ADAPTIVE BAYESIAN INFERENCE FOR SPACE OBJECT TRACKING

Kyle J. DeMars and William N. Fife

Texas A&M University, College Station, TX, USA, Email: {demars, wfife}@tamu.edu

ABSTRACT

The determination of the state (e.g., position and velocity) of an orbiting object using limited data (e.g., line-of-sight measurements from a ground-based sensor) is a challenging problem in light of the, in general, nonlinear dynamics and measurements models. This can be overcome by taking more data, or by developing and applying more sophisticated tracking algorithms. This work pursues the latter and develops an adaptive, approximately Bayesian, recursive filter for space object tracking. The method is demonstrated for a near-Earth angles-only orbit determination problem and is shown to produce improvements to the accuracy and precision of the orbit estimates as well as the statistical consistency of the solution.

Keywords: Bayesian estimation; nonlinear filtering; space domain awareness; short-arc tracking.

1. INTRODUCTION

Tracking algorithms provide a fundamental element of space domain awareness by producing up-to-date information on the state (e.g., position and velocity) of objects of interest, including operational satellites and trackable fragments of space debris. The information desired is typically an estimate of the state and an understanding of the uncertainty in this estimate. One way to obviate the need for advanced tracking algorithms is to attempt to overcome any lack of knowledge through the deployment and acquisition of an ever-increasing number of sensors to collect more data to process. Space domain awareness is–almost by definition–a realm in which it is impossible to dominate the problem through more information; instead, it becomes necessary to embrace uncertainty and advance the state of the art in an alternative manner. The need for advancements in tracking algorithms for functional, decommissioned, and defunct space objects is greatly exacerbated by the growing pressure being placed on the space domain by the presence of large satellite constellations and the growing interest in the cislunar orbital regime.

Approaches for tracking space objects are typically comprised of two phases: prediction and inference. The prediction phase leverages dynamics modeling to provide an estimated state and uncertainty prior to the incorporation of new measurement information acquired from ground-based or space-based sensors. The current state of the art focuses on prediction in the presence of nonlinear dynamics, and it has been shown that Gaussian mixture representations are better able to approximate the underlying probability distribution for space object tracking [6, 10, 23]. The inference phase ingests the prediction and new measurement observations to produce an improved estimate and uncertainty of the state of the object. The focus of the current work is to develop and analyze a new approach for approximate Bayesian inference for space object tracking that leverages Gaussian mixture representations of the uncertainty.

In this work, the Bayesian update is transformed from a single-point solution to multi-point and continuous solutions through a homotopic continuation of Bayes' rule via manipulation of the measurement likelihood function [4]. Ideas along these lines have been explored previously, including the progressive Bayes framework of [8]. In [5, 25], the Bayesian update is transformed via a linear-log homotopy and used to formulate alternatives to the particle filter that are known as particle flow filters. Recently, [17, 18] have introduced partitioning of the measurement likelihood to achieve multi-step and continuous solutions to approximate Bayesian inference. This work develops an adaptive approach to approximate Bayesian inference that is first formulated by partitioning the measurement likelihood via a generalization of the method presented in [17] and then extended to a continuous solution in a manner similar to that pursued in [18] and [4].

The resulting filtering approach is shown to be implementable as a recursion of partial updates that better approximates the true Bayesian posterior in the presence of nonlinear measurements. The developed approach is applied to a short-arc space object tracking problem in the near-Earth domain and analyzed through simulation studies. Results are presented that highlight the power of the method, especially the ability to obtain accurate, precise, and statistically consistent estimates after long periods of uncertainty propagation.

Proc. 9th European Conference on Space Debris, Bonn, Germany, 1–4 April 2025, published by the ESA Space Debris Office
Editors: S. Lemmens, T. Flohrer & F. Schmitz, (http://conference.sdo.esoc.esa.int, April 2025)

2. BAYESIAN INFERENCE

Consider the scenario in which newly acquired measurement data (e.g., line-of-sight measurements), represented by $\boldsymbol{z}$, are to be used to estimate the state of a system (e.g., the position and velocity of a space object), represented by $\boldsymbol{x}$. There exist a variety of ways in which this can be accomplished, such as batch (least-squares) estimators, Kalman filters, and Bayesian filters. In this work, Bayesian approaches are leveraged for performing statistical inference, which rely on Bayes' rule, i.e.,

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x}) \,, \tag{1}$$

where $p(\boldsymbol{x})$ represents the prior probability density function (pdf), $p(\boldsymbol{z}|\boldsymbol{x})$ represents the conditional likelihood, and the posterior pdf is given by $p(\boldsymbol{x}|\boldsymbol{z})$. Proportionality in Bayes' rule is resolved by requiring that the posterior pdf is a valid pdf in $\boldsymbol{x}$, i.e. that $\int p(\boldsymbol{x}|\boldsymbol{z})\mathrm{d}\boldsymbol{x} = 1$. The normalization process can be performed by computing the evidence, $p(\boldsymbol{z}) = \int p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, which allows Bayes' rule to also be expressed as

$$p(\boldsymbol{x}|\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{z}) \,. \tag{2}$$

Alternatively, Bayes' rule can be rearranged as

$$p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x}) \,, \tag{3}$$

which is the equivalence of two versions of the joint probability of the measurement and the state.

As noted by Jaynes, "[a]n acceptable inference procedure should have the property that it neither ignores any of the input information nor injects any false information" [12]. To wit, there should be an underlying conservation of information that is exhibited by a proper inference procedure. Equation (3) is useful for establishing an equivalence of information that forms the basis of the principle of information conservation that lies at the heart of this work. The right-hand side of Eq. (3) is the information that is input into Bayes' rule, and the left-hand side of Eq. (3) is the output information of Bayes' rule. As is evident from Eq. (3), these two sources of information should be equivalent.

To establish an information conservation principle, it is necessary to to quantify the amount of information in a variety of ways. In this work, the information-theoretic principles pioneered by Shannon are leveraged [21]. The Shannon entropy, given by [3]

$$H_S[q] = -\int q(\boldsymbol{y}) \log q(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \,,$$

provides a quantifiable measure of the amount of information contained in the arbitrary pdf $q(\boldsymbol{y})$, provided the integral exists. The Shannon entropy (also called the differential Shannon entropy) measures the expected amount of surprisal or disorder of a random variable. Random variables that are more certain have lower Shannon entropy than those that are less certain. To exemplify the Shannon entropy, let $\boldsymbol{y}$ be Gaussian-distributed with mean $\boldsymbol{m}_y$ and covariance $\boldsymbol{P}_{yy}$, such that the pdf of $\boldsymbol{y}$ is given by $q(\boldsymbol{y}) = p_g(\boldsymbol{y}; \boldsymbol{m}_y, \boldsymbol{P}_{yy})$. The Shannon entropy for the Gaussian is given by $H_S[q] = \frac{1}{2}\log|2\pi e\boldsymbol{P}_{yy}|$, from which it follows that $H_S[q] \to -\infty$ as $|\boldsymbol{P}_{yy}| \to 0$ and $H_S[q] \to \infty$ as $|\boldsymbol{P}_{yy}| \to \infty$. As the volume of the Gaussian, which is measured by the determinant of the covariance matrix, decreases (increases), the Shannon entropy decreases (increases).

The Kullback-Leibler (KL) divergence (also called the relative differential Shannon entropy), which is defined as [15, 3]

$$D_{KL}[q_1||q_2] = \int q_1(\boldsymbol{y}) \log(q_1(\boldsymbol{y})/q_2(\boldsymbol{y}))\mathrm{d}\boldsymbol{y} \,,$$

quantifies the information lost in representing one pdf, $q_1(\boldsymbol{y})$, by another pdf, $q_2(\boldsymbol{y})$. The KL divergence satisfies two key properties: 1) it is self-identifying ($D_{KL}[q||q] = 0$), and 2) it is non-negative ($D_{KL}[q_1||q_2] \geq 0$). Unfortunately, the KL divergence is not symmetric (in general, $D_{KL}[q_1||q_2] \neq D_{KL}[q_2||q_1]$), and it does not satisfy the triangle inequality. As such, the KL divergence does not satisfy the properties to be a metric; nevertheless, self-identification and non-negativity are powerful properties of the KL divergence. The KL divergence can also be expressed in terms of the Shannon entropy as

$$D_{KL}[q_1||q_2] = H_C[q_1||q_2] - H_S[q_1] \,, \tag{4}$$

where $H_C[q_1||q_2]$ is the Shannon cross-entropy that is given by

$$H_C[q_1||q_2] = -\int q_1(\boldsymbol{y}) \log q_2(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \,.$$

Comparing the forms of the cross-entropy and entropy, it is clear that $H_C[q||q] = H_S[q]$.

Consider an inference procedure that ingests input information from $p(\boldsymbol{x})$ and $p(\boldsymbol{z}|\boldsymbol{x})$ and outputs information into $p(\boldsymbol{z})$ and $\tilde{p}(\boldsymbol{x}|\boldsymbol{z})$, where $\tilde{p}(\boldsymbol{x}|\boldsymbol{z})$ is not necessarily the result of Bayes' rule. Let the information contained in each of these sources be measured in the sense of Shannon with respect to some arbitrary pdf, $q(\boldsymbol{x})$, such that the input information is represented using cross-entropy as[1].

$$I_0 = H_C[q(\boldsymbol{x})||p(\boldsymbol{x})] + H_C[q(\boldsymbol{x})||p(\boldsymbol{z}|\boldsymbol{x})] \,;$$

similarly, the output information is represented using cross-entropy as

$$I_f = H_C[q(\boldsymbol{x})||p(\boldsymbol{z})] + H_C[q(\boldsymbol{x})||\tilde{p}(\boldsymbol{x}|\boldsymbol{z})] \,.$$

It is worth noting that the lack of dependence of $p(\boldsymbol{z})$ on $\boldsymbol{x}$ means that $H_C[q(\boldsymbol{x})||p(\boldsymbol{z})] \equiv -\log p(\boldsymbol{z})$, which is the negative log-evidence. Let the difference in information between the input and output be defined as $\Delta I \triangleq I_f - I_0$; then, by the properties of logarithms,

$$\Delta I = H_C\left[q(\boldsymbol{x})\middle|\middle|\frac{\tilde{p}(\boldsymbol{x}|\boldsymbol{z})}{p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{z})}\right] \,,$$

[1] Note that this is the opposite sign convention as the one used in [26].

which, from the definition of Bayes' rule in Eq. (2), is equivalent to

$$\Delta I = H_C[q(\boldsymbol{x})||\tilde{p}(\boldsymbol{x}|\boldsymbol{z})] - H_C[q(\boldsymbol{x})||p(\boldsymbol{x}|\boldsymbol{z})] \,.$$

At this point, it is instructive to choose a particular pdf against which the Shannon information is measured; to that end, let $q(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{z})$, such that

$$\Delta I = H_C[p(\boldsymbol{x}|\boldsymbol{z})||\tilde{p}(\boldsymbol{x}|\boldsymbol{z})] - H_S[p(\boldsymbol{x}|\boldsymbol{z})] \,,$$

or, from Eq. (4),

$$\Delta I = D_{KL}[p(\boldsymbol{x}|\boldsymbol{z})||\tilde{p}(\boldsymbol{x}|\boldsymbol{z})] \,, \tag{5}$$

where it is helpful to recall that $p(\boldsymbol{x}|\boldsymbol{z})$ represents the (exact) Bayesian posterior and $\tilde{p}(\boldsymbol{x}|\boldsymbol{z})$ represents an arbitrary posterior that may or may not be the result of an approximate solution to Bayes' rule.

Equation (5) states that the change in information from the inference procedure can be expressed as the KL divergence between the true Bayesian posterior and the (potentially) arbitrary posterior. By the non-negativity and self-identification properties of the KL divergence, the minimum value attainable for the change of information is zero, which is achieved if and only if $\tilde{p}(\boldsymbol{x}|\boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})$. Put simply, if it is possible to exactly produce the Bayesian posterior, then there is no loss of information. That is, Bayes' rule does not ignore any of the input information or inject any extraneous information into the posterior pdf.

3. DISCRETE PARAMETER FLOW

Except in special cases, such as linear measurements and Gaussian distributions, implementations of Bayes' rule require approximations to facilitate solutions. For instance, when presented with nonlinear measurements of the form

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{v} \,, \tag{6}$$

approximations have to be made to arrive at a tractable posterior. In this model, $\boldsymbol{x}$ is the state, which is distributed according to $p(\boldsymbol{x})$, $\boldsymbol{v}$ is additive measurement noise that is uncorrelated with $\boldsymbol{x}$ and distributed according to $p(\boldsymbol{v})$. One common approach is to linearize (in some manner) the measurement and produce an approximation to the true Bayesian posterior. If the prior uncertainty, represented by $p(\boldsymbol{x})$, is "small enough," reasonable results can be obtained via variations on the Kalman filter [14] and the Gaussian mixture filter [22]. This is, however, not always the case. As such, it is conducive to partition the update and perform a sequence of updates that progressively incorporate the measurement information into distribution of the state.

Let a sequence of points through "pseudotime" be given by $s_k \in [0, 1]$, such that $0 = s_1 < s_2 < \cdots < s_{M+1} = 1$. Let the widths of the intervals of pseudotime be denoted by $\Delta s_k = s_{k+1} - s_k$, such that $\sum \Delta s_k = 1$. Note that, by the strict monotonicity of s_k, $\Delta s_k > 0$. Since the sum of the interval widths is equal to one, the conditional likelihood is equivalently expressed as

$$p(\boldsymbol{z}|\boldsymbol{x}) = \prod_{k=1}^{M} p^{\Delta s_k}(\boldsymbol{z}|\boldsymbol{x}) \,.$$

This partitioning process effectively subdivides the measurement into M pieces that can be folded into the posterior one by one. In doing so, the effects of nonlinearity can be reduced.

In this work, Gaussian mixture (GM) models are leveraged as representations of the state pdf; as such, the prior pdf is given by an L_x-component mixture of the form

$$p(\boldsymbol{x}) = \sum_{\ell=1}^{L_x} w_x^{(\ell)-} p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)-}, \boldsymbol{P}_{xx}^{(\ell)-}) \,, \tag{7}$$

where $w_x^{(\ell)-}$, $\boldsymbol{m}_x^{(\ell)-}$, and $\boldsymbol{P}_{xx}^{(\ell)-}$ represent the prior weights, means, and covariances of each component within the mixture. The weights are constrained to be non-negative, and the covariance matrices are required to be symmetric and positive definite. Similarly, let the likelihood be given by the Gaussian pdf

$$p(\boldsymbol{z}|\boldsymbol{x}) = p_g(\boldsymbol{z}; \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{P}_{vv}) \,, \tag{8}$$

which is the probabilistic state space model representation of the nonlinear measurement model of Eq. (6), in which the measurement noise is taken to be zero-mean, white, and Gaussian with covariance $\boldsymbol{P}_{vv}$. Substituting the Gaussian likelihood into the partitioning equation and simplifying, it follows that

$$\begin{aligned} p(\boldsymbol{z}|\boldsymbol{x}) = \left|2\pi\boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \prod_{k=1}^{M} \left|2\pi\boldsymbol{P}_{vv}/\Delta s_k\right|^{\frac{1}{2}} \\ \times p_g(\boldsymbol{z}; \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{P}_{vv}/\Delta s_k) \,. \end{aligned} \tag{9}$$

It is worth noting that, since $\Delta s_k \leq 1$, $(\boldsymbol{P}_{vv}/\Delta s_k) \geq \boldsymbol{P}_{vv}$; furthermore, since $\Delta s_k > 0$, there are no issues with singularity of $(\boldsymbol{P}_{vv}/\Delta s_k)$ so long as $\boldsymbol{P}_{vv}$ is nonsinuglar.

The partitioned Gaussian likelihood of Eq. (9) and GM prior of Eq. (7) can now be substituted into Bayes' rule (Eq. (1)) to give

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left|2\pi\boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \left[\prod_{k=1}^{M} \left|2\pi\boldsymbol{P}_{vv}/\Delta s_k\right|^{\frac{1}{2}}\right] \\ \times \left[\prod_{k=1}^{M} p_g(\boldsymbol{z}; \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{P}_{vv}/\Delta s_k)\right] p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)-}, \boldsymbol{P}_{xx}^{(\ell)-}) \,, \end{aligned}$$

which is a weighted sum, where each element in the sum is formulated as a series of products of two Gaussian pdfs. The product of the prior and the modified likelihood for $k = 1$ produces a pseudoposterior that then becomes the prior for the $k = 2$ elements of the modified likelihood, and so on. This process is repeated until all M terms of the likelihood are applied to produce the approximate Bayesian posterior. Once the process of folding in each partition of the likelihood is completed, the scaling terms that arise from the process are used to define posterior weights. The result is that the posterior is a GM of the form

$$p(\boldsymbol{x}|\boldsymbol{z}) = \sum_{\ell=1}^{L_x} w_x^{(\ell)+} p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)+}, \boldsymbol{P}_{xx}^{(\ell)+}) \, ,$$

where the weights, means, and covariances are determined by the iterative relationships

$$w_x^{(\ell,i)} = k^{(\ell,i)} w_x^{(\ell,i-1)} \, / \, \textstyle\sum_{\ell'=1}^{L_x} k^{(\ell',i)} w_x^{(\ell',i-1)} \tag{10a}$$

$$\boldsymbol{m}_x^{(\ell,i)} = \boldsymbol{m}_x^{(\ell,i-1)} + \boldsymbol{K}^{(\ell,i)}(\boldsymbol{z} - \boldsymbol{m}_h^{(\ell,i-1)}) \tag{10b}$$

$$\begin{aligned} \boldsymbol{P}_{xx}^{(\ell,i)} &= \boldsymbol{P}_{xx}^{(\ell,i-1)} - \boldsymbol{K}^{(\ell,i)} \boldsymbol{P}_{hh}^{(\ell,i-1)} (\boldsymbol{K}^{(\ell,i)})^T \\ &\quad - \boldsymbol{K}^{(\ell,i)} (\boldsymbol{P}_{vv}/\Delta s_i)(\boldsymbol{K}^{(\ell,i)})^T \, , \end{aligned} \tag{10c}$$

which are applied for $1 \leq i \leq M$. This iterative procedure is initialized with $w_x^{(\ell,0)} = w_x^{(\ell)-}$, $\boldsymbol{m}_x^{(\ell,0)} = \boldsymbol{m}_x^{(\ell)-}$ and $\boldsymbol{P}_{xx}^{(\ell,0)} = \boldsymbol{P}_{xx}^{(\ell)-}$. When the last iteration is completed, the output is $w_x^{(\ell)+} = w_x^{(\ell,M)}$, $\boldsymbol{m}_x^{(\ell)+} = \boldsymbol{m}_x^{(\ell,M)}$ and $\boldsymbol{P}_{xx}^{(\ell)+} = \boldsymbol{P}_{xx}^{(\ell,M)}$. The weight and state gains that appear in Eqs. (10) are given by

$$k^{(\ell,i)} = p_g(\boldsymbol{z}; \boldsymbol{m}_h^{(\ell,i-1)}, \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv}/\Delta s_i) \tag{11a}$$

$$\boldsymbol{K}^{(\ell,i)} = \boldsymbol{P}_{xh}^{(\ell,i-1)} (\boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv}/\Delta s_i)^{-1} \, , \tag{11b}$$

and the mean, cross-covariance (with the state) and covariance of the nonlinear function $\boldsymbol{h}(\boldsymbol{x})$ that are required to complete each iteration are defined in terms of expected values as

$$\boldsymbol{m}_h^{(\ell,\cdot)} = \mathbb{E}\{\boldsymbol{h}(\boldsymbol{x})\} \tag{12a}$$

$$\boldsymbol{P}_{xh}^{(\ell,\cdot)} = \mathbb{E}\{(\boldsymbol{x} - \boldsymbol{m}_x^{(\ell,\cdot)})(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell,\cdot)})^T\} \tag{12b}$$

$$\boldsymbol{P}_{hh}^{(\ell,\cdot)} = \mathbb{E}\{(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell,\cdot)})(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell,\cdot)})^T\} \, , \tag{12c}$$

where the expectations are taken with respect to $p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell,\cdot)}, \boldsymbol{P}_{xx}^{(\ell,\cdot)})$. The method defined by the general application of Eqs. (10)–(12) is referred to as discrete parameter flow (DPF). It is termed “parameter flow” due to the governing equations that dictate the motion of the parameters that define the underlying pdf through pseudotime. This motion is similar to the motion of particles in the particle flow approach [5].

The results of Eqs. (10)–(12) are obtained by substituting Eqs. (7) and (9) into Eq. (1), and then applying a generalized form of the Ho-Lee equation [9] for products of Gaussians in conjunction with statistical linearization [7, 16] to handle the nonlinearity of the measurement model. A proof of Eqs. (10) is provided in Appendix A.

3.1. Linearized Formulation

The expected values of Eqs. (12) that are required for the discrete parameter flow approach can be approximated via analytical linearization (i.e., first-order Taylor series linearization), such that

$$\begin{aligned} \boldsymbol{m}_h^{(\ell,i-1)} &\approx \boldsymbol{h}(\boldsymbol{m}_x^{(\ell,i-1)}) \\ \boldsymbol{P}_{xh}^{(\ell,i-1)} &\approx \boldsymbol{P}_{xx}^{(\ell,i-1)} (\boldsymbol{H}_x^{(\ell,i-1)})^T \\ \boldsymbol{P}_{hh}^{(\ell,i-1)} &\approx \boldsymbol{H}_x^{(\ell,i-1)} \boldsymbol{P}_{xx}^{(\ell,i-1)} (\boldsymbol{H}_x^{(\ell,i-1)})^T \end{aligned}$$

where $\boldsymbol{H}_x^{(\ell,i-1)}$ is the Jacobian of $\boldsymbol{h}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, evaluated at $\boldsymbol{m}_x^{(\ell,i-1)}$. Using these approximations in conjunction with the discrete parameter flow approach for GM densities gives rise to the discrete parameter flow extended Gaussian mixture filter (DPF-EGMF).

3.2. Quadrature Formulation

Consider a set of N_q weights and points that comprise a quadrature rule for expectations with respect to $p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell,i-1)}, \boldsymbol{P}_{xx}^{(\ell,i-1)})$ of the form $\{w_m^{(\ell,j)}, w_c^{(\ell,j)}, \boldsymbol{\mathcal{X}}^{(\ell,i-1,j)}\}_{j=0}^{N_q-1}$ that are constructed so that

$$\begin{aligned} \boldsymbol{m}_x^{(\ell,i-1)} &= \sum_{j=0}^{N_q-1} w_m^{(\ell,j)} \boldsymbol{\mathcal{X}}^{(\ell,i-1,j)} \\ \boldsymbol{P}_{xx}^{(\ell,i-1)} &= \sum_{j=0}^{N_q-1} w_c^{(\ell,j)} (\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)} - \boldsymbol{m}_x^{(\ell,i-1)}) \\ &\qquad \times (\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)} - \boldsymbol{m}_x^{(\ell,i-1)})^T \, . \end{aligned}$$

Note that the quadrature weights carry no dependence on i; that is, the quadrature scheme is taken to have constant weights through the update, which is common. The quadrature-based approximation of the expected values of Eqs. (12) that are required for the discrete parameter flow approach are then given by

$$\begin{aligned} \boldsymbol{m}_h^{(\ell,i-1)} &\approx \sum_{j=0}^{N_q-1} w_m^{(\ell,j)} \boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)}) \\ \boldsymbol{P}_{xh}^{(\ell,i-1)} &\approx \sum_{j=0}^{N_q-1} w_c^{(\ell,j)} (\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)} - \boldsymbol{m}_x^{(\ell,i-1)}) \\ &\qquad \times (\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)}) - \boldsymbol{m}_h^{(\ell,i-1)})^T \\ \boldsymbol{P}_{hh}^{(\ell,i-1)} &\approx \sum_{j=0}^{N_q-1} w_c^{(\ell,j)} (\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)}) - \boldsymbol{m}_h^{(\ell,i-1)}) \\ &\qquad \times (\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,i-1,j)}) - \boldsymbol{m}_h^{(\ell,i-1)})^T \, . \end{aligned}$$

Using these approximations in conjunction with the discrete parameter flow approach for GM densities gives rise to the discrete parameter flow quadrature Gaussian mixture filter (DPF-QGMF).

If a certain quadrature rule is employed, such as Gauss-Hermite quadrature [2, 11] or the unscented transform [13, 20, 24], then specific construction rules for the weights and points can be given. Using the unscented transform (in any of its variants) for the quadrature rule is referred to as the discrete parameter flow unscented Gaussian mixture filter (DPF-UGMF).

3.3. Numerical Consideration

The iterative update procedure indicated by Eqs. (10) and supported by Eqs. (11) and Eqs. (12) can be interpreted as M applications of a form of the Gaussian mixture filter (GMF), such as the extended GMF (EGMF) or unscented GMF (UGMF), with the substitution of $(\boldsymbol{P}_{vv}/\Delta s_i)$ in place of $\boldsymbol{P}_{vv}$. This property is highly attractive, as it allows existing implementations of the EGMF, UGMF, or other versions of the GMF to be used with minimal modification. It is, however, imperative to note that the use of $(\boldsymbol{P}_{vv}/\Delta s_i)$ has some limitations. Notably, since $0 < s_i \leq 1$, it has previously been noted that $(\boldsymbol{P}_{vv}/\Delta s_i) > \boldsymbol{P}_{vv}$, which indicates that singularity is not impacted by Δs_i. That being said, $\lim_{\Delta s_i \to 0} |\boldsymbol{P}_{vv}/\Delta s_i| \to \infty$. Thus, too small of a step through pseudotime can lead to numerical issues in the modified measurement noise covariance matrix that is used. As this happens, the resulting iterative update procedure of Eqs. (10)–(12) begins to break down.

4. CONTINUOUS PARAMETER FLOW

The iterative procedure can be applied for any number of intervals, M, that define the partitioning of the likelihood function. These intervals, which are characterized by their widths, Δs_i, may be uniformly or non-uniformly spaced, as long as $\Delta s_i > 0 \;\; \forall\, i$ and $\sum_{i=1}^{M} \Delta s_i = 1$.

An alternative approach is to seek governing differential equations for the means, covariances, and weights. This can be done for the weights, means, and covariances by taking the limit of ratio of the change in each parameter to the interval width as the interval width goes to zero, i.e.,

$$\frac{\mathrm{d}(\star)}{\mathrm{d}s} = \lim_{\Delta s_i \to 0} \frac{(\star)^{(\ell,i)} - (\star)^{(\ell,i-1)}}{\Delta s_i} \,. \tag{13}$$

Carrying out this procedure, it can be shown that the weights, means, and covariances obey the ordinary differential equations

$$\begin{aligned}\frac{\mathrm{d}w_x^{(\ell)}(s)}{\mathrm{d}s} &= -\frac{1}{2} w_x^{(\ell)}(s) \Bigg[\mathrm{tr}\big\{\boldsymbol{P}_{vv}^{-1}\boldsymbol{P}_{hh}^{(\ell)}(s)\big\} \qquad (14\mathrm{a}) \\ &\quad + (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell)}(s))^T \boldsymbol{P}_{vv}^{-1} (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell)}(s)) \\ &\quad - \sum_{\ell'=1}^{L_x} w_x^{(\ell')}(s) \Big\{ \mathrm{tr}\big\{\boldsymbol{P}_{vv}^{-1}\boldsymbol{P}_{hh}^{(\ell')}(s)\big\} \\ &\qquad + (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell')}(s))^T \boldsymbol{P}_{vv}^{-1} (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell')}(s)) \Big\} \Bigg]\end{aligned}$$

$$\frac{\mathrm{d}\boldsymbol{m}_x^{(\ell)}(s)}{\mathrm{d}s} = \boldsymbol{P}_{xh}^{(\ell)}(s) \boldsymbol{P}_{vv}^{-1} (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell)}(s)) \tag{14b}$$

$$\frac{\mathrm{d}\boldsymbol{P}_{xx}^{(\ell)}(s)}{\mathrm{d}s} = -\boldsymbol{P}_{xh}^{(\ell)}(s) \boldsymbol{P}_{vv}^{-1} (\boldsymbol{P}_{xh}^{(\ell)}(s))^T \,. \tag{14c}$$

The results of Eqs. (14b) and (14c) are relatively straightforward. Obtaining the result of Eq. (14a) requires more care and application of de l'Hôpital's rule. A proof of Eqs. (14) is provided in Appendix B.

The mean, cross-covariance (with the state) and covariance of the nonlinear function $\boldsymbol{h}(\boldsymbol{x})$ that are required by Eqs. (14) are the continuous-time versions of Eqs. (12), i.e.,

$$\boldsymbol{m}_h^{(\ell)}(s) = \mathbb{E}\big\{\boldsymbol{h}(\boldsymbol{x})\big\} \tag{15a}$$

$$\boldsymbol{P}_{xh}^{(\ell)}(s) = \mathbb{E}\big\{(\boldsymbol{x} - \boldsymbol{m}_x^{(\ell)}(s))(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell)}(s))^T\big\} \tag{15b}$$

$$\boldsymbol{P}_{hh}^{(\ell)}(s) = \mathbb{E}\big\{(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell)}(s))(\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{m}_h^{(\ell)}(s))^T\big\} \,, \tag{15c}$$

where the expectations are taken with respect to $p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)}(s), \boldsymbol{P}_{xx}^{(\ell)}(s))$. The method defined by the general application of Eqs. (14) and (15) is referred to as continuous parameter flow (CPF).

4.1. Linearized Implementation

The expected values required for the continuous parameter flow approach can be approximated via analytical linearization, such that

$$\begin{aligned}\boldsymbol{m}_h^{(\ell)}(s) &\approx \boldsymbol{h}(\boldsymbol{m}_x^{(\ell)}(s)) \\ \boldsymbol{P}_{xh}^{(\ell)}(s) &\approx \boldsymbol{P}_{xx}^{(\ell)}(s)(\boldsymbol{H}_x^{(\ell)}(s))^T \\ \boldsymbol{P}_{hh}^{(\ell)}(s) &\approx \boldsymbol{H}_x^{(\ell)}(s)\boldsymbol{P}_{xx}^{(\ell)}(s)(\boldsymbol{H}_x^{(\ell)}(s))^T \,,\end{aligned}$$

where $\boldsymbol{H}_x^{(\ell)}(s)$ is the Jacobian of $\boldsymbol{h}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, evaluated at $\boldsymbol{m}_x^{(\ell)}(s)$. Using these approximations in conjunction with the discrete parameter flow approach for GM densities gives rise to the continuous parameter flow extended Gaussian mixture filter (CPF-EGMF).

4.2. Quadrature Implementation

The expected values can also be approximated by a quadrature rule, such as

$$\boldsymbol{m}_h^{(\ell)}(s) \approx \sum_{j=0}^{N_q-1} w_m^{(\ell,j)} \boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,j)}(s))$$

$$\boldsymbol{P}_{xh}^{(\ell)}(s) \approx \sum_{j=0}^{N_q-1} w_c^{(\ell,j)}(\boldsymbol{\mathcal{X}}^{(\ell,j)}(s) - \boldsymbol{m}_x^{(\ell)}(s)) \times (\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,j)}(s)) - \boldsymbol{m}_h^{(\ell)}(s))^T$$

$$\boldsymbol{P}_{hh}^{(\ell)}(s) \approx \sum_{j=0}^{N_q-1} w_c^{(\ell,j)}(\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,j)}(s)) - \boldsymbol{m}_h^{(\ell)}(s)) \times (\boldsymbol{h}(\boldsymbol{\mathcal{X}}^{(\ell,j)}(s)) - \boldsymbol{m}_h^{(\ell)}(s))^T ,$$

where $\{w_m^{(\ell,j)}, w_c^{(\ell,j)}, \boldsymbol{\mathcal{X}}^{(\ell,j)}(s)\}_{j=0}^{N_q-1}$ defines the quadrature rule in terms of mean weights, covariance weights, and quadrature points for a given set of N_q points. For general quadrature rules, this leads to the continuous parameter flow quadrature Gaussian mixture filter (CPF-QGMF); for the specific case of using the unscented transform, the result is the continuous parameter flow unscented Gaussian mixture filter (CPF-UGMF).

4.3. Weight Normalization

In both the discrete and continuous forms of the GM parameter flow, the normalized weights, $w_x^{(\ell,i)}$ or $w_x^{(\ell)}(s)$, are used. This means that at any point in pseudotime, the resulting GM representation of the pdf is a proper pdf; the weights sum to one, so the pdf integrates to one. The advantage of not normalizing the weights until it is required is that the computational burden is greatly lessened.

It is possible to also work with non-normalized weights, $\tilde{w}_x^{(\ell,i)}$ for discrete parameter flow and $\tilde{w}_x^{(\ell)}(s)$ for continuous parameter flow. These weights do not necessarily sum to one, but brute-force normalization can be used on non-normalized weights to ensure a valid pdf.

For discrete parameter flow, the unnormalized weights are found via

$$\tilde{w}_x^{(\ell,i)} = k^{(\ell,i)} \tilde{w}_x^{(\ell,i-1)} ,$$

which is applied for $1 \leq i \leq M$ starting from $\tilde{w}_x^{(\ell,0)} = w_x^{(\ell)-}$ and ending with $\tilde{w}_x^{(\ell)+} = \tilde{w}_x^{(\ell,M)}$. After the last iteration is complete, the weights are normalized such that $w_x^{(\ell)+} = \tilde{w}_x^{(\ell)+} / \sum_{\ell'=1}^{L_x} \tilde{w}_x^{(\ell')+}$.

The equivalent implementation for continuous parameter flow is

$$\frac{\mathrm{d}\tilde{w}_x^{(\ell)}(s)}{\mathrm{d}s} = -\frac{1}{2}\tilde{w}_x^{(\ell)}(s)\Big[\mathrm{tr}\Big\{\boldsymbol{P}_{vv}^{-1}\boldsymbol{P}_{hh}^{(\ell)}(s)\Big\} + (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell)}(s))^T \boldsymbol{P}_{vv}^{-1}(\boldsymbol{z} - \boldsymbol{m}_h^{(\ell)}(s))\Big] ,$$

which is initialized with $\tilde{w}_x^{(\ell)}(s = 0) = w_x^{(\ell)-}$. After integrating from $s = 0$ to $s = 1$, the result is $\tilde{w}_x^{(\ell)+} = \tilde{w}_x^{(\ell)}(s = 1)$. These unnormalized weights are then normalized as $w_x^{(\ell)+} = \tilde{w}_x^{(\ell)+} / \sum_{\ell'=1}^{L_x} \tilde{w}_x^{(\ell')+}$.

For either the discrete or continuous versions of parameter flow, various approximations of the expected values, such as analytical linearization or quadrature approximation can be used.

5. DEMONSTRATION

As a demonstration of the capabilities of the discrete parameter flow (DPF) and continuous parameter flow (CPF) methods, a low-dimensional test case is considered in which a range measurement is used to estimate the planar position of an uncertain object. The prior knowledge, represented by $p(\boldsymbol{x})$, is defined by the Gaussian pdf, $p(\boldsymbol{x}) = p_g(\boldsymbol{x}; \boldsymbol{m}_x, \boldsymbol{P}_{xx})$, where the state is $\boldsymbol{x} = [x \ \ y]^T$, and the mean and covariance of the prior pdf are, respectively,

$$\boldsymbol{m}_x = \begin{bmatrix} 15 \\ 15 \end{bmatrix} \quad \text{and} \quad \boldsymbol{P}_{xx} = \begin{bmatrix} 10^2 & 0 \\ 0 & 15^2 \end{bmatrix} .$$

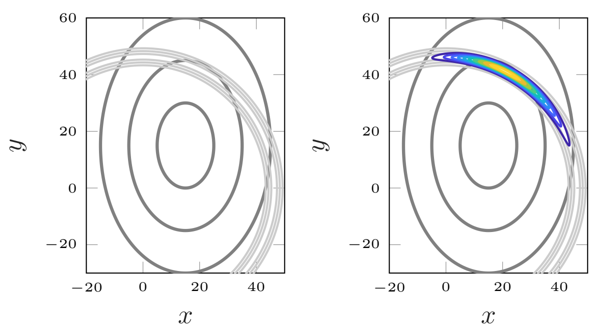
No units are given, but all units in the problem may be considered to be consistent distance units.

A range measurement is taken to be modeled as

$$z = \|\boldsymbol{x}\| + v ,$$

such that the range is taken from the origin, where v represents measurement noise that is modeled according to $p(v) = p_g(z; 0, P_{vv})$, where $P_{vv} = 1$ is the measurement noise covariance. Equivalently, the measurement model may be represented as the conditional likelihood $p(z|\boldsymbol{x}) = p_g(z; \|\boldsymbol{x}\|, P_{vv})$. For this demonstration the measurement used is sampled from the measurement likelihood to be $z = 46.2891$. The same measurement is used in all of the following results.

The prior pdf, $p(\boldsymbol{x})$, and the measurement likelihood, $p(z|\boldsymbol{x})$, are depicted in Fig. 1a, where the dark gray lines are isoprobability contours of $p(\boldsymbol{x})$ (i.e., the 1, 2, and 3σ contours of $p(\boldsymbol{x})$) and the light gray lines are isoprobability contours of $p(z|\boldsymbol{x})$ (also 1, 2, and 3σ contours of $p(z|\boldsymbol{x})$).

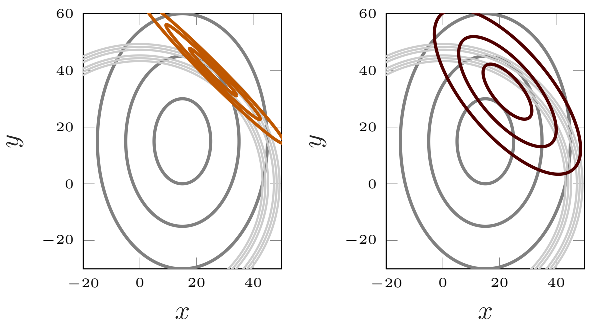


(a) Problem setup (b) Bayesian solution

Figure 1: Uncertainty contours for the prior, measurement likelihood, and the result of Bayes' rule

Figure 1b depicts the result of a grid-based implementation of Bayes' rule. At each point in the state space, Eq. (2) is applied to determine the value of the posterior pdf at each point. This result is then normalized so that the grid-based implementation of Bayes' rule produces a valid pdf. The result of Fig. 1b highlights the "and" property of Bayesian inference; to wit, posterior density is non-zero at locations in the state space where there is both prior probability *and* evidence supported by the new measurement. The result is a distinctly non-Gaussian shape, and it is this result that is desired to be approximated by other methods. It is also important to note that this result is not parameterized; all other methods examined in this work are, i.e., their posterior pdfs are represented using weights, means, and/or covariances, rather than relying on grid-based computations.

The first two methods implemented for comparison are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), under the Bayesian interpretation of Gaussianity. The EKF uses a first-order Taylor series approach, whereas the UKF uses the unscented transform with $\alpha = 0.1$, $\beta = 2$, and $\kappa = 1$. The posterior pdf for the EKF is depicted in Fig. 2a, and the poster pdf for the UKF is depicted in Fig. 2b. Neither result is able to reproduce the curved nature of the Bayesian solution, but the UKF clearly captures more of the region of posterior probability that is indicated by the Bayesian result of Fig. 1b.

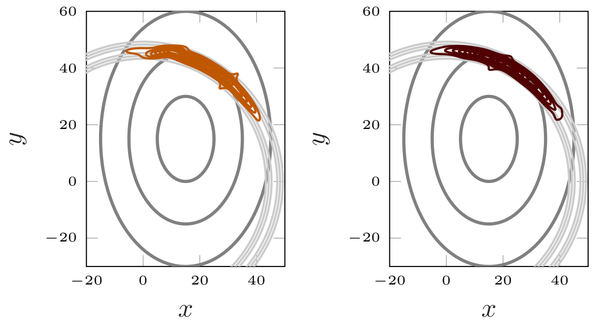


(a) EKF (b) UKF

Figure 2: Posterior uncertainty contours for Kalman filter based approaches

The Gaussian posterior densities of the EKF and UKF operating within the Bayesian paradigm do not permit any flexibility in the shape of the posterior pdf. To combat this, the EGMF and UGMF (with the same parameters as the UKF) are applied to the same problem. The results are depicted in Figs. 3a and 3b. Unlike the EKF and UKF, the EGMF and UGMF can produce curved pdfs. A visual comparison of each one to the Bayesian pdf from Fig. 1b shows that there is some similarity, but that there are also some undesirable artifacts in the posterior pdfs of the EGMF and UGMF.

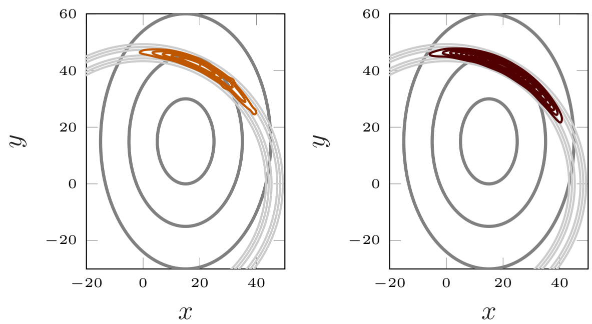
The next level of sophistication is to use the DPF approach with both the EGMF and UGMF. In this case, a



(a) EGMF (b) UGMF

Figure 3: Posterior uncertainty contours for Gaussian mixture filter based approaches

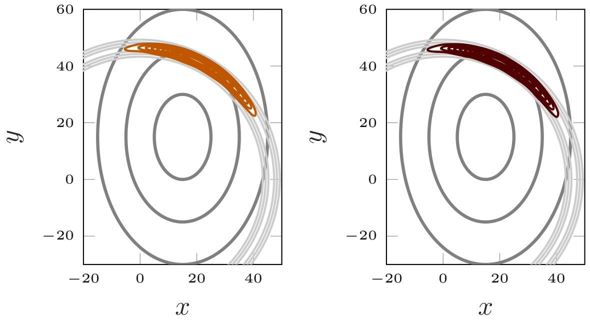
set of $M = 10$ steps is applied via a cubic rule; that is, $M = 10$ steps spaced according to a cubic progression through pseudotime. The DPF-UGMF uses the same parameters as those used in the UKF. The results of the DPF-EGMF are shown in Fig. 4a, and the results of the DPF-UGMF are shown in Fig. 4b. Compared to the results of Figs. 3a and 3b, the posterior pdfs obtained by the DPF approach are much smoother and remove many of the artifacts observed in the results of the GMF approach. Careful inspection of Fig. 4b also shows that the tails of the unscented version more closely mirror the desired Bayesian tails, as compared to the tails of the extended version of Fig. 4a.



(a) DPF-EGMF (b) DPF-UGMF

Figure 4: Posterior uncertainty contours for discrete parameter flow Gaussian mixture filter based approaches

The final methods considered implement the CPF approach. The results of the CPF-EGMF and the CPF-UGMF are illustrated in Figs. 5a and 5b, respectively. Unlike the DPF methods, no step size is dictated; instead a 4$^{\text{th}}$-order Runge-Kutta method with 5$^{\text{th}}$-order step size control is used to numerical integrate Eqs. (14). The same parameters as used for the other unscented implementations are used for the CPF-UGMF. Of all the methods considered, the results of Figs. 5a and 5b most closely mirror the Bayesian posterior of Fig. 1b. Additionally, of all the extended/unscented implementations, the CPF-based ones also exhibit the most visual similarity.



(a) CPF-EGMF (b) CPF-UGMF

Figure 5: Posterior uncertainty contours for continuous parameter flow Gaussian mixture filter based approaches

To provide a quantitative measure of performance for the eight different approximate Bayesian posteriors applied in this demonstration, the information degradation defined in Eq. (5) is computed via Monte Carlo sampling with 1×10^8 samples. The tabulated results are shown in Table 1. As expected from the illustrated results, the EKF and UKF perform the worst, with the UKF significantly outperforming the EKF. The GM-based techniques are all much better, with the parameter flow approaches obtaining the best performance overall. In every case, the unscented implementation outperforms its extended counterpart. The CPF-UGMF is the overall best performer, with about a 15% improvement over the DPF-UGMF, which is the second-best performer (tied with the UGMF).

Table 1: Information degradation relative to Bayesian posterior for the applied approaches

Method	ΔI [nats]	Method	ΔI [nats]
EKF	22.2289	UKF	2.0846
EGMF	0.1434	UGMF	0.1133
DPF-EGMF	0.1316	DPF-UGMF	0.1133
CPF-EGMF	0.1320	CPF-UGMF	0.0966

6. RESULTS AND DISCUSSION

The proposed parameter flow approach is applied to a space object tracking problem by considering the simulated processing of a single line-of-sight measurement. The simulated object considered is a GPS satellite with a semimajor axis of $a = 26560.185$ km, an eccentricity of $e = 0.008$, and an inclination of $i = 55.674°$. The orbital state of the object is represented using alternate equinoctial orbital elements (AEOEs), which are defined in terms of the standard Keplerian orbital elements, $\{a, e, i, \Omega, \omega, M\}$, as [10, 19]

$$\begin{aligned} n &= \sqrt{\mu/a^3} \\ h &= e\sin(\Omega+\omega) \\ k &= e\cos(\Omega+\omega) \\ p &= \tan(i/2)\sin\Omega \\ q &= \tan(i/2)\cos\Omega \\ \ell &= \Omega+\omega+M\,. \end{aligned}$$

To determine the initial state uncertainty of the object, the Cartesian coordinates (inertial position and velocity) representation of the object is used. Uncorrelated, zero-mean, Gaussian errors of 1 km (1σ) in position and 1 m/s (1σ) in velocity are added to the Cartesian coordinates for a set of 1×10^5 samples. Each sample is converted to AEOE coordinates, and the sample mean, $\boldsymbol{m}_{x,0}$, and covariance, $\boldsymbol{P}_{xx,0}$, are determined. This provides the initial Gaussian distribution in AEOE coordinates, $p(\boldsymbol{x}_0) = p_g(\boldsymbol{x}_0; \boldsymbol{m}_{x,0}, \boldsymbol{P}_{xx,0})$.

To simulate the measurement of right-ascension and declination, a random sample of the state is drawn from $p(\boldsymbol{x}_0)$. This sample is propagated according to Keplerian motion represented in AEOE coordinates, which is governed by the linear dynamical system

$$\begin{aligned} \dot{n} &= 0 \\ \dot{h} &= 0 \\ \dot{k} &= 0 \\ \dot{p} &= 0 \\ \dot{q} &= 0 \\ \dot{\ell} &= n\,, \end{aligned}$$

to a time t_k. At time t_k, it is assumed that a ground-based sensor located on the surface of a spherical Earth is directly beneath the object and can take a line-of-sight measurement. The subsequent right-ascension and declination measurements are subjected to uncorrelated, zero-mean, Gaussian measurement noise with standard deviations of $3''$, each. The time at which the simulated measurement is processed, t_k, is varied from one to fourteen days after t_0.

The inference methods employed in this study are the standard EKF and UKF updates, as well as the CPF-EGMF and CPF-UGMF updates. The unscented implementations all use the same parameters of $\alpha = 0.1$, $\beta = 2.0$, and $\kappa = -3.0$. The GM-based implementations all use a single Gaussian, as do the EKF and UKF implementations. The input into each update scheme is the same propagated Gaussian, which is obtained by propagating $\boldsymbol{m}_{x,0}$ and $\boldsymbol{P}_{xx,0}$ from t_0 to t_k. Since the dynamics of the AEOEs are linear, the input properly describes the distribution of the prior. The only difference is the way in which the nonlinear measurement is processed.

Performance of the methods is analyzed by computing the posterior estimation error,

$$\boldsymbol{e}^+_{x,k} = \boldsymbol{x}_k - \boldsymbol{m}^+_{x,k}\,,$$

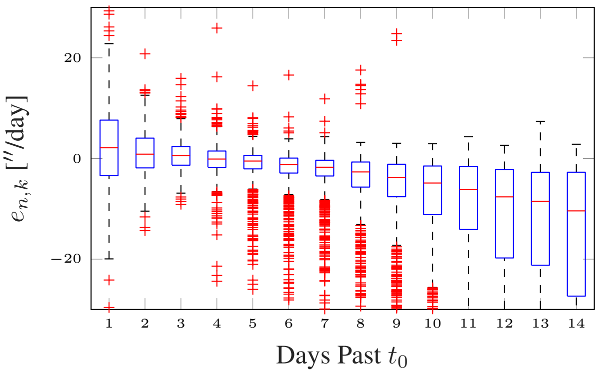
and the normalized estimation error squared (NEES),

$$d_{x,k}^2 = \frac{1}{n_x}(\boldsymbol{x}_k - \boldsymbol{m}_{x,k}^+)^T \boldsymbol{P}_{xx,k}^+ (\boldsymbol{x}_k - \boldsymbol{m}_{x,k}^+)\,,$$

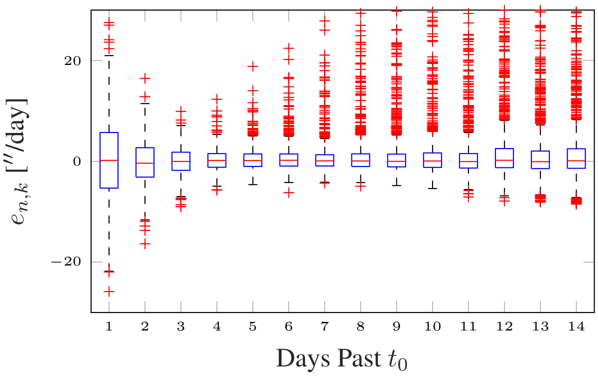
where $n_x = 6$ is the state dimension, $\boldsymbol{m}_{x,k}^+$ is posterior mean (taken to be the estimate), and $\boldsymbol{P}_{xx,k}^+$ is the posterior covariance. Each of the four inference procedures produces, in general, a different posterior mean and covariance. A well-behaved inference procedure will have the characteristics that the posterior estimation error is unbiased (i.e., has a mean of zero) and that the posterior covariance provides a proper statistical quantification of the posterior estimation err (i.e., that the mean NEES is one).

The estimation errors obtained for each of the implemented methods (the EKF, the UKF, the CPF-EGMF, and the CPF-UGMF) for the mean motion state are illustrated in Fig. 6. This state is selected for analysis due to its relevance in the dynamics of the AEOEs. The results for each approach are illustrated via box plots. The box plot includes the median of the data as the central mark within the box. The box itself represents the data residing within the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. The dashed lines extended away from the box indicate the extent of the most extreme data points still considered inliers, and the individual marks indicate data classified as outliers. Figure 6a indicates that the EKF cannot achieve median estimation errors near zero beyond more than a few days of propagation time. Even then, the influence of outliers is significant, indicating non-robust performance of the EKF. Figure 6b shows improved median estimation error performance, with the median remaining near zero for the entire time period considered. The outliers, however, indicate that the UKF also performs with a lack of robustness. Figures 6c and 6d, on the other hand, indicate extremely robust performance for all propagation times considered when processing a single line-of-sight measurement. It is also worth noting that the spread of the posterior estimation errors (as indicated by the box) shrinks as the propagation time before the measurement processing increases. The increased propagation time leads to more direct correlation between the mean motion and the line-of-sight measurement, which makes the processed measurement more informative in estimating the mean motion.

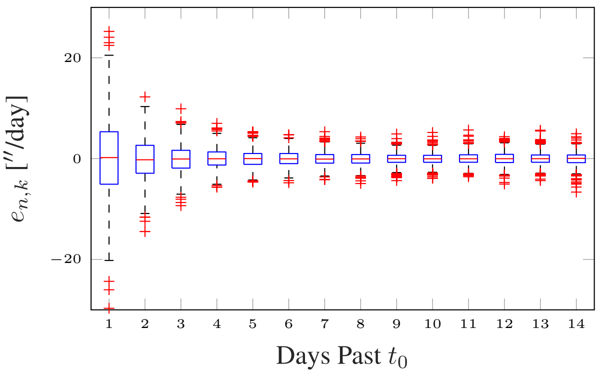
The NEES for each of the four implemented methods are illustrated as box plots in Fig. 7. Figure 7a indicates that the EKF is incapable of producing statistically consistent results at even one day of propagation time. As the propagation time increases, the consistency of the filter continues to degrade. Figure 7b shows that the UKF can produce statistically consistent results up to about three days of propagation time. After that, the consistency of the UKF consistently degrades. Figures 7c and 7d, however, show that the CPF-based approaches are successful in producing statistically consistent results, even when the measurement is fourteen days past the initial estimate of the state of the object. Comparing Figs. 7c and 7d, it is seen that there is very little difference between the CPF-EGMF and CPF-UGMF performance.



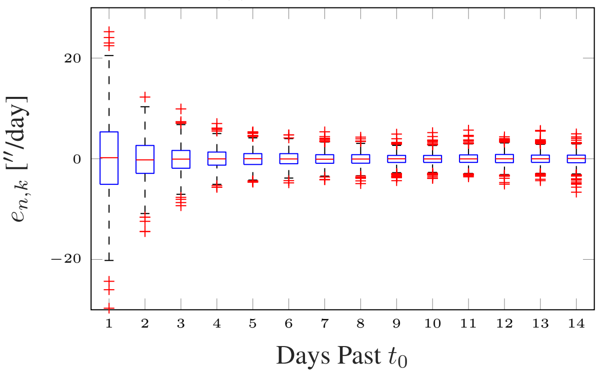
(a) EKF



(b) UKF

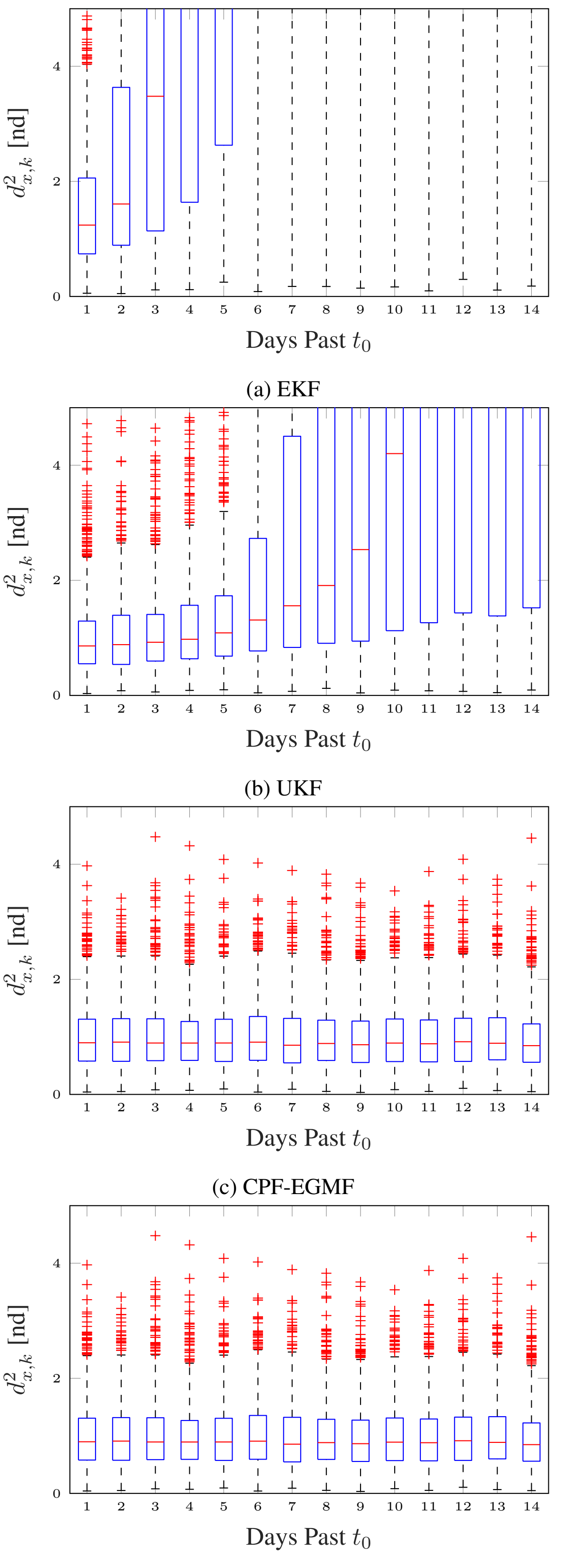


(c) CPF-EGMF



(d) CPF-UGMF

Figure 6: Estimation error in the mean motion for each of the implemented filters



(a) EKF

(b) UKF

(c) CPF-EGMF

(d) CPF-UGMF

Figure 7: Statistical consistency via the NEES for each of the implemented filters

REFERENCES

1. Alspach, D. L. and Sorenson, H. W., (1972), Nonlinear Bayesian Estimation using Gaussian Sum Approximations, *IEEE Transactions on Automatic Control*, **17**(4), 439-448.
2. Arasaratnam, I., et al., (2007), Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature, *Proceedings of the IEEE*, **95**(5), 953-977.
3. Cover, T. M. and Thomas, J. A., (2006), Elements of Information Theory, 2nd ed., Wiley-Interscience.
4. Craft, K. J. and DeMars, K. J., (2025), Homotopic Gaussian Mixture Filtering for Applied Bayesian Inference, *IEEE Transactions on Automatic Control*, to appear.
5. Daum F. and Huang, J. (2011), Particle degeneracy: root cause and solution, *Proc. SPIE 8050, Signal Processing, Sensor Fusion, and Target Recognition XX, 80500W*.
6. DeMars, K. J., et al., (2013), Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems, *Journal of Guidance, Control, and Dynamics*, **36**(4), 1047-1057.
7. García-Fernández, Á. F., et al., (2014), Iterated statistical linear regression for Bayesian updates, *17th International Conference on Information Fusion (FUSION)*.
8. Hanebeck, U. D., et al., (2003), Progressive Bayes: A New Framework for Nonlinear State Estimation, *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*, 256-267.
9. Ho, Y. C. and Lee, R. C. K., (1964), A Bayesian Approach to Problems in Stochastic Estimation and Control, *IEEE Transactions on Automatic Control*, **9**(4), 333-339.
10. Horwood, J. T., et al., (2011), Gaussian Sum Filters for Space Surveillance: Theory and Simulations, *Journal of Guidance, Control, and Dynamics*, **34**(6), 1839-1851.
11. Ito, K. and Xiong, K., (2000), Gaussian Filters for Nonlinear Filtering Problems, *IEEE Transactions on Automatic Control*, **45**(5), 910–927.
12. Jaynes, E. T., (1988), [Optimal Information Processing and Bayes's Theorem]: Comment, *The American Statistician*, **42**(4), 280-281.
13. Julier, S. J., and Uhlmann, J. K., (2004), Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, **92**(3), 401-422.
14. Kalman, R. E. (1960), A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME–Journal of Basic Engineering*, **82**(Series D), 35-45.
15. Kullback, S. and Leibler, R. A., (1951), On Information and Sufficiency, *The Annals of Mathematical Statistics*, **22**(1), 79-86.
16. Kullberg, A., et al., (2023), On the Relationship Between Iterated Statistical Linearization and Quasi–Newton Methods, *IEEE Signal Processing Letters*, **30**, 1777-1781.

17. Michaelson, K., et al., (2023), Ensemble Kalman Filter with Bayesian Recursive Update, *26th International Conference on Information Fusion (FUSION)*.

18. Michaelson, K., et al., (2024), Particle Flow with a Continuous Formulation of the Nonlinear Measurement Update, *27th International Conference on Information Fusion (FUSION)*.

19. Montenbruck, O. and Gill, E., (2000), Satellite Orbits: Models, Methods, and Applications, Springer.

20. Särkkä, S., (2013), Bayesian Filtering and Smoothing, Cambridge.

21. Shannon, C. E. and Weaver, W., (1949), The Mathematical Theory of Communication, The University of Illinois Press.

22. Sorenson, H. W. and Alspach, D. L., (1971), Recursive Bayesian Estimation Using Gaussian Sums, *Automatica*, **7**(4), 465-479.

23. Terejanu, G., et al., (2008), Uncertainty Propagation for Nonlinear Dynamic Systems Using Gaussian Mixture Models, *Journal of Guidance, Control, and Dynamics*, **31**(6), 1623-1633.

24. van der Merwe, R., (2004), Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models, Ph.D. thesis, Oregon Health and Science University.

25. Ward, K. C. and DeMars, K. J., (2022), Information-based Particle Flow with Convergence Control, *IEEE Transactions on Aerospace and Electronic Systems*, **58**(2), 1377–1390.

26. Zellner, A., (1988), Optimal Information Processing and Bayes's Theorem, *The American Statistician*, **42**(4), 278-280.

A. PROOF OF EQS. (10)

It is known that the partitioned Gaussian likelihood of Eq. (9) and the GM prior of Eq. (7) can substituted into Bayes' rule to give

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \left[\prod_{k=1}^{M} \left|2\pi \boldsymbol{P}_{vv}/\Delta s_k\right|^{\frac{1}{2}}\right] \times \left[\prod_{k=1}^{M} p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_k)\right] p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell)-},\boldsymbol{P}_{xx}^{(\ell)-})\,.$$

If the first term of the product of partial likelihoods is separated out from the others, it follows that

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} &\left[\prod_{k=1}^{M} \left|2\pi \boldsymbol{P}_{vv}/\Delta s_k\right|^{\frac{1}{2}}\right] \\ &\times \left[\prod_{k=2}^{M} p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_k)\right] \\ &\times p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_1) p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell)-},\boldsymbol{P}_{xx}^{(\ell)-})\,, \end{aligned}$$

where the last two terms represent a product of Gaussians in which the first term is a nonlinear function of the state. Following along the lines of [1], this product is approximated by another product of Gaussians in which the conditional dependence is reversed.

To facilitate such an approximation, let $\boldsymbol{m}_x^{(\ell)-} \to \boldsymbol{m}_x^{(\ell,0)}$ and $\boldsymbol{P}_{xx}^{(\ell)-} \to \boldsymbol{P}_{xx}^{(\ell,0)}$. Applying statistical linearization [7, 16] with respect to $p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell,0)},\boldsymbol{P}_{xx}^{(\ell,0)})$,

$$\begin{aligned} &p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_k) \\ &\quad\approx p_g(\boldsymbol{z};\boldsymbol{A}^{(\ell,0)}\boldsymbol{x}+\boldsymbol{b}^{(\ell,0)},(\boldsymbol{P}_{vv}/\Delta s_1)+\boldsymbol{P}_{ee}^{(\ell,0)})\,, \end{aligned}$$

where

$$\boldsymbol{A}^{(\ell,i)} = (\boldsymbol{P}_{xh}^{(\ell,i)})^T(\boldsymbol{P}_{xx}^{(\ell,i)})^{-1} \tag{16a}$$

$$\boldsymbol{b}^{(\ell,i)} = \boldsymbol{m}_h^{(\ell,i)} - \boldsymbol{A}^{(\ell,i)}\boldsymbol{m}_x^{(\ell,i)} \tag{16b}$$

$$\boldsymbol{P}_{ee}^{(\ell,i)} = \boldsymbol{P}_{hh}^{(\ell,i)} - \boldsymbol{A}^{(\ell,i)}\boldsymbol{P}_{xx}^{(\ell,i)}(\boldsymbol{A}^{(\ell,i)})^T\,, \tag{16c}$$

and $\boldsymbol{m}_h^{(\ell,i)}$, $\boldsymbol{P}_{xh}^{(\ell,i)}$, and $\boldsymbol{P}_{hh}^{(\ell,i)}$ are given by Eqs. (12).

It is useful to note that Eqs. (16) can be rearranged to produce

$$\begin{aligned} \boldsymbol{A}^{(\ell,i)}\boldsymbol{m}_x^{(\ell,i)} + \boldsymbol{b}^{(\ell,i)} &= \boldsymbol{m}_h^{(\ell,i)} \\ \boldsymbol{A}^{(\ell,i)}\boldsymbol{P}_{xx}^{(\ell,i)}(\boldsymbol{A}^{(\ell,i)})^T + \boldsymbol{P}_{ee}^{(\ell,i)} &= \boldsymbol{P}_{hh}^{(\ell,i)}\,. \end{aligned}$$

The Gaussian product,

$$p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_1)p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell,0)},\boldsymbol{P}_{xx}^{(\ell,0)})\,,$$

can now be addressed using the approximation provided by statistical linearization and the approach of [1], which yields

$$\begin{aligned} p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_1)&p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell,0)},\boldsymbol{P}_{xx}^{(\ell,0)}) \\ &= p_g(\boldsymbol{z};\boldsymbol{m}_h^{(\ell,0)},\boldsymbol{P}_{hh}^{(\ell,0)}+(\boldsymbol{P}_{vv}/\Delta s_1)) \\ &\quad\times p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell,1)},\boldsymbol{P}_{xx}^{(\ell,1)})\,, \end{aligned} \tag{17}$$

where

$$\begin{aligned} \boldsymbol{m}_x^{(\ell,1)} &= \boldsymbol{m}_x^{(\ell,0)} + \boldsymbol{K}^{(\ell,1)}(\boldsymbol{z}-\boldsymbol{m}_h^{(\ell,0)}) \\ \boldsymbol{P}_{xx}^{(\ell,1)} &= \boldsymbol{P}_{xx}^{(\ell,0)} - \boldsymbol{K}^{(\ell,1)}\boldsymbol{P}_{hh}^{(\ell,0)}(\boldsymbol{K}^{(\ell,1)})^T \\ &\quad - \boldsymbol{K}^{(\ell,1)}(\boldsymbol{P}_{vv}/\Delta s_1)(\boldsymbol{K}^{(\ell,1)})^T\,, \end{aligned}$$

and $\boldsymbol{K}^{(\ell,i)}$ is given by Eq. (11b).

Returning to the posterior with the result of Eq. (17), it follows that

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} &\left[\prod_{k=1}^{M} \left|2\pi \boldsymbol{P}_{vv}/\Delta s_k\right|^{\frac{1}{2}}\right] \\ &\times p_g(\boldsymbol{z};\boldsymbol{m}_h^{(\ell,0)},\boldsymbol{P}_{hh}^{(\ell,0)}+(\boldsymbol{P}_{vv}/\Delta s_1)) \\ &\times \left[\prod_{k=2}^{M} p_g(\boldsymbol{z};\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{P}_{vv}/\Delta s_k)\right] \\ &\times p_g(\boldsymbol{x};\boldsymbol{m}_x^{(\ell,1)},\boldsymbol{P}_{xx}^{(\ell,1)})\,. \end{aligned}$$

The next element of the partitioned likelihood can be removed from the product of remaining partial likelihoods, and the previous process of approximating the Gaussian product by another Gaussian product can be repeated. Carrying out this process for all of the partitions of the likelihood produces

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \left[\prod_{k=1}^{M} \left|2\pi(\boldsymbol{P}_{vv}/\Delta s_k)\right|^{\frac{1}{2}} \times k^{(\ell,k)} \right] p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)+}, \boldsymbol{P}_{xx}^{(\ell)+}) \,,$$

where $k^{(\ell,k)}$ is given by Eq. (11a) and the mean and covariance are determined by the complete application of the iterative relationships given in Eqs. (10b) and (10c). Note that this result is a non-normalized Gaussian mixture. Alternatively, the factors of $\left|2\pi \boldsymbol{P}_{vv}\right|^{-1/2}$ and $\prod_{k=1}^{M} \left|2\pi(\boldsymbol{P}_{vv}/s_k)\right|^{1/2}$ can be removed, as they have no dependence on the component index of the mixture, leading to

$$p(\boldsymbol{x}|\boldsymbol{z}) \propto \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \left[\prod_{k=1}^{M} k^{(\ell,k)} \right] p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)+}, \boldsymbol{P}_{xx}^{(\ell)+}) \,.$$

To resolve the proportionality in Bayes' rule and formulate a proper posterior pdf, the evidence is computed as

$$\int p(\boldsymbol{z}|\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \sum_{\ell=1}^{L_x} w_x^{(\ell)-} \prod_{k=1}^{M} k^{(\ell,k)} \,.$$

The fully normalized Bayesian posterior pdf can be expressed as

$$p(\boldsymbol{x}|\boldsymbol{z}) = \sum_{\ell=1}^{L_x} w_x^{(\ell)+} p_g(\boldsymbol{x}; \boldsymbol{m}_x^{(\ell)+}, \boldsymbol{P}_{xx}^{(\ell)+}) \,,$$

where the posterior weights are such that

$$w_x^{(\ell)+} \propto w_x^{(\ell)-} \prod_{k=1}^{M} k^{(\ell,k)} \,.$$

Ensuring that $\sum_{\ell=1}^{L_x} w_x^{(\ell)+} = 1$, or, equivalently, that $p(\boldsymbol{x}|\boldsymbol{z})$ is a valid pdf, is accomplished by normalizing the weights as

$$w_x^{(\ell)+} = \frac{w_x^{(\ell)-} \prod_{k=1}^{M} k^{(\ell,k)}}{\sum_{\ell'=1}^{L_x} w_x^{(\ell')-} \prod_{k'=1}^{M} k^{(\ell',k')}} \,.$$

To define a recursion for the weights through the partitioned likelihood, each of the terms in the numerator and denominator products can be applied one by one to produce

$$w_x^{(\ell,i)} = \frac{k^{(\ell,i)} w_x^{(\ell,i-1)}}{\sum_{\ell'=1}^{L_x^-} k^{(\ell',i)} w_x^{(\ell',i-1)}} \,,$$

which is initialized with $w_x^{(\ell,0)} = w_x^{(\ell)-}$; after the last iteration ($i = M$) is completed, $w_x^{(\ell)+} = w_x^{(\ell,M)}$. This is the result of Eq. (10a), and ensures that $\sum_{\ell=1}^{L_x} w_x^{(\ell,i)} = 1$ at each iteration. As discussed elsewhere, normalization at every iteration is not required.

B. PROOF OF EQS. (14)

The governing ordinary differential equations for CPF come from taking the limit of the iterative relationships for DPF, as in Eq. (13).

First, consider the mean of the ℓ^{th} component. The discrete iteration is given by Eq. (10b), such that

$$\frac{\mathrm{d}\boldsymbol{m}_x^{(\ell)}(s)}{\mathrm{d}s} = \lim_{\Delta s_i \to 0} \frac{\boldsymbol{K}^{(\ell,i)}(\boldsymbol{z} - \boldsymbol{m}_h^{(\ell,i-1)})}{\Delta s_i} \,.$$

Substituting for $\boldsymbol{K}^{(\ell,i)}$ from Eq. (11b) and simplifying, it follows that

$$\frac{\mathrm{d}\boldsymbol{m}_x^{(\ell)}(s)}{\mathrm{d}s} = \lim_{\Delta s_i \to 0} \boldsymbol{P}_{xh}^{(\ell,i-1)} \big(\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv}\big)^{-1} (\boldsymbol{z} - \boldsymbol{m}_h^{(\ell,i-1)}) \,.$$

Evaluating the limit, noting that the the expected values go to continuous versions of their respective equations, the differential equation for each mean is given by Eq. (14b).

Next, consider the covariance of the ℓ^{th} component. Applying the discrete iteration of Eq. (10c) to the limit of Eq. (13) leads to

$$\frac{\mathrm{d}\boldsymbol{P}_{xx}^{(\ell)}(s)}{\mathrm{d}s} = -\lim_{\Delta s_i \to 0} \frac{\boldsymbol{K}^{(\ell,i)}(\boldsymbol{P}_{hh}^{(\ell,i-1)} + (\boldsymbol{P}_{vv}/\Delta s_i))(\boldsymbol{K}^{(\ell,i)})^T}{\Delta s_i} \,.$$

Substituting for $\boldsymbol{K}^{(\ell,i)}$ from Eq. (11b) and simplifying, it follows that

$$\frac{\mathrm{d}\boldsymbol{P}_{xx}^{(\ell)}(s)}{\mathrm{d}s} = -\lim_{\Delta s_i \to 0} \boldsymbol{P}_{xh}^{(\ell,i-1)} \big(\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv}\big)^{-1} (\boldsymbol{P}_{xh}^{(\ell,i-1)})^T \,.$$

In a similar manner to the mean, evaluating the limit yields Eq. (14c).

Finally, consider the weights and the discrete iteration for the ℓ^{th} component given by Eq. (10a). Before attempting to find a governing differential equation, a few new terms and variations on expressions are introduced to simplify further developments. Let $(d_z^{(\ell,i-1)})^2$ be the squared Mahalanobis distance of the measurement for the ℓ^{th} component at the $(i-1)^{\text{th}}$ iteration; by definition,

$$(d_z^{(\ell,i-1)})^2 = (\boldsymbol{e}_z^{(\ell,i-1)})^T (\boldsymbol{P}_{hh}^{(\ell,i-1)} + (\boldsymbol{P}_{vv}/\Delta s_i))^{-1} \boldsymbol{e}_z^{(\ell,i-1)} \,,$$

where $\boldsymbol{e}_z^{(\ell,i-1)} = \boldsymbol{z} - \boldsymbol{m}_h^{(\ell,i-1)}$ is the innovation for the ℓ^{th} component at the $(i-1)^{\text{th}}$ iteration. Equivalently, the squared Mahalanobis distance may be written as

$$(d_z^{(\ell,i-1)})^2 = \Delta s_i (\boldsymbol{e}_z^{(\ell,i-1)})^T (\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv})^{-1} \boldsymbol{e}_z^{(\ell,i-1)} \,,$$

which makes it clear to see that

$$\lim_{\Delta s_i \to 0} (d_z^{(\ell,i-1)})^2 = 0 \,. \tag{18}$$

From the definition of $k^{(\ell,i)}$ in Eq. (11a) and the form of the Gaussian pdf, it follows that

$$\bar{k}^{(\ell,i)} = \frac{k^{(\ell,i)}}{\Delta s_i^{n_z/2}} = \left|2\pi(\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv})\right|^{-\frac{1}{2}} \times \exp\left\{ -(d_z^{(\ell,i-1)})^2/2 \right\} , \tag{19}$$

where n_z is the dimension of $\boldsymbol{z}$; furthermore,

$$\lim_{\Delta s_i \to 0} \bar{k}^{(\ell,i)} = \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} , \tag{20}$$

which relies on Eq. (18).

The difference of the weights between two successive steps, $\Delta w_x^{(\ell,i)} = w_x^{(\ell,i)} - w_x^{(\ell,i-1)}$, is given by

$$\Delta w_x^{(\ell,i)} = \left[\frac{k^{(\ell,i)} - \sum_{\ell'=1}^{L_x} k^{(\ell',i)} w_x^{(\ell',i-1)}}{\sum_{\ell'=1}^{L_x} k^{(\ell',i)} w_x^{(\ell',i-1)}} \right] w_x^{(\ell,i-1)} ,$$

or, equivalently,

$$\Delta w_x^{(\ell,i)} = \left[\frac{\bar{k}^{(\ell,i)} - \sum_{\ell'=1}^{L_x} \bar{k}^{(\ell',i)} w_x^{(\ell',i-1)}}{\sum_{\ell'=1}^{L_x} \bar{k}^{(\ell',i)} w_x^{(\ell',i-1)}} \right] w_x^{(\ell,i-1)} .$$

To find the time rate of change of the weights, the limit of the ratio of the change in weights to the interval width is taken as the interval width goes to zero, i.e.,

$$\frac{\mathrm{d}w_x^{(\ell)}(s)}{\mathrm{d}s} = \lim_{\Delta s_i \to 0} \frac{\Delta w_x^{(\ell,i)}}{\Delta s_i} \,. \tag{21}$$

Leveraging Eq. (20) and $\sum_{\ell=1}^{L_x} w_x^{(\ell,i-1)} = 1$, it can be shown that

$$\lim_{\Delta s_i \to 0} \frac{\Delta w_x^{(\ell,i)}}{\Delta s_i} = \frac{\left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} - \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}}}{0 \cdot \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}}} w_x^{(i,\ell-1)} ,$$

which is clearly indeterminate and necessitates the application of de l'Hôpital's rule.

To apply de l'Hôpital's rule, a few limits of derivatives are required. The key element that appears in the change in the weights are the $\bar{k}^{(\ell,i)}$ terms. As such, starting from Eq. (19), it can be shown that

$$\begin{aligned}
\frac{\mathrm{d}\bar{k}^{(\ell,i)}}{\mathrm{d}\Delta s_i} &= \\
&- \frac{1}{2}\bar{k}^{(\ell,i-1)} \Bigg[\mathrm{tr}\Big\{ \big(\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv}\big)^{-1} \boldsymbol{P}_{hh}^{(\ell,i-1)} \Big\} \\
&+ (\boldsymbol{e}_z^{(\ell,i-1)})^T (\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv})^{-1} \boldsymbol{e}_z^{(\ell,i-1)} \\
&- \Big(\Delta s_i (\boldsymbol{e}_z^{(\ell,i-1)})^T (\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv})^{-1} \\
&\quad \times \boldsymbol{P}_{hh}^{(\ell,i-1)} (\Delta s_i \boldsymbol{P}_{hh}^{(\ell,i-1)} + \boldsymbol{P}_{vv})^{-1} \boldsymbol{e}_z^{(\ell,i-1)} \Big) \Bigg] ,
\end{aligned}$$

which leads to the limit

$$\lim_{\Delta s_i \to 0} \frac{\mathrm{d}\bar{k}^{(\ell,i)}}{\mathrm{d}\Delta s_i} = -\frac{1}{2}\left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \times \left[\mathrm{tr}\Big\{ \boldsymbol{P}_{vv}^{-1} \boldsymbol{P}_{hh}^{(\ell)}(s) \Big\} + (\boldsymbol{e}_z^{(\ell)}(s))^T \boldsymbol{P}_{vv}^{-1} (\boldsymbol{e}_z^{(\ell)}(s)) \right] . \tag{22}$$

From the result of Eq. (22), it directly follows that

$$\begin{aligned}
&\lim_{\Delta s_i \to 0} \frac{\mathrm{d}}{\mathrm{d}\Delta s_i} \left\{ \sum_{\ell'=1}^{L_x} \bar{k}^{(\ell',i)} w_x^{(\ell',i-1)} \right\} \\
&= -\frac{1}{2}\left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} \left\{ \sum_{\ell'=1}^{L_x} \left[\mathrm{tr}\Big\{ \boldsymbol{P}_{vv}^{-1} \boldsymbol{P}_{hh}^{(\ell')}(s) \Big\} \right.\right. \\
&\quad \left.\left. + (\boldsymbol{e}_z^{(\ell')}(s))^T \boldsymbol{P}_{vv}^{-1} (\boldsymbol{e}_z^{(\ell')}(s)) \right] w_x^{(\ell')}(s) \right\} ,
\end{aligned} \tag{23}$$

and that

$$\lim_{\Delta s_i \to 0} \frac{\mathrm{d}}{\mathrm{d}\Delta s_i} \left\{ \Delta s_i \sum_{\ell'=1}^{L_x} \bar{k}^{(\ell',i)} w_x^{(\ell',i-1)} \right\} = \left|2\pi \boldsymbol{P}_{vv}\right|^{-\frac{1}{2}} . \tag{24}$$

Applying de l'Hôpital's rule to Eq. (21) and leveraging the results of Eqs. (22)–(24) leads to the differential equation given in Eq. (14a).