frame and its shape is a disc with radii $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ alligned to the axis of the reference frame. In this case a random point $\boldsymbol{z} \in \mathbb{R}^2$ on the object surface can be expressed as follows

$$\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix} = \begin{bmatrix} \boldsymbol{l}_1 \boldsymbol{h}_1 \\ \boldsymbol{l}_2 \boldsymbol{h}_2 \end{bmatrix} \qquad \equiv \qquad \underbrace{\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{l}_1 & 0 \\ 0 & \boldsymbol{l}_2 \end{bmatrix}}_{E(\boldsymbol{l}_1, \boldsymbol{l}_2)} \underbrace{\begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix}}_{\boldsymbol{h}} \tag{8.11}$$

note that now the multiplicative error $\boldsymbol{h}$ is scaled by the extension matrix $E(\boldsymbol{l}_1, \boldsymbol{l}_2)$: for this fact is called *multiplicative*, while is referred as an *error* because it is random.

Since $\boldsymbol{h} \sim \mathcal{U}(C_{0,1})$, the generic point $\boldsymbol{z}$ is uniformly distributed over the ellipsoidal disc with radii $\boldsymbol{l}_1$, $\boldsymbol{l}_2$

$$\begin{gathered} \boldsymbol{z} \sim \mathcal{U}(\mathrm{Ell}_{\boldsymbol{l}_1, \boldsymbol{l}_2}) \\ \mathrm{Ell}_{\boldsymbol{l}_1, \boldsymbol{l}_2} \triangleq \left\{ \xi, \eta \in \mathbb{R} : \left(\frac{\xi}{\boldsymbol{l}_1} \right)^2 + \left(\frac{\eta}{\boldsymbol{l}_2} \right)^2 \leq 1 \right\} \end{gathered} \tag{8.12}$$

thus $\boldsymbol{z}$ represents a random point of the extended object.

8.3.3 Misaligned ellipsoidal disc

Now assume that the object is still positioned in the origin and its shape is ellipsoidal, but assume that the object is misaligned with respect the reference frame according to an orientation angle $\boldsymbol{\theta}$. In this case holds

$$\begin{gathered} \begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix} = \begin{bmatrix} \cos\boldsymbol{\theta}\, \boldsymbol{l}_1 \boldsymbol{h}_1 - \sin\boldsymbol{\theta}\, \boldsymbol{l}_2 \boldsymbol{h}_2 \\ \sin\boldsymbol{\theta}\, \boldsymbol{l}_1 \boldsymbol{h}_1 + \cos\boldsymbol{\theta}\, \boldsymbol{l}_2 \boldsymbol{h}_2 \end{bmatrix} \qquad \equiv \\ \underbrace{\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} \cos\boldsymbol{\theta} & -\sin\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} \end{bmatrix}}_{R(\boldsymbol{\theta})} \underbrace{\begin{bmatrix} \boldsymbol{l}_1 & 0 \\ 0 & \boldsymbol{l}_2 \end{bmatrix}}_{E(\boldsymbol{l}_1, \boldsymbol{l}_2)} \underbrace{\begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix}}_{\boldsymbol{h}} \end{gathered} \tag{8.13}$$

and, as a consequence of the randomness of $\boldsymbol{h}$, $\boldsymbol{z}$ is uniformly distributed over the ellipsoidal disc $\mathrm{Ell}_{\boldsymbol{\theta}, \boldsymbol{l}_1, \boldsymbol{l}_2}$

$$\begin{gathered} \boldsymbol{z} \sim \mathcal{U}(\mathrm{Ell}_{\boldsymbol{\theta}, \boldsymbol{l}_1, \boldsymbol{l}_2}) \\ \mathrm{Ell}_{\boldsymbol{\theta}, \boldsymbol{l}_1, \boldsymbol{l}_2} \triangleq \left\{ \xi, \eta \in \mathbb{R} : \left(\frac{\cos\boldsymbol{\theta}\, \xi - \sin\boldsymbol{\theta}\, \eta}{\boldsymbol{l}_1} \right)^2 + \left(\frac{\sin\boldsymbol{\theta}\, \xi + \cos\boldsymbol{\theta}\, \eta}{\boldsymbol{l}_2} \right)^2 \leq 1 \right\} \end{gathered} \tag{8.14}$$

if, moreover, the object is positioned in a generic point $\boldsymbol{m} \in \mathbb{R}^2$, the model for $\boldsymbol{z}$ gets the form

$$\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix} = \begin{bmatrix} \boldsymbol{m}_\xi + \cos\boldsymbol{\theta}\, \boldsymbol{l}_1 \boldsymbol{h}_1 - \sin\boldsymbol{\theta}\, \boldsymbol{l}_2 \boldsymbol{h}_2 \\ \boldsymbol{m}_\eta + \sin\boldsymbol{\theta}\, \boldsymbol{l}_1 \boldsymbol{h}_1 + \cos\boldsymbol{\theta}\, \boldsymbol{l}_2 \boldsymbol{h}_2 \end{bmatrix} \equiv$$
$$\underbrace{\begin{bmatrix} \boldsymbol{z}_\xi \\ \boldsymbol{z}_\eta \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{m}_\xi \\ \boldsymbol{m}_\eta \end{bmatrix}}_{\boldsymbol{m}} + \underbrace{\begin{bmatrix} \cos\boldsymbol{\theta} & -\sin\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} \end{bmatrix}}_{R(\boldsymbol{\theta})} \underbrace{\begin{bmatrix} \boldsymbol{l}_1 & 0 \\ 0 & \boldsymbol{l}_2 \end{bmatrix}}_{E(\boldsymbol{l}_1, \boldsymbol{l}_2)} \underbrace{\begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix}}_{\boldsymbol{h}} \tag{8.15}$$

from which follows that $\boldsymbol{z}$ is uniformly distributed over the ellipsoidal disc $\mathrm{Ell}_{\boldsymbol{m},\boldsymbol{\theta},\boldsymbol{l}_1,\boldsymbol{l}_2}$

$$\boldsymbol{z} \sim \mathcal{U}(\mathrm{Ell}_{\boldsymbol{m},\boldsymbol{\theta},\boldsymbol{l}_1,\boldsymbol{l}_2})$$
$$\mathrm{Ell}_{\boldsymbol{m},\boldsymbol{\theta},\boldsymbol{l}_1,\boldsymbol{l}_2} \triangleq \Bigg\{ \xi, \eta \in \mathbb{R} : \left(\frac{\cos\boldsymbol{\theta}\,(\xi - \boldsymbol{m}_\xi) - \sin\boldsymbol{\theta}\,(\eta - \boldsymbol{m}_\eta)}{\boldsymbol{l}_1} \right)^2 + \left(\frac{\sin\boldsymbol{\theta}\,(\xi - \boldsymbol{m}_\xi) + \cos\boldsymbol{\theta}\,(\eta - \boldsymbol{m}_\eta)}{\boldsymbol{l}_2} \right)^2 \leq 1 \Bigg\} \tag{8.16}$$

meaning that $\boldsymbol{z}$ represents a random point on the surface of the extended object considered.

8.3.4 MEM equation

The simplest form of the MEM is the following

$$\boldsymbol{y}_k = \boldsymbol{m}_k + R(\boldsymbol{\theta}_k)\, E(\boldsymbol{l}_1, \boldsymbol{l}_2)\, \boldsymbol{h}_k + \boldsymbol{v}_k \tag{8.17}$$

which can be written in a more compact and general form. By introducing the following quantities

- **kinematic state**: $2s$-dimensional vector containing the position $\boldsymbol{m}$ of the object and its first $s-1$ derivatives (whereas s is a design parameter)
$$\boldsymbol{r} \triangleq \begin{bmatrix} \boldsymbol{m}' & \dot{\boldsymbol{m}}' & \dots & (\boldsymbol{m}^{(s-1)})' \end{bmatrix}' \tag{8.18}$$
- **shape state**: 3-dimensional vector containing the shape parameters
$$\boldsymbol{p} \triangleq \begin{bmatrix} \boldsymbol{\theta} & \boldsymbol{l}_1 & \boldsymbol{l}_2 \end{bmatrix}' \tag{8.19}$$
- **observation matrix**: matrix that maps $\boldsymbol{r}$ to $\boldsymbol{m}$
$$H \triangleq \begin{bmatrix} I_2 & 0_{2\times(s-1)} \end{bmatrix} \tag{8.20}$$

- **profile matrix**: matrix that encodes the roto-dilation perfomed over the measurement error $\boldsymbol{h}$

$$S(\boldsymbol{p}) \triangleq R(\boldsymbol{\theta})\, E(\boldsymbol{l}_1, \boldsymbol{l}_2) = \begin{bmatrix} \cos\boldsymbol{\theta}\,\boldsymbol{l}_1 - \sin\boldsymbol{\theta}\,\boldsymbol{l}_2 \\ \sin\boldsymbol{\theta}\,\boldsymbol{l}_1 + \cos\boldsymbol{\theta}\,\boldsymbol{l}_2 \end{bmatrix} \tag{8.21}$$

  note that $S(\boldsymbol{p})$ is a random matrix and its distribution, which is significantly complex to be derived, is induced by the distribution of $\boldsymbol{p}$.

turns out the final expression of the MEM

$$\boldsymbol{y}_k = H\boldsymbol{r}_k + S(\boldsymbol{p}_k)\,\boldsymbol{h}_k + \boldsymbol{v}_k \tag{8.22}$$

in conclusion, the MEM splits the measure $\boldsymbol{y}_k$ in the sum of three terms:

1. the first one, $H\boldsymbol{r}_k$, expresses the location where $\boldsymbol{y}_k$ will be generated. Such location is a neighborhood of the position of the extended object, i.e. the centroid $\boldsymbol{m}_k$;
2. the second, $S(\boldsymbol{p}_k)\,\boldsymbol{h}_k$, expresses the fact the $\boldsymbol{y}_k$ can be generated by any point $\boldsymbol{z}_k$ of the extended object, which is not necessarely the centroid $\boldsymbol{m}_k$. This term encodes the shape of the object;
3. the third, $\boldsymbol{v}_k$, expresses the fact that $\boldsymbol{y}_k$ is measured with some error, so $\boldsymbol{y}_k$ does not contain the exact position $\boldsymbol{z}_k$ of a random point on the surface of extended object.

8.4 Linearized multiplicative error model

According to the MEM, the state of an extended object is the $(2s + 3)$-dimensional vector

$$\boldsymbol{x} \triangleq \begin{bmatrix} \boldsymbol{r}' & \boldsymbol{p}' \end{bmatrix}' \tag{8.23}$$

and its easy to see that the relationship between the state $\boldsymbol{x}$ and the measure $\boldsymbol{y}$, because of the shape term $S(\boldsymbol{p})$, is non-linear. Due to this fact, the MEM-EFK* corrector requires a linear approximation of the term $S(\boldsymbol{p})\boldsymbol{h}$. By considering $\hat{p}$ as the center of the linearization, follows

$$S(\boldsymbol{p})\boldsymbol{h} \approx S(\hat{p})\boldsymbol{h} + J_{\hat{p}}(\boldsymbol{p} - \hat{p}) \tag{8.24}$$

where the Jacobian J of $S(\boldsymbol{p})\boldsymbol{h}$ evaluated in $\hat{p}$, as one can show, has the following structure

$$J_{\hat{p}} \triangleq \left.\frac{\partial S(\boldsymbol{p})\boldsymbol{h}}{\partial \boldsymbol{p}}\right|_{\hat{p}} = \begin{bmatrix} \boldsymbol{h}' J_1 \\ \boldsymbol{h}' J_2 \end{bmatrix} \tag{8.25}$$

and J_1 and J_2 are respectively the Jacobians in $\hat{p} = [\hat{\theta}\ \hat{l}_1\ \hat{l}_2]'$ of the first row $S_1(\boldsymbol{p})$ and second row $S_2(\boldsymbol{p})$ of the profile matrix $S(\boldsymbol{p})$, i.e.

$$\begin{aligned} J_1 &\triangleq \left.\frac{\partial (S_1(\boldsymbol{p}))'}{\partial \boldsymbol{p}}\right|_{\hat{p}} = \frac{\partial}{\partial \boldsymbol{p}} \begin{bmatrix} \cos\boldsymbol{\theta}\,\boldsymbol{l}_1 \\ -\sin\boldsymbol{\theta}\,\boldsymbol{l}_2 \end{bmatrix}_{\hat{p}} = \begin{bmatrix} -\sin\hat{\theta}\ \hat{l}_1 & \cos\hat{\theta} & 0 \\ -\cos\hat{\theta}\ \hat{l}_2 & 0 & -\sin\hat{\theta} \end{bmatrix} \\ J_2 &\triangleq \left.\frac{\partial (S_2(\boldsymbol{p}))'}{\partial \boldsymbol{p}}\right|_{\hat{p}} = \frac{\partial}{\partial \boldsymbol{p}} \begin{bmatrix} \sin\boldsymbol{\theta}\,\boldsymbol{l}_1 \\ \cos\boldsymbol{\theta}\,\boldsymbol{l}_2 \end{bmatrix}_{\hat{p}} = \begin{bmatrix} -\cos\hat{\theta}\ \hat{l}_1 & \sin\hat{\theta} & 0 \\ -\sin\hat{\theta}\ \hat{l}_2 & 0 & \cos\hat{\theta} \end{bmatrix} \end{aligned} \tag{8.26}$$

the linear form of the MEM is

$$\boldsymbol{y}_k \approx H\boldsymbol{r}_k + S(\hat{p}_k)\,\boldsymbol{h}_k + J_{\hat{p}_k}(\boldsymbol{p}_k - \hat{p}_k) + \boldsymbol{v}_k \tag{8.27}$$

As a final remark note that, on the other hand, the GIW corrector does not require any approximation on its measurement model (which is linear by itself).

8.5 Pseudo-measurement model

Besides the 2-dimensional measure $\boldsymbol{y}$, the MEM-EKF* filter considers an additional vector $\boldsymbol{Y} \in \mathbb{R}^3$, called *pseudo-measure*, to perform the correction step. As will be shown, $\boldsymbol{y}$ will be used to get the corrected estimate of the kinematic state $\boldsymbol{r}$, while $\boldsymbol{Y}$ will be used to get the corrected estimate of the shape state $\boldsymbol{p}$. For this reason $\boldsymbol{Y}$ is introduced besides $\boldsymbol{y}$.

In the next subsections is discussed the definition of the model for $\boldsymbol{Y}$ and a an intuitive explanation about why $\boldsymbol{Y}$ is used to perform the shape correction.

8.5.1 Definition

Let $\mu_{\boldsymbol{y}} = [\mu_{\boldsymbol{y},\xi}\ \mu_{\boldsymbol{y},\eta}]'$ be the expected value of the generic measure vector $\boldsymbol{y} = [\boldsymbol{y}_\xi\ \boldsymbol{y}_\eta]'$, then the pseudo-measure $\boldsymbol{Y} \in \mathbb{R}^3$ is defined as

$$\boldsymbol{Y} \triangleq \begin{bmatrix} (\boldsymbol{y}_\xi - \mu_{\boldsymbol{y},\xi})^2 \\ (\boldsymbol{y}_\eta - \mu_{\boldsymbol{y},\eta})^2 \\ (\boldsymbol{y}_\xi - \mu_{\boldsymbol{y},\xi})(\boldsymbol{y}_\eta - \mu_{\boldsymbol{y},\eta}) \end{bmatrix} \tag{8.28}$$

in words, the pseudo measures is a vector that contains the quadratic deviations of the measure from its expected value.

Its easy to see that the pseudo-measure can be written in the following algebraic form

$$\boldsymbol{Y} = F[(\boldsymbol{y} - \mu_{\boldsymbol{y}}) \otimes (\boldsymbol{y} - \mu_{\boldsymbol{y}})] \tag{8.29}$$

where

$$F \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{8.30}$$

8.5.2 Motivation

Assume for simplicity that $R_v = 0$ (no measurement noise) and assume to know exactly the kinematic state r and the shape p of the object. Despite these positions, the measure $\boldsymbol{y}$ is still uncertain because it is not known what specific point $\boldsymbol{z} = Hr + S(p)\boldsymbol{h}$ of the object will be observed. This fact can be easily seen mathematically from the expression of MEM, which now reduces to

$$\boldsymbol{y} = Hr + S(p)\boldsymbol{h} \tag{8.31}$$

here the multiplicative error $\boldsymbol{h}$, which acts as a point selector, is still random, so $\boldsymbol{y}$ is random as well. It follows immediately that the covariance of $\boldsymbol{y}$, denoted as $\Sigma_{\boldsymbol{y}}$, is

$$\Sigma_{\boldsymbol{y}} = S(p)R_h S(p)' = \frac{S(p)S(p)'}{4} \tag{8.32}$$

On the other hand, according to the general definition of covariance matrix,

$$\Sigma_{\boldsymbol{y}} \triangleq \mathbb{E}[(\boldsymbol{y} - \mu_{\boldsymbol{y}})(\boldsymbol{y} - \mu_{\boldsymbol{y}})'] \tag{8.33}$$

thus by comparing the two expressions of $\Sigma_{\boldsymbol{y}}$ and by switching the representation from matricial to vectorial follows

$$\mathbb{E}[\boldsymbol{Y}] = \frac{F}{4} \operatorname{vec}[S(p)S(p)'] \tag{8.34}$$

which shows that the pseudo-measures $\boldsymbol{Y}$ is a random vector that is disperse around a particular transformation of the shape p of the tracked object. Hence $\boldsymbol{Y}$, up to the corrupting noise, contains information about $\boldsymbol{p}$ and thus it make sense to use it to get the corrected estimate of p.

8.6 Linearized pseudo-measurement model

It is clear that the pseudo-measure $\boldsymbol{Y}$ is related to the object state $\boldsymbol{x}$ through a non-linear function. The MEM-EKF* corrector requires also a linear approximation of the pseudo-measurement model. In order to get such approximation, start by observing that, as $\boldsymbol{h}$ and $\boldsymbol{v}$ are zero-mean noises, holds

$$\mu_{\boldsymbol{y}} = H\mu_{\boldsymbol{r}} \tag{8.35}$$

where $\mu_{\boldsymbol{r}}$ is the expected value of the kinematic state $\boldsymbol{r}$. Consequently, it follows that the non-linear relation between $\boldsymbol{Y}$ and $\boldsymbol{x}$ is

$$g(\boldsymbol{r},\boldsymbol{p}) \triangleq \begin{bmatrix} (H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi)^2 \\ (H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta)^2 \\ (H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi)(H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta) \end{bmatrix} \tag{8.36}$$

where H_1 and H_2 are the first and the second row of the observation matrix H. By applying the chain rule, turns out that the Jacobians of $g(\cdot,\cdot)$ are respectively

$$\begin{aligned} \frac{\partial g_1}{\partial \boldsymbol{r}} &= 2\left(H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi\right) H_1 \\ \frac{\partial g_2}{\partial \boldsymbol{r}} &= 2(H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta) H_2 \\ \frac{\partial g_3}{\partial \boldsymbol{r}} &= (H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi) H_2 \\ &\quad + (H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta) H_1 \\ \frac{\partial g_1}{\partial \boldsymbol{p}} &= 2(H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi)\boldsymbol{h}' J_1 \\ \frac{\partial g_2}{\partial \boldsymbol{p}} &= 2(H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta)\boldsymbol{h}' J_2 \\ \frac{\partial g_3}{\partial \boldsymbol{p}} &= (H_1(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_1(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\xi)\boldsymbol{h}' J_2 \\ &\quad + (H_2(\boldsymbol{r}-\mu_{\boldsymbol{r}}) + S_2(\boldsymbol{p})\boldsymbol{h} + \boldsymbol{v}_\eta)\boldsymbol{h}' J_1 \end{aligned} \tag{8.37}$$

thus, in conclusion, if $\hat{x} \triangleq [\hat{r}'\ \hat{p}']'$ is the center of the linearization, the linearized pseudo-measurement model is

$$\boldsymbol{Y} \approx g(\hat{x}) + \left.\frac{\partial g}{\partial \boldsymbol{x}}\right|_{\hat{x}} (\boldsymbol{x} - \hat{x}) \tag{8.38}$$

where

$$\frac{\partial g}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial g_1}{\partial \boldsymbol{r}} & \frac{\partial g_1}{\partial \boldsymbol{p}} \\ \frac{\partial g_2}{\partial \boldsymbol{r}} & \frac{\partial g_2}{\partial \boldsymbol{p}} \\ \frac{\partial g_3}{\partial \boldsymbol{r}} & \frac{\partial g_3}{\partial \boldsymbol{p}} \end{bmatrix} \tag{8.39}$$

8.7 Motion model

The motion model considered by the MEM-EKF* filter consists in two parts, which are the kinematic motion model, which tries to represent the time evolution of the position of the object, and the shape motion model, which tries to represent the time evolution of the shape of the object.

8.7.1 Kinematic motion model

It is defined as the following linear model

$$\boldsymbol{r}_{k+1} = A_r\, \boldsymbol{r}_k + \boldsymbol{w}_k^r \tag{8.40}$$

where A_r is constant matrix and $\boldsymbol{w}^r \sim \mathcal{N}(0, Q^r)$ is a zero-mean Gaussian white noise. The typical choice for the transition matrix A_r is the following

$$A_r \triangleq \mathrm{Toep}(s, T) \otimes I_2 \tag{8.41}$$

where T is the sampling interval and $\mathrm{Toep}(s, T)$ denotes a Toeplitz matrix[2] with first row $\mathrm{r}_{0,s-1,T}$ and first colum c_s defined as follows

$$\begin{aligned} \mathrm{r}_{a,b,T} &\triangleq \begin{bmatrix} \frac{T^a}{a!} & \frac{T^{a+1}}{(a+1)!} & \frac{T^{a+2}}{(a+2)!} & \cdots & \frac{T^b}{b!} \end{bmatrix} \\ \mathrm{c}_c &\triangleq \begin{bmatrix} 1 & 0_{1\times(c-1)} \end{bmatrix}' \end{aligned} \tag{8.42}$$

for example, if $s = 2$ the transition matrix gets the explicit form

$$A_r \triangleq \mathrm{Toep}(2, T) \otimes I_2 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.43}$$

in other words, the choice for the transition matrix corresponds to model the kinematic state $\boldsymbol{r}_k$ with a so-called *linear cynematic motion model.* For $s \triangleq 2$ the generic linear cynematic motion model is referred as *nearly constant velocity* (NCV) motion model, for $s \triangleq 3$ as *nearly constant acceleration* (NCA) motion model.

8.7.2 Shape motion model

It is defined as the following linear motion model

$$\boldsymbol{p}_{k+1} = A_p\, \boldsymbol{p}_k + \boldsymbol{w}_k^p \tag{8.44}$$

where A_p is constant matrix and $\boldsymbol{w}^p \sim \mathcal{N}(0, Q^p)$ is a zero-mean Gaussian white noise. The typical choice for the transition matrix is

$$A_p \triangleq I_3 \tag{8.45}$$

[2] a Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant. Consequently, a Toeplitz matrix is identified by only its first row and its first column (with the constraint that the first element of the previous row and column has to be the same)

encoding the fact that the shape parameters change slowly in time. This is particularly true for the lenght and the width variables, but for the orientation angle this model can be inaccurate if the tracked object exhibits a fast-maneuvering motion.

8.8 Measurement model

The measurement model considered by the MEM-EKF* filter consists in two parts, which are the MEM and pseudo-measurement model

$$\begin{aligned} \boldsymbol{y}_k &= H\boldsymbol{r}_k + S(\boldsymbol{p}_k)\boldsymbol{h}_k + \boldsymbol{v}_k \\ \boldsymbol{Y}_k &= F[(\boldsymbol{y}_k - \mu_{\boldsymbol{y},k}) \otimes (\boldsymbol{y}_k - \mu_{\boldsymbol{y},k})] \end{aligned} \tag{8.46}$$

the MEM-EKF* filter deals with the non-linearities of this model similarly to an extended Kalman filter because it uses the Jacobians $J_{\hat{p}}$ and $\partial g/\partial \boldsymbol{p}|_{\hat{p}}$ to compute the covariance matrices of the corrected estimates. For this reason the shorthand EKF, which indeed stands for *extended Kalman filter*, figures in the name MEM-EKF*. The shorthand MEM, clearly, stands for *multiplicative error model* and figures in the name MEM-EKF* in order to remind the special model employed to represent the relation between a measure and the state of the object.

8.9 MEM-EKF* predictor

Since the motion model is linear, the MEM-EKF* compute a standard Kalman prediction. By denoting as $\hat{r}$ and $\hat{p}$ the estimates and as P^r, P^p their covariances, the MEM-EKF* predictor gets the form

$$\begin{aligned} \hat{r}_{k|k-1} &= A_r\, \hat{r}_{k-1|k-1} \\ P^r_{k|k-1} &= A_r\, P^r_{k-1|k-1}\, A_r' + Q^r \end{aligned} \tag{8.47}$$

for the kinematic state, and gets the form

$$\begin{aligned} \hat{p}_{k|k-1} &= A_p\, \hat{p}_{k-1|k-1} \\ P^p_{k|k-1} &= A_p\, P^p_{k-1|k-1}\, A_p' + Q^p \end{aligned} \tag{8.48}$$

for the shape state. Note that, since the covariance matrix Q^p is a design parameter, the MEM-EKF* can filter out the noise from the corrected estimates $\hat{\theta}_{k|k}$, $\hat{l}_{1,k|k}$, $\hat{l}_{2,k|k}$ with different intensities. This effect can be obtained by defining, for example,

$$Q^p \triangleq \operatorname{diag}(\sigma_\theta^2, \sigma_l^2, \sigma_l^2) \tag{8.49}$$

with $\sigma_\theta^2 \neq \sigma_l^2$ (typicially $\sigma_\theta^2 \gg \sigma_l^2$). As just mentioned before, this is the most important feature that allows the MEM-EKF* filter to achieve better performance with respect the GIW filter.

8.10 MEM-EKF* corrector

The correction step is way less straightforward since the measurement model is not linear. First of all, since the tracked object is assumed to be extended, at every time step k are available n_k measures

$$\mathsf{y}_k \triangleq \{y_k^{(1)}, \ldots, y_k^{(n_k)}\} \tag{8.50}$$

which are, according to the MEM, 2-dimensional position-only measures. Likewise the GIW filter, the MEM-EKF* corrector assumes that n_k is known and that the measurements are statistically independent. Due to this positions, the MEM-EKF* corrector performs sequentially $|\mathsf{y}_k| = n_k$ single-measurement corrections,

$$\begin{array}{ccccccccc} \hat{r}_{k|k}^{(0)} \triangleq \hat{r}_{k|k-1} & \xrightarrow{y_k^{(1)}} & \hat{r}_{k|k}^{(1)} & \xrightarrow{y_k^{(2)}} & \hat{r}_{k|k}^{(2)} & \cdots & \xrightarrow{y_k^{(n_k)}} & \hat{r}_{k|k} \triangleq \hat{r}_{k|k}^{(n_k)} \\ \hat{p}_{k|k}^{(0)} \triangleq \hat{p}_{k|k-1} & & \hat{p}_{k|k}^{(1)} & & \hat{p}_{k|k}^{(2)} & & & \hat{p}_{k|k} \triangleq \hat{p}_{k|k}^{(n_k)} \end{array}$$

For each observed measurement $y_k^{(i)}$, the MEM-EKF* corrector performs two operations:

- **kinematic state correction**: compute $\hat{r}_{k|k}^{(i)}$, $P_{k|k}^{(i),r}$ according to $\hat{r}_{k|k}^{(i-1)}$, $P_{k|k-1}^{(i-1),r}$ and the observed measurement $y_k^{(i)}$;
- **shape state correction**: compute firstly $Y_k^{(i)}$ according to $y_k^{(i)}$ and $\hat{r}_{k|k}^{(i)}$, then compute $\hat{p}_{k|k}^{(i)}$, $P_{k|k}^{(i),p}$ according to $\hat{p}_{k|k-1}^{(i-1)}$, $P_{k|k-1}^{(i-1),p}$ and the observed pseudo-measurement $Y_k^{(i)}$.

8.10.1 Kinematic state correction

The correction of the kinematic state is given by the so-called *best linear unbiased estimation* (BLUE) correction equations

$$\begin{aligned} \hat{r}_{k|k}^{(i)} &= \hat{r}_{k|k-1}^{(i-1)} + \Sigma_{\boldsymbol{ry}}\, \Sigma_y^{-1}\, (y_k^{(i)} - \hat{y}_{k|k-1}^{(i-1)}) \\ P_{k|k}^{(i),r} &= P_{k|k-1}^{(i-1),r} - \Sigma_{\boldsymbol{ry}}\, \Sigma_{\boldsymbol{y}}^{-1}\, \Sigma_{\boldsymbol{ry}}' \end{aligned} \tag{8.51}$$

at his point the problem is to derive the explicit expressions of the moments $\hat{y}_{k|k-1}^{(i)}$, $\Sigma_{\boldsymbol{ry}}$, $\Sigma_{\boldsymbol{y}}$ according to the MEM.

- **predicted measurement**: by expressing equation in terms of the accumulated information available up to time k and up to measurement $y_k^{(i-1)}$ turns out

$$\hat{y}_{k|k-1}^{(i)} = H\hat{r}_{k|k-1}^{(i-1)} \tag{8.52}$$

- **measurement covariance**: according to the linearized form of the MEM, one can show that

$$\Sigma_{\boldsymbol{y}} \approx HP_{k|k-1}^{(i-1)}H' + C^I + C^{II} + R_v \tag{8.53}$$

  where

  - the first term is the covariance of the predicted measurement $\hat{y}_{k|k-1}^{(i-1)} = H\hat{r}_{k|k-1}^{(i-1)}$;
  - the second term C^I is defined as

$$C^I \triangleq \boldsymbol{S}(\hat{p}_{k|k-1}^{(i-1)})R_h\boldsymbol{S}(\hat{p}_{k|k-1}^{(i-1)})' = \frac{\boldsymbol{S}(\hat{p}_{k|k-1}^{(i-1)})\boldsymbol{S}(\hat{p}_{k|k-1}^{(i-1)})'}{4} \tag{8.54}$$

    and it is the covariance of the term $\boldsymbol{S}(\hat{p}_{k|k-1}^{(i-1)})\boldsymbol{h}_k$;
  - the third term C^{II} is defined as

$$C^{\mathrm{II}} \triangleq \begin{bmatrix} C_{11}^{\mathrm{II}} & C_{12}^{\mathrm{II}} \\ C_{21}^{\mathrm{II}} & C_{22}^{\mathrm{II}} \end{bmatrix} \quad \text{with} \quad \begin{aligned} C_{ij}^{\mathrm{II}} &\triangleq \mathrm{tr}\left[P_{k|k-1}^{(i-1),p}\, J_j' R_h\, J_i\right] \\ &= \frac{\mathrm{tr}\left[P_{k|k-1}^{(i-1),p}\, J_j' J_i\right]}{4} \end{aligned} \tag{8.55}$$

    and it is the covariance of the term $J_{\hat{p}_k}(\boldsymbol{p}_k - \hat{p}_k)$;
  - the fourth term R_v is the covariance of $\boldsymbol{v}_k$.

- **kinematic cross-covariance**: trivially is given by

$$\Sigma_{\boldsymbol{ry}} = P_{k|k-1}^{(i-1),r}H' \tag{8.56}$$

8.10.2 Shape state correction

Likewise teh kinematic state correction, the correction of the shape state is performed using the BLUE equations, with the difference that now the considered observation is the pseudo-measurement $Y_k^{(i)}$ rather than the measurement $y_k^{(i)}$. Hence, the correction equations gets the following form

$$\begin{aligned} \hat{p}_{k|k}^{(i)} &= \hat{p}_{k|k-1}^{(i-1)} + \Sigma_{\boldsymbol{pY}}\, \Sigma_{\boldsymbol{Y}}^{-1}\, (Y_k^{(i)} - \hat{Y}_{k|k-1}^{(i-1)}) \\ P_{k|k}^{(i),p} &= P_{k|k-1}^{(i-1),p} - \Sigma_{\boldsymbol{pY}}\, \Sigma_{\boldsymbol{Y}}^{-1}\, \Sigma_{\boldsymbol{pY}}' \end{aligned} \tag{8.57}$$

now the problem consists into compute the moments $\hat{Y}^{(i-1)}_{k|k-1}$, $\Sigma_{\boldsymbol{Y}}$, $\Sigma_{\boldsymbol{pY}}$ according to the pseudo-measurement model. Note that, given the observed measurement $y^{(i)}_k$ and its prediction $y^{(i-1)}_{k|k-1}$, the theoretical pseudo-measurement vector (???) reduces to the observed pseudo-measurement vector

$$Y^{(i)}_k = F[(y^{(i)}_k - y^{(i-1)}_{k|k-1}) \otimes (y^{(i)}_k - y^{(i-1)}_{k|k-1})] \tag{8.58}$$

- **predicted pseudo-measurement**:, the predicted pseudo-measurement is

$$Y^{(i)}_{k|k-1} = F\text{vec}[\Sigma_{\boldsymbol{y}}] \tag{8.59}$$

;

- **pseudo-measurement covariance**: as one can show, according to a result due to Isserli, the exact expression of the pseudo-measurement covariance is

$$\Sigma_{\boldsymbol{Y}} = F(\Sigma_{\boldsymbol{y}} \otimes \Sigma_{\boldsymbol{y}})(F + \tilde{F})' \tag{8.60}$$

where

$$\tilde{F} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{8.61}$$

- **shape cross-covariance**: according to the linearized form of the pseudo-measurement model, one can show that

$$\Sigma_{\boldsymbol{pY}} \approx P^{(i-1)}_{k|k-1} M'_{\hat{p}} \tag{8.62}$$

where

$$\begin{aligned} M_{\hat{p}} &\triangleq \begin{bmatrix} 2S_1(\hat{p}^{(i-1)}_{k|k-1})R_h J_1 \\ 2S_2(\hat{p}^{(i-1)}_{k|k-1})R_h J_2 \\ S_1(\hat{p}^{(i-1)}_{k|k-1})R_h J_2 + S_2(\hat{p}^{(i-1)}_{k|k-1})R_h J_1 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} 2S_1(\hat{p}^{(i-1)}_{k|k-1})J_1 \\ 2S_2(\hat{p}^{(i-1)}_{k|k-1})J_2 \\ S_1(\hat{p}^{(i-1)}_{k|k-1})J_2 + S_2(\hat{p}^{(i-1)}_{k|k-1})J_1 \end{bmatrix} \end{aligned} \tag{8.63}$$

8.11 MEM-EKF* filter for manuvering objects

One limitation of the standard MEM-EKF* filter is that, since it consider a linear motion model, it can not track accurately manuevering objects that change rapidly (in relation to the sampling interval T) their orientation. This problem can be solved easily by changing the definition of the kinematic state $\boldsymbol{r}$ with the introduction of the steering speed

$$\boldsymbol{\omega} \triangleq \dot{\boldsymbol{\theta}} \tag{8.64}$$

and its first $O-1$ derivatives $\dot{\boldsymbol{\omega}}, \ldots, \boldsymbol{\omega}^{(O-1)}$ as a new kinematic state variables.

8.11.1 Constant turn MEM-EKF* filter

In the simple case $O = 1$ the new definition of kinematic state is

$$\boldsymbol{r} \triangleq \begin{bmatrix} \boldsymbol{m}' & \dot{\boldsymbol{m}}' & \ldots & (\boldsymbol{m}^{(s-1)})' & \boldsymbol{\omega} \end{bmatrix}' \tag{8.65}$$

accordingly, the motion models get the new forms

$$\begin{aligned} \boldsymbol{r}_{k+1} &= A_r \boldsymbol{r}_k + \boldsymbol{w}_k^r \\ \boldsymbol{p}_{k+1} &= A_{pr} \boldsymbol{r}_k + A_p \boldsymbol{p}_k + \boldsymbol{w}_k^p \end{aligned} \tag{8.66}$$

where the transitions matrix are

$$A_r \triangleq \operatorname{diag}(\operatorname{Toep}(s,T) \otimes I_2, 1) \qquad A_p \triangleq I_3 \qquad A_{pr} \triangleq \begin{bmatrix} 0_{1\times s} & T \\ 0_{2\times s} & 0_{2\times 1} \end{bmatrix} \tag{8.67}$$

consequently, the new prediction equations are

$$\begin{aligned} \hat{r}_{k|k-1} &= A_r \hat{r}_{k-1|k-1} \\ P^r_{k|k-1} &= A_r P^r_{k-1|k-1} A_r' + Q^r \\ \hat{p}_{k|k-1} &= A_{pr} \hat{r}_{k-1|k-1} + A_p \hat{p}_{k|k-1} \\ P^p_{k|k-1} &= A_{pr} P^{rp}_{k-1|k-1} A_{pr}' + A_p P^p_{k-1|k-1} A_p' + Q^p \end{aligned} \tag{8.68}$$

on the other hand, the measurement actual model still holds but with the new convention

$$H \triangleq \begin{bmatrix} I_2 & 0_{2\times 2(s-1)} & 0 \end{bmatrix} \tag{8.69}$$

hence the new definition of the kinematic state does not change the correction equations.

The resulting algorithm is called *constant turn* (CT) MEM-EKF* filter, and can handle objects that performs manuevers characterized by steering speeds that change slowly in time.

8.11.2 General MEM-EKF* filter

If the steering speed $\boldsymbol{\omega}$ change rapidly (in relation to the sampling interval T), one can consider $O > 1$ big enough. In this case the kinematic state is

$$\boldsymbol{r} \triangleq \begin{bmatrix} \boldsymbol{m}' & \dot{\boldsymbol{m}}' & \dots & (\boldsymbol{m}^{(s-1)})' & \boldsymbol{\omega} & \dot{\boldsymbol{\omega}} & \dots & \boldsymbol{\omega}^{(O-1)} \end{bmatrix}' \tag{8.70}$$

and the new motion and measurement models still holds, with the conventions

$$\begin{gathered} A_r \triangleq \mathrm{diag}(\mathrm{Toep}(s,T) \otimes I_2, \mathrm{Toep}(O,T)) \\ A_{pr} \triangleq \begin{bmatrix} 0_{1\times s} & \mathrm{r}_{1,O-1,T} \\ 0_{2\times s} & 0_{2\times O} \end{bmatrix} \qquad A_p \triangleq I_3 \end{gathered} \tag{8.71}$$

and

$$H \triangleq \begin{bmatrix} I_2 & 0_{2\times 2(s-1)} & 0_{2\times O} \end{bmatrix} \tag{8.72}$$

the resulting algorithm is the most general case of the MEM-EKF* filter and, for this reason, is called *general* MEM-EKF* filter. This filter can track reasonably well objects that performs manuevers characterized by a $\boldsymbol{\omega}^{(O-1)}$ that changes slowly in time.

8.12 PHD implementation

While the PHD extension of the GIW filter relys on the rigourous equations of the general PHD filter for extended objects, the PHD extension of the MEM-EKF* filter follows a more direct way by considering the GM-PHD equations for extended objects.

In other words, the PHD filter based on the MEM-EKF* filter is a special GM-PHD filter for extended objects (meaning that the GM-APB-PHD corrector is employed) where the state of an object does not consist only on the kinematic variables but also on the shape variables.

8.12.1 Predictor

The state of a generic extended object is defined as

$$\boldsymbol{x} \triangleq \begin{bmatrix} \boldsymbol{r}' & \boldsymbol{p}' \end{bmatrix}' \tag{8.73}$$

and the relative motion model as

$$\begin{gathered} \boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{w}_k \sim \mathcal{N}(0, Q) \end{gathered} \tag{8.74}$$

where

$$A \triangleq \operatorname{diag}(A_r, A_p) \qquad Q \triangleq \operatorname{diag}(Q^r, Q^p) \tag{8.75}$$

consequently the transition density gets the following Gaussian form

$$\varphi_{k|k-1}(x|w) = \mathcal{N}(x; Aw, Q) \tag{8.76}$$

and so, by assuming a Gaussian mixture prior, the GM-PHD predictor (???) holds. The predicted PHD is

$$D_{k|k-1}(x) = D_{\mathrm{B}}(x) + D_{\mathrm{S}}(x) \tag{8.77}$$

where the new born objects PHD $D_{\mathrm{B}}(\cdot)$ is

$$D_{\mathrm{B}}(x) = \sum_{i=1}^{\nu_{\mathrm{B}}} w_{\mathrm{B}} \, \mathcal{N}(x; x_{\mathrm{B},i}, P_{\mathrm{B},i}) \tag{8.78}$$

while the survived objects $D_{\mathrm{S}}(\cdot)$, assuming p_{S} constant, is given by

$$\begin{aligned} D_{\mathrm{S}}(x) = & \sum_{i=1}^{\nu_{k-1|k-1}} w_{k|k-1,i} \, \mathcal{N}(x; x_{k|k-1,i}, P_{k|k-1,i}) \\ w_{k|k-1,i} & \triangleq p_{\mathrm{S}} \, w_{k-1|k-1,i} \\ x_{k|k-1,i} & \triangleq A x_{k-1|k-1,i} \\ P_{k-1|k-1,i} & \triangleq A P_{k-1|k-1,i} A' + Q \end{aligned} \tag{8.79}$$

8.12.2 Corrector

The corrector is defined more heuristically with respect to the predictor. The idea is to embed the MEM-EKF* corrector into the corrected PHD provided by the GM-APB-PHD corrector, which is

$$D_{k|k}(x) = D_{\mathrm{ND}}(x) + \sum_{\mathcal{P} \boxminus \mathsf{y}} \sum_{\mathsf{w} \in \mathcal{P}} D_{\mathrm{D}}(x; \mathcal{P}, \mathsf{w}) \tag{8.80}$$

where

- assuming p_{D} and λ_{D} constant, the non-detected objects PHD $D_{\mathrm{ND}}(\cdot)$ is

$$\begin{aligned} D_{\mathrm{ND}}(x) = & \sum_{i=1}^{\nu_{k|k-1}} w_{k|k,i} \mathcal{N}(x; x_{k|k-1,i}, P_{k|k-1,i}) \\ w_{k|k,i} & \triangleq \left[1 - (1 - \exp(-\lambda_{\mathrm{D}})) \, p_{\mathrm{D}}\right] w_{k|k-1,i} \end{aligned} \tag{8.81}$$

- the detected objects PHD $D_{\mathrm{D}}(\cdot;\mathcal{P},\mathsf{w})$ is given by

$$\begin{aligned}
D_{\mathrm{D}}(x;\mathcal{P},\mathsf{w}) &= \sum_{i=1}^{\nu_{k|k-1}} w_{k|k}^{\mathcal{P},\mathsf{w},i}\mathcal{N}(x;x_{k|k}^{\mathsf{w},i},P_{k|k}^{\mathsf{w},i}) \\
w_{k|k}^{\mathcal{P},\mathsf{w},i} &\triangleq \omega_{\mathcal{P}}\frac{\tilde{\ell}_{\mathsf{w},i}}{d_{\mathsf{w}}} \\
\tilde{\ell}_{\mathsf{w},i} &\triangleq p_{\mathrm{D}}\exp(-\lambda_{\mathrm{D}})\left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|}\mathcal{L}_{\mathsf{w},i} \\
d_{\mathsf{w}} &\triangleq \delta_1(|\mathsf{w}|)+\sum_{i=1}^{\nu_{k|k-1}}\tilde{\ell}_{\mathsf{w},i} \\
\omega_{\mathcal{P}} &\triangleq \frac{\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}}{\sum_{\mathcal{P}\boxminus y}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}}
\end{aligned} \tag{8.82}$$

  where the cell likelihood[3] is defined as

$$\mathcal{L}_{\mathsf{w},i} \triangleq \prod_{y\in\mathsf{w}}\mathcal{N}(y;y_{k|k-1}^{i},S^{i}) \tag{8.83}$$

  and the corrected parameters $x_{k|k}^{\mathsf{w},i}$, $P_{k|k}^{\mathsf{w},i}$, $y_{k|k-1}^{\mathsf{w},i}$, $S_{\mathsf{w},i}$ are given by the Kalman corrector equations (by correcting the state of object i according to cell of measures w, which is represented by the joint vector y_{w}).

The PHD extension of the MEM-EKF* changes the definitions of the cell likelihood and the corrected parameters.

Cell likelihood

Since the MEM-EKF* perform the correction according to $\boldsymbol{y}$, $\boldsymbol{Y}$, the cell likelihood is defined according to the model of the joint vector

$$\boldsymbol{\mathcal{Y}} \triangleq \begin{bmatrix}\boldsymbol{y}' & \boldsymbol{Y}'\end{bmatrix}' \tag{8.84}$$

here there as some critical observations about the model of $\mathcal{Y}$:

- the model for $\boldsymbol{y}$ is

$$p_{\boldsymbol{y}|i}(y) \triangleq \mathcal{N}(y;y_{k|k-1}^{i},\Sigma_{\boldsymbol{y}}^{i}) \tag{8.85}$$

[3]the cell likelihood is a function that measures how much likely is the event "object i have generated the cell of measurements w"

where, according to the MEM moments,

$$\begin{aligned} y^i_{k|k-1} &\triangleq H\hat{r}_{k|k-1,i} \\ \Sigma^i_{\boldsymbol{y}} &\triangleq HP^{r,i}_{k|k-1}H' + C^{I,i} + C^{II,i} + R_v \end{aligned} \tag{8.86}$$

and $\hat{r}_{k|k-1,i}$, $P^{r,i}_{k|k-1}$, $C^{I,i}$, $C^{II,i}$ are moments relative to the predicted object i. This model is reasonable if the covariance of $\boldsymbol{p}_{k|k-1,i}$ is small: in this case $\boldsymbol{y}$ is, with a good approximation, Gaussian;

- the model for $\boldsymbol{Y}$ is

$$p_{\boldsymbol{Y}|i}(Y) \triangleq \mathcal{N}(Y; Y^i_{k|k-1}, \Sigma^i_{\boldsymbol{Y}}) \tag{8.87}$$

where, according to the pseudo-measurement moments,

$$\begin{aligned} Y^i_{k|k-1} &\triangleq F\text{vec}[\Sigma^i_{\boldsymbol{y}}] \\ \Sigma^i_{\boldsymbol{Y}} &\triangleq F(\Sigma^i_{\boldsymbol{y}} \otimes \Sigma^i_{\boldsymbol{y}})(F + \tilde{F})' \end{aligned} \tag{8.88}$$

This model is a rough approximation because, clearly, $\boldsymbol{Y}$ is not Gaussian. To see why, it is sufficient to observe that the first two componenst of $\boldsymbol{Y}$ are non-negative random variables because are squared innovations. On the other hand, a Gaussian random variable can get a realization that is negative, hence $\boldsymbol{Y}$ is not Gaussian.

then, by exploiting the fact that $\boldsymbol{y}$ and $\boldsymbol{Y}$ are independently distributed (because, as one can show, $\boldsymbol{Y}$ is an uncorrelated transformation of $\boldsymbol{y}$), the joint model, which defines also the cell likelihood, gets the factorized form

$$\begin{aligned} \mathcal{L}_{\mathsf{w},i} &\triangleq p_{\mathcal{Y}|i}(\mathcal{Y}) = p_{\boldsymbol{y}|i}(y)p_{\boldsymbol{Y}|i}(Y) \\ &= \mathcal{N}\left(\mathcal{Y}; \begin{bmatrix} y^i_{k|k-1} \\ Y^i_{k|k-1} \end{bmatrix}, \begin{bmatrix} \Sigma^i_{\boldsymbol{y}} & 0 \\ 0 & \Sigma^i_{\boldsymbol{Y}} \end{bmatrix}\right) \end{aligned} \tag{8.89}$$

Corrected parameters

Finally, the $|\mathcal{P}| \cdot \nu_{k|k-1}$ corrected parameters $x^{\mathsf{w},i}_{k|k}$, $P^{\mathsf{w},i}_{k|k}$ are computed by an ensamble of $|\mathcal{P}| \cdot \nu_{k|k-1}$ MEM-EKF* filters, where each filter performs the correction of the predicted object i, which is characterized by the predicted parameters $x^{\mathsf{w},i}_{k|k-1}$, $P^{\mathsf{w},i}_{k|k-1}$, according to the cell of measurements w.

Chapter 9

LO-MEM filter

9.1 Summary

This chapter, which contains the main contribution of this thesis, is dedicated to the derivation of the LO-MEM filter, a new algorithm for the extended object tracking. In short, the LO-MEM filter is a slight modification of the MEM-EKF* filter, where are considered a new prediction model, called *Lambda:Omicron* model, and a new measurement model (always based on the MEM).

The chapter has the following structure:

- in the first part is defined the Lambda:Omicron model. Firstly it is introduced in two simple forms (the 1:0 model and the 2:1 model), then its general form is derived;
- in the second part is discussed the new measurement model, which is slight modification of the measurement model considered by the MEM-EKF* filter;
- in the final part are derived the equations of the LO-MEM predictor and corrector, then the PHD extension is discussed.

9.2 Lambda:Omicron motion model

9.2.1 Motivation

The two extended object trackers, i.e. the GIW filter and the MEM-EKF* filter, share the fact that the estimation process of the object shape depends on the estimate of the object center. In fact,

- for the GIW filter, the corrected shape matrix $\hat{X}_{k|k}$ can be written as

$$\hat{X}_{k|k} = \frac{X_{k|k-1} + (\bar{y} - \bar{y}_{k|k-1})(\bar{y} - \bar{y}_{k|k-1})' S_k^{-1} + \overline{Y}}{\nu_{k|k} - 2p - 2} \tag{9.1}$$

  which depends on $\bar{y}_{k|k-1} = Cx_{k|k-1}$, i.e. the predicted estimation of the object center;

- for the MEM-EKF* filter, the corrected shape vector $p_{k|k}$ is computed according the following BLUE equation

$$p^{(i)}_{k|k-1} = p^{(i-1)}_{k|k-1} + \Sigma_{\boldsymbol{rY}} \Sigma_{\boldsymbol{Y}}^{-1} F \left[(y^{(i)}_k - y^{(i)}_{k|k-1}) \otimes (y^{(i)}_k - y^{(i-1)}_{k|k-1}) - \Sigma_{\boldsymbol{y}} \right] \tag{9.2}$$

  which depends on $\bar{y}^{(i-1)}_{k|k-1} = Hr^{(i-1)}_{k|k-1}$, that is the predicted estimate of the object center.

If the object center is poorly estimated then, for both filters, the shape of the object cannot be estimated with accuracy. As a consequence, a key issue is the estimation of the object kinematic state because on it both localization accuracy and shape estimation depend crucially.

The GIW and the MEM-EKF* filters assume a linear motion model for the kinematic state, which implies that the tracked object moves along a line. Clearly this is a limiting factor from the point of view of the kinematic state estimation, because in general the trajectory of an object, represented by the trajectory of its center, can be curved.

Idea

The Lambda:Omicron model is a new type of motion model designed for the MEM-EKF* filter that, in order to increase the accuracy of the predictions of the kinematic state (which, as just explained, reflects also in a better estimation of the object shape), tries to represent curved trajectories.

In order to do that, the starting point is the so-called *unicycle motion model*

$$\begin{aligned}
\boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + T \begin{bmatrix} \cos \boldsymbol{\theta}_k \\ \sin \boldsymbol{\theta}_k \end{bmatrix} \boldsymbol{v}_k + \boldsymbol{w}_{m,k} \\
\boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + \boldsymbol{w}_{v,k} \\
\boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + T\boldsymbol{\omega}_k + \boldsymbol{w}_{\theta,k} \\
\boldsymbol{\omega}_{k+1} &= \boldsymbol{\omega}_k + \boldsymbol{w}_{\omega,k}
\end{aligned} \tag{9.3}$$

where:

- $\boldsymbol{m} = [\boldsymbol{\xi}\ \boldsymbol{\eta}]'$ is the object position expressed in Cartesian coordinates;

- $\boldsymbol{v} = \|\dot{\boldsymbol{m}}\|$ is magnitude of the velocity vector $\dot{\boldsymbol{m}}$;
- $\boldsymbol{\theta} = \tan^{-1}(\boldsymbol{\eta}/\boldsymbol{\xi})$ is the angle of the velocity vector $\dot{\boldsymbol{m}}$;
- $\boldsymbol{\omega} = \dot{\boldsymbol{\theta}}$ is the time-derivative of the velocity vector angle $\boldsymbol{\theta}$.

This model can be cast in the extended object tracking problem by assuming that $\boldsymbol{\theta}$ is the orientation angle of object or, in other words, that the extended object can only move along its longitudinal direction (lateral movements are not considered by the unicycle model). A limitation of the unicycle model is that, unlike a linear kinematic model, the model state is fixed to $\boldsymbol{r} \triangleq [\boldsymbol{m}' \ \boldsymbol{v} \ \boldsymbol{\theta} \ \boldsymbol{\omega}]'$. The Lambda:Omicron model tries to solve this limitation by including in the model higher order derivatives of $\boldsymbol{v}$ and $\boldsymbol{\omega}$.

9.2.2 Drawback

The major drawback of the Lambda:Omicron model is that, likewise the unicycle model, it cannot represents lateral motions. This is a clear limitation because, in general, it is not true that an object moves only along its longitudinal direction.

For example, consider a boat that moves from an edge of a river to the other one. Due to the flow of the river, the velocity of the boat is not perfectly aligned to the boat heading direction.

Despite this negative fact, in some contexts the assumption that the velocity vector is aligned to the object orientation is reasonable or at least a good first order approximation.

9.2.3 Nomenclature

The name *Lambda:Omicron*, which is directly inspired by the name *Alpha-Beta filter* used to refer a special stationary Kalman filter, arises from the fact that, as will be shown (see section 1.2.6), the new motion model consists of two independent parts:

- **Lambda model**: describes the dynamics of the longitudinal speed

$$\boldsymbol{v} \triangleq \|\dot{\boldsymbol{m}}\| \tag{9.4}$$

  where $\boldsymbol{m}$ is the object position, and its first $\Lambda - 1$ derivatives $\dot{\boldsymbol{v}}, \ldots, \boldsymbol{v}^{(\Lambda-1)}$. Here $\Lambda \geq 0$ is a design parameter called *linear order*. If only this part is considered (for convention $O = 0$, where O is defined in Omicron part) then the trajectory is a line. Hence the name "Lambda" recalls this fact by abbreviating the word "line". For $\Lambda = 0$, the model represents a stationary trajectory (the object does not move);

- **Omicron model**: describes the dynamics of the steering speed

$$\boldsymbol{\omega} \triangleq \dot{\boldsymbol{\theta}} \tag{9.5}$$

where $\boldsymbol{\theta}$ is the orientation of the object, and its first $O-1$ derivatives $\dot{\boldsymbol{\omega}}, \ldots, \boldsymbol{\omega}^{(O-1)}$. Here $O \geq 0$ is a design parameter, which can be chosen independently from Λ, called *angular order*. If $\Lambda = 1$, then the trajectory is a circle. Here the name "Omicron" recalls this fact because the letter Omicron resembles the shape of a circle. For $O = 0$, the model represents a rectilinear trajectory (the object moves on a straight line).

9.2.4 Examples of Lambda:Omicron motion models

Naturally, each choice of Λ and O gives rise to a different type of motion model. The simplest examples of Lambda:Omicron models are the following:

- **0:0 motion model**: represents the trajectory of a completely stationary object which does not change its position or orientation. Clearly, this is not an useful model for a tracking problem;

- **0:1 motion model**: this model represents the trajectory of a stationary object which does not change its position but changes its orientation with a constant turning rate. Once again, this is not a useful model for a tracking problem;

- **1:0 motion model**: this is the simplest useful model and represents a uniform rectilinear motion. This model is suitable if it is known that the object moves along a line with a constant speed;

- **1:1 motion model**: represents a uniform circular motion. This model is suitable if it is known that the object moves on a circle with a constant speed. It turns out that the 1:1 motion model is nothing but more the well-known *unicycle motion model*;

- **2:0 motion model**: represents a rectilinear uniformly accelerated motion. This model is suitable if it is known that the object moves on a line with a constant acceleration;

- **2:1 motion model**: this model is suitable if it is known that the object moves on a cirlce with a constant acceleration and steering speed. This is the simplest model that can represent a trajectory with non-constant curvature.

9.2.5 1:0 motion model

Vanilla 1:0 motion model

Consider the following NCV motion model, where the center of the object is $\boldsymbol{m} \triangleq [\boldsymbol{\xi}\ \boldsymbol{\eta}]'$ and the sampling interval is T,

$$\begin{aligned} \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + T\,\dot{\boldsymbol{\xi}}_k + \boldsymbol{w}_{\xi,k} \\ \boldsymbol{\eta}_{k+1} &= \boldsymbol{\eta}_k + T\,\dot{\boldsymbol{\eta}} + \boldsymbol{w}_{\eta,k} \\ \dot{\boldsymbol{\xi}}_{k+1} &= \dot{\boldsymbol{\xi}}_k + \boldsymbol{w}_{\dot{\xi},k} \\ \dot{\boldsymbol{\eta}}_{k+1} &= \dot{\boldsymbol{\eta}}_k + \boldsymbol{w}_{\dot{\eta},k} \end{aligned}. \tag{9.6}$$

By expressing the velocity vector $\dot{\boldsymbol{m}} \triangleq [\dot{\boldsymbol{\xi}}\ \dot{\boldsymbol{\eta}}]'$ in polar coordinates $\boldsymbol{v}$, $\boldsymbol{\theta}$, i.e. by using the change of variables

$$\begin{aligned} \dot{\boldsymbol{\xi}} &= \boldsymbol{v}\cos\boldsymbol{\theta} \\ \dot{\boldsymbol{\eta}} &= \boldsymbol{v}\sin\boldsymbol{\theta} \end{aligned} \tag{9.7}$$

where it is assumed that the velocity vector $\boldsymbol{m}$ is aligned with the orientation of the object, so that $\boldsymbol{\theta}$ represents the orientation angle, the NCV equations can be written in the following compact form

$$\begin{aligned} \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + T f_{\parallel}^{(1)}(\boldsymbol{v}_k, \boldsymbol{\theta}_k) + \boldsymbol{w}_{m,k} \\ \boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + \boldsymbol{w}_{v,k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + \boldsymbol{w}_{\theta,k} \end{aligned} \tag{9.8}$$

where it is introduced the nonlinear function

$$f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta}) \triangleq \begin{bmatrix} \cos\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} \end{bmatrix} \boldsymbol{v}. \tag{9.9}$$

These Equations define the 1:0 motion model and

- the first two equations in $\boldsymbol{m}$ and $\boldsymbol{v}$ define the *first order lambda model*;
- the third equation in $\boldsymbol{\theta}$ defines the *zero order omicron model.*

Augmented 1:0 motion model

At this point, in order to integrate the 1:0 model with the formalism of the MEM-EKF* filter, the model is augmented with the two following additional motion equations for the lenght $\boldsymbol{l}_1$ and width $\boldsymbol{l}_2$ of the object

$$\begin{aligned} \boldsymbol{l}_{1,k+1} &= \boldsymbol{l}_{1,k} + \boldsymbol{w}_{l_1,k} \\ \boldsymbol{l}_{2,k+1} &= \boldsymbol{l}_{2,k} + \boldsymbol{w}_{l_2,k} \end{aligned} \tag{9.10}$$

by introducing the kinematic state $\boldsymbol{r} \triangleq [\boldsymbol{m}' \ \boldsymbol{v}]'$ and the shape state $\boldsymbol{p} \triangleq [\boldsymbol{\theta} \ \boldsymbol{l}_1 \ \boldsymbol{l}_2]'$, the model can be written more coincisely as

$$\begin{aligned} \boldsymbol{r}_{k+1} &= f_r(\boldsymbol{r}_k, \boldsymbol{p}_k) + \boldsymbol{w}_{r,k} \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + \boldsymbol{w}_{p,k} \end{aligned} \tag{9.11}$$

where

$$f_r(\boldsymbol{r}, \boldsymbol{p}) \triangleq \begin{bmatrix} \boldsymbol{m} + T f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta}) \\ \boldsymbol{v} \end{bmatrix}. \tag{9.12}$$

Note that, unlike the MEM-EKF* motion model, which treats $\boldsymbol{r}$ and $\boldsymbol{p}$ separately, now the first equation depends on $\boldsymbol{p}$, so $\boldsymbol{r}$ and $\boldsymbol{p}$ are coupled. In order to treat $\boldsymbol{r}$ and $\boldsymbol{p}$ jointly, it is convenient to define the state of the extended object as $\boldsymbol{x} \triangleq [\boldsymbol{r}' \ \boldsymbol{p}']'$, so that the final expression of the augmented 1:0 motion model is

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{9.13}$$

where the global transition function is

$$f(\boldsymbol{x}) \triangleq \begin{bmatrix} f_r(\boldsymbol{r}, \boldsymbol{p}) \\ \boldsymbol{p} \end{bmatrix} \tag{9.14}$$

9.2.6 2:1 motion model

Vanilla 2:1 motion model

Consider the following NCA motion model,

$$\begin{aligned} \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + T\,\dot{\boldsymbol{\xi}}_k + \frac{T^2}{2}\,\ddot{\boldsymbol{\xi}}_k + \boldsymbol{w}_{\boldsymbol{\xi},k} \\ \boldsymbol{\eta}_{k+1} &= \boldsymbol{\eta}_k + T\,\dot{\boldsymbol{\eta}} + \frac{T^2}{2}\,\ddot{\boldsymbol{\eta}}_k + \boldsymbol{w}_{\eta,k} \\ \dot{\boldsymbol{\xi}}_{k+1} &= \dot{\boldsymbol{\xi}}_k + T\,\ddot{\boldsymbol{\xi}}_k + \boldsymbol{w}_{\dot{\xi},k} \\ \dot{\boldsymbol{\eta}}_{k+1} &= \dot{\boldsymbol{\eta}}_k + T\,\ddot{\boldsymbol{\eta}}_k + \boldsymbol{w}_{\dot{\eta},k} \\ \ddot{\boldsymbol{\xi}}_{k+1} &= \ddot{\boldsymbol{\xi}}_k + \boldsymbol{w}_{\ddot{\xi},k} \\ \ddot{\boldsymbol{\eta}}_{k+1} &= \ddot{\boldsymbol{\eta}}_k + \boldsymbol{w}_{\ddot{\eta},k} \end{aligned} \tag{9.15}$$

by expressing the velocity vector $\dot{\boldsymbol{m}} \triangleq [\dot{\boldsymbol{\xi}} \ \dot{\boldsymbol{\eta}}]'$ in polar coordinates $\boldsymbol{v}$, $\boldsymbol{\theta}$, where as the previous case it is assumed that the velocity vector $\boldsymbol{m}$ is aligned to the object orientation, follows that

$$\begin{aligned} \ddot{\boldsymbol{\xi}} &= \dot{\boldsymbol{v}} \cos \boldsymbol{\theta} + \boldsymbol{v}\boldsymbol{\omega}(-\sin \boldsymbol{\theta}) \\ \ddot{\boldsymbol{\eta}} &= \dot{\boldsymbol{v}} \sin \boldsymbol{\theta} + \boldsymbol{v}\boldsymbol{\omega} \cos \boldsymbol{\theta} \end{aligned} \tag{9.16}$$

where it is introduced the longitudinal acceleration $\dot{\boldsymbol{v}}$ and the steering speed $\boldsymbol{\omega}$, defined as

$$\dot{\boldsymbol{v}} \triangleq \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \qquad \boldsymbol{\omega} \triangleq \frac{\mathrm{d}\boldsymbol{\theta}}{\mathrm{d}t}. \tag{9.17}$$

Sampling these equations according to the sampling interval T and assuming that $\boldsymbol{\omega}$ and $\dot{\boldsymbol{v}}$ are stationary up to small zero-mean Gaussian fluctuations $\boldsymbol{w}_{\omega,k}$, $\boldsymbol{w}_{\dot{v},k}$, the following equations are obtained

$$\begin{aligned} \boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + T\dot{\boldsymbol{v}}_k + \boldsymbol{w}_{v,k} \\ \dot{\boldsymbol{v}}_{k+1} &= \dot{\boldsymbol{v}}_k + \boldsymbol{w}_{\dot{v},k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + T\boldsymbol{\omega}_{k+1} + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{\omega}_{k+1} &= \boldsymbol{\omega}_k + \boldsymbol{w}_{\omega,k} \end{aligned} \tag{9.18}$$

consequently,

$$\begin{aligned} \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + T\,\boldsymbol{v}_k \cos\boldsymbol{\theta}_k + \frac{T^2}{2}\left[\dot{\boldsymbol{v}}_k \cos\boldsymbol{\theta}_k + \boldsymbol{v}_k\boldsymbol{\omega}_k(-\sin\boldsymbol{\theta}_k)\right] + \boldsymbol{w}_{\xi,k} \\ \boldsymbol{\eta}_{k+1} &= \boldsymbol{\eta}_k + T\,\boldsymbol{v}_k \sin\boldsymbol{\theta}_k + \frac{T^2}{2}\left[\dot{\boldsymbol{v}}_k \sin\boldsymbol{\theta}_k + \boldsymbol{v}_k\boldsymbol{\omega}_k \cos\boldsymbol{\theta}_k\right] + \boldsymbol{w}_{\eta,k} \\ \boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + T\dot{\boldsymbol{v}}_k + \boldsymbol{w}_{v,k} \\ \dot{\boldsymbol{v}}_{k+1} &= \dot{\boldsymbol{v}}_k + \boldsymbol{w}_{\dot{v},k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + T\boldsymbol{\omega}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{\omega}_{k+1} &= \boldsymbol{\omega}_k + \boldsymbol{w}_{\omega,k} \end{aligned} \tag{9.19}$$

which can be expressed coincisely as

$$\begin{aligned} \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + T f_{\parallel}^{(1)}(\boldsymbol{\lambda}_k, \boldsymbol{\theta}_k) + \frac{T^2}{2}\left(f_{\parallel}^{(2)}(\boldsymbol{\lambda}_k, \boldsymbol{\theta}_k) + f_{\perp}^{(2)}(\boldsymbol{\lambda}_k, \boldsymbol{\omega}_k, \boldsymbol{\theta}_k)\right) + \boldsymbol{w}_{m,k} \\ \boldsymbol{\lambda}_{k+1} &= A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + T\boldsymbol{\omega}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{\omega}_{k+1} &= \boldsymbol{\omega}_k + \boldsymbol{w}_{\omega,k} \end{aligned} \tag{9.20}$$

where

$$\boldsymbol{\lambda} \triangleq \begin{bmatrix} \boldsymbol{v} \\ \dot{\boldsymbol{v}} \end{bmatrix} \qquad A_\lambda \triangleq \mathrm{Toep}\,(1, T) \tag{9.21}$$

and the nonlinear functions

$$f_{\parallel}^{(1)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \triangleq \begin{bmatrix} \cos\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} \end{bmatrix} \boldsymbol{v} \qquad f_{\parallel}^{(2)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \triangleq \begin{bmatrix} \cos\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} \end{bmatrix} \dot{\boldsymbol{v}} \qquad f_{\perp}^{(2)}(\boldsymbol{\lambda}, \boldsymbol{\omega}, \boldsymbol{\theta}) \triangleq \begin{bmatrix} -\sin\boldsymbol{\theta} \\ \cos\boldsymbol{\theta} \end{bmatrix} \boldsymbol{v}\boldsymbol{\omega} \tag{9.22}$$

These equations are the 2:1 model and

- the first two equations in $\boldsymbol{m}$ and $\boldsymbol{\lambda}$ define, according to, the *second order lambda model*;
- the last two equations in $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ define the *first order omicron model.*

Augmented 2:1 motion model

Now, by including in the model the radii equations and by defining the kinematic state $\boldsymbol{r} \triangleq [\boldsymbol{m}' \ \boldsymbol{\lambda}' \ \boldsymbol{\omega}]'$ and the shape state $\boldsymbol{p} \triangleq [\boldsymbol{\theta} \ \boldsymbol{l_1} \ \boldsymbol{l_2}]'$, follows that the augmented 2:1 model can be written as

$$\begin{aligned} \boldsymbol{r}_{k+1} &= f_r(\boldsymbol{r}_k, \boldsymbol{p}_k) + \boldsymbol{w}_{r,k} \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + A_{pr}\boldsymbol{r}_k + \boldsymbol{w}_{p,k} \end{aligned} \tag{9.23}$$

where

$$A_{pr} \triangleq \begin{bmatrix} 0_{1\times 4} & T \\ 0_{1\times 4} & 0 \\ 0_{1\times 4} & 0 \end{bmatrix} \qquad f_r(\boldsymbol{r}, \boldsymbol{p}) \triangleq \begin{bmatrix} \boldsymbol{m} + T f_{\parallel}^{(1)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) + \frac{T^2}{2}\left(f_{\parallel}^{(2)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) + f_{\perp}^{(2)}(\boldsymbol{\lambda}, \boldsymbol{\omega}, \boldsymbol{\theta})\right) \\ A_{\lambda}\boldsymbol{\lambda} \\ \boldsymbol{\omega} \end{bmatrix}. \tag{9.24}$$

Finally, in terms of the global state $\boldsymbol{x} \triangleq [\boldsymbol{r}' \ \boldsymbol{p}']'$, the augmented 2:1 motion model assumes the form

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{9.25}$$

where the global transition function is

$$f(\boldsymbol{x}) \triangleq \begin{bmatrix} f_r(\boldsymbol{r}, \boldsymbol{p}) \\ \boldsymbol{p} + A_{pr}\boldsymbol{r} \end{bmatrix} \tag{9.26}$$

9.2.7 General Lambda:Omicron motion model

Vanilla N:$N-1$ motion model

Consider the following linear kinematic model of order $N+1$, where $N \geq 0$ is a design parameter, for the center $\boldsymbol{m} \triangleq [\boldsymbol{\xi} \ \boldsymbol{\eta}]'$ of the object,

$$\tilde{\boldsymbol{m}}_{k+1} = A_{\tilde{m}}\tilde{\boldsymbol{m}}_k + \boldsymbol{w}_{\tilde{m},k} \tag{9.27}$$

where

$$\tilde{\boldsymbol{m}} \triangleq \left[\boldsymbol{m}' \quad \ldots \quad (\boldsymbol{m}^{(N)})'\right]' \qquad A_{\tilde{\mu}} \triangleq \text{Toep}\,(N, T) \otimes I_2 \tag{9.28}$$

by introducing the polar change of variables (with the usual assumption that the velocity vector $\boldsymbol{m}$ is alligned to the angle orientation)

$$\dot{\boldsymbol{m}} = f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta}) \tag{9.29}$$

and by expressing also $\ddot{\boldsymbol{m}}$, $\dddot{\boldsymbol{m}}$, $\ldots$, $\boldsymbol{m}^{(N)}$ in terms of the new variables

$$\begin{aligned} \boldsymbol{\lambda} &\triangleq \begin{bmatrix} \boldsymbol{v} & \dot{\boldsymbol{v}} & \ldots & \boldsymbol{v}^{(\Lambda \triangleq N)} \end{bmatrix}' \\ \boldsymbol{o} &\triangleq \begin{bmatrix} \boldsymbol{\omega} & \dot{\boldsymbol{\omega}} & \ldots & \boldsymbol{\omega}^{(O \triangleq N-1)} \end{bmatrix}' \end{aligned} \tag{9.30}$$

via the relations

$$\dot{\boldsymbol{m}} = \frac{\mathrm{d} f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta})}{\mathrm{d}t} \qquad \ldots \qquad \boldsymbol{m}^{(N)} = \frac{\mathrm{d}^N f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta})}{\mathrm{d}t^N} \tag{9.31}$$

it turns out, by assuming the model

$$\begin{aligned} \boldsymbol{\lambda}_{k+1} &= A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \qquad & A_\lambda &\triangleq \mathrm{Toep}(\Lambda \triangleq N, T) \\ \boldsymbol{o}_{k+1} &= A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k} \qquad & A_o &\triangleq \mathrm{Toep}(O \triangleq N-1, T) \end{aligned}, \tag{9.32}$$

that equation is equivalent to

$$\begin{aligned} \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + \sum_{i=1}^{N} \frac{T^i}{i!} f^{(i)}(\boldsymbol{\lambda}_k, \boldsymbol{o}_k, \boldsymbol{\theta}_k) + \boldsymbol{w}_{m,k} \\ \boldsymbol{\lambda}_{k+1} &= A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + A_{\theta o} \boldsymbol{o}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{o}_{k+1} &= A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k} \end{aligned} \tag{9.33}$$

where

$$\begin{aligned} A_{\theta o} &\triangleq \mathrm{r}_{1, O \triangleq N-1, T} \\ f^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) &\triangleq \frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}} f_{\parallel}^{(1)}(\boldsymbol{v}, \boldsymbol{\theta}) \qquad i = 1, 2, \ldots N \end{aligned}. \tag{9.34}$$

Observation 2. The generic nonlinear function $f^{(i)}(\cdot)$ can always be decomposed in the following form

$$f^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) = f_{\parallel}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) + f_{\perp}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) \tag{9.35}$$

where, for suitable scalars $k_{\parallel}^{(i)}(\cdot)$ and $k_{\perp}^{(i)}(\cdot)$ depending on $\boldsymbol{\lambda}$ and $\boldsymbol{o}$,

$$f_{\parallel}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) = \begin{bmatrix} \cos \boldsymbol{\theta} \\ \sin \boldsymbol{\theta} \end{bmatrix} k_{\parallel}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}) \qquad f_{\perp}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) = \begin{bmatrix} -\sin \boldsymbol{\theta} s \\ \cos \boldsymbol{\theta} \end{bmatrix} k_{\perp}^{(i)}(\boldsymbol{\lambda}, \boldsymbol{o}). \tag{9.36}$$

A simple induction shows that the decomposition holds for any i. The base of the induction is expressed, for example, by the 2:1 model where $f^{(1)}$, $f^{(2)}$

are explicitly computed and then decomposed in $f_\parallel^{(1)}$, $f_\perp^{(1)}$ and $f_\parallel^{(2)}$, $f_\perp^{(2)}$. To show the induction step, suppose that the decomposition holds for a generic i and compute $f^{(i+1)}$ as follows

$$\begin{aligned} f^{(i+1)} &\triangleq \frac{\mathrm{d}f^{(i)}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\begin{bmatrix}\cos\boldsymbol{\theta}\\ \sin\boldsymbol{\theta}\end{bmatrix} k_\parallel^{(i)}\right] + \frac{\mathrm{d}}{\mathrm{d}t}\left[\begin{bmatrix}-\sin\boldsymbol{\theta}\\ \cos\boldsymbol{\theta}\end{bmatrix} k_\perp^{(i)}\right] \\ &= \underbrace{\begin{bmatrix}\cos\boldsymbol{\theta}\\ \sin\boldsymbol{\theta}\end{bmatrix} \underbrace{(\dot{k}_\parallel^{(i)} - \boldsymbol{\omega} k_\perp^{(i)})}_{\triangleq k_\parallel^{(i+1)}}}_{\triangleq f_\parallel^{(i+1)}} + \underbrace{\begin{bmatrix}-\sin\boldsymbol{\theta}\\ \cos\boldsymbol{\theta}\end{bmatrix} \underbrace{(\boldsymbol{\omega} k_\parallel^{(i)} + \dot{k}_\perp^{(i)})}_{\triangleq k_\perp^{(i+1)}}}_{f_\perp^{(i+1)}}. \end{aligned} \tag{9.37}$$

Note that the decomposition still holds also for $i = 1$ by considering the convention

$$f_\perp^{(1)}(\boldsymbol{\lambda}, \boldsymbol{o}, \boldsymbol{\theta}) \triangleq 0_{2\times 1} = \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix} \underbrace{k_\perp^{(1)}}_{\triangleq 0}. \tag{9.38}$$

According to the decomposition, it turns out that

$$\begin{aligned} \boldsymbol{m}_{k+1} = \boldsymbol{m}_k + \sum_{i=1}^{N} \frac{T^i}{i!}\left[f_\parallel^{(i)}(\boldsymbol{\lambda}_k, \boldsymbol{o}_k, \boldsymbol{\theta}_k) + f_\perp^{(i)}(\boldsymbol{\lambda}_k, \boldsymbol{o}_k, \boldsymbol{\theta}_k)\right] + \boldsymbol{w}_{m,k} \\ \boldsymbol{\lambda}_{k+1} = A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \\ \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + A_{\theta o}\boldsymbol{o}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{o}_{k+1} = A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k} \end{aligned} \tag{9.39}$$

which is the N:$N-1$ model and

- the first two equations in $\boldsymbol{m}$ and $\boldsymbol{\lambda}$ define the *lambda model* of order $\Lambda \triangleq N$;
- the last two equations in $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ define the *omicron model* of order $O \triangleq N-1$.

In some scenarios one can be interested in more accurate estimation of $\boldsymbol{\lambda}$ rather than $\boldsymbol{o}$ (or viceversa). In this case (or the other one), the quality of the estimate of $\boldsymbol{\lambda}$ ($\boldsymbol{o}$) can be improved by increasing Λ (O). However, the $N : N-1$ model does not allow the designer to increase the linear order Λ (angular order O) without increasing also the angular order O (linear order Λ).

In order to overcome this limitation, the general Lambda:Omicron model generalizes the $N : N-1$ motion model in a way such that Λ and O can be potentially chosen without satisfying the constraint $\Lambda = O+1$. On the other hand, the major flexibility of the generalized model is paid with the price of not having a direct correspondence with a familiar motion model such as the generic linear kinematic motion model.

Vanilla Lambda:Omicron motion model

Given two generic orders Λ, O, let

$$N \triangleq \max(\Lambda, O-1) \tag{9.40}$$

and observe the following fact:

- **case** $\Lambda = O-1$: trivially, $N = \Lambda$ and the $N : N-1$ model holds;
- **case** $\Lambda > O-1$: in this case $N = \Lambda$ and by defining

$$\tilde{\boldsymbol{o}} \triangleq \begin{bmatrix} \boldsymbol{o}' & 0_{1\times\Lambda-O-1} \end{bmatrix}' \tag{9.41}$$

  the $N : N-1$ model applys with the following convention

$$\begin{aligned} \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + \sum_{i=1}^{\Lambda} \frac{T^i}{i!} \left[f_{\parallel}^{(i)}(\boldsymbol{\lambda}_k, \tilde{\boldsymbol{o}}_k, \boldsymbol{\theta}_k) + f_{\perp}^{(i)}(\boldsymbol{\lambda}_k, \tilde{\boldsymbol{o}}_k, \boldsymbol{\theta}_k) \right] + \boldsymbol{w}_{m,k} \\ \boldsymbol{\lambda}_{k+1} &= A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + A_{\theta o} \boldsymbol{o}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{o}_{k+1} &= A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k} \end{aligned} \tag{9.42}$$

- **case** $\Lambda < O-1$: in this case $N = O-1$ and by defining

$$\tilde{\boldsymbol{\lambda}} \triangleq \begin{bmatrix} \boldsymbol{\lambda}' & 0_{1\times O-\Lambda+1} \end{bmatrix}' \tag{9.43}$$

  the $N : N-1$ model applys with the following convention

$$\begin{aligned} \boldsymbol{m}_{k+1} &= \boldsymbol{m}_k + \sum_{i=1}^{\Lambda} \frac{T^i}{i!} \left[f_{\parallel}^{(i)}(\tilde{\boldsymbol{\lambda}}_k, \boldsymbol{o}_k, \boldsymbol{\theta}_k) + f_{\perp}^{(i)}(\tilde{\boldsymbol{\lambda}}_k, \boldsymbol{o}_k, \boldsymbol{\theta}_k) \right] + \boldsymbol{w}_m \\ \boldsymbol{\lambda}_{k+1} &= A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_\lambda \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + A_{\theta o} \boldsymbol{o}_k + \boldsymbol{w}_\theta \\ \boldsymbol{o}_{k+1} &= A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k} \end{aligned} \tag{9.44}$$

The (general) Lambda:Omicron model, which applys in all three cases, is the following

$$\begin{aligned}\boldsymbol{m}_{k+1} = \boldsymbol{m}_k + \sum_{i=1}^{\max(\Lambda,O-1)} \frac{T^i}{i!}\left[f_{\parallel}^{(i)}(\tilde{\boldsymbol{\lambda}}_k, \tilde{\boldsymbol{o}}_k, \boldsymbol{\theta}_k) + f_{\perp}^{(i)}(\tilde{\boldsymbol{\lambda}}_k, \tilde{\boldsymbol{o}}_k, \boldsymbol{\theta}_k)\right] + \boldsymbol{w}_{m,k} \\ \boldsymbol{\lambda}_{k+1} = A_\lambda \boldsymbol{\lambda}_k + \boldsymbol{w}_{\lambda,k} \\ \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + A_{\theta o}\boldsymbol{o}_k + \boldsymbol{w}_{\theta,k} \\ \boldsymbol{o}_{k+1} = A_o \boldsymbol{o}_k + \boldsymbol{w}_{o,k}\end{aligned} \tag{9.45}$$

where are defined the padded lambda vector $\tilde{\boldsymbol{\lambda}}$ and the padded omicron vector $\tilde{\boldsymbol{o}}$ as

$$\begin{aligned}\tilde{\boldsymbol{\lambda}}_k &\triangleq \left[\boldsymbol{\lambda}' \quad 0_{1\times\max(O-\Lambda+1,0)}\right]' \\ \tilde{\boldsymbol{o}}_k &\triangleq \left[\boldsymbol{o}' \quad 0_{1\times\max(\Lambda-O-1,0)}\right]'.\end{aligned} \tag{9.46}$$

In conclusion,

- the first two equations are the general *Lambda model of order* Λ;
- the last two equations are the general *Omicron model of order* O.

Augmented Lambda:Omicron motion model

By including in the Lambda:Omicron model the radii equations and by defining the kinematic state $\boldsymbol{r} \triangleq [\boldsymbol{m}'\ \tilde{\boldsymbol{\lambda}}'\ \tilde{\boldsymbol{o}}']'$ and the shape state $\boldsymbol{p} \triangleq [\boldsymbol{\theta}\ \boldsymbol{l}_1\ \boldsymbol{l}_2]'$, turns out the augmented model

$$\begin{aligned}\boldsymbol{r}_{k+1} &= f_r(\boldsymbol{r}_k, \boldsymbol{p}_k) + \boldsymbol{w}_{r,k} \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + A_{pr}\boldsymbol{r}_k + \boldsymbol{w}_{p,k}\end{aligned} \tag{9.47}$$

where

$$A_{pr} \triangleq \begin{bmatrix} 0_{1\times 2\Lambda} & A_{\theta o} \\ 0_{1\times 2\Lambda} & 0_{1\times O} \\ 0_{1\times 2\Lambda} & 0_{1\times O}\end{bmatrix} \qquad f_r(\boldsymbol{r},\boldsymbol{p}) \triangleq \begin{bmatrix} \boldsymbol{m} + \sum_{i=1}^{\max(\Lambda,O-1)} \frac{T^i}{i!}\left(f_{\parallel}^{(i)}(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{o}}, \boldsymbol{\theta}) + f_{\perp}^{(i)}(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{o}}, \boldsymbol{\theta})\right) \\ A_\lambda \lambda \\ A_o o \end{bmatrix} \tag{9.48}$$

now, by defining the global state $\boldsymbol{x} \triangleq [\boldsymbol{r}'\ \boldsymbol{p}']'$ follows the compact expression

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{9.49}$$

where the transition function is

$$f(\boldsymbol{x}) \triangleq \begin{bmatrix} f_r(\boldsymbol{r},\boldsymbol{p}) \\ \boldsymbol{p} + A_{pr}\boldsymbol{r}\end{bmatrix} \tag{9.50}$$

9.3 Measurement model

9.3.1 Motivations

The MEM-EKF* filter considers as measurement model the couple given by the MEM and the pseudo-measurement model. One characteristic of the correction law is that, given the sample of measurements

$$\mathsf{y}_k \triangleq \{y_k^{(1)}, \ldots, y_k^{(n_k)}\}, \tag{9.51}$$

the filtered estimates are computed by processing sequentially the n_k measurements in the sample. As a result, the MEM-EKF* at each sampling step k performs n_k single-measurement corrections (which consists in a kinematic correction based directly on the measure y and in a shape correction based on the pseudo-measurement Y relative to y).

The LO-MEM filter uses the same measurement model of the MEM-EKF* filter to represent the measurements and the pseudo-measurements but, in order to obtain a corrector with low computational burden, tries to simplify the sequential strategy by replacing it with a single-shot strategy.

idea

Instead to consider the available measurements independently, the correction is performed, by using the BLUE equations, according to the mean measurement and the mean pseudo-measurement[1], defined respectively as

$$\bar{y}_k \triangleq \frac{1}{n_k}\sum_{i=1}^{n_k} y_k^{(i)} \qquad \overline{Y}_k \quad \triangleq \frac{1}{n_k - 1}\sum_{i=1}^{n_k} Y_k^{(i)}. \tag{9.52}$$

This idea is inspired to the correction strategy used by the GIW filter, which choose to process the mean measure $\bar{y}$ and the scatter matrix $\overline{Y}$ rather than the single measurements individually.

As a result, at each time step k the LO-MEM filter performs only one single measurement correction (which consists in a kinematic correction based directly on the mean measure $\bar{y}$ and in a shape correction based on the mean pseudo-measurement $\overline{Y}$). In other terms, the LO-MEM corrector is characterized by a constant computational cost, while the MEM-EKF* requires a computational cost that grows with the number n_k of available measurements.

[1]the normalizing factor is $n_k - 1$ rather than n_k due the so-called *Bessel correction*, which improves the estimation process (as will be shown in section*)

9.3.2 Drawback

It should be noted that, despite the clear advantage of the low computation cost, the LO-MEM corrector is less accurate than the MEM-EKF* because in general the transformations

$$\begin{aligned} y_k^{(i)}, \dots, y_k^{(i)} &\mapsto \bar{y} \\ Y_k^{(i)}, \dots, Y_k^{(i)} &\mapsto \overline{Y} \end{aligned} \tag{9.53}$$

implies an inevitable loss of information.

9.3.3 Mean MEM

The MEM, according to the definition of the kinematic state used by the Lambda:Omicron model, assumes the form

$$\boldsymbol{y}_k = H_r \boldsymbol{r}_k + S(\boldsymbol{p}_k)\boldsymbol{h}_k + \boldsymbol{v}_k \tag{9.54}$$

where the obsevation matrix H_r is defined as

$$H_r \triangleq \begin{bmatrix} I_2 & 0_{2\times N} & 0_{2\times N} \end{bmatrix} \tag{9.55}$$

now, by observing that in the sample each measurement, since are originated by the same object, shares the same kinematic state $\boldsymbol{r}_k$ and the same shape state $\boldsymbol{p}_k$, the mean MEM is given by

$$\begin{aligned} \bar{\boldsymbol{y}}_k &\triangleq \frac{1}{n_k} \sum_{i=1}^{n_k} \left(H_r \boldsymbol{r}_k + S(\boldsymbol{p}_k)\boldsymbol{h}_k^{(i)} + \boldsymbol{v}_k^{(i)} \right) \\ &= H_r \boldsymbol{r}_k + \frac{1}{n_k} \sum_{i=1}^{n_k} \left(S(\boldsymbol{p}_k)\boldsymbol{h}_k^{(i)} + \boldsymbol{v}_k^{(i)} \right) \end{aligned} \tag{9.56}$$

9.3.4 Mean MEM distribution

In order to apply the BLUE equations, the moments $\bar{y}_{k|k-1}$, $\Sigma_{\bar{\boldsymbol{y}}}$, $\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}}$ are required. In this section the entire distribution $p_{\bar{\boldsymbol{y}}}(\cdot)$ of $\bar{\boldsymbol{y}}_k$ is derived and consequently the three moments $\bar{y}_{k|k-1}$, $\Sigma_{\bar{\boldsymbol{y}}}$, $\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}}$ are extracted from $p_{\bar{\boldsymbol{y}}}(\cdot)$.

Distribution

By considering the following moment-matched Gaussian approximations[2]

$$\begin{aligned} \boldsymbol{h}_k &\sim \mathcal{N}(0, R_h) \\ \boldsymbol{r}_k &\sim \mathcal{N}(\hat{r}_{k|k-1}, P^r_{k|k-1}) \end{aligned} \tag{9.57}$$

[2] by definition, the multiplicative error $\boldsymbol{h}_k$ is not Gaussian, while the kinematic state $\boldsymbol{r}_k$ is not Gaussian due to the non-linearity of the Lambda:Omicron motion model

and by assuming that the covariance $P^p_{k|k-1}$ of $\boldsymbol{p}_k$ is sufficiently small to justify the following additional moment-matched Gaussian approximation[3]

$$S(\boldsymbol{p}_k)\boldsymbol{h}_k \sim \mathcal{N}(0, C^I + C^{II}) \tag{9.58}$$

follows, for the elementary sum rule for Gaussian distributions, that the mean MEM measurement $\bar{\boldsymbol{y}}_k$ is approximately distributed like

$$p_{\bar{\boldsymbol{y}}_k}(\bar{y}_k) \triangleq \mathcal{N}\left(\bar{y}_k; H_r\hat{r}_{k|k-1}, H_r P^r_{k|k-1} H'_r + \frac{+C^I + C^{II} + R_v}{n_k}\right) \tag{9.59}$$

Moments

According to $p_{\bar{\boldsymbol{y}}_k}(\cdot)$, the prediction $\bar{y}_{k|k-1}$ of $\bar{\boldsymbol{y}}_k$ and its covariance $\Sigma_{\bar{\boldsymbol{y}}}$ are respectively

$$\begin{aligned} \bar{y}_{k|k-1} &\triangleq H_r\hat{r}_{k|k-1} \\ \Sigma_{\bar{\boldsymbol{y}}} &\triangleq H_r P^r_{k|k-1} H'_r + \frac{C^I + C^{II} + R_v}{n_k} \end{aligned} \tag{9.60}$$

the cross-covariance $\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}}$ can be computed via the definition

$$\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}} \triangleq \mathbb{E}[\tilde{\boldsymbol{x}}_k \tilde{\bar{\boldsymbol{y}}}'_k] \tag{9.61}$$

where

$$\begin{aligned} \tilde{\boldsymbol{x}}_k &\triangleq \boldsymbol{x}_k - \hat{x}_{k|k-1} = \begin{bmatrix} \boldsymbol{r}_k - \hat{r}_k \\ \boldsymbol{p}_k - \hat{p}_k \end{bmatrix} \triangleq \begin{bmatrix} \tilde{\boldsymbol{r}}_{k|k-1} \\ \tilde{\boldsymbol{p}}_{k|k-1} \end{bmatrix} \\ \tilde{\bar{\boldsymbol{y}}}_k &\triangleq \bar{\boldsymbol{y}}_k - \bar{y}_{k|k-1} = H_r\tilde{\boldsymbol{r}}_k + \frac{1}{n_k}\sum_{i=1}^{n_k}\left(S(\boldsymbol{p}_k)\boldsymbol{h}^{(i)}_k + \boldsymbol{v}^{(i)}_k\right) \end{aligned} \tag{9.62}$$

its easy to see that

$$\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}} = \begin{bmatrix} P^r_{k|k-1}H'_r \\ P^{pr}_{k|k-1}H'_r \end{bmatrix} = P_{k|k-1}H' \tag{9.63}$$

where it is introduced the global observation matrix

$$H \triangleq \begin{bmatrix} H_r & 0_{2\times(2N_{\Lambda O}+1)} \end{bmatrix} \tag{9.64}$$

[3] the idea behind this approximation is that, if $\boldsymbol{p}_k$ is deterministic, i.e. $P^p_{k|k-1} = 0$ and $\boldsymbol{p}_k = \hat{p}_{k|k-1}$, then, according to the Gaussian approximation for $\boldsymbol{h}_k$, the distribution of $S(\hat{p}_{k|k-1})\boldsymbol{h}_k$ is trivially $\mathcal{N}(0, S(\hat{p}_{k|k-1})R_h S(\hat{p}_{k|k-1})') = \mathcal{N}(0, C^I)$. In order to take into account the (assumed to be small) covariance $P^p_{k|k-1} \approx 0$, the second factor C^{II} is included in the final approximation for the distribution of $S(\boldsymbol{p}_k)\boldsymbol{h}_k$

9.3.5 Mean pseudo-measurement model

Likewise in the previous section, the distribution $p_{\overline{\boldsymbol{Y}}}(\cdot)$ and the moments $\overline{Y}_{k|k-1}$, $\Sigma_{\overline{\boldsymbol{Y}}}$, $\Sigma_{\boldsymbol{x}\overline{\boldsymbol{Y}}}$ associated to the mean pseudo-measurement $\overline{\boldsymbol{Y}}_k$ are derived in what follows in order to apply the BLUE equations.

Preliminary discussion

According to its definition, the mean pseudo-measurement $\overline{\boldsymbol{Y}}_k$ can be written in the following form

$$\overline{\boldsymbol{Y}}_k = F\,\text{vec}[\boldsymbol{S}_{\text{cov}}] \tag{9.65}$$

where $\boldsymbol{S}_{\text{cov}}$ is the so-called *sample covariance*, defined as

$$\boldsymbol{S}_{\text{cov}} \triangleq \frac{1}{n_k - 1}\sum_{i=1}^{n_k}(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)' \tag{9.66}$$

in fact,

$$\begin{aligned}
\overline{\boldsymbol{Y}}_k &\triangleq \frac{1}{n_k - 1}\sum_{i=1}^{n_k}\boldsymbol{Y}_k^{(i)} \\
&= \frac{1}{n_k - 1}\sum_{i=1}^{n_k}F[(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)\otimes(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)] \\
&= \frac{1}{n_k - 1}\sum_{i=1}^{n_k}F\,\text{vec}\left[(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)'\right] \\
&= F\,\text{vec}\left[\frac{1}{n_k - 1}\sum_{i=1}^{n_k}(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)'\right] \triangleq F\,\text{vec}[\boldsymbol{S}_{\text{cov}}]
\end{aligned} \quad . \tag{9.67}$$

In the Gaussian case holds the following famous result regarding the sample covariance

Theorem 16. Let $\{\boldsymbol{y}^{(i)}\}_{i=1}^{n}$ be a sample of n IID p-variate measurements distributed according to

$$\boldsymbol{y}^{(i)} \sim \mathcal{N}(\mu_{\boldsymbol{y}}, \Sigma_{\boldsymbol{y}} > 0) \qquad i = 1, 2, \ldots, n \tag{9.68}$$

then

- the sample mean $\bar{\boldsymbol{y}} \triangleq \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}^{(i)}$ follows the Gaussian distribution

$$p_{\bar{\boldsymbol{y}}}(\bar{y}) = \mathcal{N}(\bar{y}; \mu_{\boldsymbol{y}}, \Sigma_{\boldsymbol{y}}) \tag{9.69}$$

- the sample covariance $\boldsymbol{S}_{\text{cov}} \triangleq \frac{1}{n_k-1}\sum_{i=1}^{n_k}(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_k^{(i)} - \bar{\boldsymbol{y}}_k)'$ follows the Wishart distribution

$$p_{\boldsymbol{S}_{\text{cov}}}(S_{\text{cov}}) = \mathcal{W}_p\left(S_{\text{cov}}; n-1, \frac{\Sigma_{\boldsymbol{y}}}{n-1}\right) \tag{9.70}$$

- the sample mean $\bar{\boldsymbol{y}}$ and the sample covariance $\boldsymbol{S}_{\text{cov}}$ are independently distributed.

Distribution

In the MEM case, the sample is (approximately) Gaussian but not IID because each measurement shares the factor $H_r\boldsymbol{r}$, which correlates the measurements to each other. Despite this negative fact, the same factor $H_r\boldsymbol{r}$ is also present in the sample mean and the correlation effect partially cancels out. Hence thereom applys approximately to the actual case.

Since between the sample covariance $\boldsymbol{S}_{\text{cov}}$ and the mean pseudo-measurement $\overline{\boldsymbol{Y}}$ there is a 1-to-1 relationship (which says that $\overline{\boldsymbol{Y}}$ is nothing but more than the vectorial representation of $\boldsymbol{S}_{\text{cov}}$), follows that the distribution of $\overline{\boldsymbol{Y}}$ is the same as the distribution of $\boldsymbol{S}_{\text{cov}}$, so the distribution of the mean pseudo-measurement is approximately given by

$$p_{\overline{\boldsymbol{Y}}}(\overline{Y}) = \mathcal{W}_2\left(S_{\text{cov}} \triangleq \begin{bmatrix} \overline{Y}_1 & \overline{Y}_3 \\ \overline{Y}_3 & \overline{Y}_2 \end{bmatrix}; n-1, \frac{\Sigma_{\boldsymbol{y}}}{n-1}\right) \tag{9.71}$$

where $\overline{Y} = [\overline{Y}_1 \; \overline{Y}_2 \; \overline{Y}_3]'$.

Moments

According to the distribution $p_{\overline{\boldsymbol{Y}}}(\cdot)$ and the formulae for the Wishart moments, the moments $\overline{Y}_{k|k-1}$ and $\Sigma_{\overline{\boldsymbol{Y}}}$ are respectively[4]

$$\begin{aligned} \overline{Y}_{k|k-1} &= F\,\text{vec}[\mathbb{E}[\boldsymbol{S}_{\text{cov}}]] \\ &= F\,\text{vec}\left[(n_k-1)\frac{\Sigma_{\boldsymbol{y}}}{n_k-1}\right] \\ &= F\,\text{vec}[\Sigma_{\boldsymbol{y}}] \end{aligned} \tag{9.72}$$

[4] The Bessel correction simplifies the expression of the prediction $\overline{Y}_{k|k-1}$. By considering the true mean pseudo-measurement $\frac{1}{n_k}\sum_{i=1}^{n_k} Y_k^{(i)}$, the prediction would be $\frac{n_k-1}{n_k}F\text{vec}[\Sigma_{\boldsymbol{y}}]$ (instead of the simpler $F\text{vec}[\Sigma_{\boldsymbol{y}}]$)

and

$$
\begin{aligned}
\Sigma_{\overline{\boldsymbol{Y}}} &= \mathrm{Cov}[F\mathrm{vec}[\boldsymbol{S}_{\mathrm{cov}}]] \\
&= F(\mathrm{Cov}[\mathrm{vec}[\boldsymbol{S}_{\mathrm{cov}}]])F' \\
&= F\left((n_k-1)(I_4+K)\left(\frac{\Sigma_{\boldsymbol{y}}}{n_k-1}\otimes\frac{\Sigma_{\boldsymbol{y}}}{n_k-1}\right)\right)F' \\
&= F\left(\frac{I_4+K}{n_k-1}(\Sigma_{\boldsymbol{y}}\otimes\Sigma_{\boldsymbol{y}})\right)F'
\end{aligned}
\tag{9.73}
$$

where K is the commutation matrix given.

In order to computate of the cross-covariance, start by observing that for linearity

$$
\Sigma_{\boldsymbol{x}\overline{\boldsymbol{Y}}} = \frac{1}{n_k-1}\sum_{i=1}^{n_k}\Sigma_{\boldsymbol{x}\boldsymbol{Y}} = \frac{n_k}{n_k-1}\Sigma_{\boldsymbol{x}\boldsymbol{Y}} \tag{9.74}
$$

now, the cross-covariance between $\boldsymbol{x}$ and $\boldsymbol{Y}$ can be computed approximately as

$$
\Sigma_{\boldsymbol{x}\boldsymbol{Y}} \approx \mathbb{E}\left[\tilde{\boldsymbol{x}}_k\left(\left.\frac{\partial g}{\partial \boldsymbol{x}}\right|_{\hat{x}_{k|k-1}}\tilde{\boldsymbol{x}}\right)'\right] \tag{9.75}
$$

by assuming that the Jacobian $\frac{g}{\boldsymbol{x}}$ is independent from $\tilde{\boldsymbol{x}}$ and by observing that

$$
\mathbb{E}\left[\left.\frac{\partial g}{\partial \boldsymbol{x}}\right|_{\hat{x}_{k|k-1}}\right] = \left[\mathbb{E}\left[\left.\frac{\partial g}{\partial \boldsymbol{r}}\right|_{\hat{r}_{k|k-1}}\right] \quad \mathbb{E}\left[\left.\frac{\partial g}{\partial \boldsymbol{p}}\right|_{\hat{p}_{k|k-1}}\right]\right] = \underbrace{\left[0_{2\times 2} \quad M_{\hat{p}}\right]}_{\triangleq M_{\hat{x}}} \tag{9.76}
$$

follows the final expression

$$
\Sigma_{\boldsymbol{x}\overline{\boldsymbol{Y}}} = P_{k|k-1}\left(\frac{n_k}{n_k-1}M_{\hat{x}}\right)' \tag{9.77}
$$

9.4 LO-MEM predictor

9.4.1 Extended Kalman predictor

The Lambda:Omicron motion model is not linear. The LO-MEM predictor is thus defined as the following EKF predictor

$$
\begin{aligned}
\hat{x}_{k+1|k} &\triangleq f(x_{k|k}) \\
P_{k+1|k} &\triangleq \hat{J}_x P_{k|k}\hat{J}_x' + Q
\end{aligned}
\tag{9.78}
$$

where

$$\hat{J}_x \triangleq \left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\hat{x}_{k|k}} \tag{9.79}$$

the main problem, addressed in the next section, is to find the explicit expression of the motion Jacobian $J_x \triangleq \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$

9.4.2 Motion Jacobian

Start from the Jacobians of the non linear functions $f_{\|}^{(i)}$ and $f_{\perp}^{(i)}$, which are defined by

$$\begin{aligned}
J_{\theta\|}^{(i)} &\triangleq \frac{\partial f_{\|}^{(i)}}{\partial \boldsymbol{\theta}} & J_{\theta\perp}^{(i)} &\triangleq \frac{\partial f_{\perp}^{(i)}}{\partial \boldsymbol{\theta}} \\
J_{\lambda\|}^{(i)} &\triangleq \frac{\partial f_{\|}^{(i)}}{\partial \tilde{\boldsymbol{\lambda}}} & J_{\lambda\perp}^{(i)} &\triangleq \frac{\partial f_{\perp}^{(i)}}{\partial \tilde{\boldsymbol{\lambda}}} \\
J_{o\|}^{(i)} &\triangleq \frac{\partial f_{\|}^{(i)}}{\partial \tilde{\boldsymbol{o}}} & J_{o\perp}^{(i)} &\triangleq \frac{\partial f_{\perp}^{(i)}}{\partial \tilde{\boldsymbol{o}}}
\end{aligned} \tag{9.80}$$

then the equation in $\boldsymbol{m}$ is linearized near $\hat{x}_{k|k}$ as follow

$$\begin{aligned}
\boldsymbol{m}_{k+1} = \boldsymbol{m}_k + \sum_{i=1}^{N} \frac{T^i}{i!}\left(\hat{f}_{\|}^{(i)} + \hat{f}_{\perp}^{(i)}\right) + \underbrace{\left[\sum_{i=1}^{N} \frac{T^i}{i!}\left(\hat{J}_{\lambda\|}^{(i)} + \hat{J}_{\lambda\perp}^{(i)}\right)\right]}_{\triangleq \hat{J}_\lambda}(\tilde{\boldsymbol{\lambda}}_k - \tilde{\lambda}_{k|k}) \\
+ \underbrace{\left[\sum_{i=1}^{N} \frac{T^i}{i!}\left(\hat{J}_{o\|}^{(i)} + \hat{J}_{o\perp}^{(i)}\right)\right]}_{\triangleq \hat{J}_o}(\tilde{\boldsymbol{o}}_k - \tilde{o}_{k|k}) + \underbrace{\left[\sum_{i=1}^{N} \frac{T^i}{i!}\left(\hat{J}_{\theta\|}^{(i)} + \hat{J}_{\theta\perp}^{(i)}\right)\right]}_{\triangleq \hat{J}_\theta}(\boldsymbol{\theta}_k - \theta_{k|k}) + \boldsymbol{w}_m
\end{aligned} \tag{9.81}$$

where

$$\hat{f}_{\|}^{(i)} \triangleq f_{\|}^{(i)}(\hat{x}_{k|k}) \qquad \hat{f}_{\perp}^{(i)} \triangleq f_{\|}^{(i)}(\hat{x}_{k|k}) \tag{9.82}$$

and

$$\begin{aligned}
\hat{J}_{\theta\|}^{(i)} &\triangleq \left.\frac{\partial f_{\|}^{(i)}}{\partial \boldsymbol{\theta}}\right|_{\hat{x}_{k|k}} & \hat{J}_{\theta\perp}^{(i)} &\triangleq \left.\frac{\partial f_{\perp}^{(i)}}{\partial \boldsymbol{\theta}}\right|_{\hat{x}_{k|k}} \\
\hat{J}_{\lambda\|}^{(i)} &\triangleq \left.\frac{\partial f_{\|}^{(i)}}{\partial \tilde{\boldsymbol{\lambda}}}\right|_{\hat{x}_{k|k}} & \hat{J}_{\lambda\perp}^{(i)} &\triangleq \left.\frac{\partial f_{\perp}^{(i)}}{\partial \tilde{\boldsymbol{\lambda}}}\right|_{\hat{x}_{k|k}} . \\
\hat{J}_{o\|}^{(i)} &\triangleq \left.\frac{\partial f_{\|}^{(i)}}{\partial \tilde{\boldsymbol{o}}}\right|_{\hat{x}_{k|k}} & \hat{J}_{o\perp}^{(i)} &\triangleq \left.\frac{\partial f_{\perp}^{(i)}}{\partial \tilde{\boldsymbol{o}}}\right|_{\hat{x}_{k|k}}
\end{aligned} \tag{9.83}$$

Consequently,

$$\begin{aligned} \boldsymbol{r}_{k+1} &= \hat{f}_r + \hat{J}_r(\boldsymbol{r}_k - r_{k|k}) + \hat{J}_{rp}(\boldsymbol{p}_k - p_{k|k}) + \boldsymbol{w}_r \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + A_{pr}(\boldsymbol{r}_k - r_{k|k}) + \boldsymbol{w}_p \end{aligned} \tag{9.84}$$

where $\hat{f}_r \triangleq f_r(\hat{r}_{k|k}, \hat{p}_{k|k})$ and

$$\hat{J}_r \triangleq \begin{bmatrix} I_2 & J_\lambda & J_o \\ 0_{\Lambda\times 2} & A_\lambda & 0_{\Lambda\times O} \\ 0_{O\times 2} & 0_{O\times\Lambda} & A_o \end{bmatrix}_{\hat{x}_{k|k}} \qquad \hat{J}_{rp} \triangleq \begin{bmatrix} J_\theta & 0_{2\times 1} & 0_{2\times 1} \\ 0_{\Lambda\times 1} & 0_{\Lambda\times 1} & 0_{\Lambda\times 1} \\ 0_{O\times 1} & 0_{O\times 1} & 0_{O\times 1} \end{bmatrix}_{\hat{x}_{k|k}} \tag{9.85}$$

In conclusion, the linearized form of the Lambda:Omicron model is

$$\boldsymbol{x}_{k+1} = f(\hat{x}_{k|k}) + \hat{J}_x(\boldsymbol{x}_k - \hat{x}_{k|k}) + \boldsymbol{w} \tag{9.86}$$

where $\hat{J}_x = \hat{J}_x(\hat{x}_{k|k})$ and the searched motion Jacobian is

$$J_x = \begin{bmatrix} J_r & J_{rp} \\ A_{pr} & I_3 \end{bmatrix} \tag{9.87}$$

9.4.3 Basic factors and Jacobians

Due to the their implicit definitions, explicit formulae for the motion function $f(\cdot)$ and for the motion Jacobian J_x are not yet found. In this section this problem is solved by employing the results expressed by the special decomposition of the motion functions.

Basic factors

The motion function $f(\cdot)$ requires the explicit expressions of the scalars $k_\parallel^{(i)}$ and $k_\perp^{(i)}$, here called *basic factors*. A practical way compute such factors is by using the recursion

$$\begin{aligned} k_\parallel^{(i+1)} &= \dot{k}_\parallel^{(i)} - \boldsymbol{\omega} k_\perp^{(i)} \\ k_\perp^{(i+1)} &= \boldsymbol{\omega} k_\parallel^{(i)} + \dot{k}_\perp^{(i)} \end{aligned} \tag{9.88}$$

initialized with

$$k_\parallel^{(1)} \triangleq \boldsymbol{v} \qquad k_\perp^{(1)} \triangleq 0 \tag{9.89}$$

Basic Jacobians

The motion Jacobian J_x requires the explicit expressions of the Jacobians. Such Jacobians can be computed via the following practical formulae

$$\begin{aligned}
\hat{J}^{(i)}_{\theta\|} &= \begin{bmatrix} -\sin\theta_{k|k} \\ \cos\theta_{k|k} \end{bmatrix} k^{(i)}_{\|} & \hat{J}^{(i)}_{\theta\perp} &= \begin{bmatrix} \cos\theta_{k|k} \\ \sin\theta_{k|k} \end{bmatrix} (-k^{(i)}_{\perp}) \\
\hat{J}^{(i)}_{\lambda\|} &= \begin{bmatrix} \cos\theta_{k|k} \\ \sin\theta_{k|k} \end{bmatrix} \hat{j}^{(i)}_{\lambda\|} & \hat{J}^{(i)}_{\lambda\perp} &= \begin{bmatrix} -\sin\theta_{k|k} \\ \cos\theta_{k|k} \end{bmatrix} \hat{j}^{(i)}_{\lambda\perp}. \\
\hat{J}^{(i)}_{o\|} &= \begin{bmatrix} \cos\theta_{k|k} \\ \sin\theta_{k|k} \end{bmatrix} \hat{j}^{(i)}_{o\|} & \hat{J}^{(i)}_{o\perp} &= \begin{bmatrix} -\sin\theta_{k|k} \\ \cos\theta_{k|k} \end{bmatrix} \hat{j}^{(i)}_{o\perp}
\end{aligned} \tag{9.90}$$

where are introduced the *basic Jacobians*

$$\begin{aligned}
\hat{j}^{(i)}_{\lambda\|} &\triangleq \left.\frac{\partial k^{(i)}_{\lambda\|}}{\partial \boldsymbol{\lambda}}\right|_{\lambda_{k|k}, o_{k|k}} & \hat{j}^{(i)}_{\lambda\perp} &\triangleq \left.\frac{\partial k^{(i)}_{\lambda\perp}}{\partial \boldsymbol{\lambda}}\right|_{\lambda_{k|k}, o_{k|k}} \\
\hat{j}^{(i)}_{o\|} &\triangleq \left.\frac{\partial k^{(i)}_{o\|}}{\partial \boldsymbol{o}}\right|_{\lambda_{k|k}, o_{k|k}} & \hat{j}^{(i)}_{o\perp} &\triangleq \left.\frac{\partial k^{(i)}_{o\perp}}{\partial \boldsymbol{o}}\right|_{\lambda_{k|k}, o_{k|k}}
\end{aligned} \tag{9.91}$$

Explicit expressions

- **initialisation** $i = 1$: basic factors

$$\begin{aligned}
k^{(1)}_{\|} &\triangleq v \\
k^{(1)}_{\perp} &\triangleq 0
\end{aligned} \tag{9.92}$$

  basic Jacobians

$$\begin{aligned}
j^{(1)}_{\|\bar{\lambda}} &\triangleq 1 \\
j^{(1)}_{\|\bar{o}} &\triangleq 0 \\
j^{(1)}_{\perp\bar{\lambda}} &\triangleq 0 \\
j^{(1)}_{\perp\bar{o}} &\triangleq 0
\end{aligned} \tag{9.93}$$

- **case** $i = 2$: basic factors

$$\begin{aligned}
k^{(2)}_{\|} &\triangleq \dot{v} \\
k^{(2)}_{\perp} &\triangleq \omega v
\end{aligned} \tag{9.94}$$

basic Jacobians

$$\begin{aligned}
j_{\|\bar{\lambda}}^{(2)} &\triangleq \begin{bmatrix} 0 & 1 \end{bmatrix} \\
j_{\|\bar{o}}^{(2)} &\triangleq 0 \\
j_{\perp\bar{\lambda}}^{(2)} &\triangleq \begin{bmatrix} \omega_{k|k} & 0 \end{bmatrix} \\
j_{\perp\bar{o}}^{(2)} &\triangleq v_{k|k}
\end{aligned} \tag{9.95}$$

- **case** $i = 3$: basic factors

$$\begin{aligned}
k_{\|}^{(3)} &\triangleq \ddot{v} - \omega^2 v \\
k_{\perp}^{(3)} &\triangleq 2\omega\dot{v} + \dot{\omega} v
\end{aligned} \tag{9.96}$$

basic Jacobians

$$\begin{aligned}
j_{\|\bar{\lambda}}^{(3)} &\triangleq \begin{bmatrix} -\omega_{k|k}^2 & 0 & 1 \end{bmatrix} \\
j_{\|\bar{o}}^{(3)} &\triangleq \begin{bmatrix} -2\omega_{k|k} v_{k|k} & 0 \end{bmatrix} \\
j_{\perp\bar{\lambda}}^{(3)} &\triangleq \begin{bmatrix} \dot{\omega}_{k|k} & 2\omega_{k|k} & 0 \end{bmatrix} \\
j_{\perp\bar{o}}^{(3)} &\triangleq \begin{bmatrix} 2\dot{v}_{k|k} & v_{k|k} \end{bmatrix}
\end{aligned} \tag{9.97}$$

- **case** $i = 4$: basic factors

$$\begin{aligned}
k_{\|}^{(4)} &\triangleq \dddot{v} - 3\omega\dot{\omega} v - 3\omega^2 \dot{v} \\
k_{\perp}^{(4)} &\triangleq 3\omega\ddot{v} + 3\dot{\omega}\dot{v} + \ddot{\omega} v - \omega^3 v
\end{aligned} \tag{9.98}$$

basic Jacobians

$$\begin{aligned}
j_{\|\bar{\lambda}}^{(4)} &\triangleq \begin{bmatrix} -3\omega_{k|k}\dot{\omega}_{k|k} & -3\omega_{k|k}^2 & 0 & 1 \end{bmatrix} \\
j_{\|\bar{o}}^{(4)} &\triangleq \begin{bmatrix} -3\dot{\omega}_{k|k} v_{k|k} - 6\omega_{k|k}\dot{v}_{k|k} & -3\omega_{k|k} v_{k|k} & 0 \end{bmatrix} \\
j_{\perp\bar{\lambda}}^{(4)} &\triangleq \begin{bmatrix} \ddot{\omega}_{k|k} - \omega_{k|k}^3 & 3\dot{\omega}_{k|k} & 3\omega_{k|k} \end{bmatrix} \\
j_{\perp\bar{o}}^{(4)} &\triangleq \begin{bmatrix} 3\ddot{v}_{k|k} - 3\omega_{k|k}^2 v_{k|k} & 3\dot{v}_{k|k} & v_{k|k} \end{bmatrix}
\end{aligned} \tag{9.99}$$

9.5 LO-MEM corrector

At each time step $k + 1$, the LO-MEM corrector consists in the following BLUE operations

$$\begin{aligned}
\hat{x}_{k+1|k+1} &= \hat{x}_k + \Sigma_{\boldsymbol{x}\bar{\boldsymbol{\mathcal{Y}}}} \Sigma_{\overline{\boldsymbol{\mathcal{Y}}}}^{-1} (\overline{\mathcal{Y}}_{k+1} - \overline{\mathcal{Y}}_{k+1|k}) \\
P_{k+1|k+1} &= P_{k+1|k} - \Sigma_{\boldsymbol{x}\bar{\boldsymbol{\mathcal{Y}}}} \Sigma_{\overline{\boldsymbol{\mathcal{Y}}}}^{-1} \Sigma'_{\boldsymbol{x}\bar{\bar{\mathcal{Y}}}}
\end{aligned} \tag{9.100}$$

where the considered observation is

$$\overline{\boldsymbol{\mathcal{Y}}} \triangleq \left[\bar{\boldsymbol{y}}' \quad \overline{\boldsymbol{Y}}'\right]' \tag{9.101}$$

and the relative moments are computed, according to the mean MEM and the mean pseudo-measurement model, as follows

- **prediction**: the prediction $\overline{\mathcal{Y}}_{k+1|k}$ has the following structure

$$\overline{\mathcal{Y}}_{k+1|k} = \left[\bar{y}'_{k+1|k} \quad \overline{Y}'_{k+1|k}\right]' \tag{9.102}$$

  ;

- **covariance**: the covariance $\Sigma_{\overline{\boldsymbol{\mathcal{Y}}}}$ has the following structure

$$\Sigma_{\overline{\boldsymbol{\mathcal{Y}}}} = \begin{bmatrix} \Sigma_{\bar{\boldsymbol{y}}} & \Sigma_{\bar{\boldsymbol{y}}\overline{\boldsymbol{Y}}} = 0 \\ \cdot & \Sigma_{\overline{\boldsymbol{Y}}} \end{bmatrix} \tag{9.103}$$

  The mixed term $\Sigma_{\bar{\boldsymbol{y}}\overline{\boldsymbol{Y}}}$ is zero because $\bar{\boldsymbol{y}}$ and $\overline{\boldsymbol{Y}}$ are independently distributed;

- **cross-covariance**: the cross-covariance $\Sigma_{\boldsymbol{x}\overline{\boldsymbol{\mathcal{Y}}}}$ has the following structure

$$\Sigma_{\boldsymbol{x}\overline{\boldsymbol{\mathcal{Y}}}} = \left[\Sigma_{\boldsymbol{x}\bar{\boldsymbol{y}}} \quad \Sigma_{\boldsymbol{x}\overline{\boldsymbol{Y}}}\right] \tag{9.104}$$

9.6 PHD implementation

Since the LO-MEM filter considers a non-linear motion model and a non-linear measurement model, both the PHD predictor and corrector are defined heuristically. The idea is to embed the estimates generated by the LO-MEM filter in the Gaussian mixture defined by the GM-PHD filter.

9.6.1 Predictor

The state of a generic extended object is defined as

$$\boldsymbol{x} \triangleq \left[\boldsymbol{r}' \quad \boldsymbol{p}'\right]' \tag{9.105}$$

and the considered predicted PHD is

$$D_{k|k-1}(x) = D_{\mathrm{B}}(x) + D_{\mathrm{S}}(x) \tag{9.106}$$

where the new born objects PHD $D_{\mathrm{B}}(\cdot)$ is

$$D_{\mathrm{B}}(x) = \sum_{i=1}^{\nu_{\mathrm{B}}} w_{\mathrm{B}}\, \mathcal{N}(x; x_{\mathrm{B},i}, P_{\mathrm{B},i}) \tag{9.107}$$

while the survived objects $D_{\mathrm{S}}(\cdot)$, assuming p_{S} constant, is given by

$$\begin{aligned} D_{\mathrm{S}}(x) = &\sum_{i=1}^{\nu_{k-1|k-1}} w_{k|k-1,i}\, \mathcal{N}(x; x_{k|k-1,i}, P_{k|k-1,i}) \\ w_{k|k-1,i} &\triangleq p_{\mathrm{S}}\, w_{k-1|k-1,i} \\ x_{k|k-1,i} &\triangleq f(x_{k-1|k-1,i}) \\ P_{k-1|k-1,i} &\triangleq \hat{J}_x P_{k-1|k-1,i} \hat{J}'_x + Q \end{aligned} \tag{9.108}$$

where $f(\cdot)$ is the Lambda:Omicron motion function and $\hat{J}_x$ is its Jacobian evaluated in $\hat{x}_{k-1|k-1,i}$

9.6.2 Corrector

The considered corrected PHD is

$$D_{k|k}(x) = D_{\mathrm{ND}}(x) + \sum_{\mathcal{P} \boxminus \mathsf{y}} \sum_{\mathsf{w} \in \mathcal{P}} D_{\mathrm{D}}(x; \mathcal{P}, \mathsf{w}) \tag{9.109}$$

where

- assuming p_{D} and λ_{D} constant, the non-detected objects PHD $D_{\mathrm{ND}}(\cdot)$ is

$$\begin{aligned} D_{\mathrm{ND}}(x) = &\sum_{i=1}^{\nu_{k|k-1}} w_{k|k,i} \mathcal{N}(x; x_{k|k-1,i}, P_{k|k-1,i}) \\ w_{k|k,i} &\triangleq [1 - (1 - \exp(-\lambda_{\mathrm{D}}))\, p_{\mathrm{D}}]\, w_{k|k-1,i} \end{aligned} \tag{9.110}$$

- the detected objects PHD $D_{\mathrm{D}}(\cdot;\mathcal{P},\mathsf{w})$ is given by

$$\begin{aligned}
D_{\mathrm{D}}(x;\mathcal{P},\mathsf{w}) &= \sum_{i=1}^{\nu_{k|k-1}} w_{k|k}^{\mathcal{P},\mathsf{w},i}\mathcal{N}(x;x_{k|k}^{\mathsf{w},i},P_{k|k}^{\mathsf{w},i}) \\
w_{k|k}^{\mathcal{P},\mathsf{w},i} &\triangleq \omega_{\mathcal{P}}\frac{\tilde{\ell}_{\mathsf{w},i}}{d_{\mathsf{w}}} \\
\tilde{\ell}_{\mathsf{w},i} &\triangleq p_{\mathrm{D}}\exp(-\lambda_{\mathrm{D}})\left(\frac{\lambda_{\mathrm{D}}}{I_{\mathrm{C}}}\right)^{|\mathsf{w}|}\mathcal{L}_{\mathsf{w},i} \\
d_{\mathsf{w}} &\triangleq \delta_1(|\mathsf{w}|)+\sum_{i=1}^{\nu_{k|k-1}}\tilde{\ell}_{\mathsf{w},i} \\
\omega_{\mathcal{P}} &\triangleq \frac{\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}}{\sum_{\mathcal{P}\boxminus y}\prod_{\mathsf{w}\in\mathcal{P}} d_{\mathsf{w}}}
\end{aligned} \tag{9.111}$$

where the cell likelihood and the corrected parameters are defined as follows.

Cell likelihood

The cell likelihood is defined according to the model of the joint vector

$$\overline{\boldsymbol{\mathcal{Y}}} \triangleq \begin{bmatrix}\bar{\boldsymbol{y}}' & \overline{\boldsymbol{Y}}'\end{bmatrix}' \tag{9.112}$$

now consider the following facts:

- the model for $\bar{\boldsymbol{y}}$ is

$$p_{\bar{\boldsymbol{y}}|i}(\bar{y}) \triangleq \mathcal{N}(\bar{y};\bar{y}_{k|k-1}^{i},\Sigma_{\bar{\boldsymbol{y}}}^{i}) \tag{9.113}$$

  where, according to the mean MEM moments,

$$\begin{aligned}
\bar{y}_{k|k-1}^{i} &\triangleq H\hat{r}_{k|k-1,i} \\
\Sigma_{\bar{\boldsymbol{y}}}^{i} &\triangleq HP_{k|k-1}^{r,i}H' + \frac{C^{I,i}+C^{II,i}+R_v}{|\mathsf{w}|}
\end{aligned} \tag{9.114}$$

  and $\hat{r}_{k|k-1,i}$, $P_{k|k-1}^{r,i}$, $C^{I,i}$, $C^{II,i}$ are moments relative to the predicted object i. This model is reasonable if the covariance of $\boldsymbol{p}_{k|k-1,i}$ is small: in this case $\boldsymbol{y}$ is, with a good approximation, Gaussian;

- the model for $\overline{\boldsymbol{Y}}$ is

$$p_{\overline{\boldsymbol{Y}}|i}(\overline{Y}) \triangleq \mathcal{W}_2\left(Y;|\mathsf{w}|-1,\frac{\Sigma_{\boldsymbol{y}}^{i}}{|\mathsf{w}|-1}\right) \tag{9.115}$$

where, according to the pseudo-measurement model,

$$\Sigma_{\boldsymbol{y}}^{i} = HP_{k|k-1}^{r,i}H' + C^{I,i} + C^{II,i} + R_v \tag{9.116}$$

then, by exploiting the fact that $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{Y}}$ are independently distributed (due to theorem (???)), the cell likelihood gets the explicit expression

$$\begin{aligned}\mathcal{L}_{\mathsf{w},i} &\triangleq p_{\overline{\boldsymbol{\mathcal{Y}}}|i}(\overline{\mathcal{Y}}) = p_{\bar{\boldsymbol{y}}|i}(\bar{y})p_{\overline{\boldsymbol{Y}}|i}(\overline{Y}) \\ &= \mathcal{N}(\bar{y};\bar{y}_{k|k-1}^{i},\Sigma_{\bar{\boldsymbol{y}}}^{i})\mathcal{W}_2(\overline{Y};|\mathsf{w}|-1,\Sigma_{\boldsymbol{y}}^{i})\end{aligned} \tag{9.117}$$

Corrected parameters

The $|\mathcal{P}|\cdot\nu_{k|k-1}$ corrected parameters $x_{k|k}^{\mathsf{w},i}$, $P_{k|k}^{\mathsf{w},i}$ are computed by an ensamble of $|\mathcal{P}|\cdot\nu_{k|k-1}$ LO-MEM filters, where each filter performs the correction of the predicted object i, which is characterized by the predicted parameters $x_{k|k-1}^{\mathsf{w},i}$, $P_{k|k-1}^{\mathsf{w},i}$, according to the cell of measurements w.

Part III

Simulations

Chapter 10

Simulations

10.1 Summary

In this final chapter are shown the numerical results about two different simulations modelling a naval tracking problem where different boats moves in fixed sourveilled area (which is a square of $[-1500, 1500]^2$ meters).

- **simulation 1**: in this simulation is considered a single boat that performs an highly manuevering trajectory in the sourveilled area. The trajectory is characterized by relevant variations of longitudinal speed and steering speed;

- **simulation 2**: in this simulation are considered five different boats that moves around the sourveilled area. In this simulation the number of boats simultaneously present in the scene is a time varying quantity because the boats can enter or leave the sourvelleid area.

In the first simulation are compared the LO-MEM filter, in its version 2:1-MEM, with the MEM-EKF* filter, in its version CT-NCV. In particular, it is shown that the LO-MEM filter can achieve better performance than the MEM-EKF* filter. In the second simulation is shown the effectiveness of the PHD extesion of the LO-MEM filter in a multiobject scenario.

The GIW filter is not considered because it was already prooved that the MEM-EKF* filter achieves better performance.

The interessed reader can find the source code, written in MATLAB, on the GitHub page of the author.

10.2 Simulation 1

10.2.1 Ground truth

motion equations

The trajectory of the tracked boat is given by the unicycle model

$$\begin{aligned} m_{k+1} &= m_k + T \begin{bmatrix} \cos\theta_k \\ \sin\theta_k \end{bmatrix} u_k^v \\ \theta_{k+1} &= \theta_k + T u_k^\omega \end{aligned} \tag{10.1}$$

where u^v and u^ω, seen as driving inputs, are the longitudinal velocity and the steering speed of the boat. In order to define smooth u^v and u^ω in a simple manner, it is considered the following two additional equations

$$\begin{aligned} u_{k+1}^v &= u_k^v + T u_k^{\dot{v}} \\ u_{k+1}^\omega &= u_k^\omega + T u_k^{\dot{\omega}} \end{aligned} \tag{10.2}$$

which allows to define u^v and u^ω in terms of integrals of the longitudinal acceleration $u^{\dot{v}}$ and the of the steering acceleration $u^{\dot{\omega}}$.

initial state

The initial kinematic state considered is the following

$$\begin{aligned} m_0 &\triangleq \begin{bmatrix} -562.5 & -562.5 \end{bmatrix}' \qquad \text{(in meters)} \\ \theta_0 &\triangleq \frac{\pi}{2} \qquad \text{(in radiants)} \\ v_0 &\triangleq 24 \qquad \text{(in kilonodes / seconds)} \\ \omega_0 &\triangleq 0 \qquad \text{(in radiants / seconds)} \end{aligned} \tag{10.3}$$

the dimensions (width and lenght) of the boat are respectively

$$\begin{aligned} l_1 &\triangleq 100 \qquad \text{(in meters)} \\ l_2 &\triangleq 10 \qquad \text{(in meters)} \end{aligned} \tag{10.4}$$

input signals

The sampling interval considered is $T = 1$ second and the length of the simulation is 180 seconds. The driving inputs $u_k^{\dot{v}}$ and $u_k^{\dot{\omega}}$ are defined as the

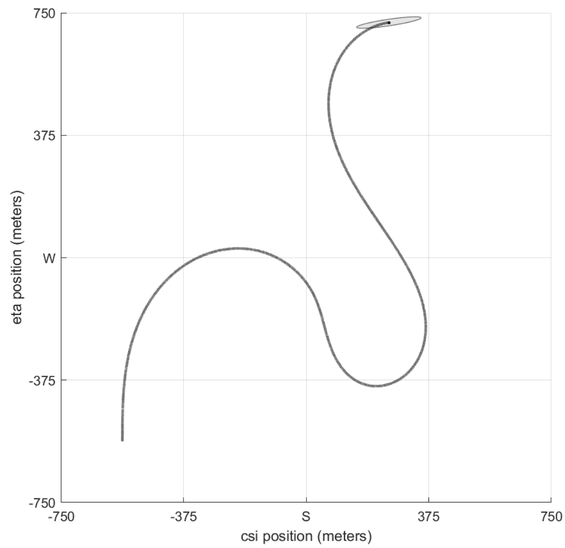

Figure 10.1: Ground truth trajectory. At the ending point is shown the ground truth ellipse representing the tracked boat.

following piece-wise constant signals

$$u_k^{\dot{v}} \triangleq \begin{cases} 0.1 & \text{if} \quad 0 \leq k < 60 \\ 0 & \text{if} \quad 60 \leq k < 75 \\ 0.1 & \text{if} \quad 75 \leq k < 95 \\ 0 & \text{if} \quad 95 \leq k < 180 \end{cases} \qquad \text{(in kilonodes / seconds}^2\text{)} \tag{10.5}$$

$$u_k^{\dot{\omega}} \triangleq \begin{cases} -0.001 & \text{if} \quad 0 \leq k < 60 \\ 0 & \text{if} \quad 60 \leq k < 75 \\ 0.01 & \text{if} \quad 75 \leq k < 95 \\ -0.004 & \text{if} \quad 95 \leq k < 125 \\ -0.002 & \text{if} \quad 125 \leq k < 180 \end{cases} \qquad \text{(in radiants / seconds}^2\text{)} \tag{10.6}$$

measurement generation

The measurement equation employed is the MEM equation

$$\begin{aligned} y_k^{(i)} &= m_k + S(p_k)h_k^{(i)} + v_k^{(i)} \\ h_k^{(i)} &\sim \mathcal{U}(C_{0,1}) \\ v_k^{(i)} &\sim \mathcal{N}(0, R_v) \end{aligned} \tag{10.7}$$

where m_k and p_k are the position and shape state of the tracked boat.

At each sampling instant k are generated 5 different measurements by generating stochasticly 5 different multiplicative noises and measurement noises $\{h_k^{(i)}, v_k^{(i)}\}_{i=1}^5$.

The choosen measurement covariance is

$$R_v \triangleq \text{diag}(15^2, 15^2) \qquad (\text{in meters}^2) \tag{10.8}$$

At each sampling instant k no clutter measurements are generated.

10.2.2 PHD MEM-EKF* setup

PHD setup

Since in the actual simulation there is only one object that does not leave the scene and since are not modelled blind spots in the sourveilled area, the considered PHD parameters are the following

$$p_{\text{D}} \triangleq 1 \qquad p_{\text{S}} \triangleq 1 \tag{10.9}$$

Since no additional objects enter in the scene during the simulation, the new born object mixture is neglected.

Due to the fact that no clutter measurements are generated, it is considered

$$\lambda_{\text{C}} \triangleq 0 \tag{10.10}$$

MEM-EKF* setup

The filter is initialized with a mixture composed by only one component, which is characterized by the following parameters

$$w_{0|0} \triangleq 1$$

$$\hat{x}_{0|0} = \begin{bmatrix} \hat{\xi}_0 \\ \hat{\eta}_0 \\ \hat{\dot{\xi}}_0 \\ \hat{\dot{\eta}}_0 \\ \hat{\omega}_0 \\ \hat{\theta}_0 \\ \hat{l}_{1,0} \\ \hat{l}_{2,0} \end{bmatrix} \triangleq \begin{bmatrix} -700 \\ -700 \\ 0 \\ 0 \\ 0 \\ \frac{\pi}{3} \\ 50 \\ 25 \end{bmatrix} \tag{10.11}$$

$$P_{0|0} \triangleq \text{diag}\,(1000, 1000, 100, 100, 0.01, 0.01, 200, 100)$$

the considered process noise covariance is

$$Q \triangleq \text{diag}\,(100, 100, 0.01, 0.01, 0.00005, 0, 0.000001, 0.00000001) \tag{10.12}$$

10.2.3 PHD LO-MEM setup

PHD setup

See the PHD setup of the previous PHD-MEM-EKF* filter.

LO-MEM setup

The filter is initialized with a mixture composed by only one component, which is characterized by the following parameters

$$w_{0|0} \triangleq 1$$

$$\hat{x}_{0|0} = \begin{bmatrix} \hat{\xi}_0 \\ \hat{\eta}_0 \\ \hat{v}_0 \\ \hat{\dot{v}}_0 \\ \hat{\omega}_0 \\ \hat{\theta}_0 \\ \hat{l}_{1,0} \\ \hat{l}_{2,0} \end{bmatrix} \triangleq \begin{bmatrix} -700 \\ -700 \\ 0 \\ 0 \\ 0 \\ \frac{\pi}{3} \\ 50 \\ 25 \end{bmatrix} \tag{10.13}$$

$$P_{0|0} \triangleq \text{diag}\,(1000, 1000, 100, 100, 0.01, 0.01, 200, 100)$$

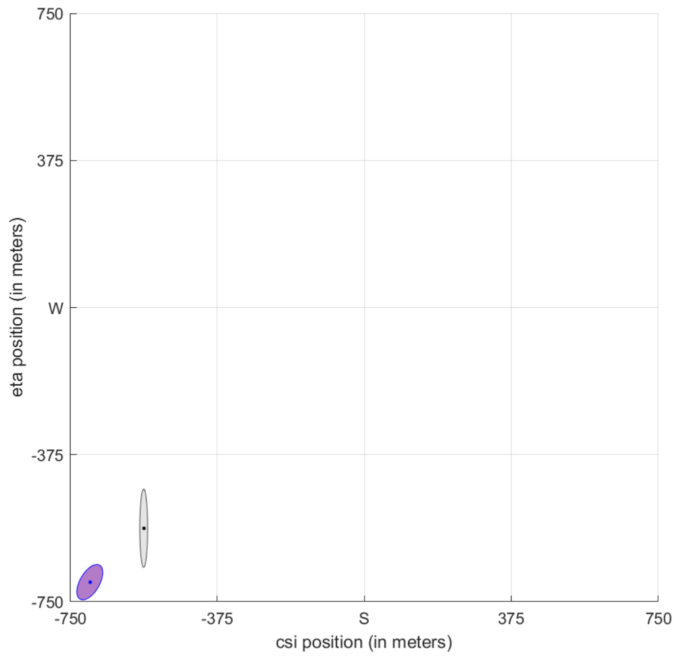

Figure 10.2: Initial condition. The grey ellipse represents the ground truth, the blue ellipse represents the initialization of the MEM-EKF* filter, the red ellipse represents the initialization of the 2:1-MEM filter

the considered process noise covariance is

$$Q \triangleq \text{diag}\,(0, 0, 0.01, 0.005, 0.00005, 0, 0.000001, 0.00000001) \qquad (10.14)$$

10.2.4 Error metric

At each sampling instant k, the performances of the two filters are measured with the so-called *Wasserstein distance* between the ground truth ellipse and the estimated ellipses generated by the filters.

The Wasserstein distance is a metric that measures the difference between two ellipses $E_1 = (m_1, \Sigma_1)$, $E_2 = (m_2, \Sigma_2)$, where m_1, m_2 are the centers and Σ_1, Σ_2 are the shape matrix of the ellipses.

The square of the Wasserstein distance, which is measured in meters^2, is defined as

$$W^2(E_1, E_2) \triangleq \|m_1 - m_2\|^2 + \mathrm{tr}\left[\Sigma_1 + \Sigma_2 - 2\sqrt{\sqrt{\Sigma_1}\Sigma_2\sqrt{\Sigma_1}}\right] \tag{10.15}$$

where $\sqrt{\Sigma_1}$ is a matrix such that $\sqrt{\Sigma_1}\sqrt{\Sigma_1} = \Sigma_1$. The first term $\|m_1 - m_2\|^2$ takes into account the position error between the two ellipses, the second term $\mathrm{tr}[\cdots]$ takes into account the misalignement and the differences in the extensions of the two ellipses.

Let $E_{\mathrm{g},k}$, $E_{\text{MEM-EKF*},k}$, $E_{\text{LO-MEM},k}$ be the ground truth ellipse, the estimated ellipse by the MEM-EKF* filter and the estimated ellipse by the LO-MEM filter at time k. The performance indexes considered are

$$\begin{aligned} CW_{\text{MEM-EKF*}} &\triangleq \sum_{k=0}^{180} W^2(E_{\mathrm{g},k}, E_{\text{MEM-EKF*},k}) \\ CW_{\text{LO-MEM}} &\triangleq \sum_{k=0}^{180} W^2(E_{\mathrm{g},k}, E_{\text{LO-MEM},k}) \end{aligned} \tag{10.16}$$

referred as *cumulative Wasserstein errors.*

10.2.5 Result

The cumulative Wasserstein errors produced by the two filters are respectively

$$\begin{aligned} CW_{\text{MEM-EKF*}} &\approx 5314 \\ CW_{\text{LO-MEM}} &\approx 3193 \end{aligned} \tag{10.17}$$

This result suggests that, in particular circumstances such as simulation 1, the LO-MEM filter can achieve better performance than the MEM-EKF*.

10.3 Simulation 2

10.3.1 Ground truth

motion equations

The 5 boats follow the same motion model defined in simulation 1.

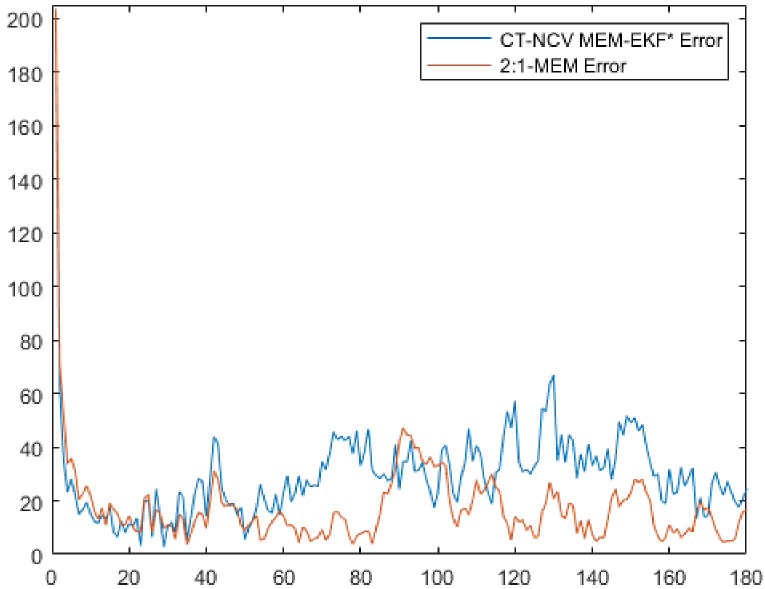

Figure 10.3: Time evolution of the Wasserstein errors generated by the two filters.

initial states

The considered initial states are the following

$$x_0^1 \triangleq \begin{bmatrix} -562.5 \\ -562.5 \\ 24 \\ 0.05 \\ \frac{\pi}{2} \\ 100 \\ 10 \end{bmatrix} \quad x_0^2 \triangleq \begin{bmatrix} 300 \\ -900 \\ 24 \\ -0.02 \\ \frac{\pi}{2} \\ 60 \\ 10 \end{bmatrix} \quad x_0^3 \triangleq \begin{bmatrix} 780 \\ -300 \\ 24 \\ 0.05 \\ \frac{2\pi}{3} \\ 70 \\ 15 \end{bmatrix} \quad x_0^4 \triangleq \begin{bmatrix} -300 \\ -750 \\ 24 \\ -0.03 \\ -\frac{\pi}{2} \\ 80 \\ 20 \end{bmatrix} \quad x_0^5 \triangleq \begin{bmatrix} 600 \\ 1200 \\ 24 \\ -0.01 \\ \frac{4\pi}{3} \\ 60 \\ 20 \end{bmatrix} \tag{10.18}$$

where x_0^i denotes the initial state of boat i.

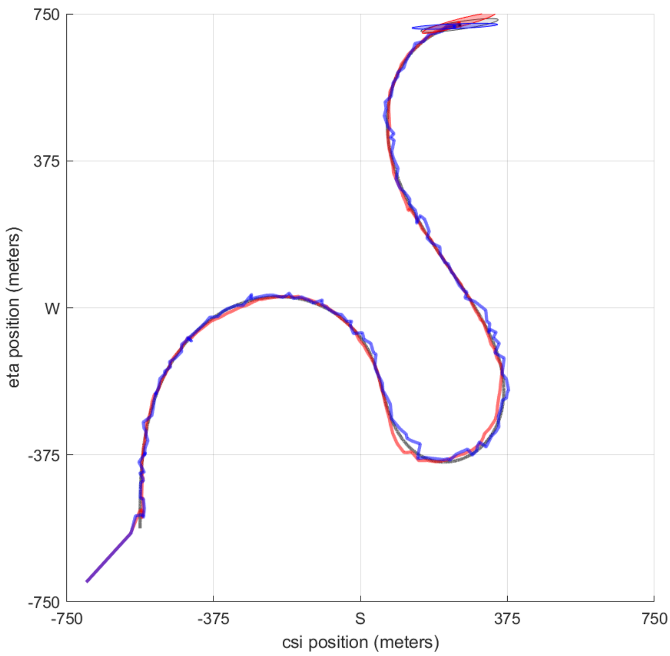

Figure 10.4: Final result of the simulation. In grey is represented the ground truth trajectory, in blue is represented the estimated trajectory generated by the MEM-EKF* filter, in red is represented the estimated trajectory generated by the 2:1-MEM filter.

input signals

The considered sampling interval is $T = 1$ second and the length of the simulation is 90 seconds. The driving inputs $u_k^{\dot{v},i}$, $u_k^{\dot{\omega},i}$ of boat i are defined as follows

$$u_k^{\dot{v},i} = u_k^{\dot{\omega},i} \triangleq 0 \qquad k = 0, 1, \ldots, 89, \qquad i = 1, 2, 3, 4, 5 \tag{10.19}$$

measurement generation

It is considered the same measurement model of simulation 1, with the following additional aspects:

- the number of measurements generated by a boat is Poisson with expected value $\lambda_{\mathrm{D}} \triangleq 5$;

- clutter is included in the simulation. It is assumed that the number of clutter measurements is Poisson with expected value $\lambda_{\mathrm{C}} \triangleq 10$. The clutter measurements are assumed to be IID random variables uniformly distributed over the entire scene, so that the true clutter intensity is $10/1500^2$.

10.3.2 PHD LO-MEM setup

PHD setup

The considered PHD parameters are

$$p_{\mathrm{D}} \triangleq 1 \qquad p_{\mathrm{D}} \triangleq 0.7 \tag{10.20}$$

The considered new born object PHD is composed by the following 4 components

$$\begin{aligned}
w_{\mathrm{B}}^{1} &\triangleq 10^{-5} \\
x_{\mathrm{B}}^{1} &\triangleq \begin{bmatrix}-750 & -750 & 0 & 0 & 0 & \frac{\pi}{4} & 60 & 15\end{bmatrix}' \\
P_{\mathrm{B}}^{1} &\triangleq \operatorname{diag}(100, 100, 15, 15, 0.1, 0.01, 1, 0.1) \\
w_{\mathrm{B}}^{2} &\triangleq 10^{-5} \\
x_{\mathrm{B}}^{2} &\triangleq \begin{bmatrix}750 & -750 & 0 & 0 & 0 & \frac{3\pi}{4} & 60 & 15\end{bmatrix}' \\
P_{\mathrm{B}}^{2} &\triangleq \operatorname{diag}(100, 100, 15, 15, 0.1, 0.01, 1, 0.1) \\
w_{\mathrm{B}}^{3} &\triangleq 10^{-5} \\
x_{\mathrm{B}}^{3} &\triangleq \begin{bmatrix}750 & 750 & 0 & 0 & 0 & \frac{5\pi}{4} & 60 & 15\end{bmatrix}' \\
P_{\mathrm{B}}^{3} &\triangleq \operatorname{diag}(100, 100, 15, 15, 0.1, 0.01, 1, 0.1) \\
w_{\mathrm{B}}^{4} &\triangleq 10^{-5} \\
x_{\mathrm{B}}^{4} &\triangleq \begin{bmatrix}-750 & 750 & 0 & 0 & 0 & \frac{7\pi}{4} & 60 & 15\end{bmatrix}' \\
P_{\mathrm{B}}^{4} &\triangleq \operatorname{diag}(100, 100, 15, 15, 0.1, 0.01, 1, 0.1)
\end{aligned} \tag{10.21}$$

The considered clutter intensity is

$$I_{\mathrm{C}} \triangleq \frac{10}{1500^2} \tag{10.22}$$

LO-MEM setup

The filter is initialized with the void initialization, i.e. the starting mixture does not contain any components.

10.3.3 Results

The performance of the filter is qualitatively represented by the following snapshots of the simulation. In such snapshots the ground truth ellipses are drawn in grey, while the estimated ellipse are drawn in red. Moreover are shown the measurements with the following convention:

- if the measurement is a detection, i.e. is generated by a boat, then it is drawn as a yellow circle;
- if the measurement is a clutter measurement then it is drawn as a red circle.

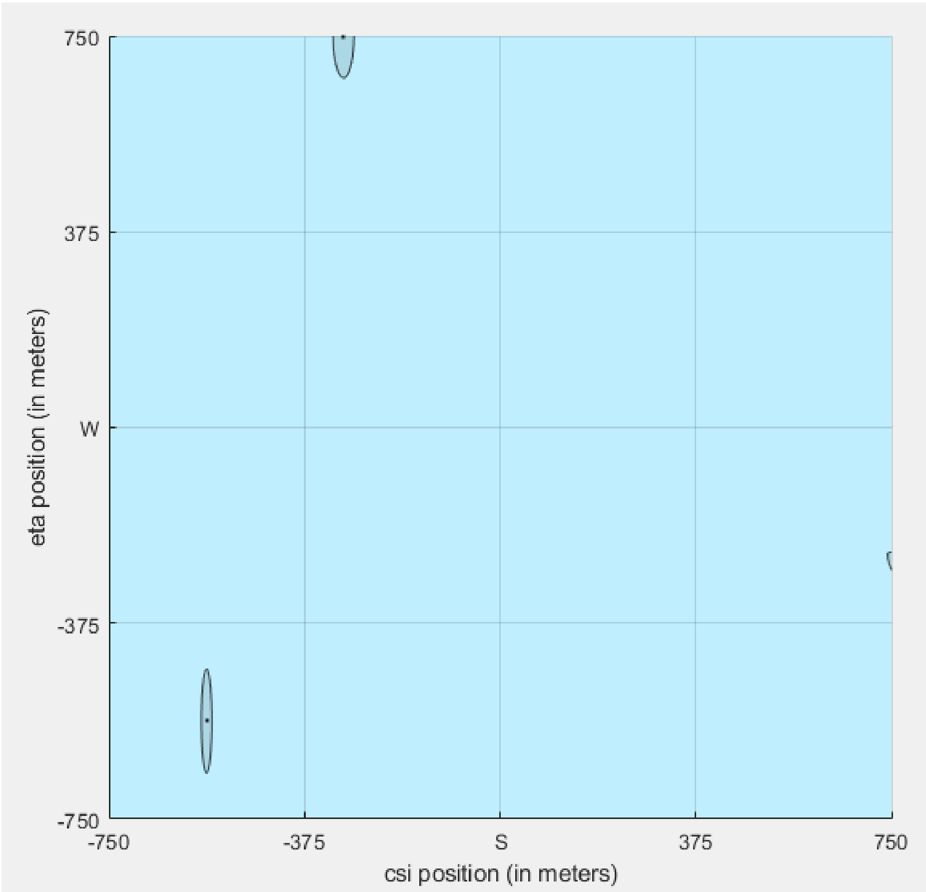

Figure 10.5: Initial condition

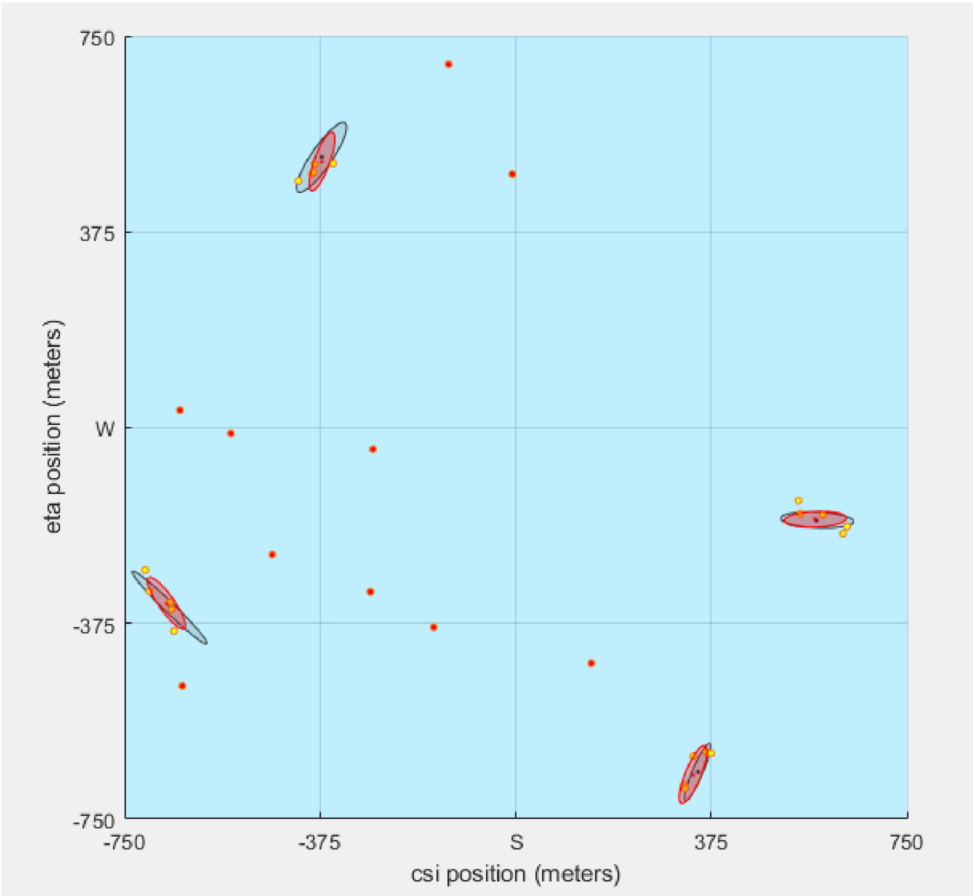

Figure 10.6: Simulation at time $k = 20$

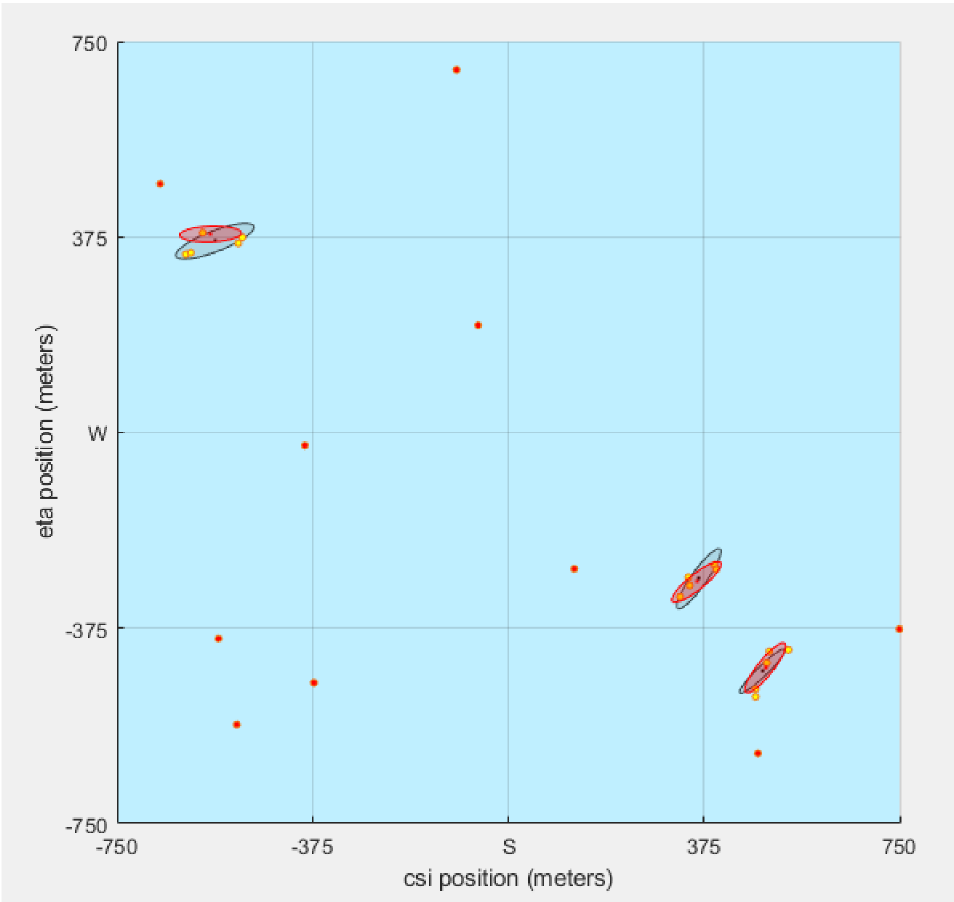

Figure 10.7: Simulation at time $k = 40$

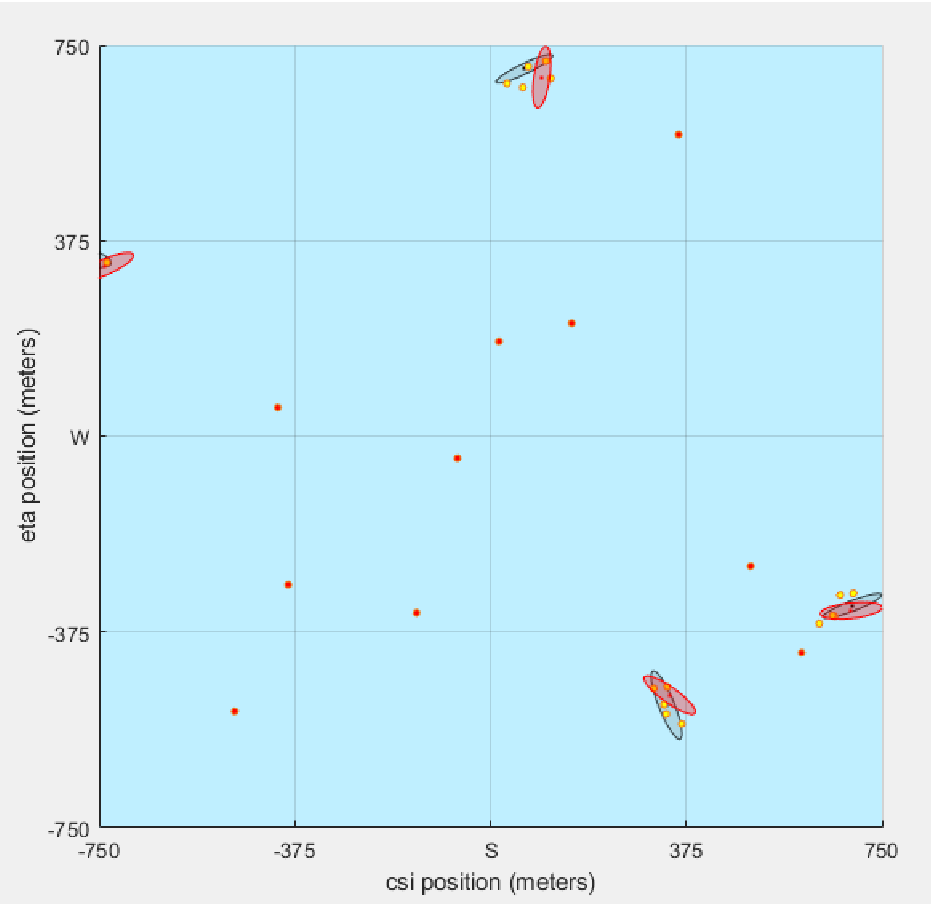

Figure 10.8: Simulation at time $k = 60$

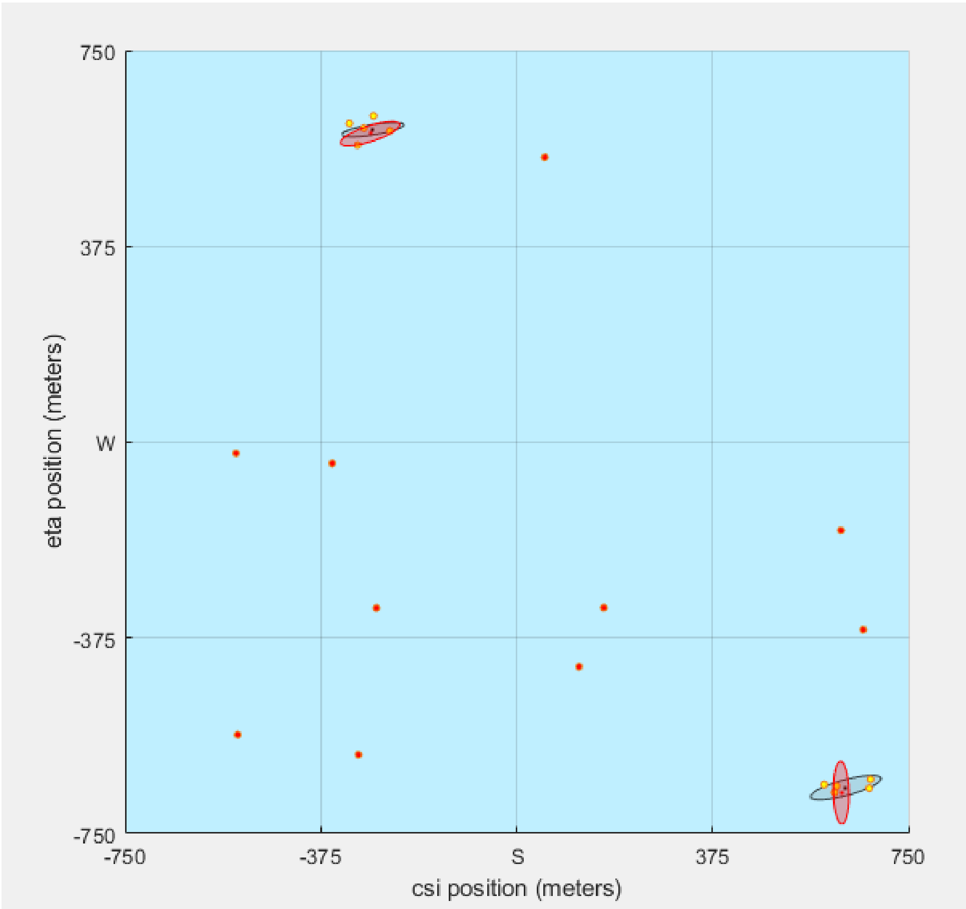

Figure 10.9: Simulation at time $k = 90$

Chapter 11

Conclusions and future work

The present work of thesis has dealt with the problem of *Multiple Extended Object Tracking* (MEOT), i.e. the joint estimation of the kinematic state and the shape state of an unknown and time varying number of extended objects. Such a problem involves two sources of difficulty:

- **1)** at any given sampling instant, likewise their states, it is not known the number of objects present in the surveilled scene;
- **2)** at any given sampling instant, an object can produce more than one measurement.

In order to deal with the first difficulty, the concept of random finite set was introduced. Thanks to the tools provided by FISST, i.e. the multiobject calculus and its generalization, a feasible RFS algorithm, i.e. the PHD filter, was obtained to solve the MEOT problem. More precisely, two different versions of the PHD filter were discussed:

- **standard PHD filter**: this type of filter can handle the simultaneous estimation of the kinematic states of an unknown and time varying number of point objects, i.e. objects that can generate no more than one measurement per time step;
- **extended object PHD filter**: this type of filter is the natural extension of the standard PHD filter, where the objects are considered extended, i.e. the simplifying assumption that an object can generate no more than one measurement per time step no longer holds.

Both the standard PHD filter and the extended object PHD filter are not natively designed to estimate the shape states of the objects present in the surveilled scene. In order to get a PHD filter capable to address this limitation, the problem of the shape estimation of a single extended object was discussed in detail, leading to two different algorithms:

- **GIW filter**: this type of filter is the simplest solution to the joint estimation of the kinematic and shape states of a single extended object. As a result, the GIW filter is computationally cheap but achieves poor performance in practical applications. The drawback is a consequence of the underlying simplifying assumptions. For example such filter, in order to get closed-form formulae, assumes that the covariance of the kinematic state is proportional to the extension (i.e., the area of the surface) of the object. Consequently, the bigger is the tracked object, the smaller is the precision of kinematic state estimation.

- **MEM-EKF* filter**: this type of filter solves some issues of the GIW filter, such as, for example, the just mentioned problem arising from the proportionality between the covariance of the kinematic state and the extension of the object. The MEM-EKF* filter is more accurate than the GIW filter, but it is characterized by an heavier computational burden. In general, the gain in the estimation accuracy justifies the additional computational cost, hence, the MEM-EKF* filter has to be preferred to the GIW filter.

For both filters, the PHD implementations was devised, leading to two algorithms that, finally, effectively solve the MEOT problem, i.e. they can estimate the kinematic state and the shape state of an unknown and time varying number of multiple extended objects.

The final topic of the present work was an attempt to improve the PHD filter based on the MEM-EKF* filter, i.e. the so-called LO-MEM filter. The present work leaves some open problems, which are discussed below and left for future work.

- **Sampling interval problem**: the LO-MEM filter is numerically unstable when the considered sampling interval is large. In particular, numerical simulations show that, when the sampling time is large, often the associated covariance matrix to the corrected estimate of the shape loses positive-definiteness;

- **Cell likelihood verification**: in the present work a new model for the cell likelihood, which is the kernel of a PHD filter, is proposed. However, the new model was not numerically tested, so that it is not clear if it is effective;

- **Longitudinal velocity assumption**: the LO-MEM motion model makes the strong assumption that the tracked object cannot move along the lateral direction. Due to this assumption, the LO-MEM filter can achieve better performance than MEM-EKF* but, clearly, in a real scenario this assumption does not hold. An interesting problem is to extend the LO-MEM filter in a way such that the limiting assumption can be avoided.

Bibliography

[1] I. Goodman, R. Mahler, H. Nguyen. *Mathematics of Data Fusion.* Kluwer Academic Publisher, 1997.

[2] D. Daley, D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition.* Springer, 2003.

[3] D. Daley, D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure, Second Edition.* Springer, 2008.

[4] R. Mahler. *Statistic 101 for Multisensor, Multitarget Data Fusion.* IEEE Aerospace and Electronic Systems Magazine, 2004.

[5] R. Mahler. *Statistic 102 for Multisource-Multitarget Detection and Tracking.* IEEE Journal of Selected Topics in Signal Processing, 2013.

[6] R. Mahler. *Statistic 103 for Multitarget Tracking.* MDPI, 2019.

[7] R. Mahler. *Statistical Multisource-Multitarget Information Fusion.* Artech House, 2007.

[8] R. Mahler. *Advances in Statistical Multisource-Multitarget Information Fusion.* Artech House, 2014.

[9] B. Vo, W. Ma. *The Gaussian Mixture Probability Hypothesis Density Filter.* IEEE Transactions on Signal Processing, 2006.

[10] R. Mahler. *PHD filters for nonstandard targets, I: Extended targets.* 12th International Conference on Information Fusion, 2009.

[11] K. Granstrom, C. Lundquist, U. Orguner. *Extended Target Tracking using a Gaussian-Mixture PHD filter.* IEEE Transactions on Aerospace and Electronic Systems, 2012.

[12] R Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, 1982.

[13] M. DeGroot. *Optimal Statistical Decisions*. John Wiley and Sons, 2004.

[14] A. Gupta, D. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000.

[15] W. Koch. *On Bayesian Tracking of Extended Objects*. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2006.

[16] W. Koch. *Bayesian Approach to Extended Object and Cluster Tracking using Random Matrices*. IEEE Transactions on Aerospace and Electronic Systems, 2008.

[17] W. Koch. *Comments on Bayesian Approach to Extended Object and Cluster Tracking using Random Matrices*. IEEE Transactions on Aerospace and Electronic Systems, 2012.

[18] K Granstrom, U. Orguner. *A PHD filter for tracking Multiple Extended Targets Using Random Matrices*. IEEE Transactions on Signal Processing, 2012.

[19] M. Baum, F. Faion, U. Hanebeck. *Modeling the target extent with multiplicative noise*. 15th International Conference on Information Fusion, 2012.

[20] M. Baum, S. Yang. *Extended Kalman Filter for Extended Object Tracking*. IEEE International Conference on Acoustics, Speech and Signal Processing, 2017.

[21] M. Baum, S. Yang. *Tracking the Orientation and Axes Lengths of an Elliptical Extended Object*. IEEE Transactions on Signal Processing, 2019.

[22] M. Baum, S. Yang, F Teich. *GM-PHD filter for multiple extended object tracking based on the multiplicative error shape model and network flow labeling*. IEEE Intelligent Vehicles Symposium, 2017.