Chapter 3
Estimation with Random Finite Sets

3.1 Introduction

The previous chapter provided the motivation to adopt an RFS representation for the map in both FBRM and SLAM problems. The main advantage of the RFS formulation is that the dimensions of the measurement likelihood and the predicted FBRM or SLAM state do not have to be compatible in the application of Bayes theorem, for optimal state estimation. The *implementation* of Bayes theorem with RFSs (equation 2.15) is therefore the subject of this chapter. It should be noted that in any realistic implementation of the *vector* based Bayes filter, the recursion of equation 2.13 is, in general, intractable. Hence, the well known extended Kalman filter (EKFs), unscented Kalman filter (UKFs) and higher order filters are used to approximate multi-feature, vector based densities. Unfortunately, the general RFS recursion in equation 2.15 is also mathematically intractable, since multiple integrals on the space of features are required. This chapter therefore introduces principled approximations which propagate approximations of the full multi-feature posterior density, such as the expectation of the map. Techniques borrowed from recent research in point process theory known as the *probability hypothesis density* (PHD) filter, *cardinalised probability hypothesis density* (C-PHD) filter, and the *multi-target, multi-Bernoulli* (MeMBer) filter, all offer principled approximations to RFS densities. A discussion on Bayesian RFS estimators will be presented, with special attention given to one of the simplest of these, the PHD filter. In the remaining chapters, variants of this filter will be explained and implemented to execute both FBRM and SLAM with simulated and real data sets.

The notion of Bayes optimality is equally as important as the Bayesian recursion of equation 2.15 itself. The following section therefore discusses optimal feature map estimation in the case of RFS based FBRM and SLAM, and once again, for clarity, makes comparisons with vector based

J. Mullane et al.: Random Finite Sets for Robot Mapping & SLAM, STAR 72, pp. 27–42.
springerlink.com

estimators. Issues with standard estimators are demonstrated, and optimal solutions presented.

3.2 Classical State Estimators

In this section, we pose a fundamental question: "Given the posterior distribution of the map/SLAM state, what is the Bayes "optimal" estimate?" While an RFS map representation can jointly encapsulate feature number and location uncertainty, the problem of extracting the optimal estimate from the posterior density (in the case of RFS SLAM), $p_{k|k}(X_{0:k}, \mathcal{M}_k | \mathcal{Z}_{0:k}, U_{0:k-1}, X_0)$ of equation 2.15, is not straight forward. This section therefore outlines certain technical inconsistencies of traditional estimators, leading to summaries of principled approaches in Section 3.3 (for more details see [3], [27]).

3.2.1 Naive Estimators

The difficulty of applying standard estimators to RFSs arises because they represent information on the number of their elements (features) which is a dimensionless quantity, and the elements themselves, which can have dimensions (in the case of features – their location, in units of distance from a globally defined origin). To demonstrate some of the difficulties in deriving useful estimators for RFSs, consider the following example in which a PDF $p()$ on the RFS $\mathcal{M}$, representing an entire, unknown map, is assumed to be available. Intuitive, and standard, expected a posteriori (EAP) and maximum a posteriori (MAP) estimators are applied to a seemingly simple estimation problem [3].

Consider a simplistic situation in which there is at most one feature located in the map. Suppose that a corresponding feature existence filter [28] reports a 0.5 probability of the feature being present. Suppose also that, if the feature is present, the posterior density of the corresponding spatial state, $p(\mathcal{M})$, indicates that it is equally likely to be found anywhere in the one-dimensional interval $[0, 2]$, with the unit of distance given in meters. It should be noted here that, already at this simplistic level, vector approaches cannot jointly model this feature state. Under an RFS representation however, the map state $\mathcal{M}$ can be defined as Bernoulli RFS, with probability density,

$$p(\mathcal{M}) = \begin{cases} 0.5 & \mathcal{M} = \emptyset \\ 0.25 & \mathcal{M} = \{m\}, 0 \leq m \leq 2 \\ 0 & \text{otherwise} \end{cases} . \tag{3.1}$$

Note that the density, $p(\mathcal{M})$, is still a valid PDF, since its integral, with respect to $\mathcal{M}$, equates to unity. However, in this case, it is difficult to define an expected a posteriori (EAP) estimate since the addition of sets is not meaningfully defined. Instead, a naive maximum a posteriori (MAP) estimate could be constructed as,

$$\widehat{\mathcal{M}}^{MAP} \stackrel{?}{=} \arg\sup_{\mathcal{M}} p(\mathcal{M}) = \emptyset, \tag{3.2}$$

(where $\stackrel{?}{=}$ represents a question "Is it equal to?"), since $p(\emptyset) > p(\{m\})$ $(0.5 > 0.25)$. If the unit of distance is changed from meters to kilometres, the spatial probability density consequently becomes $p(m) = U(0, 0.002)$, and the probability density of the map state $\mathcal{M}$ is,

$$p(\mathcal{M}) = \begin{cases} 0.5 & \mathcal{M} = \emptyset \\ 250 & \mathcal{M} = \{m\}, 0 \leq m \leq 0.002 \\ 0 & \text{otherwise} \end{cases},$$

and the naive MAP estimate then becomes,

$$\widehat{\mathcal{M}}^{MAP} \stackrel{?}{=} \arg\sup_{\mathcal{M}} p(\mathcal{M}) = \{m\} \tag{3.3}$$

for any $0 \leq m \leq 0.002$ since $p(\{m\}) > p(\emptyset)$ $(250 > 0.5)$. This leads to the conclusion, that a target is now present, even though only the units of measurement have changed. This arises since the naive MAP yields a mathematical paradigm which compares a dimensionless quantity $p(\mathcal{M})$ (when $\mathcal{M} = \emptyset$) to a quantity $p(\mathcal{M})$ with dimensions (when $\mathcal{M} = m$). Such an MAP estimate is not well-defined since a change in the units of measurement results in a dramatic change in the estimate. In fact the MAP is only defined if the units of all possibilities are the same, such as in discretised state spaces, divided into cells.

This example has shown that standard estimators (EAP and MAP) are not well defined in the presence of non-unity target existence probability. It is therefore the aim of the next section to introduce new multi-target state estimators which are well behaved.

3.3 Bayes Optimal RFS Estimators

Several principled solutions to performing multi-object state estimation are now presented, in the form of two statistical estimators and a first order moment technique (the PHD filter) with desirable properties.

We begin by opening the discussion on the full SLAM problem in terms of joint Bayes optimal estimators for the vehicle trajectory and the map. The

Bayes risk is then defined for the map along with corresponding feature map estimators. Finally, Bayes optimal estimation approximations for FBRM and SLAM are derived.

This section discusses various Bayesian estimators for the SLAM problem and their optimality, based on a vector representation of the vehicles trajectory, and a finite-set representation of the map. The notion of Bayes optimal estimators is fundamental to the Bayesian estimation paradigm. In general, if the function $\hat{\theta} : z \mapsto \hat{\theta}(z)$ is an estimator of a parameter θ, based on a measurement z, and $C(\hat{\theta}(z), \theta)$ is the penalty for using $\hat{\theta}(z)$ to estimate θ, then the Bayes risk $R(\hat{\theta})$ is the expected penalty over all possible realisations of the measurement and parameter, i.e

$$R(\hat{\theta}) = \int\int C(\hat{\theta}(z), \theta) p(z, \theta) d\theta dz \tag{3.4}$$

where $p(z, \theta)$ is the joint probability density of the measurement z and the parameter θ. A Bayes optimal estimator is any estimator that minimises the Bayes risk.

In the SLAM context, relevant Bayes optimal estimators are those for the vehicle trajectory and the map. The posterior densities[1] $p_k(X_{1:k}) \triangleq p_k(X_{1:k}|\mathcal{Z}_{0:k}, U_{0:k-1}, X_0)$ and $p_k(\mathcal{M}_k) \triangleq p_k(\mathcal{M}_k|\mathcal{Z}_{0:k}, U_{0:k-1}, X_0)$ for the vehicle trajectory and map, can be obtained by marginalising the joint posterior density, $p_k(\mathcal{M}_k, X_{1:k}|\mathcal{Z}_{0:k}, U_{0:k-1}, X_0)$. For the vehicle trajectory, the posterior mean, which minimises the mean squared error (MSE), is a widely used Bayes optimal estimator. However, since the map is a finite set, the notion of MSE does not apply. Moreover, standard Bayes optimal estimators are defined for vectors and subsequently do not apply to finite-set feature maps. To the best of authors' knowledge, there is no work which establishes the Bayes optimality of estimators for finite-set feature maps (and consequently feature-based SLAM). Therefore the following sections propose frameworks for Bayes optimal estimation with RFSs, which assume varying degrees of approximation to the statistical representations of sets.

3.3.1 Bayes Risk in Feature Map Estimation

The convergence of the vehicle location estimation aspect, of feature-based frameworks, has received a great deal of attention to date [1]. However, to the authors' knowledge, the convergence of the corresponding map estimate, particularly with regards to converging to the true number of features, has never before been addressed or proven. Therefore, to address the optimal map estimation problem, as before, let $\mathcal{M}_k$ denote the feature-based map state at time k comprising $\mathfrak{m}_k$ features and $p_k(\mathcal{M}_k)$ denote its posterior density.

[1] Note that henceforth for compactness, $p_k(\cdot) = p_{k|k}(\cdot)$.

If $\widehat{\mathcal{M}}_k : \mathcal{Z}_{1:k} \mapsto \widehat{\mathcal{M}}_k(\mathcal{Z}_{1:k})$ is an estimator of the feature map $\mathcal{M}_k$, and $C(\widehat{\mathcal{M}}_k(\mathcal{Z}_{1:k}), \mathcal{M}_k)$ is the penalty for using $\widehat{\mathcal{M}}_k(\mathcal{Z}_{1:k})$ to estimate $\mathcal{M}_k$, then the Bayes risk for mapping is given by

$$R(\widehat{\mathcal{M}}_k) = \int \int C(\widehat{\mathcal{M}}_k(\mathcal{Z}_{1:k}), \mathcal{M}_k) p_k(\mathcal{M}_k, \mathcal{Z}_{1:k}) \delta \mathcal{M}_k \delta \mathcal{Z}_{1:k}.$$

where $p_k(\mathcal{M}_k, \mathcal{Z}_{1:k})$ is the joint density of the map and measurement history. Note that since the map and measurements are finite sets, standard integration for vectors is not appropriate for the definition of the Bayes risk. Subsequently the Bayes risk above is defined in terms of set integrals. Several principled solutions to performing feature map estimation are presented next, with the main focus of attention being on the PHD filter in Section 3.3.4, which is used widely throughout this book. The following estimators are Bayes optimal given the definition of an appropriate Bayes risk as just described.

3.3.2 Marginal Multi-Object Estimator

The Marginal Multi-Object (MaM) estimator is defined via a two-step estimation procedure. The number of features is first estimated using a maximum a posterior (MAP) estimator on the posterior cardinality distribution, ρ,

$$\hat{\mathfrak{m}}_k = \arg \sup_m \rho_k(|\mathcal{M}_k| = m). \tag{3.5}$$

Second, the individual feature states are estimated by searching over all maps with cardinality $\hat{\mathfrak{m}}_k$, using a MAP criteria,

$$\widehat{\mathcal{M}_k}^{MaM} = \arg \sup_{\mathcal{M}:|\mathcal{M}_k| = \hat{\mathfrak{m}}_k} p_k(\mathcal{M}). \tag{3.6}$$

It has been shown that the MaM estimator is Bayes optimal, however convergence results are not currently known.

3.3.3 Joint Multi-Object Estimator

In contrast to the MaM estimator, which first estimates the number of features and restricts its feature state estimation process to maps with only that number of features, the Joint Multi-Object (JoM) estimator executes its feature state estimation process on maps of all possible feature number.

The JoM estimator is defined as

$$\widehat{\mathcal{M}}_{k,s}^{JoM} = \arg\sup_{\mathcal{M}_k} \left(p_k(\mathcal{M}_k) \frac{s^{|\mathcal{M}_k|}}{|\mathcal{M}_k|!} \right), \tag{3.7}$$

where s is a constant with units of volume in the feature space, arg sup denotes the argument of the supremum, and $|\mathcal{M}_k|$ denotes the cardinality of $\mathcal{M}_k$. Notice that the fundamental difference between this estimator and the MAP estimator of equations 3.2 and 3.3 is that the factor $s^{|\mathcal{M}_k|}/|\mathcal{M}_k|!$ allows target based attributes of differing dimensions (e.g. spatial and non-spatial) to be "compared" in a principled manner.

First, to consider all possible sizes $\mathfrak{m}$ of the feature map for each $\mathfrak{m} \geq 0$, determine the MAP estimate,

$$\widehat{\mathcal{M}}^{(\mathfrak{m})} = \arg \sup_{\mathcal{M}:|\mathcal{M}|=\mathfrak{m}} p_k(\mathcal{M}|\mathcal{Z}_{0:k}). \tag{3.8}$$

Second, set

$$\widehat{\mathcal{M}}_s^{JoM} = \widehat{\mathcal{M}}^{(\hat{\mathfrak{m}})} \quad \text{where} \quad \hat{\mathfrak{m}} = \arg\sup_{\mathfrak{m}} p_k(\widehat{\mathcal{M}}^{(\mathfrak{m})}|\mathcal{Z}_{0:k}) \frac{s^{\mathfrak{m}}}{\mathfrak{m}!}. \tag{3.9}$$

It has been shown that the JoM estimator is Bayes optimal and is statistically consistent i.e. the feature map error distance (to be discussed in Section 4.3), between the optimal estimate and the true map, tends to zero as data accumulates [23], [3], [27]. Hence,

- "The JoM estimator determines the number $\hat{\mathfrak{m}}$ and the locations $\widehat{\mathcal{M}}$ of features optimally and simultaneously without resorting to optimal data association." [3].

Additionally, the value of s in equation 3.9 should be made equal to the desired accuracy for the state estimate. The smaller s is, the greater the accuracy of the estimate, but the rate of convergence of the estimator will be compromised. Because of this, while JoM is a theoretically attractive estimator, it is computationally expensive.

3.3.4 The Probability Hypothesis Density (PHD) Estimator

A simple approach to set-based estimation, is to exploit the physical intuition of the first moment of an RFS, known as its PHD or *intensity function*. This corresponds to the multi-feature equivalent of an expected value – the expectation of an RFS.

This section starts by giving an explanation of what the PHD is, and how it should be statistically interpreted in Section 3.3.4.1. This is followed by two intuitive derivations of the PHD in Sections 3.3.4.2 and 3.3.4.3.

3.3.4.1 Interpretation of the PHD

The intensity function at a point, gives the *density* of the expected number of features occurring at that point and therefore the mass (integral of the density over the volume of the space) of the PHD gives the expected number of features. The peaks of the intensity function indicate locations with relatively high concentration of expected number of features, in other words locations with high probability of feature existence. To provide an intuitive interpretation, consider a simple 1D example of two targets located at $x = 1$ and $x = 4$ each with spatial variance $\sigma^2 = 1$ taken from page 569, [3]. A corresponding Gaussian mixture representation of the PHD for this problem is:

$$\mathrm{PHD}(x) = \frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(x-1)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x-4)^2}{2\sigma^2}\right)\right]. \tag{3.10}$$

$\mathrm{PHD}(x)$ versus x is plotted in figure 3.1. Note that the maxima of $\mathrm{PHD}(x)$

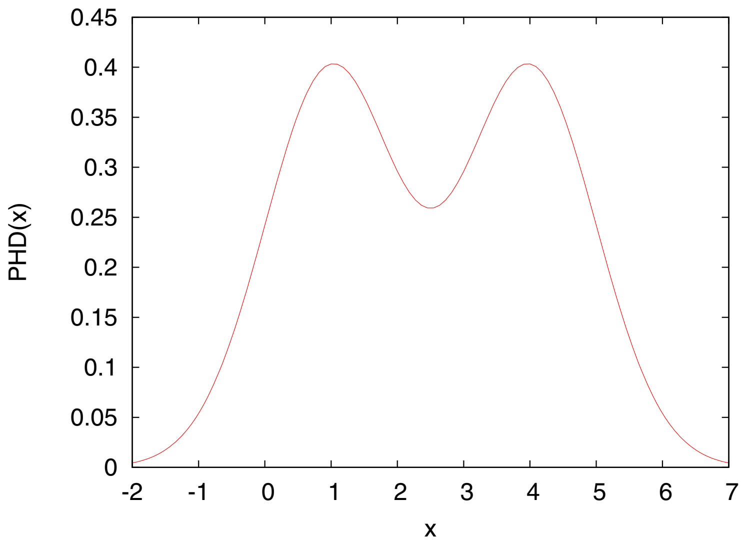

Fig. 3.1 A PHD for a 1D, 2 target problem of equation 3.10

occur near the target locations ($x = 1,\ 4$)). The integral of $\mathrm{PHD}(x)$ is $\mathfrak{m}$ where

$$\begin{aligned}\mathfrak{m} = \int \mathrm{PHD}(x)dx &= \int \mathcal{N}(1,\sigma^2)dx + \int \mathcal{N}(4,\sigma^2)dx \\ &= \quad 1 \quad + \quad 1 \quad = 2\end{aligned} \tag{3.11}$$

i.e. $\mathfrak{m}$ equals the actual number of targets. Here we note that a PHD is not a PDF, since its integral over the space of its variable is not, in general, unity.

For a 2D, robotic feature based map, graphical depictions of posterior PHDs after two consecutive measurements, approximated by Gaussian mixtures, are shown in figures 3.2 and 3.3. In each figure the intensity function

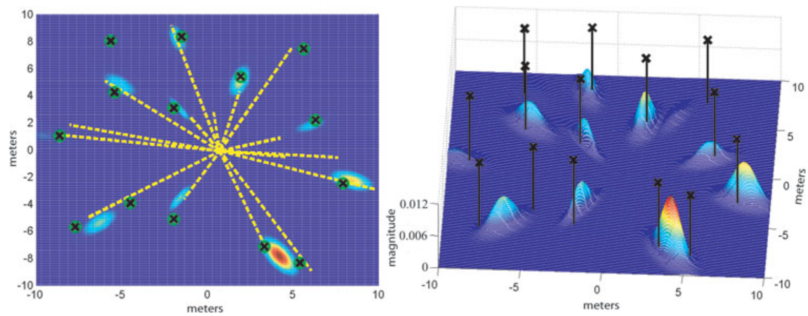

Fig. 3.2 A sample map PHD at time $k-1$, with the true map represented by black crosses. The measurement at $k-1$ is represented by the yellow dashed lines. The peaks of the PHD represent locations with the highest concentration of expected number of features. The local PHD mass in the region of most features is 1, indicating the presence of 1 feature. The local mass close to some unresolved features (for instance at (5,-8)) is closer to 2, demonstrating the unique ability of the PHD function to jointly capture the number of features.

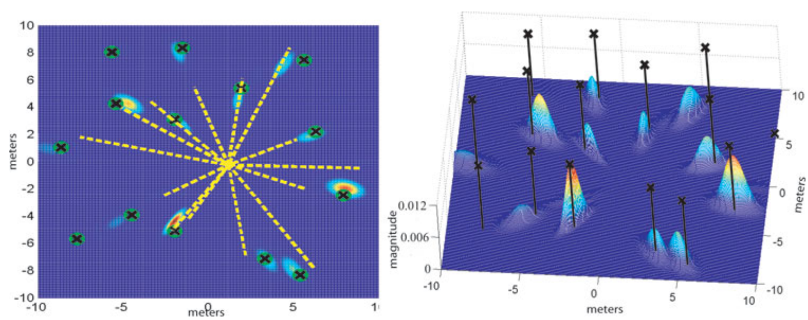

Fig. 3.3 A sample map PHD and measurement at time k. Note that the features at (5,-8) are resolved due to well separated measurements, while at (-3,-4), a lone false alarm close to the feature measurement contributes to the local PHD mass. At (-5,-4) a small likelihood over all measurements, coupled with a moderate $c_k(z|X_k)$ results in a reduced local mass.

is plotted as a function of the spatial coordinates. Since the integral of the intensity function (or PHD) is, by definition, the estimated number of features in the map, the mass (or integral) of each Gaussian can be interpreted as the number of features it represents. In the case of closely lying features (and large measurement noise), the PHD approach may not be able to resolve the features, as demonstrated for the right hand feature of Figure 3.2 at approximate coordinates (5, -8). However the PHD will represent the spatial density of L features by a singular Gaussian with a corresponding mass of L, which may improve the feature number estimate. This is only theoretically possible using the RFS framework. A graphical example for $L = 2$ is illustrated in Figure 3.2, which is then resolved through measurement updates into individual Gaussian components for each feature of mass $L \approx 1$, as shown in Figure 3.3 (the two peaks at approximate coordinates (5, -8)).

The PHD estimator has recently been proven to be Bayes optimal [29] and has been proven to be powerful and effective in multi-target tracking [3].

3.3.4.2 The PHD as the Limit of an Occupancy Probability

Intuitively, the PHD can be derived as a limiting case of the occupancy probability used in grid based methods. Following [30], consider a grid system and let m_i denote the centre and $B(m_i)$ the region defined by the boundaries of the ith grid cell. Let $P^{(occ)}(B(m_i))$ denote the occupancy probability and $\lambda(B(m_i))$ the area of the ith grid cell. Assume that the grid is sufficiently fine so that each grid cell contains at most one feature, then the expected number of features over the region $S_J = \bigcup\limits_{i \in J} B(m_i)$ is given by,

$$\begin{aligned}\mathbb{E}\left[|\mathcal{M} \cap S_J|\right] &= \sum_{i \in J} P^{(occ)}(B(m_i)) \\ &= \sum_{i \in J} v(m_i)\lambda(B(m_i)).\end{aligned} \tag{3.12}$$

where $v(m_i) = \frac{P^{(occ)}(B(m_i))}{\lambda(B(m_i))}$. Intuitively any region S_J can be represented by $\bigcup\limits_{i \in J} B(m_i)$, for some J. As the grid cell area tends to zero (or for an infinitesimally small cell), $B(m_i) \to dm$. The sum then becomes an integral and the expected number of features in S becomes,

$$\mathbb{E}\left[|\mathcal{M} \cap S|\right] = \int_S v(m)dm. \tag{3.13}$$

$v(m)$ defines the PHD and it can be interpreted as the occupancy probability density at the point m. The (coordinates of the) peaks of the intensity are points (in the space of features) with the highest local concentration of expected number of features and hence can be used to generate optimal estimates for the elements of $\mathcal{M}$. The integral of the PHD gives the expected number of features and the peaks of the PHD function can be used as estimates of the positions of the features.

3.3.4.3 The PHD as the Density of the Expectation of a Point Process

An alternative derivation of the PHD now follows. An analogous notion to the 'expectation' of an RFS can be borrowed from point process theory. This construct treats the random set as a random counting measure or a point process (a random finite set and a simple finite point process are equivalent [31]).

Let $p(\mathcal{M})$ be the multi-feature probability distribution of the map RFS $\mathcal{M}$. A somewhat naive interpretation of its expected value $\widehat{\mathcal{M}}$ would then be

$$\widehat{\mathcal{M}}^{\text{naive}} \triangleq \int \mathcal{M}_i p(\mathcal{M}) \delta \mathcal{M}_i. \tag{3.14}$$

where $\mathcal{M}_i$ represents the ith subset of $\mathcal{M}$. Since the addition of finite subsets of $\mathcal{M}$ is undefined, the above integral is also undefined. It can be solved by defining a transformation which maps finite subsets $\mathcal{M}_i$ into vectors M_i in some vector space. This transformation must maintain the set theoretic structure by transforming unions into sums - i.e. $\mathcal{M}_i \cup \mathcal{M}_j = M_i + M_j$, if $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$. In this case, an expected value can be defined in terms of the "equivalent" vectors

$$\widehat{\mathcal{M}} \triangleq \mathbb{E}[M] = \int M_i p(\mathcal{M}) dM_i \tag{3.15}$$

The point process literature [32] uses a transformation $M_i = \delta_{\mathcal{M}_i}$ where

$$\begin{aligned} \delta_{\mathcal{M}_i} &\triangleq 0 & \text{if } \mathcal{M}_i = \emptyset \\ \delta_{\mathcal{M}_i} &\triangleq \sum_{m \in \mathcal{M}_i} \delta(x - m) & \text{otherwise.} \end{aligned} \tag{3.16}$$

where x is the vector space of the features and $\delta(x - m)$ is the Dirac delta density concentrated at each feature m within the random finite subset $\mathcal{M}_i$. Taking the expectation of equation 3.16 gives

$$v(m) \triangleq \mathbb{E}[\delta_{\mathcal{M}}] = \int \delta_{\mathcal{M}_i} p(\mathcal{M}) \delta \mathcal{M}_i \tag{3.17}$$

which is the multi-feature equivalent of the expected value. This is called the probability hypothesis density (PHD), also known as the intensity density, or intensity function $v(m)$ of $\mathcal{M}$.

Note that while $v(m)$ is a density, it is *not* a PDF, since it may not necessarily integrate to 1. This is clear, as the integral of $v(m)$ over any region S gives the expected number of features in that region - i.e.

$$\int_S v(m) dm = \mathbb{E}\left[|\mathcal{M} \cap S|\right] \tag{3.18}$$

Hence, the integral of the non-negative intensity function $v(m)$ over $\mathcal{M}$ gives the expected number of features in the map.

Note that we have arrived at the same result as equation 3.13, in which the PHD was considered to be the limit of an occupancy probability.

3.3.4.4 Recovering the Map from the PHD Intensity Function

Since $v(m)$ is a density, it can be readily approximated by standard Sequential Monte Carlo (SMC) or Gaussian Mixture methods as described later in Chapter 4. The PHD filter recursion therefore propagates the intensity function $v(m)$ of the RFS state and uses the RFS measurement in the update stage. Since the intensity is the first order statistic of a random finite set, the PHD filter is analogous to the constant gain Kalman filter, which propagates the first order statistic (the mean) of the vector-based state. However, the *intensity*, $v(m)$, can be used to estimate both the number of features in the map, and their corresponding states (along with the uncertainty in the state estimates) [2].

If the RFS, $\mathcal{M}_k$, is Poisson, i.e. the number of points is Poisson distributed and the points themselves are independently and identically distributed (IID), then the probability density of $\mathcal{M}_k$ can be constructed exactly from the PHD.

$$p_k(\mathcal{M}_k) = \frac{\prod\limits_{m \in \mathcal{M}_k} v_k(m)}{\exp(\int v_k(m)dm)}. \tag{3.19}$$

where $v_k(m)$ is the map intensity function at time k and $\mathcal{M}_k$ is the RFS map which has passed through the field of view (FoV) of the vehicle's on board sensor(s) up to and including time k. In this sense, the PHD can be thought of as a 1st moment approximation of the probability density of an RFS.

Under these approximations, it has been shown [2] that, similar to standard recursive estimators, the PHD recursion has a *predictor-corrector* form.

3.4 The PHD Filter

As defined in Section 3.2.1, $\mathcal{M}$ is the RFS representing the entire unknown map. Let $\mathcal{M}_{k-1}$ be the RFS representing the explored map, with trajectory $X_{0:k-1} = [X_0, X_1, \ldots, X_{k-1}]$ at time $k-1$, i.e.

$$\mathcal{M}_{k-1} = \mathcal{M} \cap FoV(X_{0:k-1}). \tag{3.20}$$

Note that $FoV(X_{0:k-1}) = FoV(X_0) \cup FoV(X_1) \cup \cdots \cup FoV(X_{k-1})$. $\mathcal{M}_{k-1}$ therefore represents the set on the space of features which intersects with the union of individual FoVs, over the vehicle trajectory up to and including time $k-1$. Given this representation, $\mathcal{M}_{k-1}$ evolves in time according to,

$$\mathcal{M}_k = \mathcal{M}_{k-1} \cup \left(FoV(X_k) \cap \bar{\mathcal{M}}_{k-1} \right) \tag{3.21}$$

where $\bar{\mathcal{M}}_{k-1} = \mathcal{M} - \mathcal{M}_{k-1}$ represents the unexplored map (note the difference operator used here is the set difference, sometimes referred to as $\mathcal{M} \backslash \mathcal{M}_{k-1}$ or the relative complement of $\mathcal{M}_{k-1}$ in $\mathcal{M}$), i.e the set of features that are not in $\mathcal{M}_{k-1}$. Let the newly explored features which have entered the FoV be modelled by the independent RFS, $\mathcal{B}_k(X_k)$. In this case, the RFS map transition density is given by,

$$f_{\mathcal{M}}(\mathcal{M}_k|\mathcal{M}_{k-1}, X_k) = \sum_{\mathcal{W} \subseteq \mathcal{M}_k} f_{\mathcal{M}}(\mathcal{W}|\mathcal{M}_{k-1}) f_{\mathcal{B}}(\mathcal{M}_k - \mathcal{W}|X_k) \tag{3.22}$$

where $f_{\mathcal{M}}(\mathcal{W}|\mathcal{M}_{k-1})$ is the transition density of the set of features that are in $FoV(X_{0:k-1})$ at time $k-1$ to time k, and $f_{\mathcal{B}}(\mathcal{M}_k - \mathcal{W})|X_k)$ is the density of the RFS, $\mathcal{B}(X_k)$, of the new features that pass within the field of view at time k. To define the PHD filter in a form general enough to be applied to FBRM and SLAM, we now define a state variable Γ_k which corresponds to the state of interest to be estimated. In the case of FBRM, Γ_k would be replaced by "$m|X_k$" i.e. the feature at m, given the vehicle location X_k. This implementation of the PHD filter will be the subject of Chapter 4. In the case of SLAM a "brute force" approach is implemented in Chapter 5 which replaces Γ_k with each feature augmented with a hypothesised vehicle trajectory. Its implementation is shown to demonstrate a viable, and theoretically simple, SLAM implementation. A more elegant, Rao-Blackwellised implementation of the PHD filter is implemented in Chapter 6 which propagates N conditionally independent PHDs, based on each of the N hypothesised trajectories, represented as particles. In this case, Γ_k is effectively replaced by "$m|X_{0:k}$" i.e. the feature at m conditioned on the vehicle trajectory $X_{0:k}$. This will demonstrate a more computationally efficient SLAM implementation, which allows the Bayes optimal, expected trajectory and expected map to be evaluated.

In terms of the general state variable Γ_k the prediction of the map intensity function $v_{k|k-1}(\Gamma_k)$, is given by

$$v_{k|k-1}(\Gamma_k) = v_{k-1|k-1}(\Gamma_{k-1}) + b(\Gamma_k) \tag{3.23}$$

where $b(\Gamma_k)$ is the PHD of the new feature RFS, $\mathcal{B}(X_k)$. Note that $v_{k-1|k-1}()$ corresponds to the estimate of the intensity function at time $k-1$, given all observations up to, and including, time $k-1$. For ease of notation however, this will be referred to as $v_{k-1}()$ in future instances.

The PHD corrector equation is then [2],

$$v_k(\Gamma_k) = v_{k|k-1}(\Gamma_k)\Bigg[1 - P_D(\Gamma_k) + \sum_{z\in\mathcal{Z}_k} \frac{P_D(\Gamma_k)g_k(z|\Gamma_k)}{c_k(z) + \int P_D(\xi_k)g_k(z|\xi_k)v_{k|k-1}(\xi_k)d\xi_k}\Bigg] \quad (3.24)$$

where $v_{k|k-1}(\Gamma_k)$ is the predicted intensity function from equation 3.23, ξ is a subset of $\mathcal{M}_k$ and,

$P_D(\Gamma_k)$ = the probability of detecting a feature at m, from vehicle location X_k (encapsulated in Γ_k),
$g_k(z|\Gamma_k)$ = the measurement model of the sensor at time k,
$c_k(z)$ = intensity of the clutter RFS $\mathcal{C}_k(X_k)$ (in equation 2.10) at time k.

3.4.1 Intuitive Interpretation of the PHD Filter

An intuitive interpretation of the PHD filter equations 3.23 and 3.24 is given in Chapter 16 of [3]. The predictor equation 3.23 comprises the addition of the previous PHD correction and the PHD of the set of features hypothesised to enter the sensor's FoV. The corrector equation 3.24, can be more clearly interpreted in its integrated form since, by definition

$$\int v_k(\Gamma_k)d\Gamma_k = \mathfrak{m}_k \quad (3.25)$$

where $\mathfrak{m}_k$ is the number of estimated features at time k. To simplify the interpretation further, a constant (state independent) probability of detection is assumed in this section - i.e.

$$P_D(\Gamma_k) = P_D. \quad (3.26)$$

Therefore, from equation 3.24,

$$\begin{aligned}\mathfrak{m}_k &= \int v_k(\Gamma_k)d\Gamma_k \\ &= (1 - P_D)\mathfrak{m}_{k|k-1} + \\ &\quad P_D \sum_{z\in\mathcal{Z}_k} \frac{\int g_k(z|\Gamma_k)v_{k|k-1}(\Gamma_k)d\Gamma_k}{c_k(z) + P_D \int g_k(z|\xi_k)v_{k|k-1}(\xi_k)d\xi_k}\end{aligned} \quad (3.27)$$

Notice that the integrals in the numerator and denominator of the final term within the summation of equation 3.27 are identical and to simplify the equation we introduce

$$D_{k|k-1}(g_k, v_{k|k-1}) \triangleq \int g_k(z|\Gamma_k)v_{k|k-1}(\Gamma_k)d\Gamma_k = \int g_k(z|\xi_k)v_{k|k-1}(\xi_k)d\xi_k \tag{3.28}$$

where g_k abbreviates $g_k(z|\Gamma_k)$ and $v_{k|k-1}$ abbreviates $v_{k|k-1}(\Gamma_k)$. Therefore the integral of the PHD corrector equation 3.24, with constant P_D, can be written as the feature number corrector equation

$$\mathfrak{m}_k = (1 - P_D)\mathfrak{m}_{k|k-1} + P_D \sum_{z \in \mathcal{Z}_k} \frac{D_{k|k-1}(g_k, v_{k|k-1})}{c_k(z) + P_D D_{k|k-1}(g_k, v_{k|k-1})} \tag{3.29}$$

Equation 3.29 is useful for intuitively interpreting the PHD corrector equation, and is governed by the following effects:

1. *Probability of detection* P_D. If the map feature at m is not in the FoV of the sensor, it could not have been observed, thus $P_D = 0$. Therefore, from equation 3.29
$$\mathfrak{m}_k = (1 - 0)\mathfrak{m}_{k|k-1} + 0 = \mathfrak{m}_{k|k-1} \tag{3.30}$$
i.e. the updated number of features simply equals the predicted number, since no new information is available. Similarly from from equation 3.24,
$$v_k(\Gamma_k) = v_{k|k-1}(\Gamma_k)[1 - 0 + 0] = v_{k|k-1}(\Gamma_k) \tag{3.31}$$
i.e. the updated PHD will simply equal the predicted value.
On the other hand, if m is within the sensor FoV and if $P_D \approx 1$, the summation in equation 3.24, tends to dominate the PHD update and
$$v_k(\Gamma_k) \quad \approx \quad v_{k|k-1}(\Gamma_k)\left[\sum_{z \in \mathcal{Z}_k} \frac{g_k(z|\Gamma_k)}{c_k(z) + \int g_k(z|\xi_k)v_{k|k-1}(\xi_k)d\xi_k}\right] \tag{3.32}$$
Then the predicted PHD is modified by the sum of terms dependent on the measurement likelihood and clutter PHD.
2. *False alarms* $c_k(z)$. A particular feature observation could have originated from a feature or as a false alarm. Assume that the number of false alarms λ (represented by its intensity $c_k(z)$) is large and uniformly distributed in some region R. If the observed feature is in R, the term within the summation of equation 3.29 becomes
$$\begin{aligned} & P_D \frac{D_{k|k-1}(g_k, v_{k|k-1})}{c_k(z) + P_D D_{k|k-1}(g_k, v_{k|k-1})} \\ &= P_D \frac{D_{k|k-1}(g_k, v_{k|k-1})}{\frac{\lambda}{|R|} + P_D D_{k|k-1}(g_k, v_{k|k-1})} \approx P_D \frac{|R|}{\lambda} D_{k|k-1}(g_k, v_{k|k-1}) \approx 0 \end{aligned} \tag{3.33}$$
since λ is so large that it dominates the denominator. Therefore, if an observation originates from R, it is likely to be a false alarm, and it contributes almost nothing to the total posterior feature count, as it should.

On the other hand if a measurement originates from a region other than R, with low clutter, then it is unlikely to be a false alarm. This means that $c_k(z) \approx 0$ and

$$P_D \frac{D_{k|k-1}(g_k, v_{k|k-1})}{c_k(z) + P_D D_{k|k-1}(g_k, v_{k|k-1})} \approx P_D \frac{D_{k|k-1}(g_k, v_{k|k-1})}{0 + P_D D_{k|k-1}(g_k, v_{k|k-1})} = 1 \tag{3.34}$$

so that the measurement now contributes one feature to the total feature number. In general, if the number of false alarms, governed by the clutter PHD $c_k(z|X_k)$, is high, this increases the denominator of the summation, thus lowering the effect of the sensor update, as it should.

3. *Prior information* $P_D g_k(z|\Gamma_k)$. Assume that the sensor model is good, and $P_D g_k(z|\Gamma_k)$ is large for a particular state Γ_k which produces z. If z is consistent with prior information (the observation model), $P_D D_{k|k-1}(g_k, v_{k|k-1})$ will tend to dominate the denominator of the summation in equation 3.29, and the term corresponding to that feature in the summation will become

$$P_D \frac{D_{k|k-1}(g_k, v_{k|k-1})}{c_k(z) + P_D D_{k|k-1}(g_k, v_{k|k-1})} \approx 1 \tag{3.35}$$

Hence, a feature which is consistent with the observation model tends to contribute one feature to the total feature count.

Conversely if the observation z is inconsistent with prior information (is unlikely according to the sensor model), then the product $P_D g_k(z|\Gamma_k)$ will be small, and its corresponding term in the summation in equation 3.29 will tend to be ignored.

Equations 3.23 and 3.24 which comprise the PHD filter have been shown to be Bayes optimal, assuming that the RFS observation and map statistics can be represented by their first moments only [3].

3.5 Summary

This chapter addressed the issues of estimation with RFSs. Initially, the traditional MAP and EAP estimators were applied to a simple, single feature problem with both feature existence and spatial uncertainty. It was demonstrated that such estimators are not suitable in such applications, and new multi-feature estimators were defined, which minimised the Bayes risk in feature map estimation.

The main focus of attention of the chapter was on the PHD filter. An RFS map density can be represented by its first moment, the intensity function. Brief derivations for the PHD estimator (intensity function) were shown based on the PHD as the limit of an occupancy probability, and the density of the expectation of a point process.

The PHD recursion is far more numerically tractable than propagating the RFS map densities of equation 2.15. In addition, the recursion can be readily extended to incorporate multiple sensors/vehicles by sequentially updating the map PHD with the measurement from each robot.