Optimal Sampling for Point Mass Filtering with Applications to Cislunar Orbit Determination

Felipe Giraldo-Grueso[1*], Andrey A. Popov[2], Uwe D. Hanebeck[3], Renato Zanetti[1]

[1*]Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, 78712, TX, United States.
[2]Department of Information and Computer Sciences, The University of Hawai'i at Mānoa, Honolulu, 96822, HI, United States.
[3]Department of Computer Science, Karlsruhe Institute of Technology, Karlsruhe, Germany.

*Corresponding author(s). E-mail(s): fgiraldo@utexas.edu;
Contributing authors: apopov@hawaii.edu; uwe.hanebeck@kit.edu; renato@utexas.edu;

Abstract

The point mass filter (PMF) is a widely used strategy for solving the state estimation problem. In this filter, a deterministic grid of point particles is used to numerically solve the Bayesian recursive relations. Since the probability density function of the state is only approximated at the grid points, the placement of the grid is crucial for the performance of the filter. A new strategy, named the Silverman mass filter (SMF), has recently been developed, in which a Gaussian sum filter (GSF) update is performed before placing a new grid. In the SMF, the grid is placed at the mean of the GSF-updated points and is oriented and expanded to match their covariance. While the SMF has been shown to improve upon the standard PMF, the grid placement in the SMF can be inefficient for highly multimodal problems. This work introduces two new techniques for grid placement in the SMF: a clustering-based approach and an optimal deterministic sampling technique. These new techniques are shown to improve grid placement in the SMF when applied to a bimodal distribution example. Furthermore, the techniques are evaluated in three sequential filtering problems, including: the univariate nonstationary growth model, the Ikeda map, and a cislunar orbit determination example, where they show improved performance over the standard SMF by providing a more consistent and accurate state estimate.

Keywords: point mass filtering, clustering, deterministic sampling

1 Introduction

The state estimation problem for discrete-time continuous-amplitude systems involves solving Bayesian recursive relations (BRRs) [1]. The BRRs first describe the time evolution of a state probability density function (PDF) using the Chapman-Kolmogorov equation. After the PDF has been propagated and a measurement obtained, the PDF is updated using Bayes' rule. These relations can be solved analytically for linear dynamics, measurement models, and Gaussian additive noise. However, when either the dynamics or the measurement models are nonlinear, the solution can become intractable [1]. Since nonlinear dynamics and measurement models are of interest for aerospace applications, such as atmospheric entry [2, 3] and terrain relative navigation [4, 5], various alternatives must be explored to numerically approximate the solution to the BRRs.

Point mass filters (PMFs) are a strategy for numerically solving the BRRs [6]. Instead of describing the state PDF with random samples, as done in particle filters, the state PDF is discretized using a *deterministic* grid, where each grid point is assigned a probability value. This deterministic approach makes the approximation of the posterior PDF significantly superior to random sampling approaches [7]. As a result, PMFs have been shown to outperform particle filters in nonlinear systems [7–9].

In the PMF, the BRRs are numerically solved by first propagating an initial grid, representing the previous posterior PDF, using the system dynamics. A new grid is then placed at the mean and covariance of the approximated prior distribution, calculated using the propagated grid points. Once a new measurement is obtained, the grid is appropriately weighted using Bayes' rule. Since the PMF approximates the state PDF at these discrete points, placement of the grid can severely affect the estimation performance. Consequently, much of the research on this type of filter has focused on finding the optimal placement of the grid to improve estimation performance [7, 10–14].

A variant of the PMF, called the Silverman mass filter (SMF), has recently been proposed [4]. This new filter considers the propagated points as a Gaussian mixture (GM), allowing the use of a Gaussian sum filter (GSF) update before placing the new grid. Performing a GSF update leads to a better approximation of the mean and covariance of the posterior PDF, which can then be used to place the new grid closer to the true posterior distribution. These techniques are similar to the ensemble Gaussian mixture filter (EnGMF) [15–18], but have the advantage of using deterministic points.

If the posterior PDF is approximately unimodal, placing a new grid at the updated mean and covariance from the GSF-updated points (as done in the SMF) can be an improvement over placing a grid at the approximated prior distribution (as done in the PMF). However, for multimodal distributions, using only the mean and covariance of the approximated posterior distribution can be an inefficient approach to place the new grid. Since the points from the GSF update each carry their own mean, covariance,

and probability mass, more efficient approaches can use the full information from these updated points to find higher-quality grids.

This work presents two different strategies for generating the new grid in the SMF from the GSF-updated points.[1] In the first strategy, a density-based clustering algorithm is used to cluster the GSF-updated points. Once clustered, the individual mean and covariance of each cluster are calculated. A new grid is then placed on top of each cluster, improving the placement of the deterministic points and resulting in an overall grid that better describes the posterior distribution. Given that the clustering algorithm used is purely density-based, a second strategy is presented that considers the probability masses of the updated points. In this strategy, a set number of deterministic points are sampled from each Gaussian component describing the approximated posterior PDF, which are then optimally reduced to obtain a higher-quality approximation.

Throughout this work, a 'dual banana' distribution [20] is used to illustrate the limitations of the standard grid in the SMF and the advantages of the two proposed techniques for grid construction. For this example, the prior distribution is assumed to be Gaussian and centered at the origin, with a range measurement also assumed to be obtained from the origin. This setup results in a bimodal posterior distribution that resembles the shape of two bananas. Although the standard grid in the SMF fails to efficiently capture the true posterior distribution, the new points obtained using the two new approaches result in a better representation of the posterior PDF. The performance of the standard SMF and its two new variants is also evaluated using a sequential filtering problem involving nonlinear and chaotic dynamics, as described by the Ikeda map. Both new techniques demonstrate better estimation performance and consistency compared to the standard SMF. Additionally, this manuscript has been extended to include a summary of the new algorithms in pseudocode, and a performance study of the standard SMF and its two variants in two new sequential filtering problems: the univariate nonstationary growth model and a cislunar orbit determination example.

The remainder of this paper is organized as follows. Section 2 provides a brief introduction to state estimation for discrete-time continuous-amplitude systems. Section 3 introduces the SMF and uses the 'dual banana' distribution to demonstrate the limitations in grid construction. Section 4 presents the clustering-based approach, and the optimal deterministic sampling technique is presented in Section 5. Both techniques are evaluated using the 'dual banana' distribution in their respective sections. The proposed techniques are then compared in Section 6 using three sequential filtering problems, including the univariate nonstationary growth model, the Ikeda map, and a cislunar orbit determination example. Finally, Section 7 draws conclusions.

2 State Estimation for Discrete-Time Systems

The solutions to the BRRs are the foundation of the state estimation problem. In this section, these relations are presented and discussed in the case of discrete-time

[1] This paper is an extension to the work presented at the 2025 AAS/AIAA Spaceflight Mechanics Meeting [19].

continuous-amplitude systems, as they are fundamental for the derivation of the SMF. Let $\boldsymbol{x}_k$ be the state at time k, evolving according to discrete dynamics model $\boldsymbol{f}_k$ and following the measurement model $\boldsymbol{h}_k$, where

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}_k\left(\boldsymbol{x}_k, \boldsymbol{q}_k\right), \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}_k\left(\boldsymbol{x}_k, \boldsymbol{\eta}_k\right), \tag{2}$$

with $\boldsymbol{q}_k$ representing the process noise, $\boldsymbol{y}_k$ the measurement, and $\boldsymbol{\eta}_k$ the measurement noise.

The first step in the state estimation problem is to use the Chapman-Kolmogorov equation to propagate the state PDF [1]. Thus, given an initial state PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$,

$$p\left(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}\right) = \int\limits_{\mathcal{S}(\boldsymbol{x}_k)} p\left(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k\right) p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right)\, d\boldsymbol{x}_k, \tag{3}$$

with $\mathcal{S}\left(\boldsymbol{x}_k\right)$ denoting the support of $\boldsymbol{x}_k$, and $p\left(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k\right)$ the transition density. Assuming that a measurement is available ($k \leftarrow k+1$), the second step is to update the PDF using Bayes' rule [1], where

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) = \frac{p\left(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{y}_{1:k-1}\right) p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}\right)}{p\left(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}\right)}. \tag{4}$$

Therefore, the Chapman-Kolmogorov equation is used to propagate the PDF, and Bayes' rule is used to update the PDF, together forming the BRRs. In the state estimation problem, these equations are solved recursively until a desired time step. However, as noted earlier, the solution can become intractable when the dynamics and measurement models are nonlinear.

3 Point Mass Filtering

The PMF is a filtering strategy used to solve the state estimation problem by discretizing the support of the state PDF on a deterministic grid. This section provides a summary of the SMF, a new variant of the standard PMF, which uses GMs and kernel density estimation (KDE) techniques to achieve better estimation performance. This summary is included in order for this work to be self-contained, but the readers are encouraged to refer to the original paper for a more detailed explanation [4]. The SMF approach begins by approximating the initial posterior state PDF with a grid of point particles $\boldsymbol{\mathcal{X}}_{k|k}^{(i)}$,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} w_{k|k}^{(i)}\, \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}_{k|k}^{(i)}\right), \tag{5}$$

where $w_{k|k}^{(i)}$ represents the assigned probabilities for each point, and N is the number of points.

The SMF proposes to describe the initial Dirac mixture (DM) as an equivalent GM,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} w_{k|k}^{(i)} \lim_{\rho(P)\to 0} \mathcal{N}\left(\boldsymbol{x}_k; \boldsymbol{\mathcal{X}}_{k|k}^{(i)}, P\right), \tag{6}$$

where $\rho(P)$ is the spectral radius of the covariance matrix P. By describing the PDF as a GM and assuming zero-mean additive process noise with covariance Q_k, the GSF algorithm [21, 22] and KDE techniques can be used to propagate the PDF to the next time step, resulting in a predictive distribution,

$$p\left(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} w_{k+1|k}^{(i)} \mathcal{N}\left(\boldsymbol{x}_{k+1}; \boldsymbol{f}_k\left(\boldsymbol{\mathcal{X}}_{k|k}^{(i)}\right), B_k\right). \tag{7}$$

Note that other kernels can be used [23], but this work uses Gaussian kernels. For this propagation step, $w_{k+1|k}^{(i)} = w_{k|k}^{(i)}$, and,

$$B_k = \beta^2 \hat{P}_{k+1|k} + Q_k, \tag{8}$$

$$\hat{P}_{k+1|k} = \sum_{i=1}^{N} w_{k|k}^{(i)} \left[\boldsymbol{f}_k\left(\boldsymbol{\mathcal{X}}_{k|k}^{(i)}\right) - \sum_{i=1}^{N} w_{k|k}^{(i)} \boldsymbol{f}_k\left(\boldsymbol{\mathcal{X}}_{k|k}^{(i)}\right)\right] [\cdots]^{\mathrm{T}}, \tag{9}$$

where $[\cdots]$ signifies the same quantity as the preceding parentheses, and β^2 is the bandwidth parameter, calculated using Silverman's rule of thumb [24],

$$\beta_{\text{sil}}^2 = \left(\frac{4}{N\left(N_s+2\right)}\right)^{\frac{2}{N_s+4}}, \tag{10}$$

where N_s is the state space dimension. The bandwith parameter, β_{sil}^2, can be scaled by a factor $0 < \alpha$, such that

$$\beta^2 = \alpha \beta_{\text{sil}}^2. \tag{11}$$

As Silverman's rule of thumb is known to be conservative for some non-Gaussian distributions [25], the scaling factor is usually chosen to satisfy $0 < \alpha \leq 1$. As α increases, the filter becomes more conservative, preventing divergence in challenging scenarios. As the filter becomes more conservative, estimation accuracy might improve. However, if the filter is too conservative, the estimation accuracy might degrade. The optimal α value depends on the application, but $\alpha = 1$ is usually a good starting point. For more information on how to pick α, readers are encourage to refer to the SMF paper [4].

When a measurement is obtained ($k \leftarrow k+1$), a GSF update [21, 22] is performed to obtain an approximation to the posterior distribution, such that

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} \widetilde{w}_{k|k}^{(i)} \mathcal{N}\left(\boldsymbol{x}_k; \widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}, \widetilde{\mathcal{P}}_{k|k}^{(i)}\right), \tag{12}$$

where the notation $\widetilde{\cdot}$ represents auxiliary variables obtained with a GSF update. Each auxiliary mean and covariance of the approximated posterior distribution is given by

$$\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)} = \boldsymbol{\mathcal{X}}_{k|k-1}^{(i)} + K_k^{(i)} \boldsymbol{\nu}_k^{(i)}, \tag{13}$$

$$\widetilde{\mathcal{P}}_{k|k}^{(i)} = B_{k-1} - K_k^{(i)} W_k^{(i)} K_k^{(i)\mathrm{T}}, \tag{14}$$

with the rest of the variables defined as

$$\boldsymbol{\mathcal{X}}_{k|k-1}^{(i)} = \boldsymbol{f}_{k-1}\left(\boldsymbol{\mathcal{X}}_{k-1|k-1}^{(i)}\right), \tag{15}$$

$$\boldsymbol{\nu}_k^{(i)} = \boldsymbol{y}_k - \boldsymbol{h}_k\left(\boldsymbol{\mathcal{X}}_{k|k-1}^{(i)}\right), \tag{16}$$

$$K_k^{(i)} = B_{k-1} H_k^{(i)\mathrm{T}} \left(W_k^{(i)}\right)^{-1}, \tag{17}$$

$$W_k^{(i)} = H_k^{(i)} B_{k-1} H_k^{(i)\mathrm{T}} + R_k, \tag{18}$$

$$H_k^{(i)} = \frac{\partial \boldsymbol{h}_k(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=\boldsymbol{\mathcal{X}}_{k|k-1}^{(i)}}. \tag{19}$$

If $\boldsymbol{h}_k(\cdot)$ is highly nonlinear, the GSF update can be performed in a unscented Kalman filter (UKF) fashion [26]. In this new GM, each Gaussian component is weighted proportional to the probability of the measurement,

$$\widetilde{w}_{k|k}^{(i)} \propto w_{k|k-1}^{(i)} \mathcal{N}\left(\boldsymbol{y}_k; \boldsymbol{h}_k\left(\boldsymbol{\mathcal{X}}_{k|k-1}^{(i)}\right), W_k^{(i)}\right). \tag{20}$$

With the approximated posterior distribution, a new grid is constructed in order to continue with the recursion. In the SMF, a new rectangular grid is constructed after the GSF update, by centering the grid at the updated mean and aligning it with the updated covariance, given by

$$\widetilde{\boldsymbol{x}}_{k|k} = \sum_{i=1}^{N} \widetilde{w}_{k|k}^{(i)} \widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}, \tag{21}$$

$$\widetilde{P}_{k|k} = \sum_{i=1}^{N} \widetilde{w}_{k|k}^{(i)} \left(\widetilde{\mathcal{P}}_{k|k}^{(i)} + \widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)} \left(\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}\right)^{\mathrm{T}} - \widetilde{\boldsymbol{x}}_{k|k} \widetilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}}\right), \tag{22}$$

With this new grid, the approximate posterior distribution is once again described as a DM,

$$p\left(\boldsymbol{x}_k | \boldsymbol{y}_{1:k}\right) \approx \sum_{j=1}^{M} w_{k|k}^{(j)} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}_{k|k}^{(j)}\right), \tag{23}$$

where $\boldsymbol{\mathcal{X}}_{k|k}^{(j)}$ are the new M discretization points with their corresponding weights $w_{k|k}^{(j)}$.

To obtain the new weights for each point, the following process is used. First, $\boldsymbol{x}_k$ is marginalized to obtain an expression for the posterior distribution,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) = \int\limits_{\mathcal{S}(\boldsymbol{x}_{k-1})} p\left(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k}\right)\, d\boldsymbol{x}_{k-1}, \tag{24}$$

$$= \int\limits_{\mathcal{S}(\boldsymbol{x}_{k-1})} p\left(\boldsymbol{y}_k|\boldsymbol{x}_k\right) p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}\right)\, d\boldsymbol{x}_{k-1}, \tag{25}$$

$$= p\left(\boldsymbol{y}_k|\boldsymbol{x}_k\right) \int\limits_{\mathcal{S}(\boldsymbol{x}_{k-1})} p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}\right)\, d\boldsymbol{x}_{k-1}. \tag{26}$$

At this point, note that the new grid ($\boldsymbol{\mathcal{X}}_{k|k}$) and the previous posterior grid ($\boldsymbol{\mathcal{X}}_{k-1|k-1}$) with its corresponding weights ($w_{k-1|k-1}$) are available. Since the previous posterior is approximated via a DM (or a GM with $\rho\left(P\right) \to 0$), the sifting property can be used to evaluate the obtained integral. Given

$$p\left(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}\right) \approx \sum_{i=1}^{N} w_{k-1|k-1}^{(i)} \delta\left(\boldsymbol{x}_{k-1} - \boldsymbol{\mathcal{X}}_{k-1|k-1}^{(i)}\right), \tag{27}$$

then Eq. (26) can be rewritten as

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx p\left(\boldsymbol{y}_k|\boldsymbol{x}_k\right) \int\limits_{\mathcal{S}(\boldsymbol{x}_{k-1})} p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}\right) \sum_{i=1}^{N} w_{k-1|k-1}^{(i)} \delta\left(\boldsymbol{x}_{k-1} - \boldsymbol{\mathcal{X}}_{k-1|k-1}^{(i)}\right)\, d\boldsymbol{x}_{k-1}, \tag{28}$$

$$= p\left(\boldsymbol{y}_k|\boldsymbol{x}_k\right) \sum_{i=1}^{N} p\left(\boldsymbol{x}_k|\boldsymbol{\mathcal{X}}_{k-1|k-1}^{(i)}\right) w_{k-1|k-1}^{(i)}, \tag{29}$$

resulting in weights proportional to

$$w_{k|k}^{(j)} \propto p\left(\boldsymbol{y}_k|\boldsymbol{\mathcal{X}}_{k|k}^{(j)}\right) \sum_{i=1}^{N} p\left(\boldsymbol{\mathcal{X}}_{k|k}^{(j)}|\boldsymbol{\mathcal{X}}_{k-1|k-1}^{(i)}\right) w_{k-1|k-1}^{(i)}. \tag{30}$$

Note that the weighting of the new grid makes no assumption about the location or distribution of the points. In addition, from Eq. (30), it can be seen that this filtering algorithm has a linear time complexity in both N and M, $\mathcal{O}\left(N \cdot M\right)$. These equations summarize the standard SMF methodology. Figure 1 shows a visual summary of the key steps in this new technique.

Although the SMF has been shown to outperform the standard PMF [4], this framework has clear limitations. The described procedure works well for approximately unimodal PDFs, as in this case, the mean and covariance of the GSF-updated points are indicative of where the grid should be placed. As the approximate posterior PDF

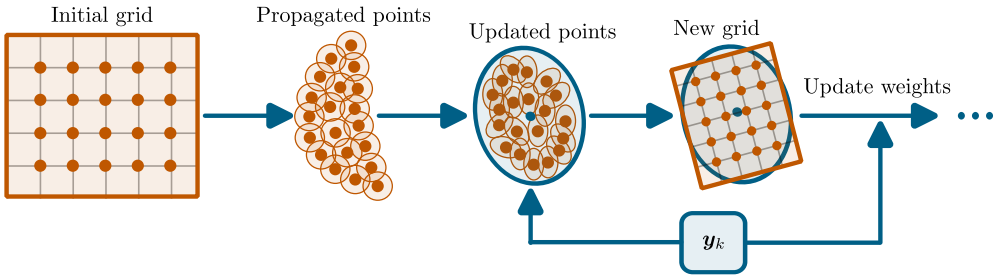


Fig. 1: Flowchart summarizing the four key steps in the SMF: First, the initial grid representing the previous posterior is propagated. Then, the propagated points are updated using a GSF update. From the updated points, a new rectangular grid is placed at the updated mean and covariance. Finally, the grid is weighted accordingly.

starts to become multimodal, the SMF grid placement becomes inefficient. If the modes of the PDF are close together, the mean and covariance can still be indicative of where the grid should be placed. However, as the modes start to move apart, using only the mean and covariance can result in large grids, covering areas of very low or zero probability. Therefore, other techniques have to be used to obtain better grid point positions.

The problem of accurately placing a grid in a multimodal distribution has been studied before. In [27], grid splitting is proposed after the measurement update, as some areas of the state space may drop in probability when new information is included. Therefore, after the measurement update step, the grid is split by finding the locations where the approximate marginal PDFs are negligible. The new grids are then used independently for the next time step. An improvement to the overall splitting method is presented in [28], by merging overlapping grids using the Mahalanobis distance between their local means. The complete algorithm, using both splitting and merging techniques, is presented in [7]. Instead of splitting the grid, in [29], a density difference grid design is proposed. For this technique, the grid is designed to be sparse in the low-probability regions and denser in the higher-probability regions, calculated by the spatial variation of the PDF. These strategies are proposed in the framework of the standard PMF, where a new grid is placed at the approximated prior. In the SMF, an approximated posterior PDF is available before constructing the new grid, via the GSF-updated points. Thus, compared to the other strategies, this work provides two new multigrid approaches by considering the approximated posterior given by the GM in Eq. (12).

3.1 Limitations in Grid Construction

This section provides an example where the SMF fails to obtain an accurate representation of the posterior distribution in the case of a bimodal PDF [20]. Consider the

following system where the prior distribution is given by

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}\right). \tag{31}$$

If a range measurement is assumed to be obtained from the origin ($k \leftarrow k+1$),

$$y_k = \sqrt{\left(x_k^{(1)}\right)^2 + \left(x_k^{(2)}\right)^2}, \tag{32}$$

with a value of $y_k = 3$ and a measurement error covariance of $R = 0.1^2$, the true posterior distribution will have two modes that resemble the shape of two bananas, as shown in Fig. 2. If a GSF update is performed on an initial grid with $N = 400$ points, centered on the prior mean and expanding the prior covariance (simulating the propagated points from the previous posterior distribution), the GSF-updated points cluster around the two modes of the true posterior distribution, as illustrated in the left panel of Fig. 2. If the new grid, with $M = 400$ points, is created using only the first two moments of the GSF-updated points (as is done in the SMF), a significant amount of information is lost, as seen in the right panel of Fig. 2. In this case, using the standard PMF approach, where the grid is centered at the prior mean and expands the prior covariance, could yield a better grid approximation.

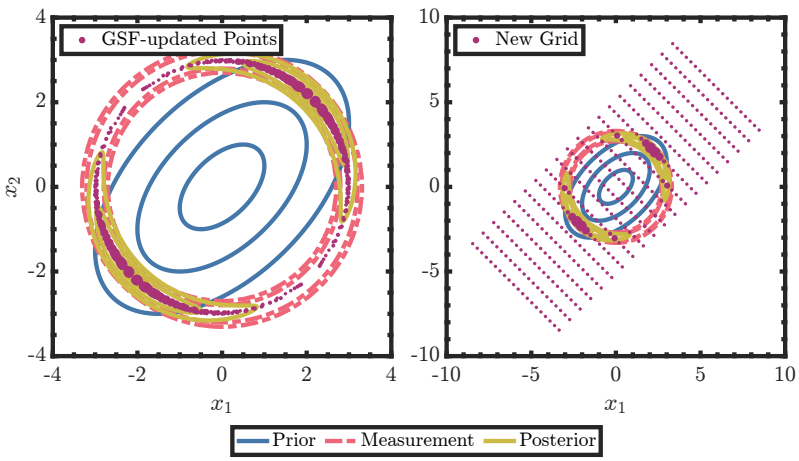


Fig. 2: GSF-updated points in purple (left) and new grid in purple (right). The 1σ, 2σ, and 3σ contours of the prior distribution are shown in blue. The 2σ and 3σ contours of the measurement likelihood are plotted in red. The true posterior distribution is shown in yellow.

Therefore, a more sophisticated method that considers all the information provided by the GSF update is required to create the new grid. Consequently, the problem to

be solved in this work can be formulated as follows.

Given the information from the GSF-updated points, what is the optimal grid?

In other words, this work addresses the challenge of optimally representing the GSF-approximated posterior distribution,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} \tilde{w}_{k|k}^{(i)} \mathcal{N}\left(\boldsymbol{x}_k; \tilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}, \tilde{\mathcal{P}}_{k|k}^{(i)}\right) \tag{33}$$

as a DM approximation,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{j=1}^{M} w_{k|k}^{(j)} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}_{k|k}^{(j)}\right). \tag{34}$$

A possible solution would be to simply use the GSF-updated points for the next iteration. However, using these points to represent the GM in Eq. (33) as a DM in Eq. (34) can result in a poor approximation. As mentioned earlier, each GSF-updated point represents a Gaussian distribution rather than a single point, meaning that their locations are not optimized to be treated as such. Furthermore, points with low probability would be fully considered, which could lead to filter degeneracy or divergence, particularly in cases of poor propagation or low-quality measurements.

4 A First Approach Using Clustering Algorithms

By examining Fig. 2, it is clear that the GSF-updated points are clustered around the two modes of the true posterior distribution. These points suggest using cluster-based algorithms to autonomously determine the number of statistically significant clusters and place independent grids on each. This section explores the possibility of using clustering algorithms to obtain higher quality grids.

Density-based spatial clustering of applications with noise (DBSCAN) [30] is a commonly used clustering algorithm. As its name suggests, this is a density-based clustering algorithm that is used to find clusters and noise in a set of points. The advantage of this algorithm is that, since it is purely density-based, there is no need to pre-define the number of clusters desired, unlike other clustering methods such as k-means [31]. In this algorithm, each point is classified as core, border, or noise. Given the parameters `epsilon` and `minPts`, the algorithm works by following the next steps: [30]

1. An unlabeled point is selected as the current point and the first cluster is initialized.
2. The neighboring points within an `epsilon` distance from the current point are found.

3. If the number of neighbors is less than `minPts`, the current point is labeled as noise. Otherwise, the point is labeled as a core point that belongs to the current cluster.
4. The same process is repeated over each neighbor until no new neighbors can be added to the current cluster.
5. Steps 1 to 4 are repeated until all points are labeled.

Using DBSCAN, this approach proposes first to classify the GSF-updated points into a set of clusters such that

$$\widetilde{\boldsymbol{\mathcal{X}}}_{k|k} = \left\{ {}^{(1)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}, {}^{(2)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}, \ldots, {}^{(L)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k} \right\}, \tag{35}$$

where ${}^{(i)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}$ represents each cluster and L is the total number of clusters found. With the clusters identified, the mean and covariance of each one can be calculated as

$${}^{(i)}\widetilde{\boldsymbol{x}}_{k|k} = \sum_{j=1}^{{}^{(i)}N} {}^{(i)}\widetilde{w}_{k|k}^{(j)} \, {}^{(i)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(j)}, \tag{36}$$

$${}^{(i)}\widetilde{P}_{k|k} = \sum_{j=1}^{{}^{(i)}N} {}^{(i)}\widetilde{w}_{k|k}^{(j)} \left({}^{(i)}\widetilde{\mathcal{P}}_{k|k}^{(j)} + {}^{(i)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(j)} \left({}^{(i)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(j)} \right)^{\mathrm{T}} - {}^{(i)}\widetilde{\boldsymbol{x}}_{k|k} \, {}^{(i)}\widetilde{\boldsymbol{x}}_{k|k}^{\mathrm{T}} \right), \tag{37}$$

where ${}^{(i)}N$ are the number of points in the i-th cluster. In this case, the weights of the clusters need to be re-normalized, such that

$$\sum_{j=1}^{{}^{(i)}N} {}^{(i)}\widetilde{w}_{k|k}^{(j)} = 1. \tag{38}$$

With the mean and covariance of each cluster, new grids are placed at each cluster center, aligned with its mean and covariance. The number of grid points for every new grid is set to be equal to the number of points in its corresponding cluster,

$$\sum_{i=1}^{L} {}^{(i)}N = N, \tag{39}$$

making $M = N$. The union of all grids is then used as the new grid, which can be weighted as in Eq. (30). It is important to mention that, since DBSCAN can classify points as noise, the GSF-updated points classified as noise are not used to place the grids. Instead, the total number of noise points is distributed among the new grids to satisfy the constraint $M = N$.

As mentioned above, the number of grid points for every new grid is set to be equal to the number of points in its corresponding cluster. Satisfying this constraint can be difficult using rectangular grids, as not every number can be expressed as a

sum of squares (in the case of two-dimensional grids). For this work, instead of using rectangular grids, generalized Fibonacci grids [32] are used to place a new grid on top of each cluster. Generalized Fibonacci grids are a simple method to obtain equally weighted deterministic samples of a multivariate Gaussian PDF, allowing the sampling of an arbitrary number of homogeneously distributed points, even in high-dimensional spaces. Figure 3 shows three Fibonacci grids, each obtained for a different number of grid points in the case of a standard two-dimensional normal distribution. Note that as fewer points are used, fewer points lie outside the 2σ bounds. Therefore, if a smaller number of points is used, the grid points can be stretched without losing homogeneity, to cover up to an arbitrary number of σ, resulting in a more conservative approach.

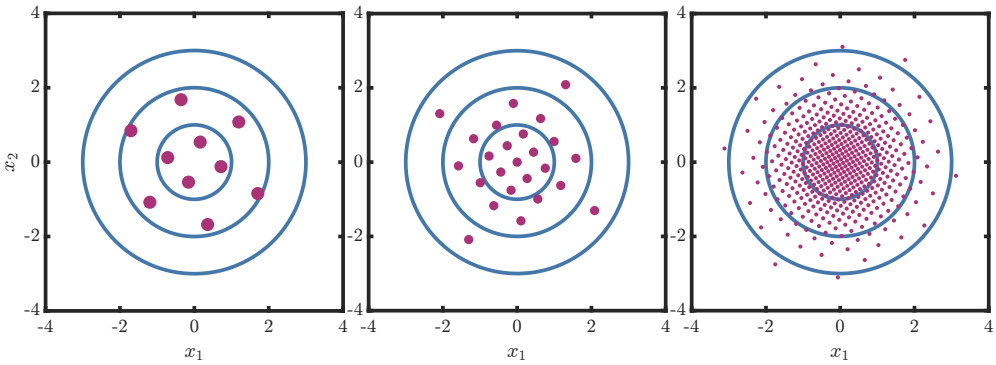


Fig. 3: Fibonacci grids for a standard two-dimensional normal distribution with 10 grid points (left), 25 grid points (center), 500 grid points (right). The circles represent the 1σ, 2σ, and 3σ bounds.

It is important to note that this strategy can result in suboptimal grids for two reasons. First, the results obtained with DBSCAN depend highly on the choice of the parameters `epsilon` and `minPts`. Second, since DBSCAN is a density-based algorithm, it does not consider the probability mass of each GSF-updated point, which represents a Gaussian distribution rather than a single point. However, this clustering algorithm has the potential to produce higher-quality grids compared to the original strategy in the SMF. In addition, since DBSCAN has a worst-case time complexity of $\mathcal{O}(N^2)$, using this framework will not affect the overall complexity of the filtering algorithm.

To study the performance of this new grid construction scheme, consider the example presented in the previous section. In this case, to evaluate the quality of the new grid, the effective grid size is used. This metric is defined as

$$M_{\text{eff}} = \frac{1}{M \sum_{i=1}^{M} \left(w_{k|k}^{(i)} \right)^2}. \tag{40}$$

If M_{eff} is close to 1, then the new grid accurately represents the approximate posterior distribution. In the case where $M_{\text{eff}} \ll 1$, the new grid is not an appropriate

representation of the underlying PDF. Figure 4 shows the different grids obtained for the dual banana problem using the standard SMF approach and the clustering strategy with different values for `epsilon` and `minPts`. More specifically, the top-left figure shows the GSF-updated points with the traditional grid at the bottom-left. The top-center figure shows the clusters obtained (in different colors) for the GSF-updated points using `epsilon` = 0.50 and `minPts` = 5, with the corresponding new grid at the bottom-center. The top-right figure shows the clusters obtained for the GSF-updated points using `epsilon` = 0.17 and `minPts` = 3, with the corresponding new grid at the bottom-right.

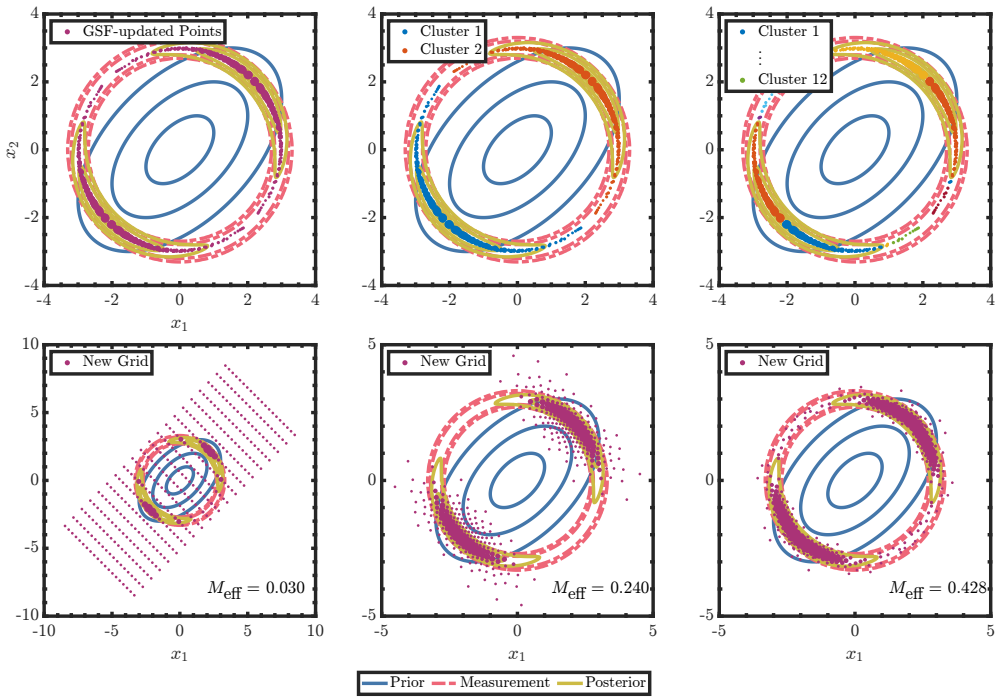


Fig. 4: Standard grid placement (left), grid placement using DBSCAN with `epsilon` = 0.50 and `minPts` = 5 (center), and grid placement using DBSCAN with `epsilon` = 0.17 and `minPts` = 3 (right). The top panels show the GSF-updated points and clusters, while the bottom panels show the grid placement.

Figure 4 shows the advantages of using the clustering algorithm over the standard SMF method. By using the clustering algorithm with `epsilon` = 0.50 and `minPts` = 5, the GSF-updated points are divided into two clusters that accurately represent the two modes of the posterior distribution. Placing the new grid on top of the two identified clusters significantly increases the effective grid size compared to the standard grid. If `epsilon` and `minPts` are decreased, the GSF-updated points are divided into more clusters. By having more clusters, additional grids are superimposed, better capturing

the underlying shape of the posterior PDF, and thus further increasing the effective grid size. Although the use of more clusters may increase the effective grid size, it is important to mention that this can affect the consistency of the overall filter, as the new grid can become too narrow.

5 A Solution Based on Optimal Point Sampling

As mentioned in the previous section, although using clustering-based algorithms might yield higher-quality grids when compared to the standard SMF grid, relying solely on density-based clustering methods can lead to overall suboptimal grids. Since the GSF-updated points carry an individual mean, an individual covariance, and an associated probability mass, density-based clustering methods will not use the full information provided by the GM. This section presents an alternative solution that uses all available information to find an optimal set of points using a reduction of DMs by minimizing a distance measure.

Recall that the goal of this work is to represent the GSF-approximated posterior distribution,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} \tilde{w}_{k|k}^{(i)} \mathcal{N}\left(\boldsymbol{x}_k; \tilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}, \tilde{\mathcal{P}}_{k|k}^{(i)}\right) \tag{41}$$

as a DM approximation,

$$p\left(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right) \approx \sum_{j=1}^{M} w_{k|k}^{(j)} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}_{k|k}^{(j)}\right). \tag{42}$$

Therefore, previously developed tools can be combined to find optimal points by maximizing the similarity between the two distributions. Specifically, the strategy used in this work consists of two distinct steps. First, deterministic samples are obtained from each of the Gaussian components that represent the GSF-approximated posterior distribution. Second, the union of these deterministic samples is optimally reduced to obtain a high-quality DM approximation.

To explain the process of obtaining the high-quality DM approximation, two important concepts have to be introduced first, the localized cumulative density (LCD) and the modified Cramér-von Mises distance (MCVMD) [33–35].

5.1 Localized Cumulative Distribution

Let $p\left(\boldsymbol{x}\right) : \mathbb{R}^{N_s} \to \mathbb{R}_+$ be the PDF of a random vector $\boldsymbol{x} \in \mathbb{R}^{N_s}$. The corresponding LCD is defined as

$$P\left(\boldsymbol{m}, b\right) = \int_{\mathbb{R}^{N_s}} p\left(\boldsymbol{x}\right) \mathcal{K}\left(\boldsymbol{x} - \boldsymbol{m}, \boldsymbol{b}\right) d\boldsymbol{x}, \tag{43}$$

where $\mathcal{K}\left(\boldsymbol{x} - \boldsymbol{m}, \boldsymbol{b}\right)$ is a kernel centered on $\boldsymbol{m}$ with bandwidth $\boldsymbol{b}$ [35]. Following the formulation in [35], where the kernel has to be separable and have equal width in every

dimension, this work considers Gaussian kernels of the form:

$$\mathcal{K}\left(\boldsymbol{x}-\boldsymbol{m}, b\right)=\prod_{i=1}^{N_s}\exp\left(\frac{-\left(x^{(i)}-m^{(i)}\right)^2}{2b^2}\right). \tag{44}$$

Note that other kernels could be used, but the Gaussian kernel simplifies the optimization step defined in the next sections. With this definition and kernel, the LCD for a DM is given by

$$P\left(\boldsymbol{m}, b\right)=\sum_{j=1}^{N} w_x^{(j)}\prod_{i=1}^{N_s}\exp\left(\frac{-\left(x^{(i)}-m^{(i)}\right)^2}{2b^2}\right). \tag{45}$$

5.2 Modified Cramér-von Mises Distance

The MCVMD allows for the comparison of similarity between two different distributions without restrictions on whether they are continuous or discrete. This distance is defined as the integral of the squared difference between the LCDs of two PDFs [35],

$$d_{\mathrm{MCVM}}=\int_{\mathbb{R}_+} w(b)\int_{\mathbb{R}^{N_s}}\left(P_1\left(\boldsymbol{m}, b\right)-P_2\left(\boldsymbol{m}, b\right)\right)^2 d\boldsymbol{m}\, db, \tag{46}$$

where $w(b)\colon \mathbb{R}_+ \mapsto [0,1]$ is an appropriate weighting function. If the weighting function for the MCVMD is defined as

$$w(b)=\begin{cases} b^{1-n}, & b\in[0, b_{\max}] \\ 0, & \text{otherwise,}\end{cases} \tag{47}$$

then the distance between the two distributions and its corresponding gradient can be easily approximated for large values of $b_{\max}$. In the case of two DMs and a large value of $b_{\max}$, the MCVMD between these two distributions has a closed-form solution and a closed-form solution gradient as well [35].

Let two DMs be expressed as

$$p_1\left(\boldsymbol{x}\right)=\sum_{i=1}^{N} w_{\boldsymbol{\mathcal{X}}_1}^{(i)}\delta\left(\boldsymbol{x}-\boldsymbol{\mathcal{X}}_1^{(i)}\right), \tag{48}$$

$$p_2\left(\boldsymbol{x}\right)=\sum_{i=1}^{M} w_{\boldsymbol{\mathcal{X}}_2}^{(i)}\delta\left(\boldsymbol{x}-\boldsymbol{\mathcal{X}}_2^{(i)}\right), \tag{49}$$

where $\boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1}$ and $\boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2}$ are column vectors that represent the weights of each distribution, such that

$$\boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1}=\begin{bmatrix} w_{\boldsymbol{\mathcal{X}}_1}^{(1)} & w_{\boldsymbol{\mathcal{X}}_1}^{(2)} & \cdots & w_{\boldsymbol{\mathcal{X}}_1}^{(N)}\end{bmatrix}^{\mathrm{T}}, \tag{50}$$

$$\boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2} = \begin{bmatrix} w_{\boldsymbol{\mathcal{X}}_2}^{(1)} & w_{\boldsymbol{\mathcal{X}}_2}^{(2)} & \cdots & w_{\boldsymbol{\mathcal{X}}_2}^{(M)} \end{bmatrix}^{\mathrm{T}}, \tag{51}$$

and $\boldsymbol{\mathcal{X}}_1$, $\boldsymbol{\mathcal{X}}_2$, are the matrices representing the set of points in the DM,

$$\boldsymbol{\mathcal{X}}_1 = \begin{bmatrix} \boldsymbol{\mathcal{X}}_1^{(1)} & \boldsymbol{\mathcal{X}}_1^{(2)} & \cdots & \boldsymbol{\mathcal{X}}_1^{(N)} \end{bmatrix}, \tag{52}$$

$$\boldsymbol{\mathcal{X}}_2 = \begin{bmatrix} \boldsymbol{\mathcal{X}}_2^{(1)} & \boldsymbol{\mathcal{X}}_2^{(2)} & \cdots & \boldsymbol{\mathcal{X}}_2^{(M)} \end{bmatrix}. \tag{53}$$

where $\boldsymbol{\mathcal{X}}_1$ is a $N_s \times N$ matrix and $\boldsymbol{\mathcal{X}}_2$ is a $N_s \times M$ matrix. With this notation, the MCVMD for these two DMs can be written as

$$\begin{aligned} d_{\mathrm{MCVM}}\left(\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1}, \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2}\right) = {} & \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1}^{\mathrm{T}} \boldsymbol{M}_{\boldsymbol{\mathcal{X}}_1\boldsymbol{\mathcal{X}}_1} \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1} - 2\boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1}^{\mathrm{T}} \boldsymbol{M}_{\boldsymbol{\mathcal{X}}_1\boldsymbol{\mathcal{X}}_2} \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2} \\ & + \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2}^{\mathrm{T}} \boldsymbol{M}_{\boldsymbol{\mathcal{X}}_2\boldsymbol{\mathcal{X}}_2} \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2} + K \cdot ||\boldsymbol{\mathcal{X}}_1 \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_1} - \boldsymbol{\mathcal{X}}_2 \boldsymbol{w}_{\boldsymbol{\mathcal{X}}_2}||_2^2, \end{aligned} \tag{54}$$

where $K > 0$ is a constant that can be tuned so that all terms have similar significance, and,

$$\boldsymbol{M}_{\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_j} = \mathrm{xlog}\left(\boldsymbol{D}_{\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_j}\right), \tag{55}$$

with $\mathrm{xlog}\left(z\right) = z \cdot \log\left(z\right)$ and $\boldsymbol{D}_{\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_j}$ representing the Euclidean distance matrix between $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$. The gradient with respect to $\boldsymbol{\mathcal{X}}_1$ or $\boldsymbol{\mathcal{X}}_2$ is not presented for brevity, although it also has a closed-form solution that can be obtained by differentiating Eq. (54) [35].

5.3 Optimal Sampling of Gaussian Mixtures

With the LCD and MCVMD defined, the process of obtaining high-quality points can be presented. As mentioned before, the first step is to obtain equally weighted deterministic samples from each Gaussian component, such that

$$\mathcal{N}\left(\boldsymbol{x}_k; \tilde{\boldsymbol{\mathcal{X}}}_{k|k}^{(i)}, \tilde{\mathcal{P}}_{k|k}^{(i)}\right) \approx \frac{1}{D} \sum_{j=1}^{D} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{D}}_{k|k,i}^{(j)}\right), \tag{56}$$

where D are the total number of points to be sampled from each Gaussian component and $\boldsymbol{\mathcal{D}}_{k|k,i}^{(j)}$ are the deterministic samples. With this approximation, the GM obtained with the GSF update can be expressed as,

$$p\left(\boldsymbol{x}_k | \boldsymbol{y}_{1:k}\right) \approx \sum_{i=1}^{N} \frac{\tilde{w}_{k|k}^{(i)}}{D} \sum_{j=1}^{D} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{D}}_{k|k,i}^{(j)}\right). \tag{57}$$

To find the deterministic points, the previously mentioned Fibonacci grids are used. First, a set of samples $\boldsymbol{\mathcal{D}}$, representing the Fibonacci grid of a standard multivariate

Gaussian (i.e., zero mean and identity covariance), are obtained. Once the deterministic samples have been obtained, they can be scaled and shifted by the corresponding mean and covariance,

$$\boldsymbol{\mathcal{D}}_{k|k,i} = \left[\tilde{\mathcal{P}}^{(i)}_{k|k}\right]^{1/2} \boldsymbol{\mathcal{D}} + \tilde{\boldsymbol{\mathcal{X}}}^{(i)}_{k|k}. \tag{58}$$

For this work, the Cholesky decomposition is used to compute the matrix square root. The process of obtaining the unscaled deterministic samples $\boldsymbol{\mathcal{D}}$ can be performed offline, thus not increasing the overall complexity of the filtering algorithm.

Once the GM has been represented as a union of DMs, the final step is to optimally reduce the approximation to obtain the desired high-quality approximation. This final step can be expressed as the approximation of

$$\sum_{i=1}^{N} \frac{\tilde{w}^{(i)}_{k|k}}{D} \sum_{j=1}^{D} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{D}}^{(j)}_{k|k}\right) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}^{(j)}_{k|k}\right). \tag{59}$$

To unclutter notation, let the left-hand side of Eq. (59) be rewritten as

$$\sum_{i=1}^{N} \frac{\tilde{w}^{(i)}_{k|k}}{D} \sum_{j=1}^{D} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{D}}^{(j)}_{k|k}\right) = \sum_{i=1}^{D\cdot N} w^{(i)}_{k|k} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{Y}}^{(i)}_{k|k}\right), \tag{60}$$

where $\boldsymbol{\mathcal{Y}}_{k|k}$ represents the union of DMs with corresponding weights $w_{k|k}$, such that

$$\sum_{i=1}^{D\cdot N} w^{(i)}_{k|k} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{Y}}^{(i)}_{k|k}\right) \approx \frac{1}{M} \sum_{j=1}^{M} \delta\left(\boldsymbol{x}_k - \boldsymbol{\mathcal{X}}^{(j)}_{k|k}\right). \tag{61}$$

In this representation, the discrete density obtained in the previous step is reduced to M equally weighted discretization points, representing the optimal points.

Using the notation presented for the MCVMD definition, the location of the new points is found by solving the following minimization problem

$$\boldsymbol{\mathcal{X}}_{k|k} = \arg\min_{\boldsymbol{\mathcal{X}}} d_{\mathrm{MCVM}}\left(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}_{k|k}, \frac{1_{M\times 1}}{M}, w_{k|k}\right). \tag{62}$$

In this work, the MCVMD between the two DMs is minimized using Matlab's `fminunc` function with default settings, which uses the BFGS quasi-Newton method with a cubic line search procedure. This step has to be performed online as it requires the current GSF-approximated posterior. When used within the SMF framework, this sampling technique, like the others, uses $M = N$.

Before presenting a numerical example of this technique, three remarks must be made. First, it is important to discuss the initialization of the points $\boldsymbol{\mathcal{X}}$ when solving Eq. (62). To keep the deterministic nature of the filter, these points can be initialized using a single Fibonacci grid, with M points, centered at the mean and covariance of the GSF-updated points given by Eq. (36) and Eq. (37). This Fibonacci grid can

also be obtained offline and scaled by the mean and covariance of the GSF-updated points when the reduction is performed. If $M = N$, the GSF-updated points can be used to initialize the minimization problem instead. Second, note that the new points $\boldsymbol{\mathcal{X}}_{k|k}$ are no longer weighted as in Eq. (30), since they are set to be equally weighted. Therefore, using this sampling solution makes the SMF equivalent to a deterministic EnGMF. Third, solving Eq. (62) increases the computational complexity of the overall algorithm. If $D \cdot N \gg M$, the time complexity of evaluating the MCVMD and its gradient is $\mathcal{O}(D \cdot N \cdot M \cdot N_s^2)$ [35]. The complexity increases linearly with the number of optimizer iterations, which can vary widely with respect to the specific routine used. However, the main time complexity, $\mathcal{O}(D \cdot N \cdot M)$, can be reduced using other techniques, such as divide and conquer. These factors must be considered when comparing with the $\mathcal{O}(N^2)$ complexity of the SMF and DBSCAN.

Figure 5 shows a two-dimensional example of the optimal point sampling procedure discussed. The underlying distribution is a GM with four equally weighted components, each representing a petal of the distribution in the form of a clover. For each Gaussian component, 144 deterministic samples are obtained using scaled Fibonacci grids ($N = 4$, $D = 144$). These points, shown in gold in the figure, are the same standard points used for each component, but each set is shifted by its corresponding mean and covariance. The samples of each Gaussian component are joined to form the first DM approximation to the GM. Finally, the DM is optimally reduced to 50 equally weighted points, shown in purple ($M = 50$). For this example, since $M > N$, the reduced points are initialized using a Fibonacci grid with 50 points, centered at the mean of the GM and oriented with respect to its covariance.

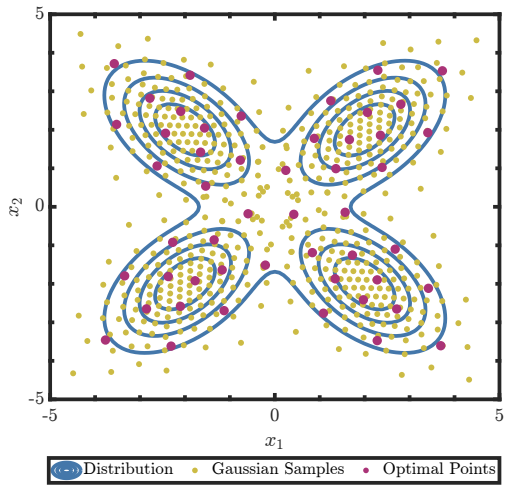


Fig. 5: Optimal point sampling of a GM. The distribution is shown with solid blue lines, the first DM approximation is shown with yellow points, and the optimally reduced approximation is shown with purple points.

Returning to the dual banana problem, Fig. 6 shows the points obtained by the optimal point sampling solution. In this case, 5 deterministic points are sampled from each Gaussian component ($N = 400$, $D = 5$). Since $M = N$, the reduced points are initialized to be the GSF-updated points. As the final reduced points are set to be equally weighted, $M_{\text{eff}} = 1$. From this figure, it can be seen that the optimal points are able to capture the entirety of the posterior distribution, showing a clear improvement over both the standard solution and the clustering approaches. It is important to note that these points are completely deterministic, meaning that no random number generator is used, unlike the various resampling strategies commonly used in particle filters.

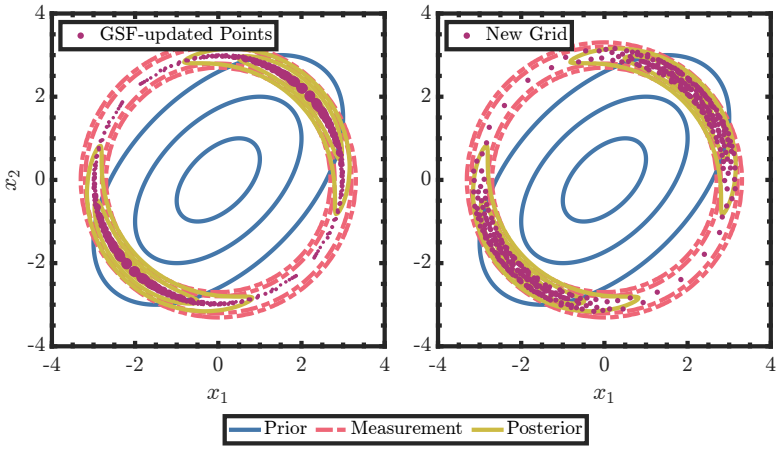


Fig. 6: GSF-updated points in purple (left) and new optimal grid in purple (right).

6 Numerical Examples

6.1 Univariate Nonstationary Growth Model

The univariate nonstationary growth model describes a nonlinear system governed by the following dynamics and measurement model:

$$x_{k+1} = \frac{x_k}{2} + \frac{25x_k}{1 + x_k^2} + 8\cos(1.2k), \tag{63}$$

$$y_k = \frac{x_k^2}{20} + \eta_k, \tag{64}$$

where η_k represents measurement noise. This problem has been widely used to analyze various nonlinear and particle filters [36–38]. Due to the nature of the measurement model, the posterior PDF is bimodal, making it an ideal candidate for studying the performance of the strategies presented. Instead of reporting estimation accuracy and

consistency, which is not particularly informative for this bimodal example [38], this problem is used to show how each strategy evolves at each time step.

For this problem, the three strategies presented in this work are used and evaluated:

1. **SMF:** This represents the standard SMF grid design, where the new grid is placed at the mean of the GSF-updated points and aligned with the updated covariance.
2. **SMF-DBS:** This strategy involves the clustering methodology, where the GSF-updated points are clustered using DBSCAN, and a Fibonacci grid is placed on top of each cluster.
3. **SMF-OPS:** This strategy uses the optimal point sampling approach, where deterministic points are first sampled from each Gaussian component represented by the GSF-updated points, and then optimally reduced by minimizing the MCVMD.

The three filtering strategies are presented in pseudocode in Algorithm 1. The only difference between the three algorithms is the generation of the new grid after the GSF update.

Algorithm 1 SMF/SMF-DBS/SMF-OPS

Input: $\boldsymbol{\mathcal{X}}_{k-1|k-1}$,$\boldsymbol{w}_{k-1|k-1}$,$Q_{k-1}$,$\boldsymbol{y}_k$, R_k,α
Output: $\boldsymbol{\mathcal{X}}_{k|k}, \boldsymbol{w}_{k|k}$

▷ *Propagate the points with the dynamics:* ◁
for $i = 1 : N$ **do**
  $\boldsymbol{\mathcal{X}}^{(i)}_{k|k-1} = f_{k-1}\left(\boldsymbol{\mathcal{X}}^{(i)}_{k-1|k-1}\right)$
  $w^{(i)}_{k|k-1} = w^{(i)}_{k-1|k-1}$
▷ *Calculate covariance for predictive distribution:* ◁
$B_{k|k-1} = \alpha\beta^2_{\text{sil}} \cdot \hat{P}_{k|k-1} + Q_{k-1}$
▷ *Perform GSF update:* ◁
Use Eqs. (13) to (20) to perform a GSF update: $\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}, \widetilde{\boldsymbol{w}}_{k|k}$
▷ *Generate the new grid* ◁
if SMF **then**
  Center a new grid at $\widetilde{\boldsymbol{x}}_{k|k}$ and align with $\widetilde{P}_{k|k}$ (Eqs. (21) and (22)): $\boldsymbol{\mathcal{X}}_{k|k}$
  Use Eq. (30) to obtain the new weights: $\boldsymbol{w}_{k|k}$
if SMF-DBS **then**
  Cluster the GSF-updated points $\widetilde{\boldsymbol{\mathcal{X}}}_{k|k} = {}^{(1)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}, \ldots, {}^{(L)}\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}$
  Center a new grid at ${}^{(i)}\widetilde{\boldsymbol{x}}_{k|k}$ and align with ${}^{(i)}\widetilde{P}_{k|k}$ (Eqs. (36) and (37))
  Join the collection of grids to obtain the new grid: $\boldsymbol{\mathcal{X}}_{k|k}$
  Use Eq. (30) to obtain the new weights: $\boldsymbol{w}_{k|k}$
if SMF-OPS **then**
  Generate first DM approximation (Eq. (60)): $\boldsymbol{\mathcal{Y}}_{k|k}, \boldsymbol{w}_{k|k}$
  Initialize the new points: $\boldsymbol{\mathcal{X}}$
  Reduce the approximation using Eq. (62) to obtain the new grid: $\boldsymbol{\mathcal{X}}_{k|k}$
  $w^{(j)}_{k|k} = 1/N$

To evaluate the performance of each strategy, all filters begin from an initial state $x_0 \sim \mathcal{N}(0, 1)$. The dynamics are propagated using additive white Gaussian noise with a scalar covariance matrix $Q = 10$, while the measurements are corrupted by additive white Gaussian noise with a scalar covariance matrix $R = 1$. The three strategies start with a rectangular one-dimensional grid, centered on the mean of the initial state expanding to $3\sigma_0$, with a total of 14 grid points ($N = 14$). The Fibonacci grids used in the SMF-DBS for each cluster are stretched to expand up to 6σ in each direction, and the DBSCAN parameters are set to `epsilon` $= 4.5$ and `minPts` $= 3$. The Fibonacci grids used in the SMF-OPS to sample each Gaussian component are composed of 5 points without applying any stretching, and $\boldsymbol{\mathcal{X}}$ is initialized to be the GSF-updated points, $\widetilde{\boldsymbol{\mathcal{X}}}_{k|k}$. The scaling factor for Silverman's rule of thumb is set to $\alpha = 0$ for all strategies, as the process noise is considerably high in this problem.

Figure 7 shows the evolution of the approximated posterior PDF, estimated using the GSF-updated points plotted in light blue. The new grid obtained with each strategy from the GSF points is shown with red stems, and the true state is marked with a black vertical line. Each grid is plotted according to its weights, with taller stems indicating higher weights. The left panels show the standard SMF, the center panels show the SMF-DBS, and the right panels show the SMF-OPS. The time index k increases downward.

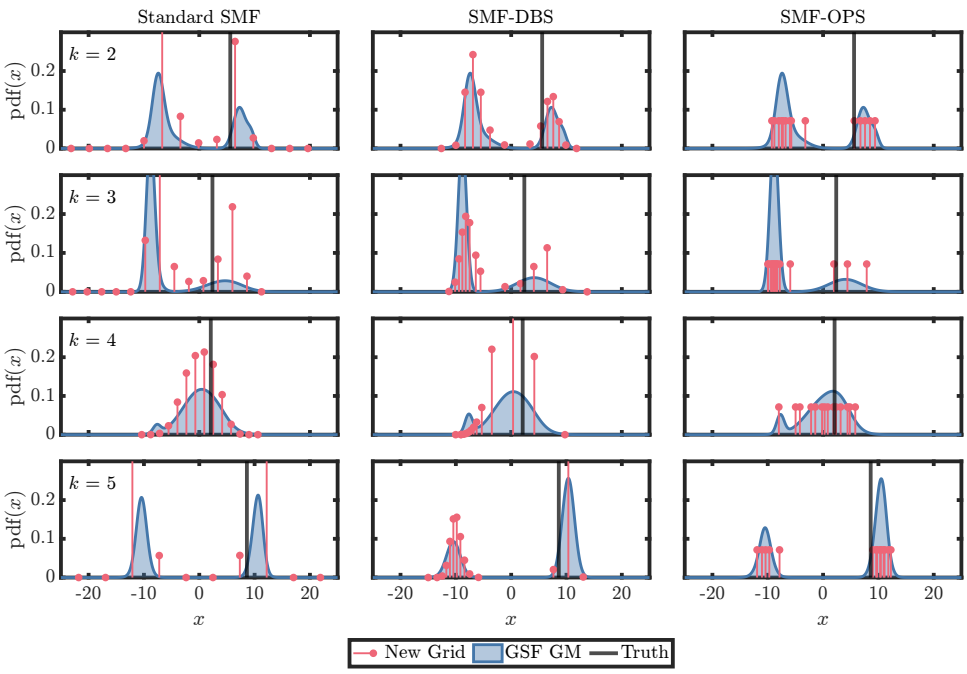


Fig. 7: Evolution of the posterior PDF, estimated via the GSF GM plotted in light blue. The new grid obtained with each strategy is shown with red stems. The standard SMF is shown on the left, the SMF-DBS in the center, and the SMF-OPS on the right.

In Fig. 7, the problems with the standard SMF grid placement can be seen. When the approximated posterior PDF is highly bimodal, the rectangular grid poorly approximates the underlying distribution, as observed at time steps $k = 2$, $k = 3$, and $k = 5$. In contrast, the SMF-DBS is able to identify two clusters and place grid points at appropriate locations in these bimodal time steps. However, the drawbacks of using a density-based clustering algorithm are noticeable. For example, at time step $k = 5$, the SMF-DBS places more points in the low-probability mode near $x = -10$ than in the higher-probability mode near $x = 10$. This behavior does not occur with the SMF-OPS, as the optimal sampling algorithm deterministically samples points directly from the GM described by the GSF-updated points. As a result, fewer points are sampled from the low-probability mode and more points are sampled from the high-probability mode.

6.2 Sequential Filtering with the Ikeda Map

To evaluate the performance of the proposed approaches in a sequential filtering problem, the Ikeda map is used, where

$$x_{k+1}^{(1)} = 1 + u\left(x_k^{(1)} \cos t_k - x_k^{(2)} \sin t_k\right), \tag{65}$$

$$x_{k+1}^{(2)} = u\left(x_k^{(1)} \sin t_k + x_k^{(2)} \cos t_k\right), \tag{66}$$

$$t_k = 0.4 - \frac{6}{1 + \left(x_k^{(1)}\right)^2 + \left(x_k^{(2)}\right)^2}. \tag{67}$$

This highly nonlinear function is designed to model light circulating within a nonlinear optical resonator [39, 40]. To induce a chaotic behavior and have a more difficult filtering problem, $u = 0.9$. Figure 8 shows 1,000 trajectories for different initial states propagated using $u = 0.9$. In addition, for this filtering example, the dynamics are propagated with additive white Gaussian process noise with covariance matrix $Q = 1 \times 10^{-2} I_{2\times 2}$ and a range measurement is assumed,

$$y_k = \sqrt{\left(x_k^{(1)}\right)^2 + \left(x_k^{(2)}\right)^2} + \eta_k, \tag{68}$$

where η_k is white Gaussian noise with scalar covariance matrix $R = 1$.

Each strategy is used to estimate the state in 1,000 different trajectories ($N_m = 1000$). The trajectories are propagated for 50 time steps, starting from an initial true state $\boldsymbol{x}_0 \sim \mathcal{N}(0_{2\times 1}, I_{2\times 2})$. All filters use grids with 25 points ($N = 25$), and start with a rectangular grid centered on the initial state estimate. The grids in the standard SMF are set to expand up to 3σ in each direction. The Fibonacci grids used in the SMF-DBS for each cluster are stretched to expand up to 6σ in each direction, `epsilon` $= 0.275$, and `minPts` $= 10$. The Fibonacci grids used in the SMF-OPS to sample each Gaussian component are composed of 5 points with no stretching applied ($D = 5$), and $\boldsymbol{\mathcal{X}}$ is initialized using the GSF-updated points. The scaling factor used to scale Silverman's rule of thumb is set to 0.3 for all strategies ($\alpha = 0.3$).

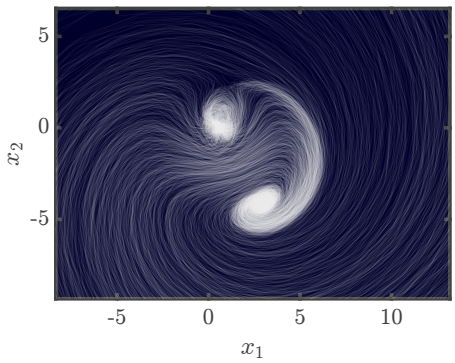


Fig. 8: Multiple trajectories for different initial states of the Ikeda map with $u = 0.9$.

To compare the accuracy of each strategy, the root mean square error (RMSE) is used. For this work, the RMSE is defined as follows

$$\mathrm{RMSE}(k) = \sum_{j=1}^{N_m} \frac{1}{N_m} \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} \left(x_{k,j}^{(i)} - \hat{x}_{k|k,j}^{(i)} \right)^2}, \tag{69}$$

where $\hat{x}_{k|k,j}^{(i)}$ is the estimated state and $x_{k,j}^{(i)}$ is the true state. To evaluate the consistency of each strategy, the scaled normalized estimation error squared (SNEES) is used,

$$\mathrm{SNEES}(k) = \frac{1}{N_s N_m} \sum_{j=1}^{N_m} \left(\boldsymbol{x}_{k,j} - \hat{\boldsymbol{x}}_{k|k,j} \right)^{\mathrm{T}} \left(P_{k|k,j} \right)^{-1} \left(\boldsymbol{x}_{k,j} - \hat{\boldsymbol{x}}_{k|k,j} \right), \tag{70}$$

where $P_{k|k,j}$ is the posterior estimated covariance. If the SNEES is close to one, the estimator is consistent. A SNEES lower than one suggests a conservative estimator, while a SNEES higher than one indicates an overconfident estimator [1]. For this section, any SNEES value greater than 1×10^3 was disregarded as a numerical instability.

Figure 9 shows the RMSE as a function of time for the three strategies presented. As seen in the figure, the standard SMF yields the highest RMSE, indicating the lowest accuracy in the estimates. The SMF-DBS increases accuracy compared to the standard SMF, approaching the SIR $\to \infty$ line. Regardless, by not considering the probability mass of the GSF-updated points, accuracy is lost as the new grids are not directly sampling the GM represented by the updated points. In this sense, since the SMF-OPS is optimally sampling the GM, it achieves the best accuracy among the three strategies, closely following the best achievable performance line converging to Bayesian inference.

In Fig. 10, the SNEES is shown as a function of time steps for the three strategies. In this case, it can be observed that the standard SMF has high SNEES values, suggesting overconfidence in its estimates. This is mainly due to the low number of grid points used in this problem setting. With an increasing number of grid points, the

standard SMF has been shown to produce an accurate and consistent estimator [4]. However, both the SMF-DBS and SMF-OPS outperform the standard SMF in low grid point settings. The SMF-DBS achieves a SNEES value closer to one compared to the standard SMF. The SMF-OPS obtains a SNEES very close to one at all time steps, indicating both an accurate and consistent filter. It is important to mention that the factor α can be increased to obtain better consistency with both the SMF and the SMF-DBS, but this can affect the accuracy of the filter [4].

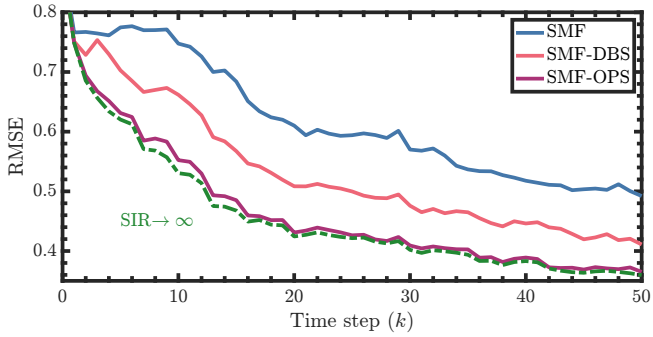


Fig. 9: RMSE as a function of time steps for the three strategies presented. The green dashed line shows the results obtained with a regularized particle filter with 10,000 particles and serves as a reference.

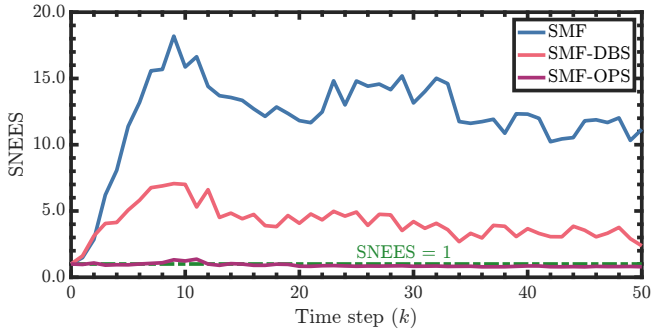


Fig. 10: SNEES as a function of time steps for the three strategies presented. The green dashed line shows an SNEES of one and serves as a reference.

From these results, it is important to mention that each strategy has a different level of implementation complexity. The grid design in the standard SMF is by far the simplest, as its implementation is straightforward and only requires the tuning two parameters, specifically the extent of the grid and the bandwidth parameter α. The SMF-DBS is more complex to tune, as the clustering algorithm is highly sensitive to the two tuning parameters, `epsilon` and `minPts`. The SMF-OPS is less complicated

in terms of tuning, as the only parameters that need adjustment are those in the optimization scheme used to minimize the MCVMD and reduce the set of deterministic points. These are additional factors that must be taken into account when comparing the three strategies.

6.3 Cislunar Orbit Determination

As seen from the previous two examples, the SMF-OPS is a more accurate and consistent estimator than the standard SMF and the SMF-DBS. In order to evaluate the performance of the SMF-OPS against other state-of-the-art filters used for aerospace applications, a cislunar orbit determination problem is considered. In this scenario, a spacecraft is assumed to orbit the Earth-Moon system. The dynamics of the spacecraft are expressed in the rotating synodic reference frame using the circular restricted three-body problem (CR3BP). These dynamics are derived by expressing the motion of the spacecraft relative to the barycenter of the Earth-Moon system. In this problem, the x axis is aligned from the center of the Earth to the center of the Moon, the z axis is aligned with the orbital plane of the Moon relative to the Earth, and the y axis completes the coordinate system [41].

To improve numerical stability, the variables are normalized to be unitless. In this specific setup, the location of the barycenter is defined by the mass ratio between the Earth and the Moon, such that

$$\mu = \frac{\mu_{☾}}{\mu_{☾} + \mu_{\oplus}}, \tag{71}$$

making the distance between Earth and the Moon, 1 LU $= 389\,703 \times 10^3$ meters, and 1 TU $= 382\,981$ seconds [42]. Under these conditions, the dynamics of the spacecraft, parametrized by the vector $\boldsymbol{x} = \begin{bmatrix} r^{(1)} & r^{(2)} & r^{(3)} & \dot{r}^{(1)} & \dot{r}^{(2)} & \dot{r}^{(3)} \end{bmatrix}^{\mathrm{T}}$, can be written as [41]

$$\ddot{r}^{(1)} = r^{(1)} + 2\dot{r}^{(2)} - \frac{(1-\mu)\left(r^{(1)} + \mu\right)}{r_{\oplus}^3} - \frac{\mu\left(r^{(1)} - 1 + \mu\right)}{r_{☾}^3}, \tag{72}$$

$$\ddot{r}^{(2)} = r^{(2)} - 2\dot{r}^{(1)} - \frac{(1-\mu)\, r^{(2)}}{r_{\oplus}^3} - \frac{\mu r^{(2)}}{r_{☾}^3}, \tag{73}$$

$$\ddot{r}^{(3)} = -\frac{(1-\mu)\, r^{(3)}}{r_{\oplus}^3} - \frac{\mu r^{(3)}}{r_{☾}^3}, \tag{74}$$

where $r^{(i)}$ represents the corresponding scaled Cartesian positions of the spacecraft relative to the barycenter, $\dot{r}_i$ represents its corresponding velocities, $r_{\oplus}$ is the distance from Earth to the spacecraft, and $r_{☾}$ is the distance from the Moon to the spacecraft. Both $r_{\oplus}$ and $r_{☾}$ are defined as

$$r_{\oplus} = \sqrt{\left(r^{(1)} + \mu\right)^2 + \left(r^{(2)}\right)^2 + \left(r^{(3)}\right)^2}, \tag{75}$$

$$r_{☾} = \sqrt{\left(r^{(1)} - 1 + \mu\right)^2 + \left(r^{(2)}\right)^2 + \left(r^{(3)}\right)^2}. \tag{76}$$

For this work, the spacecraft is propagated from an initial distribution, in normalized units, given by

$$\boldsymbol{x}_0 \sim \mathcal{N}\left([1.155\,682\,17 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^{\mathrm{T}}, 1 \times 10^{-6} I_{6\times 6}\right). \tag{77}$$

The mean of this initial distribution corresponds to the second Lagrange point (L2) in the Earth-Moon system [42]. This point is known to be unstable, meaning that the selected initial condition can lead to a bimodal distribution [43]. Based on the initial condition for $\boldsymbol{x}_0$, as sampled from the initial PDF, the trajectory can integrate into two different families of orbits, one orbiting the Moon and the other escaping the system. The two possible trajectories make this a challenging problem for particle-based filters, as some particles may follow one trajectory, while others may follow the second, causing the filter to diverge in the initial steps [43]. Figure 11 shows the two different families of orbits that result from the propagation of different initial conditions close to the L2 point.

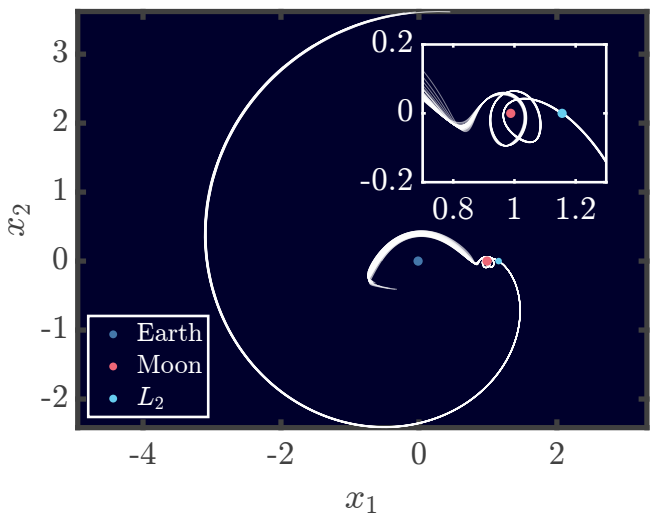


Fig. 11: Different orbits resulting from propagating different initial conditions close to the L2 point.

In addition, the dynamics are propagated with discrete process noise. Let F_t represent the flow of dynamics of the system, obtained by propagating the trajectory numerically. Therefore, process noise is added by

$$\boldsymbol{x}_{t+\Delta t} = F_t\left(\boldsymbol{x}_t\right) + \boldsymbol{q}_t, \tag{78}$$

where $\boldsymbol{q}_t$ represents white Gaussian process noise with covariance matrix [43]

$$Q_t = \sigma_Q^2 \begin{bmatrix} \frac{\Delta t^2}{2} I_{3\times3} & \Delta t\, I_{3\times3} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{\Delta t^2}{2} I_{3\times3} & \Delta t\, I_{3\times3} \end{bmatrix}, \tag{79}$$

with $\sigma_Q = 1 \times 10^{-6}$.

The measurement setup in this problem simulates observations from a single sensor on Earth's surface, providing range, azimuth, and elevation data, as given by

$$y_k^{(1)} = r_\oplus + \eta_k^{(1)}, \tag{80}$$

$$y_k^{(2)} = \arctan\left(\frac{r^{(2)}}{r^{(1)} + \mu}\right) + \eta_k^{(2)}, \tag{81}$$

$$y_k^{(3)} = \arcsin\left(\frac{r^{(3)}}{r_\oplus}\right) + \eta_k^{(3)}, \tag{82}$$

where $\boldsymbol{\eta}_k = \begin{bmatrix} \eta_k^{(1)} & \eta_k^{(2)} & \eta_k^{(3)} \end{bmatrix}^{\mathrm{T}}$ represents the white Gaussian measurement noise with covariance matrix [44],

$$R = \mathrm{diag}\left(\begin{bmatrix} 1^2\,\mathrm{m}^2 & 29^2\,\mathrm{arcsec}^2 & 29^2\,\mathrm{arcsec}^2 \end{bmatrix}\right). \tag{83}$$

Measurements are assumed to be sparse, with one set of measurements every 4 days.

In this problem, the SMF-OPS is tested and evaluated in 1,000 distinct trajectories, each propagated for 40 days. The standard SMF and the SMF-DBS are not included since both strategies diverge within the first few steps of the problem. The Fibonacci grids used in the SMF-OPS consist of 25 points for this example ($D = 25$), Silverman's rule of thumb is not scaled, $\alpha = 1$, and $\boldsymbol{\mathcal{X}}$ is initialized with a Fibonacci grid centered at the mean and covariance of the GSF-updated points[1]. Additionally, the UKF and the standard EnGMF with stochastic resampling are included as reference points to compare the proposed solution with well-established linear and nonlinear filters. Recall that the SMF-OPS is equivalent to a deterministic EnGMF. Therefore, both the SMF-OPS and the EnGMF are implemented with 100 points/ensemble members to be directly comparable ($N = 100$). Instead of using a rectangular grid, the SMF-OPS starts with a Fibonacci grid matching the initial distribution due to the high dimensionality of the problem.

For all filters, the state is propagated at a rate of 1 day ($\Delta t = 1$ day), with an update to the estimate once a measurement is obtained every 4 days. This is meant to simulate a tracking scenario where the state and its uncertainty are needed at a higher frequency than the measurements. Since the SMF-OPS points are propagated without process noise, the accumulated process noise covariance, calculated using the state transition matrix, is used when a measurement is obtained. For the EnGMF, process noise samples are added to the propagated ensemble members after each propagation. For the UKF, the sigma points are propagated, then re-Gaussianized.

[1]For this example, this proved to behave better numerically than initializing $\boldsymbol{\mathcal{X}}$ at the GSF-updated points.

To assess the accuracy and consistency of the filters tested, the standard deviation profiles [43] for the position and velocity of the spacecraft are used as shown in Fig. 12. In this figure, the dashed lines show the standard deviations calculated from the Monte Carlo (MC) samples at each time step, while the solid lines show the standard deviations calculated from the diagonal elements of the posterior covariance, where

$$\sigma_{\boldsymbol{r}}(k) = \frac{1}{N_m} \sum_{j=1}^{N_m} \sqrt{P_{k|k,j}^{(1,1)} + P_{k|k,j}^{(2,2)} + P_{k|k,j}^{(3,3)}}, \tag{84}$$

$$\sigma_{\boldsymbol{v}}(k) = \frac{1}{N_m} \sum_{j=1}^{N_m} \sqrt{P_{k|k,j}^{(4,4)} + P_{k|k,j}^{(5,5)} + P_{k|k,j}^{(6,6)}}. \tag{85}$$

For a filter to be consistent, the dashed and solid lines should align. If the dashed line is above the solid line, the filter is overconfident in its estimate, indicating potential divergence. If the dashed line is below the solid line, the filter is conservative [43]. For this problem set-up, both the UKF and EnGMF start to diverge after the second measurement is obtained, as the estimated error decreases while the true error begins to increase. This result is expected, given the low frequency of measurements and the small ensemble size used for the EnGMF. In contrast, the SMF-OPS, even with a small number of points, provides a consistent and accurate solution, as indicated by the alignment of the solid and dashed lines.

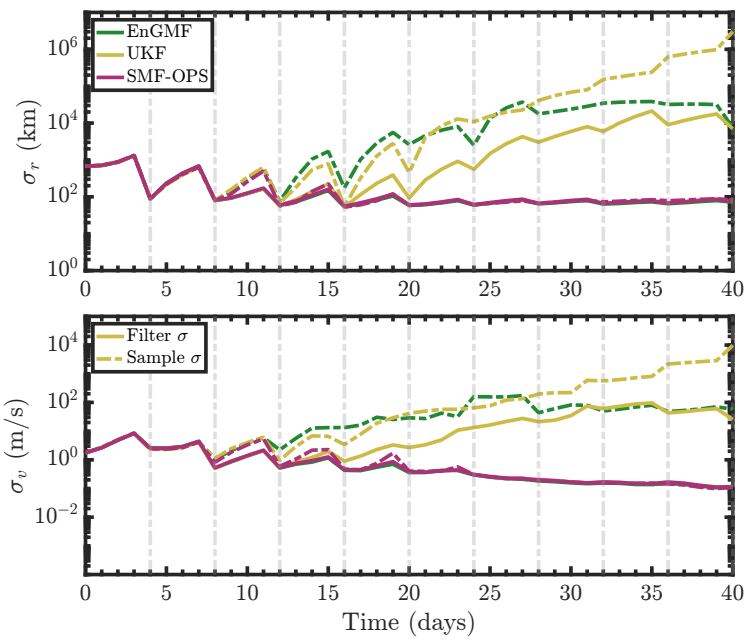
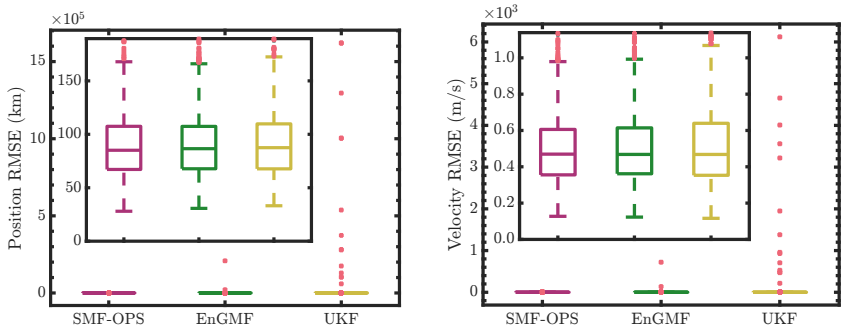


Fig. 12: Standard deviation profiles of position (top) and velocity (bottom) as a function of time. Measurement intervals are indicated by light gray vertical lines.

Figure 13 shows the box-plots for the position and velocity RMSE averaged over time for each MC trajectory. In Fig. 14 the effective grid size (Eq. (40)) for the SMF-OPS and the EnGMF calculated after the GSF update is shown, averaged over time for each MC trajectory as well. With these plots, it is easier to understand the behavior shown in Fig. 12, as the discrepancy between the dashed and solid lines for the EnGMF and UKF is due to outliers. For some trajectories, the EnGMF and UKF are unable to track the spacecraft, resulting in very high errors, as seen by the outliers plotted in red. Although the median accuracy difference between the three filters is comparable, the SMF-OPS obtains more robust results, with a lower number of significant outliers compared to the other two filters. In addition, the effective grid size after the GSF update of the SMF-OPS, is on average, higher than the EnGMF. Therefore, the deterministic and optimal sampling method used in the SMF-OPS can prove to be a more reliable strategy for nonlinear problems coupled with sparse measurements. It is important to note that measurement rejection strategies could be used to improve the results obtained with the EnGMF and the UKF. However, the SMF-OPS does not need to use other techniques to be a robust strategy in such cases.



(a) Position RMSE (b) Velocity RMSE

Fig. 13: Box-plots for position and velocity. Outliers are plotted in red.

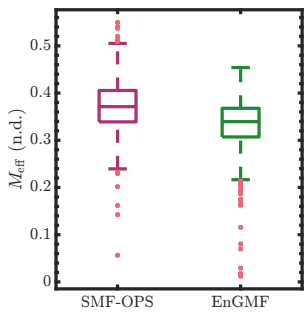


Fig. 14: Box plots for the effective grid size. Outliers are plotted in red.

7 Conclusions

In this work, two new techniques are presented for generating a new grid in the Silverman mass filter, a variation of the point mass filter. In the Silverman mass filter, a Gaussian sum filter update is performed before generating a new grid and continuing with the recursion. The standard approach places a new grid at the mean and covariance of these updated points. In contrast, the two new techniques leverage the additional information provided by the updated points to generate a better placement of the new points.

The first technique uses a density-based clustering approach in which the updated points are first clustered. Once the points have been clustered, the individual mean and covariance of each cluster are obtained. From the mean and covariance of each cluster, a new grid is placed, with the union of all grids considered as the new grid for the next iteration. This technique only considers the location of the updated points to generate the new grid, thus not using the full information provided by the Gaussian sum filter update.

The second technique takes advantage of the full information provided by the update. Since each updated point carries a mean, covariance, and probability mass, a modified distance measure is used to optimally generate samples from the Gaussian mixture. First, deterministic points are sampled from each updated point (representing a Gaussian distribution), and the union of all sampled points is then optimally reduced by minimizing a distance measure. This technique results in better points as it considers the full information from the Gaussian sum update.

The two techniques were evaluated in a static problem using a bimodal posterior distribution. Both techniques were shown to describe the posterior distribution better than the standard approach in the Silverman mass filter. Using the univariate nonstationary growth model, the evolution of each strategy was studied at each time step. This example showed some the limitations of the clustering-based approach, particularly due to its density-based algorithm. In addition, the new techniques were evaluated in a sequential filtering problem with the Ikeda map. The clustering-based approach was shown to outperform the standard technique, while the optimal sampling approach was shown to converge to Bayesian inference, producing the most accurate and consistent state estimate among the three approaches considered.

Finally, the optimal sampling technique was tested on a cislunar orbit determination example. The optimal sampling technique was shown to outperform the ensemble Gaussian mixture filter and the unscented Kalman filter in a scenario with sparse observations, including range and angles.

Acknowledgements. The first author would like to thank Dalton Durant for the insightful discussion on initializing the points for the optimal point sampling.

Declarations

Funding. This material is based on research sponsored by the Air Force Office of Scientific Research (AFOSR) under award numbers: FA9550-22-1-0419 and FA9550-23-1-0646, and by the J. Tinsley Oden Faculty Fellowship Research (Visitors) Program.

Conflict of Interest. The authors have no conflicts of interest to disclose.

Data availability statement. Not applicable.

Author Contribution. All authors contributed to the design of the research. F.G. wrote the main manuscript text. All authors reviewed the manuscript.

References

[1] Bar-Shalom, Y., Li, X.R., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation. John Wiley & Sons, Inc, New York (2001)

[2] Giraldo-Grueso, F., Popov, A.A., Zanetti, R.: Precision mars entry navigation with atmospheric density adaptation via neural networks. Journal of Aerospace Information Systems **21**(12) (2024) https://doi.org/10.2514/1.I011426

[3] Giraldo-Grueso, F., Popov, A.A., Zanetti, R.: Adaptive mars entry guidance with atmospheric density estimation. In: Proceedings of the AAS/AIAA Astrodynamics Specialist Conference, Broomfield, Colorado (2024)

[4] Giraldo-Grueso, F., Popov, A.A., Zanetti, R.: Gaussian mixture-based point mass filtering with applications to terrain relative navigation. IEEE Transactions on Aerospace and Electronic Systems, 1–15 (2025)

[5] Michaelson, K., Wang, F., Zanetti, R.: Terrain-relative navigation with neuro-inspired elevation encoding. In: 2023 IEEE/ION Position, Location and Navigation Symposium (PLANS) (2023)

[6] Bucy, R.S., Senne., K.D.: Digital synthesis of non-linear filters. Automatica **7**(3), 287–298 (1971)

[7] Šimandl, M., Královec, J., Söderström, T.: Advanced point-mass method for nonlinear state estimation. Automatica **42**(7), 1133–1145 (2006)

[8] Ånonsen, K.B., Hallingstad, O.: Terrain aided underwater navigation using point mass and particle filters. In: 2006 IEEE/ION Position, Location, And Navigation Symposium, pp. 1027–1035 (2006). https://doi.org/10.1109/PLANS.2006.1650705

[9] Ånonsen, K.B., Hagen, O.K.: An analysis of real-time terrain aided navigation results from a hugin auv. In: OCEANS 2010 MTS/IEEE SEATTLE, pp. 1–9 (2010). https://doi.org/10.1109/OCEANS.2010.5664076

[10] Bergman, N.: Recursive bayesian estimation: Navigation and tracking applications. PhD thesis, Linköping University (1999)

[11] Šimandl, M., Královec, J., Söderström, T.: Anticipative grid design in point-mass approach to nonlinear state estimation. IEEE Transactions on Automatic Control **47**(4), 699–702 (2002)

[12] Duník, J., Straka, O., Matoušek, J.: Conditional density driven grid design in point-mass filter. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 9180–9184 (2020)

[13] Choe, Y., Park, C.G.: Point-mass filtering with boundary flow and its application to terrain referenced navigation. IEEE Transactions on Aerospace and Electronic Systems **57**(6), 3600–3613 (2021)

[14] Giraldo-Grueso, F., Popov, A.A., Zanetti, R.: Gaussian mixture-based point mass filtering. In: Proceedings of the 2024 27th International Conference on Information Fusion (FUSION), Venice, Italy (2024)

[15] Yun, S., Zanetti, R., Jones, B.A.: Kernel-based ensemble gaussian mixture filtering for orbit determination with sparse data. Advances in Space Research **69**(12), 4179–4197 (2022)

[16] Popov, A.A., Zanetti, R.: An adaptive covariance parameterization technique for the ensemble gaussian mixture filter. arXiv preprint arXiv:2212.10323 (2022)

[17] Popov, A.A., Zanetti, R.: Ensemble gaussian mixture filtering with particle-localized covariances. In: Proceedings of the 2023 26th International Conference on Information Fusion (FUSION), Charleston, South Carolina (2023)

[18] Popov, A.A., Zanetti, R.: Ensemble-localized kernel density estimation with applications to the ensemble gaussian mixture filter. arXiv preprint arXiv:2308.14143 (2023)

[19] Giraldo-Grueso, F., Popov, A.A., Hanebeck, U.D., Zanetti, R.: Optimal grid point sampling for point mass filtering. In: Proceedings of the AAS/AIAA Spaceflight Mechanics Meeting, Kaua'i, Hawaii (2025)

[20] Michaelson, K., Popov, A.A., Zanetti, R., DeMars, K.: Particle flow with a continuous formulation of the nonlinear measurement update. In: Proceedings of the 2024 27th International Conference on Information Fusion (FUSION), Venice, Italy (2024)

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

[22] Alspach, D., Sorenson, H.: Nonlinear bayesian estimation using gaussian sum approximations. IEEE Transactions on Automatic Control **17**(4), 439–448 (1972)

[23] Popov, A.A., Zanetti, R.: The ensemble epanechnikov mixture filter. arXiv

preprint arXiv:2408.11164 (2024)

[24] Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)

[25] P. Janssen, N.V. J. S. Marron, Sarle, W.: Scale measures for bandwidth selection. Journal of Nonparametric Statistics **5**(4), 359–380 (1995)

[26] Durant, D., Popov, A.A., Zanetti, R.: Updating gaussian mixture weights using posterior estimates. IEEE Transactions on Aerospace and Electronic Systems, 1–15 (2025)

[27] Ŝimandl, M., Královec, J.: Multigrid design in point-mass approach to nonlinear state estimation. IFAC Proceedings Volumes **36**(16), 1693–1698 (2003)

[28] Královec, J., Ŝimandl, M.: Numerical solution of filtering problem with multimodal densities. IFAC Proceedings Volumes **38**(1), 242–247 (2005)

[29] Matoušek, J., Duník, J., Straka, O.: Density difference grid design in a point-mass filter. Energies **13**(16), 4080 (2020)

[30] Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, pp. 226–231 (1996)

[31] Lloyd, S.: Least squares quantization in pcm. IEEE Transactions on Information Theory **28**(2), 129–137 (1982)

[32] Frisch, D., Hanebeck, U.D.: Deterministic gaussian sampling with generalized fibonacci grids. In: Proceedings of the 2021 IEEE International Conference on Information Fusion (FUSION), Sun City, South Africa (2021)

[33] Hanebeck, U.D., Klumpp, V.: Localized cumulative distributions and a multivariate generalization of the cramér-von mises distance. In: 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 33–39 (2008)

[34] Gilitschenski, I., Steinbring, J., Hanebeck, U.D., Simandl, M.: Deterministic dirac mixture approximation of gaussian mixtures. In: 17th International Conference on Information Fusion (FUSION), pp. 1–7 (2014)

[35] Hanebeck, U.D.: Optimal reduction of multivariate dirac mixture densities. arXiv preprint arXiv:1411.4586 (2014) 1411.4586

[36] Pogorelsky, B., Michaelson, K., Zanetti, R.: Particle filter with lmmse importance sampling. In: Proceedings of the 2022 25th International Conference on Information Fusion (FUSION) (2022)

[37] Servadio, S.: Likelihood scouting via map inversion for a posterior-sampled particle filter. IEEE Transactions on Aerospace and Electronic Systems, 1–17 (2024)

[38] Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing **50**(2), 174–188 (2002)

[39] Ikeda, K.: Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. Optics Communications **30**(2), 257–261 (1979)

[40] Ikeda, K., Daido, H., Akimoto, O.: Optical turbulence: Chaotic behavior of transmitted light from a ring cavity. Phys. Rev. Lett. **45**, 709–712 (1980)

[41] Vallado, D.A.: Fundamentals of Astrodynamics and Applications. Microcosm Press/Springer, Hawthorne (2007)

[42] Three-Body Periodic Orbits. https://ssd.jpl.nasa.gov/tools/periodic_orbits.html. Accessed: 2025-03-24

[43] Servadio, S., Zanetti, R.: Differential algebra-based multiple gaussian particle filter for orbit determination. Journal of Optim. Theory Appl. **101**, 459–485 (2021)

[44] Koblick, D., Choi, J.: Cislunar orbit determination benefits of moon-based sensors. In: Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS) (2022)