Preprint: IEEE Trans. Signal Processing, Vol. 63, No. 21, pp. 5812-5820, 2015.

Derivation of the PHD and CPHD filters based on direct Kullback-Leibler divergence minimisation

Ángel F. García-Fernández, Ba-Ngu Vo

***Abstract*—In this paper, we provide novel derivations of the probability hypothesis density (PHD) and cardinalised PHD (CPHD) filters without using probability generating functionals or functional derivatives. We show that both the PHD and CPHD filters fit in the context of assumed density filtering and implicitly perform Kullback-Leibler divergence (KLD) minimisations after the prediction and update steps. We perform the KLD minimisations directly on the multitarget prediction and posterior densities.**

***Index Terms*—Random finite sets, PHD filter, CPHD filter, multiple target tracking, Kullback-Leibler divergence**

I. Introduction

Inference in multi-target systems has a host of applications in many different disciplines such as radar/sonar tracking, navigation, air traffic control, computer vision and robotics [1]–[5]. The random finite set (RFS) formulation of the multi-target tracking problem is a widely used approach that allows us to model the appearance and disappearance of targets, misdetection of measurements and false alarms within the Bayesian framework [3]. More specifically, the objective is to estimate the current state of a dynamic system, which is a set that contains target states at the current time step, based on a sequence of measurements. State estimation is based on the posterior probability density function (PDF), i.e., the PDF of the current state given the sequence of measurements, as it contains all information of interest about the target states. Theoretically, the posterior can be calculated recursively by the prediction and update steps [3]. However, in general, these steps cannot be computed in closed-form so approximations are necessary [3].

Two popular approximations to the posterior are provided by the probability hypothesis density (PHD) filter [6] and the cardinalised PHD (CPHD) filter [7]. There are some extensions of the PHD/CPHD filters [8] but, in this paper, we always refer to their classical forms [6], [7]. These filters have been succesfully applied in different fields such as multitarget tracking [9], [10], robotics [4], computer vision [5], road mapping [11] and sensor control [12], [13]. Importantly, they admit elegant expressions which avoid the computational complexity of evaluating measurement-to-target association hypotheses. Implementations of these filters based on Gaussian mixtures, sequential Monte Carlo methods or splines have been proposed [10], [14]–[16]. However, the topic of this paper is not PHD and CPHD filter implementations but their derivation.

Both filters were originally derived by Mahler using probability generating functionals (PGFLs) and functional derivatives [6], [7]. The Faà di Bruno's formulae for functional derivatives [8], [17], [18] provide elegant derivations of the original, as well as more general PHD/CPHD filters, based on PGFLs. To facilitate implementations, the resulting CPHD equations are usually rewritten in the form given in [10]. Alternative derivations for the PHD filter based on measure theory have been proposed in [19], [20]. While these derivations are mathematically rigorous, they have not been extended to the CPHD filter. An interesting interpretation of the PHD/CPHD formulae in terms of the probability existence of targets on infinitesimal regions of the state space is given in [21].

In this paper, we present novel derivations of the PHD and CPHD filters that do not require the use of PGFLs or functional derivatives. We believe that the new derivations are more accessible, thereby bringing the PHD/CPHD filters to a wider audience. In addition, we cast these filters into the assumed density filtering (ADF) framework [22], [23]. In ADF, we propagate a certain type of PDF in the Bayesian filtering recursion. As the output PDF of the prediction and/or update steps might not be of the considered type, we have to approximate it by a PDF of the type under consideration to continue with the filtering recursion. Ideally, this approximation is obtained by some optimality criterion such as minimising the Kullback-Leibler divergence (KLD) w.r.t. the true PDF [24], [25]. KLD minimisations are sometimes referred to as moment projection (M-projection) or information-projection (I-projection) depending on the order of the PDFs in the KLD [24] and are also used in consensus algorithms [26], [27]. The PHD and CPHD filters follow this scheme with M-projections for Poisson and independent identically distributed (IID) cluster PDFs, respectively [3]. The KLD minimisation property of the PHD filter has been known since its inception [6] but, for the CPHD filter, it is proved in this paper and also independently in [28]. We proceed to explain this more thoroughly.

The Poisson point process is perhaps the best known of the RFSs [29]. In this case, the number of elements in the set is Poisson distributed and its elements are IID. This is the type of PDF that the PHD filter propagates. If the prior is Poisson, the result of Bayes' rule is no longer Poisson so the PHD filter performs KLD minimisation to approximate the posterior [6]. In the prediction step with the usual modelling assumptions and Poisson input, the output is Poisson if there is

The authors are with the Department of Electrical and Computer Engineering, Curtin University, Perth, WA 6102, Australia (emails: {angel.garciafernandez, ba-ngu.vo}@curtin.edu.au).

This work was supported in part by the Australian Research Council under Discovery Project DP130104404.

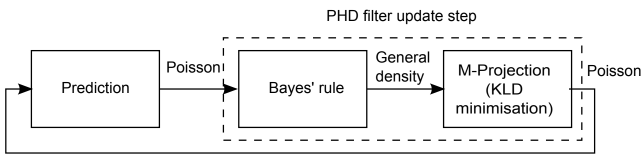


Figure 1: PHD filter with Poisson births and no target spawning diagram. The PHD filter assumes that the PDFs involved are Poisson. The output of Bayes' rule, which is given by Eq. (1), is no longer Poisson but, in order to be able to perform filtering, it obtains the best Poisson approximation to the posterior by minimising the KLD divergence (M-projection).

no target spawning and the RFS of new born targets is Poisson. Otherwise, KLD minimisation is performed to obtain a Poisson PDF. The resulting recursions are illustrated in Figures 1 and 2. By applying KLD minimisation to the output PDF of the prediction and update steps, we directly obtain the PHD filter recursion. In contrast, the CPHD filter propagates an IID cluster PDF, which is more general than the Poisson PDF. In an IID cluster PDF, targets are also IID but the cardinality distribution is arbitrary. We show in this paper that both the prediction and update steps of the CPHD filter can be directly obtained by applying KLD minimisation to the output PDF of the prediction and update steps. The resulting recursion is illustrated in Figure 3.

The rest of this paper is organised as follows. In Section II, we review the Bayesian filtering recursion using RFSs. In Section III, we provide two useful theorems for KLD minimisation. The PHD and CPHD filter equations are derived in Sections IV and V, respectively. Finally, conclusions are given in Section VI.

II. Bayesian filtering with random finite sets

In this section we review the Bayesian filtering recursion with RFSs, which consists of the usual prediction and update steps. As we only need to consider one prediction and update step, we omit the time index of the filtering recursion for notational simplicity.

In the standard RFS framework to target tracking, a single target state $x \in \mathbb{R}^{n_x}$ and the state $X \in \mathcal{F}(\mathbb{R}^{n_x})$, which denotes the space of all finite subsets of $\mathbb{R}^{n_x}$, so X is a set whose elements are single target state vectors. In the update step, the state is observed by measurements that are represented as a set $Z \in \mathcal{F}(\mathbb{R}^{n_z})$. Given a prior PDF $\nu(\cdot)$ and the PDF $f(Z|X)$ of the measurement Z given the state X, the posterior PDF of X after observing Z is given by Bayes' rule [6]

$$q(X) = \frac{f(Z|X)\nu(X)}{p(Z)} \tag{1}$$

where the PDF $p(Z)$ of the measurement is given by the set integral

$$\begin{aligned} p(Z) &= \int f(Z|X)\nu(X)\delta X \qquad (2)\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int f\left(Z\left|\{x_1,...,x_n\}\right.\right)\\ &\quad \times \nu\left(\{x_1,...,x_n\}\right)d\left(x_1,...,x_n\right). \qquad (3)\end{aligned}$$

The Bayesian filtering recursion is completed with the prediction step. Given a posterior PDF $q(\cdot)$, the prior PDF $\omega(\cdot)$ at the next time step is given by the Chapman-Kolmogorov equation

$$\omega(X') = \int \gamma(X'|X)\, q(X)\,\delta X \tag{4}$$

where $X' \in \mathcal{F}(\mathbb{R}^{n_x})$ denotes the state at the next time step and $\gamma(X'|X)$ is the PDF of the state X' given the state X.

As in single target filtering, the prediction and update steps cannot be computed in closed-form in general. In single target filtering, a well-known technique is assumed density filtering [22], [23], in which the PDF before the prediction and update step is assumed to be of a certain form. Then, we compute the output PDF via Bayes' rule or Chapman-Kolmogorov equation and project it to the same family of PDFs so that the Bayesian recursion can be performed. Ideally, the projection should be performed by minimising the Kullback-Leibler divergence [25]. This is exactly what the PHD and CPHD filters do with Poisson and IID cluster PDFs, respectively, see Figures 1, 2 and 3. An equivalent idea in single target filtering with Gaussian PDFs was proposed in [30].

We also want to remark at this point the similarity between the PDF of the measurement (2) and the prediction step (4). As we will see, unless we consider target spawning/non-Poisson births in the PHD filter (target spawning is not considered in the CPHD filter), both equations are identical. Therefore, when we derive the filters, we first compute the update step, which requires the calculation of (2), and use this knowledge to derive the prediction step.

III. Kullback-Leibler minimisation

As indicated in the previous section, in both the PHD and CPHD filtering recursions, there are several KLD minimisations. As a result, in order to derive the filters based on this framework, we need to indicate how KLD minimisations are performed. In Section III-A, we first review the concepts of PHD and cardinality distribution of an RFS density. In Section III-B, we review a known result for KLD minimisation for Poisson RFS. In Section III-C, we provide a theorem that indicates how KLD minimisations are performed for IID cluster RFS.

A. PHD and cardinality distribution

Given an RFS density $\pi(\cdot)$, its PHD is [3, Eq. (16.33)]

$$\begin{aligned} D_\pi(x) &= \int \pi(\{x\}\cup X)\,\delta X\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int \pi(\{x,x_1,...,x_n\})\,d(x_1,...,x_n). \qquad (5)\end{aligned}$$

The PHD is also called intensity function in stochastic geometry [31]. The cardinality distribution of $\pi(\cdot)$ is [3, Eq. (11.115)]

$$\rho_\pi(n) = \frac{1}{n!}\int \pi(\{x_1,...,x_n\})\,d(x_1,...,x_n). \tag{6}$$

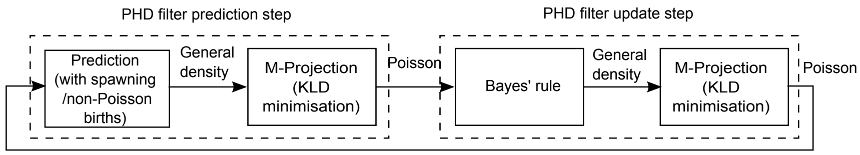


Figure 2: PHD filter with non-Poisson births and/or target spawning diagram. The PHD filter assumes that the PDFs involved are Poisson. The output of the prediction and Bayes' rule, which are given by Eqs. (4) and (1), are no longer Poisson but, in order to be able to perform filtering, it obtains the best Poisson approximation to the corresponding density by minimising the KLD divergence (M-projection).

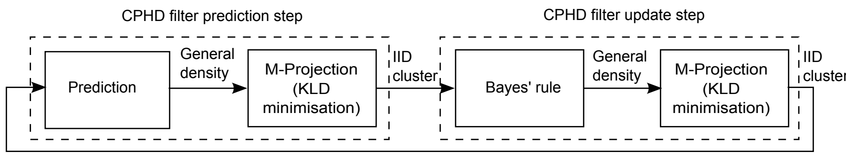


Figure 3: CPHD filter diagram. The CPHD filter assumes that the PDFs involved are IID cluster. The output of the prediction and Bayes' rule, which are given by Eqs. (4) and (1), are no longer IID clusters but the CPHD filter obtains the best IID cluster approximation by minimising the KLD divergence (M-projection).

B. Poisson RFS

The PHD filter propagates a Poisson PDF. If $\nu(\cdot)$ is Poisson, it can be written as [3, Eq. (11.122)]

$$\nu(\{x_1,...,x_n\}) = e^{-\lambda_\nu}\lambda_\nu^n \prod_{j=1}^{n} \breve{\nu}(x_j) \tag{7}$$

where $\breve{\nu}(\cdot)$ is a PDF on the single target state space and $\lambda_\nu \geq 0$. A Poisson PDF is characterised by its PHD $D_\nu(x) = \lambda_\nu \breve{\nu}(x)$.

Theorem 1. *Given an RFS density $\pi(\cdot)$, the Poisson density $\nu(\cdot)$ that minimises the KLD*

$$\mathrm{D}(\pi \| \nu) = \int \pi(X) \log \frac{\pi(X)}{\nu(X)} \delta X$$

is characterised by its PHD $D_\nu(\cdot) = D_\pi(\cdot)$.

Theorem 1 is proved in [6, Theorem 4].

C. IID cluster RFS

The CPHD filter propagates an IID cluster PDF. If $\nu(\cdot)$ is the density of an IID cluster RFS, it can be written as [3, Eq. (11.121)]

$$\nu(\{x_1,...,x_n\}) = \rho_\nu(n)\, n! \prod_{j=1}^{n} \breve{\nu}(x_j) \tag{8}$$

where $\rho_\nu(\cdot)$ is the cardinality distribution and $\breve{\nu}(\cdot)$ is a PDF on the single target state space. It should be noted that if $\rho_\nu(\cdot)$ is Poisson, then (8) reduces to (7). The PHD of (8) is given by [8]

$$D_\nu(x) = \breve{\nu}(x) \sum_{n=0}^{\infty} n\rho_\nu(n)$$

where the second term corresponds to the expected number of targets. Clearly, an IID cluster density can be characterised either by $\rho_\nu(\cdot)$ and $\breve{\nu}(\cdot)$ or $\rho_\nu(\cdot)$ and $D_\nu(\cdot)$.

Theorem 2. *Given an RFS density $\pi(\cdot)$, the IID cluster density $\nu(\cdot)$ that minimises the KLD $\mathrm{D}(\pi \| \nu)$ is characterised by its PHD $D_\nu(\cdot) = D_\pi(\cdot)$ and its cardinality distribution $\rho_\nu(\cdot) = \rho_\pi(\cdot)$.*

Theorem 2 is proved in the Appendix.

IV. PHD FILTER DERIVATION

In this section, we provide the derivation of the PHD filter based on direct KLD minimisation. In Section IV-A, we review the PHD filter update and the proposed proof is provided in Section IV-B. Lastly, in Section IV-C, we address the prediction step.

A. Review of the PHD filter update

The PHD filter update is developed under the assumptions

- U1 The set $Z = Z_1 \cup Z_2$ where Z_1 and Z_2 are the independent RFSs of measurements coming from targets and clutter, respectively.
- U2 Given $X = \{x_1,...,x_n\}$, $Z_1 = \cup_{i=1}^{n} \tilde{Z}_{1,i}$ where $\tilde{Z}_{1,i} = \oslash$ with probability $1 - p_D(x_i)$, otherwise $\tilde{Z}_{1,i} = \{z_i\}$ where z_i has a PDF $l(\cdot|x_i)$ and sets $\tilde{Z}_{1,i}$ $i = 1,...,n$ are independent.
- U3 The set Z_2 of clutter measurements is Poisson with RFS density $c(\cdot)$.
- U4 The prior is Poisson.

Under Assumptions U1, U2 and U3, which define the standard measurement model, the PDF of the measurement given the state is [3, Eq. (12.42)]

$$\begin{aligned} f(\{z_1,...,z_{n_z}\} | \{x_1,...,x_n\}) \\ = e^{\lambda_c} \left[\prod_{i=1}^{n_z} \lambda_c \breve{c}(z_i)\right] \left[\prod_{i=1}^{n} (1 - p_D(x_i))\right] \\ \times \sum_{\sigma \in \Xi_{n,n_z}} \prod_{i:\sigma_i > 0} \frac{p_D(x_i)\, l(z_{\sigma_i}|x_i)}{(1 - p_D(x_i))\, \lambda_c \breve{c}(z_{\sigma_i})} \end{aligned} \tag{9}$$

where Ξ_{n,n_z} is a set which contains all the vectors $\sigma = (\sigma_1,...,\sigma_n)$ that indicate associations of n_z measurements to n targets, which can be either detected or undetected, taking into account that only one measurement can be associated with a given target. If $\sigma \in \Xi_{n,n_z}$, $\sigma_i \in \{1,...,n_z\}$ indicates the measurement associated with target i and $\sigma_i = 0$ indicates that target i has not been detected.

The PHD filter update equation is [6]

$$D_q(x) = (1 - p_D(x)) D_\nu(x) + p_D(x) \times \sum_{z\in Z} \frac{l(z|x) D_\nu(x)}{D_c(z) + \int p_D(x') l(z|x') D_\nu(x') dx'} \quad (10)$$

where $x' \in \mathbb{R}^{n_x}$. Note that (10) is the PHD of the posterior $q(\cdot)$, see (1), under Assumptions U1-U4. This is proved in the next subsection.

B. Proof of the PHD filter update

The aim is to compute the PHD of the posterior, which characterises the best Poisson approximation to the posterior in the KLD sense, see Theorem 1. Using (1) and (5), we get

$$D_q(x) = \frac{1}{p(Z)} \sum_{n=0}^{\infty} \frac{1}{n!} \int f(Z|\{x, x_1, ..., x_n\}) \times \nu(\{x, x_1, ..., x_n\}) d(x_1, ..., x_n). \quad (11)$$

First, we compute the denominator in Section IV-B1 and then we complete the proof in Section IV-B2.

1) Density of the measurement: The denominator of (11) corresponds to the PDF $p(\cdot)$ of Z. This PDF can be obtained from well-known results of Poisson point processes theory [20]. According to Assumption U1, Z is the union of two independent sets Z_1 and Z_2. Under Assumption U2, set Z_1 comes from performing thinning [31] on X with a probability $p_D(\cdot)$ followed by a displacement [29] with Markov transition $l(\cdot|x)$. As X is Poisson distributed, see Assumption U4, we can apply the thinning and displacement theorems [29], [31] so that we get that Z_1 is Poisson distributed with intensity

$$\lambda_\nu \int p_D(x) l(z|x) \breve{\nu}(x) dx.$$

Under Assumption U1, Z is the union of two independent Poisson RFSs. Consequently, we can apply the superposition theorem [29], which says that Z is Poisson distributed with intensity given by the sum of the intensities

$$D_p(z) = \lambda_c \breve{c}(z) + \lambda_\nu \int p_D(x) l(z|x) \breve{\nu}(x) dx. \quad (12)$$

Therefore, the denominator of (11) is given by

$$p(Z) = e^{-\lambda_\nu \int p_D(x)\breve{\nu}(x)dx - \lambda_c} \prod_{z\in Z} \left[\lambda_c \breve{c}(z) + \lambda_\nu \int p_D(x) l(z|x) \breve{\nu}(x) dx\right]. \quad (13)$$

2) Rest of the proof: We perform the following decomposition

$$\begin{aligned} &f(Z|\{x, x_1, ..., x_n\}) \\ &= (1 - p_D(x)) f(Z|\{x_1, ..., x_n\}) \\ &\quad + p_D(x) \sum_{z\in Z} l(z|x) f(Z \setminus \{z\} |\{x_1, ..., x_n\}) \end{aligned} \quad (14)$$

where $B \setminus A = \{z \in B | z \notin A\}$. Note that $f(Z|X)$ goes through all the possible data association hypotheses, see (9). There are two hypotheses for target x, it can be either detected or not detected. If it is not detected, which happens with probability $(1 - p_D(x))$, all the measurements have originated from the rest of the targets or clutter. This is represented by the first term of (14). The other hypothesis is that x is detected, which happens with probability $p_D(x)$. If it is detected, it can be associated with any of the $z \in Z$ measurements and the rest $Z \setminus \{z\}$ of the measurements have originated from the rest of the targets or clutter. This is represented by the second term of (14).

Substituting (7) and (14) into (11), we obtain

$$\begin{aligned} D_q(x) = &\frac{1}{p(Z)} \sum_{n=0}^{\infty} \frac{1}{n!} (1 - p_D(x)) \int f(Z|\{x_1, ..., x_n\}) \\ &\times e^{-\lambda_\nu} \lambda_\nu^{n+1} \breve{\nu}(x) \prod_{j=1}^{n} \breve{\nu}(x_j) d(x_1, ..., x_n) \\ &+ \frac{1}{p(Z)} \sum_{n=0}^{\infty} \frac{1}{n!} p_D(x) \sum_{z\in Z} l(z|x) \\ &\times \int f(Z \setminus \{z\} |\{x_1, ..., x_n\}) \\ &\times e^{-\lambda} \lambda_\nu^{n+1} \breve{\nu}(x) \prod_{j=1}^{n} \breve{\nu}(x_j) d(x_1, ..., x_n). \end{aligned} \quad (15)$$

Using (3) and (7) in (15), we obtain

$$D_q(x) = (1 - p_D(x)) \lambda_\nu \breve{\nu}(x) + \frac{1}{p(Z)} p_D(x) \sum_{z\in Z} l(z|x) \lambda_\nu \breve{\nu}(x) f(Z \setminus \{z\}). \quad (16)$$

Finally, we substitute (13) into (16) to get

$$D_q(x) = (1 - p_D(x)) \lambda_\nu \breve{\nu}(x) + p_D(x) \lambda_\nu \breve{\nu}(x) \sum_{z\in Z} \frac{l(z|x)}{\lambda_c \breve{c}(z) + \int p_D(x') l(z|x') \lambda_\nu \breve{\nu}(x') dx'}.$$

which completes the proof of (10).

C. PHD filter prediction

The PHD prediction assumes

- P1 The set $X' = X'_1 \cup X'_2 \cup X'_3$ where X'_1, X'_2 and X'_3 are the independent sets of surviving targets, newborn targets and spawned targets, respectively.
- P2 Given $X = \{x_1, ..., x_n\}$, $X'_1 = \cup_{i=1}^{n} \tilde{X}'_{1,i}$ where $\tilde{X}'_{1,i} = \oslash$ with probability $1 - p_S(x_i)$, otherwise $\tilde{X}'_{1,i} = \{x'_i\}$ where x'_i has a PDF $g(\cdot|x_i)$ and sets $\tilde{X}'_{1,i}$ $i = 1, ..., n$ are independent.
- P3 The set X'_2 of new born targets has an RFS density $b(\cdot)$.
- P4 Given $X = \{x_1, ..., x_n\}$, $X'_3 = \cup_{i=1}^{n} X''_{3,i}$ where $X''_{3,i}$ has PHD $D_\xi(\cdot|x_i)$ and $X''_{3,i}$ $i = 1, ..., n$ are independent.
- P5 The posterior $q(\cdot)$ is Poisson.

The PHD of the prior at the next time step is [6]

$$D_\omega(x') = D_b(x')$$

$$+\int\left[p_S\left(x\right)g\left(x'|x\right)+D_\xi\left(x'\left|x\right.\right)\right]D_q\left(x\right)dx. \tag{17}$$

If there is no target spawning and the density $b(\cdot)$ is Poisson, we see that U1-U4 are equivalent to P1-P3 and P5, therefore, the PDF of the measurement (2) and the prediction step (4) are analogous. Consequently, in this case, it can be directly established that the density $\omega(\cdot)$ is Poisson, with PHD given by (17) setting $D_\xi(\cdot)=0$, and there is no need to perform KLD minimisation, as illustrated by Figure 1. Under Assumption P4, the set X_3' of spawned targets constitutes a cluster process with centers given by $q(\cdot)$ [32, Chap. 6]. As a result, its PHD is given by

$$D_\xi\left(x'\right)=\int D_\xi\left(x'\left|x\right.\right)D_q\left(x\right)dx$$

where we have used (6.3.3) in [32]. By the superposition theorem [31] and Assumption P1, the PHD of X' is given by the sum of the PHDs of X_1', X_2', X_3' so we get (17). Due to the fact that (17) is the PHD of the prior, it represents the best Poisson fit to the prior in the KLD sense, see Theorem 1, as illustrated in Figure 2.

V. CPHD FILTER DERIVATION

In this section, we provide the derivation of the CPHD filter based on direct KLD minimisation. In Section V-A, we review the CPHD filter update and the proposed proof is provided in Section V-B. Lastly, in Section V-C, we address the prediction step.

A. Review of the CPHD filter update

The CPHD filter update is developed under Assumptions U1-U2 and

- U5 The set Z_2 of clutter measurements is IID cluster with RFS density $c(\cdot)$.
- U6 The prior is IID cluster.

Given two sequences $a(n)$ and $b(n)$ $n\in\mathbb{N}\cup\{0\}$, we denote

$$\langle a,b\rangle=\sum_{n=0}^{\infty}a\left(n\right)b\left(n\right).$$

The CPHD filter update equation for the cardinality and PHD are [10]

$$\rho_q\left(n\right)=\frac{\Upsilon^0\left[D_\nu,Z\right]\left(n\right)\rho_\nu\left(n\right)}{\left\langle\Upsilon^0\left[D_\nu,Z\right],\rho_\nu\right\rangle} \tag{18}$$

$$D_q\left(x\right)=\frac{\left\langle\Upsilon^1\left[D_\nu,Z\right],\rho_\nu\right\rangle}{\left\langle\Upsilon^0\left[D_\nu,Z\right],\rho_\nu\right\rangle}\left(1-p_D\left(x\right)\right)D_\nu\left(x\right)+\sum_{z\in Z}\frac{\left\langle\Upsilon^1\left[D_\nu,Z\setminus\{z\}\right],\rho_\nu\right\rangle}{\left\langle\Upsilon^0\left[D_\nu,Z\right],\rho_\nu\right\rangle}\frac{l\left(z|x\right)}{\breve{c}\left(z\right)}p_D\left(x\right)D_\nu(x) \tag{19}$$

where

$$\Upsilon^u\left[D_\nu,Z\right]\left(n\right)=\sum_{j=0}^{\min(|Z|,n-u)}(|Z|-j)!\rho_c\left(|Z|-j\right)\times\frac{\left[\int\left(1-p_D(x)\right)D_\nu(x)dx\right]^{n-(j+u)}}{\left[\int D_\nu(x)dx\right]^n}\times\frac{n!}{(n-j-u)!}e_j\left(\Xi\left(D_\nu,Z\right)\right) \tag{20}$$

$$\Xi\left(D_\nu,Z\right)=\left\{\int\frac{l\left(z|x\right)}{\breve{c}\left(z\right)}p_D\left(x\right)D_\nu(x)dx:z\in Z\right\}$$

and the elementary symmetric function is

$$e_j\left(Z\right)=\sum_{S\subseteq Z,|S|=j}\left(\prod_{\zeta\in S}\zeta\right) \tag{21}$$

with $e_0(Z)=1$ by convention. Note that there are two typographical errors in the definition of $\Upsilon^u\left[D_\nu,Z\right](n)$ in [10]. Value $(|Z|-j)$ should have a factorial and the upper value of the sum should be $\min(|Z|,n-u)$ instead of $\min(|Z|,n)$. While the first one has been usually corrected in later sources, the second still appears in many of them [21], [33]–[38].

B. Proof of the CPHD filter update

1) Density of the measurement: We first proceed to calculate the density of the measurement. Under Assumptions U1-U2, given $\{x_1,...,x_n\}$, $Z=Z_1\cup Z_2$ where Z_1 are the measurements originated from the targets, Z_2 are the clutter measurements, and these two sets are independent. We use $t(\cdot)$ to denote the RFS density of Z_1 given $\{x_1,...,x_n\}$ and $c(\cdot)$ for the clutter measurements, as stated in Assumption U5. Then, we apply the convolution formula to obtain the PDF of Z given $\{x_1,...,x_n\}$ [3, page 385]:

$$f\left(Z\left|\{x_1,...,x_n\}\right.\right)=\sum_{S\subseteq Z}t\left(S\left|\{x_1,...,x_n\}\right.\right)c\left(Z\setminus S\right) \tag{22}$$

$$=\sum_{j=0}^{\min(|Z|,n)}\sum_{S\subseteq Z,|S|=j}t\left(S\left|\{x_1,...,x_n\}\right.\right)c\left(Z\setminus S\right). \tag{23}$$

In the sum in (23), we use the auxiliary variable j that indicates the number of measurements associated with targets. The minimum number value of j is zero and its maximum value is the minimum between the number of measurements or number of targets.

Under Assumption U2, the density of the measurement set $S=\{s_1,...,s_j\}$ coming from target set $\{x_1,...,x_n\}$ can be written as [8, Eq. (7.24)]

$$t\left(\{s_1,...,s_j\}\left|\{x_1,...,x_n\}\right.\right)=\left[\prod_{i=1}^{n}\left(1-p_D\left(x_i\right)\right)\right]\sum_{\sigma\in\Gamma_{n,j}}\prod_{i=1}^{j}\frac{l\left(s_i\left|x_{\sigma_i}\right.\right)p_D\left(x_{\sigma_i}\right)}{\left(1-p_D\left(x_{\sigma_i}\right)\right)}$$

$$=\sum_{\sigma\in\Gamma_{n,j}}\prod_{i=1}^{n_z}l\left(s_i\left|x_{\sigma_i}\right.\right)p_D\left(x_{\sigma_i}\right)\prod_{i=1}^{n-j}\left(1-p_D\left(x_{\sigma_i'}\right)\right) \tag{24}$$

where $\Gamma_{n,j}$ is the set that contains all possible selections $\sigma=(\sigma_1,...,\sigma_j)$ of j ordered elements from $(1,...,n)$ and $\sigma'=\{1,...,n\}\setminus\{\sigma_1,...,\sigma_j\}$. It should be noted that the cardinality of set $\Gamma_{n,j}$ is

$$\left|\Gamma_{n,j}\right|=j!\left(\begin{array}{c}n\\j\end{array}\right). \tag{25}$$

That is, there are $\begin{pmatrix} n \\ j \end{pmatrix}$ possible ways to select j elements from n elements and the resulting j elements can be ordered in $j!$ different ways. First, we calculate the density of j measurements generated by n targets, without considering clutter. According to (8), given n targets, they are distributed according to a PDF $\prod_{j=1}^{n} \breve{\nu}(\cdot)$ so we can use (24) and compute

$$\begin{aligned}
&\int t\left(\{s_1,...,s_j\}\left|\{x_1,...,x_n\}\right.\right)\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\
&=\sum_{\sigma\in\Gamma_{n,j}}\int\prod_{i=1}^{j}l\left(z_i\left|x_{\sigma_i}\right.\right)p_D\left(x_{\sigma_i}\right)\prod_{i=1}^{n-j}\left(1-p_D\left(x_{\sigma'_i}\right)\right)\\
&\quad\times\prod_{l=1}^{n}\breve{\nu}(x_l)\,d(x_1,...,x_n) \qquad (26)\\
&=\sum_{\sigma\in\Gamma_{n,j}}\prod_{i=1}^{j}\left[\int l\left(z_i\left|x_{\sigma_i}\right.\right)p_D\left(x_{\sigma_i}\right)\breve{\nu}\left(x_{\sigma_i}\right)dx_{\sigma_i}\right]\\
&\quad\times\prod_{i=1}^{n-j}\int\left[\left(1-p_D\left(x_{\sigma'_i}\right)\right)\breve{\nu}\left(x_{\sigma'_i}\right)dx_{\sigma'_i}\right] \qquad (27)\\
&=\left[\int\left(1-p_D(x)\right)\breve{\nu}(x)\,dx\right]^{n-j}\\
&\quad\times\prod_{i=1}^{j}\left[\int l\left(s_i\left|x\right.\right)p_D(x)\,\breve{\nu}(x)\,dx\right]\left(\sum_{\sigma\in\Gamma_{n,j}}1\right) \qquad (28)\\
&=\left[\int\left(1-p_D(x)\right)\breve{\nu}(x)\,dx\right]^{n-j}j!\begin{pmatrix} n \\ j \end{pmatrix}\\
&\quad\times\prod_{i=1}^{j}\left[\int l\left(s_i\left|x\right.\right)p_D(x)\,\breve{\nu}(x)\,dx\right]. \qquad (29)
\end{aligned}$$

From Eq. (26) to (27), we write the integral as the product of n integrals. In the next step, we use the fact that the integrals inside the summatory do not depend on σ so the last term in (28) represents the number of elements in $\Gamma_{n,j}$, which is given by (25).

Second, we calculate the density of the measurements given that there are n targets, considering clutter. We use (23) to get

$$\begin{aligned}
&\int f(Z|\{x_1,...,x_n\})\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\
&=\sum_{j=0}^{\min(|Z|,n)}\sum_{S\subseteq Z,|S|=j}c(Z\setminus S)\int t\left(S\left|\{x_1,...,x_n\}\right.\right)\\
&\quad\times\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n). \qquad (30)
\end{aligned}$$

Under Assumption U5 and using (29), (30) becomes

$$\begin{aligned}
&\int f(Z|\{x_1,...,x_n\})\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\
&=\prod_{z\in Z}\breve{c}(z)\sum_{j=0}^{\min(|Z|,n)}(|Z|-j)!\rho_c(|Z|-j)\,j!\begin{pmatrix} n \\ j \end{pmatrix}\\
&\quad\times\left[\int\left(1-p_D(x)\right)\breve{\nu}(x)\,dx\right]^{n-j}\\
&\quad\times\sum_{S\subseteq Z,|S|=j}\frac{\prod_{z\in S}\left[\int l\left(z\left|x\right.\right)p_D(x)\,\breve{\nu}(x)\,dx\right]}{\prod_{z\in S}\breve{c}(z)}\\
&=\prod_{z\in Z}\breve{c}(z)\,\Upsilon^0\left[D_\nu,Z\right](n) \qquad (31)
\end{aligned}$$

where we have used (20) in the last equality.

Under Assumption U6, we can use (3) and (31) to obtain the PDF of the measurement, which becomes

$$\begin{aligned}
p(Z)&=\sum_{n=0}^{\infty}\rho_\nu(n)\int f(Z|\{x_1,...,x_n\})\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\
&=\prod_{z\in Z}\breve{c}(z)\sum_{n=0}^{\infty}\rho_\nu(n)\,\Upsilon^0\left[D_\nu,Z\right](n)\\
&=\prod_{z\in Z}\breve{c}(z)\left\langle\rho_\nu,\Upsilon^0\left[D_\nu,Z\right]\right\rangle. \qquad (32)
\end{aligned}$$

2) Cardinality of the posterior: The cardinality of the posterior is obtained by (1) and (6):

$$\begin{aligned}
\rho_q(n)&=\frac{1}{n!}\int q\left(\{x_1,...,x_n\}\right)d(x_1,...,x_n)\\
&=\frac{1}{n!p(Z)}\int f(Z|\{x_1,...,x_n\})\\
&\quad\times\nu\left(\{x_1,...,x_n\}\right)d(x_1,...,x_n). \qquad (33)
\end{aligned}$$

Under Assumption U6, we can write (33) as

$$\begin{aligned}
\rho_q(n)&=\frac{\rho_\nu(n)}{p(Z)}\int f(Z|\{x_1,...,x_n\})\\
&\quad\times\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n). \qquad (34)
\end{aligned}$$

We finish the proof of (18) by substituting (31) and (32) into (34).

3) PHD of the posterior: The PHD of the posterior can be obtained using (11) and Assumption U6 as

$$\begin{aligned}
D_q(x)&=\frac{1}{p(Z)}\sum_{n=0}^{\infty}(n+1)\int f\left(Z|\{x,x_1,...,x_n\}\right)\\
&\quad\times\rho_\nu(n+1)\,\breve{\nu}(x)\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n). \qquad (35)
\end{aligned}$$

As in the PHD filter derivation, we use decomposition (14) in (35) to obtain

$$\begin{aligned}
D_q(x)&=\frac{\left(1-p_D(x)\right)\breve{\nu}(x)}{p(Z)}\sum_{n=0}^{\infty}(n+1)\,\rho_\nu(n+1)\\
&\quad\times\int f\left(Z|\{x_1,...,x_n\}\right)\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\
&\quad+\frac{p_D(x)\,\breve{\nu}(x)}{p(Z)}\sum_{z\in Z}l(z|x)\sum_{n=0}^{\infty}(n+1)\,\rho_\nu(n+1)\\
&\quad\times\int f\left(Z\setminus\{z\}\left|\{x_1,...,x_n\}\right.\right)\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n). \qquad (36)
\end{aligned}$$

Before simplifying (36), from (20), we get the equality

$$\Upsilon^1\left[D_\nu,Z\right](n)=\frac{n}{\int D_\nu(x)dx}\Upsilon^0\left[D_\nu,Z\right](n-1). \quad (37)$$

We now proceed to simplify one of the terms in (36)

$$\begin{aligned}&\sum_{n=0}^{\infty}(n+1)\,\rho_\nu(n+1)\\&\quad\times\int f\left(Z|\left\{x_1,...,x_n\right\}\right)\prod_{j=1}^{n}\breve{\nu}\left(x_j\right)d\left(x_1,...,x_n\right)\\&\quad=\prod_{z\in Z}\breve{c}\left(z\right)\sum_{n=0}^{\infty}(n+1)\,\rho_\nu(n+1)\,\Upsilon^0\left[D_\nu,Z\right](n) \quad (38)\\&\quad=\prod_{z\in Z}\breve{c}\left(z\right)\sum_{n=1}^{\infty}n\rho_\nu(n)\,\Upsilon^0\left[D_\nu,Z\right](n-1)\\&\quad=\prod_{z\in Z}\breve{c}\left(z\right)\int D_\nu(x)dx\sum_{n=1}^{\infty}\rho_\nu(n)\,\Upsilon^1\left[D_\nu,Z\right](n)\\&\quad=\prod_{z\in Z}\breve{c}\left(z\right)\int D_\nu(x)dx\left\langle\rho_\nu,\Upsilon^1\left[D_\nu,Z\right]\right\rangle \quad (39)\end{aligned}$$

where we have used that $\Upsilon^1\left[D_\nu,Z\right](0)=0$, see (20).

Plugging (32) and (39) into the first and second term of (36) completes the proof

$$\begin{aligned}D_q(x)=&\frac{\left\langle\rho_\nu,\Upsilon^1\left[D_\nu,Z\right]\right\rangle}{\left\langle\rho_\nu,\Upsilon^0\left[D_\nu,Z\right]\right\rangle}\left(1-p_D(x)\right)D_\nu(x)\\&+\sum_{z\in Z}\frac{\left\langle\Upsilon^1\left[D_\nu,Z\setminus\{z\}\right],\rho_\nu\right\rangle}{\left\langle\Upsilon^0\left[D_\nu,Z\right],\rho_\nu\right\rangle}\frac{l\left(z|x\right)}{\breve{c}\left(z\right)}p_D\left(x\right)D_\nu(x).\end{aligned}$$

4) Interpretation: We can easily interpret some of the terms of the CPHD filter update (18)-(19). From (31), we see that $\Upsilon^0\left[D_\nu,Z\right](n)$ is proportional to the density of the measurements given that there are n targets evaluated at Z so $\left\langle\Upsilon^0\left[D_\nu,Z\right],\rho_\nu\right\rangle$ in (18)-(19) is proportional to $p(Z)$. From (23), variable j in $\Upsilon^0\left[D_\nu,Z\right](n)$, see (20), indicates the number of measurements associated to targets. Therefore, the cardinality update (18) can be seen as a Bayes' update, as pointed out in [10]. On the contrary, $\Upsilon^1\left[D_\nu,Z\right](n)$ does not admit such an easy interpretation. From (37), $\Upsilon^1\left[D_\nu,Z\right](n)$ is proportional to the density of the measurements given that there are $n-1$ targets evaluated at Z multiplied by n and divided by the a priori expected number of targets.

It is also interesting to see the steps of the derivation in which the data association problem is removed and only one hypothesis given the number of target generated measurements must be considered by the use of elementary symmetric functions, which are defined by (21). This can be clearly seen when we derive the PDF of the measurement from Equations (26) to (29). It is a consequence of the fact that targets are IID and, therefore, all data associations in (27) are alike so we just need to evaluate one. The same effect appears in the updated PHD in (38).

C. CPHD filter prediction

The CPHD filter prediction is obtained under Assumption P2 and

- P6 The set $X'=X'_1\cup X'_2$ where X'_1 and X'_2 are the independent sets of surviving targets and newborn targets, respectively.
- P7 The set X'_2 of new born targets is IID cluster with RFS density $b(\cdot)$.
- P8 The posterior $q(\cdot)$ is IID cluster.

The CPHD filter prediction equations for the cardinality and PHD are [10]

$$\begin{aligned}\rho_\omega(m)=&\sum_{j=0}^{m}\rho_b(m-j)\sum_{n=j}^{\infty}\binom{n}{j}\rho_q(n)\\&\times\frac{\left[\int\left(1-p_S(x)\right)D_q(x)dx\right]^{n-j}}{\left[\int D_q(x)dx\right]^n} \quad (40)\\&\times\left[\int p_S(x)\,D_q(x)dx\right]^j \quad (41)\end{aligned}$$

$$D_\omega(x')=D_b(x')+\int p_S(x)\,g(x'|x)\,D_q(x)\,dx. \quad (42)$$

It should be noted that the assumptions of the CPHD filter prediction (P2, P6-P8) are analogous of the CPHD filter update (U1-U2, U5-U6). Therefore, the prior PDF $\omega(\cdot)$ is analogous to the PDF of the measurements (32). Clearly, this density is not IID cluster but we can obtain the IID cluster approximation that minimises the KLD by using Theorem 2. The result for the PHD (42) can be established directly based on thinning, displacement and superposition of point processes [31]. In the rest of the section, we derive the equation for the cardinality. With the objective of not introducing more notation, we calculate the cardinality of the density of the measurements (32), which is an equivalent formula to the cardinality of the prior.

1) Cardinality of the prior: We calculate the cardinality of the measurements, whose PDF is given by (32), which is analogous to the cardinality of the prior. We use (6) and (32) to get

$$\begin{aligned}&\rho_p(m)\\&\quad=\frac{1}{m!}\int p\left(\{z_1,...,z_m\}\right)d\left(z_1,...,z_m\right)\\&\quad=\frac{1}{m!}\int\prod_{i=1}^{m}\breve{c}\left(z_i\right)\sum_{n=0}^{\infty}\rho_\nu(n)\sum_{j=0}^{\min(m,n)}(m-j)!\rho_c(m-j)\\&\qquad\times\frac{n!}{(n-j)!}\frac{\left[\int\left(1-p_D(x)\right)D_\nu(x)dx\right]^{n-j}}{\left[\int D_\nu(x)dx\right]^n}\\&\qquad\times\sum_{S\subseteq\{z_1,...,z_m\},|S|=j}\frac{\prod_{z\in S}\left[\int l\left(z\,|x\right)p_D\left(x\right)D_\nu(x)dx\right]}{\prod_{z\in S}\breve{c}\left(z\right)}\\&\qquad\times d\left(z_1,...,z_m\right). \quad (43)\end{aligned}$$

Integrating out $z_1...z_m$ in (43), we obtain

$$\begin{aligned}\rho_p(m)=&\frac{1}{m!}\sum_{n=0}^{\infty}\rho_\nu(n)\sum_{j=0}^{\min(m,n)}(m-j)!\rho_c(m-j)\frac{n!}{(n-j)!}\\&\times\frac{\left[\int\left(1-p_D(x)\right)D_\nu(x)dx\right]^{n-j}}{\left[\int D_\nu(x)dx\right]^n}\end{aligned}$$

$$\times \sum_{S\subseteq\{z_1,...,z_m\},|S|=j} \left[\int p_D(x) D_\nu(x)dx\right]^j. \quad (44)$$

As there are $\begin{pmatrix} m \\ j \end{pmatrix}$ subsets of $\{z_1, ..., z_m\}$ with cardinality j, (44) becomes

$$\begin{aligned}\rho_p(m) &= \frac{1}{m!}\sum_{n=0}^{\infty}\rho_\nu(n)\sum_{j=0}^{\min(m,n)}(m-j)!\rho_c(m-j)\\ &\times \frac{n!}{(n-j)!}\frac{\left[\int(1-p_D(x))D_\nu(x)dx\right]^{n-j}}{\left[\int D_\nu(x)dx\right]^n}\\ &\times\left[\int p_D(x)D_\nu(x)dx\right]^j\begin{pmatrix} m \\ j \end{pmatrix}\\ &=\sum_{n=0}^{\infty}\rho_\nu(n)\sum_{j=0}^{\min(m,n)}\rho_c(m-j)\begin{pmatrix} n \\ j \end{pmatrix}\\ &\times\frac{\left[\int(1-p_D(x))D_\nu(x)dx\right]^{n-j}}{\left[\int D_\nu(x)dx\right]^n}\\ &\times\left[\int p_D(x)D_\nu(x)dx\right]^j. \end{aligned} \quad (45)$$

We use the fact that $\sum_{n=0}^{\infty}\sum_{j=0}^{\min(m,n)} = \sum_{j=0}^{m}\sum_{n=j}^{\infty}$ in (45) so

$$\begin{aligned}\rho_p(m) = \sum_{j=0}^{m}\rho_c(m-j)\sum_{n=j}^{\infty}\begin{pmatrix} n \\ j \end{pmatrix}\rho_\nu(n)\\ \times\frac{\left[\int(1-p_D(x))D_\nu(x)dx\right]^{n-j}}{\left[\int D_\nu(x)dx\right]^n}\\ \times\left[\int p_D(x)D_\nu(x)dx\right]^j.\end{aligned}$$

By changing $\rho_c(\cdot)$, $p_D(\cdot)$ and $\nu(\cdot)$ for $\rho_b(\cdot)$, $p_S(\cdot)$ and $q(\cdot)$, respectively, we obtain (41).

VI. Conclusions

In this paper, we have examined the PHD and CPHD filters in the context of assumed density filtering, which enabled us to derive these filters based on KLD minimisation of the PDFs, without resorting to PGFLs or functional derivatives. Our derivations use an intuitive decomposition of the multi-target likelihood function and shed some insights into how data association is avoided in the CPHD filter update, which performs a Bayes update with an IID prior followed by a KLD minimisation. Avoiding data association has important practical implications as the computational burden of the filter is significantly lowered, albeit with some drawbacks such as the spooky effect [39]. For the CPHD filter, we only need to evaluate one hypothesis for target generated measurements. For the PHD filter, further simplifications are possible.

It should also be noted that the resulting minimised KLD is an indicator of the performance of the filter. The lower this KLD is, the closer the approximated posterior is to the true posterior and performance improves. This analysis for the (non-linear) Kalman filter update in single target applications was performed in [40]. Unfortunately, even though we can minimise the KLD to obtain the PHD/CPHD filters, we cannot obtain closed-form formulas for the resulting KLD. Nevertheless, we could always approximate the resulting KLD using Monte Carlo integration by drawing samples from the prior so that we could analyse the performance for different parameters/implementations in a mathematically rigorous way.

Appendix

In this appendix we prove Theorem 2. A general RFS density $\pi(\cdot)$ can be written as [6]

$$\pi(\{x_1, ..., x_n\}) = \rho_\pi(n)\, n!\pi_n(x_1, ..., x_n) \quad (46)$$

where $\pi_n(\cdot)$ is a permutation invariant vector density. Note that given $\pi(\cdot)$ we can recover $\rho_\pi(\cdot)$ by (6) [7].

Using (8), we get

$$\begin{aligned}\mathrm{D}(\pi\,\|\nu) &= \int \pi(X)\log\frac{\pi(X)}{\nu(X)}\delta X\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int\pi(\{x_1,...,x_n\})\\ &\quad\times\log\frac{\pi(\{x_1,...,x_n\})}{\nu(\{x_1,...,x_n\})}d(x_1,...,x_n)\\ &=\sum_{n=0}^{\infty}\rho_\pi(n)\int\pi_n(x_1,...,x_n)\\ &\quad\times\log\frac{\rho_\pi(n)\pi_n(x_1,...,x_n)}{\rho_\nu(n)\prod_{j=1}^{n}\breve{\nu}(x_j)}d(x_1,...,x_n)\\ &=\sum_{n=0}^{\infty}\rho_\pi(n)\log\frac{\rho_\pi(n)}{\rho_\nu(n)}\\ &\quad+\sum_{n=0}^{\infty}\rho_\pi(n)\int\pi_n(x_1,...,x_n)\\ &\quad\times\log\frac{\pi_n(x_1,...,x_n)}{\prod_{j=1}^{n}\breve{\nu}(x_j)}d(x_1,...,x_n).\end{aligned}$$

The objective is to find $\rho_\nu(\cdot)$ and $\breve{\nu}(\cdot)$ that minimise $\mathrm{D}(\pi\,\|\nu)$. From KLD minimisation over discrete variables, it is clear that $\rho_\nu(n) = \rho_\pi(n)$ minimises the KLD. Minimising the KLD w.r.t. $\breve{\nu}(\cdot)$ is equivalent to minimising the functional

$$\begin{aligned}L[\breve{\nu}] &= -\sum_{n=0}^{\infty}\rho_\pi(n)\int\pi_n(x_1,...,x_n)\\ &\quad\times\log\prod_{j=1}^{n}\breve{\nu}(x_j)\,d(x_1,...,x_n)\\ &=-\sum_{n=0}^{\infty}\rho_\pi(n)\sum_{j=1}^{n}\int\pi_n(x_1,...,x_n)\\ &\quad\times\log\breve{\nu}(x_j)\,d(x_1,...,x_n).\end{aligned}$$

As $\pi_n(\cdot)$ is permutation invariant, the n integrals have the same value. In addition, the marginal PDF of $\pi_n(\cdot)$ over any of its variables is the same. Therefore, if we denote the marginal as

$$\begin{aligned}\tilde{\pi}_n(x) &= \int\pi_n(x, x_2..., x_n)\,d(x_2,...,x_n)\\ &=\frac{1}{\rho_\pi(n)\,n!}\int\pi(\{x,x_2,...,x_n\})\,d(x_2,...,x_n)\end{aligned}$$

we obtain

$$L\left[\breve{\nu}\right]=-\int\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n\tilde{\pi}_{n}\left(x\right)\log\breve{\nu}\left(x\right)dx.$$

By KLD minimisation over vector densities, we know that this functional is minimised if [41]

$$\begin{aligned}\breve{\nu}\left(x\right) & =\frac{\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n\tilde{\pi}_{n}\left(x\right)}{\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n}\\ & =\frac{\sum_{n=0}^{\infty}\frac{n}{n!}\int\pi\left(\left\{ x,x_{2},...,x_{n}\right\} \right)d\left(x_{2},...,x_{n}\right)}{\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n}\\ & =\frac{\sum_{n=0}^{\infty}\frac{1}{n!}\int\pi\left(\left\{ x,x_{1},...,x_{n}\right\} \right)d\left(x_{1},...,x_{n}\right)}{\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n}\\ & =\frac{D_{\pi}\left(x\right)}{\sum_{n=0}^{\infty}\rho_{\pi}\left(n\right)n}.\end{aligned}$$

As in the minimum $\rho_{\nu}\left(\cdot\right)=\rho_{\pi}\left(\cdot\right)$, we get $D_{\nu}\left(\cdot\right)=D_{\pi}\left(\cdot\right)$, which completes the proof of Theorem 2.

REFERENCES

[1] M. Skolnik, *Introduction to Radar Systems*. McGraw-Hill, 2001.

[2] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.

[3] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[4] J. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, "A random-finite-set approach to Bayesian SLAM," *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 268–282, April 2011.

[5] E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1016–1027, Aug. 2008.

[6] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[7] R. Mahler, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, October 2007.

[8] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. Artech House, 2014.

[9] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.

[10] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553–3567, July 2007.

[11] C. Lundquist, L. Hammarstrand, and F. Gustafsson, "Road intensity based mapping using radar measurements with a probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1397–1408, April 2011.

[12] B. Ristic, B.-N. Vo, and D. Clark, "A note on the reward function for PHD filters with sensor control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1521–1529, April 2011.

[13] H. Hoang, B.-N. Vo, B.-T. Vo, and R. Mahler, "The Cauchy-Schwarz divergence for Poisson point processes," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4475–4485, Aug. 2015.

[14] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[15] N. Whiteley, S. Singh, and S. Godsill, "Auxiliary particle implementation of probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1437–1454, July 2010.

[16] R. Sithiravel, X. Chen, R. Tharmarasa, B. Balaji, and T. Kirubarajan, "The spline probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6188–6203, Dec. 2013.

[17] D. Clark and R. Mahler, "Generalized PHD filters via a general chain rule," in *15th International Conference on Information Fusion*, July 2012, pp. 157–164.

[18] D. E. Clark and J. Houssineau, "Faà di Bruno's formula and spatial cluster modelling," *Spatial Statistics*, vol. 6, pp. 109–117, 2013.

[19] F. Caron, P. D. Moral, A. Doucet, and M. Pace, "On the conditional distributions of spatial point processes," *Advanced Applied Probability*, vol. 43, no. 2, pp. 301–307, 2011.

[20] S. S. Singh, B.-N. Vo, A. Baddeley, and S. Zuyev, "Filters for spatial point processes," *SIAM Journal of Control and Optimization*, vol. 48, no. 4, pp. 2275–2295, 2009.

[21] O. Erdinc, P. Willett, and Y. Bar-Shalom, "The bin-occupancy filter and its connection to the PHD filters," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4232–4246, Nov. 2009.

[22] P. Maybeck, *Stochastic models, estimation and control*. Academic Press, 1982, vol. 2.

[23] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[24] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[25] T. P. Minka, "Expectation propagation and approximate Bayesian inference," in *Proceedings of the 17th Conference in Uncertainty in Artifical Intelligence*, 2001, pp. 362–369.

[26] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and A. Graziano, "Consensus CPHD filter for distributed multitarget tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 508–520, June 2013.

[27] G. Battistelli and L. Chisci, "Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.

[28] J. Williams, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, Jan. 2015.

[29] J. F. C. Kingman, *Poisson processes*. Oxford University Press, 1993.

[30] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[31] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 1995.

[32] D. Daley and D. Vere-Jones, *An introduction to the theory of point processes: Volume I: Elementary Theory and methods*. Springer, 2003.

[33] B. Ristic, *Particle filters for random finite set models*. Springer, 2013.

[34] S. Challa, M. R. Morelande, D. Musicki, and R. J. Evans, *Fundamentals of Object Tracking*. Cambridge University Press, 2011.

[35] M. Mallick, V. Krishnamurthy, and B.-N. Vo, *Integrated Tracking, Classification, and Sensor Management*. John Wiley & Sons, 2013.

[36] M. Lundgren, L. Svensson, and L. Hammarstrand, "A CPHD filter for tracking with spawning models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 496–507, June 2013.

[37] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1656–1668, April 2012.

[38] R. Mahler, B.-T. Vo, and B.-N. Vo, "CPHD filtering with unknown clutter rate and detection profile," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3497–3513, Aug. 2011.

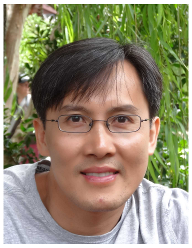
[39] D. Franken, M. Schmidt, and M. Ulmke, ""Spooky action at a distance" in the cardinalized probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 4, pp. 1657–1664, Oct. 2009.

[40] M. R. Morelande and A. F. García-Fernández, "Analysis of Kalman filter approximations for nonlinear measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5477–5484, Nov. 2013.

[41] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science + Business Media, 2006.

Ángel F. García-Fernández received the telecommunication engineering degree (with honours) and the Ph.D. degree from Universidad Politécnica de Madrid, Madrid, Spain, in 2007 and 2011, respectively.

He is currently a Research Associate in the Department of Electrical and Computer Engineering at Curtin University, Perth, Australia. His main research activities and interests are in the area of Bayesian nonlinear inference.



Ba-Ngu Vo received his Bachelor degrees jointly in Science and Electrical Engineering with first class honors in 1994, and PhD in 1997. He had held various research positions before joining the department of Electrical and Electronic Engineering at the University of Melbourne in 2000. In 2010, he joined the School of Electrical Electronic and Computer Engineering at the University of Western Australia as Winthrop Professor and Chair of Signal Processing. Since 2012 he is Professor and Chair of Signals and Systems in the Department of Electrical and Computer Engineering at Curtin University. Prof. Vo is a recipient of the Australian Research Council's inaugural Future Fellowship and the 2010 Australian Museum Eureka Prize for Outstanding Science in support of Defence or National Security. His research interests are Signal Processing, Systems Theory and Stochastic Geometry with emphasis on target tracking, robotics, computer vision and space situational awareness.