Set-membership PHD filter

Alessio Benavoli
Dalle Molle Institute for Artificial Intelligence (IDSIA),
Manno, Switzerland, alessio@idsia.ch

Francesco Papi
Joint Research Centre, IPSC, European Commission,
Ispra, Italy, francesco.papi@jrc.ec.europa.eu

***Abstract*— The paper proposes a novel Probability Hypothesis Density (PHD) filter for linear system in which initial state, process and measurement noises are only known to be bounded (they can vary on compact sets, e.g., polytopes). This means that no probabilistic assumption is imposed on the distributions of initial state and noises besides the knowledge of their supports. These are the same assumptions that are used in set-membership estimation. By exploiting a formulation of set-membership estimation in terms of set of probability measures, we derive the equations of the set-membership PHD filter, which consist in propagating in time compact sets that include with guarantee the targets' states. Numerical simulations show the effectiveness of the proposed approach and the comparison with a sequential Monte Carlo PHD filter which instead assumes that initial state and noises have uniform distributions.**

Keywords: multi-target tracking, Probability Hypothesis Density filter, set-membership estimation.

I. Introduction

The Probability Hypothesis Density (PHD) filter [1], [2] is an algorithm for tracking multiple targets in presence of missed detections and clutter. Two main implementations of the PHD filter have been proposed: the first uses sequential Monte Carlo methods [3], the second uses Gaussian mixtures [4]. In particular in [4] it is shown that, for linear Gaussian system and Gaussian birth model, it is possible to derive an analytic solution to the PHD recursion.

In this paper, we show that there is another case in which it is possible to derive an analytic solution to the PHD recursion. This is the case in which we have a linear system with bounded noises. This means that instead of assuming that initial state and noises are Gaussian, we assume to only know that they are bounded. Bounds can be expressed in terms of the supremum norm, 1-norm or more in general by polytopic constraints. This kind of model for the noises is used in *robust filtering* and in particular in *set-membership estimation*. Here, the term robust refers to the fact that the probability distributions of the noises are unknown - only the supports (membership-sets) are assumed to be known.

Set-membership techniques are based on the construction of a compact set that includes, with guarantee, the states of the system that are consistent with the measured output and the bounded noise. In [5], [6], an ellipsoidal bounding of the state of the dynamic system is provided. The application of ellipsoidal sets to filtering has also been studied by other authors, for example [7], [8]. In order to improve the estimation accuracy, the use of a convex polytope instead of an ellipsoid has been proposed in [9], [10]. Unfortunately such a polytope may be extremely complex and the corresponding polytopic updating algorithms may require an excessive amount of calculations and storage (without approximations the number of vertices of the polytope increases exponentially in time). For this reason, it has been suggested to outer approximate the true polytope with a simpler polytope, i.e. possessing a limited number of vertices. In this way, good approximation results can be obtained [11]. However, the complexity of the corresponding algorithms may still be too high and, in any case, is data-dependent. An alternative approach based on a parallelotopic approximation was presented in [12]. Notice that a parallelotope is a set described by an 1-norm bound and it better represents uncertainty expressed by componentwise bounds. Minimum-volume bounding parallelotopes are then used to estimate the state of a discrete-linear dynamical system with polynomial complexity [12].

Set-membership estimation is in general referred in literature as the deterministic approach to filtering, since its solution can be formulated in the realm of set-valued calculus and no stochastic calculations are necessary. Recently, it has been shown [13] that this is not completely true. Set-membership estimation can also be formulated in the realm of probability by considering set of distributions instead of a single distribution. In particular, by employing the theory of *Imprecise Probability* [14] and its application to the filtering problem [15], it can be shown that the prediction and updating steps in set-membership estimation can be reformulated by applying Chapman-Kolmogorov equation (for prediction) and Bayes' rule (for updating) to a particular set of probability measures, i.e., the set of all Dirac's deltas inside the membership-set.

By exploiting this result and the formulation of the PHD as Gaussian Mixture we have derived a *Set-Membership* based PHD filter. This *Set-Membership* based PHD filter has the following properties.
(1) The membership-sets that are computed at each time instant include with guarantee, the states of the targets that are consistent with the measured outputs and the bounded noises.
(2) The Set-Membership based PHD filter outperforms the SMC-PHD filter which uses uniform distributions as true distributions for initial state and noises.

It should be pointed out that the approach proposed in this paper is different from the box-particle PHD filter developed in [16] from previous work [17], [18]. This difference is mainly due to the fact that set-membership estimation is different from box-particle filtering [19]. In [20] in fact it has been shown that the box-particle filter can be interpreted in the

Bayesian filtering framework as a mixture of uniform probability density functions (PDF). Set-membership estimation cannot be interpreted in the Bayesian framework, but only in the framework of Imprecise probability (filtering with set of probability measures). We discuss with more details this differences in Section V.

II. Set-membership estimation

Consider the following linear, time-invariant, discrete-time system

$$\left\{ \begin{array}{rcl} x_k & = & Fx_{k-1} + \omega_{k-1}, \\ z_k & = & Hx_k + \lambda_k, \end{array} \right. \tag{1}$$

where: k is the time; $x_k \in \mathbb{R}^n$ is the state; $z_k \in \mathbb{R}^p$ is the measured output; $\omega_{k-1} \in \mathbb{R}^n$ is the process noise; $\lambda_k \in \mathbb{R}^p$ is the measurement noise; F and H are matrices of compatible dimensions. In Kalman filtering, the initial state and noise signals are assumed to be:

$$\begin{array}{rcl} p(\omega_{k-1}) & = & \mathcal{N}(\omega_{k-1}; \hat{\omega}, Q), \\ p(\lambda_k) & = & \mathcal{N}(\lambda_k; \hat{\lambda}, R), \\ p(x_0) & = & \mathcal{N}(x_0; \hat{x}_0, P_0), \end{array} \tag{2}$$

where $\mathcal{N}(\rho; \mu, \Sigma)$ denotes that ρ is a Gaussian random variable with mean μ and variance Σ. Note that in (2) the noises have a mean different from zero. We exploit this fact to derive the following result.

Lemma 1. *Consider the process noise and assume that*

$$\hat{\omega} \in \Omega, \tag{3}$$

where Ω is some polytope in $\mathbb{R}^n$, i.e., we only know that the mean of the process noise belongs to the set Ω, then for $Q \to 0$ (i.e., Q is scaled by a constant ϵ and $\epsilon \to 0$), one has that

$$\{\mathcal{N}(\omega; \hat{\omega}, Q) : \ \hat{\omega} \in \Omega\} \overset{Q \to 0}{=} \{\delta_{\hat{\omega}}(\omega) : \ \hat{\omega} \in \Omega\},$$

where $\delta_{\hat{\omega}}$ denotes a Dirac's delta centred at $\hat{\omega}$. □

The proof is obvious and omitted. Lemma 1 states that if we only know that the mean of the process noise belongs to a polytope Ω and we take the limit for $Q \to 0$ of the set of Gaussian PDFs with means varying in this polytope, then we obtain the set of all Dirac's deltas on Ω. Note that the closed convex set containing all Dirac's deltas on Ω and the closed set containing all (finitely additive) probabilities with support on Ω are equivalent.[1] Thus, from Lemma 1, it also follows that the limit of the set of Gaussian PDFs with means in Ω is equivalent to the set of all probability measures with support on Ω. We can state this result differently by saying that the only information about the variable ω_{k-1} is that it takes values in Ω. This is the condition which is assumed for initial state and noises in set-membership estimation.

The aim of this section is to show that using the same limit procedure described in Lemma 1 we can recover the formulas of set-membership estimation. This is not a rigorous proof, we point the reader to [13] for a more rigorous derivation. Let us start with the prediction step.

[1]With equivalent we mean that they give the same lower and upper expectations w.r.t. real-valued function defined on Ω [14, Sec. 3.6].

Theorem 1. *Assume that the posterior distribution of the state estimate at time $k-1$ belongs to the set*

$$\left\{\mathcal{N}(x_{k-1}; \hat{x}_{k-1}, P_{k-1}) : \ \hat{x}_{k-1} \in \hat{X}_{k-1}\right\}, \tag{4}$$

where $\hat{X}_{k-1}$ is a polytope in $\mathbb{R}^n$ and consider the state equation in (1) in which the distribution of the process noise belongs to the set

$$\{\mathcal{N}(\omega_{k-1}; \hat{\omega}, Q) : \ \hat{\omega} \in \Omega\}.$$

By taking the limit for $(FP_{k|k-1}F^T + Q) \to 0$,[2] we obtain the set of predicted distributions of the set-membership estimate at time k, i.e.,:

$$\left\{\delta_{\hat{x}_{k|k-1}}(x_{k|k-1}) : \ \hat{x}_{k|k-1} \in \hat{X}_{k|k-1}\right\}, \tag{5}$$

with

$$\hat{X}_{k|k-1} = F\hat{X}_{k-1} \oplus \Omega, \tag{6}$$

where $\oplus$ denotes the Minkowski sum of two sets and $F\hat{X}_{k-1}$ denotes the product of the matrix F for the elements of the set $\hat{X}_{k-1}$. □

Proof: Choose a value for $\hat{x}_{k-1}$ in $\hat{X}_{k-1}$ and for $\hat{\omega}$ in Ω and apply Kalman filtering equations for non-zero mean process noise. The state prediction is then $\mathcal{N}(x_{k|k-1}; \hat{x}_{k|k-1}, P_{k|k-1})$ where $\hat{x}_{k|k-1} = F\hat{x}_{k-1} + \hat{\omega}$ and $P_{k|k-1} = FP_{k|k-1}F^T + Q$. By taking the limit for $(FP_{k|k-1}F^T + Q) \to 0$ of the predictive Gaussian, we obtain $\delta_{\hat{x}_{k|k-1}}(x_{k|k-1})$. Finally, by repeating the previous step for all values of $\hat{x}_{k-1}$ in $\hat{X}_{k-1}$ and $\hat{\omega}$ in Ω, we can prove the Theorem. □

Thus, at the end of the prediction step we only know that the value of the state prediction variable $x_{k|k-1}$ belongs to the set $\hat{X}_{k|k-1}$ in (6).

Theorem 2. *Assume that the predicted distribution of the state estimate at time $k|k-1$ belongs to the set*

$$\left\{\mathcal{N}(x_{k|k-1}; \hat{x}_{k|k-1}, P_{k|k-1}) : \ \hat{x}_{k|k-1} \in \hat{X}_{k|k-1}\right\}, \tag{7}$$

where $\hat{X}_{k|k-1}$ is a polytope in $\mathbb{R}^n$ and consider the measurement equation in (1) in which the distribution of the measurement noise is assumed to belong to the set

$$\left\{\mathcal{N}(\lambda_k; \hat{\lambda}, R) : \ \hat{\lambda} \in \Lambda\right\}.$$

Define $K_k = P_{k|k-1}H^T S_k^{-1}$, $S_k = HP_{k|k-1}H^T + R$, $P_{k|k} = (I - K_k H)P_{k|k-1}$ By taking the limit for $S_k, P_{k|k} \to 0$ of KF updating equation, we obtain the set of updated distributions of the set-membership estimate at time k:

$$\left\{\delta_{\hat{x}_k}(x_k) : \ \hat{x}_k \in \hat{X}_k\right\}, \tag{8}$$

with

$$\hat{X}_k = \left\{x \in \hat{X}_{k|k-1} \ \text{s.t.} \ Hx \in \{z_k - \hat{\lambda} : \ \hat{\lambda} \in \Lambda\}\right\}. \tag{9}$$

□

[2]Again we can assume that $P_{k|k-1}, Q$ are scaled by a constant ϵ and $\epsilon \to 0$.

Proof: The KF updated posterior distribution is $\mathcal{N}(x_k; \hat{x}_k, P_k)$ where $\hat{x}_k = \hat{x}_{k|k-1} + K_k(z_k - \hat{\lambda} - H\hat{x}_{k|k-1})$. The above Gaussian has been obtained by applying Bayes'rule, i.e.,

$$\mathcal{N}(x_k; \hat{x}_k, P_k) \propto \frac{\mathcal{N}(z_k - \hat{\lambda}; H\hat{x}_{k|k-1}, R)\mathcal{N}(x_{k|k-1}; \hat{x}_{k|k-1}, P_{k|k-1})}{\mathcal{N}(z_k - \hat{\lambda}; H\hat{x}_{k|k-1}, S_k)}.$$

A condition to apply Bayes'rule is that the denominator must be positive; we must meet this condition when we take the limit $S_k \to 0$. By considering only the exponential kernel of $\mathcal{N}(z_k - \hat{\lambda}; H\hat{x}_{k|k-1}, S_k)$, this condition can be satisfied at the limit provided that $z_k - \hat{\lambda} = H\hat{x}_{k|k-1}$, since $\exp(-(z_k - \hat{\lambda} - H\hat{x}_{k|k-1})^T S_k^{-1}(z_k - \hat{\lambda} - H\hat{x}_{k|k-1})) = 1$ in this case.[3] This means that, given z_k and $\hat{\lambda}$ in Λ, we cannot arbitrarily choose $\hat{x}_{k|k-1}$ in $\hat{X}_{k|k-1}$ but we must choose this value to satisfy the positiveness of the denominator.[4] There exist values that verify this condition iff:

$$H\hat{X}_{k|k-1} \cap \{z_k - \hat{\lambda} : \ \hat{\lambda} \in \Lambda\} \neq \emptyset.$$

Then assuming that there exists some $\hat{x}_{k|k-1}$ which satisfies the above condition and if we take $z_k - \hat{\lambda} = H\hat{x}_{k|k-1}$ we obtain

$$\begin{aligned} \hat{x}_k &= \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) - K_k\Lambda \\ &= \hat{x}_{k|k-1} + K_k(H\hat{x}_{k|k-1} - H\hat{x}_{k|k-1}) = \hat{x}_{k|k-1}, \end{aligned} \tag{10}$$

and this holds for all $\hat{x}_{k|k-1}$ such that $z_k - \hat{\lambda} = H\hat{x}_{k|k-1}$. Then by taking the limit $P_{k|k} \to 0$ we obtain the set of Dirac's deltas with support in (9). □

Theorems 1 and 2 give the theoretical solution of the problem of set-membership estimation, i.e., to compute the support of the posterior distribution of x_k given all the past observations under the assumption that the only information about the distributions of initial state, process and measurement noises is their support. To practically solve the prediction and updating steps, it is in necessary to assume that the borders of the supports of initial state and noises can be described by simple shapes, for instance polytopes. In this case, the prediction and updating steps reduce to propagate in time the vertices or the linear constraints that characterize these polytopes [11], [12].

Example 1. Consider the following example

$$F = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \quad H = [1, \ 0]$$

$\Omega = Co\{[-0.5, -0.1]^T, [-0.5, 0.1]^T, [0.5, -0.1]^T, [0.5, 0.1]^T\}$, $\hat{X}_0 = Co\{[-1, 0.3]^T, [-1, 0.7]^T, [1, 0.3]^T, [1, 0.7]^T\}$, $\Lambda = [-0.3, 0.3]$, *where Co denotes convex hull (note that* $\Omega, \hat{X}_0$ *are boxes centred at* $[0, 0]^T$*). Fig. 1 shows the sets* $F\hat{X}_0$ *(dashed line) and* $\hat{X}_{1|0} = F\hat{X}_0 \oplus \Omega$ *(solid line) for the first* (x_1) *and second* (x_2) *component of the state. The updated membership-set* $\hat{X}_1$ *is shown in Fig. 1 (red thick line) - it has been computed assuming the observation* $z_1 = 3.5$. □

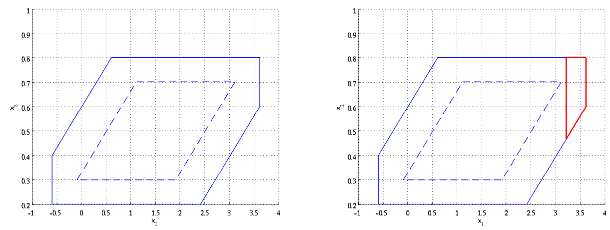

Fig. 1. Predicted and updated membership-set.

Form the previous example, it can be noticed that the prediction and updating steps change the shape of the membership-set $\hat{X}_k$. As the time increases, this set becomes more and more complicated (many vertices and facets). Therefore, in order to reduce the complexity of the prediction and updating steps, at each time instant it is necessary to approximate the membership-set with a simpler shaped region. Usually, outer-approximating regions are provided by ellipsoids or by polytopes with predefined shape (fixed number of vertices or parallel edges), e.g., [12].

III. The Gaussian mixture PHD filter

Hereafter we briefly review the Gaussian mixture PHD filter; we point the reader to [4] for more details. In the the Gaussian mixture PHD filter it is assumed that (1) the state dynamics and measurement equation are given by (1); (2) the survival p_s and detection probability p_d are state independent; (3) the clutter is Poisson and independent of target-originated measurements; (4) the intensity of the birth process is a Gaussian mixture:

$$\gamma_k(x) = \sum_{i=1}^{J_{\gamma,k}} w_{\gamma,k}^{(i)} \mathcal{N}(x; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}),$$

where $m_{\gamma,k}^{(i)}$ are the peaks of the birth intensity (the locations where new targets are more likely to appear), $P_{\gamma,k}^{(i)}$ determines the spread of the birth intensity, the weight $w_{\gamma,k}^{(i)}$ gives the expected number of new targets originated from $m_{\gamma,k}^{(i)}$.

Suppose that these assumptions hold and that the posterior intensity at time $k-1$ is a Gaussian mixture of the form

$$v_{k-1}(x) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \mathcal{N}(x; m_{k-1}^{(i)}, P_{k-1}^{(i)}), \tag{11}$$

then the predicted intensity for time k is also a Gaussian mixture

$$v_{k|k-1}(x) = v_{S,k|k-1}(x) + \gamma_k(x), \tag{12}$$

where

$$v_{S,k|k-1}(x) = p_s \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \mathcal{N}(x; m_{S,k|k-1}^{(i)}, P_{S,k|k-1}^{(i)}), \tag{13}$$

$$m_{S,k|k-1}^{(i)} = F m_{k-1}^{(i)}, \qquad P_{S,k|k-1}^{(i)} = F P_{k-1}^{(i)} F^T + Q. \tag{14}$$

[3] If $z_k - \hat{\lambda} \neq H\hat{x}_{k|k-1}$, $\exp(-(z_k - \hat{\lambda} - H\hat{x}_{k|k-1})^T S_k^{-1}(z_k - \hat{\lambda} - H\hat{x}_{k|k-1})) \to 0$ for $S_k \to 0$.

[4] The application of Bayes' rule only to the probabilities that assign positive mass (density) to the observation is called by Walley "regular extension" [14, Appendix J].

Observe that $v_{k|k-1}(x)$ is again a Gaussian mixture, which can be rewritten as

$$v_{k|k-1}(x) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} \mathcal{N}(x; m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}). \qquad (15)$$

Then, let $Z_k = \{z_1, \ldots, z_{l_k}\}$ be the set of observations at time k, then the posterior intensity at time k is also a Gaussian mixture and is given by:

$$v_k(x) = (1 - p_d) v_{k|k-1}(x) + \sum_{z \in Z_k} v_{D,k}(x; z), \qquad (16)$$

with

$$v_{D,k}(x; z) = \sum_{i=1}^{J_{k|k-1}} w_k^{(i)} \mathcal{N}(x; m_k^{(i)}, P_k^{(i)}), \qquad (17)$$

$$w_k^{(i)} = \frac{p_d w_{k|k-1}^{(i)} q_k^{(i)}(z)}{\kappa_k(z) + p_d \sum_{j=1}^{J_{k|k-1}} w_{k|k-1}^{(j)} q_k^{(j)}(z)}, \qquad (18)$$

$$\begin{aligned} q_k^{(i)}(z) &= \mathcal{N}(z; H m_{k|k-1}^{(i)}, R + H P_{k|k-1}^{(i)} H^T), \qquad (19)\\ m_k^{(i)} &= m_{k|k-1}^{(i)} + K_k(z - H m_{k|k-1}^{(i)}),\\ P_k^{(i)} &= (I - K_k H) P_{k|k-1}^{(i)}, \quad K_k^{(i)} = P_{k|k-1}^{(i)} H^T S_k^{-1},\\ S_k^{(i)} &= H P_{k|k-1}^{(i)} H^T + R, \end{aligned}$$

Note that $\kappa_k(z)$ is the clutter intensity which is in general assumed to be:

$$\kappa_k(z) = \phi_c V u(z), \qquad (20)$$

$u(\cdot)$ is the uniform density over the surveillance region, V is the volume of the surveillance region and ϕ_c is the average number of clutter returns per unit volume.

It should be noticed that because of (12) and (16) the number of components of the Gaussian mixture increases at each time step. Pruning and merging procedures are thus necessary for computational feasibility. A pruning and merging algorithm is described in [4] together with an algorithm to extract the peaks of the mixture. This latter algorithm is used to provide a point estimate of the target locations.

IV. Set-membership PHD filter

The aim of this section is to derive a PHD filter in case we assume that

1) initial state, measurement and process noise are modelled by a set of Dirac's deltas over polytopic regions;
2) the clutter distribution $u(\cdot)$ in the surveillance region U is
$$u(z) \in \{\delta_{\hat{z}}(z) : \hat{z} \in U\}; \qquad (21)$$
3) the set of intensity of the birth process is $\Gamma_{s,k}(x) = \{\gamma_{s,k}(x) : \forall s\}$ with:
$$\gamma_{s,k}(x) = \sum_{i=1}^{J_{\gamma,k}} w_{\gamma,k}^{(i)} \delta_{m_{\gamma,k}^{(i,s_i)}}(x), \quad m_{\gamma,k}^{(i,s_i)} \in M_{\gamma,k}^{(i)},$$
where s_i indexes the choice of one element of the membership set $M_{\gamma,k}^{(i)}$ and $s = \{s_1, \ldots, s_{J_{\gamma,k}}\}$ the set of all the choices in each membership-set. Thus, $\Gamma_{s,k}(x)$ is the set of all possible mixture of Dirac's deltas that we obtain by selecting $m_{\gamma,k}^{(i,s_i)}$ to be any of the point in the membership set $M_{\gamma,k}^{(i)}$ for all elements of the mixture $i = 1, \ldots, J_{\gamma,k}$.

Theorem 3. *Suppose that the above assumptions hold and that the set of posterior intensity at time $k-1$ is $V_{k-1}(x) = \{v_{s,k-1}(x) : \forall j\}$ with*

$$v_{s,k-1}(x) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} \delta_{m_{k-1}^{(i,s_i)}}(x), \quad m_{k-1}^{(i,s_i)} \in M_{k-1}^{(i)}, \qquad (22)$$

then the set of predicted intensity for time k is $V_{k|k-1}(x) = \{v_{s,k|k-1}(x) : \forall j\}$ where

$$v_{s,k|k-1}(x) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} \delta_{m_{k|k-1}^{(i,s_i)}}(x), \quad m_{k|k-1}^{(i,s_i)} \in M_{k|k-1}^{(i)}, \qquad (23)$$

with $J_{k|k-1} = J_{k-1} + J_{\gamma,k}$,

$$M_{k|k-1}^{(i)} = \begin{cases} F M_{k-1}^{(i)} \oplus \Omega & \text{for } i = 1, \ldots, J_{k-1}, \\ M_{\gamma,k-1}^{(i)} & \text{for } i = J_{k-1}+1, \ldots, J_{k-1}+J_{\gamma,k}. \end{cases} \qquad (24)$$

$$w_{k|k-1}^{(i)} = \begin{cases} p_s w_{k-1}^{(i)} & \text{for } i = 1, \ldots, J_{k-1}, \\ w_{\gamma,k}^{(i)} & \text{for } i = J_{k-1}+1, \ldots, J_{k-1}+J_{\gamma,k}. \end{cases} \qquad (25)$$

□

Proof: The result follows directly from the proof of Theorem 1 and (11)–(15) exploiting the limit procedure described in Theorem 1. Observe that in (24) new membership-sets are added to account for the birth of new targets. □

Theorem 4. *Suppose that the set of predicted posterior intensity at time k is $V_{k|k-1}(x) = \{v_{s,k|k-1}(x) : \forall j\}$ with $v_{s,k|k-1}(x)$ given in (23). Let $Z_k = \{z_1, \ldots, z_{l_k}\}$ be the set of observations at time k, then the set of posterior intensity at time k is $V_k(x) = \{v_{s,k}(x) : \forall j\}$ with*

$$v_{s,k}(x) = \sum_{i=1}^{J_k} w_k^{(i)} \delta_{m_k^{(i,s_i)}}(x), \quad m_k^{(i,s_i)} \in M_k^{(i)}, \qquad (26)$$

where $J_k = (1 + |Z_k|) J_{k|k-1}$,

$$M_k^{(i)} = \begin{cases} M_{k-1}^{(i)}, i = 1, \ldots, J_{k|k-1}, \\ M_{z_1,k|k-1}^{(i)}, i = J_{k|k-1}+1, \ldots, 2J_{k|k-1}, \\ \ldots \\ M_{z_{l_k},k|k-1}^{(i)}, i = |Z_k| J_{k|k-1}+1, \ldots, |Z_k+1| J_{k|k-1}, \end{cases} \qquad (27)$$

with $M^{(i)}_{z_r,k|k-1} = \{M^{(i)}_{k|k-1} : \ HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}$*,* $\hat{Z}_r = \{z_r - \hat{\lambda} : \ \hat{\lambda} \in \Lambda\}$*,*

$$w^{(i)}_k = \begin{cases} (1-p_d)w^{(i)}_{k|k-1} \ \ \textit{for } i = 1,\dots,J_{k|k-1}, \\ w^{(i)}_{z_1,k} \ \ \textit{for } i = J_{k|k-1}+1,\dots,2J_{k|k-1}, \\ \cdots \\ w^{(i)}_{z_{l_k},k} \ \ \textit{for } i = |Z_k|J_{k|k-1}+1,\dots,|Z_k+1|J_{k|k-1}, \end{cases} \tag{28}$$

and

$$w^{(i)}_{z_r,k} = \frac{p_d I_{\{HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}}}{\phi_c V + p_d \sum_{l=1}^{J_{k|k-1}} I_{\{HM^{(l)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}}} \tag{29}$$

where $I_{\{HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}}$ *is the indicator function which is one when its argument is satisfied and zero otherwise.* □

Proof: The result follows directly from the proof of Theorem 2 and (15)–(20). We discuss the derivation of (29) with more details. From Theorem 2, we know that to apply Bayes'rule we must ensure that the denominator $\mathcal{N}(z_r - \hat{\lambda}; Hm^{(i)}_{k|k-1}, S^{(i)}_k)$ is positive at the limit $S^{(i)}_k \to 0$, which implies that

$$z_r - \hat{\lambda} = Hm^{(i)}_{k|k-1}.$$

This means that, given z_k and $\hat{\lambda}$ in Λ, we cannot arbitrarily choose $m^{(i)}_{k|k-1}$ in $M^{(i)}_{k|k-1}$ but we must choose this value to satisfy the above condition. There exist values that verify this condition iff:

$$HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset.$$

From (19), this implies that

$$q^{(i)}_k(z_r) = \begin{cases} \dfrac{1}{\sqrt{det(2\pi S^{(i)}_k)}} & \text{if } HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset \\ 0 & \text{other.} \end{cases}$$

the first row is obtained by taking $z_r - \hat{\lambda} = Hm^{(i)}_{k|k-1}$. Thus, (18) is equal to

$$w^{(i)}_{z_r,k} = \frac{p_d I_{\{HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}}}{\kappa_k(z_r)\sqrt{det(2\pi S_k)} + p_d \sum_{l=1}^{J_{k|k-1}} I_{\{HM^{(l)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}}}, \tag{30}$$

where we have assumed that all the $S^{(i)}_k$ are equal to S_k. Now we must take the limit $S_k \to 0$. Since $\kappa_k(z_r) = \phi_c V u(z_r)$, $u(z_r) \in \{\delta_{\hat{z}}(z_r), \quad \hat{z} \in U\}$, if $\delta_{\hat{z}}(z_r)$ is obtained as limit for $S_k \to 0$ of $\mathcal{N}(z_r; \hat{z}, S_k)$, then we obtain (29) by taking $\hat{z} = z_r$.

□

It is worth to point our the non-obvious differences between the linear Gaussian mixture PHD and the set-membership PHD filter.

1) In the set-membership PHD filter, the prediction and updating steps practically consist on computing $M^{(i)}_{k|k-1}, w^{(i)}_{k|k-1}$ and $M^{(i)}_k, w^{(i)}_k$.[5]
2) Because of the boundedness of the process and measurement noise, the set-membership PHD filter naturally performs a gating procedure on the measurements: if $I_{\{HM^{(i)}_{k|k-1} \cap \hat{Z}_r \neq \emptyset\}} = 0$, the weight $w^{(i)}_{z_r,k}$ is zero. This means that, after the updating step, the intensity mixture $v_{s,k}(x)$ may practically have much less components than $J_k = (1 + |Z_k|)J_{k|k-1}$.
3) The set-membership PHD filter always guarantees that the true values of the target states are included in some of the polytopes $M^{(i)}_k$.
4) As for the linear Gaussian mixture PHD, the sum of the weights of the mixture gives the number of expected targets.

[5]The expression of the predicted and updated intensity as mixture of Dirac's is used only in the proof, for the practical implementation we just propagate the membership-sets.

A. Pruning, merging, bounding

Although the gating procedure of the membership-set guarantees that the set-membership PHD filter has less components than the linear Gaussian mixture PHD, a reduction of the number of terms of the mixture may be necessary for $p_d < 1$ and large number of false alarms. As for the the linear Gaussian mixture PHD, we have implemented three reduction strategies. First, there is a pruning strategy which deletes all the terms in the mixture in $v_{s,k}(x)$ whose weights $w^{(i)}_k$ are less than a fixed threshold. Second, there is a merging strategy, which combines the polytopes that are close. We have used the Hausdorff metric to measure the distance of two polytopes. If this distance is less than a fixed threshold, the polytope with smaller weight $w^{(r)}_k$ is deleted and the weight of the other polytope is incremented by $w^{(r)}_k$. Finally, the total number of terms of the mixture cannot exceed a fixed maximum value. It it does, the exceeding terms with smaller weights are deleted.

Another issue concerns the number of vertices of the polytopes of the membership sets. In this paper, we have used a bounding-box approximation to prevent that the polytopes become too complex. The set-membership filter has been implemented using the routines for computational geometry of the Multi-Parametric Toolbox [21].

V. Difference with the Box Particle filter approach

The aim of this section is to point out the differences between the set-membership PHD filter proposed in this paper and the Box Particle PHD filter presented in [16].

We start this comparison by first highlighting the differences between Box Particle filter (Box-PF) and set-membership estimation. Box-PF has been proposed in [19] with the goal of reducing the computational complexity in PF. It is well known that PF approximates the posterior PDF of the state by a mixture of Dirac's deltas centred on the computed particles. The idea of the Box-PF is to approximate this PDF by a mixture of uniform distributions with box supports [20]. Then, assuming that the process and measurement noises are also enclosed in boxes, at each time instant the predicted and updated supports of the boxes in the mixture are computed by using interval-analysis and the weights of the mixture are updated using Bayes' rule under the further assumptions that the noises have uniform distributions (with box supports). With this latter assumption, Box-PF can be completely formulated in the realm of Bayesian filtering [20], since the distributions of the noises are completely known. In this context, the advantage of Box-PF is to use interval-analysis to perform prediction/updating and, thus, to obtain a good approximation of the solution of Bayesian filtering by using a relatively small number of box particles.

Conversely, in set-membership estimation no assumption is imposed on the distributions of the noises (the distributions of the noises are unknown - only their supports are known). The lack of information on the distributions of the noises can be modelled by considering the set of all distributions which are zero outside the given supports.[6] In this case, we cannot use Bayesian filtering to perform prediction and updating but we need to work in the realm of imprecise probability. The general solution of the filtering problem in this case has been presented in [15] and it has also been shown that the imprecise probability based filter coincides with set-membership estimation in case only the supports of the noises are known [13]. This result has been exploited in Section II to derive set-membership estimation formulas in a more direct (but less rigorous) way.

Set-membership estimation guarantees with probability one that the posterior membership set computed at each time instant always includes the true state. This is the main difference between the two approaches - Box-PF does not guarantee such inclusion. Because of the resampling step in Box-PF, it may happen that a component in the mixture of uniform distributions is discarded because it has a relatively small weight. The union of the boxes of the Box-PF may not include with certainty the true value of the state and, thus, they are not supports (in the probabilistic sense) but they are credible regions in the Bayesian filtering sense, i.e., regions that includes the value of the state with a given probability (e.g., 90%, 95% etc., the value depends on the resampling strategy and on the number of particles). These main differences also appear in the implementation for the PHD filter.

VI. Numerical simulations

In this section, the performance of the proposed set-membership PHD filter is assessed by means of Monte Carlo simulations, concerning different scenarios. The simulated targets move in the xy Cartesian plane and are, therefore, characterized, at discrete time k, by the state vector $x_t = [p_x, v_x, p_y, v_y]^T$, where (p_x, p_y) provides the position and (v_x, v_y) the velocity in Cartesian coordinates at time t. The following motion model has been considered for the targets:

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{31}$$

where $T = 1$ is the sampling interval;

$$\begin{aligned} \omega_k &\in \Omega = Box([-1,1],[-0.1,0.1],[-1,1],[-0.1,0.1]), \\ \lambda_k &\in \Lambda = Box([-1,1],[-1,1]), \end{aligned}$$

where $Box([-1,1],[-0.1,0.1],[-1,1],[-0.1,0.1])$ denotes a Box in which the first component of ω_k is bounded in the interval $[-1,1]$, the second in $[-0.1,0.1]$ etc. (similar for λ_k). For the initial state, we have assumed that $x_0 \in \hat{X}_0 = \hat{x}_0 + \Omega$, where $\hat{x}_0 = [10,5,80,2]^T$ for target 1, $\hat{x}_0 = [80,2,10,5]^T$ for target 2. We are considering a scenario with 2 targets, trajectories' length of 120 time instants, number of clutter measurements $V\phi_c \sim poiss(20)$ (on average 20 clutter measurements at each time instant), detection probability $p_d = 0.95$, and 100 Monte Carlo (MC) runs.

To evaluate the performance of the proposed algorithm, we have used the OSPA distance [22], and inclusion metric. Briefly, denote by $d^{(c)}(x,y) := \min(c, d(x,y))$ the distance between $x, y \in W \subset \mathbb{R}^N$ with cut off at c, and by Π_k the set of permutations on $\{1,\ldots,k\}$ for any $k \in \mathbb{N}$. For $1 \le p \le \infty$, $c > 0$, and arbitrary finite subsets $X = \{x_1,\ldots,x_n\}$ and $Y = \{y_1,\ldots,y_m\}$ of W, where $m, n \in \mathbb{N}_0 = \{0,1,2,\ldots\}$, then the OSPA distance is defined as:

$$d_p^{(c)}(X,Y) := \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m) \right) \right)^{1/p} \tag{32}$$

if $m \le n$, and $d_p^{(c)}(X,Y) := d_p^{(c)}(Y,X)$ if $m > n$.[7] Observe that, set-membership estimation does not return a point estimate but a set-estimate (the set that includes the target's state), so to compute the OSPA metric we must extract a point estimate. As point estimate we have selected the centre of the minimum-volume ellipsoid outer-bounding the membership-set. The inclusion is a binary variable taking value 1 if the target's state at time k is included in some of the membership-sets and value 0 otherwise. The inclusion has been averaged over the number of targets, and the number of MC trials.

For comparison, we have also implemented a SMC-PHD filter, which assumes that initial state, process and measurement noises have uniform distribution inside the respective membership-sets $\hat{X}_0$, Ω and Λ. Also for this filter, we have computed the OSPA metric and the inclusion metric. The inclusion has been computed by considering credible regions around the extracted estimates, which are obtained following the approach described in [23]. Specifically, to calculate the inclusion metric, we construct polytopes from the particle sets used to extract the estimates, and then assign value 1 if the

[6]The uniform distribution is one of such distributions but it is not the only one. For instance, a Dirac's delta on one of the vertices of the box also belongs to this set and represents the most critical case in which the noise is maximum.

[7]In multi-target tracking, x_i, y_i are respectively the estimated and true positions of the targets.

target's state is inside any of the constructed polytopes or zero otherwise.

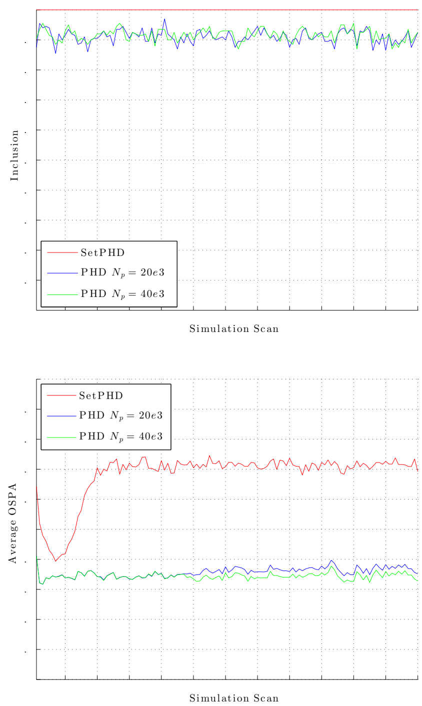

Fig. 2. Inclusion and OSPA (uniform noise, $p_d = 1$).

In the first scenario we assume a unitary detection probability, $p_d = 1$, and that process and measurement noises are uniformly distributed. That is, the scenario perfectly fits the assumptions underlying the SMC-PHD filter. Results are reported in Figs. 2 in terms of inclusion and OSPA distance. The estimated number of targets is not reported since in this case both filters perfectly estimate this parameter. As expected, the SMC-PHD filter can cope with the scenario and guarantees a lower OSPA distance than the set-membership PHD.

Results from a second scenario are reported in figs. 3. Here we have a unitary detection probability, $p_d = 1$, maximum process noise, i..e., $\omega_k = [1, 0.1, 1, 0.1]^T$, (in this way the two targets cross their trajectories at about time $k = 10$) and measurement noise uniformly distributed over the positive subset of Λ, i.e., $\lambda_k = u[1, 1]^T$ with u is uniform distributed in $[0, 1]$. We compare the set-membership PHD filter against 4 slightly different implementations of the SMC-PHD filter. As detailed in the legend, we consider different numbers of particles and process noise intensity for the SMC-PHD in order to guarantee a sufficient coverage of the posterior PHD, which is more critical in the case of maximum process noise intensity. Here, $Q_c = 2\Omega$ means that the SMC-PHD is using a uniform distribution for the process noise with a support that is twice larger than the true one used to generate the trajectories. The inclusion metric and OSPA distance show that the set-membership PHD filter in this case obtains a better performance. This means that the uniform distribution assumption in SMC-PHD filter cannot cope with a deterministic process noise (which is constantly equal to the allowed maximum value). The SMC-PHD filter with uniform distribution is thus not robust to the choice of the process noise: it performs well when the process noise is (close to) uniform but degrades its performance (in some case it diverges for $N_p = 20000$) in the case the process noise is maximum. Conversely, the inclusion performance of the set-membership approach are not affected by the choice of the process noise.

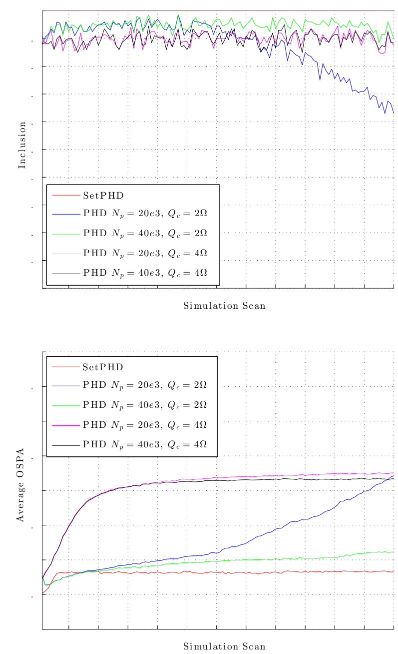

Fig. 3. Inclusion and OSPA (maximum noise, $p_d = 1$).

A more interesting scenario is obtained considering a non-unitary detection probability, $p_d = 0.95$ and the same conditions as in the second simulation case above. In Figs. 4 we have reported the estimated number of targets, the inclusion metric, and the OSPA distance. It is immediate to verify that the proposed set-membership PHD filter is still better than the SMC-PHD filter. The fact that the inclusion metric is not one for the set-membership estimation in this case, it due to the way we extract the targets' tracks to compute the metrics. In the simulations, we have extracted the best $\hat{N}_t$ tracks, where $\hat{N}_t$ is the number of estimated target. This track extraction is not very efficient. We have verified that the set-membership PHD filter in practice has inclusion equal to one also in this case, but we have not yet found a good algorithm to extract the track/tracks that include/s the targets.

VII. Conclusions

The paper has proposed a Probability Hypothesis Density filter for linear system in which initial state, process and measurement noises are only known to be bounded. No probabilistic assumption is imposed on the distributions of initial state and noises and set-membership estimation is used to

Fig. 4. Estimated number of targets, inclusion, OSPA (maximum noise, $p_d = 0.95$).

propagate in time the compact sets that include with guarantee the targets' states. Numerical simulations have shown that the proposed approach outperforms a sequential Monte Carlo PHD filter which instead assumes that initial state and noises have uniform distributions. As future work, it may be interesting to implement set-membership estimation in terms of parallelotopes, which provides a smaller outer-approximation of the membership-set than the bounding-boxes used in this paper. New algorithms for tracks' extraction should also be derived to take full advantage of the guarantee of inclusion of the set-membership approach. It may also be interesting to practically compare the set-membership PHD with the Box Particle PHD filter.

Acknowledgements

This work has been partially supported by the Swiss NSF grant n. 200020-137680/1

References

[1] I. R. Goodman, R. P. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Norwell, MA, USA: Kluwer Academic Publishers, 1997.

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

[3] B.-N. Vo, S. Singh, and A. Doucet, "Sequential monte carlo implementation of the phd filter for multi-target tracking," in *Information Fusion, 2003. Proceedings of the Sixth International Conference of*, vol. 2, pp. 792–799, 2003.

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

[5] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," in *Adaptive Processes, Sixth Symposium on*, vol. 6, pp. 102 –107, oct. 1967.

[6] D. Bertsekas and I. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *Automatic Control, IEEE Transactions on*, vol. 16, pp. 117 – 128, apr 1971.

[7] V. Kuntsevich and M. Lychak, *Guaranteed estimates, adaptation and robustness in control systems*. Springer-Verlag, 1992.

[8] A. Savkin and I. Petersen, "Robust state estimation and model validation for discrete-time uncertain systems with a deterministic description of noise and uncertainty," *Automatica*, vol. 34, no. 2, pp. 271–274, 1998.

[9] H. Piet-Lahanier and E. Walter, "Further results on recursive polyhedral description of parameter uncertainty in the bounded-error context," in *Decision and Control, 1989., Proceedings of the 28th IEEE Conference on*, pp. 1964 –1966 vol.3, dec 1989.

[10] S. Mo and J. Norton, "Fast and robust algorithm to compute exact polytope parameter bounds," *Mathematics and computers in simulation*, vol. 32, no. 5-6, pp. 481–493, 1990.

[11] V. Broman and M. Shensa, "A compact algorithm for the intersection and approximation of N-dimensional polytopes," *Mathematics and computers in simulation*, vol. 32, no. 5-6, pp. 469–480, 1990.

[12] L. Chisci, A. Garulli, and G. Zappa, "Recursive state bounding by parallelotopes," *Automatica*, vol. 32:7, pp. 1049–1055, 1996.

[13] A. Benavoli, "The generalised moment based filter," Tech. Rep. IDSIA-03-13, IDSIA, March 2013. http://www.idsia.ch/idsiareport/IDSIA-03-13.pdf.

[14] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. New York: Chapman and Hall, 1991.

[15] A. Benavoli, M. Zaffalon, and E. Miranda, "Robust filtering through coherent lower previsions," *Automatic Control, IEEE Transactions on*, vol. 56, pp. 1567 –1581, July 2011.

[16] M. Schikora, A. Gning, L. Mihaylova, D. Cremers, and W. Koch, "Box-particle phd filter for multi-target tracking," in *Information Fusion (FUSION), 2012 15th International Conference on*, pp. 106–113, 2012.

[17] A. Gning, B. Ristic, and L. Mihaylova, "A box particle filter for stochastic and set-theoretic measurements with association uncertainty," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pp. 1–8, 2011.

[18] A. Gning, B. Ristic, and L. Mihaylova, "Bernoulli particle/box-particle filters for detection and tracking in the presence of triple measurement uncertainty," *Signal Processing, IEEE Transactions on*, vol. 60, no. 5, pp. 2138–2151, May.

[19] F. Abdallah, A. Gning, and P. Bonnifait, "Box particle filtering for nonlinear state estimation using interval analysis," *Automatica*, vol. 44, no. 3, pp. 807 – 815, 2008.

[20] A. Gning, L. Mihaylova, and F. Abdallah, "Mixture of uniform probability density functions for non linear state estimation using interval analysis," in *Information Fusion (FUSION), 2010 13th Conference on*, pp. 1–8, 2010.

[21] M. Kvasnica, P. Grieder, and M. Baotić, "Multi-Parametric Toolbox (MPT)," 2004.

[22] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3447–3457, 2008.

[23] B. Ristic, D. Clark, and B. Vo, "Improved SMC implementation of the PHD filter," in *Proceedings of the 13th International Conference on Information Fusion*, vol. 1, pp. 1–8, 2010.