Tomograhic Multitarget Tracking and Its Relationship to PHD Filters II

R.L. Streit
Metron, Inc
1818 Freedom Drive, Suite 800
Reston, VA 20190
USA

Chapter 6
Multiple Target Tracking

"Although this may seem a paradox, all exact science is dominated by the idea of approximation."
Bertrand Russell, The Scientific Outlook, *1931*

Abstract Multitarget tracking intensity filters are closely related to imaging problems, especially PET imaging. The intensity filter is obtained by three different methods. One is a Bayesian derivation involving target prediction and information updating. The second approach is a simple, compelling, and insightful intuitive argument. The third is a straightforward application of the Shepp-Vardi algorithm. The intensity filter is developed on an augmented target state space. The PHD filter is obtained from the intensity filter by substituting assumed known target birth and measurement clutter intensities for the intensity filter's predicted target birth and clutter intensities.

To accommodate heterogeneous targets and sensor measurement models, a parameterized intensity filter is developed using a marked PPP Gaussian sum model. Particle and Gaussian sum implementations of intensity filters are reviewed. Mean-shift algorithms are discussed as a way to extract target state estimates. Grenander's method of sieves is discussed for regularization of the multitarget intensity filter estimates. Sources of error in the estimated target count are discussed. Finally, the multisensor intensity filter is developed using the same PPP target models as in the single sensor filter. Both homogeneous and heterogeneous multisensor fields are discussed. Multisensor intensity filters reduce the variance of estimated target count by averaging.

Multitarget tracking in clutter is a joint detection and estimation problem. It comprises two important inter-related tasks. One initiates and terminates targets, and the other associates, or assigns, data to specific targets and to clutter. The MHT (Multiple Hypothesis Tracking) formulation treats both tasks. It is well matched to the target physics and to the sensor signal processors of most radar and sonar systems. Unfortunately, exact MHT algorithms are intractable because the number of measurement assignment hypotheses grows exponentially with the number of measurements. These problems are

aggravated when multiple sensors are used. Circumventing the computational difficulties of MHT requires approximation.

Approximate tracking methods based on PPP models are the topics of this chapter. They show much promise in difficult problems with high target and clutter densities. The key insight is to model the distribution of targets in state space as a PPP, and then use a filter to update the defining parameter of the PPP — its intensity. To update the intensity is to update the PPP. The intensity function of the PPP approximation characterizes the multiple target tracking model. This important point is discussed further in Section 6.1.1.

The PPP intensity model uses an augmented state space, $\mathcal{S}^+$. This enables it to estimate target birth and measurement clutter processes on-line as part of the filtering algorithm.

Three approaches to the intensity filter are provided. The first and most rigorous is a Bayesian derivation given in Appendix D. The relationship between this approach and the "first moment intensity" of the posterior point process is shown. The second approach is a short but extraordinarily insightful derivation that is ideal for readers who wish to avoid the Bayesian analysis, at least on a first reading. The third approach is based on the connection to PET and the Shepp-Vardi algorithm. The PET interpretation contributes significantly to understanding the PPP target model. A special case of the intensity filter is the well known PHD filter. It is obtained by assuming a known target birth-death process, a known measurement clutter process, and restricting the intensity filter to the non-augmented target state space $\mathcal{S}$.

Implementation issues are discussed in Section 6.3. Current approaches use either particle or Gaussian sum methods. An image processing method called the mean shift algorithm is well suited to point target estimation, especially for particle methods. Observed information matrices (*cf.* Section 4.7) are proposed as surrogates for the error covariance matrices widely used in single target Bayesian filters. The underlying statistical meaning of OIM estimates is as yet unresolved.

Several sources of error in the estimated target count are discussed in Section 6.4. Target count estimates are sensitive to the probability of target detection. This function, $P_k^D(x)$, depends on target state x at measurement time t_k. It varies over time because of slowly changing sensor characteristics and environmental conditions. Monitoring these and other factors that affect target detection probability is not typically considered part of the tracking problem.

Several areas of on-going research are also discussed. One is a Gaussian sum PPP filter that enables heterogeneous target motion and sensor measurement models to be used in an intensity filter setting. See Section 6.2.2 for details. Another is the multiple sensor intensity filter described in Section 6.5. This filter is relies on the validity of the PPP approximation for every sensor. It reduces the variance of the target count by averaging over the sensors,

so that the variance of estimated target count decreases with the number of sensors.

6.1 Intensity Filters

6.1.1 PPP Model Interpretation

The points of a realization of a PPP on the target state space are a poor representation of the physical reality of a multiple target state. This is especially easy to see when exactly one target is present, for then ideally

$$\int_{\mathcal{S}} \lambda(x)\,\mathrm{d}x = 1\,. \tag{6.1}$$

From (2.4), the probability that a realization of the PPP has *exactly* one point target is

$$p_N(n=1) = e^{-1} \approx 37\%\,.$$

Hence, 63% of all realizations have either no target or two or more targets. Evidently, realizations of the PPP seriously mismodel this simple tracking problem. The problem worsens with increasing the target count: if exactly n targets are present, then the probability that a realization has exactly n points is $e^{-n}\,n^n\,/\,n! \approx (2\pi n)^{-1/2} \to 0$ as $n \to \infty$. Evidently, PPP realizations are poor models of real targets.

One interpretation of the PPP approximation is that the critical element of the multitarget model is the intensity function, *not* the PPP realizations. The shift of perspective means that the integral (6.1) is the more physically meaningful quantity. Said another way, the concept of expectation, or ensemble average over realizations, corresponds more closely to the physical target reality than do the realizations themselves.

A huge benefit comes from accepting the PPP approximation to the multiple target state — exponential numbers of assignments are completely eliminated. The PPP approximation finesses the data assignment problem by replacing it with a stochastic imaging problem, and the imaging problem is easier to solve. It is fortuitous that the imaging problem is mathematically the same problem that arises in PET; see Section 5.2. The "at most one measurement per target" rule for tracking corresponds in PET to the physics — there is at most measurement per one positron-electron annihilation.

Analogies are sometimes misleading, but consider this one: In the language of thermodynamics, the points of PPP realizations are microstates. Microstates obey the laws of physics, but are not directly observable without disturbing the system state. Physically meaningful quantities (such as temperature, etc.) are ensemble averages over the microstates. In the PPP target

model, the points of a realization are thus "microtargets", and microtargets obey the same target motion and measurement models as real targets. The ensemble average over the PPP microtargets yields the target intensity function. The language of microtargets is helpful in Section 6.5 on multisensor intensity filtering, but it is otherwise eschewed in this chapter.

6.1.2 Predicted Target and Measurement Processes

Formulation

Standard filtering notation is adopted, but modified to accommodate PPPs. The general Bayesian filtering problem is reviewed in Appendix C. Let $\mathcal{S} = \mathbb{R}^{n_x}$ denote the n_x-dimensional single target state space. The augmented space is $\mathcal{S}^+ \equiv \mathcal{S} \cup \phi$, where ϕ represents a "clutter target" state. Clutter targets account for data not assigned to real, physical targets. The augmented space $\mathcal{S}^+$ is discussed in Section 2.12.

The single target transition function from time t_{k-1} to time t_k, denoted by $\Psi_{k-1}(y|x) \equiv p_{\Xi_k|\Xi_{k-1}}(y|x)$, is assumed known for all $x, y \in \mathcal{S}^+$. The augmented state space enables both target initiation and termination to be incorporated directly into Ψ_{k-1} as specialized kinds of state transitions. Novel aspects of the transition function are:

- $\Psi_{k-1}(\phi|x)$ is the probability that a target at $x \in \mathcal{S}$ terminates;
- $\Psi_{k-1}(y|\phi)$ is the likelihood that a new target initiates at $y \in \mathcal{S}$; and
- $\Psi_{k-1}(\phi|\phi)$ is the probability that a target in ϕ remains in ϕ.

The augmented state ϕ is also used to account for measurement clutter, that is, for data that do not correspond to real targets.

The multitarget state at time t_k is Ξ_k. It is a point process on $\mathcal{S}^+$, but it is not a PPP on $\mathcal{S}^+$. Nonetheless, Ξ_k is approximated by a PPP to "close the loop" after the Bayesian information update. The multitarget state is a realization ξ_k of Ξ_k. If $\xi_k = (n, \{x_1, \ldots, x_n\})$, then every point x_j is either a point in $\mathcal{S}$ or is ϕ. It is stressed that repeated occurrences of ϕ are allowed in the list $\{x_1, \ldots, x_n\}$ to account for clutter.

The measurement at time t_k is Υ_k. It is a point process on the (nonaugmented) space $\mathcal{T} \equiv \mathbb{R}^{n_z}$, where n_z is the dimension of a sensor measurement. The measurement data set

$$\upsilon_k = (m, \{z_1, \ldots, z_m\}) \in \mathcal{E}(\mathcal{T})$$

is a realization of Υ_k. The pdf of a point measurement $z \in \mathcal{T}$ conditioned on a target in state $x \in \mathcal{S}^+$ at time t_k is the measurement pdf $p_k(z|x)$. The only novel aspect of this pdf is that $p_k(z|\phi)$ is the pdf that z is a clutter measurement.

The Bayesian posterior multitarget state point process conditioned on the data $\upsilon_1, \ldots, \upsilon_{k-1}$ is approximated by a PPP. Denote this PPP by $\Xi_{k-1|k-1}$. The

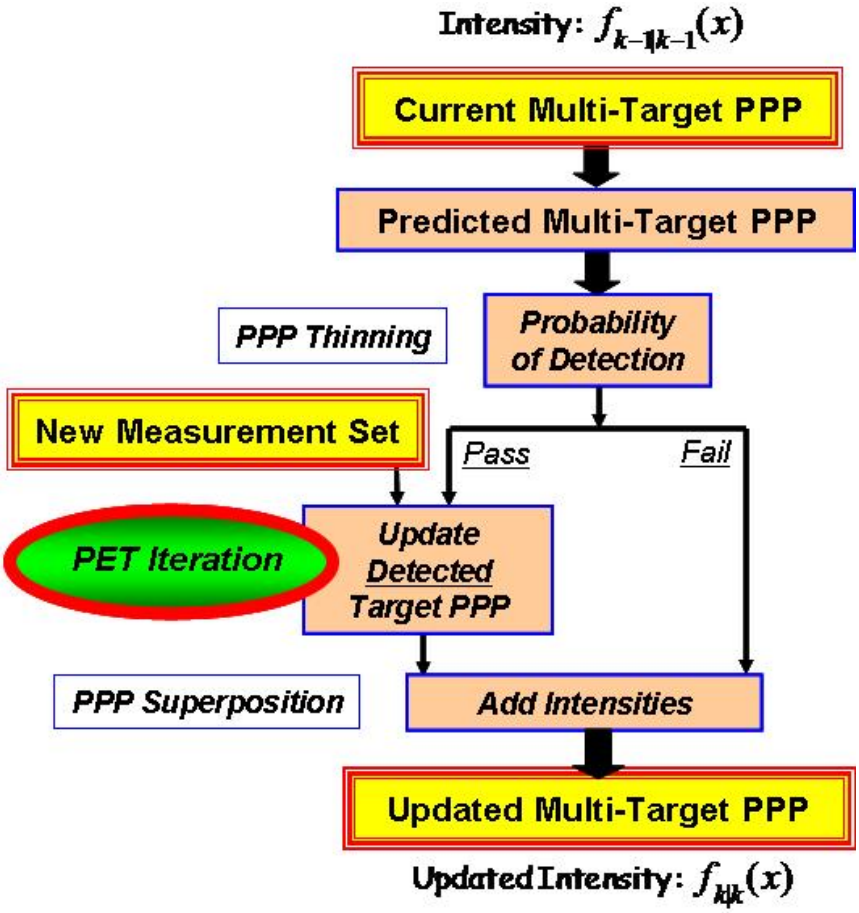

Fig. 6.1 Block diagram of the Bayes update of the intensity filter on the augmented target state space $\mathcal{S}^+$. One step of the Shepp-Vardi algorithm is used in the "PET Iteration" block. Because the null state ϕ is part of the state space, target birth and measurement clutter estimates are intrinsic to the predicted target and predicted measurement steps. The same block diagram holds for the PHD filter on the nonaugmented space $\mathcal{S}$.

intensity of $\Xi_{k-1|k-1}$ is $f_{k-1|k-1}(x)$, $x \in \mathcal{S}^+$. Let $\Xi_{k|k-1}$ denote the predicted PPP at time t_k. Its intensity is denoted by $f_{k|k-1}(x)$, and it is the integral of the intensity $f_{k-1|k-1}(x)$ of $\Xi_{k-1|k-1}$, as seen in Section 2.11.2, Eqn. (2.86).

The goal is to update the predicted PPP, $\Xi_{k|k-1}$, with the measurement data υ_k. The information updated point process is not a PPP, so it is approximated by a PPP. Let $\Xi_{k|k}$ denote the approximating PPP, and let its intensity be $f_{k|k}(x)$.

Fig. 6.1 outlines the steps of the intensity filter. The discussion below walks through the steps in the order outlined.

Target Motion and the Bernoulli Split

The first of these steps accounts for target motion and predicts the intensity at the next time step. The input is a PPP with intensity $f_{k-1|k-1}(x)$, so the transition procedure yields an output process that is also a PPP, as is seen in Section 2.11.1. Let $f_{k|k-1}(x)$ denote the intensity of the output PPP. Adapting (2.83) to $\mathcal{S}^+$ gives

$$f_{k|k-1}(x) = \int_{\mathcal{S}^+} \Psi(x|y)\, f_{k-1|k-1}(y)\, \mathrm{d}y, \tag{6.2}$$

where the integral over $\mathcal{S}^+$ is defined as in (2.97).

Target motion is followed by a Bernoulli thinning procedure using the probability of the sensor detecting a target at the point x, denoted $P_k^D(x)$. This probability is state dependent and assumed known. The input PPP intensity is $f_{k|k-1}(x)$. As seen in Section 2.9.2, thinning splits it into two PPPs – one for detected targets and the other for undetected targets, denoted by $f_{k|k-1}^D(x)$ and $f_{k|k-1}^U(x)$, respectively. These PPPs are independent (see Section 2.9.2) and, from (2.56), and their intensities are

$$f_{k|k-1}^D(x) = P_k^D(x)\, f_{k|k-1}(x)$$

and

$$f_{k|k-1}^U(x) = \left(1 - P_k^D(x)\right) f_{k|k-1}(x).$$

Both branches in Fig. 6.1 are now subjected to an information update.

Predicted Measurement PPP, and Why It Is Important

As seen in Section 2.11.2, the predicted measurement process is a PPP. Its intensity is

$$\lambda_{k|k-1}(z) = \int_{\mathcal{S}^+} p_k(z|x)\, P_k^D(x)\, f_{k|k-1}(x)\, \mathrm{d}x, \quad \text{for } z \in \mathcal{T}, \tag{6.3}$$

as is seen from (2.86). The measurement PPP is a critical component of the intensity filter because, as is seen in Eqn. (6.9), it weights the individual terms in the sum that comprises the filter.

Another way to see the importance of $\lambda_{k|k-1}(z)$ is to recall the classical single-target Bayesian tracking problem. The standard Bayesian formulation gives

$$p(x|z) = \frac{p(z|x)\,p(x)}{p(z)}, \tag{6.4}$$

where the denominator is a scale factor that makes the left hand side a true pdf. It is very easy to ignore $p(z)$ in practice because the numerator is obtained by multiplication and the product scaled so that it is a pdf. When multiple conditionally independent measurements are available, the conditional likelihood is a product and it is the same story again for the scale factor. However, *if the posterior pdfs are summed, not multiplied*, the scale factor must be included for the individual terms to be comparable. Such is the case with the intensity filter: the PPP model justifies adding Bayesian posteriors instead of multiplying them, and the scale factors are crucial to making the sum meaningful.

The scale factor clearly deserves a respectable name, and it has one. It is called the partition function in statistical physics and the machine learning communities.

6.1.3 Information Updates

It is seen in Appendix C that the mathematically correct information update procedure is to apply the Bayesian method to both the detected and undetected target PPPs to evaluate their posterior pdfs. These pdfs are defined on the event space $\mathscr{E}(\mathcal{S}^+)$. If the posterior pdfs have the proper form, then the posterior point processes are PPPs and are characterized by their intensity functions on $\mathcal{S}^+$.

The information update of the undetected target PPP is the Bayesian updated process condition on no target detection. The posterior point process is identical to the predicted target point process. It is therefore a PPP whose intensity, denoted by $f^U_{k|k}(x)$, is given by

$$\begin{aligned} f^U_{k|k}(x) &= f^U_{k|k-1}(x) \\ &= \left(1 - P^D_k(x)\right) f_{k|k-1}(x). \end{aligned} \tag{6.5}$$

This brings the right hand branch of Fig. 6.1 to the superposition stage, i.e., the block that says "Add Intensities".

The left hand branch is more difficult because, as it turns out, the information updated detected target point process is not a PPP. This is a serious dilemma since it is highly desirable both theoretically and computationally for the filter recursion to remain a closed loop. The posterior point process of the detected targets is therefore approximated by a PPP. Three methods are given for obtaining this approximation.

The first method is a Bayesian derivation of the posterior density of the point process on the event space $\mathscr{E}(\mathcal{S}^+)$, followed by a "mean field" ap-

proximation. Details about Bayesian filters in general are given in Appendix C. The Bayesian derivation is mathematically rigorous, but not particularly insightful.

To gain intuition, one need look no further than to the second method. While not rigorous, it is intuitively appealing and very convincing. The perspective is further enriched by the third method. It shows a direct connection between the information update of the Bayesian filter and the Shepp-Vardi algorithm for PET imaging. This third method also poses an interesting question about iterative updates of a Bayesian posterior density.

First Method: Bayesian Derivation and Mean Field Approximation

The pdf of the Bayesian posterior point process for detected targets is defined on the event space $\mathscr{E}(\mathcal{S}^+)$. The Bayesian posterior, or information updated, pdf is defined on this complex event space. The derivation is straightforward and a delight for Bayesians. It is relatively long and interferes with the flow of the discussion, so it is given in Appendix D where it can be read at leisure. The main points are outlined here, so readers suffer little or no loss of insight by skipping it on a first reading. Specific equation references are provided here so the precision of the mathematics is not lost in the flow of words.

The Bayesian derivation explicitly incorporates the "at most one measurement per target"rule into the measurement likelihood function. It imposes this constraint via the measurement conditional pdf (*cf.* Eqn. (D.5)). This pdf sums over all possible assignments of the given data to targets. Because of the augmented space $\mathcal{S}^+$, a clutter measurement is accounted for by assigning it to a target with state ϕ.

The usual Bayesian update (*cf.* Eqn. (D.6)) leads to the pdf of the Bayesian posterior point process (*cf.* Eqn. (D.10)). This pdf uses the facts that the *a priori* target process is a PPP with intensity $f^D_{k|k-1}(x)$ and that the predicted measurement process is a PPP with intensity $\lambda_{k|k-1}(z)$ given by (6.3). The posterior pdf is computationally intractable except in reasonably small problems. In any event, inspection of the posterior pdf clearly reveals that it does not have the form of a PPP pdf. Approximating the posterior point process with a PPP is the next step.

The Bayesian posterior pdf is replaced by a mean field approximation, a widely used method of approximation in machine learning and statistical physics problems. This approximation is the product of the one dimensional marginal pdfs of the posterior pdf. The marginal pdfs are identical. The intensity function of the approximating PPP is therefore taken to be proportional to the marginal pdf (*cf.* Eqn. (D.13)). The appropriate scale factor is a constant determined by maximum likelihood. In essence, the mean field approximation (*cf.* Eqn. (D.14)) is proportional to the intensity function of the approximating PPP.

The posterior detected target point process is a special case of the more general class of finite point processes. The general theory of these processes dates to the 1950s. An excellent text on the topic is [16, Chap. 5]. Finite point process theory leads via the so-called Janossy densities to the first moment intensity function approximation. Janossy densities are merely the joint pdfs of the points of the process (conditioned on their number). The first moment is a sum over all Janossy densities. In the tracking application, only one Janossy density is nonzero, and it is the likelihood function of updated point process. Consequently, and happily, the first moment is given by a sum comprising only one term. The mean field approximation is very closely related to the first moment intensity function. Further details are provided in the last section of Appendix D.

Second Method: Expected Target Count

Let $p_k(\cdot|x)$, $x \in S^+$, denote the conditional pdf of a measurement in the measurement space $\mathcal{T}$, so that

$$\int_{\mathcal{T}} p_k(z|x)\,\mathrm{d}z = 1\,, \quad \text{for all } x \in S^+\,. \tag{6.6}$$

The special case $p_k(z|\phi)$ is the pdf of a data point z conditioned on the target state ϕ, that is, the pdf of z given that it is clutter. The predicted intensity at time t_k of the target point process is a PPP with intensity $P_k^D(x)\,f_{k|k-1}(x)$. The intensity $f_{k|k}^D(x)$ is the intensity of a PPP that approximates the information updated, or Bayes posterior, detected target process.

The measured data at time t_k are m_k points in a measurement space $\mathcal{T}$. Denote these data points by $Z_k = (z_1, \ldots, z_{m_k})$.

The information update of the detected target PPP is obtained intuitively as follows. The best current estimate of the probability that the point measurement z_j originated from a physical target with state $x \in \mathcal{S}$ in the infinitesimal $\mathrm{d}x$ is

$$\frac{p_k(z_j|x)\,P_k^D(x)\,f_{k|k-1}(x)\,\mathrm{d}x}{\lambda_{k|k-1}(z_j)}\,, \tag{6.7}$$

where the denominator is found using (6.3). Similarly, the probability that z_j originated from a target with state ϕ is

$$\frac{p_k(z_j|\phi)\,P_k^D(\phi)\,f_{k|k-1}(\phi)}{\lambda_{k|k-1}(z_j)}\,. \tag{6.8}$$

Because of the "at most one measurement per target" rule, the sum of the ratios over all measurements z_j is the estimated number of targets at x, or targets in ϕ, that generated a measurement.

The estimated number of targets at $x \in \mathcal{S}$ is set equal to the expected number of targets conditioned on the data v_k, namely $f^D_{k|k}(x)\,\mathrm{d}x$. Cancelling $\mathrm{d}x$ gives

$$f^D_{k|k}(x) = \sum_{j=1}^{m} \frac{p_k(z_j|x)\, P^D_k(x)\, f_{k|k-1}(x)}{\lambda_{k|k-1}(z_j)}\,. \tag{6.9}$$

Eqn. (6.9) holds for all $x \in \mathcal{S}^+$, not just for $x \in \mathcal{S}$.

The expected target count method makes it clear that the expected number of detected targets in any given set $\mathcal{R} \subset \mathcal{S}$ is simply the integral of the posterior pdf

$$E[\text{Number of detected targets in } \mathcal{R}] = \int_{\mathcal{R}} f^D_{k|k}(x)\,\mathrm{d}x\,. \tag{6.10}$$

Similarly, the expected number of targets in state ϕ is the posterior intensity evaluated at ϕ, namely, $f^D_{k|k}(\phi)$. The predicted measurement process with intensity $\lambda_{k|k-1}(z)$ is a vital part of the intensity filter.

Third Method: Shepp-Vardi Iteration

The PET model is interesting here. The measurement data and multiple target models are interpreted analogously so that:

- The target state space $\mathcal{S}^+$ corresponds to the space in which the radioisotope is absorbed.
- The measured data $Z_k \subset \mathcal{T}$ correspond to the measured locations of the annihilation events. As noted in the derivation of Shepp-Vardi, the measurement space $\mathcal{T}$ need not be the same as the state space.
- The posterior target intensity $f^D_{k|k}(x)$ corresponds to the annihilation event intensity.

The analogy makes the targets mathematically equivalent to the distribution of (hypothetical) positron-electron annihilation events in the state space.

Under the annihilation event interpretation, the information update (6.9) of the detected target process is given by the Shepp-Vardi algorithm for PET using PPP sample data. The EM derivation needs only small modifications to accommodate the augmented state space. Details are left to the reader. The n-th iteration of the Shepp-Vardi algorithm is, from (5.18),

$$f^D_{k|k}(x)^{(n+1)} = f^D_{k|k}(x)^{(n)} \sum_{j=1}^{m_k} \frac{p_k(z_j|x)}{\int_{\mathcal{S}^+} p_k(z_j|s)\, f^D_{k|k}(s)^{(n)}\, \mathrm{d}s}, \tag{6.11}$$

where the predicted intensity $f^D_{k|k}(x)^{(0)} \equiv f^D_{k|k-1}(x) = P^D_k(x)\, f_{k|k-1}(x)$ initializes the algorithm. The first iteration of this version of the Shepp-Vardi algorithm is clearly identical to the Bayesian information update (6.9) of the detected target process. The second and higher iterations are *not* Bayesian intensity estimates.

The Shepp-Vardi iteration converges to an ML estimate of the target state intensity given only data at time t_k. It is independent of the data at times $t_1, \ldots, t_{k-1}$ except insofar as the initialization influences the ML estimate. In other words, the iteration leads to an ML estimate of an intensity that does *not* include the effect of a Bayesian prior. The problem lies not in the PET interpretation but in the pdf of the data. To see this it suffices to observe that the parameters of the pdf (5.7) are not constrained by a Bayesian prior and, consequently, the Shepp-Vardi algorithm converges to an estimate that is similarly unconstrained. It is, moreover, not obvious how to impose a Bayesian prior on the PET parameters that does not disappear in the small cell limit.

6.1.4 The Final Filter

Superposing the PPP approximation of the detected target process and the undetected target PPP gives

$$\begin{aligned} f_{k|k}(x) &= f^U_{k|k}(x) + f^D_{k|k}(x), \qquad x \in \mathcal{S}^+ \\ &= \left[1 - P^D_k(x) + \sum_{j=1}^{m} \frac{p_k(z_j|x)\, P^D_k(x)}{\lambda_{k|k-1}(z_j)}\right] f_{k|k-1}(x) \end{aligned} \tag{6.12}$$

as the updated intensity of the PPP approximation to $\Xi_{k|k}$.

The intensity filter comprises equations (6.2), (6.3), and (6.12). The first two equations are more insightful when written in traditional notation. Expanding the discrete-continuous integral (6.2) gives

$$f_{k|k-1}(x) = \hat{b}_k(x) + \int_{\mathcal{S}} \Psi_{k-1}(x|y)\, f_{k-1|k-1}(y)\, \mathrm{d}y, \tag{6.13}$$

where the predicted target birth intensity is

$$\hat{b}_k(x) = \Psi_{k-1}(x|\phi)\, f_{k-1|k-1}(\phi). \tag{6.14}$$

Also, from (6.3),

Table 6.1 Intensity Filter on the State Space $\mathcal{S}^+ = S \cup \phi$

INPUTS:
Data: $z_{1:m} \equiv \{z_1, \ldots, z_m\} \subset \mathcal{T}$ at time t_k
Probability of target detection: $P_k^D(x)$ at time t_k

OUTPUT: $\{f_{k|k}(x),\ f_{k|k}(\phi)\} = \text{IntensityFilter}\left[f_{k-1|k-1}(x),\ f_{k-1|k-1}(\phi),\ P_k^D(x),\ z_{1:m}\right]$

- Predicted target intensity: For $x \in \mathcal{S}$,
  - Newly born targets (target initiations): $\hat{b}_k(x) = \Psi_{k-1}(x|\phi)\, f_{k-1|k-1}(\phi)$
  - Moving targets: $\hat{d}_k(x) = \int_{\mathcal{S}} \Psi_{k-1}(x|y)\, f_{k-1|k-1}(y)\, \mathrm{d}y$
  - Target intensity: $f_{k|k-1}(x) = \hat{b}_k(x) + \hat{d}_k(x)$
- Predicted clutter intensity:
  - Clutter persistence: $\hat{b}_k(\phi) = \Psi_{k-1}(\phi|\phi)\, f_{k-1|k-1}(\phi)$
  - Newly born clutter (target terminations): $\hat{d}_k(\phi) = \int_{\mathcal{S}} \Psi_{k-1}(\phi|y)\, f_{k-1|k-1}(y)\, \mathrm{d}y$
  - Clutter intensity: $f_{k|k-1}(\phi) = \hat{b}_k(\phi) + \hat{d}_k(\phi)$
- Predicted measurement intensity: FOR $j = 1 : m$,
  - Generated by targets: $\hat{\nu}_k(z_j) = \int_{\mathcal{S}} p_k(z_j|x)\, P_k^D(x)\, f_{k|k-1}(x)\, \mathrm{d}x$
  - Generated by clutter: $\hat{\lambda}_k(z_j) = p_k(z_j|\phi)\, P_k^D(\phi)\, f_{k|k-1}(\phi)$
  - Measurement intensity: $\lambda_{k|k-1}(z_j) = \hat{\lambda}_k(z_j) + \hat{\nu}_k(z_j)$
- Information updated target intensity: For $x \in \mathcal{S}$,
  - Undetected targets: $f_{k|k}^U(x) = \left(1 - P_k^D(x)\right) f_{k|k-1}(x)$
  - Detected targets: $f_{k|k}^D(x) = \left[\sum_{j=1}^m \frac{p_k(z_j|x)\, P_k^D(x)}{\lambda_{k|k-1}(z_j)}\right] f_{k|k-1}(x)$
  - Target intensity: $f_{k|k}(x) = f_{k|k}^U(x) + f_{k|k}^D(x)$
- Information updated clutter intensity:
  - Undetected clutter targets (generate no data): $f_{k|k}^U(\phi) = \left(1 - P_k^D(\phi)\right) f_{k|k-1}(\phi)$
  - Detected clutter targets (generate data): $f_{k|k}^D(\phi) = \left[\sum_{j=1}^m \frac{p_k(z_j|\phi)\, P_k^D(\phi)}{\lambda_{k|k-1}(z_j)}\right] f_{k|k-1}(\phi)$
  - Clutter target intensity: $f_{k|k}(\phi) = f_{k|k}^U(\phi) + f_{k|k}^D(\phi)$

$$\lambda_{k|k-1}(z) = \hat{\lambda}_k(z) + \int_{\mathcal{S}} p_k(z|x)\, P_k^D(x)\, f_{k|k-1}(x)\, \mathrm{d}x, \tag{6.15}$$

where

$$\hat{\lambda}_k(z) = p_k(z|\phi)\, P_k^D(\phi)\, f_{k|k-1}(\phi) \tag{6.16}$$

is the predicted measurement clutter intensity. The probability $P_k^D(\phi)$ in (6.16) is the probability that a clutter target generates a measurement at time t_k.

The computational parts of the intensity filter are outlined in Table 6.1. The table clarifies certain interpretive issues that are glossed over in the discussion. Implementation methods are discussed elsewhere in this chapter.

Likelihood Function of the Data Set

Since $f_{k|k}(x)$ is the intensity of a PPP, it is reasonable to inquire about the pdf of the data $\eta_k = \left(m_k, \left\{z_1, \ldots, z_{m_k}\right\}\right)$. The measurement intensity after the information update of the target state intensity is, applying (2.86) on the augmented space $\mathcal{S}^+$,

$$\lambda_{k|k}(z) = \int_{\mathcal{S}^+} p_k(z|x)\, f_{k|k}(x)\, \mathrm{d}x. \tag{6.17}$$

Therefore the pdf of the data η_k is

$$\begin{aligned} p(\eta_k) &= e^{-\int_{\mathbb{R}^{n_z}} \lambda_{k|k}(z)\,\mathrm{d}z} \prod_{j=1}^{m_k} \lambda_{k|k}(z_j) \\ &= e^{-N_{k|k}} \prod_{j=1}^{m_k} \lambda_{k|k}(z_j), \end{aligned} \tag{6.18}$$

where

$$\begin{aligned} N_{k|k} &= \int_{\mathbb{R}^{n_z}} \lambda_{k|k}(z)\,\mathrm{d}z \\ &= \int_{\mathcal{S}^+} \left(\int_{\mathbb{R}^{n_z}} p_k(z|x)\,\mathrm{d}z \right) f_{k|k}(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^{n_x}} f_{k|k}(x)\,\mathrm{d}x \end{aligned} \tag{6.19}$$

is the estimated mean number of targets. The pdf of η_k is approximate because $f_{k|k}(x)$ is an approximation.

6.2 Relationship to Other Filters

The modeling assumptions of the intensity filter are very general, and specializations are possible. The most important is the PHD filter discussed in Section 6.2.1. By assuming certain kinds of *a priori* knowledge concerning target birth and measurement clutter, and adjusting the filter appropriately, the intensity filter reduces to the PHD filter. The differences between the

intensity and PHD filters are nearly all attributable to the augmented state space $\mathcal{S}^+$. That is, the intensity filter uses the augmented single target state space $\mathcal{S}^+ = \mathcal{S} \cup \phi$, while the PHD filter uses only the single target space $\mathcal{S}$. Using $\mathcal{S}$ practically forces the PHD filter to employ target birth and death processes to model initiation and termination of targets.

A different kind of specialization is the marked multitarget intensity filter. This is a parameterized linear Gaussian sum intensity filter that interprets measurements as target marks. This interpretation is interesting in the context of PPP target models because it implies that joint measurement-target point process is a PPP. Details are discussed in Section 6.2.2.

6.2.1 Probability Hypothesis Density (PHD) Filter

The state ϕ is the basis for the on-line estimates of the intensities of the target birth and measurement clutter PPPs given by (6.13) and 6.15), respectively. If, however, the birth and clutter intensities are known *a priori* to be $b_k(x)$ and $\lambda_k(z)$, then the predictions $\hat{b}_k(x)$ and $\hat{\lambda}_k(z)$ can be replaced by $b_k(x)$ and $\lambda_k(z)$. This is the basic strategy taken by the PHD filter.

The use of *a posteriori* methods makes good sense in many applications. For example, they can help regularize parameter estimates. These methods can also incorporate information not included *a priori* in the Bayes filter. For example, Jazwinski [49] uses an *a posteriori* method to derive the Schmidt-Kalman filter for bias compensation. These methods may improve performance, i.e., if the *a priori* birth and clutter intensities are more accurate or stable than their on-line estimated counterparts, the PHD filter may provide better tracking performance.

Given these substitutions, the augmented space is no longer needed and can be eliminated. This requires some care. If the recursion is simply restricted to $\mathcal{S}$ and no other changes are made, the filter will not be able to discard targets and the target count may balloon out of control. To balance the target birth process, the PHD filter uses a death probability before propagating the multitarget intensity $f_{k-1|k-1}(x)$. This probability was intentionally omitted from the intensity filter because transition into ϕ is target death, and it is redundant to have two death models.

The death process is a Bernoulli thinning process applied to the PPP at time t_{k-1} before targets transition and are possibly detected. Let $d_{k-1}(x)$ denote the probability that a target at time t_{k-1} dies before transitioning to time t_k. The surviving target point process is a PPP and its intensity is $(1 - d_{k-1}(x))\, f_{k-1|k-1}(x)$. Adding Bernoulli death and restricting the recursion to $\mathcal{S}$ reduces the intensity filter to the PHD filter.

Table 6.2 PHD Filter on the State Space $\mathcal{S}$

INPUTS:
Data: $z_{1:m} = \{z_1, \ldots, z_m\} \subset \mathcal{T}$ at time t_k
Target death probability function: $d_{k-1}(x) = \Pr[\text{target death at state } x \in \mathcal{S} \text{ at time } t_{k-1}]$
Target birth probability function: $b_k(x) = \Pr[\text{target birth at state } x \in \mathcal{S} \text{ at time } t_k]$
Probability of target detection: $P_k^D(x)$ at time t_k
Measurement clutter intensity function: $\lambda_k(z)$ at time t_k

OUTPUT: $f_{k|k}(x) = \text{PHDFilter}\left[f_{k-1|k-1}(x),\, d_{k-1}(x),\, b_k(x),\, \lambda_k(z),\, P_k^D(x),\, z_{1:m}\right]$

- Predicted target intensity : For $x \in \mathcal{S}$,
  - Surviving targets:
    $S_k(x) = (1 - d_{k-1}(x))\, f_{k-1|k-1}(y)$
  - Propagated targets:
    $S_k(x) \leftarrow \int_{\mathcal{S}} \Psi_{k-1}(x|y)\, S_k(y)\, \mathrm{d}y$
  - Predicted target intensity:
    $f_{k|k-1}(x) = b_k(x) + S_k(x)$
- Predicted measurement intensity:
  IF $m = 0$, THEN $f_{k|k}(x) = \left(1 - P_k^D(x)\right) f_{k|k-1}(x)$ STOP
  FOR $j = 1 : m$,
  - Intensity contributions from predicted target intensity:
    $\hat{v}_k(z_j) = \int_{\mathcal{S}} p_k(z_j|x)\, P_k^D(x)\, f_{k|k-1}(x)\, \mathrm{d}x$
  - Predicted measurement intensity:
    $\lambda_{k|k-1}(z_j) = \lambda_k(z_j) + \hat{v}_k(z_j)$

  END FOR
- Information updated target intensity: For $x \in \mathcal{S}$,

$$f_{k|k}(x) = \left(1 - P_k^D(x) + \sum_{j=1}^{m} \frac{p_k(z_j|x)\, P_k^D(x)}{\lambda_{k|k-1}(z_j)}\right) f_{k|k-1}(x)$$

- END

6.2.2 *Marked Multisensor Intensity Filter (MMIF)*

The intensity filter assumes targets have the same motion model, and that the sensor measurement likelihood function is the same for all targets and data. Such assumptions are idealized at best. An alternative approach is to develop a parameterized intensity filter that accommodates heterogeneous target motion models and measurement pdfs by using target-specific parameterizations. The notion of target-specific parameterizations in the context of PPP target modeling seems inevitably to lead to the idea of modeling individual targets as a PPP, and then using superposition to obtain the aggregate PPP target model. Parameter estimation using the EM method is natural to

superposition problems, as shown in Chapter 3. The marked multisensor intensity filter (MMIF) is one instance of such an approach.

The MMIF builds on the basic idea that a target at state x is "marked" with a measurement z. If the target is modeled as a PPP, then the joint measurement-target vector (z, x) is a PPP on the Cartesian product of the measurement and target spaces. This is an intuitively reasonable result, but the details needed to see that it is true are postponed to Section 8.1. The MMIF uses a linear Gaussian target motion and measurement model for each target and superposes them against a background clutter model. Since the Gaussian components correspond to different targets, they need not have the same motion model. Similarly, different sensor measurement models are possible. Superposition therefore leads to an affine Gaussian sum intensity function on the joint measurement-target space. The details of the EM method and the final MMIF recursion are given in Appendix E.

The MMIF adheres to the "at most one measurement per target rule" but only *in the mean*, or on average. It does this by reinterpreting the single target pdf as a PPP intensity function, and by interpreting measurements as the target marks. The expected number of targets that the PPP on the joint measurement-target space produces is one.

Another feature of the MMIF is that the EM weights depend on the Kalman filter innovations. The weights in other Gaussian sum filters often involve scaled multiples of the measurement variances, resulting in filters that are somewhat akin to "nearest neighbor" tracking filters.

The limitation of MMIF and other parameterized sum approaches is the requirement to use a fixed number of terms in the sum. This strongly affects its ability to model the number of targets. In practice, various devices can compensate for this limitation, but they are not intrinsic to the filter.

6.3 Implementation

Simply put, targets correspond to the local peaks of the intensity function and the areas of uncertainty correspond to the contours, or isopleths, of the intensity. Very often in practice, isopleths are approximated by ellipsoids in target state space corresponding to error covariance matrices. Methods for locating the local peak concentrations of intensity and finding appropriate covariance matrices to measure the width of the peaks are discussed in this section.

Implementation issues for intensity filters therefore concern two issues. Firstly, it is necessary to develop a computationally viable representation of the information updated intensity function of the filter. Two basic representations are proposed, one based on particles and the other on Gaussian sums. Secondly, postprocessing procedures are applied to the intensity function representation to extract the number of detected targets, together with their

estimated states and corresponding error covariance matrices. Analogous versions of both issues arise in classical single target Bayesian filters.

The fact remains, however, that a proper statistical interpretation of target point estimates and their putative error covariances is lacking for intensity filters. The concern may be dismissed in practice because they are intuitively meaningful and closely resemble their single target Bayesian analogs. The concern is nonetheless worrisome and merits further study.

6.3.1 Particle Methods

The most common and by far the easiest implementation of nonlinear filters is by particle, or sequential Monte Carlo (SMC), methods. In such methods the posterior pdf is represented nonparametrically by a set of particles in target state space, together with a set of associated weights, and estimated target count. Typically these weights are uniform, so the spatial distribution of particles represents the variability of the posterior density. An excellent discussion of SMC methods for Bayesian single target tracking applications is found in the first four chapters of [91].

Published particle methods for the general intensity filter are limited to date to the PHD filter. Extensions to the intensity filter are not reported here. An early and well described particle methodology (as well as an interesting example for tracking on roads) for PHD filters is given in [98]. Particle methods and their convergence properties for the PHD filter are discussed in detail in a series of papers by Vo *et al.* [126]. Interested readers are urged to consult them for specifics.

Tracking in a surveillance region $\mathcal{R}$ using SMC methods starts with an initial set of particles and weights at time t_{k-1} together with the estimated number of targets in $\mathcal{R}$:

$$\left\{\left(x_{k-1|k-1}(\ell), w_{k-1|k-1}(\ell)\right) : \ell = 1, L_{SMC}\right\} \text{ and } N_{k-1|k-1}\,,$$

where $w_{k-1|k-1}(\ell) = 1/L_{SMC}$ for all ℓ. For PHD filters the particle method proceeds in several steps that mimic the procedure outlined in Table 6.2:

- *Prediction.* In the sequential importance resampling (SIR) method, prediction involves thinning a given set of particles with the survival probability $1 - d_k(x)$ and then stochastically transforming the survivors into another particle set using the target motion model $\Psi_{k-1}(y|x)$ with weights adjusted accordingly. Additional new particles and associated weights are generated to model new target initializations.
- *Updating.* The particle weights are multiplicatively (Bayesian) updated using the measurement likelihood function $p_k(z|x)$ and the probability of detection $P_k^D(x)$. The factors are of the form

$$\left(1 - P_k^D(x) + \sum_{j=1}^{m} \frac{p_k(z_k(j)\,|\,x)\,P_k^D(x)}{\lambda_{k|k-1}(z_j)}\right), \tag{6.20}$$

where $z_k(1), \ldots, z_k(m)$ are the measurements at time t_k. The updated particle weights are nonuniform.

- *Normalization.* Compute the scale factor, call it $N_{k|k}$, of the sum of the updated particle weights. Divide all the particle weights by $N_{k|k}$ to normalize the weights.
- *Resampling.* Particles are resampled by choosing i.i.d. samples from the discrete pdf defined by the normalized weights. Resampling restores the particle weights to uniformity.

If the resampling step is omitted, the SMC method leads to particle weight distributions that rapidly concentrate on a small handful of particles and therefore poorly represents the posterior intensity. There are many ways to do the resampling in practice.

By computing $N_{k|k}$ before resampling, it is easy to see that

$$\begin{aligned} N_{k|k} &\approx \int_{\mathcal{R}} \lambda_{k|k}(x)\,\mathrm{d}x \\ &= E\left[\text{Number of targets in } \mathcal{R}\right]. \end{aligned} \tag{6.21}$$

The estimated number of targets in any given subset $\mathcal{R}_0 \subset \mathcal{R}$ is

$$N_{k|k}(\mathcal{R}_0) = \left(\frac{\text{Number of particles in } \mathcal{R}_0}{L_{SMC}}\right) N_{k|k}. \tag{6.22}$$

The estimator is poor for sets $\mathcal{R}_0$ that are only a small fraction of the total volume of $\mathcal{R}$.

The primary limitations of particle approaches in many applications are due to the so-called Curse of Dimensionality[1]: the number of particles needed to represent the intensity function grows exponentially as the dimension of the state space increases. Most applications to date seem to be limited to four or five dimensions. The curse is so wicked that Moore's Law (the doubling of computational capability every 18 months) by itself will do little to increase the effective dimensional limit over a human lifetime. Moore's Law and improved methods together will undoubtedly increase the number of dimensions for which particle filters are practical, but it remains to be seen if general filters of dimension much larger than say six can be treated directly.

[1] The name was first used in 1961 by Richard E. Bellman [8]. The name is apt in very many problems; however, some modern methods in machine learning actually exploit high dimensional embeddings.

6.3.2 Mean Shift Algorithm

As is clear from the earlier discussion of PET, it is intuitively reasonable to think of the multitarget intensity filter as a sequential image processing method. In this interpretation an image comprises the "gray scales" of a set of multidimensional voxels in the target state space $\mathbb{R}^{n_x}$. Such an interpretation enables a host of image processing techniques to be applied to the estimated intensity function.

One technique that lends itself immediately to extracting target point estimates from a particle set approach is the "mean-shift" algorithm. This algorithm, based on ideas first proposed in [35], is widely used in computer vision applications such as image segmentation and tracking. The mean-shift algorithm is an EM algorithm for Gaussian kernel density estimators and a generalized EM algorithm for non-Gaussian kernels [11]. The Gaussian kernel is computationally very efficient in the mean-shift method.

Denote the non-ϕ particle set representing the PPP intensity at time t_k by

$$\left\{x_{k|k}(\ell) : \ell = 1, \ldots, L_{SMC}\right\}.$$

The intensity function is modeled as a scalar multiple of the kernel estimator

$$\lambda_k(x) = I_k \sum_{\ell=1}^{L_{SMC}} \mathcal{N}(x; x_{k|k}(\ell), \Sigma_{\text{ker}}), \tag{6.23}$$

where $\mathcal{N}(x; x_{k|k}(\ell), \Sigma_{\text{ker}})$ is the kernel. The covariance matrix Σ_{ker} is specified, not estimated. Intuitively, the larger Σ_{ker}, the fewer the number of local maxima in the intensity (6.23), and conversely. The scale factor $I_k > 0$ is estimated by the particle filter and is taken as known here.

The form (6.23) has no parameters to estimate, so extend it by defining

$$\lambda_k(x; \mu) = I_k \sum_{\ell=1}^{L_{SMC}} \mathcal{N}\left(x; x_{k|k}(\ell) - \mu, \Sigma_{\text{ker}}\right), \tag{6.24}$$

where μ is an unknown rigid translation of the intensity (6.23). It is not hard to see that the ML estimate of μ is a local maximum of the kernel estimate, that is, a point estimate for a target. The vector μ is estimated from data using the EM method. The clever part is using an artificial data set with only one point in it, namely, the origin.

Let $r = 0, 1, \ldots$ denote the EM iteration index, and let $\mu^{(0)}$ be a specified initial value for the mean. The auxiliary function is given by Eqn. (3.20) with $m = 1$, $x_1 = 0$, $L = L_{SMC}$, $\theta_\ell = \mu$, and

$$\lambda_\ell(x; \mu) \equiv I_k \mathcal{N}\left(x; x_{k|k}(\ell) - \mu, \Sigma_{\text{ker}}\right).$$

The bounded surveillance region $\mathcal{R}$ is taken to be $\mathbb{R}^{\eta_x}$. Define the weights

$$w_\ell\left(\mu^{(r)}\right) = \frac{\mathcal{N}\left(0; x_{k|k}(\ell) - \mu^{(r)}, \Sigma_{\text{ker}}\right)}{\sum_{\ell'=1}^{L_{SMC}} \mathcal{N}\left(0; x_{k|k}(\ell') - \mu^{(r)}, \Sigma_{\text{ker}}\right)} = \frac{\mathcal{N}\left(x_{k|k}(\ell); \mu^{(r)}, \Sigma_{\text{ker}}\right)}{\sum_{\ell'=1}^{L_{SMC}} \mathcal{N}\left(x_{k|k}(\ell'); \mu^{(r)}, \Sigma_{\text{ker}}\right)}. \tag{6.25}$$

The auxiliary function in the present case requires no sum over j as done in (3.20), so

$$Q(\mu; \mu^{(r)}) = -I_k + \sum_{\ell=1}^{L_{SMC}} w_\ell\left(\mu^{(r)}\right) \log\left\{I_k \mathcal{N}\left(0; x_{k|k}(\ell) - \mu, \Sigma_{\text{ker}}\right)\right\}. \tag{6.26}$$

The EM update of μ is found by taking the appropriate gradient, yielding

$$\mu^{(r+1)} = \frac{\sum_{\ell=1}^{L_{SMC}} w_\ell\left(\mu^{(r)}\right) x_{k|k}(j)}{\sum_{\ell=1}^{L_{SMC}} w_\ell(\mu^{(r)})}. \tag{6.27}$$

Substituting (6.25) and canceling the common factor gives the classical mean-shift iteration:

$$\mu^{(r+1)} = \frac{\sum_{\ell=1}^{L_{SMC}} \mathcal{N}\left(x_{k|k}(\ell); \mu^{(r)}, \Sigma_{\text{ker}}\right) x_{k|k}(\ell)}{\sum_{\ell=1}^{L_{SMC}} \mathcal{N}\left(x_{k|k}(\ell); \mu^{(r)}, \Sigma_{\text{ker}}\right)}. \tag{6.28}$$

The update of the mean is a convex combination of the particle set. Convergence to a local maximum $\mu_k^{(r)} \to \hat{x}_{k|k}$ is guaranteed as $r \to \infty$.

Different initializations are needed for different targets, so the mean shift algorithm needs a preliminary clustering method to initialize it, as well as to determine the number of peaks in the data correspond to targets. Also, the size of the kernel depends somewhat on the number of particles and may need to be adjusted to smooth the intensity surface appropriately.

6.3.3 Multimode Algorithms

Identifiability remains a problem with the mean shift algorithm, that is, there is no identification of the point estimate to a target except through the way the starting point of the iteration is chosen. This may cause problems when targets are in close proximity.

One way to try to resolve the problem is to use the particles themselves as points to feed into another tracking algorithm. This method exploits se-

rial structure in the filter estimates and may disambiguate closely spaced targets. However, particles are serially correlated and do not satisfy the conditional independence assumptions of measurement data, so the resulting track estimates may be biased.

6.3.4 Covariance Matrices

Matrix CRBs

The true error covariance matrix of the ML point estimate computed by the mean shift method is not available; however, the CRB can be evaluated using (4.21). This CRB is appropriate if the available data are reasonably modeled as realizations of a PPP with intensity (6.24). If the intensity (6.24) function is a high fidelity model of the intensity, the FIM of μ is

$$J(\mu) = I_k \Sigma_{\text{ker}}^{-1} \Sigma(\mu) \Sigma_{\text{ker}}^{-1}, \tag{6.29}$$

where the matrix $\Sigma(\mu)$ is

$$\Sigma(\mu) = \sum_{\ell=1}^{L} \sum_{\ell'=1}^{L} \int_{\mathbb{R}^{n_x}} w(x, \mu; \ell, \ell') \left(x - x_{k|k}(\ell) + \mu\right) \left(x - x_{k|k}(\ell') + \mu\right)^T \mathrm{d}x \tag{6.30}$$

and the weighting function is

$$w(x, \mu; \ell, \ell') = \frac{\mathcal{N}\left(x; x_{k|k}(\ell) - \mu, \Sigma_{\text{ker}}\right) \mathcal{N}\left(x; x_{k|k}(\ell') - \mu, \Sigma_{\text{ker}}\right)}{\sum_{\ell''=1}^{L_{SMC}} \mathcal{N}\left(x; x_{k|k}(\ell'') - \mu, \Sigma_{\text{ker}}\right)}. \tag{6.31}$$

The CRB is $J^{-1}(\mu)$ evaluated at the true value of μ. Because the integral is over all of $\mathbb{R}^{n_x}$, a change of variables shows that the information matrix $J(\mu)$ is independent of the true value of μ. This means that the FIM is not target specific.

A local bound is desired, since the mean shift algorithm converges to a local peak of the intensity. By restricting the intensity model to a specified bounded gate $G \subset \mathbb{R}^{n_x}$, the integral in (6.30) is similarly restricted. The matrix $\Sigma(\mu)$ is thus a function of G. The gated CRB is local to the gate, i.e., it is a function of the target within the gate.

OIM: The Surrogate CRB

The OIM is the Hessian matrix of the negative loglikelihood evaluated at the MAP point estimate; it is often used as a surrogate for Fisher information

when the likelihood function is complicated. Its inverse is the surrogate CRB of the estimate $\hat{x}_{k|k}$. The construction of OIMs for the mean-shift algorithm is implicit in [11].

The loglikelihood function of μ using the intensity function (6.23) and the one data point $x_1 = 0$ is

$$\log p(\mu) = -I_k L_{SMC} + \log \left[\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(x_{k|k}(\ell); \mu, \Sigma_{\text{ker}}) \right].$$

Direct calculation gives the general expression

$$\begin{aligned} \nabla_\mu \left[\nabla_\mu \log p(\mu) \right]^T = \Sigma_{\text{ker}}^{-1} + \frac{1}{\kappa^2} \left[\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\mu; x_{k|k}(\ell), \Sigma_{\text{ker}}) \Sigma_{\text{ker}}^{-1} \left(\mu - x_{k|k}(\ell) \right) \right] \\ \times \left[\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\mu; x_{k|k}(\ell), \Sigma_{\text{ker}}) \Sigma_{\text{ker}}^{-1} \left(\mu - x_{k|k}(\ell) \right) \right]^T \\ - \frac{1}{\kappa} \Sigma_{\text{ker}}^{-1} \left[\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\mu; x_{k|k}(\ell), \Sigma_{\text{ker}}) \left(\mu - x_{k|k}(\ell) \right) \left(\mu - x_{k|k}(\ell) \right)^T \right] \Sigma_{\text{ker}}^{-1}, \end{aligned}$$

where the normalizing constant is

$$\kappa = \sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\mu; x_{k|k}(\ell), \Sigma_{\text{ker}}).$$

The observed information matrix is evaluated at the MAP estimate $\mu = \hat{x}_{k|k}$. The middle term is proportional to $\left[\nabla_\mu p(\mu) \right] \left[\nabla_\mu p(\mu) \right]^T$ and so is zero any stationary point of $p(\mu)$, e.g., at the MAP estimate $\mu = \hat{x}_{k|k}$. The OIM is therefore

$$\begin{aligned} OIM\left(\hat{x}_{k|k} \right) = \Sigma_{\text{ker}}^{-1} \\ - \Sigma_{\text{ker}}^{-1} \left[\frac{\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\hat{x}_{k|k}; x_{k|k}(\ell), \Sigma_{\text{ker}}) \left(\hat{x}_{k|k} - x_{k|k}(\ell) \right) \left(\hat{x}_{k|k} - x_{k|k}(\ell) \right)^T}{\sum_{\ell=1}^{L_{SMC}} \mathcal{N}(\mu; x_{k|k}(\ell), \Sigma_{\text{ker}})} \right] \Sigma_{\text{ker}}^{-1}. \end{aligned} \tag{6.32}$$

The CRB surrogate is $OIM^{-1}\left(\hat{x}_{k|k} \right)$. The inverse exists because the OIM is positive definite at $\hat{x}_{k|k}$. This matrix is, in turn, a surrogate for the error covariance matrix.

The OIM for $\hat{x}_{k|k}$ can be computed efficiently in conjunction with any EM method (see [61] for a general discussion). As noted in Section 4.7, the

statistical interpretation of the OIM is unresolved in statistical circles. Its utility should be carefully investigated in applications.

6.3.5 Gaussian Sum Methods

An alternative to representing the PPP intensity using particle filters is to use a Gaussian sum instead. An advantage of this method is that the clustering and track extraction issues are somewhat simpler, that is, target state estimates and covariance matrices are extracted from the means and variances of the Gaussian sum instead of a myriad of particles. The Gaussian sum approach is especially attractive for linear Gaussian target motion and measurement models because the prediction and information update steps are closed form, assuming constant survival and detection functions (i.e., the PPP thinning functions are independent of state).

Gaussian sum implementations of the PHD filter are carefully discussed by Vo and his colleagues [124]. In this approach, an unnormalized Gaussian sum is used to approximate the intensity. These methods are important because they have the potential to be useful in higher dimensions than particle filters. In the end, however, Gaussian sum intensity filters will also suffer from the curse of target state space dimensionality.

Gaussian sum methods for intensity estimation comprises several steps:

- *Prediction.* The target intensity at time t_{k-1} is a Gaussian sum, to which is added a target birth process that is modeled by a Gaussian sum. The prediction equation for every component in the Gaussian sum is identical to a Kalman filter prediction equation.
- *Component Update.* For each point measurement, the predicted Gaussian components are updated using the usual Kalman update equations. The update therefore increases the number of terms in the Gaussian sum if there is more than one measurement. This step has two parts. In the first, the means and covariance matrices are evaluated. In the second, the coefficients of the Gaussian sum are updated by a multiplicative procedure.
- *Merging and Pruning.* The components of the Gaussian sum are merged and pruned to obtain a "nominal" number of terms. Various reasonable strategies are available for such purposes, as detailed in [124]. This step is the analog of resampling in the particle method.

Some form of pruning is necessary to keep the size of the Gaussian sum bounded over time, so the last – and most heuristic – step cannot be omitted.

Left out of this discussion are details that relate the weights of the Gaussian components to the estimated target count. These details can be found in [124]. For nonlinear target motion and measurement models, [124] proposes both the extended and the unscented Kalman filters. Vo and his colleagues also present Gaussian sum implementations of the CPHD filter in [124] and [125].

6.3.6 Regularization

Intensity filters are in the same class of stochastic inverse problems as image reconstruction in emission tomography — the sequence $t_0, t_1, \ldots, t_k$ of intensity filter estimates $f_{k|k}(x)$ is essentially a movie (in dimension n_x) in target state space. As discussed in Section 5.7, such problems suffer from serious noise and numerical artifacts. The high dimensionality of the PPP parameter, i.e., the number of voxels of the intensity function, makes regularization a priority in all applications. Regularization for intensity filters is a relatively new subject. Methods such as cardinalization are inherently regularizing.

Grenander's method of sieves used in Section 5.7 for regularizing PET adapts to the intensity filter, but requires some additional structure. The sieve kernel $k_0(x|u)$ is a pdf on $\mathcal{S}^+$, so that

$$\int_{\mathcal{S}^+} k_0(x|u)\,\mathrm{d}x = 1 \tag{6.33}$$

for all points u in the discrete-continuous space $\mathcal{U}^+ = \mathcal{U} \cup \phi$. As before, the choice of kernel and space $\mathcal{U}^+$ is very flexible. The multitarget intensity at every time t_k, $k = 0, 1, \ldots$, is restricted to the collection of functions of the form

$$f_{k|k}(x) = \int_{\mathcal{U}^+} k_0(x|u)\,\zeta_{k|k}(u)\,\mathrm{d}u \qquad \text{for some } \zeta_{k|k}(u) > 0\,. \tag{6.34}$$

The kernel k_0 can be a function of time t_k if desired. The restriction (6.34) is also imposed on the predicted target intensity:

$$f_{k|k-1}(x) = \int_{\mathcal{U}^+} k_0(x|u)\,\zeta_{k|k-1}(u)\,\mathrm{d}u \qquad \text{for some } \zeta_{k|k-1}(u) > 0\,. \tag{6.35}$$

Substituting (6.35) into the predicted measurement intensity (6.3) gives

$$\lambda_{k|k-1}(z) = \int_{\mathcal{U}^+} \widetilde{p}_k(z|u)\,\zeta_{k|k-1}(u)\,\mathrm{d}u\,, \tag{6.36}$$

where

$$\widetilde{p}_k(z|u) = \int_{\mathcal{S}^+} p_k(z|x)\,k_0(x|u)\,\mathrm{d}x\,. \tag{6.37}$$

is the regularized measurement likelihood function.

An intensity filter is used to update $\zeta_{k|k}(u)$. This filter employs a transition function $\Phi_{k-1}(\cdot|\cdot)$ to provide a dynamic connection between the current intensity $\zeta_{k-1|k-1}(u)$ and predicted intensity $\zeta_{k|k-1}(u)$. The function $\Phi_{k-1}(\cdot|\cdot)$ is specified for all u and v in $\mathcal{U}^+$ and is, in principle, any reasonable function, but in practice is linked to the target motion model $\Psi_{k-1}(\cdot|\cdot)$.

One way to define $\Phi_{k-1}(\cdot|\cdot)$ requires defining an additional kernel. Substitute (6.34) into the predicted detected target intensity to obtain

$$\begin{aligned} f_{k|k-1}(x) &= \int_{\mathcal{S}^+} f_{k-1|k-1}(y)\,\Psi_{k-1}(x|y)\,\mathrm{d}y \\ &= \int_{\mathcal{S}^+} \left(\int_{\mathcal{U}^+} k_0(y|u)\,\zeta_{k-1|k-1}(u)\,\mathrm{d}u \right) \Psi_{k-1}(x|y)\,\mathrm{d}y \\ &= \int_{\mathcal{U}^+} \widetilde{\Psi}_{k-1}(x|u)\,\zeta_{k-1|k-1}(u)\,\mathrm{d}u\,, \end{aligned}$$

where

$$\widetilde{\Psi}_{k-1}(x|u) = \int_{\mathcal{S}^+} \Psi_{k-1}(x|y)\,k_0(y|u)\,\mathrm{d}y\,. \tag{6.38}$$

Define a Bayesian kernel $k_1(v|x)$ so that

$$\int_{\mathcal{U}^+} k_1(v|x)\,\mathrm{d}v = 1 \tag{6.39}$$

for all points $x \in \mathcal{S}^+$. Like the sieve kernel $k_0(\cdot)$, the Bayesian kernel $k_1(v|x)$ is very flexible. It is easily verified that the function

$$\Phi_{k-1}(v|u) = \int_{\mathcal{S}^+} k_1(v|x)\,\widetilde{\Psi}_{k-1}(x|u)\,\mathrm{d}x \tag{6.40}$$

is a valid transition function for all $k_0(\cdot)$ and $k_1(\cdot)$.

Given the intensity $\widehat{\zeta}_{k-1|k-1}(u)$ from time t_{k-1}, the predicted intensity $\widehat{\zeta}_{k|k-1}(u)$ at time t_k is defined by

$$\widehat{\zeta}_{k|k-1}(u) = \int_{\mathcal{U}^+} \Phi_{k-1}(u|v)\,\widehat{\zeta}_{k-1|k-1}(v)\,\mathrm{d}v\,.$$

The information updated intensity $\widehat{\zeta}_{k|k}(u)$ is evaluated via the intensity filter using the regularized measurement pdf (6.37) and the predicted measurement intensity (6.36). The regularized target state intensity at time t_k is the integral

$$f_{k|k}(x) = \int_{\mathcal{U}^+} k(x|u)\,\widehat{\zeta}_{k|k}(u)\,\mathrm{d}u\,. \tag{6.41}$$

The regularized intensity $f_{k|k}(x)$ depends on the sieve and Bayesian kernels $k_0(\cdot)$ and $k_1(\cdot)$.

The question of how best to define the $k_0(\cdot)$ and $k_1(\cdot)$ kernels depends on the application. It is common practice to define kernels using Gaussian pdfs. As mentioned in Section 5.7, the sieve kernel is a kind of measurement

smoothing kernel. If $\dim(\mathcal{U}) < \dim(\mathcal{S})$, the Bayesian kernel disguises observability issues, that is, many points $x \in \mathcal{S}$ map with the same probability to a given point $u \in \mathcal{U}$. This provides a mechanism for target state space smoothing.

6.4 Estimated Target Count

Estimating the number of targets present in the data is a difficult hypothesis testing problem. In a track before detect (TBD) approach, it is a track management function that is integrated with the tracking algorithm. Intensity filters seem to offer an alternative way to integrate, or fuse, the multitarget track management and state estimation functions.

6.4.1 Sources of Error

Accurate knowledge of the target detection probability function $P_k^D(x)$ is crucial to correctly estimating target count. An incorrect value of $P_k^D(x)$ is a source of systematic error. For example, if the filter uses the value $P_k^D(x) = .5$ but in fact *all* targets *always* show up in the measured data, the estimated mean target count will be high by a factor of two. This example is somewhat extreme but it makes the point that correctly setting the detection probability is an important task for the track management function. The task involves executive knowledge about changing sensor performance characteristics, as well as executive decisions external to the tracking algorithm about the number of targets actually present — decisions that feedback to validate estimates of $P_k^D(x)$. Henceforth, the probability of detection function $P_k^D(x)$ is assumed accurate.

There are other possible sources of error in target count estimates. Birth-death processes can be difficult to tune in practice, regardless of whether they are modeled implicitly as transitions into and out of a state ϕ in the intensity filter, or explicitly as in the PHD filter. If in an effort to detect new targets early and hold track on them as long as possible, births are too spontaneous and deaths are too infrequent, the target count will be too high on average. Conversely, it will be too low with delayed initiation and early termination. Critically damped designs, however that concept is properly defined in this context, would seem desirable in practice. In any event, tuning is a function of the track management system.

Under the PPP model, the estimated expected number of targets in a given region, $\mathcal{A}$, is the integral over $\mathcal{A}$ of the estimated multitarget intensity. Because the number is Poisson distributed, the variance of the estimated number of targets is equal to the mean number. This large variance is an

unhappy fact of life. For example, if 10 targets are present, the standard deviation on the estimated number is $\sqrt{10} \approx 3$.

It is therefore foolhardy in practice to assume that the estimated of the number of targets is the number that are *actually* present. Variance reduction in the target count estimate is a high priority from the track management point of view for both intensity and PHD filters.

6.4.2 Variance Reduction

The multisensor intensity filter discussed in Section 6.5 reduces the variance by averaging the sensor-level intensity functions over the number of sensors that contribute to the filter. Consequently, if the individual sensors estimate target count correctly, so does the multisensor intensity filter.

Moreover, and just as importantly, the variance of the target count estimate of the multisensor intensity filter is reduced by a factor of M compared to that of a single sensor, where M is the number of sensors, assuming for simplicity that the sensor variances are identical.

This important variance reduction property is analogous to estimators in other applications. An especially prominent and well known example is power spectral estimation of wideband stationary time series. For such signals the output bins of the DFT of a non-overlapped blocks of sampled data are distributed with a mean level equal to the signal power in the bin, and the variance equal to the mean. This property of the periodogram is well known, as is the idea of time averaging the periodogram, i.e., the non-overlapped DFT outputs, to reduce the variance of spectral estimates.[2] The Wiener-Khinchin theorem justifies averaging the short term Fourier transforms of nonoverlapped data records as a way to estimate the power spectrum with reduced variance. In practice, the number of DFT records averaged is often about 25.

The multisensor intensity filter is low computational complexity, and applicable to distributed heterogeneous sensor networks. It is thus practical and widely useful. Speculating now for the sheer fun of it, if the number of data records in a power spectral average carries over to multisensor multitarget tracking problems, then the multisensor intensity filter achieves satisfactory performance for many practical purposes with about 25 sensors.

[2] Averaging trades off variance reduction and spectral resolution. It was first proposed by M. S. Bartlett [6] in 1948.

6.5 Multiple Sensor Intensity Filters

To motivate the discussion, consider the SPECT application of Section 5.4. In SPECT, a single gamma camera is moved to several view angles and a snapshot is taken of light observed emanating from gamma photon absorption events. The EM recursion given by Eqn. (5.67) is the superposition of the intensity functions estimated by each of the camera view angles. Intuitively, since different snapshots *cannot* contain data from the same absorption event, the natural way to fuse the multiple images into one image is to add them. The theoretical justification is that, since the number of absorptions is unknown and Poisson distributed, the estimates of the spatial distribution of the radioisotope that are obtained from different view angles are independent, *not* conditionally independent, so the intensity functions (images) are superposed.

The general multisensor filtering problem is not concerned with gamma photon absorptions, but rather with physical entities (aircraft, ships, etc.) that persist over long periods of time — target physics are very different from the physics of photons. Nonetheless, for reasons discussed at the beginning of this chapter, a PPP multitarget model is used for single sensor multitarget tracking. It is assumed, very reasonably, that the same PPP target model holds regardless of how many sensors are employed to detect and track targets. The analogy with SPECT is now clear: each sensor in the multisensor filtering problem is analogous to a camera view angle in SPECT, and the sensor-level data are analogous to the camera snapshot data.

The PPP multitarget model has immediate consequences. The most important is that conditionally independent sensors are actually independent due to the independence property of PPPs discussed in Section 2.9. The multisensor intensity filter averages the sensor-level intensities [111]. The reason the sensor intensities are averaged and not simply added is that targets are persistent (unlike absorption events). Adding the intensities "over count" targets because each sensor provides its own independent estimate of target intensity.

The possibility of regularization at the multisensor level is not considered explicitly. Although perhaps obvious, the multisensor intensity filter is fully compatible with sensor-level regularization methods.

Let the number of sensors be $M \geq 1$. It is assumed that the target detection probability functions, $P_k^D(x;\ell)$, $\ell = 1, \ldots, M$, are specified for each sensor. The sensor-specific state space coverage is defined by

$$C_k(\ell) = \left\{ x \in \mathcal{S} : \quad P_k^D(x;\ell) > 0 \right\}. \tag{6.42}$$

In homogeneous problems the sensor coverages are identical, i.e., $C_k(\ell) \equiv C_k$ for all ℓ. Heterogeneous problems are those that are not homogeneous.

Two sensors with the same coverage need not have the same, or even closely related, probability of detection functions. As time passes, homoge-

neous problems may turn into heterogeneous ones, and vice versa. In practice, it is probably desirable to set a small threshold to avoid issues with very small probabilities of detection. Homogeneous and heterogeneous problems are discussed separately.

6.5.1 Identical Coverage Sensors

For $\ell = 1, \ldots, M$, let the measurement space of sensor ℓ be $\mathcal{Z}(\ell)$. Denote the measurement pdf by $p_k(z|x;\ell)$, where $z(\ell) \in \mathcal{Z}(\ell)$ is a point measurement. The predicted measurement intensity is

$$\lambda_{k|k-1}(z;\ell) = \int_{\mathcal{S}^+} p_k(z|x;\ell) P_k^D(x;\ell) f_{k|k-1}(x)\,\mathrm{d}x, \quad z \in \mathcal{Z}(\ell). \tag{6.43}$$

The measured data from sensor ℓ is

$$\xi_k(\ell) = (m_k(\ell), \{z_k(1;\ell), \ldots, z_k(m_k(\ell);\ell)\}). \tag{6.44}$$

The sensor-level intensity filter is

$$f_{k|k}(x;\ell) = L_k(\xi_k(\ell)|x;\ell) f_{k|k-1}(x), \tag{6.45}$$

where the Bayesian information update factor is

$$L_k(\xi_k(\ell)|x;\ell) = 1 - P_k^D(x;\ell) + \sum_{j=1}^{m_k(\ell)} \frac{p_k(z_k(j;\ell)|x;\ell) P_k^D(x;\ell)}{\lambda_{k|k-1}(z_k(j;\ell);\ell)}. \tag{6.46}$$

The multisensor intensity filter is the average:

$$\begin{aligned} f_{k|k}^{\text{Fused}}(x) &= \frac{1}{M} \sum_{\ell=1}^{M} f_{k|k}(x;\ell) \\ &= \frac{1}{M} \left(\sum_{\ell=1}^{M} L_k(\xi_k(\ell)|x;\ell) \right) f_{k|k-1}(x). \end{aligned} \tag{6.47}$$

If the sensor-level intensity filters are maintained by particles, and the number of particles is the same for all sensors, the multisensor averaging filter is implemented merely by pooling all the particles (and randomly downsampling to the desired particle size, if desired).

Multisensor fusion methods sometimes rank sensors by some relative quality measure. This is unnecessary for the multisensor intensity filter. The reason is that sensor quality, as measured by the probability of detection

functions $P_k^D(x;\ell)$ and the sensor measurement pdfs $p_k(z|x;\ell)$, is automatically included in (6.47).

The multisensor intensity filter estimates the number of targets as

$$\begin{aligned} N_{k|k}^{\text{Fused}} &= \int_{\mathcal{S}} f_{k|k}^{\text{Fused}}(x)\,\mathrm{d}x \\ &= \frac{1}{M}\sum_{\ell=1}^{M}\int_{\mathcal{S}} f_{k|k}(x;\ell)\,\mathrm{d}x \\ &= \frac{1}{M}\sum_{\ell=1}^{M} N_{k|k}(\ell), \end{aligned} \tag{6.48}$$

where $N_{k|k}(\ell)$ is the number of targets estimated by sensor ℓ. Taking the expectation of both sides gives

$$E\left[N_{k|k}^{\text{Fused}}\right] = \frac{1}{M}\sum_{\ell=1}^{M} E\left[N_{k|k}(\ell)\right]. \tag{6.49}$$

If the individual sensors are unbiased on average, or in the mean, then $E[N_{k|k}(\ell)] = N$ for all ℓ, where N is the true number of targets present. Consequently, the multisensor intensity filter is also unbiased.

The estimate $N_{k|k}(\ell)$ is Poisson distributed, and the variance of a Poisson distribution is equal to its mean, so

$$\text{Var}[N_{k|k}(\ell)] = N, \qquad \ell = 1,\ldots,M.$$

The variance of the average in (6.48) is the average of the variances, since the terms in the sum are independent. Thus,

$$\begin{aligned} \text{Var}\left[N_{k|k}^{\text{Fused}}\right] &= \frac{1}{M}\sum_{\ell=1}^{M}\text{Var}\left[N_{k|k}(\ell)\right] \\ &= \frac{N}{M}. \end{aligned} \tag{6.50}$$

In words, the standard deviation of the estimated target count in the multisensor intensity filter is smaller than that of individual sensors by a factor of $\sqrt{M}$, where M is the number of fielded sensors. This is an important result for spatially distributed networked sensors.

The averaging multisensor intensity filter is derived by Bayesian methods in [111]. It is repeated here in outline. The Bayesian derivation of the single sensor intensity filter in Appendix D is a good guide to the overall structure of most of the argument.

The key is to exploit the PPP target model on the augmented space $\mathcal{S}^+$. Following the lead of Eqn. (D.5) in Appendix D, the only PPP realizations

with nonzero likelihood have $m_k = \sum_{\ell=1}^{M} m_k(\ell)$ microtargets. The m_k PPP microtargets are paired with the m_k sensor data points, so the overall joint likelihood function is the product of the sensor data likelihoods given the microtarget assignments. This product is then summed over all partitions of the m_k microtargets into parts of size $m_k(1), \ldots, m_k(M)$.

The sum over all partitions is the Bayes posterior pdf on the event space $\mathcal{E}(\mathcal{S}^+)$. It is a very complex sum, but it has important structure. In particular, the single target marginal pdfs are identical, that is, the integrals over all but one microtarget state are all the same. After tedious algebraic manipulation, the single target marginal pdf is seen to be

$$p_X^{\text{Fused}}(x) = \frac{1}{m_k} \sum_{\ell=1}^{M} L_k(\xi_k(\ell)|x;\ell)\, f_{k|k-1}(x), \qquad x \in \mathcal{S}^+. \tag{6.51}$$

The mean field approximation is now invoked as in Eqn. (D.13). Under this approximation, $f_{k|k}^{\text{Fused}}(x) = c\, p_X^{\text{Fused}}(x)$, where the constant $c > 0$ is estimated. From (6.17) and (6.18), the measurement intensity is

$$\lambda_{k|k}^{\text{Fused}}(z) = c \int_{S^+} p_k(z|x)\, p_X^{\text{Fused}}(x)\, \mathrm{d}x, \tag{6.52}$$

so the likelihood function of c given the data sets $\xi_k(\ell)$ is

$$\begin{aligned} \mathscr{L}(c;\xi_k(1),\ldots,\xi_k(M)) &= \prod_{\ell=1}^{M} \left\{ e^{-\int_{S^+} c\, p_X^{\text{Fused}}(x)\mathrm{d}x} \prod_{j=1}^{m_k(\ell)} \lambda_{k|k}^{\text{Fused}}(z_k(j;\ell)) \right\} \\ &\propto e^{-cM} c^{m_k}. \end{aligned}$$

Setting the derivative with respect to c to zero and solving gives ML estimate $\hat{c}_{ML} = \frac{m_k}{M}$. The multisensor intensity filter is $\hat{c}_{ML}\, p_X^{\text{Fused}}(x)$. Further purely technical details of the Bayesian derivation provide little additional insight, so they are omitted.

The multiplication of the conditional likelihoods of the sensor data happens at the PPP event level, where the correct associations of sensor data to targets is assumed unknown. The result is that the PPP parameters — the intensity functions — are averaged, not multiplied. The multisensor intensity filter therefore cannot reduce the area of uncertainty of the extracted target point estimates. In other words, the multisensor intensity averaging filter cannot improve spatial resolution. Intuitively, the multisensor filter achieves variance reduction in the target count by foregoing spatial resolution of the target point estimates.

6.5.2 Heterogeneous Sensor Coverages

When the probability of detection functions are not identical, the multisensor intensity filter description is somewhat more involved. At each target state x the only sensors that are averaged are those whose detection functions are nonzero at x. This leads to a "quilt-like" fused intensity that may have discontinuities at the boundaries of sensor detection coverages.

The Bayesian derivation of (6.47) outlined above assumes that all the microtargets of the PPP realizations can be associated to any of the M sensors. If, however, any of these microtargets fall outside the coverage set of a sensor, then the assignment is not valid. The way around the problem is to partition the target state space appropriately.

The total coverage set

$$C = \cup_{\ell=1}^{M} C(\ell) \tag{6.53}$$

contains points in target state space that are covered by at least one sensor. Partition C into disjoint, nonoverlapping sets B_ρ that comprise points covered by exactly ρ sensors, $\rho = 1, \ldots, M$. Now partition B_ρ into subsets $B_{\rho,1}, \ldots, B_{\rho,j_\rho}$ that are covered by different combinations of ρ sensors. To simplify notation, denote the sets $\left\{B_{\rho j}\right\}$ by $\{\mathcal{A}_\omega\}$, $\omega = 1, 2, \ldots, \Omega$. The sets are disjoint and their union is all of C:

$$C = \cup_{\omega=1}^{\Omega} \mathcal{A}_\omega, \qquad \mathcal{A}_i \cap \mathcal{A}_j = \varnothing \ \text{ for } \ i \neq j. \tag{6.54}$$

No smaller number of sets satisfies (6.54) and also has the property that each set $\mathcal{A}_\omega$ in the partition is covered by the same subset of sensors.

The overall multisensor intensity filter operates on the partition $\{\mathcal{A}_\omega\}$. The assignment assumptions of the multisensor intensity filter are satisfied in each of the sets $\mathcal{A}_\omega$. Thus, the overall multisensor filter is

$$f_{k|k}^{\text{Fused}}(x) = \frac{1}{|\mathcal{A}_\omega|}\left(\sum_{\ell \in \mathcal{I}(\mathcal{A}_\omega)} L_k(\xi_k(\ell)|x;\ell)\right) f_{k|k-1}(x), \quad x \in \mathcal{A}_\omega. \tag{6.55}$$

where $\mathcal{I}(\mathcal{A}_\omega)$ are the indices of the sensors that contribute to the coverage of $\mathcal{A}_\omega$, and $|\mathcal{A}_\omega|$ is the number of sensors that do so.

The multisensor intensity filter is thus a kind of "patchwork" with the pieces being the sets $\mathcal{A}_\omega$ of the partition. The variance of the multisensor filter is not the same throughout C — the more sensors contribute to the coverage of a set in the partition, the smaller the variance in that set.

6.6 Historical Note

The PHD (Probability Hypothesis Density) filter ([66], [67] and references therein) was pioneered by Mahler beginning about 1994. Mahler was apparently influenced in this work [66, p. 1154] by an intriguing approach to additive evidence accrual proposed in 1993 by Stein and Winter [108]. An alternative and more insightful bin-based derivation of the PHD filter was discovered in 2008 by Erdinc, Willett, and Bar Shalom [29].

The intensity filter of Streit and Stone [117] is very similar to the PHD filter, differing from it primarily in its use of an augmented target state space, $\mathcal{S}^+$, instead of birth-death processes to model target initiation and termination. The PHD filter is recovered from the intensity filter by modifying the posterior intensity. This paper is the source of the Bayesian approach to the intensity filter given in Section 6.1. The other two approaches presented in that section follow the discussion given in [112]. It draws on the connections to PET to gain intuitive insight into the interpretation of the PPP target model. These approaches greatly simplify the mathematical discussion surrounding intensity filters since it builds on work already presented in earlier chapters for PET imaging.

The multisensor intensity filter was first derived in 2008 by Streit [111] using a rigorous Bayesian methodology, followed by the same kind of PPP approximation as is used in the single sensor intensity filter. The general multisensor problem was presented for both homogeneous and heterogeneous sensor coverage. The theoretical and practice importance of the variance reduction property of the averaging multisensor filter was also discussed in the same paper.

Mahler [66] reports a product form for the multisensor PHD filter. It is unclear if the product form estimates target count correctly. The problem arises from the need for each of the sensor-level integrals of intensity as well as the multisensor integral of intensity to estimate the target count. In any event, the multisensor intensity filter and the multisensor PHD filter take quite different forms, and therefore are different filters.

The MMIF tracking filter is new. It was developed by exploring connections between intensity filters and the PMHT (Probabilistic Multiple Hypothesis Tracking) filter, a Gaussian mixture (*not* Gaussian sum) approach to multitarget tracking developed by Streit and Luginbuhl [113, 114, 115] that dates to 1993. These connections reveal the PPP underpinnings of PMHT.

Appendix D
Bayesian Derivation of Intensity Filters

The multitarget intensity filter is derived by Bayesian methods in this appendix. The posterior point process is developed first, and then the posterior point process is approximated by a PPP. Finally, the last section discusses the relationship between this method and the "first moment"approximation of the posterior point process.

The steps of the intensity filter are outlined in Fig. 6.1. The PPP interpretations of these steps are thinning, approximating the Bayes update with a PPP, and superposition. The PPP at time t_k is first thinned by detection. The two branches of the thinning are the detected and undetected target PPPs. Both branches are important. Their information updates are different. The undetected target PPP is the lesser branch. Its information update is a PPP. The detected target branch is the main branch, and its information update comprises two key steps. Firstly, the Bayes update of the posterior point process of Ξ_k on $\mathcal{E}(\mathcal{S}^+)$ given data up to and including time t_k is obtained. The posterior is not a PPP, as is seen below from the form of its pdf in (D.10). Secondly, the posterior point process is approximated by a PPP, and a low computational complexity expression for the intensity of the approximating PPP is obtained. The two branches of detection thinning are recombined by superposition to obtain the intensity filter update.

D.1 Posterior Point Process

The random variables $\Xi_{k-1|k-1}$, $\Xi_{k|k-1}$, and $\Upsilon_{k|k-1}$ are defined as in Section C. The state space of $\Xi_{k-1|k-1}$ and $\Xi_{k|k-1}$ is $\mathcal{E}(\mathcal{S}^+)$, where $\mathcal{E}(\mathcal{S}^+)$ is a union of sets defined as in (2.1). Similarly, the event space of $\Upsilon_{k|k-1}$ is $\mathcal{E}(\mathcal{T})$, not $\mathcal{T}$.

The process $\Xi_{k-1|k-1}$ is assumed to be a PPP, so it is parameterized by its intensity $f_{k-1|k-1}(s)$, $s \in \mathcal{S}^+$. A realization $\xi_k \in \mathcal{E}(\mathcal{S}^+)$ of $\Xi_{k-1|k-1}$ is thinned by a death probability function $d_{k-1}(s)$, assumed known, and subsequently diffused to time t_k via the single target transition function $\Psi_{k-1}(y|x)$. Thinning

and diffusing $\Xi_{k-1|k-1}$ yields the predicted target PPP $\Xi_{k|k-1}$. Its intensity is, using (2.83),

$$f_{k|k-1}(x) = \int_{\mathcal{S}^+} \Psi_{k-1}(x|s)\,(1 - d_{k-1}(s))\, f_{k-1|k-1}(s)\,\mathrm{d}s. \tag{D.1}$$

The integral in (D.1) is defined as in (2.97).

The point process $\Xi_{k|k}$ is the sum of detected and undetected target processes, denoted by $\Xi^D_{k|k}$ and $\Xi^U_{k|k}$, respectively. They are obtained from the same realizations of $\Xi_{k|k-1}$, so they would seem to be highly correlated. However, the number of points in the realization is Poisson distributed, so they are actually independent. See Section 2.9.

The undetected target process $\Xi^U_{k|k}$ is the predicted target PPP $\Xi_{k|k-1}$ thinned by $1 - P^D_k(s)$, where $P^D_k(s)$ is the probability of detecting a target at s. Thus $\Xi^U_{k|k}$ is a PPP, and

$$f^U_{k|k}(x) = (1 - P^D_k(x))\, f_{k|k-1}(x) \tag{D.2}$$

is its intensity.

The detected target process $\Xi^D_{k|k}$ is the predicted target PPP $\Xi_{k|k-1}$ that is thinned by $P^D_k(s)$ and subsequently updated by Bayesian filtering. Thinning yields the predicted PPP $\Xi^D_{k|k-1}$, and

$$f^D_{k|k-1}(x) = P^D_k(x)\, f_{k|k-1}(x) \tag{D.3}$$

is its intensity.

The predicted measurement process $\Upsilon_{k|k-1}$ is obtained from $\Xi^D_{k|k-1}$ via the pdf of a single point measurement $z \in \mathcal{T}$ conditioned on a target located at $s \in \mathcal{S}^+$. The quantity $p_k(z|\phi)$ is the likelihood of z if it is a false alarm. See Section 2.12. Thus, $\Upsilon_{k|k-1}$ is a PPP on $\mathcal{T}$ and

$$\lambda_{k|k-1}(z) = \int_{\mathcal{S}^+} p_k(z|s)\, P^D_k(s)\, f_{k|k-1}(s)\,\mathrm{d}s, \tag{D.4}$$

is its intensity.

The measurement set is $\upsilon_k = \{m, \{z_1, \ldots, z_m\}$, where $z_j \in \mathcal{T}$. The conditional pdf of υ_k is defined for arbitrary target realizations $\xi_k = (n, \{x_1, \ldots, x_n\}) \in \mathcal{E}(\mathcal{S}^+)$. All the points x_j of ξ_k, whether they are a true target ($x_j \in \mathbb{R}^{n_x}$) or are clutter ($x_j = \phi$), generate a measurement so that only when $m = n$ is the measurement likelihood non-zero. The correct assignment of point measurements to targets in ξ_k is unknown. All such assignments are equally probable, so the pdf averages over all possible assignments of data to false alarms and targets. Because ϕ is a target state, the measurement pdf is

$$p_{\Upsilon_k|\Xi_k}(\upsilon_k|\xi_k) = \begin{cases} \frac{1}{m!}\sum_{\sigma\in\mathrm{Sym}(m)} \prod_{j=1}^{m} p_k(z_{\sigma(j)}|x_j), & m = n \\ 0, & m \neq n, \end{cases} \tag{D.5}$$

where $\mathrm{Sym}(m)$ is the set of all permutations on the integers $\{1, 2, \ldots, m\}$.

The lower branch of (D.5) is a consequence of the "at most one measurement per target"rule together with the augmented target state space $\mathcal{S}^+$. To elaborate, the points in a realization ξ of the *detected* target PPP are targets, some of which have state ϕ. The augmented state space accommodates clutter measurements by using targets in ϕ, so only realizations with $m = n$ points have nonzero probability.

The posterior pdf of $\Xi^D_{k|k}$ on $\mathcal{E}(\mathcal{S}^+)$ is, from (C.5),

$$p_{k|k}(\xi_k) = p_{\Upsilon_k|\Xi_k}(\upsilon_k|\xi_k)\frac{p_{k|k-1}(\xi_k)}{\pi_{k|k-1}(\upsilon_k)}. \tag{D.6}$$

The pdf's $p_{k|k-1}(\xi_k)$ and $\pi_{k|k-1}(\upsilon_k)$ of $\Xi^D_{k|k-1}$ and $\Upsilon_{k|k-1}$ are given in terms of their intensity functions using (2.12):

$$p_{k|k-1}(\xi_k) = \frac{1}{m!}\exp\left(-\int_{\mathcal{S}^+} f^D_{k|k-1}(s)\,\mathrm{d}s\right)\prod_{j=1}^{m} f^D_{k|k-1}(x_j) \tag{D.7}$$

$$\pi_{k|k-1}(\upsilon_k) = \frac{1}{m!}\exp\left(-\int_{\mathcal{T}} \lambda_{k|k-1}(z)\,\mathrm{d}z\right)\prod_{j=1}^{m} \lambda_{k|k-1}(z_j). \tag{D.8}$$

From (D.3) and (D.4),

$$\int_{\mathcal{S}^+} f^D_{k|k-1}(s)\,\mathrm{d}s = \int_{\mathcal{T}} \lambda_{k|k-1}(z)\,\mathrm{d}z. \tag{D.9}$$

Substituting (D.7), (D.8), and (D.5) into (D.6) and using obvious properties of permutations gives the posterior pdf of $\Xi^D_{k|k}$:

$$p_{k|k}(\xi_k) = \frac{1}{m!}\sum_{\sigma\in\mathrm{Sym}(m)}\prod_{j=1}^{m}\frac{p_k(z_{\sigma(j)}|x_j)\,P^D_k(x_j)\,f_{k|k-1}(x_j)}{\lambda_{k|k-1}(z_{\sigma(j)})}. \tag{D.10}$$

If ξ_k does not contain exactly m points, then $p_{k|k}(\xi_k) = 0$. Conditioning $\Xi^D_{k|k}$ on m points gives the pdf of the points of the posterior process as

$$p_{k|k}(x_1, \ldots, x_m) = \frac{1}{m!}\sum_{\sigma\in\mathrm{Sym}(m)}\prod_{j=1}^{m}\frac{p_k(z_{\sigma(j)}|x_j)\,P^D_k(x_j)\,f_{k|k-1}(x_j)}{\lambda_{k|k-1}(z_{\sigma(j)})}. \tag{D.11}$$

The pdf (D.11) holds for $x_j \in \mathcal{S}^+$, $j = 1, \ldots, m$.

D.2 PPP Approximation

The pdf of the posterior point process $\Xi^D_{k|k}$ is clearly not that of a PPP. This causes a problem for the recursion. One way around it is to approximate $\Xi^D_{k|k}$ by a PPP and recursively update the intensity of the PPP approximation.

The pdf $p_{k|k}(x_1, \ldots, x_m) = p_{k|k}(x_{\sigma(1)}, \ldots, x_{\sigma(m)})$ for all $\sigma \in \mathrm{Sym}(m)$; therefore, integrating it over all of arguments except, say, the ℓ^{th} argument gives the same result regardless of the choice of ℓ. The form of the "single target marginal" is, using (D.4),

$$\begin{aligned} p_{k|k}(x_\ell) &\equiv \int_{\mathcal{S}^+} \cdots \int_{\mathcal{S}^+} p_{k|k}(x_1, \ldots, x_m) \prod_{\substack{i=1 \\ i \neq \ell}}^{m} \mathrm{d}x_i \\ &= \frac{1}{m!} \sum_{\sigma \in \mathrm{Sym}(m)} \int_{(\mathcal{S}^+)^{m-1}} \prod_{j=1}^{m} \frac{p_k(z_{\sigma(j)}|x_j) P^D_k(x_j) f_{k|k-1}(x_j)}{\lambda_{k|k-1}(z_{\sigma(j)})} \prod_{\substack{i=1 \\ i \neq \ell}}^{m} \mathrm{d}x_i \\ &= \frac{1}{m!} \sum_{r=1}^{m} \sum_{\substack{\sigma \in \mathrm{Sym}(m) \\ \text{and } \sigma(\ell)=r}} \frac{p_k(z_{\sigma(\ell)}|x_\ell) P^D_k(x_\ell) f_{k|k-1}(x_\ell)}{\lambda_{k|k-1}(z_{\sigma(j)})} \\ &= \frac{1}{m} \sum_{r=1}^{m} \frac{p_k(z_r|x_\ell) P^D_k(x_\ell) f_{k|k-1}(x_\ell)}{\lambda_{k|k-1}(z_r)}. \end{aligned} \tag{D.12}$$

This identity holds for arbitrary $x_\ell \in \mathcal{S}^+$.

The joint conditional pdf is approximated by the product of its marginal pdf's:

$$p_{k|k}(x_1, \ldots, x_m) \approx \prod_{j=1}^{m} p_{k|k}(x_j). \tag{D.13}$$

The product approximation is called a mean field approximation in the machine learning community [50, pp. 35–36]. Both sides of (D.13) integrate to one.

The marginal pdf is proportional to the intensity of the approximating PPP. Let $f^D_{k|k}(x) = c\, p_{k|k}(x)$ be the intensity. The likelihood function of the unknown constant c is

$$\mathscr{L}(c|\xi_k) = \frac{1}{m!} e^{-\int_{\mathcal{S}^+} c\, p_{k|k}(s)\, ds} \prod_{j=1}^{m} \left(c\, p_{k|k}(x_j) \right) \propto e^{-c} c^m.$$

The maximum likelihood estimate is $\hat{c}_{ML} = m$, so that

$$f^D_{k|k}(x) = \sum_{r=1}^{m} \frac{p_k(z_r|x)\, P^D_k(x)\, f_{k|k-1}(x)}{\lambda_{k|k-1}(z_r)} \tag{D.14}$$

is the intensity of the approximating PPP.

Altogether Now

The PPP approximation to the point process $\Xi_{k|k}$ is the sum of the undetected target PPP $\Xi^U_{k|k}$ and the PPP that approximates the detected target process $\Xi^D_{k|k}$. Hence,

$$\begin{aligned} f_{k|k}(x) &= f^U_{k|k}(x) + f^D_{k|k}(x), \qquad x \in \mathcal{S}^+ \\ &= \left[1 - P^D_k(x) + \sum_{r=1}^{m} \frac{p_k(z_r|x)\, P^D_k(x)}{\lambda_{k|k-1}(z_r)}\right] f_{k|k-1}(x) \end{aligned} \tag{D.15}$$

is the updated intensity of the PPP approximation to $\Xi_{k|k}$.

The intensity filter comprises equations (D.1), (D.4), and (D.15). The first two equations are more insightful when written in traditional notation. From (D.1),

$$f_{k|k-1}(x) = \hat{b}_k(x) + \int_{\mathcal{S}} \Psi_{k-1}(x|s)\,(1 - d_{k-1}(s))\, f_{k-1|k-1}(s)\, \mathrm{d}s, \tag{D.16}$$

where the predicted target birth intensity is

$$\hat{b}_k(x) = \Psi_{k-1}(x|\phi)\,(1 - d_{k-1}(\phi))\, f_{k-1|k-1}(\phi). \tag{D.17}$$

Also, from (D.4),

$$\lambda_{k|k-1}(z) = \hat{\lambda}_k(z) + \int_{\mathcal{S}} p_k(z|s)\, P^D_k(s)\, f_{k|k-1}(s)\, \mathrm{d}s, \tag{D.18}$$

where

$$\hat{\lambda}_k(z) = p_k(z|\phi)\, P^D_k(\phi)\, f_{k|k-1}(\phi) \tag{D.19}$$

is the predicted measurement clutter intensity.

The above derivation of the intensity and PHD filters was first given in [117]. A more intuitive "physical space" approach is given by [29]. An analogous derivation for multisensor multitarget intensity filter is given in [111].

D.3 First Moment Intensity and Janossy Densities

An alternative method is often used in the literature to obtain the intensity function (D.14) for the detected target posterior point process. The process $\Xi^D_{k|k}$ is not a PPP, but it is a finite point process because its realizations contain exactly m points. The general machinery of finite point processes is thus applicable to $\Xi^D_{k|k}$. An excellent reference for general finite point process theory is [16].

Let $\xi = (N, \mathcal{X}|N)$ denote a realization of $\Xi^D_{k|k}$, where N is the number of points and $\mathcal{X}|N$ is the point set. From [16, Sect. 5.3], the Janossy probability density of a finite point process is defined by

$$j_n(x_1, \ldots, x_n) = p_N(n) p_{\mathcal{X}|N}(\{x_1, \ldots, x_n\}|n) \quad \text{for } n = 0, 1, 2, \ldots . \tag{D.20}$$

Janossy densities were encountered (but left named) early in Chapter 2, Eqn. (2.10). Using the standard argument list as in (2.13) gives

$$j_n(x_1, \ldots, x_n) = n! p_N(n) p_{\mathcal{X}|N}(x_1, \ldots, x_n|n) \quad \text{for } n = 0, 1, 2, \ldots . \tag{D.21}$$

Intuitively, from [16, p. 125],

$$j_n(x_1, \ldots, x_n) = \Pr\left[\begin{array}{c}\text{Exactly } n \text{ points in a realization} \\ \text{with one point in each infinitesimal} \\ [x_i + \mathrm{d}x_i),\ i = 1, \ldots, n\end{array}\right]. \tag{D.22}$$

Now, for the finite point process $\Xi^D_{k|k}$, $p_N(m) = 1$ and $p_N(n) = 0$ if $n \neq m$, so only one of the Janossy functions is nonzero. The Janossy densities are

$$j_n(x_1, \ldots, x_n) = \begin{cases} m! p_{k|k}(x_1, \ldots, x_m), & \text{if } n = m, \\ 0, & \text{if } n \neq m, \end{cases} \tag{D.23}$$

where $p_{k|k}(x_1, \ldots, x_m)$ is the posterior pdf given by (D.11). The first moment intensity is denoted in [16] by $m_1(x)$. From [16, Lemma 5.4.III], it is given in terms of the Janossy density functions by

$$m_1(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{S}^+} \cdots \int_{\mathcal{S}^+} j_{n+1}(x, x_1, \ldots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n . \tag{D.24}$$

From (D.23), only the term $n = m - 1$ is nonzero, so that

$$m_1(x) = \frac{1}{(m-1)!} \int_{\mathcal{S}^+} \cdots \int_{\mathcal{S}^+} m! p_{k|k}(x, x_1, \ldots, x_{m-1}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{m-1} . \tag{D.25}$$

The integral (D.25) is exactly m times the integral in Eqn. (D.12), so the first moment approximation to $\Xi^D_{k|k}$ is identical to the intensity (D.14).