A CPHD approximation based on a discrete-Gamma cardinality model

Flávio Eler De Melo[1] and Simon Maskell[2]

[1]School of Engineering and Physical Sciences, Heriot-Watt University
E-mail: flavio.eler@gmail.com

[2]Department of Electrical Engineering and Electronics, University of Liverpool
E-mail: s.maskell@liverpool.ac.uk

***Abstract*—The Cardinalized Probability Hypothesis Density (CPHD) filter has become one of the most acclaimed algorithms for multi-target Bayesian filtering due to its ability to accurately estimate the number of objects and the object states in tracking scenarios affected by clutter. The CPHD filter generalizes the Probabilistic Hypothesis Density (PHD) filter by jointly propagating the first-order multi-target moment (intensity function) along with the entire probability distribution on the number of targets (cardinality distribution). In general, the CPHD recursion is computationally intractable, however successful approximations have been devised with reported computational complexity dominated by $\mathcal{O}(m^3)$ operations per filtering iteration, where m is the number of measurements. Room for improvement was originally acknowledged by Mahler, who conceived the idea of approximating the cardinality distribution by two-parameter distributions. In this paper, we further explore this idea to provide an efficient approximation of the CPHD filter where the cardinality distribution is modeled as a discretized Gamma distribution. Experiments show that the resulting filter is less computationally complex than the standard implementation of the CPHD filter but shows similar cardinality accuracy and variance.**

I. Introduction

Multi-target tracking is concerned with estimating the states of several objects of interest from a sequence of observations corrupted by noise, in the presence of missed detections and false alarms. The classical literature on multi-target tracking relies on the use of multivariate Bayesian statistics for estimating the target states while, at the same time, resorting to a combinatorial analysis for capturing the very complex problem of associating objects to sequences of measurements. Celebrated methods, such as the Multi-Hypothesis Tracking (MHT) [1] and Joint Probabilistic Data Association (JPDA) [2] have become traditional due to three factors. First, their ability to incorporate established filtering techniques (e.g., Kalman measurement update). Second, their ability to provide multi-target state estimates that are accurate enough to be practically useful. Third, the computational power available in modern digital computers becoming sufficient to allow such methods to run in real-time.

The Probability Hypothesis Density (PHD) filter [3] has set the cornerstone for a completely new way of interpreting multi-target Bayesian inference: a view where data association is avoided by making a judicious choice related to the targets' state description while promoting a neat and elegant mathematical formulation. In this formulation, a collection of target states is considered as a random set-valued state and the collection of sensor measurements is treated as a random set-valued observation. While intimately related to the theory of stochastic population processes [4], the multi-target Bayesian framework proposed by Ronald Mahler has been rather stated in terms of random finite sets (RFS) on which operations are given in the context of finite-set statistics (FISST) [5, 6]. In this context, the PHD recursion [3] was developed as a first-moment approximation to the multi-target Bayesian filtering problem, where the algorithm propagates the posterior intensity function of the set of targets' states in time. Recent advances of filters based on finite-set statistics include the filters exactly derived for labeled RFS's [7, 8].

In their influential analysis of the PHD filter, Erdinc, Willett, and Bar-Shalom [9] pointed out that its most prominent limitation resides in that the filter is first-order on the number of targets (cardinality). Since the number of targets in the PHD filter is assumed to be Poisson-distributed, the cardinality mean equals the variance. This gives rise to very unstable cardinality estimates when considering scenarios involving large number of missed detections or false alarms. In order to address this concern, Mahler derived the Cardinalized PHD (CPHD) filter [10, 11] which propagates the first-order multi-target moment (intensity function) along with the entire cardinality distribution. However, in general, practical implementations of the CPHD filter (e.g., [12]) are deemed to demand $\mathcal{O}(m^3)$ operations per filtering iteration, where m is the number of measurements. Mahler acknowledged there was room for computational improvement via additional approximation. With this potential in mind, he conceived the idea of approximating the cardinality distribution by two-parameter distributions and devised a cardinalized filter, the "binomial filter", whose number of targets is assumed to be distributed according to a binomial distribution [13]. Some developments in the same ethos have been proposed in terms of a model based on the Panjer distribution, such as that of Franken *et al.* [14], and a recent new filter by Schlangen *et al.* [15] that assumes the cardinality distribution to be a Panjer distribution and establishes a recursion where both the first moment and a second-order entity, the regional variance [16], are propagated in the filtering procedure.

PHD filters that are second-order in target number have a clear advantage relative to the CPHD filter: they approximate a special case of the CPHD filter by reproducing, for a family of discrete probability distributions, the essential strategy of the Kalman filter, i.e., estimating sufficient statistics rather than the complete distribution. Mahler's binomial filter has particular merit in this context since the binomial distribution is the most basic discrete distribution that can describe processes where the variance is smaller than or equal to the mean, hereafter called *underdispersed* processes. Their importance of this latter aspect becomes clear when we notice that, in most practical cases, cardinality estimates should be underdispersed because with a fairly high probability of

detection and a moderate number of false alarms, the set of measurements is highly informative about the target number. Unfortunately, the binomial filter suffers from a fundamental drawback: the number of trials of a binomial distribution is strongly constrained by the filtering scenario, and the filter derived in [13] loses its validity when the predicted number of trials is smaller than the number of measurements.

This paper proposes a new filter that is second-order in target number, where the multi-target state is assumed to follow an independent and identically distributed cluster process with the cardinality distribution modeled as a discretized Gamma distribution. Although apparently complicated, a discretized Gamma distribution allows simple calculations for approximating the first- and second-order moments of the posterior cardinality distribution, and efficiently addresses tracking scenarios with underdispersed target count while avoiding the restrictions imposed by the binomial filter. This latter feature is important because, when provided an i.i.d. cluster process with overdispersed cardinality model, the filter would most probably suffer from the same instability as that found in the PHD filter.

The paper is organized as follows. Section II establishes some mathematical preliminaries and Section III briefly presents the CPHD recursion. Section IV provides an analysis of the binomial filter. Section V proposes the Discrete-Gamma CPHD recursion. Section VI presents a comparison of the proposed algorithm with the PHD and CPHD filters, in the context of an illustrative simulated scenario. Section VII concludes. Proofs associated with the derivation of the DG-CPHD algorithm are available in the supplementary material.

II. Mathematical Preliminaries

In this section we briefly present the mathematical concepts required to understand the paper: multi-target statistics, point processes, probability generating functionals, and the intensity function of a state set density. We shall use the same description and notation originally proposed by Mahler [3, 11], though we note that a description reminiscent of the seminal work by Moyal [4] and related to a similar formulation [17, 18] has been adopted in some recent papers [15, 16, 19], appealing to the measure-theoretic formalism. Our choice to adopt Mahler's description is just a matter of convenience when referring to expressions derived in [11], which may prevent the reader from having to translate between two equivalent formalisms.

A. Multi-target Statistics and Point Processes

Let $\mathrm{x}_i \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ be a d_x-dimensional vector describing the state of a single target identified by i. We assume a scenario with a random number $n \in \mathbb{N}$ of targets, and the collection of targets has no intrinsic ordering such that the joint state of this collection is represented by the finite set $\mathrm{x}_{1:n} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n\}$ where $\mathrm{x}_1, \ldots, \mathrm{x}_n$ are state vectors of all targets. By allowing such set of targets to be random in number of elements and state vectors, then the resulting set is a random finite set (RFS). Let Ξ be a random finite set and $\mathrm{X} := \{\mathrm{x}_1, \ldots, \mathrm{x}_n\}$ its realization, the multi-target probability density can then be described as

$$p_{\Xi}\left(\mathrm{X}\right) = p_N\left(n\right) p_{\mathrm{x}_{1:n}|N}\left(\{\mathrm{x}_1, \ldots, \mathrm{x}_n\}|n\right), \tag{1}$$

where $p_N(n)$ is the cardinality distribution, and $p_{\mathrm{x}_{1:n}|N}\left(\{\mathrm{x}_1, \ldots, \mathrm{x}_n\}|n\right)$ is the joint probability density of the point set $\mathrm{x}_{1:n}$. In addition, the random finite set Ξ can be formally understood as a point process that pertains to the composite space $\mathscr{X} = \bigcup_{n=0}^{\infty} \mathcal{X}^n$, i.e., $\Xi \in \mathscr{X}$, with probability density p_{Ξ}. If the point process Ξ is an i.i.d. cluster process, the multi-object probability density assumes the form $p_{\Xi}\left(\mathrm{X}\right) = p_N\left(n\right) \cdot n! \prod_{i=1}^{n} p(\mathrm{x}_i)$, where $p(\mathrm{x}_i)$ is the spatial probability density of each target $i \in \{1, \ldots, n\}$ and the term $n!$ accounts for all possible orderings of the finite set of states $\mathrm{x}_{1:n}$. Particularly, if the point process is Poisson with mean μ, then $p_{\Xi}(\mathrm{X}) = e^{-\mu} \prod_{i=1}^{n} \mu p(\mathrm{x}_i)$. We can verify that $p_{\Xi}\left(\mathrm{X}\right)$ is a probability density by identifying the value of its set integral over all possible realizations as

$$\begin{aligned}
&\int p_{\Xi}\left(\mathrm{X}\right) \delta\mathrm{X} \\
&\triangleq \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^n} p_N\left(n\right) p_{\mathrm{x}_{1:n}|N}\left(\{\mathrm{x}_1, \ldots, \mathrm{x}_n\}|n\right) d(\mathrm{x}_1, \ldots, \mathrm{x}_n) \\
&\quad = \sum_{n=0}^{\infty} p_N\left(n\right) \int_{\mathcal{X}^n} \left[\prod_{i=1}^{n} p(\mathrm{x}_i)\right] d(\mathrm{x}_1, \ldots, \mathrm{x}_n) = 1.
\end{aligned} \tag{2}$$

In a similar way, let $\mathrm{z}_j \in \mathcal{Z} \subseteq \mathbb{R}^{d_z}$ be a d_z-dimensional vector describing the jth measurement collected by one (or more) sensor(s). We also assume that there is a random number $m \in \mathbb{N}$ of measurements, possibly originated from targets or false alarms and without specific order, that can be described by a random finite set Ψ with finite point set $\mathrm{z}_{1:m} = \{\mathrm{z}_1, \ldots, \mathrm{z}_m\}$ and where $\mathrm{z}_1, \ldots, \mathrm{z}_m$ are all observation vectors. For any realization of Ψ, say Z, multi-target likelihood functions can be defined as a generalization of single-target likelihoods with properties analogous to those described for the multi-target density, i.e., likelihoods of the type $p_{\Psi|\Xi}(\mathrm{Z}|\mathrm{X})$ are possible.

B. Probability Generating Functions and Functionals

Given a probability mass function $p_N(n) = \Pr\{N = n\}$ of a discrete nonnegative random variable N, its probability generating function (p.g.f. or PGF) is defined as

$$G_N(\zeta) \triangleq \sum_{n=0}^{\infty} p_N(n) \zeta^n \equiv \mathbb{E}\left[\zeta^N\right], \tag{3}$$

which converges absolutely for $\zeta \in \mathbb{C}$ such that[1] $|\zeta| \leq 1$. From (3) one can recognize that $p_N(n) = G_N^{(n)}(0)/n!$ and $G_N(1) = 1$. Similarly, an analogous concept can be applied to a point-process random variable (random finite set), Ξ, via the probability generating functional (p.g.fl. or PGFL)

$$\begin{aligned}
G_{\Xi}[h] &= \int h^{\mathrm{X}} p_{\Xi}\left(\mathrm{X}\right) \delta\mathrm{X} \equiv \mathbb{E}\left[h^{\Xi}\right] \\
&= \sum_{n=0}^{\infty} p_N\left(n\right) \int_{\mathcal{X}^n} \left[\prod_{i=1}^{n} h(\mathrm{x}_i) p(\mathrm{x}_i)\right] d(\mathrm{x}_1, \ldots, \mathrm{x}_n),
\end{aligned} \tag{4}$$

where $h : \mathcal{X} \to [0, 1]$ is a test function analogous to ζ for the PGF, where $h^{\mathrm{X}} = 1$ if $\mathrm{X} = \emptyset$ and $h^{\mathrm{X}} = \prod_{i=1}^{n} h(\mathrm{x}_i)$ if $\mathrm{X} = \{\mathrm{x}_1, \ldots, \mathrm{x}_n\}$.

C. Functional Differentiation and the Intensity Function

The functional derivative of a probability generating functional with respect to $\mathrm{x} \in \mathcal{X}$ is defined as

$$\frac{\delta G_{\Xi}}{\delta \mathrm{x}}[h] \triangleq \lim_{\varepsilon \searrow 0} \frac{G_{\Xi}[h + \varepsilon \delta_{\mathrm{x}}] - G_{\Xi}[h]}{\varepsilon}, \tag{5}$$

[1] Other values of ζ may lead to convergence, though $|\zeta| \leq 1$ is a sufficient condition.

where $\delta_{\mathrm{x}} = \delta(\mathrm{x}' - \mathrm{x})$ is the Dirac delta function concentrated at x and the differentiation is assumed linear and continuous for a fixed h. It is worth noting that the definition (5) is heuristic because the Dirac delta is not a valid test function in view of (4). A rigorous definition of the functional derivative can be found in [20]. When differentiating with respect to a random set realization $\mathrm{X} \coloneqq \{\mathrm{x}_1, \ldots, \mathrm{x}_n\}$, the following definitions apply:

$$\frac{\delta G_{\Xi}}{\delta \mathrm{X}}[h] \triangleq \frac{\delta^n G_{\Xi}}{\delta \mathrm{x}_1 \ldots \delta \mathrm{x}_n}[h], \quad \frac{\delta G_{\Xi}}{\delta \emptyset}[h] \triangleq G_{\Xi}[h]. \tag{6}$$

As detailed in [5], functional derivatives obey rules analogous to those of elementary calculus, including the rule for a linear functional, product rule, and chain rule. The first moment of the random finite set Ξ (point process) is what is generally called the "intensity function", or the probability hypothesis density (PHD) [3]. The first moment can be derived from the first derivative of the probability generating functional as

$$D_{\Xi}(\mathrm{x}) = \frac{\delta G_{\Xi}}{\delta \mathrm{x}}[1], \tag{7}$$

or equivalently from a set expectation according to

$$D_{\Xi}(\mathrm{x}) = \int \delta_{\mathrm{X}}(\mathrm{x}) p_{\Xi}(\mathrm{X}) \delta \mathrm{X} = \mathbb{E}\left[\delta_{\Xi}(\mathrm{x})\right], \tag{8}$$

where $\delta_{\mathrm{X}}(\mathrm{x}) \triangleq \sum_{\hat{\mathrm{x}} \in \mathrm{X}} \delta(\mathrm{x} - \hat{\mathrm{x}})$. The intensity function $D_{\Xi}(\mathrm{x})$ can be interpreted as a density of objects at x, i.e., $\mathbb{E}\left[|\Xi \cap \mathcal{S}|\right] = \int_{\mathcal{S}} D_{\Xi}(\mathrm{x}) d\mathrm{x}$ gives the expected number of objects in the region $\mathcal{S} \subseteq \mathcal{X}$.

III. The Cardinalized PHD Filter

In this section we present the CPHD filter in its original form. The CPHD recursion was proposed by Mahler in [10, 11] to address the instability of the PHD filter when estimating the target number. The CPHD recursion propagates the first-order multi-target moment (intensity function) along with the entire cardinality distribution. The CPHD recursion is derived based on the following assumptions:

Assumption 1. *Each target moves independently of one another, with motion modeled by a single-target Markov transition density $p_{t,k|k-1}(\mathrm{x}|\mathrm{x}')$, which we abbreviate as $p_t(\mathrm{x}|\mathrm{x}')$.*

Assumption 2. *Existing targets may disappear between two time steps with probability $1 - p_{s,k|k-1}(\mathrm{x})$, where $p_{s,k|k-1}(\mathrm{x})$ is the probability that a target with state x will survive between time steps $k-1$ and k, hereafter abbreviated as $p_s(\mathrm{x})$.*

Assumption 3. *New targets can appear in the scene independently of the existing targets, according to a Poisson point process. New targets are realized at time step k with joint density $b_{k|k-1}(\mathrm{X})$. The intensity function of the corresponding random finite set is denoted as $D_b(\mathrm{x})$, and its cardinality distribution is denoted as $p_b(n)$ whose p.g.f. is $G_b(x)$.*

Assumption 4. *Measurements generated from targets are independent of one another, with single-target likelihood function $\ell_{k,\mathrm{z}}(\mathrm{x}) = p_{\ell,k}(\mathrm{z}|\mathrm{x})$, hereafter abbreviated as $\ell_{\mathrm{z}}(\mathrm{x})$.*

Assumption 5. *The sensor detects a single target with state x at the time step k with probability $p_{d,k}(\mathrm{x})$, abbreviated as $p_d(\mathrm{x})$ and the probability of not detecting the target is denoted as $q_d(\mathrm{x}) = 1 - p_d(\mathrm{x})$.*

Assumption 6. *False alarms may affect the observation, and are assumed to be characterized by a Poisson random finite set (or point process) and independent of the actual measurements. At the time step k, the sensor obtains a number of false alarms whose spatial distribution is individually given by the probability density $c_k(\mathrm{z})$ and whose cardinality distribution is given by $p_{c,k}(m)$, abbreviated as $c(\mathrm{z})$ and $p_c(m)$ respectively, and where the average number of false alarms is denoted by $\lambda = \mathbb{E}_{p_c}[m]$. The p.g.f. of $p_c(m)$ is denoted as $G_c(x)$.*

Assumption 7. *Both the prior and posterior multi-target random finite sets are i.i.d. cluster processes.*

A. CPHD Prediction Step

At a given time instant $k-1$ one has in possession estimates of the intensity $D_{k-1}(\mathrm{x}|\mathrm{Z}_{1:k-1})$, the expected number of targets $\hat{N}_{k-1}$, and the cardinality distribution $p_{k-1}(n|\mathrm{Z}_{1:k-1})$, conditioned on all measurements received to date, $\mathrm{Z}_{1:k-1} = \{\mathrm{Z}_1, \ldots, \mathrm{Z}_{k-1}\}$, where Z_{k-1} is the RFS realization containing all observations received at time instant $k-1$. In this section, we notationally suppress the conditioning on $\mathrm{Z}_{1:k-1}$ to express the intensity function and cardinality distribution in the concise forms $D_{k-1}(\mathrm{x})$ and $p_{k-1}(n)$, respectively.

We write $\varsigma_{k-1}(\mathrm{x}) \coloneqq \hat{N}_{k-1}^{-1} D_{k-1}(\mathrm{x})$, abbreviated as $\varsigma(\mathrm{x})$ in this section, and recall the definition of inner product between functions as $\langle f, g \rangle \triangleq \int_{\mathcal{X}} f(\mathrm{x}') g(\mathrm{x}') d\mathrm{x}'$. The prior p.g.f. corresponding to the prior cardinality distribution is given by

$$G_{k|k-1}(x) = G_b(x) \cdot G_{k-1}(\langle 1 - p_s, \varsigma \rangle + \langle p_s, \varsigma \rangle x), \tag{9}$$

where $G_{k-1}(x)$ is the p.g.f. of the cardinality distribution at time step $k-1$, and the equality[2] is valid under *Assumption 7*. Expression (9) follows from the assumption that the birth process is independent of the prior process of targets that survived, which in turn is written for a Bernoulli survival transition by using the Watson-Galton recursion [21]. The CPHD prediction step obtains the prior intensity function, prior expected number of targets, and prior cardinality distribution according to

$$D_{k|k-1}(\mathrm{x}) = D_b(\mathrm{x}) + \textstyle\int_{\mathcal{X}} p_s(\mathrm{x}')\, p_t(\mathrm{x}|\mathrm{x}') D_{k-1}(\mathrm{x}') d\mathrm{x}', \tag{10}$$

$$\hat{N}_{k|k-1} = \hat{N}_{b,k} + \hat{N}_{s,k}, \tag{11}$$

$$p_{k|k-1}(n) = \textstyle\sum_{i=0}^{n} p_b(n-i) \frac{1}{i!} G_{k-1}^{(i)}(\langle 1 - p_s, \varsigma \rangle) \langle p_s, \varsigma \rangle^i, \tag{12}$$

where $\hat{N}_{b,k} = \langle 1, D_b \rangle$ is the expected number of newborn targets, and $\hat{N}_{s,k} = \langle p_s, D_{k-1} \rangle$ is the expected number of targets that have survived from time step $k-1$.

B. CPHD Measurement Update

In this section we assume that measurements are collected from a single sensor at time instant k, as a realization Z_k of the observation RFS Ψ_k, with a finite point set $\mathrm{z}_{k,1:m_k} = \{\mathrm{z}_{k,1}, \ldots, \mathrm{z}_{k,m_k}\}$ of collected measurements. Based on the prior intensity function $D_{k|k-1}(\mathrm{x}|\mathrm{Z}_{1:k-1})$, prior expected number of targets $\hat{N}_{k|k-1}$, and prior cardinality distribution $p_{k|k-1}(n|\mathrm{Z}_{1:k-1})$ the realization Z_k is used to produce the posterior intensity function $D_k(\mathrm{x}|\mathrm{Z}_{1:k})$, posterior expected number of targets $\hat{N}_k$ and posterior cardinality distribution $p_k(n|\mathrm{Z}_{1:k})$.

We denote the intensity functions as $D_{k|k-1}(\mathrm{x})$ and $D_k(\mathrm{x})$, and the cardinality distributions as $p_{k|k-1}(n)$ and $p_k(n)$. In

[2] In practice this is an approximation.

addition, we set $\varsigma_{k|k-1}(\mathrm{x}) \coloneqq \hat{N}_{k|k-1}^{-1} D_{k|k-1}(\mathrm{x})$, abbreviated as $\varsigma(\mathrm{x})$ in this section. The p.g.f. of the posterior cardinality distribution is approximately given by

$$G_{k|k}(x) = \frac{\sum_{j=0}^{m_k} x^j G_c^{(m_k-j)}(0) G^{(j)}(x\langle q_d, \varsigma\rangle)\sigma_j(\mathrm{Z}_k)}{\sum_{i=0}^{m_k} G_c^{(m_k-i)}(0) G^{(i)}(\langle q_d, \varsigma\rangle)\sigma_i(\mathrm{Z}_k)}, \quad (13)$$

where $m_k = |\mathrm{Z}_k|$, $G(x) = G_{k|k-1}(x)$, and for a set Z with $m = |\mathrm{Z}|$,

$$\sigma_i(\mathrm{Z}) \triangleq \sigma_{m,i}\left(\frac{\langle p_d \ell_{\mathrm{z}_1}, \varsigma\rangle}{c(\mathrm{z}_1)}, \ldots, \frac{\langle p_d \ell_{\mathrm{z}_m}, \varsigma\rangle}{c(\mathrm{z}_m)}\right), \quad (14)$$

where $\sigma_{m,i}(x_1, \ldots, x_m)$ is the elementary homogeneous symmetric function of degree i in $x_1, \ldots, x_m$. By defining

$$\Upsilon_k[\mathrm{Z}] = \frac{\sum_{j=0}^{|\mathrm{Z}|} G_c^{(|\mathrm{Z}|-j)}(0) G^{(j+1)}(\langle q_d, \varsigma\rangle)\sigma_j(\mathrm{Z})}{\sum_{i=0}^{m_k} G_c^{(m_k-i)}(0) G^{(i)}(\langle q_d, \varsigma\rangle)\sigma_i(\mathrm{Z}_k)}, \quad (15)$$

the CPHD measurement-update step can be described as

$$D_k(\mathrm{x}) = \frac{q_d(\mathrm{x})}{G^{(1)}(1)} \Upsilon_k[\mathrm{Z}_k] D_{k|k-1}(\mathrm{x}) + \frac{p_d(\mathrm{x})}{G^{(1)}(1)} \sum_{\mathrm{z}\in\mathrm{Z}_k} \frac{\ell_{\mathrm{z}}(\mathrm{x})}{c(\mathrm{z})} \Upsilon_k[\mathrm{Z}_k \setminus \{\mathrm{z}\}] D_{k|k-1}(\mathrm{x}), \quad (16)$$

$$\hat{N}_k = G_k^{(1)}(1) \approx \arg\max_n p_k(n), \quad (17)$$

$$p_k(n) = \frac{\sum_{j=0}^{m_k} G_c^{(m_k-j)}(0) \frac{\langle q_d, \varsigma\rangle^{n-j}}{(n-j)!} \cdot G^{(n)}(0)\sigma_j(\mathrm{Z}_k)}{\sum_{i=0}^{m_k} G_c^{(m_k-i)}(0) G^{(i)}(\langle q_d, \varsigma\rangle)\sigma_i(\mathrm{Z}_k)}. \quad (18)$$

IV. The Binomial Filter

The binomial filter [13] was proposed as one further step towards simplification of the CPHD filter. The strategy employed was to mimic, for a hypothesized discrete distribution, the procedure of a Kalman filter for the cardinality random variable, i.e., estimating sufficient statistics. Mahler proposed that the cardinality should be distributed according to a binomial distribution,

$$p_N(n) = \binom{\nu}{n} \omega^n (1-\omega)^{\nu-n}, \quad (19)$$

which models the probability of n successes out of ν trials each with probability ω. For the purpose of counting objects in a scene, the binomial distribution has two virtues: (i) underdispersion, i.e., the variance is smaller than or equal to the mean, and (ii) simplicity, with an analytic probability generating function $G_{\mathrm{bin}}(\zeta) = (1-\omega+\omega\zeta)^{\nu}$. As touched on before, the first virtue is very important for the vast majority of practical applications since in scenarios of fairly high probability of detection and moderate number of false alarms, a set of measurements is highly informative about the target number. Under these conditions, for a sufficiently high number of targets, it is very unlikely that the cardinality variance will be perceived as greater than the expected target number. To verify the underdispersed characteristic of the binomial distribution, one just needs to note that, for $0 \leq \omega \leq 1$, $\mathrm{var}(N) = \nu\omega(1-\omega) \leq \nu\omega = \mathbb{E}[N]$ because $1-\omega \leq 1$.

On the other hand, the binomial filter has a fundamental problem: for a given number of measurements m_k received at time step k, the filter requires that $\nu_{k|k-1} \geq m_k$. This is actually an essential assumption associated with *Theorem 2* in [13], which makes possible to establish the measurement-update step. Ultimately, this assumption poses a limitation in the number of newly appearing targets that the binomial filter can cope with. Supposing that $\hat{N}_{k|k-1}$ targets have been predicted in the scene and no false alarms or missed detections took place, then the maximum number of new targets that the binomial filter could account for would be[3] $\hat{\nu}_{k|k-1} - \hat{N}_{k|k-1} \approx \mathrm{var}(N_{k|k-1})/\hat{\omega}_{k|k-1}$. The problem arises because when predicting $\nu_{k|k-1}$ and $\omega_{k|k-1}$ no information on the actual number of new targets is available, and the problem tends to be exacerbated when a high number of false alarms may be realized, i.e., most probably $\hat{\nu}_{k|k-1} \geq m_k$ would not hold.

V. The Discrete-Gamma CPHD Filter

In this section we propose a new approximation to the CPHD filter in the same ethos of the binomial filter. Such approximation is based on a new cardinality model, obtained by discretizing the Gamma distribution for describing count data. This cardinality model and its properties are explained in Sections V-A and V-B. In Section V-C we present the prediction step of the proposed filter, and in Section V-D we present its measurement-update step. Section V-E is dedicated to an implementation of the Discrete-Gamma CPHD filter based on Gaussian mixtures.

A. The Discrete Gamma Distribution

In order to derive the new filter, we assume that the prior and posterior cardinality distributions can be accurately approximated as a discretized form of the Gamma distribution, called *discrete Gamma* (*dGamma*) hereafter, with probability mass function given by

$$dGamma(n|\alpha,\beta) \triangleq \frac{Gamma(n|\alpha,\beta)}{\sum_{u=0}^{\infty} Gamma(u|\alpha,\beta)},\ n \in \mathbb{N}_0, \quad (20)$$

$$dGamma(n|\alpha,\beta) = \frac{\Gamma(\alpha)^{-1}\beta^{\alpha} n^{\alpha-1} e^{-\beta n}}{\Gamma(\alpha)^{-1}\beta^{\alpha}\sum_{u=0}^{\infty} u^{\alpha-1} e^{-\beta u}} = \frac{n^{\alpha-1} e^{-\beta n}}{\sum_{u=0}^{\infty} u^{\alpha-1} e^{-\beta u}}. \quad (21)$$

where $Gamma(x|\alpha,\beta)$ denotes the continuous Gamma probability density at $x \in \mathbb{R}_+$ with a shape parameter, $\alpha > 0$, and a rate parameter, $\beta > 0$. This form of the discrete Gamma distribution is similar to that presented in [22], except that their description involves discretizing the shape parameter as well, i.e. $\alpha \in \mathbb{N}_0$. The utility of such probability mass function will become evident later on when we present the filter equations that follow from it. For now it suffices to mention its benign characteristics, involving a simple cardinality model that does not suffer from the shortcoming shown for the binomial filter, and being suitable for inference of count phenomena. Specifically, the discrete Gamma distribution can be either underdispersed or overdispersed. The discrete Gamma distribution has the probability generating function

$$G_{dGamma}(\zeta) = \sum_{n=0}^{\infty} dGamma(n|\alpha,\beta)\cdot\zeta^n = \frac{\sum_{n=0}^{\infty} n^{\alpha-1}(e^{-\beta}\zeta)^n}{\sum_{u=0}^{\infty} u^{\alpha-1} e^{-\beta u}} = \frac{\mathrm{Li}_{1-\alpha}(e^{-\beta}\zeta)}{\mathrm{Li}_{1-\alpha}(e^{-\beta}\ )}, \quad (22)$$

where $\mathrm{Li}_s(z) \triangleq \sum_{k=1}^{\infty} k^{-s} z^k$, for $|z| < 1$, is the function known as the *polylogarithm* of order $s \in \mathbb{C}$ and argument $z \in \mathbb{C}$. Interesting particular cases involve $\mathrm{Li}_1(z) = -\log(1-$

[3] Typically $\hat{\omega}_{k|k-1}$ is of the same order of magnitude as $\mathrm{var}(N_{k|k-1})$.

z), and $\mathrm{Li}_{-\ell}(z) = (1-z)^{-\ell-1}\sum_{i=0}^{\ell} A(\ell,i)z^{\ell-i}$, $\ell \in \mathbb{N}$, where $A(\ell,i)$ is an Eulerian number[4]. This latter case is used in the description of the discrete Gamma distribution in [22], and can be computed by evaluating a finite number of terms. However, in practice, as ℓ increases, computing $A(\ell,i)$ becomes computationally expensive and increasingly prone to round-off errors. Resorting to the known identity $\mathrm{Li}_{-\ell}(z) = (z\cdot\partial_z)^{\ell}\left[z/(1-z)\right]$ could be an alternative, but the numerical evaluation of these rational expressions increasingly suffers from cancellation errors as ℓ becomes large. In those cases, truncating the polylogarithm's definition directly may give a better answer [23]. This constitutes an important aspect for the implementation of the discrete Gamma filter.

Before closing this section, it is worth remarking that other discrete distributions for count data were investigated. At a first glance, some of them appeared as suitable candidates for our endeavor, including the well-known generalized Poisson distribution [24] and, more generally, the class of Lagrangian distributions [25], the Conway-Maxwell-Poisson distribution [26] and other forms of Gamma-count distributions [27]. In our investigation, we have verified the feasibility of such distributions, observing different disadvantages in each one them: the Lagrangian distributions propose complex implicit forms of probability generating functions (based on the Lagrange-Bürmann formula), the Conway-Maxwell-Poisson distribution requires approximations for the moments that imply specific conditions on the parameters (limiting the description of overdispersion), and the Gamma-count distribution in [27] results in a probability generating function involving many evaluations of the incomplete Gamma function. Eventually, we found more favorable characteristics in our approach.

B. Moments of the Discrete Gamma Distribution

From (22), it is not clear whether closed-form solutions exist for the discrete Gamma factorial moments. We will rely on a procedure closely related to that shown in [28] in the context of inducing the Euler-Mclaurin summation formula. First observe that $\lim_{m\to\infty}\sum_{n=0}^{m} n^{\alpha-1}e^{-\beta n}$ converges for $\beta > 1$ according to the ratio criterion. Then, we shall assume $\beta > 1$ from now on. The analysis requires the continuous distribution maximum, $x_m = \arg\max_x \mathit{Gamma}(x|\alpha,\beta)$, $x_m \neq 0$, given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^{\alpha-1}e^{-\beta x}}{\beta^{-\alpha}\Gamma(\alpha)}\right)\Bigg|_{x=x_m} &= \frac{x_m^{\alpha-1}e^{-\beta x}}{\beta^{-\alpha}\Gamma(\alpha)}\left[(\alpha-1)x_m^{-1}-\beta\right]\\ &= 0,\\ x_m &= (\alpha-1)\beta^{-1}. \end{aligned} \tag{23}$$

The first two moments of $\mathit{dGamma}(n|\alpha,\beta)$ are evaluated by

$$\mu_N = G^{(1)}_{\mathit{dGamma}}(1) = \frac{\mathrm{Li}_{-\alpha}(e^{-\beta})}{\mathrm{Li}_{1-\alpha}(e^{-\beta})}, \tag{24}$$

$$\sigma_N^2 = G^{(2)}_{\mathit{dGamma}}(1) - \mu_N^2 + \mu_N. \tag{25}$$

In order to evaluate these moments, one must approximate ratios of polylogarithms when computing $G^{(1)}_{\mathit{dGamma}}(1)$ and $G^{(2)}_{\mathit{dGamma}}(1)$, whose exact values are not easily calculable. We shall not prove that $G^{(1)}_{\mathit{dGamma}}(1) \approx \alpha\beta^{-1}$ and $G^{(2)}_{\mathit{dGamma}}(1) \approx \alpha(\alpha+1)\beta^{-2} - \alpha\beta^{-1}$ to first order precision, but rather provide an intuitive argument that hopefully will convince

[4]The Eulerian number $A(\ell,i)$ is the number of permutations of the numbers 1 to ℓ in which i elements are greater than the previous element.

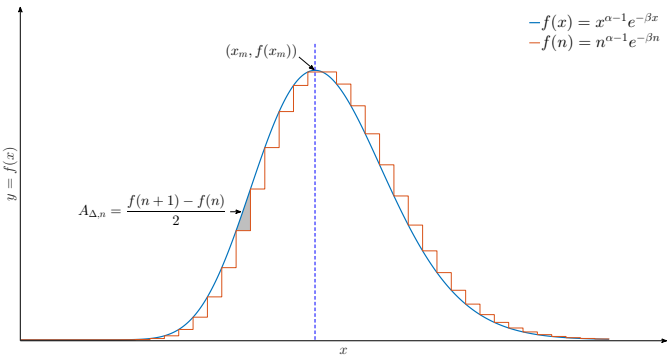

Figure 1. Summation vs. integral of $x^{\alpha-1}e^{-\beta x}$ ($\alpha = 21.5$, $\beta = 1$)

the reader that higher order terms may be neglected. Set $f(x) = x^{\alpha-1}e^{-\beta x}$ and refer to Figure 1 to observe that

$$d_{n<x_m} \triangleq \sum_{n=0}^{\lfloor x_m\rfloor-1} f(n) - \int_0^{\lfloor x_m\rfloor} f(x)dx < 0,$$

$$d_{n>x_m} \triangleq \sum_{n=\lfloor x_m\rfloor}^{\infty} f(n) - \int_{\lfloor x_m\rfloor}^{\infty} f(x)dx > 0,$$

where $\lfloor x\rfloor$ is the greatest integer less than or equal to x. Let us compute $d_{n<x_m}$ and $d_{n>x_m}$ by a first order approximation where the difference $f(n) - f(x)$, integrated within each interval $x \in \mathcal{I}_n = (n, n+1)$, is approximated by the area of a triangle (Figure 1), $A_{\Delta,n}$, with base $\Delta x_n = 1$ and height $\Delta f_n = |f(n+1) - f(n)|$, to result

$$d_{n<x_m} \approx -\frac{1}{2}\sum_{n=1}^{\lfloor x_m\rceil} (f(n) - f(n-1)) = -\frac{f(\lfloor x_m\rceil)}{2},$$

$$d_{n>x_m} \approx +\frac{1}{2}\sum_{n=\lfloor x_m\rceil}^{\infty} (f(n) - f(n+1)) = +\frac{f(\lfloor x_m\rceil)}{2},$$

since $f(0) = \lim_{x\to\infty} f(x) = 0$, and where $\lfloor x\rceil$ is the closest integer to x. Therefore, we can approximate the total difference to first order as

$$d_n = d_{n<x_m} + d_{n>x_m} \approx \frac{f(\lfloor x_m\rceil) - f(\lfloor x_m\rceil)}{2} = 0,$$

which allows us to get

$$\begin{aligned}\mathrm{Li}_{1-\alpha}(e^{-\beta}) = \sum_{n=0}^{\infty} n^{\alpha-1}e^{-\beta n} &= \int_0^{\infty} x^{\alpha-1}e^{-\beta x}dx + d_n\\ &\approx \beta^{-\alpha}\Gamma(\alpha).\end{aligned} \tag{26}$$

If one is interested in a higher-order approximation of d_n, the areas shown in Figure 1 should rather be approximated by integrals, within each interval $\mathcal{I}_n$, of a higher-order fitting polynomial. As a direct consequence of (26) and based on the property $\Gamma(\alpha+\ell) = (\alpha+\ell-1)\Gamma(\alpha+\ell-1)$, $\ell \in \mathbb{N}$, one can easily verify the validity of the approximations:

$$\mu_N = \frac{\mathrm{Li}_{-\alpha}(e^{-\beta})}{\mathrm{Li}_{1-\alpha}(e^{-\beta})} \approx \frac{\beta^{-(\alpha+1)}\Gamma(\alpha+1)}{\beta^{-\alpha}\Gamma(\alpha)} = \alpha\beta^{-1}, \tag{27}$$

$$\begin{aligned}\sigma_N^2 &= \frac{\mathrm{Li}_{-\alpha-1}(e^{-\beta}) - \mathrm{Li}_{-\alpha}(e^{-\beta})}{\mathrm{Li}_{1-\alpha}(e^{-\beta})} - \mu_N^2 + \mu_N\\ &\approx \frac{\beta^{-(\alpha+2)}\Gamma(\alpha+2) - \beta^{-(\alpha+1)}\Gamma(\alpha+1)}{\beta^{-\alpha}\Gamma(\alpha)} - \mu_N^2 + \mu_N\\ &= \frac{\beta^{-\alpha}\Gamma(\alpha)\left[\alpha(\alpha+1)\beta^{-2} - \alpha\beta^{-1}\right]}{\beta^{-\alpha}\Gamma(\alpha)} - \mu_N^2 + \mu_N\\ &= \alpha\beta^{-2}.\end{aligned} \tag{28}$$

C. Discrete-Gamma CPHD Prediction

The same assumptions used to formulate the CPHD filter, established in Section III, shall be applied to derive the Discrete-Gamma CPHD recursion. In addition to *Assumptions 1 to 7*, the following premise is also considered:

Assumption 8. *The prior and posterior multi-target random finite sets are assumed with multi-object discrete-Gamma probability densities of the form*

$$p_{\Xi}\left(\mathrm{X}\right) = dGamma_{\alpha,\beta}\left(n\right) \cdot n! \prod_{i=1}^{n} p(\mathrm{x}_i), \tag{29}$$

where $p(\mathrm{x}_i)$ is the spatial probability density of each object.

The probability generating functional of (29) can be obtained as

$$\begin{aligned} G_{\Xi}\left[h\right] &= \mathbb{E}_{\Xi}\left[e^{\sum_{i=0}^{n}\log h(\mathrm{x}_i)}\right] \\ &= \sum_{n=0}^{\infty} dGamma_{\alpha,\beta}\left(n\right) \int_{\mathcal{X}^n}\left[\prod_{i=1}^{n} h(\mathrm{x}_i)p(\mathrm{x}_i)\right] d\mathrm{x}_{1:n} \\ &= \sum_{n=0}^{\infty} \frac{n^{\alpha-1}\left[e^{-\beta}\int_{\mathcal{X}} h\left(\mathrm{x}\right)\varsigma\left(\mathrm{x}\right)d\mathrm{x}\right]^{n}}{\mathrm{Li}_{-\alpha+1}(e^{-\beta})} \\ &= \frac{\mathrm{Li}_{-\alpha+1}(e^{-\beta}\langle h,\varsigma\rangle)}{\mathrm{Li}_{-\alpha+1}(e^{-\beta})}, \end{aligned} \tag{30}$$

for $\beta \in \mathbb{R}$, $\beta > 1$, and where $\varsigma(\mathrm{x}_i) \coloneqq p(\mathrm{x}_i)$.

Now, based on previous knowledge of D_{k-1}, α_{k-1} and β_{k-1}, i.e., the posterior intensity function and cardinality parameters from the previous filtering iteration, the prediction step shall obtain the predicted intensity, $D_{k|k-1}$, and predicted cardinality parameters, $\alpha_{k|k-1}$ and $\beta_{k|k-1}$, which are functions of the prior cardinality mean, $\mu_{N,k|k-1}$, and of the prior cardinality variance, $\sigma^2_{N,k|k-1}$. The following proposition establishes how the predicted intensity and cardinality parameters are achieved. Proofs are presented in the supplementary material.

Proposition 1. *Provided the posterior intensity function at time instant $k-1$, D_{k-1}, and the parameters of a posterior discrete-Gamma cardinality distribution at time instant $k-1$, α_{k-1} (shape) and β_{k-1} (rate), then, under Assumptions 1–3 and 8, the predicted intensity function, $D_{k|k-1}$, and the predicted cardinality parameters, $\alpha_{k|k-1}$ and $\beta_{k|k-1}$, are given by*

$$D_{k|k-1}(\mathrm{x}) = D_b(\mathrm{x}) + \textstyle\int_{\mathcal{X}} p_s(\mathrm{x}')\, p_t(\mathrm{x}|\mathrm{x}')D_{k-1}(\mathrm{x}')d\mathrm{x}', \tag{31}$$

$$\alpha_{k|k-1} = \frac{\mu^2_{N,k|k-1}}{\sigma^2_{N,k|k-1}},\ \beta_{k|k-1} = \frac{\mu_{N,k|k-1}}{\sigma^2_{N,k|k-1}}, \tag{32}$$

where

$$\mu_{N,k|k-1} = \hat{N}_{k|k-1} = \hat{N}_{b,k} + \hat{N}_{s,k}, \tag{33}$$

$$\sigma^2_{N,k|k-1} = \mu_{N,k|k-1} + \langle p_s,\varsigma\rangle^2\alpha_{k-1}\beta^{-1}_{k-1}(\beta^{-1}_{k-1}-1), \tag{34}$$

given the expected number of newborn targets, $\hat{N}_{b,k} = \langle 1, D_b\rangle$, and the expected number of survived targets, $\hat{N}_{s,k} = \langle p_s, D_{k-1}\rangle$.

It is worth noting that (32) constitute approximations because $\mu_N \approx \alpha\beta^{-1}$ and $\sigma^2_N \approx \alpha\beta^{-2}$ as per the method of approximating moments of the discrete Gamma distribution presented in Section V-B. Also, note that (31) and (33) are essentially the same expressions as those found in the standard CPHD prediction. Also, similarly to the CPHD filter, the predictions of the intensity function and of the cardinality distribution are uncoupled. The main difference is in the way the predicted cardinality distribution is approximated, which for the DG-CPHD filter is modeled in terms of the discrete Gamma cardinality parameters $\alpha_{k|k-1}$ and $\beta_{k|k-1}$.

D. Discrete-Gamma CPHD Measurement Update

For the correction procedure, by knowing the prior intensity function, $D_{k|k-1}$, and the prior cardinality parameters, $\alpha_{k|k-1}$ and $\beta_{k|k-1}$, the measurement update step incorporates the set of all measurements taken at time instant k (realization Z_k) to calculate the posterior intensity function, D_k, and posterior cardinality parameters, α_k and β_k, which are functions of the posterior cardinality mean, $\mu_{N,k}$, and of the posterior cardinality variance, $\sigma^2_{N,k}$. The following proposition declares how the posterior intensity and cardinality parameters shall be calculated. Proofs are presented in the supplementary material.

Proposition 2. *Suppose that the prior intensity function at time instant k, $D_{k|k-1}$, and the parameters of a prior discrete-Gamma cardinality distribution, $\alpha_{k|k-1}$ (shape) and $\beta_{k|k-1}$ (rate), are known. Given a set of m_k collected measurements that define a realization Z_k of the observation random finite set Ψ_k, then, under Assumptions 4–8, the posterior intensity function, D_k, and the posterior cardinality parameters, α_k and β_k, are given by*

$$\begin{aligned} D_k(\mathrm{x}) &= \frac{q_d(\mathrm{x})}{\mu_{N,k|k-1}}\Theta_k\left[\mathrm{Z}_k\right]D_{k|k-1}(\mathrm{x}) \\ &+ \frac{p_d(\mathrm{x})}{\mu_{N,k|k-1}}\sum_{\mathrm{z}\in\mathrm{Z}_k}\frac{\ell_{\mathrm{z}}(\mathrm{x})}{\lambda c(\mathrm{z})}\Theta_k\left[\mathrm{Z}_k\setminus\{\mathrm{z}\}\right]D_{k|k-1}(\mathrm{x}), \end{aligned} \tag{35}$$

$$\alpha_k = \frac{\mu^2_{N,k}}{\sigma^2_{N,k}},\ \beta_k = \frac{\mu_{N,k}}{\sigma^2_{N,k}}, \tag{36}$$

where $\mu_{N,k|k-1} \coloneqq \alpha_{k|k-1}\beta^{-1}_{k|k-1}$,

$$\mu_{N,k} = \theta_{1,0} + \theta_{0,1}\cdot\langle q_d,\varsigma\rangle, \tag{37}$$

$$\begin{aligned} \sigma^2_{N,k} &= \theta_{2,0} - \theta_{1,0} + 2\theta_{1,1}\cdot\langle q_d,\varsigma\rangle \\ &+ \theta_{0,2}\cdot\langle q_d,\varsigma\rangle^2 - \mu^2_{N,k} + \mu_{N,k}, \end{aligned} \tag{38}$$

and where the following definitions apply:

$$\Theta_k\left[\mathrm{Z}\right] \triangleq \frac{\sum_{j=0}^{|\mathrm{Z}|}\hat{G}^{(j+1)}(\langle q_d,\varsigma\rangle)\bar{\sigma}_j(\mathrm{Z}\ )}{\sum_{i=0}^{m_k}\hat{G}^{(i)}(\langle q_d,\varsigma\rangle)\bar{\sigma}_i(\mathrm{Z}_k)}, \tag{39}$$

$$\theta_{u,v} \triangleq \frac{\sum_{j=0}^{m_k} j^u\hat{G}^{(j+v)}(\langle q_d,\varsigma\rangle)\bar{\sigma}_j(\mathrm{Z}_k)}{\sum_{i=0}^{m_k}\hat{G}^{(i)}(\langle q_d,\varsigma\rangle)\bar{\sigma}_i(\mathrm{Z}_k)}, \tag{40}$$

$$\bar{\sigma}_i(\mathrm{Z}) \triangleq \sigma_{m=|\mathrm{Z}|,i}\left(\frac{\langle p_d\ell_{\mathrm{z}_1},\varsigma\rangle}{\lambda c(\mathrm{z}_1)},\ldots,\frac{\langle p_d\ell_{\mathrm{z}_m},\varsigma\rangle}{\lambda c(\mathrm{z}_m)}\right), \tag{41}$$

$$\hat{G}^{(\ell)}(x) = \frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\mathrm{Li}_{-\alpha+1}(e^{-\beta}x), \tag{42}$$

for $\alpha = \alpha_{k|k-1}$, $\beta = \beta_{k|k-1}$ and $\ell \geq 1$.

The approximations in (36) arise from that $\mu_{N,k} \approx \alpha_k\beta^{-1}_k$ and $\sigma^2_{N,k} \approx \alpha_k\beta^{-2}_k$ as per the method of approximating moments of the discrete Gamma distribution presented in Section V-B. Note that equations (35) and (37) are analogous to those found in the standard CPHD correction step. In addition, as in the CPHD filter, the measurement updates of the intensity function and of the cardinality distribution are

coupled. For the measurement update step, the main difference between the DG-CPHD and CPHD filters is in the way the cardinality distribution is approximated, which for the DG-CPHD filter is in terms of the discrete Gamma cardinality parameters α_k and β_k.

At this point, it is important to clarify the essential difference between the DG-CPHD filter and the Panjer filter from [15]: the Panjer filter is derived to update the intensity function and the regional variance (analogue of the cardinality variance) without having to impose the posterior process as an i.i.d. cluster process for producing the estimates[5], whereas the DG-CPHD filter imposes the posterior as an i.i.d. cluster process in order to produce estimates of the cardinality statistics. In the Panjer filter, this posterior i.i.d. cluster model assumption is deferred to the next filter iteration in order to proceed the prediction. The Panjer update accounts for the joint effect of pairs of measurements in the second-order factorial moment, and so has an additional update term with computational complexity of order $\mathcal{O}(m^4)$, where m is the number of measurements.

It is also important to mention that computing (42) for the DG-CPHD update is not a complex operation because, typically, $\mathrm{Li}_{-\alpha+1}(e^{-\beta}x)$ can be calculated to double floating-point precision with only few hundreds terms, and the pre-evaluated terms $(n^{\alpha-1}e^{-\beta n}x^n)_{n\in\mathbb{N}}$ that compose $\mathrm{Li}_{-\alpha+1}(e^{-\beta}x)$ can be iteratively updated by element-wise vector multiplications (or sums of logarithms) in order to evaluate the sequence $(\mathrm{Li}^{(\ell)}_{-\alpha+1}(e^{-\beta}x))_{\ell\in\mathbb{N}}$. To see this just observe that

$$\mathrm{Li}^{(1)}_{-\alpha+1}(e^{-\beta}x) = x^{-1}\sum_{n=0}^{\infty} n^{\alpha-1}e^{-\beta n}x^n\cdot n, \tag{43}$$

$$\mathrm{Li}^{(2)}_{-\alpha+1}(e^{-\beta}x) = x^{-2}\sum_{n=0}^{\infty} n^{\alpha-1}e^{-\beta n}x^n\cdot n(n-1), \tag{44}$$

$$\vdots$$

$$\mathrm{Li}^{(\ell)}_{-\alpha+1}(e^{-\beta}x) = x^{-\ell}\sum_{n=0}^{\infty} n^{\alpha-1}e^{-\beta n}x^n\cdot n^{\underline{\ell}}, \tag{45}$$

where $n^{\underline{\ell}} = n(n-1)\ldots(n-\ell+1)$ is the falling factorial and whose terms can be iteratively computed as

$$\begin{aligned} D^1_x(n^{\alpha-1}e^{-\beta n}x^n)_n &= x^{-1}D^0_x(n^{\alpha-1}e^{-\beta n}x^n)_n\circ(n-0)_n,\\ D^2_x(n^{\alpha-1}e^{-\beta n}x^n)_n &= x^{-1}D^1_x(n^{\alpha-1}e^{-\beta n}x^n)_n\circ(n-1)_n,\\ &\vdots \end{aligned} \tag{46}$$

where the terms within parentheses represent sequences as $(\cdot)_{n\in\mathbb{N}}$, the symbol $\circ$ is the Hadamard product and D^{ℓ}_x is Euler's notation for differentiation.

E. DG-CPHD Filter via Gaussian Mixtures

As follows, the recursion established by *Propositions 1* and *2* will be presented in closed form by modeling the intensity functions as Gaussian mixtures. Without loss of generality in terms of applicability of the proposed filter, but in favour of illustration simplicity, we derive a solution for linear Gaussian models. In this context, the single-target state transition kernel and observation model are assumed linear and Gaussian as

$$p_t(\mathrm{x}|\mathrm{x}') = \mathcal{N}(\mathrm{x};\mathrm{F}\mathrm{x}',\mathrm{Q}), \tag{47}$$

$$\ell_{\mathrm{z}}(\mathrm{x}) \triangleq p_\ell(\mathrm{z}|\mathrm{x}\,) = \mathcal{N}(\mathrm{z};\mathrm{H}\mathrm{x}\,,\mathrm{R}). \tag{48}$$

[5]Though the Panjer filter still requires the assumption that the prior process is an i.i.d. cluster process.

Additionally, we assume that the probabilities of survival and detection, p_s and p_d respectively, are independent of the state. The state point process (RFS) at a previous time instant $k-1$, and the target birth point process (RFS) at the time instant k are characterized by the following intensity models:

$$D_{k-1}(\mathrm{x}) = \sum_{i=1}^{\mathrm{I}_{k-1}} w^{(i)}_{k-1}\mathcal{N}(\mathrm{x};\mathrm{m}^{(i)}_{k-1},\mathrm{P}^{(i)}_{k-1}), \tag{49}$$

$$D_{b,k}(\mathrm{x}) = \sum_{i=1}^{\mathrm{I}_{b,k}} w^{(i)}_{b,k}\mathcal{N}(\mathrm{x};\mathrm{m}^{(i)}_{b,k},\mathrm{P}^{(i)}_{b,k}), \tag{50}$$

where $\{(w^{(i)}_{k-1},\mathrm{m}^{(i)}_{k-1},\mathrm{P}^{(i)}_{k-1}) : i=1,\ldots,\mathrm{I}_{k-1}\}$ is the set of weights, means and covariances that describe the state intensity function, and $\{(w^{(i)}_{b,k},\mathrm{m}^{(i)}_{b,k},\mathrm{P}^{(i)}_{b,k}) : i=1,\ldots,\mathrm{I}_{b,k}\}$ is the set of weights, means and covariances that describe the target birth intensity function. We shall not prove the Gaussian-mixture DG-CPHD recursion formally, but rather refer the reader to [12] and just point out that, in essence, their proof is valid for the DG-CPHD filter, except for a few different details.

Gaussian-Mixture DG-CPHD Prediction

The following equations form the prediction step for the Gaussian-mixture DG-CPHD recursion.

$$\begin{aligned} D_{k|k-1}(\mathrm{x}) &= D_{b,k}(\mathrm{x}) + D_{s,k|k-1}(\mathrm{x})\\ &= \sum_{i=1}^{\mathrm{I}_{k|k-1}} w^{(i)}_{k|k-1}\mathcal{N}(\mathrm{x};\mathrm{m}^{(i)}_{k|k-1},\mathrm{P}^{(i)}_{k|k-1}), \end{aligned} \tag{51}$$

$$\alpha_{k|k-1} \approx \frac{\mu^2_{N,k|k-1}}{\sigma^2_{N,k|k-1}},\ \beta_{k|k-1} \approx \frac{\mu_{N,k|k-1}}{\sigma^2_{N,k|k-1}}, \tag{52}$$

$$\mu_{N,k|k-1} = \sum_{i=1}^{\mathrm{I}_{b,k}} w^{(i)}_{b,k} + p_s\alpha_{k-1}\beta^{-1}_{k-1}, \tag{53}$$

$$\sigma^2_{N,k|k-1} = \mu_{N,k|k-1} + p_s^2\alpha_{k-1}\beta^{-1}_{k-1}\left(\beta^{-1}_{k-1}-1\right), \tag{54}$$

where $\mathrm{I}_{k|k-1} = \mathrm{I}_{b,k} + \mathrm{I}_{k-1}$, and

$$\begin{pmatrix} w^{(i)}_{k|k-1},\\ \mathrm{m}^{(i)}_{k|k-1},\\ \mathrm{P}^{(i)}_{k|k-1} \end{pmatrix} = \begin{cases} \begin{pmatrix} w^{(i)}_{b,k},\\ \mathrm{m}^{(i)}_{b,k},\\ \mathrm{P}^{(i)}_{b,k} \end{pmatrix}, & i\in[1..\mathrm{I}_{b,k}];\\ \begin{pmatrix} p_s w^{(i)}_{k-1},\\ \mathrm{F}\mathrm{m}^{(i)}_{k-1},\\ \mathrm{F}\mathrm{P}^{(i)}_{k-1}\mathrm{F}^T+\mathrm{Q} \end{pmatrix}, & i\in]\mathrm{I}_{b,k}..\mathrm{I}_{k|k-1}]. \end{cases} \tag{55}$$

Gaussian-Mixture DG-CPHD Measurement Update

The equations in this section establish the correction step for the Gaussian-mixture DG-CPHD recursion.

$$\begin{aligned} D_k(\mathrm{x}) &= D_{u,k}(\mathrm{x}) + D_{d,k}(\mathrm{x})\\ &= \sum_{\ell=1}^{\mathrm{I}_k} w^{(\ell)}_k\mathcal{N}(\mathrm{x};\mathrm{m}^{(\ell)}_k,\mathrm{P}^{(\ell)}_k), \end{aligned} \tag{56}$$

$$\alpha_k \approx \frac{\mu^2_{N,k}}{\sigma^2_{N,k}},\ \beta_k \approx \frac{\mu_{N,k}}{\sigma^2_{N,k}}, \tag{57}$$

$$\mu_{N,k} = \theta_{1,0} + \theta_{0,1}\cdot(1-p_d), \tag{58}$$

$$\sigma^2_{N,k} = \theta_{2,0} - \theta_{1,0} + 2\theta_{1,1} \cdot (1 - p_d) + \theta_{0,2} \cdot (1 - p_d)^2 - \mu^2_{N,k} + \mu_{N,k}, \quad (59)$$

where $D_{u,k}$ is modeled as a Gaussian mixture describing the intensity of a random finite set composed of missed detections and $D_{d,k}$ is modeled as a Gaussian mixture describing the intensity of a random finite set composed of detected targets. For m_k collected measurements, $\mathrm{I}_k = \mathrm{I}_{k|k-1} + m_k \mathrm{I}_{k|k-1}$, where $\mathrm{I}_{k|k-1}$ components correspond to the missed detections and $m_k \mathrm{I}_{k|k-1}$ components correspond to the update of each prior component by each measurement. The missed detection components are evaluated for $i = 1, \ldots, \mathrm{I}_{k|k-1}$ as

$$w^{(i)}_{u,k} = \frac{(1 - p_d)\Theta_k\left[\mathrm{Z}_k\right]}{\alpha_{k|k-1}\beta^{-1}_{k|k-1}} w^{(i)}_{k|k-1}, \quad (60)$$

$$\mathrm{m}^{(i)}_{u,k} = \mathrm{m}^{(i)}_{k|k-1}, \quad (61)$$

$$\mathrm{P}^{(i)}_{u,k} = \mathrm{P}^{(i)}_{k|k-1}. \quad (62)$$

For each measurement $\mathrm{z} \in \mathrm{Z}_k$, for all $i \in [1..\mathrm{I}_{k|k-1}]$, the detected components are given by

$$w^{(i)}_{d,k}(\mathrm{z}) = \frac{\Theta_k\left[\mathrm{Z}_k \setminus \{\mathrm{z}\}\right] p_d\, \hat{\ell}^{(i)}_{k,\mathrm{z}}}{\alpha_{k|k-1}\beta^{-1}_{k|k-1}\lambda c(\mathrm{z})} w^{(i)}_{k|k-1}, \quad (63)$$

$$\mathrm{m}^{(i)}_{d,k}(\mathrm{z}) = \mathrm{m}^{(i)}_{k|k-1} + \mathrm{K}^{(i)}_k(\mathrm{z} - \mathrm{Hm}^{(i)}_{k|k-1}), \quad (64)$$

$$\mathrm{P}^{(i)}_{d,k}(z) = (\mathbb{I}_{d_x} - \mathrm{K}^{(i)}_k \mathrm{H})\mathrm{P}^{(i)}_{k|k-1}, \quad (65)$$

$$\hat{\ell}^{(i)}_{k,\mathrm{z}} = \mathcal{N}(\mathrm{z}; \mathrm{Hm}^{(i)}_{k|k-1}, \mathrm{S}^{(i)}_{k|k-1}), \quad (66)$$

$$\mathrm{K}^{(i)}_k = \mathrm{P}^{(i)}_{k|k-1}\mathrm{H}^T\mathrm{S}^{(i)-1}_{k|k-1}, \quad (67)$$

$$\mathrm{S}^{(i)}_{k|k-1} = \mathrm{HP}^{(i)}_{k|k-1}\mathrm{H}^T + \mathrm{R}. \quad (68)$$

Thus, the posterior components in (56) are set as

$$\begin{pmatrix} w^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k, \\ \mathrm{m}^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k, \\ \mathrm{P}^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k \end{pmatrix} = \begin{cases} \begin{pmatrix} w^{(i)}_{u,k}, \\ \mathrm{m}^{(i)}_{u,k}, \\ \mathrm{P}^{(i)}_{u,k} \end{pmatrix}, & i \in [1..\mathrm{I}_{k|k-1}], \\ & j = 0; \\ \begin{pmatrix} w^{(i)}_{d,k}(\mathrm{z}_j), \\ \mathrm{m}^{(i)}_{d,k}(\mathrm{z}_j), \\ \mathrm{P}^{(i)}_{d,k}(\mathrm{z}_j) \end{pmatrix}, & i \in [1..\mathrm{I}_{k|k-1}], \\ & j \in [1..m_k]. \end{cases} \quad (69)$$

The Gaussian-Mixture DG-CPHD filter is presented in *Algorithm 2*. Theoretically, the filter induces a geometrically increasing number of Gaussian components as time progresses in a similar way as in the Gaussian-Mixture PHD filter. Hence, practical implementations require techniques to keep the total number of components at a maximum, such as pruning and merging components as described in [29]. In this procedure, components with negligible weights are pruned and components that are close together are merged. Tracks are extracted as per the extraction algorithm presented in [29].

F. Algorithm complexity and implementation details

In the context of the original CPHD filter, the attempt of propagating a nowhere vanishing cardinality distribution, $p_N(n)$, would be intractable since this would involve estimating an infinite number of terms, for $n \in \mathbb{N}_0$. However, as touched on in [11], it is generally safe to assume that, in practical problems, cardinality distributions are short or moderate tailed, which allows them to be truncated at some number $n = n_{\max} \geq \nu$. This corresponds to say that their probability generating functions are polynomials of degrees not exceeding ν. Therefore, when talking about complexity of the CPHD filter, the usual argument observes that, by assuming the prior cardinality p.g.f. to be a polynomial of degree not exceeding ν, i.e., $\deg G_{k|k-1}(x) \leq \nu$, then $\deg G_k(x) \leq \nu$, and in the next filtering step $\deg G_{k+1|k}(x) \leq \nu + \deg G_{b,k+1}(x)$ [11]. The conclusion of this argument is that increasing the possible number of targets in the scene will not directly affect the computational requirements for cardinality.

According to the analysis provided in [11] and [12], most of the CPHD filter computational effort is primarily due to $m + 1$ evaluations of sets of elementary symmetric functions, each computed with effort bounded by $\mathcal{O}(m^2)$ operations, which would render a complexity bounded by $\mathcal{O}(m^3)$ operations, for m measurements. This complexity can be reduced to $\mathcal{O}(m^2 \log^2 m)$ for a procedure mentioned in [12]. Nevertheless, such analysis emphasizes that evaluations of elementary symmetric functions will dominate the complexity for a high number of measurements, and so disregards about $(n_{\max}m - m^2/2 + n_{\max} - m)(m + 1)$ calculations[6] for $n_{\max} > m$, involving multiple-term multiplications[7] that are necessary to compute $\Upsilon_k[\mathrm{Z}_k]$, $\Upsilon_k[\mathrm{Z}_k \setminus \{\mathrm{z}\}]$ and $p_k(n)$, according to equations (15) and (18) in Section III-B.

Not less important, although not usually taken into account, is observing that as the number of targets increases, $n_{\max}$ must be set accordingly to properly describe probability masses that become increasingly important towards higher values of n, while maintaining accuracy of the cardinality representation. How much $n_{\max}$ should increase depends on how much information measurements provide about the number of targets, which ultimately depends on the signal-to-noise ratio.

Proposition 3. *Considering a constant probability of detection $p_d(x) = p_d$, Poisson-distributed false alarms at rate λ, and setting $\varepsilon \coloneqq \Pr\{0 < n \leq n_{max}\} \in (0, 1)$, it follows that*

$$n_{max} \geq \frac{2\mu_{N,k}}{\varepsilon}\left(1 + \sqrt{(1 - \varepsilon) - \varepsilon\left(\frac{\sigma_{N,k}}{\mu_{N,k}}\right)^2}\right) \geq K \cdot \mu_{N,k|k-1}, \quad (70)$$

for $K \geq 1$ and where $0 < \sigma^2_{N,k} \leq \mu_{N,k}n_{max} - \mu^2_{N,k}$ is assumed.

The point we wish to make is that $K \cdot \mu_{N,k|k-1} > m$ is not rare (e.g., high number of targets and low number of false alarms) and, in that case, the CPHD filter complexity should be bounded by approximately $m^3 + n_{\max}m^2 - m^2/2 \sim \mathcal{O}(n_{\max}m^2)$ operations since $n_{\max} > m$. The proof of *Proposition 3* is presented in the supplementary material. In general, as $\lambda \to 0$ the CPHD filter complexity is rather dominated by the cardinality parameterization and by the birth model, and becomes increasingly complex as the number of target-generated measurements increases. This fact corroborates Mahler's perception that further approximation of the CPHD algorithm may enable computational gain [13].

In the DG-CPHD filter context, the computational complexity is alleviated by the fact that, apart from the $\mathcal{O}(m^3)$ operations for the elementary symmetric functions, only about $\bar{n}_{\max}(m + 3) \sim \mathcal{O}(\bar{n}_{\max}m)$ operations are required for evaluating derivatives of the prior cardinality PGF. In this context, $\bar{n}_{\max}$ is the number of terms used to approximate each infinite

[6] Or $(n_{\max}m/2 + n_{\max} + 1)(n_{\max} + 1)$ for $n_{\max} < m$.

[7] Exponentiating sums of logarithms and multiplying by one elementary symmetric function.

sum in (44), and about $6m+m(m-1)$ operations are needed for the terms that sum up to give $\Theta_k[\mathrm{Z}_k]$, $\Theta_k[\mathrm{Z}_k \setminus \{\mathrm{z}\}]$ and $\theta_{u,v}$, according to equations (39) and (40) in Section V-D. The overall complexity of the DG-CPHD filter is then bounded by approximately $m^3 + \bar{n}_{\max} m + m^2 \sim \mathcal{O}(m^3)$ operations, where $\bar{n}_{\max}$ must be set to meet an adequate accuracy for evaluating the prior cardinality PGFL, but has loose relationship with the number of measurements. The scheme to compute $\mathrm{Li}^{(\ell)}_{-\alpha+1}(e^{-\beta}x)$ for $\ell \in [0..m+2]$, with approximately $\bar{n}_{\max}(m+3)$ operations, evaluates $\bar{n}_{\max}$ log-terms of $\mathrm{Li}^{(0)}_{-\alpha+1}(e^{-\beta}x)$, and then applies an iterative procedure over $m+2$ steps, each calculating $\bar{n}_{\max}$ log-increments to update the log-terms in view of (46). This procedure is made explicit in *Algorithm 1*.

Algorithm 1: Normalized Derivatives of Polylogarithms

Input : α, β, ζ, ε (machine precision), m

$l_\zeta \coloneqq \log\zeta$, $l_z \coloneqq -\beta + l_\zeta$

Compute bounds for number of terms

$\bar{n}_{\min} \coloneqq 1$, $\bar{n}_0 \coloneqq \hat{N}_{k|k-1}$

/* Find $\bar{n}_{\max}$ that corresponds to the smallest representable term $n^{\alpha-1}e^{-\beta n}$ by Newton's method */

for $i = 1, \ldots, N_{iter}$ **do**

  $\bar{n}_i \leftarrow \bar{n}_{i-1} - \frac{(\alpha-1)\log\bar{n}_{i-1} - \beta\bar{n}_{i-1} - \log\varepsilon}{2((\alpha-1)/\bar{n}_{i-1} - \beta)}$

end

$\bar{n}_{\max} \coloneqq \bar{n}_i$

$(n)_n = (n)_{n\in[\bar{n}_{\min}..\bar{n}_{\max}]} \coloneqq (\bar{n}_{\min}, \bar{n}_{\min}+1, \ldots, \bar{n}_{\max})$

$(nl_z)_n \coloneqq (n)_n \cdot l_z$

$(\log n)_n \coloneqq \log(n)_n$

$((\alpha-1)\log n)_n \coloneqq (\alpha-1)\cdot(\log n)_n$

$(\vartheta(n))_n \coloneqq ((\alpha-1)\log n)_n + (nl_z)_n$

$\vartheta_{\max} \coloneqq \max(\vartheta(n))_n$

Evaluate polynomials

$L(0) \coloneqq \sum_{n=1}^{\bar{n}_{\max}} \exp((\vartheta(n))_n - \vartheta_{\max})$

$l_{\zeta,1} \coloneqq l_\zeta$

$(\log(n^{\underline{1}}))_n \coloneqq (\log n)_n$

for $\ell = 1, \ldots, m+2$ **do**

  $L(\ell) \coloneqq \sum_{n=1}^{\bar{n}_{\max}} \exp((\vartheta(n))_n - \vartheta_{\max} + (\log(n^{\underline{\ell}}))_n - l_{\zeta,\ell})$

  $(\log(n^{\underline{\ell+1}}))_n \leftarrow (\log(n^{\underline{\ell}}))_n + (\log\max(n-\ell, 0))_n$

  $l_{\zeta,\ell+1} \leftarrow l_{\zeta,\ell} + l_\zeta$

end

Output: Approximations of higher-order derivatives

$$(\mathrm{Li}^{(\ell)}_{1-\alpha}(e^{-\beta}\zeta))_{\ell\in[0..m+2]} \propto \frac{(L(\ell))_{\ell\in[0..m+2]}}{\max\left|(L(\ell))_{\ell\in[0..m+2]}\right|}$$

VI. Numerical experiment

In this section we present experimental results for a simple example, very similar to the examples presented in [12], but modulating difficulty by varying either the number of targets that appear, average number of false alarms per frame, or probability of detection. The intent is to show differences of performance and computational effort of the PHD, CPHD and DG-CPHD filters in difficult situations, where we hope to

Algorithm 2: Gaussian-Mixture DG-CPHD filter

Input : $\{w^{(i)}_{k-1}, \mathrm{m}^{(i)}_{k-1}, \mathrm{P}^{(i)}_{k-1}\}_{i\in[1..\mathrm{I}_{k-1}]}$, α_{k-1}, β_{k-1}, $\mathrm{Z}_k = \{\mathrm{z}_1, \ldots, \mathrm{z}_{m_k}\}$

Prediction Step

/* Prediction of newborn target intensity */

for $i = 1, \ldots, \mathrm{I}_{b,k}$ **do**

  $w^{(i)}_{k|k-1} \coloneqq w^{(i)}_{b,k}$, $\mathrm{m}^{(i)}_{k|k-1} \coloneqq \mathrm{m}^{(i)}_{b,k}$, $\mathrm{P}^{(i)}_{k|k-1} \coloneqq \mathrm{P}^{(i)}_{b,k}$

end

/* Prediction of surviving target intensity */

$\mathrm{I}_{k|k-1} \coloneqq \mathrm{I}_{b,k} + \mathrm{I}_{k-1}$

for $i = \mathrm{I}_{b,k}+1, \ldots, \mathrm{I}_{k|k-1}$ **do**

  $w^{(i)}_{k|k-1} \coloneqq p_s w^{(i)}_{k-1}$,

  $\mathrm{m}^{(i)}_{k|k-1} \coloneqq \mathrm{F}\mathrm{m}^{(i)}_{k-1}$, $\mathrm{P}^{(i)}_{k|k-1} \coloneqq \mathrm{F}\mathrm{P}^{(i)}_{k-1}\mathrm{F}^T + \mathrm{Q}$

end

/* Prediction of cardinality parameters */

$\mu_{N,k|k-1} \coloneqq \sum_{i=1}^{\mathrm{I}_{b,k}} w^{(i)}_{b,k} + p_s\alpha_{k-1}\beta^{-1}_{k-1}$

$\sigma^2_{N,k|k-1} \coloneqq \mu_{N,k|k-1} + p_s^2\alpha_{k-1}\beta^{-1}_{k-1}(\beta^{-1}_{k-1}-1)$

$\alpha_{k|k-1} \coloneqq \mu^2_{N,k|k-1}/\sigma^2_{N,k|k-1}$

$\beta_{k|k-1} \coloneqq \mu_{N,k|k-1}/\sigma^2_{N,k|k-1}$

Measurement Update Step

/* Pre-computations for updated components */

for $i = 1, \ldots, \mathrm{I}_{k|k-1}$ **do**

  $\bar{\mathrm{z}}^{(i)}_{k|k-1} \coloneqq \mathrm{H}\mathrm{m}^{(i)}_{k|k-1}$, $\mathrm{S}^{(i)}_{k|k-1} \coloneqq \mathrm{H}\mathrm{P}^{(i)}_{k|k-1}\mathrm{H}^T + \mathrm{R}$,

  $\mathrm{K}^{(i)}_k \coloneqq \mathrm{P}^{(i)}_{k|k-1}\mathrm{H}^T\mathrm{S}^{(i)-1}_{k|k-1}$, $\mathrm{P}^{(i)}_k \coloneqq (\mathrm{I} - \mathrm{K}^{(i)}_k\mathrm{H})\mathrm{P}^{(i)}_{k|k-1}$

end

$\{\hat{\ell}^{(i)}_{k,\mathrm{z}}\} \triangleq \{\mathcal{N}(\mathrm{z}; \bar{\mathrm{z}}^{(i)}_{k|k-1}, \mathrm{S}^{(i)}_{k|k-1})\}_{i\in[1..\mathrm{I}_{k|k-1}],\mathrm{z}\in\mathrm{Z}_k}$

For each $j \in [1..m_k]$ compute $\langle p_d\ell_{\mathrm{z}_j}, \varsigma\rangle = \alpha^{-1}_{k|k-1}\beta_{k|k-1}p_d\sum_{i=1}^{\mathrm{I}_{k|k-1}} w^{(i)}_{k|k-1}\hat{\ell}^{(i)}_{k,\mathrm{z}_j}$

Obtain $\{\bar{\sigma}_j(\mathrm{Z})\}_{j=1}^{|\mathrm{Z}|}$ (41), for $\mathrm{Z} = \mathrm{Z}_k, \mathrm{Z}_k \setminus \{\mathrm{z}\} : \mathrm{z} \in \mathrm{Z}_k$

Compute $(\mathrm{Li}^{(\ell)}_{1-\alpha_{k|k-1}}(e^{-\beta_{k|k-1}}q_d))_{\ell\in[0..m_k+2]}$ (*Alg. 1*)

Evaluate $\theta_{0,1} \equiv \Theta_k[\mathrm{Z}_k]$, $\theta_{0,2}$, $\theta_{1,0}$, $\theta_{2,0}$, $\theta_{1,1}$ (40)

/* Update of missed-detection intensity */

for $i = 1, \ldots, \mathrm{I}_{k|k-1}$ **do**

  $w^{(i)}_k \coloneqq q_d\,\alpha^{-1}_{k|k-1}\beta_{k|k-1}\Theta_k[\mathrm{Z}_k]\,w^{(i)}_{k|k-1}$,

  $\mathrm{m}^{(i)}_k \coloneqq \mathrm{m}^{(i)}_{k|k-1}$, $\mathrm{P}^{(i)}_k \coloneqq \mathrm{P}^{(i)}_{k|k-1}$

end

/* Update of detected target intensity */

$\mathrm{I}_k \coloneqq (1+m_k)\mathrm{I}_{k|k-1}$

for $j = 1, \ldots, m_k$ **do**

  Evaluate $\Theta_k[\mathrm{Z}_k \setminus \{\mathrm{z}_j\}]$ (39)

  for $i = 1, \ldots, \mathrm{I}_{k|k-1}$ **do**

    $w^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k \coloneqq \frac{\Theta_k[\mathrm{Z}_k\setminus\{\mathrm{z}_j\}]p_d\,\hat{\ell}^{(i)}_{k,\mathrm{z}_j}w^{(i)}_{k|k-1}}{\alpha_{k|k-1}\beta^{-1}_{k|k-1}\gamma c(\mathrm{z}_j)}$,

    $\mathrm{m}^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k \coloneqq \mathrm{m}^{(i)}_{k|k-1} + \mathrm{K}^{(i)}_k(\mathrm{z}_j - \mathrm{H}\mathrm{m}^{(i)}_{k|k-1})$,

    $\mathrm{P}^{(j\cdot\mathrm{I}_{k|k-1}+i)}_k \coloneqq \mathrm{P}^{(i)}_k$

  end

end

/* Update of cardinality parameters */

$\mu_{N,k} \coloneqq \theta_{1,0} + \theta_{0,1}\cdot q_d$

$\sigma^2_{N,k} \coloneqq \theta_{2,0} - \theta_{1,0} + 2\theta_{1,1}\cdot q_d + \theta_{0,2}\cdot q_d^2 - \mu^2_{N,k} + \mu_{N,k}$

$\alpha_k \coloneqq \mu^2_{N,k}/\sigma^2_{N,k}$, $\beta_k \coloneqq \mu_{N,k}/\sigma^2_{N,k}$

Output: $\{w^{(i)}_k, \mathrm{m}^{(i)}_k, \mathrm{P}^{(i)}_k\}_{i\in[1..\mathrm{I}_k]}$, α_k, β_k

demonstrate benefits of the approximations introduced by the DG-CPHD filter.

The example consists of a two-dimensional scenario in the region $[-1000, +1000] \times [-1000, +1000]$ (m × m), where a number of targets, unknown a priori, may appear and are observed under the possibility of false alarms. Each target is described by its state vector $\mathrm{x} = (p_x, p_y, v_x, v_y)^T$, where $\mathrm{p} = (p_x, p_y)^T$ is a pair that specifies a position in Cartesian coordinates and $\mathrm{v} = (v_x, v_y)^T$ is the pair specifying velocity in the same coordinates. At each time instant k, the state is written as $\mathrm{x}_k = \mathrm{x}(t_k)$. Each target is assumed to move with nearly-constant velocity, with transition matrix and state process covariance matrix given respectively by

$$\mathrm{F} = \begin{pmatrix} \mathbb{I}_2 & \mathbb{I}_2 \Delta t \\ 0_2 & \mathbb{I}_2 \end{pmatrix}, \; \mathrm{Q} = \begin{pmatrix} \mathbb{I}_2 \Delta t^3/3 & \mathbb{I}_2 \Delta t^2/2 \\ \mathbb{I}_2 \Delta t^2/2 & \mathbb{I}_2 \Delta t \end{pmatrix} \sigma_q^2,$$

where $\mathbb{I}_2$ and 0_2 are the identity and zero matrices with dimensions 2×2, respectively, $\Delta t = 1\,\mathrm{s}$ is the sampling period, and the standard deviation of velocity increments is given by $\sigma_q = 5\,\mathrm{m/s}^{\frac{3}{2}}$. Each target remains in the scene up to the next time step with probability $p_s = 0.98$. Note that the generation of targets' trajectories is performed without noise, that is, the targets (i.e., ground truth) move according to a constant velocity model.

A single sensor collects position measurements in the Cartesian space, corrupted by a Gaussian-distributed noise, characterized by the output matrix and measurement noise covariance matrix, respectively, $\mathrm{H} = \begin{pmatrix} \mathbb{I}_2 & 0_2 \end{pmatrix}$ and $\mathrm{R} = \mathbb{I}_2 \sigma_r^2$, where $\sigma_r = 5\,\mathrm{m}$ is the standard deviation of the measured positions. False alarms can be generated according to a Poisson point-process with intensity $D_c(\mathrm{z}) = \lambda \cdot c(\mathrm{z})$, where λ is the average number of false alarms per scan, and $c(\mathrm{z})$ is the spatial distribution of clutter, assumed uniform in the surveillance region with "volume" (area) $V = 2000^2\,\mathrm{m}^2$.

Each instance of the example is simulated for $T = 100\,\mathrm{s}$. Denote N_t as the total number of targets that appear and will remain till the end of a simulated instance, at $t = 100\,\mathrm{s}$. In all cases, targets appear in batches at positions uniformly sampled in the area $[-800, +800] \times [-800, +800]$ (m × m), and with random velocities uniformly sampled in the ranges $[-5, +5] \times [-5, +5]$ (m/s × m/s). The batches of target appearance are set as follows:

- $0.25N_t$ targets are already in the scene at $t = 0$ and will remain up to $t = 100\,\mathrm{s}$ with exception of 5 targets that are set to disappear at $t = 80\,\mathrm{s}$,
- $0.25N_t$ targets along with 2 other targets appear at $t = 20\,\mathrm{s}$ and are set to remain to the end,
- $0.25N_t$ of targets along with another target appear at $t = 40\,\mathrm{s}$ and are set to remain,
- $0.25N_t$ of targets along with 2 other targets appear at $t = 60\,\mathrm{s}$ and are set to remain,
- from $N_t + 5$ targets in the scene, 5 targets disappear at $t = 80\,\mathrm{s}$ and the remaining N_t stay up to $t = 100\,\mathrm{s}$.

The birth random finite set is assumed as a Poisson point-process with approximate intensity $D_b(\mathrm{x}) = \sum_{i=1}^{5} w_b^{(i)} \mathcal{N}(\mathrm{x}; \mathrm{m}_b^{(i)}, \mathrm{P}_b^{(i)})$, where

$$\mathrm{m}_b^{(1)} = (-500, -500, 0, 0)^T, \mathrm{m}_b^{(2)} = (-500, +500, 0, 0)^T,$$
$$\mathrm{m}_b^{(3)} = (+500, -500, 0, 0)^T, \mathrm{m}_b^{(4)} = (+500, +500, 0, 0)^T,$$
$$\mathrm{m}_b^{(5)} = (0, 0, 0, 0)^T,$$

and

$$w_b^{(i)} = \frac{N_t}{10T/\Delta t}\,(\text{target/frame}),$$

$$\mathrm{P}_b^{(i)} = \begin{pmatrix} 200^2 \mathbb{I}_2 & 0_2 \\ 0_2 & 4^2 \mathbb{I}_2 \end{pmatrix}, \; i = 1, \ldots, 5.$$

For each filter, at each time step, pruning is performed based on a weight threshold of $\tau_{\text{prn}} = 10^{-5}$ and merging with threshold of $\tau_{\text{mrg}} = 4\,\mathrm{m}$, and the number of maintained Gaussian components is limited at $\mathrm{J}_{\max} = 300$ (see [29] for details on the pruning and merging procedure). Measurements are gated with gate-size probability of $p_{\text{gate}} = 0.999$. For the CPHD filter, the estimated (posterior) number of targets is obtained in the maximum-a-posteriori sense. The cardinality distribution for the CPHD filter is estimated to a maximum of $n_{\max} = 2 \times N_t$ terms. This maximum number of cardinality terms has been chosen to keep the CPHD filter computational effort competitive in relation to the other filters for difficult scenarios.

We examine the DG-CPHD filter in comparison with the PHD and CPHD filters for three different cases as follows.

Case 1: Fixed probability of detection, $p_d = 0.98$, and fixed number of false alarms per scan, $\lambda = 50$, all filters are tested for different numbers of targets that appear and remain in the scene, $N_t \in \{10, 25, 50, 75, 100\}$.

Case 2: For fixed number of targets that appear and remain in the scene, $N_t = 50$, and fixed number of false alarms per scan, $\lambda = 50$, all filters are tested for different probabilities of detection, $p_d \in \{0.98, 0.90, 0.80, 0.70, 0.60\}$.

Case 3: For fixed number of targets that appear and remain in the scene, $N_t = 50$, and fixed probability of detection, $p_d = 0.80$, all filters are tested for different numbers of false alarms per frame, $\lambda \in \{10, 30, 50, 100, 200\}$.

For each case, 500 Monte Carlo (MC) runs are performed, for a sampled set of trajectories (ground truth), independently generated clutter, and independently generated (target-originated) measurements for each trial. For all filters, performance is evaluated in terms of:

- mean Optimal Subpattern Assignment (OSPA) metric [30] with a cut-off $c_{\text{OSPA}} = 100$ and a norm order $p_{\text{OSPA}} = 1$,
- root-mean-square error (RMSE) of the estimated target number,
- integrated time-weighted absolute error (ITAE) with normalized weights for the estimated target number, and
- computation time (per time step).

Consolidated cardinality RMSE and mean OSPA indexes are obtained by averaging the cardinality RMSE and MOSPA over the time steps where all filters are in steady state (disregarding the errors during the settling time). Considering a filter that estimates the target number as a "black box" for an estimated number of objects and a reference ground truth (cf. a set point) in target number, these indexes are defined invoking a common notion of steady-state error for control systems, given a step input (birth events). This implies an evaluation where any response before a settling time is not considered, and this is generally adequate for measuring non-transient precision. These steady-state time steps comprise $t \in [15, 20) \cup [30, 40) \cup [50, 60) \cup [70, 80) \cup [90, 100]$ (s) in the experiment. Since the targets (ground truth) are moving with constant velocity, these indexes at steady state measure performance under the least dynamic, least stressful parts of the scenarios with respect to the target number. However, the

fact that targets are moving along straight lines facilitates the filtering task as a whole, i.e., both when the number of targets is changing or steady.

The consolidated runtimes are obtained by taking the average of the computation times per time step. We obtain two sets of computation times: the total runtimes per filtering cycle and the runtimes without accounting for the time spent to manage the Gaussian mixtures (pruning and merging). Along with the consolidated cardinality RMSE we also provide the ITAE index, which is a performance index that favors the steady-state performance but does not disregard the transient phase. The ITAE is a weighted average of the absolute errors over time, where the weights are proportional to the elapsed time with respect to the latest birth/death event onset. The ITAE weights are normalized to sum to one. These consolidated indexes are computed for each value of the varying parameters N_t, p_d, and λ.

In our comparisons, we use implementations of the PHD and CPHD filters in Matlab language retrieved from Ba-Tuong Vo's webpage[8] but incorporating improvements in favor of numerical robustness and computational efficiency, allowing for higher numbers of false alarms per frame and reducing the computational cost by calculating the combinatorial coefficients only once. The DG-CPHD filter code was also written in Matlab with a similar structure[9].

A. Results

Case 1 : Figure 2 shows the tracks generated by the DG-CPHD filter for an exemplary run of the first case, in the subcase where $p_d = 0.98$, $\lambda = 50$ and $N_t = 100$. For the same subcase, Figures 3 and 4 present, respectively, the mean OSPA and cardinality estimates over time for the PHD, CPHD and DG-CPHD filters, where we can perceive the advantage of estimating the cardinality distribution.

The consolidated performance indexes for Case 1, comprising the average steady-state errors, ITAE and the average runtimes per time step for each $N_t \in \{10, 25, 50, 75, 100\}$, can be found in Figures 5–7. In these figures, it becomes clear that the performance of the DG-CPHD filter is similar to that of the CPHD filter but at a smaller computational cost. In terms of steady-state cardinality RMSE, ITAE, and mean OSPA, the CPHD filter is the best filter.

For all filters, the steady-state mean OSPA metric seems to converge to an asymptotic value as the target number increases. Also, the steady-state cardinality RMSE and ITAE of all filters seem to increase sub-exponentially (at a small rate) with the possible number of targets in the scene. In this case, the false alarm rate is not extremely high, thus the complexity of the DG-CPHD filter is dominated by the algorithm steps bounded by $\mathcal{O}(m^3)$ operations whereas the complexity of the CPHD filter is dominated by the algorithm steps bounded by $\mathcal{O}(n_{\max} m^2)$ operations with $n_{\max} > m$. This can be observed from the consolidated runtime trends in Figure 7 as $\mathbb{E}[m] \propto N_t$.

Case 2 : The consolidated performance indexes for Case 2, comprising the average steady-state errors, ITAE, and the average runtimes per time step for each $p_d \in \{0.98, 0.90, 0.80, 0.70, 0.60\}$, can be found in Figures 8–10. In this case, the performance of the DG-CPHD filter is very

[8]http://ba-tuong.vo-au.com/codes.html.

[9]Code available at https://github.com/femelo/dg-cphd-filter.

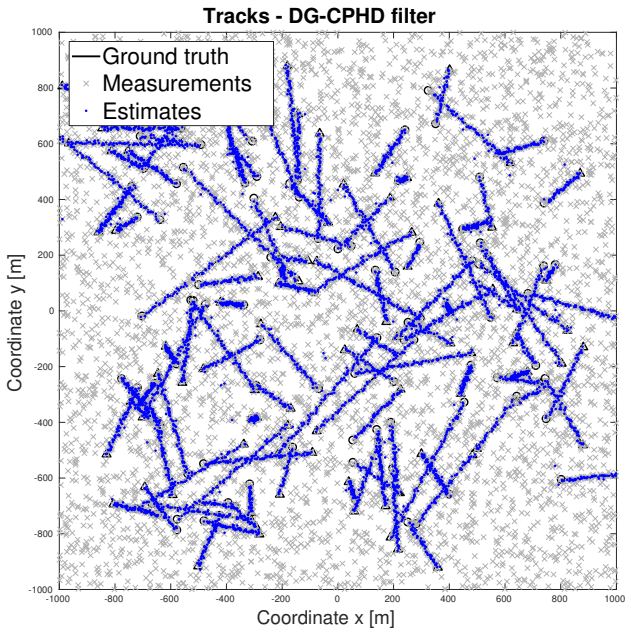

Figure 2. Exemplary run: tracks estimated by the DG-CPHD filter

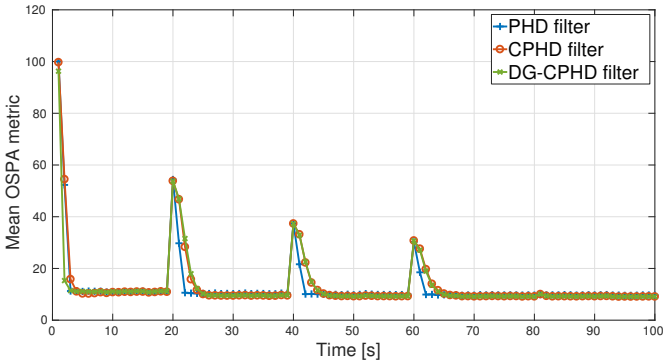

Figure 3. Mean OSPA metric over time ($p_d = 0.98$, $\lambda = 50$, $N_t = 100$)

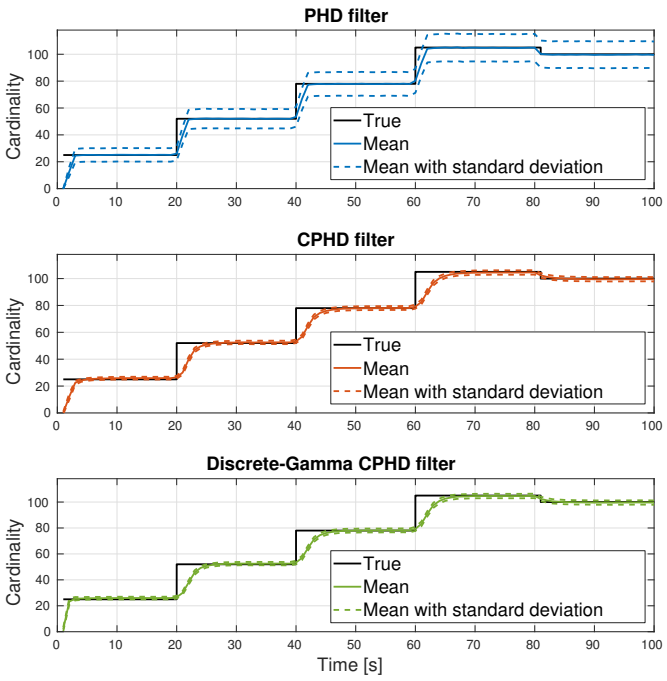

Figure 4. Cardinality estimates over time ($p_d = 0.98$, $\lambda = 50$, $N_t = 100$)

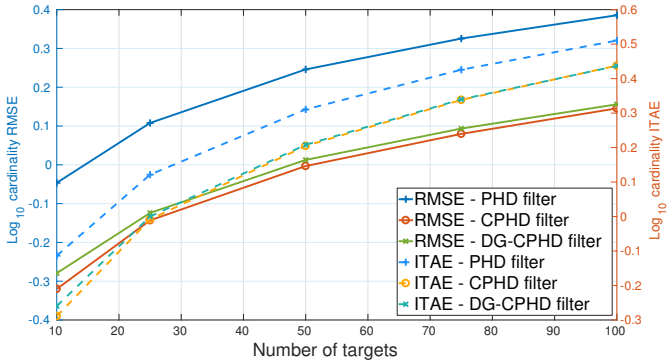

Figure 5. Case 1: Cardinality RMSE, ITAE vs. number of targets

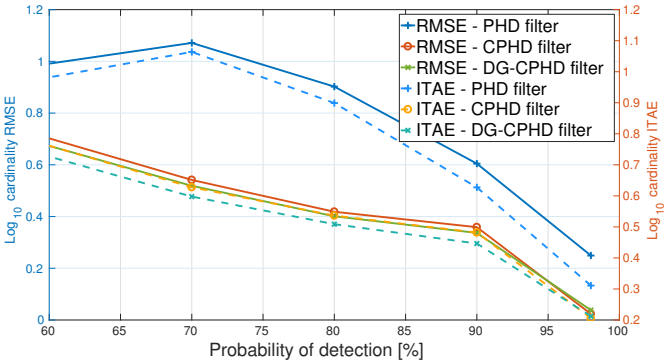

Figure 8. Case 2: Cardinality RMSE, ITAE vs. probability of detection

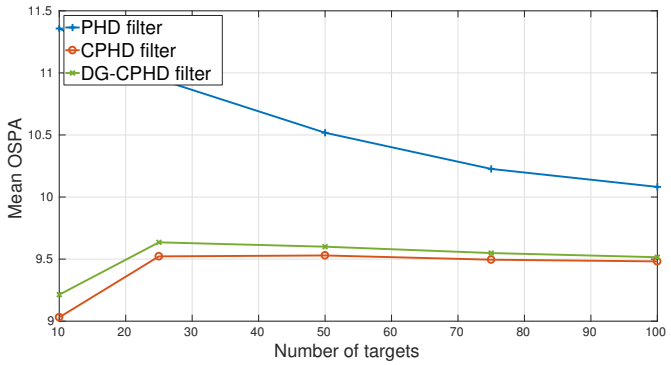

Figure 6. Case 1: Steady-state mean OSPA vs. number of targets

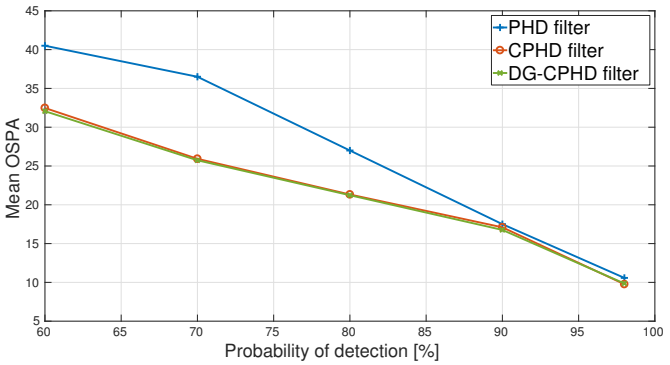

Figure 9. Case 2: Steady-state mean OSPA vs. probability of detection

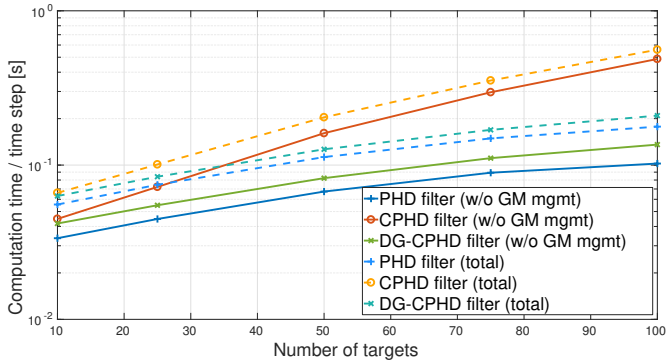

Figure 7. Case 1: Average runtime vs. number of targets

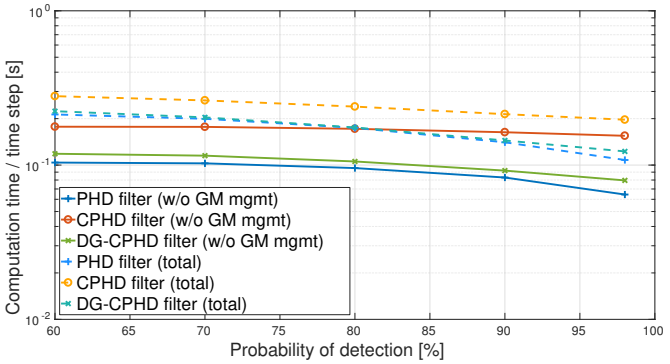

Figure 10. Case 2: Average runtime vs. probability of detection

similar to that of the CPHD filter in terms of steady-state mean OSPA, cardinality RMSE, and ITAE, with only marginal differences. As expected, for all filters, the steady-state error indexes decrease as the probability of detection increases.

As can be observed in Figure 10, the computational cost of the DG-CPHD filter is smaller than that of the CPHD filter, and greater than that of the PHD filter. It is worth noting that almost no difference of computational cost is observed over distinct detection probabilities: the slight decrease in runtimes for higher detection probabilities is associated with lower times for updating[10] Gaussian components with lesser spread.

We note that, especifically for this case, the slightly better performance of the DG-CPHD filter in comparison to the CPHD filter is surprising, but this is not central to our argument that the DG-CPHD offers performance commensurate to that of the CPHD filter at a lower computational cost. It appears that these differences are rather related to the different sensitivity of the algorithms to the filtering parameters for scenarios where the probability of detection is decreased, and most probably are not statistically significant as they are small but seem exacerbated by the logarithm scale.

Case 3 : Although the scenario is difficult for any filter, it is evident that the CPHD and DG-CPHD filters perform better than the PHD filter, being able to identify and track a greater number of targets on average. The DG-CPHD filter has virtually the same performance as that of the CPHD filter.

Figures 11–13 present the consolidated performance indexes for Case 3, for each $\lambda \in \{10, 30, 50, 100, 200\}$. From the figures, the overall performance of the DG-CPHD filter is presented as almost indistinguishable from that of the CPHD filter. The cardinality RMSE and ITAE of the DG-CPHD filter are commensurate to those of the CPHD filter. From Figure 13, one can observe that as λ increases the complexity of both the CPHD and DG-CPHD filters are dominated by the total number of measurements, i.e., the runtimes increase sub-exponentially with the number of false alarms, bounded by $\mathcal{O}(m^3)$ operations. Furthermore, the steady-state cardinality RMSE and ITAE of all filters seem to increase

[10]Kalman data update.

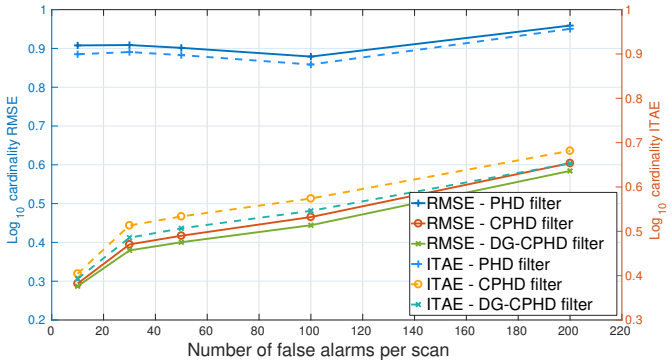

Figure 11. Case 3: Cardinality RMSE, ITAE vs. number of false alarms

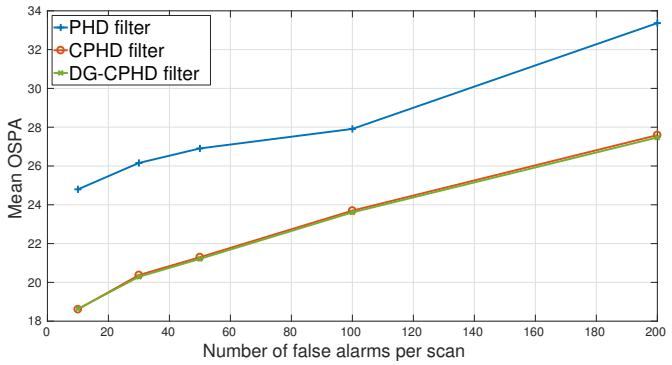

Figure 12. Case 3: Steady-state mean OSPA vs. number of false alarms

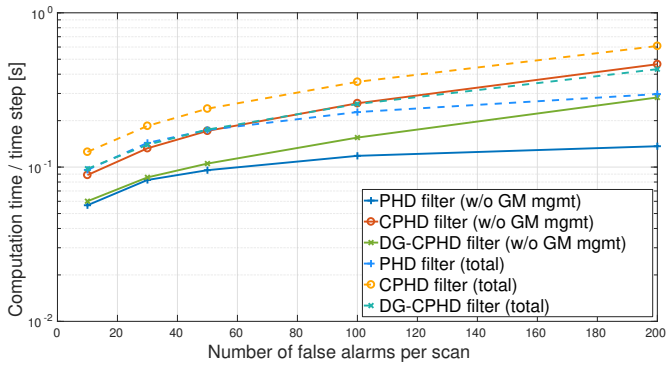

Figure 13. Case 3: Average runtime vs. number of false alarms

sub-exponentially with the average number of false alarms, suggesting a dependency of the signal-to-noise ratio that is polynomial in the number of measurements.

Also for this case, we remark that the slightly better performance of the DG-CPHD filter in comparison to the CPHD filter is surprising, but this is not central to our argument that the DG-CPHD offers performance commensurate to that of the CPHD filter at a lower computational cost. As for the previous case, the differences are likely to be related to the different sensitivity of the filters with respect to the parameters for scenarios where the average number of false alarms is increased, and most probably are not statistically significant as they are rather small but exacerbated by the logarithm scale.

It is important to observe that, in theory, when the number of targets abruptly changes at time steps $k = 20,\ 40,\ 60,\ 80$, occasionally the cardinality distribution might become briefly bimodal. In these occasions, the cardinality distribution would not be accurately approximated by a discrete Gamma distribution, which would make the variance for the DG-CPHD filter significantly larger than that for the CPHD filter. However, a thorough investigation of this issue reveals that this expected behavior is not confirmed in many practical cases. Actually, what one sees for the CPHD filter at the steps $k = 20,\ 40,\ 60,\ 80$, is a unimodal cardinality distribution with a mode peaking between the previous target number estimate and the actual (i.e., ground truth) new target number. The mode of the distribution appears to shift progressively after the birth event until it reaches the correct target number in later steps. This happens to such an extent that the usual procedure to obtain estimates, $\hat{N}_k = \operatorname{argmax}_n p_k(n|\mathrm{Z}_{1:k})$ (as recommended in [11], equation (88)), immediately after steps with birth events, always finds an estimate between the predicted target number and the ground truth. This means that the CPHD filter does not produce a MAP estimate either at 20 or 40 (for step $k = 20$), at 40 or 60 (for step $k = 40$), or at 60 or 80 (for step $k = 60$). The same behavior of the CPHD filter producing these intermediate cardinality MAP estimates from a single mode can be clearly seen in [12, 14]. This practical lack of bimodality in the cardinality distribution explains the similarity in the performance of the DG-CPHD filter and the CPHD filter.

VII. Conclusions

This paper proposes a new filter that is second-order in target number, where the multi-target state is assumed to follow an independent and identically distributed cluster process with the cardinality distribution modeled as a discretized Gamma distribution. Our work capitalizes on Ronald Mahler's perception that one more step towards simplification of the CPHD filter implementation might be interesting. The strategy employed was to mimic, based on a discrete-Gamma distribution, the procedure of a Kalman filter for the cardinality random variable, i.e., estimating sufficient statistics. As demonstrated by the numerical examples, the discrete Gamma distribution allows simple calculations for approximating the first- and second-order moments of the posterior cardinality distribution, and efficiently addresses tracking scenarios with underdispersed target counts, without the restrictions required by the binomial filter.

The results also demonstrate that the DG-CPHD filter is more computationally efficient than the CPHD filter implementation proposed in [12], especially for scenarios where a high number of CPHD cardinality terms $n_{\max}$ is necessary, i.e., in situations where the number of target-generated measurements is significantly increased as many targets appear in a scene. The experimental results support our argument that the computational complexity of the CPHD filter is dominated by $\mathcal{O}(n_{\max}m^2)$ in this case, where $n_{\max} > m$, when the CPHD filter is more computationally complex than the DG-CPHD filter. In contrast, the DG-CPHD filter must rely on a finite number of terms, $\bar{n}_{\max}$, to approximate derivatives of the prior cardinality p.g.f., $\hat{G}^{(\ell)}_{k|k-1}(\langle q_d, \varsigma\rangle)$, and as the results show, $\bar{n}_{\max}$ is much less sensitive to the number of target-generated measurements. To conclude, both our complexity analysis and the experimental results suggest that the DG-CPHD filter exhibits computational performance somewhere between that of the PHD filter and that of the CPHD filter, but cardinality accuracy and variance similar to that of the CPHD filter [12].

We conjecture that, while the true cardinality distribution might be multimodal at time steps where the target number abruptly changes, the fact that all the filters are assuming

that all targets are drawn from a common spatial distribution may add a smoothing in the process such that it induces an approximating multi-target density that has a unimodal cardinality distribution. This may explain the similarities between the CPHD and the DG-CPHD filters. While this observation is interesting, fully investigating this conjecture requires further work. We also conjecture that the procedure proposed in this paper could be extended to other cardinality distributions of interest and might be the basis for other filters.

REFERENCES

[1] D. B. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Transactions on Automatic Control*, vol. 24, pp. 843–854, December 1979.

[2] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Multi-target tracking using joint probabilistic data association," in *19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, pp. 807–812, December 1980.

[3] R. P. S. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, pp. 1152–1178, October 2003.

[4] J. E. Moyal, "The General Theory of Stochastic Population Processes," *Acta Mathematica*, vol. 108, pp. 1–31, 1962.

[5] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, Inc., 2007.

[6] R. P. S. Mahler, "'Statistics 102' for Multisource-Multitarget Detection and Tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 376–389, June 2013.

[7] B. T. Vo and B. N. Vo, "Labeled Random Finite Sets and Multi-Object Conjugate Priors," *IEEE Transactions on Signal Processing*, vol. 61, pp. 3460–3475, July 2013.

[8] B. N. Vo, B. T. Vo, and D. Phung, "Labeled Random Finite Sets and the Bayes Multi-Target Tracking Filter," *IEEE Transactions on Signal Processing*, vol. 62, pp. 6554–6567, December 2014.

[9] O. Erdinc, P. Willett, and Y. Bar-Shalom, "Probability Hypothesis Density Filter for Multitarget Multisensor Tracking," in *FUSION 2005, Proceedings of the 8th International Conference on Information Fusion*, vol. 1, p. 8, July 2005.

[10] R. P. S. Mahler, "A theory of PHD filters of higher order in target number," in *Proc. SPIE 6235, Signal Processing, Sensor Fusion, and Target Recognition XV*, vol. 6235, pp. 62350K–62350K–12, May 2006.

[11] R. P. S. Mahler, "PHD Filters of Higher Order in Target Number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, pp. 1523–1543, October 2007.

[12] 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, pp. 3553–3567, July 2007.

[13] R. P. S. Mahler, "PHD Filters of Second Order in Target Number," in *Proc. SPIE 6236, Signal and Data Processing of Small Targets*, vol. 6236, pp. 62360P–62360P–12, 2006.

[14] 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, pp. 1657–1664, October 2009.

[15] I. Schlangen, E. D. Delande, J. Houssineau, and D. E. Clark, "A Second-Order PHD Filter With Mean and Variance in Target Number," *IEEE Transactions on Signal Processing*, vol. 66, pp. 48–63, January 2018.

[16] E. Delande, M. Üney, J. Houssineau, and D. E. Clark, "Regional Variance for Multi-Object Filtering," *IEEE Transactions on Signal Processing*, vol. 62, pp. 3415–3428, July 2014.

[17] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, vol. Volume I: Elementary Theory and Methods of *Probability and its Applications*. New York, Berlin, Paris: Springer, 2003.

[18] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, vol. Volume II: General Theory and Structure of *Probability and its Applications*. New York, Berlin, Paris: Springer, 2008.

[19] P. Del Moral and J. Houssineau, *Particle Association Measures and Multiple Target Tracking*, pp. 1–30. Tokyo: Springer Japan, 2015.

[20] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. Needham, MA, USA: Artech House, 2014.

[21] H. W. Watson and F. Galton, "On the probability of the extinction of families," *The Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 4, pp. 138–144, 1875.

[22] A. M. Abouammoh and N. S. Alhazzani, "On Discrete Gamma Distribution," *Communications in Statistics - Theory and Methods*, vol. 44, no. 14, pp. 3087–3098, 2015.

[23] D. Wood, "The Computation of Polylogarithms," Tech. Rep. 15-92, University of Kent, Computing Laboratory, University of Kent, Canterbury, UK, June 1992.

[24] P. C. Consul and G. C. Jain, "A Generalization of the Poisson Distribution," *Technometrics*, vol. 15, no. 4, pp. 791–799, 1973.

[25] P. Consul, S. Kotz, and F. Famoye, *Lagrangian Probability Distributions*. Birkhäuser Boston, 2006.

[26] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright, "A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 1, pp. 127–142, 2005.

[27] S. Chakraborty and D. Chakravarty, "Discrete Gamma Distributions: Properties and Parameter Estimations," *Communications in Statistics - Theory and Methods*, vol. 41, no. 18, pp. 3301–3324, 2012.

[28] T. M. Apostol, "An Elementary View of Euler's Summation Formula," *The American Mathematical Monthly*, vol. 106, no. 5, pp. 409–418, 1999.

[29] B.-N. Vo and W. K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4091–4104, November 2006.

[30] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A Consistent Metric for Performance Evaluation of Multi-Object Filters," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3447–3457, August 2008.