Probabilistic Data Association

SCJ Robertson

11th September, 2016

Introduction

A quick look at Probabilistic Data Association Filter (PDAF). Joint PDAF (JPDAF) and Multiple Hypothesis Tracking (MHT) will not be done at any point[1].

1. The Probabilistic Data Association Filter

The PDAF can be viewed as an attempt to constrain Multiple Hypothesis Tracking (MHT) to a single target in a cluttered environment. Firstly, it only seeks plausible measurements within a gated region around a target's predicted measurement, limiting the amount of possible hypotheses. As with MHT, the measurement update will result in a mixture of the hypotheses, the PDAF simply approximates the mixture with a single Gaussian with matching moments.

To track multiple objects, each target is simply assigned its own independent filter. Due to the rough averaging process, the PDAF will to struggle to disassociate multiple targets with neighbouring trajectories and thus provided motivation for the more global minded Joint PDAF.

Figure 1.1 provides a Bayes Net representation of the filter and Figure 1.2 its cluster graph.

1.1 Probabilistic Graphical Model Development

Approach

As mentioned before, the PDAF attempts to constrain the more general MHT filter for a single target in a cluttered environment. Only a single target's track is to be investigated, no attempt will be made to initialise any new target's track. This section provides a brief outline of the PDAF's approach, before pursuing a more formal derivation.

For a single existing target's track, Figure 1.1, in steady-state:

[1]This was meant form part of an extended literature study, but as it turns out most of these filters are incredibly similar and a firm understanding of one covers a lot of ground.

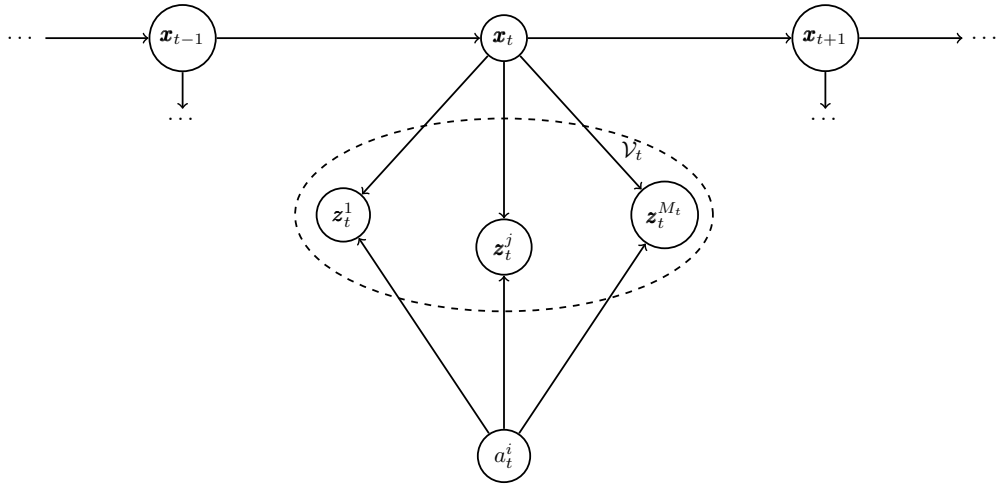

Figure 1.1: The Bayes Net representing the PDAF for a single target, $\mathcal{V}_t$ is the validation gate containing M_t measurements. The target is capable of generating at most a single measurement, the remaining measurements are assumed to be clutter.

1. The state of $\boldsymbol{x}_t$ is determined through a typical Markovian transition process.
2. The predicted measurement, $\boldsymbol{z}_t$, is determined through marginalisation of the newly constructed measurement distribution.
3. $\boldsymbol{z}_t$'s covariance is used to create a validation region in which M_t-many measurements are detected.
4. It is assumed that the target is capable of generating at most a single measurement, the remaining detections are assumed to be clutter. A latent association variable, a_t^i, with some known prior, describes the probability of a detection's origin.
5. a_t^i and $\boldsymbol{z}_t$ are used to create likelihood functions for each detection $\boldsymbol{z}_t^j$ in the validation region. As a_t^i is never directly observed, a multi-model likelihood distribution of possible hypotheses is returned and used in the measurement update.
6. Pursuing all possible hypotheses would lead to exponential explosion, therefore $\boldsymbol{x}_t$'s mixture is approximated with a single moment-matched Gaussian. The approximate state, $\overline{\boldsymbol{x}}_t$, is then propagated forward to next time step.

Motivation

The usual approach to deriving the PDAF is to assume the $\boldsymbol{x}_t$ as been updated with the multi-model measurement likelihood, and is now itself a mixture about to be reduced [1–3]. The first moment of the mixture is matched by taking a weighted sum of its component's expected

values:

$$\begin{aligned}
\overline{\boldsymbol{x}}_t &= \mathbb{E}[\boldsymbol{x}_t|Z_t] \\
&= \int \boldsymbol{x}_t p(\boldsymbol{x}_t|Z_t)d\boldsymbol{x}_t = \int \boldsymbol{x}_t \sum_{i=0}^{M_t} p(\boldsymbol{x}_t, a_t^i|Z_t)d\boldsymbol{x}_t \\
&= \int \boldsymbol{x}_t \sum_{i=0}^{M_t} p(\boldsymbol{x}_t|a_t^i, Z_t)p(a_t^i|Z_t)d\boldsymbol{x}_t \\
&= \sum_{i=0}^{M_t} p(a_t^i|Z_t) \int \boldsymbol{x}_t p(\boldsymbol{x}_t|a_t^i, Z_t)d\boldsymbol{x}_t \\
&= \sum_{i=0}^{M_t} p(a_t^i|Z_t)\mathbb{E}[\boldsymbol{x}_t|a_t^i, Z_t] = \sum_{i=0}^{M_t} \mathcal{W}_t^i \overline{\boldsymbol{x}}_t^i
\end{aligned} \tag{1.1}$$

Here $\overline{\boldsymbol{x}}_t^i$ is the expected value of the i^{th} mixture component. $\mathcal{W}_t^i$ is the weight of i^{th} component determined by the association variable, a_t^i, defined as follows:

$$\begin{aligned}
\mathcal{W}_t^i &\triangleq p(a_t^i|Z_t) \\
&= p(a_t^i|\mathcal{Z}_t, M_t, Z_{t-1})
\end{aligned} \tag{1.2}$$

$\mathcal{Z}_t = \{\boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^{M_t}\}$ are all new independent measurements introduced at time step t, $Z_t = \mathcal{Z}_t \cup Z_{t-1}$ is the cumulative set of all measurements up to and including time t. The association structure of the Bayes Net is determined by:

$$\begin{aligned}
\mathcal{W}_t^i &= \frac{p(\mathcal{Z}_t|a_t^i, M_t, Z_{t-1})p(a_t^i|M_t, Z_{t-1})}{\sum_{j=0} p(\boldsymbol{z}_t^j|a_t^i, M_t, Z_{t-1})} \\
&= \frac{\prod_{j=0}^{M_t} p(\boldsymbol{z}_i^t|a_t^i, M_t, Z_{t-1})p(a_t^i|M_t, Z_{t-1})}{\sum_{j=0}^{M_t} p(\boldsymbol{z}_t^j|a_t^i, M_t, Z_{t-1})} \\
&= \frac{\prod_{j=0}^{M_t} p(\boldsymbol{z}_i^t|a_t^i, M_t, \boldsymbol{z}_t)p(a_t^i)}{\sum_{j=0}^{M_t} p(\boldsymbol{z}_t^j|a_t^i, \boldsymbol{z}_t)p(a_t^i)}
\end{aligned} \tag{1.3}$$

The details of the required assumptions necessary to arrive at this result form the remainder of this document. (1.3) already asserts the measurement independence assumption mentioned earlier.

1.1.1 Measurement Validation

As shown in Figure 1.1, the filter is mostly Kalman. The state transition, after averaging, is typical and the predicted state is subsequently used to create the joint measurement distribution over $\boldsymbol{x}_t$ and $\boldsymbol{z}_t^j$:

$$p(\boldsymbol{x}_t, \boldsymbol{z}_t^{j=1}) = \mathcal{C}(\boldsymbol{x}_t, \boldsymbol{z}_t^j; \mathcal{P}, \boldsymbol{h}, g_{\boldsymbol{z}}) \tag{1.4}$$

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_{\boldsymbol{x},\boldsymbol{x}} & \mathcal{P}_{\boldsymbol{x},\boldsymbol{z}} \\ \mathcal{P}_{\boldsymbol{z},\boldsymbol{x}} & \mathcal{P}_{\boldsymbol{z},\boldsymbol{z}} \end{bmatrix} \tag{1.5}$$

$$\boldsymbol{h} = \mathcal{P} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{x}} \\ \overline{\boldsymbol{z}}_{\boldsymbol{t}} \end{bmatrix} \tag{1.6}$$

$$g_{\boldsymbol{z}} = -\frac{1}{2}\boldsymbol{h}_t^T \mathcal{P}^{-1}\boldsymbol{h}_t - \frac{1}{2}\ln|2\pi\mathcal{P}^{-1}| \tag{1.7}$$

This is simply a reintroduction to the canonical form notation used in [6] as it will be used during subsequent analysis. The predicted measurement is found by marginalising out the state $\boldsymbol{x}_t$:

$$\begin{aligned} p(\boldsymbol{z}_t^j) &= \int \mathcal{C}(\boldsymbol{x}_t, \boldsymbol{z}_t^j; \mathcal{P}, \boldsymbol{h}, g) d\boldsymbol{x}_t \\ &= \mathcal{C}(\boldsymbol{z}_t^j; \mathcal{S}_t^{-1}, \mathcal{S}_t^{-1}\overline{\boldsymbol{z}}_t, g_{\boldsymbol{z}}) \end{aligned} \tag{1.8}$$

This predicted measurement distribution will be used to determine possible measurement which could be target generated. It is assumed that all information about the past measurements, Z_{t-1}, will be contained contained within the prediction thus shifting any dependencies on Z_{t-1} to $\boldsymbol{x}_t$.

As shown in Figure 1.1, the validation region is the elliptical region:

$$\mathcal{V}_t(\gamma) = \{\boldsymbol{z} : [\boldsymbol{z} - \overline{\boldsymbol{z}}_t]^T \mathcal{S}_t^{-1} [\boldsymbol{z} - \overline{\boldsymbol{z}}_t] \leq \gamma\} \tag{1.9}$$

This is the region where the correct measurement is expected to arise, based on the predicted measurement $\boldsymbol{z}_t$'s covariance. This is Mahalanobis distance, measuring the distance of a point $\boldsymbol{z}_t^j$ from the measurement distribution, if the point is too far away it assumed not be generated by the distribution. γ is some appropriate gate threshold and directly determines the gate probability:

$$P_G = \int_{\mathcal{V}_t} \mathcal{C}(\boldsymbol{z}_t; \mathcal{S}_t^{-1}, \mathcal{S}_t^{-1}\overline{\boldsymbol{z}}_t, g_{\boldsymbol{z}}) d\boldsymbol{z}_t \tag{1.10}$$

P_G is the probability that the correct measurement is within the validation region, if detected. The detection probability, P_D, is the probability that the correct measurement is detected somewhere in the sensor's field-of-view. The volume of the validation region is of course determined by innovation covariance, $\mathcal{S}_t$:

$$\begin{aligned} V_t &= C_k |\gamma \mathcal{S}_t|^{\frac{1}{2}} \\ &= C_k \gamma^{\frac{k}{2}} |\mathcal{S}_t|^{\frac{1}{2}} \end{aligned} \tag{1.11}$$

C_n is the volume of a hyper-sphere in the k-dimensional measurement space.

It is assumed that searching the validation region results in a neatly arranged set of M_t-many detected measurements:

$$\mathcal{Z}_t = \{\boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^j, \ldots, \boldsymbol{z}_t^{M_t}\} \tag{1.12}$$

Each of the measurements in $\mathcal{Z}_t$ are allowed to be described by the PDF:

$$p(\boldsymbol{z}_t^j | a_t^i, M_t) = \begin{cases} \frac{1}{V_t} & i \neq j \\ \frac{1}{P_G} \mathcal{C}(\boldsymbol{z}_t^j; \mathcal{S}_t^{-1}, \mathcal{S}_t^{-1}\overline{\boldsymbol{z}}_t, g_{\boldsymbol{z}}) & i = j \end{cases} \tag{1.13}$$

If a measurement is deemed incorrect, $i \neq j$, it is assumed to be clutter, ignorance is accepted, and a uniform distribution over the validation region is returned. Should it be target generated, the measurement distribution is assigned a truncated Gaussian. It should be noted that (1.13) is the marginal of the joint conditional distribution:

$$p(\boldsymbol{x}_t, \boldsymbol{z}_t^j | a_t^i, M_t) = \begin{cases} \frac{1}{V_t} \mathcal{C}(\boldsymbol{x}_t; \Sigma_t^{-1}, \Sigma_t^{-1}\boldsymbol{\mu}_{\boldsymbol{x}}, g_{\boldsymbol{x}}) & i \neq j \\ \frac{1}{P_G} \mathcal{C}(\boldsymbol{x}_t, \boldsymbol{z}_t^j; \mathcal{P}, \boldsymbol{h}, g_{\boldsymbol{z}}) & i = j \end{cases} \tag{1.14}$$

The previous equations help motivate this approach, (1.14) is the actual joint distribution which will be created for every detected measurement during calibration.

1.1.2 Data Association

First, a formal definition of the data association variable operation is required:

$$a_t^i = \begin{cases} \textit{No target originated measurement} & i = 0 \\ \boldsymbol{z}_t^i \textit{ caused by the target} & i = 1, \ldots, M_t \end{cases} \tag{1.15}$$

If $i = 0$ it is assumed all of the detections are clutter, the correct measurement was either missed or outside of the validation region. $i \neq 0$ assumes only a single detection $\boldsymbol{z}_t^i$ is correct and the remaining $M_t - 1$ detections are clutter. The association variable must be dependent on the number of incorrect measurements, $I = n$, found within the validation region:

$$\begin{aligned} p(a_t^i | I = n, \boldsymbol{z}_t) &= p(a_t^i | I = n) \\ &= \frac{p(n|a_t^i)p(a_t^i)}{\sum_{j=0}^{M_t} p(n, a_t^j)} \\ &= \frac{p(n|a_t^i)p(a_t^i)}{\sum_{j=0}^{M_t} p(n|a_t^j)p(a_t^j)} \end{aligned} \tag{1.16}$$

The distribution is only dependent on the predicated measurement in the sense that detected measurements are only sought in the validation region. It should be legal to simply wish it away.

(1.16) requires a prior on the association variable, it natural to assume it should describe the probability of the single correct measurement being detected within the validation region. A painfully formal way of stating this would be,

$$\begin{aligned} &A = \textit{Correct measurement is within gate.} \\ &B = \textit{Correct measurement is detected.} \\ &P(A, B) = P(A)P(B) = P_G P_D \end{aligned} \tag{1.17}$$

using the gate and detection probabilities defined in the preceding section. This allows the prior on a_t^i to be defined as:

$$p(a_t^i) = \begin{cases} 1 - P_G P_D & i = 0 \\ \frac{1}{M_t} P_G P_D & i = 1, \ldots, M_t \end{cases} \tag{1.18}$$

If no measurement is correct, the correct measurement is assumed to be undetected or outside validation region.

(1.16) also requires some distribution, $\mu(I = M_t)$, capable of describing the probability of incorrect measurements distributed within the validation region, given M_t detected measurements. Only two scenarios are of interest, the case where all the measurements are incorrect and where all but a single measurement are incorrect. Therefore:

$$p(M_t | a_t^i) = \begin{cases} \mu(M_t) & \text{i} = 0 \\ \mu(M_t - 1) & \text{i} = 1, \ldots, M_t \end{cases} \tag{1.19}$$

In the case of no correct measurements, (1.16) can now be rearranged as follows:

$$\begin{aligned} p(a_t^0 | I = M_t) &= \frac{p(M_t|a_t^0)p(a_t^0)}{\sum_{j=1}^{M_t} p(M_t|a_t^j)p(a_t^j)} \\ &= \frac{\frac{1}{M_t}(1 - P_G P_D)\mu(M_t)}{(1 - P_G P_D)\mu(M_t) + M_t(\frac{1}{M_t} P_G P_D)\mu(M_t - 1)} \\ &= \frac{(1 - P_G P_D)\frac{\mu(M_t)}{\mu(M_t - 1)}}{P_G P_D + (1 - P_G P_D)\frac{\mu(M_t)}{\mu(M_t - 1)}} \end{aligned} \tag{1.20}$$

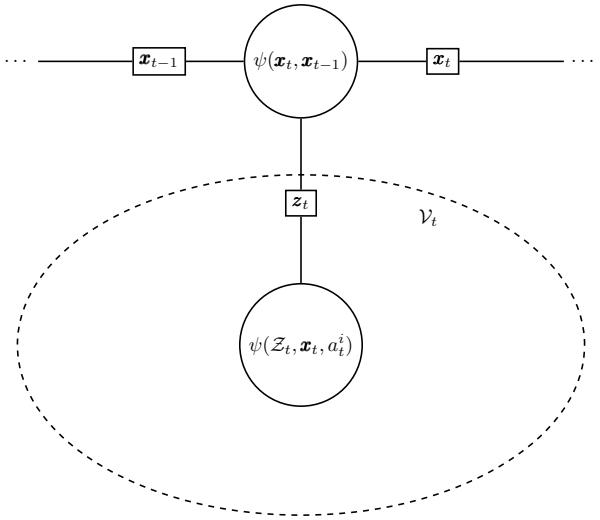

Figure 1.2: The clique tree representing the PDAF, derived from Figure 1.1. The assignment of the distributions to their cluster is handled in Section 1.2.1.

In the general case:

$$\begin{aligned} p(a_t^i|I = M_t - 1) &= \frac{p(M_t - 1|a_t^i)p(a_t^i)}{\sum_{j=1}^{M_t} p(M_t|a_t^j)p(a_t^j)} \\ &= \frac{\frac{1}{M_t}(P_G P_D)\mu(M_t - 1)}{(1 - P_G P_D)\mu(M_t) + M_t(\frac{1}{M_t}P_G P_D)\mu(M_t - 1)} \\ &= \frac{\frac{1}{M_t}(P_G P_D)\mu(M_t - 1)}{P_G P_D + (1 - P_G P_D)\frac{\mu(M_t)}{\mu(M_t - 1)}} \end{aligned} \tag{1.21}$$

Two models are usually considered for $\mu(n)$, a Poisson or a Diffuse Prior model [2, 3]. The Poisson model proved more popular, therefore:

$$\mu(I = n) = \mu(n) = e^{-\lambda V_t}\frac{(\lambda V_t)^n}{n!} \tag{1.22}$$

Here λ is the spacial density, the average rate of clutter per unit volume. In this case the distribution is quite manageable:

$$\frac{\mu(M_t)}{\mu(M_t - 1)} = \frac{\lambda V_t}{M_t} \tag{1.23}$$

The distribution over a_t^i is now well defined for all associations. In the graphical model, Figure 1.1, the distribution's dependants are incorporated into $p(a_t^i|I = n)$, therefore:

$$\therefore p(a_t^i) = p(a_t^i|I = M_t) = \begin{cases} \frac{(1-P_G P_D)\lambda V_t}{P_G P_D M_t + (1-P_G P_D)\lambda V_t} & i = 0 \\ \frac{P_G P_D}{P_G P_D M_t + (1-P_G P_D)\lambda V_t} & i = 1, \ldots, M_t \end{cases} \tag{1.24}$$

1.1.3 Cluster Graph Representation

The Bayes Net of Figure 1.1 now has well defined distributions and can now be more usefully represented as the cluster graph in Figure 1.3. Since this is a derivation, exact inference is required and the PGM will have to be a clique tree. The compact representation of Figure 1.2 will be used for the remainder of this derivation.

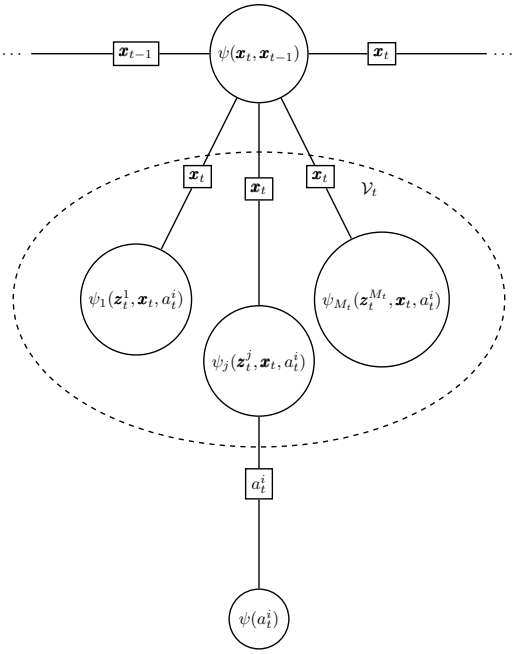

Figure 1.3: A clique tree for the Bayes Net in Figure 1.1. This is a more distributed form of Figure 1.2's tree.

1.2 Analysis

This section is fairly Spartan, just proving that the PGM is equivalent to [2, 3]. It assumes the PDAF operates on some type of Kalman Filter, but the exact details of the filter will be ignored.

1.2.1 Cluster Initialisation

It is to be assumed that that the graph in Figure 1.2 has been calibrated. The distribution construction process is identical to that of the Kalman filter and will be ignored. The measurement clique $\psi(\mathcal{Z}_t, \boldsymbol{x}_t, a_t^i)$ now holds the following belief:

$$\begin{aligned}\beta(\mathcal{Z}_t, \boldsymbol{x}_t, a_t^i) &= p(a_t^i) \prod_{j=1}^{M_t} p(\boldsymbol{x}_t, \boldsymbol{z}_t^j | a_t^i, M_t) \\ &= \begin{cases} \frac{(1-P_G P_D)\lambda}{V_t^{M_t-1}[P_G P_D M_t + (1-P_G P_D)\lambda V_t]} \mathcal{C}(\boldsymbol{x}_t; \Sigma_t^{-1}, \Sigma_t^{-1}\boldsymbol{\mu}_{\boldsymbol{x}}, g_{\boldsymbol{x}}) & i = 0 \\ \frac{P_D}{V_t^{M_t-1}[P_G P_D M_t + (1-P_G P_D)\lambda V_t]} \mathcal{C}(\boldsymbol{x}_t, \boldsymbol{z}_t^i; \mathcal{P}, \boldsymbol{h}, g_{\boldsymbol{z}}) & i = 1, \ldots, M_t \end{cases} \end{aligned} \tag{1.25}$$

After the introduction of evidence, the belief will be normalized and will be equivalent to the desired weights in (1.3).

1.2.2 Measurement update

In the BUP framework there will be in actually be several small interchanges of information during measurement update before the information can be propagated to the next state. These

pointless exchanges can easily be avoided as follows:

$$\begin{aligned}\delta_{\boldsymbol{x}_t\to\boldsymbol{x}_{t+1}}(\boldsymbol{x}_t) &= \int \beta(\boldsymbol{x}_t,\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1}\\ &= \int \psi(\boldsymbol{x}_t,\boldsymbol{x}_{t-1})\delta_{\boldsymbol{z}_t\to\boldsymbol{x}_t}(\boldsymbol{x}_t)d\boldsymbol{x}_{t-1}\\ &= \delta_{\boldsymbol{z}_t\to\boldsymbol{x}_t}(\boldsymbol{x}_t)\int \psi(\boldsymbol{x}_t,\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1}\\ &= \delta_{\boldsymbol{z}_t\to\boldsymbol{x}_t}(\boldsymbol{x}_t)\delta_{\boldsymbol{x}_t\to\boldsymbol{z}_t}(\boldsymbol{x}_t) \end{aligned} \tag{1.26}$$

Here $\delta_{\boldsymbol{x}_t\to\boldsymbol{z}_t}(\boldsymbol{x}_t)$ is the initial message passed to the measurement clique during calibration, therefore the message passed back from the measurement clique must divide it out:

$$\delta_{\boldsymbol{z}_t\to\boldsymbol{x}_t}(\boldsymbol{x}_t) = \frac{1}{\delta_{\boldsymbol{x}_t\to\boldsymbol{z}_t}(\boldsymbol{x}_t)}\int \beta(\mathcal{Z}_t=\hat{\mathcal{Z}}_t,\boldsymbol{x}_t,a_t^i)da_t^i \tag{1.27}$$

$$\therefore \delta_{\boldsymbol{x}_t\to\boldsymbol{x}_{t+1}}(\boldsymbol{x}_t) = \int \beta(\mathcal{Z}_t=\hat{\mathcal{Z}}_t,\boldsymbol{x}_t,a_t^i)da_t^i \tag{1.28}$$

Therefore the outgoing message, before it is approximated, is directly from the measurement clique. The usual approach is to introduce evidence and then normalize the resulting distribution [2,3]. The normalising constant is determined by introducing the evidence then marginalising out $\boldsymbol{x}_t$ and a_t^i. The evidence, $\boldsymbol{z}_t^j = \hat{\boldsymbol{z}}_t^j$, is introduce as follows:

$$\begin{aligned}\beta(\mathcal{Z}_t=\hat{\mathcal{Z}}_t,\boldsymbol{x}_t,a_t^i) &= p(a_t^i)\prod_{j=1}^{M_t} p(\boldsymbol{x}_t,\boldsymbol{z}_t^j=\hat{\boldsymbol{z}}_t^j|a_t^i,M_t)\\ &= \begin{cases}\frac{(1-P_GP_D)\lambda}{V_t^{M_t-1}[P_GP_DM_t+(1-P_GP_D)\lambda V_t]}\mathcal{C}(\boldsymbol{x}_t;\Sigma_t^{-1},\Sigma_t^{-1}\boldsymbol{\mu}_{\boldsymbol{x}},g_{\boldsymbol{x}}) & i=0\\ \frac{P_D}{V_t^{M_t-1}[P_GP_DM_t+(1-P_GP_D)\lambda V_t]}\mathcal{C}(\boldsymbol{x}_t;\hat{\mathcal{S}}_t,\hat{\boldsymbol{h}}_t^i,\hat{g}_t^i) & i=1,\dots,M_t\end{cases}\end{aligned} \tag{1.29}$$

$\boldsymbol{x}_t$ and a_t^i are then marginalised out of (1.30), giving the normalising constant C:

$$\begin{aligned}p(a_t^i|\mathcal{Z}_t=\hat{\mathcal{Z}}_t) &= \int \beta(\mathcal{Z}_t=\hat{\mathcal{Z}}_t,\boldsymbol{x}_t,a_t^i)d\boldsymbol{x}_t\\ &= \begin{cases}\frac{(1-P_GP_D)\lambda}{V_t^{M_t-1}[P_GP_DM_t+(1-P_GP_D)\lambda V_t]} & i=0\\ \frac{P_D}{V_t^{M_t-1}[P_GP_DM_t+(1-P_GP_D)\lambda V_t]}e_i & i=1,\dots,M_t\end{cases}\end{aligned} \tag{1.30}$$

$$\begin{aligned}C &= \sum_{i=0}^{M_t} p(a_t^i|\mathcal{Z}_t=\hat{\mathcal{Z}}_t)\\ &= \frac{(1-P_GP_D)\lambda+\sum_{i=1}^{M_t}e_i}{V_t^{M_t-1}[P_GP_DM_t+(1-P_GP_D)\lambda V_t]}\end{aligned} \tag{1.31}$$

Where,

$$e_i = \frac{1}{|2\pi\mathcal{S}_t|^{\frac{1}{2}}}e^{-\frac{1}{2}(\hat{\boldsymbol{z}}_t^i-\overline{\boldsymbol{z}}_t)^T\mathcal{S}_t^{-1}(\hat{\boldsymbol{z}}_t^i-\overline{\boldsymbol{z}}_t)} \tag{1.32}$$

and was entirely contained within the normalisation constant $\hat{g}_t^i$ of (1.30). Now (1.27) can be given explicitly as:

$$\begin{aligned}\delta_{\boldsymbol{z}_t\to\boldsymbol{x}_t}(\boldsymbol{x}_t) &= \mathcal{W}_t^0\mathcal{C}(\boldsymbol{x}_t;\Sigma_t^{-1},\Sigma_t^{-1}\boldsymbol{\mu}_{\boldsymbol{x}},g_{\boldsymbol{x}})+\sum_{i=1}^{M_t}\mathcal{W}_t^i\mathcal{C}(\boldsymbol{x}_t;\hat{\mathcal{S}}_t,\hat{\boldsymbol{h}}_t^i,\tilde{g}_t^i)\\ &= \sum_{i=0}^{M_t}W_t^i\mathcal{C}_i(\boldsymbol{x}_t;\mathcal{P}_t^i,\boldsymbol{h}_t^i,g_t^i)\end{aligned}$$

Where,

$$\mathcal{W}_t^i = \begin{cases} \frac{b}{b+\sum_{i=0}^{M_t}} & i = 0 \\ \frac{e_i}{b+\sum_{i=0}^{M_t}} & i = 1, \ldots, M_t \end{cases} \tag{1.33}$$

and,

$$b = \lambda |2\pi \mathcal{S}_t|^{\frac{1}{2}} \frac{1 - P_D P_G}{P_D} \tag{1.34}$$

It is assumed that $\tilde{g}_t^i$ is some appropriate normalising constant with e_i removed. $\hat{\mathcal{S}}_t$ is the final updated covariance for the given filter and is the same for each measurement distribution, $\hat{\boldsymbol{h}}_t^j$ contains the actual measurement information $\hat{\boldsymbol{z}}_t^j$.

1.2.3 Approximation

As mentioned countless times before, the mixture in (1.33) will be approximated with a single Gaussian with matching moments. The expected value of the distribution has been given in (1.1). The covariance can now be determined as follows:

$$\begin{aligned}
\overline{\Sigma}_t &= \mathbb{E}[(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_t)(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_t)^T] \\
&= \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^T] - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \int \boldsymbol{x}_t \boldsymbol{x}_t^T \delta_{\boldsymbol{z}_t \to \boldsymbol{x}_t}(\boldsymbol{x}_t) d\boldsymbol{x}_t - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \int \boldsymbol{x}_t \boldsymbol{x}_t^T \sum_{i=0}^{M_t} \mathcal{W}_t^i \mathcal{C}_i(\boldsymbol{x}_t; \mathcal{P}_t^i, \boldsymbol{h}_t^i, g_t^i) d\boldsymbol{x}_t - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \sum_{i=0}^{M_t} \mathcal{W}_t^i \left[\int \boldsymbol{x}_t \boldsymbol{x}_t^T \mathcal{C}_i(\boldsymbol{x}_t; \mathcal{P}_t^i, \boldsymbol{h}_t^i, g_t^i) d\boldsymbol{x}_t \right] - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \sum_{i=0}^{M_t} \mathcal{W}_t^i \left[\int \boldsymbol{x}_t \boldsymbol{x}_t^T \mathcal{C}_i(\boldsymbol{x}_t; \mathcal{P}_t^i, \boldsymbol{h}_t^i, g_t^i) d\boldsymbol{x}_t - (\overline{\boldsymbol{x}}_t^i)(\overline{\boldsymbol{x}}_t^i)^T + (\overline{\boldsymbol{x}}_t^i)(\overline{\boldsymbol{x}}_t^i)^T \right] - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \sum_{i=0}^{M_t} \mathcal{W}_t^i \left[\int \boldsymbol{x}_t \boldsymbol{x}_t^T \mathcal{C}_i(\boldsymbol{x}_t; \mathcal{P}_t^i, \boldsymbol{h}_t^i, g_t^i) d\boldsymbol{x}_t - (\overline{\boldsymbol{x}}_t^i)(\overline{\boldsymbol{x}}_t^i)^T \right] + \sum_{i=0}^{M_t} \mathcal{W}_t^i (\overline{\boldsymbol{x}}_t^i)(\overline{\boldsymbol{x}}_t^i)^T - \overline{\boldsymbol{x}}_t \overline{\boldsymbol{x}}_t^T \\
&= \sum_{i=0}^{M_t} \mathcal{W}_t^i \Sigma_t^i + \sum_{i=0}^{M_t} W_t^i \left[(\overline{\boldsymbol{x}}_t^i - \overline{\boldsymbol{x}}_t)(\overline{\boldsymbol{x}}_t^i - \overline{\boldsymbol{x}}_t)^T \right]
\end{aligned} \tag{1.35}$$

Using the previous developments and some stock definition of Kalman Gain, K_t, (1.35) can be written as:

$$\overline{\Sigma}_t = \mathcal{W}_t^0 \Sigma_t + (1 - \mathcal{W}_t^0) \hat{\mathcal{S}}_t + K_t \sum_{i=0}^{M_t} W_t^i \left[(\hat{\boldsymbol{z}}_t^i - \overline{\boldsymbol{z}}_t)(\hat{\boldsymbol{z}}_t^i - \overline{\boldsymbol{z}}_t)^T \right] K_t^T \tag{1.36}$$

[1–3] used a linear Kalman Filter, but this document neglects the finicky details of specific filters. (1.1) and (1.36) prove that the PGM is equivalent to the established PDAF.

References

[1] Bar-Shalom, Yaakov, and Edison Tse. "Tracking in a cluttered environment with probabilistic data association." *Automatica*, vol. 11.5, pp. 451-460, 1975.

[2] Kirubarajan, Thiagalingam, and Yaakov Bar-Shalom. "Probabilistic data association techniques for target tracking in clutter." Proceedings of the IEEE 92.3, 2004, pp. 536-557.

[3] Bar-Shalom, Yaakov, Fred Daum, and Jim Huang. "The probabilistic data association filter." *IEEE Control Systems*, vol 29.6, pp. 82-100, 2009.

[4] D. Koller and N. Friedman. 'Exact Inference: Clique Trees' in *Probabilistic Graphical Models: Principles and Techniques*, 1st ed., vol. 1. Cambridge, Massachusetts: The MIT Press, 2009, pp. 345-376.

[5] D. Koller and N. Friedman. 'Inference as Optimization' in *Probabilistic Graphical Models: Principles and Techniques*, 1st ed., vol. 1. Cambridge, Massachusetts: The MIT Press, 2009, pp. 391-428.

[6] D. Koller and N. Friedman. 'Inference in Hybrid Networks' in *Probabilistic Graphical Models: Principles and Techniques*, 1st ed., vol. 1. Cambridge, Massachusetts: The MIT Press, 2009, pp. 609-611.

[7] S. Thrun et al. 'Recursive State Estimation' in *Probabilistic Robotics*, 1st ed., vol. 1. Cambridge, Massachusetts: The MIT Press, 2009, pp. 14-35.

[8] S. Thrun et al. 'Gaussian Filters' in *Probabilistic Robotics*, 1st ed., vol. 1. Cambridge, Massachusetts: The MIT Press, 2009, pp. 39-53.