Automatica 98 (2018) 331–340

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# Particle Gaussian mixture filters-I☆

Dilshad Raihan, Suman Chakravorty *

*Department of Aerospace Engineering, Texas A & M University, College Station, TX, 77840, United States*

ARTICLE INFO

*Article history:*
Received 18 January 2017
Received in revised form 23 March 2018
Accepted 5 July 2018
Available online 14 September 2018

*Keywords:*
Estimation algorithms
Nonlinear filters
Gaussian mixture models
Multimodality
Particle filtering
State estimation
Machine learning
Curse of dimensionality
Kalman filters

ABSTRACT

In this paper, we propose a particle based Gaussian mixture filtering approach for nonlinear estimation that is free of the particle depletion problem inherent to most particle filters. We employ an ensemble of possible state realizations for the propagation of state probability density. A Gaussian mixture model (GMM) of the propagated uncertainty is then recovered by clustering the ensemble. The posterior density is obtained subsequently through a Kalman measurement update of the mixture modes. We prove the convergence in probability of the resultant density to the true filter density assuming exponential forgetting of initial conditions. The performance of the proposed filtering approach is demonstrated through several test cases and is extensively compared to other nonlinear filters.



## 1. Introduction

Rapid advances in the fields of control and automation have inspired a significant interest in recursive and computationally efficient algorithms for estimating the state and associated uncertainty in higher dimensional nonlinear systems. The Kalman filter provides the unbiased minimum variance estimator for linear dynamical systems perturbed by additive Gaussian noise (Kalman, 1960; Kalman & Bucy, 1961). The extended Kalman filter (EKF) was introduced to incorporate nonlinear systems into the Kalman filtering framework (Smith, Schmidt, & McGee, 1962). However, the limitations of the Jacobian linearization assumptions and the accumulation of linearization errors can result in the divergence of EKF estimates. The Unscented Kalman Filter (UKF) and the broader class of sigma point Kalman filters provided a derivative free alternative to the EKF (Julier, Uhlmann, & Durrant-Whyte, 1995; Wan & Van Der Merwe, 2001). Both EKF and UKF approximate the posterior probability density function (pdf) with a single Gaussian pdf. However, the state pdf in a general nonlinear filtering problem can be non-Gaussian and multimodal. Cubature Kalman filters (CKF) that rely on a spherical–radial cubature rule to evaluate the integrals involved in the estimation have been proposed (Arasaratnam & Haykin, 2009). A variant of cubature filters, that perform random scaling and rotation of cubature points and axes, known as stochastic integration filter (SIF) has also been proposed recently (Dunik, Straka, Simandl, & Blasch, 2015). A Gaussian mixture approximation of the state pdf was proposed to incorporate the multimodality of the problem in nonlinear settings (Alspach & Sorenson, 1972; Sorenson & Alspach, 1971). These approaches however had a major shortcoming as the number of Gaussian components were fixed initially and kept constant through out the estimation process. Also the component weights were updated only during the measurement update. Approaches to adapting the weights of individual Gaussian modes by minimizing the propagation error committed in the Gaussian mixture model (GMM) approximation have been proposed recently (Terejanu, Singla, Singh, & Scott, 2011). A different approach to improving the accuracy of GMM filters is by splitting the Gaussian components during the propagation based on nonlinearity induced distortion (DeMars, Bishop, & Jah, 2013). Both of these approaches require frequent optimizations, or entropy calculations, to be performed during the propagation, which significantly add to the overall computational requirement. A Gaussian mixture 'blob' filter that relies on EKF for propagation and update has been proposed recently (Psiaki, 2016). It performs a resampling step between the propagation and update stages. The resampling step ensures that the component covariances of

DOI of original article: http://dx.doi.org/10.1016/j.automatica.2018.07.024.
☆ This work was supported by the AFOSR grants FA9550-13-1-0074, FA9550-17-1-0068 under the Dynamic Data Driven Application Systems (DDDAS) program and the NSF award ECCS1637889. The material in this paper was partially presented at the 19th International Conference on Information Fusion (FUSION), July 5–8, 2016, Heidelberg, Germany. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Torsten Söderström.
* Corresponding author.
*E-mail addresses:* dilshadraihan@tamu.edu (D. Raihan), schakrav@tamu.edu (S. Chakravorty).

https://doi.org/10.1016/j.automatica.2018.07.023
0005-1098/

the propagated pdf all obey a linear matrix inequality (LMI) based upper bound.

The particle filters (PF) are a class of sequential Monte Carlo methods that employ an ensemble of possible state realizations known as particles to represent the state pdf (Arulampalam, Maskell, Gordon, & Clapp, 2001; Gordon, Salmond, & Smith, 1993). These states are sampled from the initial pdf and propagated forward in time based on the nonlinear system model. The measurement updates are performed by assigning weights to individual particles which may then be resampled. The PF does not enforce restrictive assumptions on the nature of dynamics or pdf. However Particle filters are subject to the curse of dimensionality due to the particle depletion problem wherein a significant fraction of particles lose their importance weights during the measurement update. Preventing depletion requires the number of particles to be increased exponentially with the dimension of state space (Bengtsson, Bickel, & Li, 2008). Particle based approaches such as the Ensemble Kalman filter (EnKF) and the Feedback particle filter (FPF) that forego the resampling based measurement update have been demonstrated to be effective in higher dimensional filtering problems involving unimodal pdfs (Evensen, 2002; Yang, Mehta, & Meyn, 2013). The Gaussian sum particle filter (GSPF) is a nonlinear filter that uses a GMM representation of the state pdf (Kotecha & Djuric, 2003). It obtains an ensemble of particles from each GMM component. The ensembles are then propagated forward separately like a parallel bank of filters. The GSPF relies on an importance sampling based approach to perform the measurement update.

In this paper, we propose a particle Gaussian mixture (PGM) filter for nonlinear estimation. The PGM filter design is inspired by a previous work on a UKF–PF hybrid filter that was proposed for space object tracking (Dilshad Raihan & Chakravorty, 2015). The PGM filter employs an ensemble of possible state realizations for performing the uncertainty propagation. A functional form of the propagated pdf is recovered as a GMM by clustering the particles. The posterior pdf is obtained by performing a Kalman measurement update on the GMM. The PGM filter is conceived to keep track of the nonlinear uncertainty propagation, without performing any additional optimization or splitting operations during the propagation step. As the posterior pdf is obtained without employing the particle measurement update, the PGM filter is not prone to the particle depletion problem and the associated curse of dimensionality. Since the additional clustering step is performed only during the measurement update step, the PGM filter is especially suitable for filtering in the sparse measurement scenario.

The remainder of this article is organized as follows. An introductory discussion on mixture model filtering is given in Section 2. The PGM filter algorithm, and an associated convergence result, are presented in Section 3. Details pertaining to the actual implementation of the proposed filter are given in Section 3.2. The PGM filter is applied to three test cases and compared extensively with other nonlinear filters in Section 4.

## 2. Preliminaries: mixture model filtering

Let the state of the dynamical system of interest be denoted by $x \in \Re^d$. We assume that the state of the system evolves according to a Markov chain whose transition density is known and is specified by $p_n(x'|x)$. We also obtain measurements of the state at discrete times $n$ and the observation model is specified by the following: $z_n = h(x_n) + v_n$, where $h(x)$ is a nonlinear measurement function and $\{v_n\}$ is a discrete time Gaussian white noise process with zero mean and covariance $R_n$. Let $\pi_{n-1}(x)$ represent the conditional state pdf $p_{n-1}(x|Z_{n-1})$ where $Z_{n-1}$ is the sequence of all measurements recorded until time $n-1$. Additionally, let the prediction of the pdf before the measurement $z_n$ at time $n$ (the predicted prior pdf) $p_n(x|Z_{n-1})$ be represented by $\pi_n^-(x)$. Then, $\pi_n^-(x) = \int p_n(x|x')\pi_{n-1}(x')dx'$. Further, after measurement $z_n$ is received, the posterior pdf of the state is obtained according to the Bayes rule as follows: $\pi_n(x) = \frac{p_n(z_n|x)\pi_n^-(x)}{\int p_n(z_n|x')\pi_n^-(x')dx'}$, $p_n(z_n|x)$ is the measurement likelihood function and can be inferred from the measurement model above.

The prediction and the update steps above are the key steps to any recursive filtering algorithm. Let us assume that a mixture representation has been chosen for the predicted and posterior pdfs. In particular, let $\pi_n^-(x) = \sum_{i=1}^{M^-(n)} \omega_i^-(n) p_{i,n}^-(x)$, $\pi_n(x) = \sum_{i=1}^{M(n)} \omega_i(n) p_{i,n}(x)$, where $p_i^-(.), p_i(.)$ are pdfs, and $\{\omega_i^-(n)\}, \{\omega_i(n)\}$ are positive sets of weights that both add up to unity. The terms $M^-(n)$ and $M(n)$ represent the number of components used in the mixture representation. The prediction equation for the mixture model then boils down to the following: $\pi_n^-(x) = \sum_{i=1}^{M(n-1)} \underbrace{\omega_i(n-1)}_{\omega_i^-(n)} \underbrace{\int p_n(x|x') p_{i,n-1}(x')dx'}_{p_{i,n}^-(x)}$. Explicitly, the mixture prediction step can be split into the following two steps: $\omega_i^-(n) = \omega_i(n-1)$, $p_{i,n}^-(x) = \int p_n(x|x') p_{i,n-1}(x')dx'$. Given an observation $z_n$, the prior mixture $\pi_n^-(x)$ is transformed into the posterior mixture $\pi_n(x)$ as follows:

$\pi_n(x) = \frac{\sum_{i=1}^{M^-(n)} \omega_i^-(n) p_n(z_n|x) p_{i,n}^-(x)}{\sum_{i=1}^{M^-(n)} \omega_i^-(n) \int p_n(z_n|x') p_{i,n}^-(x')dx'}$. Define the likelihood that $z_n$ comes from the $i$th mixture component as follows: $l_i(n) \equiv \int p_n(z_n|x') p_{i,n}^-(x')dx'$. Rearranging the above mixture expression using the definition of the component/mode likelihood gives us $\pi_n(x) = \sum_{i=1}^{M^-(n)} \underbrace{\frac{w_i^-(n) l_i(n)}{\sum_j w_j^-(n) l_j(n)}}_{w_i(n)} \underbrace{\frac{p_n(z_n|x) p_{i,n}^-(x)}{l_i(n)}}_{p_{i,n}(x)}$. This expression shows that the measurement update has a hybrid nature, a standard update of the individual modes of the mixture with the measurement $z_n$, and a discrete Bayesian update of the mode weights using the mode likelihoods $l_i(n)$. Note that the mode likelihoods are the Bayes normalization factors for the individual modes. Explicitly, we delineate the discrete and continuous updates of the mixture model below:

$$\omega_i(n) = \frac{w_i^-(n) l_i(n)}{\sum_j w_j^-(n) l_j(n)}, \tag{1}$$

$$p_{i,n}(x) = \frac{p_n(z_n|x) p_{i,n}^-(x)}{l_i(n)}. \tag{2}$$

Let us now assume that we have fixed the form of the mixture model to a GMM, i.e., the posterior pdf at time $n-1$ can be represented by the GMM:

$$p_{i,n-1}(x) = \mathcal{G}(x; \mu_i(n-1), P_i(n-1)), \tag{3}$$

where $\mathcal{G}(x; \mu, P)$ represents the Gaussian pdf with mean $\mu$ and covariance $P$. Consider first the prediction equations. Note that the number of mixture components at time $n-1$, $M(n-1)$, is the same as the number of mixture components of the prediction at time $n$, $M^-(n)$. However, this assumes that the prediction of the $i^{\text{th}}$ Gaussian component $p_{i,n-1}$ of the posterior pdf at time $n-1$ remains a single Gaussian at time $n$, $p_{i,n}^-$. However, this is, in general, not true. The number of mixture components necessary to approximate the state pdf may vary from one time step to the other. For example, consider the nonlinear dynamical system given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\frac{x_1}{2} \\ \sin(\frac{x_2}{2}) \end{bmatrix} + \Gamma(t), \tag{4}$$

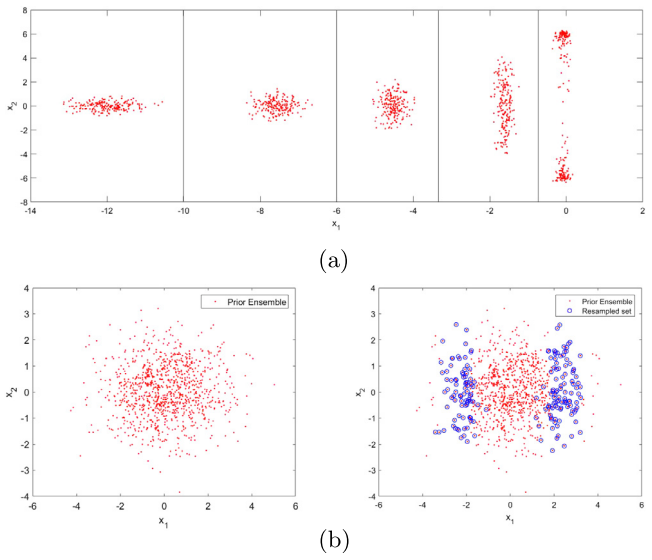



**Fig. 1.** Formation of multimodality: (a) through dynamics; (b) through measurement update.

$$\pi_0(X) = \mathcal{G}(X, \begin{bmatrix} -12 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.2 & 0 \\ 0 & 1 \end{bmatrix}), \tag{5}$$

where $\Gamma(t)$ is a white noise process. In Fig. 1(a), the locations of 200 particles sampled from the unimodal initial pdf $\pi_0(x_0)$ are seen to separate into two distinct modes as time progresses. Hence, in order to use mixture models for prediction, we have to find a way to deal with time varying number of GMM components.

Next, let us consider the measurement update (2). Since the prior component is Gaussian, and if the update (2) is approximated using the Kalman/linear minimum mean squared error (LMMSE) update (Dilshad Raihan & Chakravorty, 2015; Lefebvre, Bruyninckx, & De Schutter, 2002), we have

$$\mu_i(n) = \mu_i^-(n) + P_{i,zx}^T(n)P_{i,zz}^{-1}(n)(z_n - E_i[h(X)]), \tag{6}$$

$$P_i(n) = P_i^-(n) - P_{i,zx}^T(n)P_{i,zz}^{-1}(n)P_{i,zx}(n), \tag{7}$$

where

$$P_{i,zx}(n) = E_i\Big[\Big(h(X) - E_i\big(h(X)\big)\Big)\Big(X - E_i(X)\Big)^T\Big], \tag{8}$$

$$P_{i,zz}(n) = E_i\Big[\Big(h(X) - E_i\big(h(X)\big)\Big)\Big(h(X) - E_i\big(h(X)\big)\Big)^T\Big] \tag{9}$$

and $E_i[f(X)]$ represents an expectation of the function $f(X)$ with respect to the random variable $X$ where $X \sim \mathcal{G}(x; \mu_i^-(n), P_i^-(n))$. However, similar to the prediction case, in general, a single predicted Gaussian component can split into multiple modes after the update (2). An illustration of this is given in Fig. 1(b). In this case we have a prior ensemble generated from $\pi(x) = \mathcal{G}(X, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix})$. Then, a noisy measurement $z = 2$ is recorded where $z = x_1^2 + \tau$, $\tau \sim \mathcal{G}(x, 0, 2)$. An ensemble for the posterior pdf $\pi(x|z)$ is obtained through resampling and is seen to split into two separate modes. Hence, just as in the prediction step, there is a need to deal with the time varying number of GMM components after an update.

We shall not consider the measurement update aspect of the GMM filtering problem in this paper, which will be treated in a companion paper. Hence, we make the following assumption for the remainder of the paper.

**Assumption 1.** We shall assume a Gaussian mixture representation for the predicted and posterior filtered densities. Further, we assume that given a predicted mixture component at time $n$, $\mathcal{G}(x; \mu_i^-(n), P_i^-(n))$, the update (2) after an observation $z_n$ is approximated arbitrarily well by the Least Squares/Kalman update (6)–(7).

## 3. The particle Gaussian mixture (PGM) filter

In this section, we first present the PGM filter. The basic assumption underlying the PGM algorithm is that the predicted prior and posterior filter densities can be represented using a GMM. In particular, let $\pi_n^-(x) = \sum_{i=1}^{M^-(n)} \omega_i^-(n)\mathcal{G}_i^-(x; \mu_i^-(n), P_i^-(n))$, $\pi_n(x) = \sum_{i=1}^{M(n)} \omega_i(n)\mathcal{G}_i(x; \mu_i(n), P_i(n))$. In general, $M^-(n)$ and $M(n)$ need not be the same, however, owing to Assumption 1, they are assumed to be equal for the purposes of this paper. For instance, given a linear measurement function, this is true. The PGM filtering algorithm is composed of three basic steps that are described below.

**(1) Sampling**: The first step in the PGM algorithm is the use of the transition kernel to generate a set of samples at the next time step (which is the same as in a Particle filter). In practice, we first draw an ensemble $S_n$ of $N_p$ states $\{x_n^1, \ldots, x_n^i, \ldots, x_n^{N_p}\}$ from the GMM $\pi_n(x)$ and $N_p$ independent samples of the process noise term $w(n)$ from its density $P_W(w)$ to get $Sw_n = \{w_n^1, \ldots, w_n^i, \ldots, w_n^{N_p}\}$, and we let $x_{n+1}^{i-} = f(x_n^i) + w_n^i$, where $f(.)$ denotes the dynamics.

**(2) Clustering**: Then, we use a clustering algorithm $\mathcal{C}$ to partition the set of points into $M^-(n+1)$ different clusters whose means and covariances can be evaluated using sample averaging. Clustering is a field of Machine learning termed as Unsupervised Learning (Duda, Hart, & Stork, 2000; Jain, Murthy, & Flynn, 1999).

**Algorithm 1** PGM Algorithm

Given $\pi_0(x_0) = \sum_{i=1}^{M(0)} \omega_i(0)\mathcal{G}_i(x_0; \mu_i(0), P_i(0))$, transition density kernel $p_n(x|x')$, n = 1.

(1) Sample $N_p$ particles $X^{(i)}$ from $\pi_{n-1}$ and the transition kernel $p_n(x|x')$ as follows:
   (a) Sample $X^{(i)'}$ from $\pi_{n-1}(.)$.
   (b) Sample $X^{(i)}$ from $p(.|X^{(i)'})$.

(2) Use a clustering algorithm $\mathcal{C}$ to cluster the set of particles $\{X^{(i)}\}$ into $M^-(n)$ Gaussian clusters with weights, mean and covariance given by $\{w_i^-(n), \mu_i^-(n), P_i^-(n)\}$.

(3) Update the mixture weights and the mixture means and covariances to $\{\omega_i(n), \mu_i(n), P_i(n)\}$, given the observation $z_n$, utilizing the Kalman update (6), (7).

(4) n = n+1, go to Step 1.

In the experimental results presented in this paper, we use the simple k-means clustering algorithm (Lloyd, 1982), which is computationally very inexpensive while still being able to give good results for well separated clusters. The k-means clustering is a popular approach to partitioning wherein the dataset is grouped into different clusters so that the sum of squares of within-group distances is minimized, i.e, the dataset $S$ is partitioned into $M$ clusters $G_M^* = \{S_1, \ldots, S_M\}$ such that $G_L^* = \underset{G_L}{\text{argmin}} \sum_{i=1}^{M} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$. Here $G_M$ denotes any partition of the set S into $M$ clusters and $\mu_i$ represents the mean of the elements of the $i$th cluster in that partition. Once the vectors $x_i$ are assigned into different clusters, an $M$ mode GMM describing the set S may be derived as follows. $n_i = \sum_{j=1}^{N} \mathbb{1}(x_j \in S_i)$, $w_i = \frac{n_i}{N}$, $\mu_i = \frac{1}{n_i}\sum_{x_j \in S_i} x_j$, $C_i = \frac{\sum_{x_j \in S_i}(x_j - \mu_i)(x_j - \mu_i)^T}{n_i - 1}$. Here $\mathbb{1}(.)$ represents the indicator function.

**(3) Measurement update**: Incorporate the measurement information by updating the means and covariances of all $M$ modes individually using a Kalman measurement update/LMMSE update. Also update the mixture weights using the mode likelihoods $l_i(n+1)$ as in (1). In the present work we have considered two different approaches to computing the covariance terms $\left(P_{i,ZX}(n+1), P_{i,ZZ}(n+1)\right)$ and the expectations $\left(E_i\left(h(X)\right)\right)$ required for performing the Kalman update.

(a) Update 1 (PGM1-UT): In this approach, we compute the statistics of the posterior random variable with the unscented transform using a set of $2d+1$ sigma points that are distributed symmetrically. The covariance terms and the expectations required for computing the Kalman gain and posterior statistics are then computed as the weighted sample averages from the sigma points.

(b) Update 2 (PGM1): In this approach, the covariances and cross covariances required for computing the gain matrix are evaluated directly from the particles. Let $S^{-1}_{j,n+1} = \{x^{1-}_{j,n+1}, \ldots, x^{i-}_{j,n+1}, \ldots, x^{Nj-}_{j,n+1}\}$ denote the set of particles that form the $j$th cluster. Then the mean and covariance terms required for updating the cluster $j$ are assigned the corresponding sample averages computed from $S^{-1}_{j,n+1}$. The statistics of the measurement random variable are computed from the sample $Y^{-1}_{j,n+1} = \{h(x^{1-}_{j,n+1}), \ldots, h(x^{i-}_{j,n+1}), \ldots, h(x^{Nj-}_{j,n+1})\}$

Recursive implementation of the prediction, clustering and update steps as described here constitutes the PGM filter.

### 3.1. Analysis of the PGM filter

In the following, we analyze the PGM filter. We show that under the assumption of a perfect clustering scheme $\mathcal{C}$, the PGM filter density converges in probability to the true filter density.

Let $F_{z_n}(\pi_{n-1}) = \pi_n$ denote the true filter density at time $n$ given that the filter density at time $n-1$ is $\pi_{n-1}$ and the observation at time $n$ is $z_n$. Further, let $\hat{F}_{z_n}(\pi_{n-1})$ denote the filter density approximated by the PGM filter. We make the following exponential forgetting assumption on the true filter.

**Assumption 2.** We assume that there exists $C < \infty$ and $\rho < 1$ such that: $\|F_{z_n}(F_{z_{n-1}}(..(F_{z_1}(\pi_0))..)) - F_{z_n}(F_{z_{n-1}}(..(F_{z_1}(\pi'_0))..))\| \leq C\rho^n\|\pi_0 - \pi'_0\|$, for any measurement sequence $\{z_1, z_2, \ldots, z_n\}$, any $\pi_0, \pi'_0$, and where $\|.\|$ denotes the $L_1$ norm.

The conditions under which the exponential forgetting assumption holds is an active area of research (Douc, Moulines, & Ritov, 2007). However, the rate of forgetting is not always exponential. For example, parameter estimation problems do not satisfy the exponential forgetting assumption.

Let $\hat{F}_{z_n}(\hat{F}_{z_{n-1}}(\cdots(\hat{F}_{z_0}(\pi_0))\cdots))$ denote the filtered density approximated by the PGM filter given the measurement sequence $\{z_1, z_2, \ldots, z_n\}$ and the initial density $\pi_0$.

**Assumption 3.** Let $Prob(\|\hat{F}_{z_n}(\hat{\pi}_{n-1}) - F_{z_n}(\hat{\pi}_{n-1})\| > \epsilon) < \delta$, for all $n$. Further, we assume that $Prob(\|\hat{F}_{z_n}(\hat{\pi}_{n-1}) - F_{z_n}(\hat{\pi}_{n-1})\| > M) = 0$, for all $n$, for some $M < \infty$ (the error in a one step approximation of the filter density is almost surely uniformly bounded over all time).

The true posterior density at time $n$, $F_{z_n}(\hat{\pi}_{n-1})$, given the posterior from the previous time step, $\hat{\pi}_{n-1}$, and the current observation $z_n$, is deterministic. However, we cannot exactly reproduce the true posterior $F_{z_n}(\hat{\pi}_{n-1})$, and hence, we have the approximation $\hat{F}_{z_n}(\hat{\pi}_{n-1})$, which itself is the result of a random sampling experiment as done by the PGM algorithm. Therefore, we require that the approximation be a good enough representation of the true posterior with a high probability, where the probability now is over the randomness of the sampling that is required to generate the approximation.

**Lemma 1.** *Let $\|\hat{F}_{z_n}(\hat{\pi}_{n-1}) - F_{z_n}(\hat{\pi}_{n-1})\| \leq \epsilon$, for all $n$. Under Assumption 2, it follows that $\|\hat{\pi}_n - \pi_n\| \leq \frac{(C+1)\epsilon}{1-\rho}$.*

**Proof.** We have

$$\begin{aligned}
\hat{\pi}_n - \pi_n &= \hat{F}_{z_n}(\hat{F}_{z_{n-1}}(..(\hat{F}_{z_1}(\pi_0))..)) \\
&\quad - F_{z_n}(F_{z_{n-1}}(...(F_{z_1}(\pi_0))..)), \\
&= \underbrace{[\hat{F}_n(\hat{F}_{n-1}(..(\hat{F}_1(\pi_0))..)) - F_n(\hat{F}_{n-1}(..(\hat{F}_1(\pi_0))..))]}_{\Delta_n} \\
&\quad + \underbrace{[F_n(\hat{F}_{n-1}(..(\hat{F}_1(\pi_0))..)) - F_n(F_{n-1}(\hat{F}_{n-2}(..(\hat{F}_1(\pi_0))..)))]}_{\Delta_{n-1}} \\
&\quad + \cdots + \underbrace{[F_n(F_{n-1}(....(\hat{F}_1(\pi_0))...)) - F_n(..(F_1(\pi_0))..)]}_{\Delta_1}. \qquad (10)
\end{aligned}$$

Note that the different terms on the RHS above are $\Delta_n = \hat{F}_n(\hat{\pi}_{n-1}) - F_n(\hat{\pi}_{n-1})$, $\Delta_{n-1} = F_n(\hat{F}_{n-1}(\hat{\pi}_{n-2})) - F_n(F_{n-1}(\hat{\pi}_{n-2}))$, .... $\Delta_1 = F_n(..(F_2(\hat{F}_1(\pi_0)))..) - F_n(..(F_2(F_1(\pi_0)))..)$. Using Assumption 2 and the fact that $\|\hat{F}_{z_n}(\hat{\pi}_{n-1}) - F_{z_n}(\hat{\pi}_{n-1})\| \leq \epsilon$, for all $n$, it follows that $\|\Delta_i\| \leq C\rho^{n-i}\epsilon$, and thus, $\|\hat{\pi}_n - \pi_n\| \leq \sum_{i=1}^{n-1} C\rho^{n-i}\epsilon + \epsilon \leq \frac{(C+1)\epsilon}{1-\rho}$. The above result also holds for initial conditions in the infinite past, i.e., at $n = -\infty$. In the following, we assume that the initial condition was in the infinite past.

**Lemma 2.** *Let Assumptions 2 and 3 hold. Given any $\delta, \nu > 0$, there exists an $\bar{N} < \infty$, such that $Prob(\|\hat{\pi}_n - \pi_n\| > \frac{(1+\nu)(1+C)\epsilon}{1-\rho}) \leq \bar{N}\delta$, where $\bar{N} = n - n'$, and $n'$ is such that $\sum_{i=-\infty}^{n'} C\rho^{n-i} \leq q$, and $q = \frac{\nu(C+1)\epsilon}{M(1-\rho)}$.*

**Proof.** Let $e_n = \|\hat{\pi}_n - \pi_n\|$, and let $\epsilon_k = \|\hat{F}_{z_k}(\hat{\pi}_{k-1}) - F_{z_k}(\hat{\pi_{k-1}})\|$. It follows that $e_n \leq \sum_{k=-\infty}^{n} C\rho^{n-k}\epsilon_k$. Choose $n'$ such that $\sum_{i=-\infty}^{n'} C\rho^{n-i} \leq q$, where $q = \frac{\nu(C+1)\epsilon}{M(1-\rho)}$. Then,

$e_n = \underbrace{\sum_{k=n'}^{n} C\rho^{n-k}\epsilon_k}_{\bar{e}_n} + \underbrace{\sum_{k=-\infty}^{n'} C\rho^{n-k}\epsilon_k}_{\Delta^*_n}$. From Assumption 3, it follows that $Prob(\|\Delta^*_n\| > qM) = 0$, and thus, $Prob(\|\Delta_n\| > \frac{\nu(C+1)\epsilon}{1-\rho}) = 0$. Similarly, from Lemma 1, it follows that $Prob(\bar{e}_n > \frac{(C+1)\epsilon}{1-\rho}) \leq (n-n')\delta \equiv \bar{N}\delta$. From the above two inequalities, it follows that $Prob(e_n > \frac{(1+\nu)(C+1)\epsilon}{1-\rho}) \leq \bar{N}\delta$.

The two results above establish that if the sampling error at each step in the filter is small enough, and under the condition of exponential forgetting of initial conditions, the true filter density can be approximated arbitrarily closely with arbitrary high confidence. In the following, we establish that the sampling error at each step in the PGM filtering process can be arbitrarily small and thus, it follows from the two results above that the PGM filter can approximate the true filter density with arbitrarily high accuracy and arbitrarily high confidence. First, based on Assumption 1, we define the following:

$$P(\hat{\pi}_{n-1}) \equiv \hat{\pi}^-_n = \sum_{i=1}^{M^-(n)} \hat{\omega}^-_i(n)\mathcal{G}_i(x; \hat{\mu}^-_i(n), \hat{P}^-_i(n)),$$

$$\hat{P}(\hat{\pi}_{n-1}) \equiv \hat{\hat{\pi}}^-_n = \sum_{i=1}^{M^-(n)} \hat{\hat{\omega}}^-_i(n)\mathcal{G}(x; \hat{\hat{\mu}}^-_i(n), \hat{\hat{P}}^-_i(n)),$$

$$F_{z_n}(\hat{\pi}_{n-1}) = \sum_{i=1}^{M(n)} \hat{\omega}_i(n)\mathcal{G}(x;\hat{\mu}_i(n),\hat{P}_i(n)),$$

$$\hat{F}_{z_n}(\hat{\pi}_{n-1}) = \sum_{i=1}^{M(n)} \hat{\hat{\omega}}_i(n)\mathcal{G}(x;\hat{\hat{\mu}}_i(n),\hat{\hat{P}}_i(n)).$$

The above results represent the true and the approximate PGM predicted and filtered densities at time $n$ given the approximate density $\hat{\pi}_{n-1}$ at time $n-1$. We have the following result:

**Lemma 3.** *Given the GMM representation of the prior pdf above, and a perfect Clustering algorithm $\mathcal{C}$, given any $\epsilon' > 0$, and $\delta' > 0$, there exists an $N_{\epsilon',\delta'}(n) < \infty$ such that: if the number of samples used to approximate the predicted pdf at time n is greater than $N_{\epsilon',\delta'}(n)$ then,*

$$Prob(|\hat{\hat{\omega}}_i^-(n) - \hat{\omega}_i^-(n)| > \epsilon') < \delta', \tag{11}$$

$$Prob(\|\hat{\hat{\mu}}_i^-(n) - \hat{\mu}_i^-(n)\|_2 > \epsilon') < \delta', \tag{12}$$

$$Prob(\|\hat{\hat{P}}_i^-(n) - \hat{P}_i^-(n)\| > \epsilon') < \delta', \tag{13}$$

*for all i.*

**Proof.** In the following, we suppress the explicit dependence on time $n$ to ease the notation. Under the assumption of a perfect clustering scheme, and given that the number of samples in each cluster is large enough, the mixture weight estimates $\hat{\hat{\omega}}_i^-(n)$ are normal distributed, and one can use the central limit theorem (CLT) (Kallenberg, 1997) to always find an $N_{\epsilon',\delta'}(\omega_j)$ such that (11) is satisfied.

For a Gaussian random vector $X \in \mathbf{R}^n$ with independent components $\{x_1,\ldots,x_n\}$ and $E\{x_i^2\} = 1$, it can be shown that $Prob\{\left|\|X\|_2 - \sqrt{d}\right| \geq t\} \leq 2e^{\frac{-ct^2}{k^4}}, \forall t \geq 0$. Here $c$ is a positive constant and $k = \max_i \|x_i\|_{\psi_2}$ where $\|x_i\|_{\psi_2} = \inf\{t; E\{e^{(\frac{x_i}{t})^2}\} \leq 2\}$ (Vershynin, 2017). If $\hat{\hat{\mu}}_i^-(n)$ is determined as the sample mean of points assigned to cluster $C_i$, then we have

$$Prob\left\{\left|\left\|\left(\frac{\hat{P}_i^-(n)}{N_j}\right)^{\frac{-1}{2}}\left(\hat{\hat{\mu}}_i^-(n) - \hat{\mu}_i^-(n)\right)\right\|_2 - \sqrt{d}\right| \geq t\right\} \leq 2e^{\frac{-ct^2}{k^4}}, \tag{14}$$

where $N_j$ is the number of points assigned to cluster $j$. This can be manipulated to show that

$$Prob\left\{\left\|\left(\hat{\hat{\mu}}_i^-(n) - \hat{\mu}_i^-(n)\right)\right\|_2 \geq \frac{t+\sqrt{d}}{\left\|\left(\hat{P}_i^-(n)\right)^{\frac{-1}{2}}\right\|_2\sqrt{N_j}}\right\} \leq 2e^{\frac{-ct^2}{k^4}}. \tag{15}$$

Using the above inequality and by setting $\frac{t+\sqrt{d}}{\left\|\left(\hat{P}_i^-(n)\right)^{\frac{-1}{2}}\right\|_2\sqrt{N_j}} < \epsilon'$, $2e^{\frac{-ct^2}{k^4}} < \delta'$, it is possible to choose an $N_j$ so that (12) is satisfied. The minimum value of $N_j$ that satisfies the above set of equations is chosen as $N_{\epsilon',\delta'}(\mu_j)$.

Let $\hat{\hat{P}}_i^-(n)$ be the sample average estimate of $i$th modal covariance. Then it can be shown that $Prob\left\{\frac{\left\|\hat{\hat{P}}_i^-(n) - \hat{P}_i^-(n)\right\|}{\left\|\hat{P}_i^-(n)\right\|} \geq ck^2\left(\sqrt{\frac{d+u}{N_j}} + \frac{d+u}{N_j}\right)\right\} \leq 2e^{-u}$, where $k$ is a constant greater than or equal to one and $c \geq 0$ (Vershynin, 2017). Here $\|\|$ represents the spectral norm of the covariance matrix. Hence, by picking an $N_j$ such that $ck^2\left(\sqrt{\frac{d+u}{N_j}} + \frac{d+u}{N_j}\right)\left\|(\hat{P}_i^-(n))\right\| < \epsilon'$, $2e^{-u} < \delta'$, the condition given in (13) can be satisfied. Let $N_{\epsilon',\delta'}(P_j)$ be the minimum $N_j$ that satisfies this condition. Pick $N_{\epsilon',\delta'}(j) = \max\left(N_{\epsilon',\delta'}(\omega_j), \frac{N_{\epsilon',\delta'}(\mu_j)}{\hat{\omega}_j^-}, \frac{N_{\epsilon',\delta'}(P_j)}{\hat{\omega}_j^-}\right)$. Such $N_{\epsilon',\delta'}(j)$ can be found for all clusters $C_j$ and given that we choose $N_{\epsilon',\delta'}$ as follows:

$$N_{\epsilon',\delta'} = \max_j N_{\epsilon',\delta'}(j), \tag{16}$$

it is guaranteed that all the elements of the mean vector $\hat{\mu}^-(n)$ and the covariance matrix $\hat{P}^-(n)$ can be estimated to an accuracy of $\epsilon'$ with confidence of at least $1-\delta'$, which completes the proof of the result. Note that the above $N_{\epsilon',\delta'}$ is a function of time $N_{\epsilon',\delta'}(n)$ owing to the time dependence of the GMM that is estimated.

It may be shown that under Assumption 1, the error incurred in estimating the posterior mean and covariance $\hat{\mu}_i(n)$, $\hat{P}_i(n)$ is at most $K(n)\epsilon'$, for some time varying $K(n) < \infty$ which depends on the posterior mean and covariance, given that the predicted prior means and covariances of the clusters of the GMM have been approximated to an accuracy of $\epsilon'$. This can be summarized in the following result:

**Lemma 4.** *Given any $\epsilon'$, $\delta' > 0$, choose $N_{\epsilon',\delta'}(n)$ according (16). If the number of samples used in the PGM filter to approximate the predicted prior pdf at time n is greater than $N_{\epsilon',\delta'}(n)$ then, there exists $k(n) < \infty$ s.t:*

$$Prob(|\hat{\hat{\omega}}_i(n) - \hat{\omega}_i(n)| > K(n)\epsilon') < \delta',$$

$$Prob(\left\|\hat{\hat{\mu}}_i(n) - \hat{\mu}_i(n)\right\| > K(n)\epsilon') < \delta',$$

$$Prob(\left\|\hat{\hat{P}}_i(n) - \hat{P}_i(n)\right\| > K(n)\epsilon') < \delta',$$

*for all i.*

Next, we find a bound on the $L_1$ error between the estimated and true filtered densities given the error between the parameters of the GMM representing the true and the approximate filtered densities.

**Lemma 5.** *Let $|\hat{\hat{\omega}}_i(n) - \hat{\omega}_i(n)| < \epsilon'$, $\left\|\hat{\hat{\mu}}_i(n) - \hat{\mu}_i(n)\right\| < \epsilon'$, and $\left\|\hat{\hat{P}}_i(n) - \hat{P}_i(n)\right\| < \epsilon$ for all i. Then, given that the state of the system $x \in \Re^d$, there exists $C(n) < \infty$ such that $\|\hat{\hat{\pi}}_n - \hat{\pi}_n\| < C(n)d\epsilon'$.*

**Proof.** We show the result for the case of a simple one component Gaussian with an error in the covariance, it can be generalized to the GMM in a relatively straightforward fashion but at the expense of a very tedious derivation which we forego here for clarity. We also suppress the explicit dependence on time $n$ in the following for notational convenience.

$$\begin{aligned}\hat{\hat{\pi}}(x) - \hat{\pi}(x) &= \frac{1}{(2\pi)^{d/2}|\hat{\hat{P}}|^{1/2}}e^{-\frac{1}{2}(x-\mu)^T\hat{\hat{P}}^{-1}(x-\mu)} - \\ &\quad \frac{1}{(2\pi)^{d/2}|\hat{P}|^{1/2}}e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)}, \\ &\approx \frac{1}{(2\pi)^{d/2}|\hat{P}|^{1/2}}e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)} \\ &\quad \times \frac{1}{2}(x-\mu)^T(\hat{P}^{-1}\Delta\hat{P}^{-1})(x-\mu),\end{aligned} \tag{17}$$

where $\hat{\hat{P}} = \hat{P} + \Delta$ and since $e^{-\frac{1}{2}(x-\mu)^T(\hat{P}+\Delta)^{-1}(x-\mu)} \approx e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)}e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}\Delta\hat{P}^{-1}(x-\mu)}$, which in turn implies (17). This in turn implies that $\|\hat{\hat{\pi}} - \hat{\pi}\| \approx \frac{1}{(2\pi)^{d/2}|\hat{P}|^{1/2}} \times \int e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)}\frac{1}{2}(x-\mu)^T\hat{P}^{-1}\Delta\hat{P}^{-1}(x-\mu)dx \leq \frac{C(\hat{P})\epsilon'}{(2\pi)^{d/2}|\hat{P}|^{1/2}} \times \int e^{-\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)}\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)dx$, since there exists $C(\hat{P}) < \infty$ such that $\frac{1}{2}(x-\mu)^T\hat{P}^{-1}\Delta\hat{P}^{-1}(x-\mu) \leq C(\hat{P})\epsilon'\frac{1}{2}(x-\mu)^T\hat{P}^{-1}(x-\mu)$, owing to the fact that $\|\Delta\| < \epsilon'$.

Now, let $Y = \hat{P}^{-1/2}(X - \mu)$. Then, it follows that $\|\hat{\hat{\pi}} - \hat{\pi}\| \leq C(\hat{P})\epsilon' \frac{1}{(2\pi)^{d/2}} \int e^{-1/2y^T y} y^T y dy = C(\hat{P})\epsilon' d$. The last step in the above equation follows from noting that $Y'Y$ is a chi-squared random variable of degree of freedom $d$ and thus, its expected value is $d$. This establishes our result. In general for a GMM, the constant $C(n)$ would depend on the means and covariances of all the GMM components and their weights.

Lemmas 4 and 5 immediately lead us to the following corollary.

**Corollary 1.** *Let $\epsilon'(n)$ be the desired accuracy in estimating the parameters of the GMM representing $F_{z_n}(\hat{\pi}_{n-1})$, i.e., the true filtered density given observation $z_n$ and the PGM posterior pdf at the previous time $\hat{\pi}_{n-1}$. Let $\delta'(n)$ be the desired confidence of the estimate. IF $\epsilon'(n)$ and $\delta'(n)$ are chosen such that*

$$C(n)K(n)\epsilon'(n)d = \epsilon, \tag{18}$$

$$\delta'(n) = \frac{\delta}{N}, \tag{19}$$

*and the corresponding number of samples $N_{\epsilon'(n),\delta'(n)}(n)$ be chosen according to (16), then it follows that $\|Prob\|\hat{F}_{z_n}(\hat{\pi}_{n-1}) - F_{z_n}(\hat{\pi}_{n-1})\| > \epsilon) \leq \frac{\delta}{N}$.*

**Proof.** Recall that $\hat{\pi}_n = F_{z_n}(\hat{\pi}_{n-1})$, and $\hat{\hat{\pi}}_n = \hat{F}_{z_n}(\hat{\pi}_{n-1})$. Then, from Lemma 5 we have that $\|\hat{\pi}_n - \hat{\hat{\pi}}_n\| \leq C(n)K(n)d\epsilon'(n)$ if $|\hat{\theta}_i(n) - \hat{\hat{\theta}}_i(n)| < \epsilon'(n)$ for all $i$, where $\hat{\theta}_i(n)$ represents the true parameters underlying the GMM representation of $\hat{\pi}_n$ and $\hat{\hat{\theta}}_i(n)$ represents their PGM approximation. Hence: $Prob(\|\hat{\pi}_n - \hat{\hat{\pi}}_n\| > C(n)K(n)d\epsilon'(n)) < \delta'(n)$, which owing to the definition of $\epsilon'(n)$ and $\delta'(n)$ leads us to the desired result.

Hence, using Corollary 1 and Lemma 2, it follows that if the number of samples used to approximate the parameters of the predicted GMM pdf at time $n$ is greater than the $N_{\epsilon'(n),\delta'(n)}$, then it follows that $Prob(\|\hat{\hat{\pi}}_n - \hat{\pi}_n\| > \frac{(1+\nu)C\epsilon}{1-\rho}) \leq \delta$, for all $n$ for any arbitrarily small $\epsilon, \delta, \nu > 0$. However, in order for Assumption 3 to be valid, the sample averages $\hat{\hat{\theta}}_n$ have to be almost surely bounded. Here $\hat{\hat{\theta}}_n$ represents any parameter that is computed from the sample and used to specify the pdf such as the component weights, means or covariances. To show this, due to the Strong Law of Large Numbers, it is also true that $\hat{\hat{\theta}}_n^N \to \hat{\theta}_n$ as $N \to \infty$, where $\hat{\hat{\theta}}_n^N$ represents the estimate of the parameters after $N$ samples. Given the sample size is large enough, the estimate $\hat{\hat{\theta}}_n^N$ is arbitrarily close to the true parameters $\hat{\theta}_n$ almost surely, and thus, since the true parameters are bounded, so are the estimates. This may be summarized in the following result.

**Proposition 1.** *Let Assumptions 1 and 2 hold. Given a perfect clustering algorithm $\mathcal{C}$, and any $\epsilon, \delta, \nu > 0$, at every time step $n$, choose the required accuracy of the approximation $\epsilon'(n)$ from (18), the required confidence $\delta'(n)$ from (19), and the corresponding minimum number of samples $N_{\epsilon'(n),\delta'(n)}$ from (16), then, $Prob(\|\hat{\hat{\pi}}_n - \hat{\pi}_n\| > \frac{(1+\nu)C\epsilon}{1-\rho}) \leq \delta$.*

**Remark 1** (*The Curse of Dimensionality*)**.** The number of samples required to estimate the mixture weights does not depend on the dimension of the state space. Additionally, it can be seen that the number of samples required to estimate the component means and covariances increase only as $O(d)$. From the above analysis, (16), and Lemma 4 it can be concluded that the number of samples required to estimate the parameters of the predicted and posterior pdfs accurately increases only linearly with the dimension of the state space, and thus, is free from the "Curse of Dimensionality". However, we have to be more careful regarding the functional $L_1$ error in the PGM density: (18) shows that the accuracy parameter required at every time step is inversely proportional to the dimension of the state space since $\epsilon'(n) = \frac{\epsilon}{C(n)K(n)d}$, and thus, in order to attain the same accuracy in terms of the functional error of the filtered density, the number of samples have to increase as $O(d^2)$ where $d$ is the dimension of the problem. Further, it should also be noted that the computation of the sample averages required by the PGM filter grows as $O(d^2)$.

### 3.2. Implementation

In this section, we discuss certain steps involved in the practical implementation of the PGM filter in detail.

**Modified k-means clustering**: The k-means algorithm requires the total number of clusters to be specified externally. To work around this limitation, we have implemented a strategy which only requires the upper bound $M^-_{max}(n+1)$ as the external input instead of $M^-(n+1)$. We define the likelihood agreement measure ($L_{mes}$) (DeMars et al., 2013) as the measure of fitness of the parametric model $\theta_a$ in describing the dataset $S$. Let $\theta_{a,M}$ be an M-component mixture model indexed by $a$ and arrived at from $k-means$ clustering. Then $L_{mes}(\theta_{a,M})$ may be computed as $L_{mes}(\theta_{a,M}) = \sum_{i=1}^{N_p} \pi_{\theta^a}(x^{i-}_{n+1})$, where $\pi_{\theta^a}(x)$ is the mixture pdf derived from the parametric model $\theta_{a,M}$. Let $\theta_{a^*,M^*}$ be the optimal parametric model with $M^-_{n+1} = M^*$ components that maximizes the $L_{mes}$ given the bound $M^-_{max}(n+1)$. Then, the proposed strategy for clustering is presented in the following algorithm.

**Algorithm 2** Clustering Strategy

**Input:** $S^{-1}_{n+1} = \{x^{1-}_{n+1}, \cdots, x^{i-}_{n+1}, \cdots, x^{N_p-}_{n+1}\}$, $M^-_{max}(n+1)$
**Output:** $\theta_{a^*,M*}$, $M* \leq M^-_{max}(n+1)$

```
M ← M^-_max(n+1)
θ_{a*,M*} ← θ_{a,M^-_max(n+1)}
L*_mes ← L_mes(θ_{a,M^-_max})
while M > 1 do
    M ← M − 1
    Compute θ_{a,M} using k-means
    if L_mes(θ_{a,M}) ≥ L*_mes then
        θ_{a*,M*} ← θ_{a,M}
        L*_mes ← L_mes(θ_{a,M})
    end if
end while
```

**Merging**: Depending on the clustering scheme, dynamics and measurement models, one may observe several closely distributed mixture modes in the posterior pdf. To identify the right modes to be merged, we define the following normalized error metric (Hanebeck & Briechle, 2003) as a measure of similarity between modes $i$ and $j$. $D(i,j) = \frac{\int (\mathcal{G}_i(x,\mu_i,P_i) - \mathcal{G}_j(x,\mu_j,P_j))^2 dx}{\int \mathcal{G}_i(x,\mu_i,\Sigma_i)^2 dx + \int \mathcal{G}_j(x,\mu_j,\Sigma_j)^2 dx}$. Clearly, $D(i,j) = 0$ when the components $i, j$ are identical. It also has an upper bound at 1. Mixture modes that are closely spaced can be merged whenever the value of normalized error metric falls below a predetermined tolerance ($tol$).

## 4. Numerical examples

In this section, the particle Gaussian mixture filter is applied to three test case problems to evaluate the filtering performance. Due to their ease of implementation and extensive use as benchmark filters, the UKF and PF are included for comparison. The PGM filter is also compared with Gaussian mixture filters such as a GMUKF

and the blob filter. The estimation results are assessed for accuracy, consistency and informativeness as described below.

**(1) Accuracy**: A Monte Carlo averaged root mean squared error ($E_{rms}(t)$) is considered for evaluating the accuracy of the estimates. The value of $E_{rms}(t)$ is computed over a set of $N_{Mo}$ Monte Carlo runs as $E_{rms}(t) = \sqrt{\frac{1}{N_{Mo}}\sum_{j=1}^{N_{Mo}} \|\hat{X}^j(t) - \hat{\mu}^j(t)\|_2^2}$.

Here, $\hat{X}^j(t)$ and $\hat{\mu}^j(t)$ represent the actual and estimated states at the time instant $t$ during the $j$th Monte Carlo run. The time averaged error ($\overline{E_{rms}}$) can be computed from $E_{rms}(t)$ as $\overline{E_{rms}} = \frac{1}{T}\sum_{t=1}^{T} E_{rms}(t)$.

**(2) Consistency**: The consistency of the filtered pdf is examined using the normalized estimation error squared (NEES) test. For a unimodal state pdf, the NEES test is evaluated using the $\chi^2$ test statistic ($\beta_{j,t}$) given by $\beta_{j,t} = (\hat{X}^j(t) - \hat{\mu}^j(t))^T(\hat{P}^j(t))^{-1}(\hat{X}^j(t) - \hat{\mu}^j(t))$. The term $\hat{P}^j(t)$ in the above expression represents the covariance of the unimodal filtered pdf at time $t$ during $j$th Monte Carlo run. The Monte Carlo averaged NEES test ($\beta_t$) is computed from this expression as $\beta_t = \frac{1}{N_{Mo}}\sum_{j=1}^{N_{Mo}} \beta_{j,t}$. When the state vector $x \in \Re^d$ is normally distributed, the product $N_{Mo}\beta_t$ has a $\chi^2$ distribution with $dN_{Mo}$ degrees of freedom. Hence, the consistency of the filtered pdf can be tested by determining whether $\beta_t$ falls within probable bounds determined from the corresponding $\chi^2$ random variable.

**(3) Informativeness**: Two separate metrics are considered for evaluating the informativeness of estimates in the present work namely the averaged likelihood of the truth ($L(t)$) and volume of $2-\sigma$ uncertainty region ($V\sigma_2$). The averaged likelihood of the truth over $N_{Mo}$ Monte Carlo runs may be computed as $L(t) = \frac{1}{N_{Mo}}\sum_{j=1}^{N_{Mo}} \pi_t^j(\hat{X}^j(t))$. Here $\pi_t^j$ represents the conditional state pdf at time $t$ in the $j$th Monte Carlo run. The time averaged likelihood is computed from the above expression as $\hat{L} = \frac{1}{T}\sum_{t=1}^{T} L(t)$. When the state pdf is in the ensemble form, the likelihood is computed using a unimodal Gaussian pdf characterized by the sample mean and covariance of the collection of states. For a well separated GMM pdf, the value of $V\sigma_2$ can be computed as $V^j\sigma_2(t) = \sum_{i=1}^{M_t} |2\Sigma_i|$, where $M_t$ is the number of modes. Here $|.|$ represents the determinant of the enclosed square matrix. We compute the Monte Carlo averaged $2-\sigma$ volume as $V\sigma_2(t) = \sum_{j=1}^{N_{Mo}} V^j\sigma_2(t)$. We compute the corresponding time averaged value $\hat{V}\sigma_2$ as $\hat{V}\sigma_2 = \frac{1}{T}\sum_{i=1}^{T} V\sigma_2(t)$.

### 4.1. Example 1

In this problem, we consider the estimation of the one dimensional discrete time nonlinear dynamic system given by

$$x_{k+1} = \frac{x_k}{2} + \frac{25x_k}{(1+x_k^2)} + 8\cos(1.2k) + \nu_k \tag{20}$$

A measurement model aiding the estimation of the system is specified as $z_k = \frac{x_k^2}{20} + n_k$.

The process noise term $\nu_k$ and measurement noise term $n_k$ are assumed to be independent zero mean Gaussian random variables with covariances $Q = 10$ and $R = 1$, respectively. This example or its variants have been studied in several publications before (Arulampalam et al., 2001; Gordon et al., 1993). Two variants of the PGM filter, i.e, PGM1-UT and PGM1, an SIR filter, blob filter and a UKF are simulated to estimate the test case 1 system for a duration of 52 time steps over 50 Monte Carlo runs. The initial state of the system is assumed to be distributed as $P_0(x) = \mathcal{N}(0, 2)$. Measurements are recorded at every other instant. The SIR and the PGM filters are implemented with a set of 50 particles. The upper bound on the number of mixture components $M_{max}$ is set to be 3. For blob filter, 50 Gaussian components were used with a covariance upper bound $P_{max} = 10^{-6}$. The parameter values used in the implementation of the UKF may be found in Table 1.

**Table 1**
UKF parameters.

| $\alpha$ | $\beta$ | $\lambda$ |
|---|---|---|
| 1.3 | 1.5 | 0.2 |

**Table 2**
Example 1: Results.

| | $\overline{E_{rms}}$ | $\beta_{t,c}(\%)$ | $\hat{L}$ | $\hat{V}\sigma_2$ |
|---|---|---|---|---|
| PGM1-UT | 6.4513 | 78.85 | 0.1209 | 62.1753 |
| PGM1 | 6.2859 | 84.62 | 0.1253 | 60.3453 |
| PF | 6.5488 | 46.15 | 0.1063 | 79.8123 |
| UKF | 8.3279 | 36.54 | 0.0488 | 103.8405 |
| Blob Filter | 6.5243 | 46.15 | 0.1827 | 0.0001 |

The values of $E_{rms}(t)$ plotted in Fig. 2(a) indicate good tracking performance by the PGM filters. The PGM filters, the PF and the blob filter are seen to offer comparable tracking performance. The Monte Carlo averaged NEES results plotted in Fig. 2(b) show that the UKF and PF frequently oversteps the upper bound which marks inconsistent estimates. Furthermore, $\beta_t$ computed using the PF estimates are found to frequently exhibit peaks several orders of magnitude larger than the 99% upper bound, indicating covariance collapse. The averaged likelihood $L(t)$ and the volume $V_{2\sigma}$ plotted in Fig. 2(c) and (d) show that only the blob Filter provides more informative estimates than the PGM-1 and PGM-1 UT.

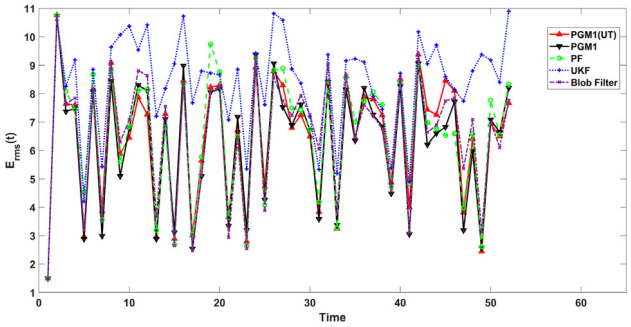
The time averaged values of RMSE $\overline{E_{rms}}$, likelihood $\hat{L}$, and the $2\sigma$ volume for each filter are listed in Table 2. Also included is the fraction ($\beta_{c\%}$) of the time instants during which the computed averaged NEES result stayed within the 99% limits, i.e., $\beta_{c\%} = \frac{\sum_{t=1}^{T} \mathbb{1}_{Ub0.99}(\beta_t)}{T}$ where $\mathbb{1}_{Ub0.99}(\beta_t)$ is the indicator function which equals 1 when $\beta_t < Ub_{0.99}$ and zero otherwise. The results presented in Table 2 clearly show that the PGM filter implementations offer accurate, consistent and informative estimates.

### 4.2. Example 2

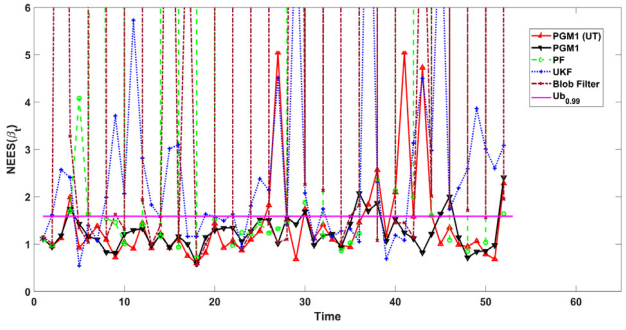
In this example, the PGM filters are employed in the estimation of a 3 dimensional Lorenz 63 model for atmospheric convection. The noise perturbed dynamics of the Lorenz 63 system is described in the following set of equations:

$$\begin{aligned} \dot{x_1} &= \alpha(-x_1 + x_2), \quad \alpha = 10 \\ \dot{x_2} &= \beta x_1 - x_2 - x_1x_3, \quad \beta = 28 \\ \dot{x_3} &= -\gamma x_3 + x_1x_2 + \Gamma(t), \quad \gamma = 8/3 \end{aligned} \tag{21}$$

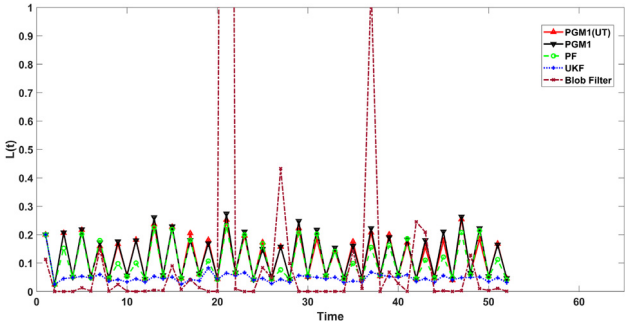
A scalar nonlinear measurement model ($z_k$) is considered which is given by $z_k = \sqrt{x_1(t)^2 + x_2(t)^2 + x_3(t)^2} + \nu_k$. Note that since the measurement function is nonlinear, the PGM filter will not be asymptotically exact. However it is still of interest to study the performance of the filter in the nonlinear case. The process and measurement noise covariances are both set to be equal to 1. The initial state of the system is characterized by the bimodal pdf $p_0(x) = 0.9\mathcal{G}(x, [-0.2, -0.2, 8]^T, \sqrt{0.35}I_{3\times3}) + 0.1\mathcal{G}(x, [0.2, 0.2, 8]^T, \sqrt{0.35}I_{3\times3})$. The state of the system is updated at a time step $\Delta t = 0.01$ s. The measurements are recorded at the interval of ten time steps. This example has been considered previously in Terejanu et al. (2011). The PGM1 filter, PGM1-UT filter, the PF, the blob filter and a conventional Gaussian mixture UKF (Alspach & Sorenson, 1972) are employed in the estimation of the Lorenz63 system. The PGM filters and the SIR filter are implemented with 300 particles and $M_{max}$ is set to be 3. The UKF is implemented using the parameters listed in Table 1. The blob filter is implemented by re-approximating the initial pdf using 300 Gaussians with a maximum covariance $P_{max} = 0.0005$



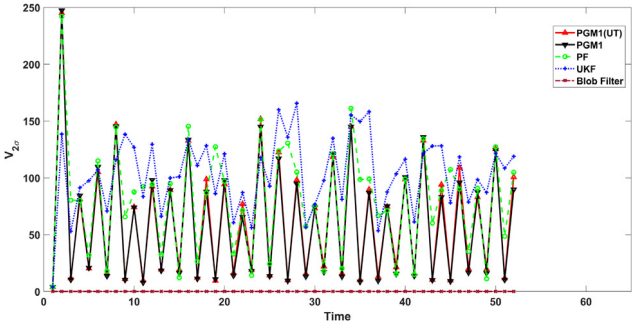

(a) $E_{rms}(t)$




(b) $\beta_t$

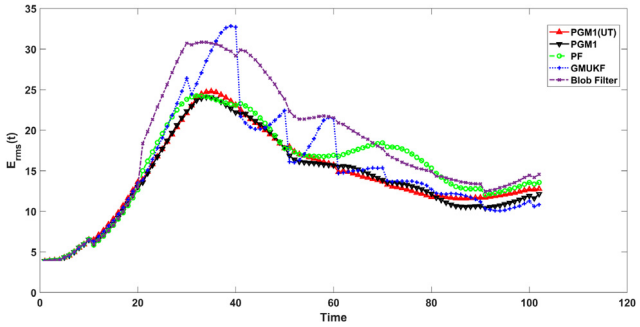



(c) $L(t)$

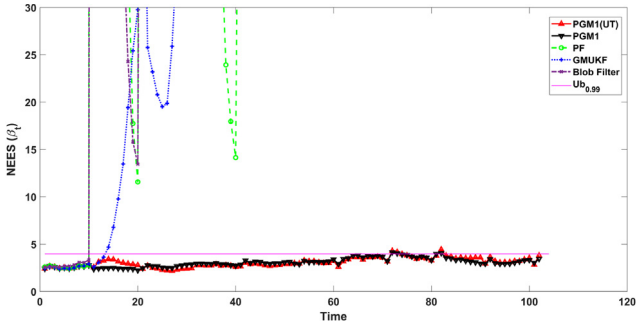



(d) $V_{2\sigma}$

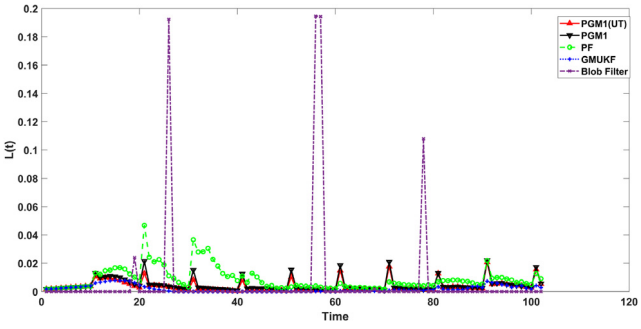
**Fig. 2.** Example 1: Results.



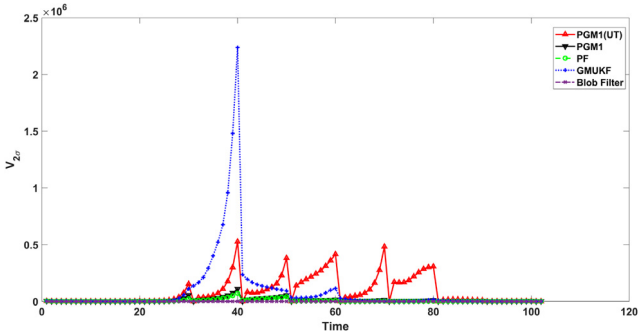

(a) $E_{rms}(t)$




(b) NEES($\beta_t$)




(c) $L(t)$




(d) $V_{2\sigma}$

**Fig. 3.** Example 2: Results.

$\times I_3$. The values of $E_{rms}(t)$ computed over 50 runs and plotted in Fig. 3(a) show that estimation errors for PGM1 and PGM1-UT are the smallest among the four filters. The Monte Carlo averaged NEES results are plotted in Fig. 3(b). It is observed that the NEES test statistic $\beta_t$ for the PF, the blob filter and the mixture UKF overstep the $y = Ub_{0.99}$ line early in the simulation.

The averaged likelihoods ($L(t)$) and $V\sigma_2$ volumes plotted in Fig. 3(c), (d) show that the PF has the highest average likelihood whereas the Blob filter has the smallest $V\sigma_2$ volume. However, as the NEES results of the PF and the blob filter are seen to stay above the $10^2$ for around 85% of the time. The higher likelihoods and the small $V\sigma_2$ of the PF should be understood as a consequence of its covariance collapse. The consistency fractions ($\beta_{c\%}$) and the time averaged values of other performance metrics for each filter are listed in Table 3. The results clearly indicate that the PGM filters

**Table 3**
Example 2: Results.

| | $\overline{E_{rms}}$ | $\beta_{t,c}(\%)$ | $\hat{L}$ | $\hat{V}\sigma_2(\times 10^4)$ |
|---|---|---|---|---|
| PGM1-UT | 14.3886 | 97.06 | 0.0038 | 8.644 |
| PGM1 | 14.1148 | 97.06 | 0.0045 | 1.046 |
| PF | 15.5425 | 11.76 | 0.0088 | 0.6737 |
| GMUKF | 15.3528 | 13.73 | 0.0019 | 9.3291 |
| Blob Filter | 17.9477 | 9.80 | 0.0070 | $1.09 \times 10^{-11}$ |

are more accurate and consistent than the PF, mixture UKF and blob filter.

### 4.3. Example 3

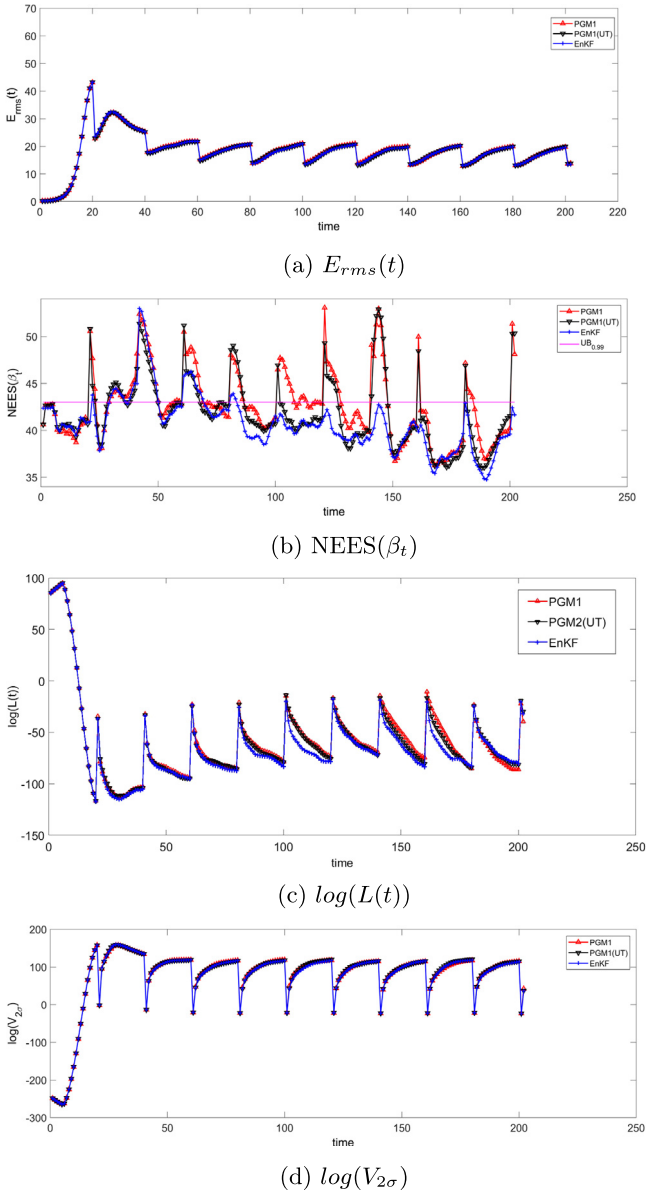
In this test case, the PGM filters are employed in the estimation of a Lorenz96 system. The noise perturbed dynamics of the Lorenz96 system is given by

$$\dot{x}_i = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \Gamma(t), \tag{22}$$

where $i = 1, 2, \ldots, 40$ (Lorenz, 1995). The state variables are assumed to be cyclical so that $x_0 = x_{40}, x_{-1} = x_{39}, x_{41} = x_1$. The term $F$ represents a constant external forcing. In the present work, we set $F = 8$ at which the system is chaotic. The covariance of the zero mean Gaussian white noise is assumed to be $Q = 10^{-2}$. A linear measurement model is employed in the estimation of the Lorenz96 system and it is defined as

$$z_k = HX_k + v_k, \quad H_{i,j} = \begin{cases} 1, & j = 2i - 1 \\ 0, & \text{otherwise.} \end{cases}$$

where $H \in \mathbb{R}_{20\times 40}$. The measurement noise is assumed to be a zero mean Gaussian random vector with a covariance $R = 10^{-2} I_{20\times 20}$ where $I_{i,j} = \delta_{i,j}$. The initial state pdf is given by $p_0(x) = \mathcal{G}(x, \mu_0, P_0)$, where $\mu_0 = F\left[1 \cdots 1 \cdots 1\right]^T$, $\mu_0 \in \mathbb{R}_{40\times 1}$ and $P_0 = 10^{-3} I_{40\times 40}$. The state of the system is updated at $\Delta t = 0.05$ time units and measurements are recorded at the interval of 1 time unit. The performance of the PGM-1 filters is compared to that of an EnKF (Burgers, Van Leeuwen, & Evensen, 1998), an SIR filter and a blob filter. The PGM-1 filters and the EnKF were equipped with a set of 2000 particles. The value of $M_{max}$ is kept at 2 in order to keep the computational cost low. The SIR filter was implemented with 2000 particles. The blob filter was employed with 2000 Gaussians each having a maximum covariance $P_{max} = 10^{-4} \times I_{40}$. The filters were used to estimate the state of the system for a duration of 200 time steps over 50 Monte Carlo runs. The number of time steps is kept low in order to keep the computational cost of Monte Carlo simulations manageable. The PF and the blob filter were found to undergo covariance collapse after the first measurement was recorded. This was seen to be the case even after using 4000 particles for the PF and using 4000 Gaussians with a maximum covariance of $I_{40}$ for the blob filter. From the $E_{rms}(t)$ plots in 4(a), it can be observed that the tracking performance of the PGM filters and the EnKF are comparable. The Monte Carlo averaged NEES test statistic $\beta_t$ is plotted in Fig. 4(b). show that the EnKF and the PGM1 filter offer comparable performance. The plots of $log(L(t))$ and $V\sigma_2$ given in Fig. 4(c), (d) show that, in comparison to the EnKF, the PGM filters perform better in terms of the $v_{2\sigma}$ volume, whereas the EnkF estimates have the highest averaged likelihoods. The time averaged values of the performance metrics are listed in Table 4 along with the consistency fractions. The performances of the EnkF and the PGM filters are seen to be comparable. It may be observed that the EnKF is quite similar to a unimodal PGM Filter as they both rely on particle uncertainty propagation and Kalman measurement update.




(a) $E_{rms}(t)$

(b) NEES($\beta_t$)

(c) $log(L(t))$

(d) $log(V_{2\sigma})$

**Fig. 4.** Example 3: Results.

**Table 4**
Example 3: Results.

| | $\overline{E_{rms}}$ | $\beta_{t,c}(\%)$ | $\widehat{logL}$ | $\widehat{logV}\sigma_2$ |
|---|---|---|---|---|
| PGM1-UT | 18.0069 | 80.69 | 89.6553 | 152.8588 |
| PGM1 | 18.0452 | 70.30 | 89.6227 | 152.7732 |
| EnKF | 18.1055 | 81.19 | 89.8193 | 152.8034 |

## 5. Conclusions

A novel Gaussian mixture-particle PGM algorithm for nonlinear filtering has been presented. During the prediction step, the PGM filter uses an ensemble of particles to propagate the prior uncertainty and recovers a GMM representation of the propagated pdf by clustering. Measurements are incorporated through a Kalman update of the mixture modes to arrive at the posterior pdf. The PGM approach is not susceptible to the issue of particle depletion associated with particle filters. The PGM filter is applied to three test cases and offers superior estimation performance in comparison to UKF, PF the blob filter and a mixture UKF.

## References


Alspach, D. L., & Sorenson, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, *17*(4), 439–448.

Arasaratnam, Ienkaran, & Haykin, Simon (2009). Cubature Kalman filters. *IEEE Transactions on Automatic Control*, *54*(6), 1254–1269.

Arulampalam, S., Maskell, S., Gordon, N., & Clapp, T. (2001). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Bengtsson, T., Bickel, P., & Li, B. (2008). Curse-of-dimensionality revisited: Collapse of particle filter in very large scale systems. In *Probability and statistics: essays in honor of David A Freedman* (Vol. 2) (pp. 316–334).

Burgers, G., Van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724.

DeMars, K. J., Bishop, R. H., & Jah, M. K. (2013). Entropy-based approach for uncertainty propagation of nonlinear dynamical systems. *Journal of Guidance, Control, and Dynamics*, *36*(4), 1047–1056.

Dilshad Raihan, Akkam Veettil, & Chakravorty, Suman (2015). A ukf-pf based hybrid estimation scheme for space object tracking. In *Proceedings of the AAS/AIAA astrodynamics specialist conference* (pp. 4221–4240).

Douc, Randal, Moulines, Eric, & Ritov, Ya'Acov (2007). *Forgetting of the initial condition for the filter in general state-space hidden Markov chain: A coupling approach. Technical report*.

Duda, Richard O., Hart, Peter E., & Stork, David G. (2000). *Pattern classification* (2nd ed.). New York: Wiley-Interscience.

Dunik, Jindrich, Straka, Ondrej, Simandl, Miroslav, & Blasch, Erik (2015). Random-point-based filters: Analysis and comparison in target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, *51*(2), 1403–1421.

Evensen, G. (2002). *Data assimilation: The ensemble Kalman filter*. Berlin: Springer.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F, Radar and Signal Processing*, *140*(2), 107–113.

Hanebeck, Uwe D., & Briechle, Kai (2003). Progressive bayes: a new framework for nonlinear state estimation. In *SPIE Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications* (Vol. 5099) (pp. 256–267).

Jain, A. K., Murthy, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *13*(3), 264–323.

Julier, Simon J., Uhlmann, Jeffrey K., & Durrant-Whyte, H. F. (1995). A new approach for filtering nonlinear systems. In *Proceedings of the american control conference* (pp. 1628–1632).

Kallenberg, Olav (1997). *Foundations of modern probability*. Springer.

Kalman, Rudolph Emil (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*, 35–45.

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of ASME-Journal of Basic Engineering*, *83*, 96–108.

Kotecha, Jayesh H., & Djuric, Petar M. (2003). Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, *51*(10), 2602–2612.

Lefebvre, Tine, Bruyninckx, Herman, & De Schutter, Joris (2002). Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators". *IEEE Transactions on Automatic Control*, *47*(8), 1406–1409.

Lloyd, Stuart P. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Lorenz, E. N. (1995). Predictability: a problem partly solved. In *Seminar on predictability, 4–8 1995* (Vol. 1) (pp. 1–18). Reading: Shinfield Park, ECMWF, ECMWF.

Psiaki, Mark L. (2016). Gaussian mixture nonlinear filtering with resampling for mixand narrowing. *IEEE Transactions on Signal Processing*, *64*(21), 1047–1056.

Smith, G. L., Schmidt, S. F., & McGee, L. A. (1962). *Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle. Technical report NASA TR-135*. NASA.

Sorenson, H. W., & Alspach, D. L. (1971). Recursive bayesian estimation using gaussian sums. *Automatica*, *7*(4), 465–479.

Terejanu, G., Singla, P., Singh, T., & Scott, P. D. (2011). Adaptive gaussian sum filter for nonlinear bayesian estimation. *IEEE Transactions on Automatic Control*, *56*(9), 2151–2156.

Vershynin, Roman (2017). *High dimensional probability*. Cambridge University Press.

Wan, E. A., & Van Der Merwe, R. (2001). The unscented Kalman filter. In Simon Haykin (Ed.), *Kalman filter and neural networks*. New York: J. Wiley and Sons.

Yang, Tao, Mehta, Prashant G., & Meyn, Sean P. (2013). Feedback particle filter. *IEEE Transactions on Automatic Control*, *58*(10), 2465–2480.


**Dilshad Raihan A.V.** received his B.Tech in Mechanical Engineering in 2013 from the Indian Institute of Technology, Madras. He is now a Ph.D. student in the department of Aerospace Engineering at Texas A&M University, College Station. His main research interests are in Bayesian estimation and machine learning.

**Suman Chakravorty** obtained his B.Tech in Mechanical Engineering in 1997 from the Indian Institute of Technology, Madras and his Doctoral degree in Aerospace Engineering from the University of Michigan, Ann Arbor in 2004. From August 2004–August 2010, he was an Assistant Professor with the Aerospace Engineering Department at Texas A&M University, College Station and since August 2010, he has been an Associate Professor in the department. Dr. Chakravorty's broad research interests lie in the estimation and control of stochastic dynamical systems with application to autonomous, distributed robotic mapping and planning and situational awareness problems. He is a member of AIAA, ASME and IEEE. He is an Associate Editor for the ASME Journal on Dynamical Systems, Measurement and Control and the IEEE Robotics and Automation Letters.