JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 41, No. 3, March 2018

Adaptive Split/Merge-Based Gaussian Mixture Model Approach for Uncertainty Propagation

Kumar Vishwajeet* and Puneet Singla†
Pennsylvania State University, University Park, Pennsylvania 16802

DOI: 10.2514/1.G002801

This paper presents an adaptive splitting and merging scheme for dynamic selection of Gaussian kernels in a Gaussian mixture model. The Gaussian kernel in the Gaussian mixture model is split into multiple components if the Kolmogorov equation error exceeds a prescribed threshold. Two different splitting mechanisms are presented in this work. The first splitting mechanism corresponds to splitting one Gaussian kernel in all directions, whereas the second splitting mechanism corresponds to splitting in only the direction of maximum nonlinearity. The state transition matrix in conjunction with unscented transformation is used to compute the departure from linearity, and hence approximate the direction of maximum nonlinearity. The merging mechanism exploits the angle between eigenvectors corresponding to the maximum eigenvalue of covariance matrices corresponding to two different Gaussian kernels to find candidate components for merging. Finally, a sparse approximation problem is defined to provide a mechanism to trade off between the number of Gaussian kernels and the Kolmogorov equation error in a mixture model. The uncertainty propagation problem for a satellite motion in a low Earth orbit is considered to show the efficacy of the proposed ideas.

I. Introduction

NUMEROUS fields of science and engineering present the problem of uncertainty propagation through nonlinear dynamic systems with stochastic excitation and uncertain initial conditions [1]. One may be interested in predicting the probability of collision of an asteroid with Earth [2–4], diffusion of chemical plumes through air [5,6], control of movement and planning of actions of autonomous systems [7], the optimization of financial policies [8], active control of structural vibrations [9,10], the motion of particles under the influence of stochastic force fields [11,12], or simply the computation of the prediction step in the design of a Bayes filter [13,14]. All these applications require the study of the relevant stochastic dynamic system.

For a linear stochastic dynamical system driven by Gaussian white noise with Gaussian state uncertainty, the probability density function (PDF) remains Gaussian for all time. Because a Gaussian PDF can be completely characterized using its first two moments (i.e., mean and covariance), the analysis of a linear stochastic dynamical system is relatively less challenging as compared to the analysis of a nonlinear stochastic dynamical system. A variety of algorithms has been developed by researchers to characterize and propagate uncertainty through nonlinear dynamical systems. It is well known that the nonlinearity of system dynamics invalidates the applicability of linear error theory in uncertainty propagation [15–17]. The non-Gaussian evolution of the joint state PDF from a given density function at epoch time and uncertain forcing functions is captured by solving the Kolmogorov equation for the joint density function at any time.

Analytical solutions for the Kolmogorov equation exist only for certain classes of dynamical systems [1,18–23]. The solution of the Kolmogorov equation for linear systems in conjunction with Bayes's rule to assimilate measurement data with linear model results in the Kalman filter [24,25], which is also an optimal mean square error filter. In general, the solution of the Kolmogorov equation requires various numerical approximations [26–32] involving the space discretization. These methods are severely handicapped for higher dimensions because the discretization of the space over which PDF lives is computationally impractical.

Various studies have endeavored to exploit the knowledge of statistics, dynamic systems, and numerical analysis to develop techniques [4,13,14,19,33–58] that cater to the various classes of problems of interest. All of these algorithms, except Monte Carlo (MC) methods, are similar in several respects; and they are suitable only for linear or moderately nonlinear systems because the effect of higher-order terms can lead to significant errors. In particular, regularized particle filters [55,56], mesh-free adjoint methods [52–54], polynomial chaos [51], and optimal transport-based methods [57,58] have gained special attention and have shown promising results in accurately capturing the non-Gaussian behavior of the density function. However, all of these methods suffer from the curse of dimensionality, and their numerical burden increases with the state dimension.

Recently, different variants of Gaussian mixture models (GMMs) have found special attention for uncertainty propagation through nonlinear systems [38,59–63]. All GMM-based algorithms are based upon the assumption that the probability density function of the state can be approximated by a mixture of Gaussian kernels. The mean and covariance information for each Gaussian kernel is computed independently. Different algorithms incorporate various criteria and corresponding rationale for selection of the Gaussian kernels and the computation of weight of each Gaussian kernel.

In earlier work, a variant of the GMM algorithm known as the adaptive Gaussian mixture model (AGMM) [5,61,64–66] was developed. The main feature of the AGMM algorithm is that it incorporates the solution of the integral or differential form of the Kolmogorov equation during the propagation step of the filter to update the weights of individual Gaussian kernels. This makes it an apt choice when measurement data are sparse in nature. Although this approach has been successfully used to solve many benchmark and engineering problems [5,64,65,67–69] (including the Lorenz system, nonlinear oscillators, the two-body problem, and rigid-body attitude estimation), the challenge lies in choosing appropriate Gaussian components.

The number of components in a GMM determines the accuracy and the computational complexity of the mixture. In prior work [61,65], the Gaussian kernels in the GMM were predetermined and were placed along the principal axis of the covariance matrix corresponding to the initial Gaussian density function. Initially, these Gaussian kernels were assigned zero weight, except for the Gaussian kernel, which corresponded to true initial Gaussian density function. The zero weights for these fictitious Gaussian kernels made sure that they did not participate in the GMM representation of the initial Gaussian density function. Over time, these zero weights became nonzero to minimize the Kolmogorov equation error. However, these

Received 31 January 2017; revision received 20 September 2017; accepted for publication 26 September 2017; published online 8 December 2017.

*Sensor Fusion Engineer, Delphi Advanced Engineering Center; kumarvis@buffalo.edu. Member AIAA.

†Associate Professor, Department of Aerospace Engineering, Delphi Advanced Engineering Center; psingla@psu.edu. Associate Fellow AIAA.

naively chosen Gaussian kernels could result in either an unnecessary increase in the number of Gaussian kernels in the GMM or poor approximation of the state PDF globally. Hence, it was desired to adapt the size of the GMM and the location of the GMM components over time while maintaining a tradeoff between the accuracy of the GMM approximation and the computational complexity. Such an approach should be able to split initial Gaussian kernels into multiple Gaussian kernels and merge/prune the components that do not contribute to the GMM approximation.

In the past, researchers [70–74] have provided different methods of selecting Gaussian kernels in a GMM by splitting the original component into many smaller components. Hanebeck et al. [70] proposed a method for scalar dynamical system by using a precalculated library of the mean, covariance, and weights of components formed due to splitting. The measurement feedback was used to adaptively split Gaussian components in the mixture model. The squared integral of the difference between the true posterior PDF and the approximated posterior PDF was used to identify Gaussian kernels to be split when measurements were available at every time step. Vittaldev and Russell [75] extended the library to a higher number of components; however, the Gaussian components were not adapted over time. Richardson and Green [73] introduced the concept of the reversible jump Markov chain Monte Carlo method, which was later used in [76] for adapting the number of components over time. Runnalls [74] used the Kullback–Leibler (KL) distance-based approach for merging Gaussian components. Zhang et al. [71] extended the work of Richardson in multidimensions in the context of pattern recognition. DeMars et al. [77] used the concept of minimization of entropy to split a Gaussian component. Many of these methods [70,71,75,76] compared the true posterior PDF with the approximated PDF to identify critical components to be split and the corresponding direction of splitting. This reduced the utility of these methods for pure forecast applications or when measurement data were sparse. These methods did not take into account the effect of split on the other components of the mixture.

One of the challenges in the splitting of a Gaussian component is the identification of the critical component. A critical component in a GMM can be defined as the Gaussian kernel, which shows very high error in the approximation of the PDF locally. In this work, the Kolmogorov equation error feedback is used to identify the critical Gaussian components to be split. Because the proposed error measure uses the Kolmogorov equation error, it can be used even when measurement data are sparse or not available at all. A Gaussian component can be split along each of its axes or along the direction of maximum nonlinearity. However, splitting along all directions may lead to an exponential increase in the number of components, and it may introduce unwanted spread in the wrong direction. So, it is important to identify the direction of maximum nonlinearity (i.e., the direction in which the state space motion has maximum departure from the linear approximation of the system) and split along this direction while conserving the first two moments. In this work, we use the nonlinearity index introduced in [16] to find the direction of maximum nonlinearity and, subsequently, the direction of splitting. The conservation of first two moments is used to find the location and amplitude of new components introduced during the splitting process. Furthermore, a combination of the Kolmogorov equation error, the Euclidean distance between two Gaussian components, and their orientation with respect to each other is used to find Gaussian components to be merged in a mixture model. The merging of similar Gaussian components provides a mechanism to control the growth of the mixture model. Every time new Gaussian components are introduced or pruned in the mixture model due to splitting/merging, one needs to recompute the weights of the mixture components to minimize the Kolmogorov equation error. The recomputation of the mixture component weight is necessary because the splitting and merging of mixture components are performed to conserve the first two moments in the local domain represented by the mixture component being split/merged. The recomputation of mixture weights ensures that the Kolmogorov equation error is minimized over the whole domain to provide the best approximation of the PDF, which was not discussed in previous work. In this work, the weights after the splitting/merging process are redistributed by solving a convex optimization problem to satisfy the Kolmogorov equation [64,65,78]. Finally, a sparse optimization-based framework is used to trade off between the GMM accuracy and the number of Gaussian components in a mixture model [79].

The rest of the paper is organized as follows: In the next section, a brief summary of the AGMM method is presented. Section III introduces two different schemes for splitting a Gaussian component into multiple Gaussian components. Section IV presents a scheme to identify and merge similar Gaussian components in a GMM. A sparse AGMM scheme is discussed in Sec. V to trade off between the number of Gaussian kernels and the Kolmogorov equation error. Finally, numerical simulation results are presented in Sec. VI to show the efficacy of the presented ideas, followed by the conclusions in Sec. VII.

II. Adaptive Gaussian Mixture Model

Consider the following continuous-time nonlinear stochastic dynamical system:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(t, \boldsymbol{x}(t)) + \boldsymbol{g}(t, \boldsymbol{x}(t))\boldsymbol{\eta}(t), \qquad \boldsymbol{x}(t_0) = \bar{\boldsymbol{x}}_0 \tag{1}$$

where $\boldsymbol{x}(t) \in R^n$ represents the system state vector. The random function $\boldsymbol{\eta}(t)$ is a stochastic forcing term that represents modeling errors and is modeled as a Gaussian white noise process with the specified correlation function $\boldsymbol{Q}(t) \in R^{n\times n}$. The total uncertainty associated with the state vector $\boldsymbol{x}(t)$ is characterized by the probability density function $p(t, \boldsymbol{x})$. The key objective of this work is to accurately propagate the initial state PDF through the system equations of Eq. (1) over time. The time evolution of the state PDF through a nonlinear continuous dynamical system is provided by the Kolmogorov forward equation, which is also known as the Fokker–Planck–Kolmogorov equation [80]:

$$\frac{\partial}{\partial t} p(t, \boldsymbol{x}) = \mathfrak{L}_{\mathcal{FP}}[p(t, \boldsymbol{x})] \tag{2}$$

where $\mathfrak{L}_{\mathcal{FP}}[\cdot]$ is the Fokker–Planck operator or forward diffusion operator defined as follows:

$$\mathfrak{L}_{\mathcal{FP}}[.] = -\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{x}_i} \boldsymbol{f}(t, \boldsymbol{x})[\cdot] + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \boldsymbol{x}_i \partial \boldsymbol{x}_j} \boldsymbol{g}(t, \boldsymbol{x}) \boldsymbol{Q} \boldsymbol{g}^T(t, \boldsymbol{x})[\cdot] \tag{3}$$

Traditional numerical approaches to solve the Kolmogorov equation, which discretize the space in which the PDF lies, suffer from the "curse of dimensionality." To overcome this obstacle, adaptive Gaussian mixture models were used in prior work [61,64,65] to solve the Kolmogorov equation in a computationally efficient manner. The key idea of the AGMM is to approximate the state PDF by the following GMM approximation (denoted by the caret ^):

$$\hat{p}(\boldsymbol{x}(t)) = \sum_{i=1}^{N} w_t^{(i)} p_{g_i}(\boldsymbol{x}(t) | \boldsymbol{\mu}^{(i)}(t), \boldsymbol{P}^{(i)}(t)) \tag{4}$$

where $\boldsymbol{\mu}^{(i)}(t)$ and $\boldsymbol{P}^{(i)}(t)$ represent the mean and covariance of the ith component of the Gaussian mixture:

$$p_{g_i}(\boldsymbol{x}(t) | \boldsymbol{\mu}^{(i)}(t), \boldsymbol{P}^{(i)}(t)) = \frac{1}{|2\pi \boldsymbol{P}^{(i)}(t)|^{1/2}} \exp\left(-\frac{1}{2} \left(\boldsymbol{x} - \boldsymbol{\mu}^{(i)} \right)^T \left(\boldsymbol{P}^{(i)} \right)^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}^{(i)} \right) \right) \tag{5}$$

Note that $w_t^{(i)}$ denotes the amplitude (or weight) of the ith Gaussian kernel in the mixture model at time t. The positivity and normalization constraint of the state PDF leads to the following constraint equations:

$$\sum_{i=1}^{N} w_t^{(i)} = 1 \quad \text{and} \quad w_t^{(i)} \geq 0, \forall i \tag{6}$$

The mean and covariance of each Gaussian component are propagated from one time step to the next locally while using the propagation steps of the extended Kalman filter (EKF) or the unscented transform (UT) as described in [61]. The propagation steps of the EKF involve linearization of the system dynamics equations around the mean of the Gaussian component, whereas the UT provides sigma points to compute the mean and covariance with second-order accuracy. The weights of the Gaussian kernels are updated at every time step by requiring the sum to satisfy the Kolmogorov equation. The substitution of the GMM approximation in the Kolmogorov equation leads to the following expression for the Kolmogorov equation error:

$$e(t,\boldsymbol{x}) = \frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t} - \mathfrak{L}_{\mathcal{FP}}(\hat{p}(t,\boldsymbol{x})) \tag{7}$$

where, $\frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t}$, is obtained by differentiating Eq. (4) as follows:

$$\begin{aligned}\frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t} &= \frac{\partial}{\partial t}\sum_{i=1}^{N} w_t^{(i)} p_{g_i}(\boldsymbol{x}(t)|\boldsymbol{\mu}^{(i)}(t), \boldsymbol{P}^{(i)}(t)) \\ &= \sum_{i=1}^{N}\left(\frac{\partial w_t^{(i)}}{\partial t} p_{g_i} + w_t^{(i)}\frac{\partial p_{g_i}}{\partial t}\right)\end{aligned} \tag{8}$$

Because $p_{g_i}(\boldsymbol{x}(t)|\boldsymbol{\mu}^{(i)}(t), \boldsymbol{P}^{(i)}(t))$ is a function of $\boldsymbol{\mu}^{(i)}(t)$ and $\boldsymbol{P}^{(i)}(t)$, the equation on the right-hand side can be modified as follows:

$$\frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t} = \sum_{i=1}^{N}\left(\dot{w}_t^{(i)} p_{g_i} + w_t^{(i)}\left(\frac{\partial p_{g_i}}{\partial \boldsymbol{\mu}_t^{(i)}}\frac{\partial \boldsymbol{\mu}_t^{(i)}}{\partial t} + \mathrm{Tr}\left[\frac{\partial p_{g_i}}{\partial \boldsymbol{P}_t^{(i)}}\frac{\partial \boldsymbol{P}_t^{(i)}}{\partial t}\right]\right)\right) \tag{9}$$

where $\boldsymbol{\mu}_t^{(i)} = \boldsymbol{\mu}^{(i)}(t)$ and $\boldsymbol{P}_t^{(i)} = \boldsymbol{P}^{(i)}(t)$. Let $\dot{w}_t^{(i)}$ be approximated as the following finite difference:

$$\dot{w}_t^{(i)} = \frac{w_{t+\Delta t}^{(i)} - w_t^{(i)}}{\Delta t} \tag{10}$$

where Δt is the time difference between successive weight updates. Substituting this value in the previous equation and rearranging the terms, we get the following:

$$\begin{aligned}&\frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t} \\ &= \sum_{i=1}^{N}\left(\frac{1}{\Delta t} w_{t+\Delta t}^{(i)} p_{g_i} + w_t^{(i)}\left(\frac{\partial p_{g_i}}{\partial \boldsymbol{\mu}_t^{(i)}}\dot{\boldsymbol{\mu}}_t^{(i)} + \mathrm{Tr}\left[\frac{\partial p_{g_i}}{\partial \boldsymbol{P}_t^{(i)}}\dot{\boldsymbol{P}}_t^{(i)}\right] - \frac{1}{\Delta t} p_{g_i}\right)\right)\end{aligned} \tag{11}$$

Using vector notation, the terms on the right-hand side can be rewritten as follows:

$$\frac{\partial \hat{p}(t,\boldsymbol{x})}{\partial t} = \frac{1}{\Delta t}\boldsymbol{p}_g^T \boldsymbol{w}_{t+\Delta t} + \boldsymbol{m}_{DT}^T \boldsymbol{w}_t$$

where

$$\begin{aligned} m_{DT}^i &= \frac{\partial p_{g_i}}{\partial \boldsymbol{\mu}_t^{(i)}}\dot{\boldsymbol{\mu}}_t^{(i)} + \mathrm{Tr}\left[\frac{\partial p_{g_i}}{\partial \boldsymbol{P}_t^{(i)}}\dot{\boldsymbol{P}}_t^{(i)}\right] - \frac{1}{\Delta t} p_{g_i} \\ \boldsymbol{p}_g &= \begin{bmatrix} p_{g_1} & p_{g_2} & \cdots & p_{g_N}\end{bmatrix}^T \\ \boldsymbol{w}_t &= \begin{bmatrix} w_t^{(1)} & w_t^{(2)} & \cdots & w_t^{(N)}\end{bmatrix}^T \end{aligned} \tag{12}$$

where m_{DT}^i is the ith element of $\boldsymbol{m}_{DT}$. Similarly, using Eq. (3), $\mathfrak{L}_{\mathcal{FP}}(\hat{p}(t,\boldsymbol{x}))$ in Eq. (7) can be written as follows:

$$\mathfrak{L}_{\mathcal{FP}}(\hat{p}(t,\boldsymbol{x})) = \sum_{i=1}^{N} w_t^{(i)}\mathfrak{L}_{\mathcal{FP}}(p_{g_i}) = \boldsymbol{L}_{\mathrm{FP}}^T \boldsymbol{w}_t$$

where

$$\boldsymbol{L}_{\mathrm{FP}} = \begin{bmatrix}\mathfrak{L}_{\mathcal{FP}}(p_{g_1}) & \mathfrak{L}_{\mathcal{FP}}(p_{g_2}) & \cdots & \mathfrak{L}_{\mathcal{FP}}(p_{g_N})\end{bmatrix}^T \tag{13}$$

Substituting Eqs. (12) and (13) into Eq. (7) and minimizing $e(t,\boldsymbol{x})$ in the least-squares sense leads to the following convex optimization problem [61,64,65]:

$$\min_{w_{t+\Delta t}^i} \boldsymbol{J} = \frac{1}{2}\boldsymbol{w}_{t+\Delta t}^T \boldsymbol{M}_c \boldsymbol{w}_{t+\Delta t} + \boldsymbol{w}_{t+\Delta t}^T \boldsymbol{N}_c \boldsymbol{w}_t$$

subject to

$$\begin{aligned}\boldsymbol{1}_{N\times 1}^T \boldsymbol{w}_{t+\Delta t} &= 1 \\ \boldsymbol{w}_{t+\Delta t} &\geq \boldsymbol{0}_{N\times 1}, \quad i = 1, 2, \ldots, N\end{aligned} \tag{14}$$

where

$$\boldsymbol{M}_c = \frac{1}{\Delta t^2}\int_V \boldsymbol{p}_g \boldsymbol{p}_g^T \,\mathrm{d}\boldsymbol{x} \tag{15}$$

$$\begin{aligned}M_{c_{ij}} &= \frac{1}{\Delta t^2}\left|2\pi\left(\boldsymbol{P}_t^{(i)} + \boldsymbol{P}_t^{(j)}\right)\right|^{-(1/2)} \\ &\times \exp\left[-\frac{1}{2}(\boldsymbol{\mu}_t^{(i)} - \boldsymbol{\mu}_t^{(j)})^T (\boldsymbol{P}_t^{(i)} + \boldsymbol{P}_t^{(j)})^{-1}(\boldsymbol{\mu}_t^{(i)} - \boldsymbol{\mu}_t^{(j)})\right]\end{aligned} \tag{16}$$

$$\boldsymbol{N}_c = \frac{1}{\Delta t}\int_V \boldsymbol{p}_g (\boldsymbol{m}_{DT} - \boldsymbol{L}_{\mathrm{FP}})^T \,\mathrm{d}\boldsymbol{x} \tag{17}$$

It should be noted that $\boldsymbol{N}_c$ has to be computed numerically. The numerical solution involves approximate evaluation of the integral over volume V using Gaussian quadrature, the Monte Carlo method, or unscented transformation as described in earlier work [61]. A critical step in the successful implementation of the AGMM approach is the selection of points at which the local Gaussian approximations are made. Naively chosen points can result in either an unnecessary increase in the number of local approximations or poor approximation of the state PDF globally. If the initial state PDF is Gaussian, then only one component is sufficient at the initial time to accurately represent the state PDF. However, the non-Gaussian effects become prevalent over time and the original Gaussian component must be split to introduce new components in the mixture model approximation. On the other hand, the similar Gaussian components in a mixture model should be merged together to avoid unnecessary computational burden. There may also be some components for which the amplitude (or weight) will be very low. Such components can be dropped without much loss in accuracy. This provides the motivation for the adaptive split/merge algorithm to adaptively select the appropriate components in a mixture model.

III. Adaptive Splitting

In this section, two different splitting mechanisms are discussed to adapt the number of Gaussian kernels in a mixture model. Both the methods make use of the Kolmogorov equation error feedback to select the critical component to be split. One of the approaches split the critical Gaussian components in all directions, whereas the other approach split the critical components only in the direction of high nonlinearity.

A. Splitting Criteria

As discussed earlier, each Gaussian kernel in a GMM captures the state PDF locally. The change in the weight (amplitude) of each component from t to $t + \Delta t$ depends upon the nonlinearity of the system and the number of Gaussian components at time t. The Kolmogorov equation error due to the finite number of Gaussian components can be written as follows:

$$e' = \sum_{i=1}^{N_t} e_{A_i} - e_{B_i} \tag{18}$$

where

$$e_{A_i} = \left(\frac{1}{\Delta t} w_{t+\Delta t}^{(i)} p_{g_i} + w_t^{(i)} \left(\frac{\partial p_{g_i}^T}{\partial \boldsymbol{\mu}_t^{(i)}} \dot{\boldsymbol{\mu}}_t^{(i)} + \mathrm{Tr}\left[\frac{\partial p_{g_i}}{\partial \boldsymbol{P}_t^{(i)}} \dot{\boldsymbol{P}}_t^{(i)}\right] - \frac{1}{\Delta t} p_{g_i}\right)\right)$$

$$e_{B_i} = w_t^{(i)} \mathfrak{L}_{\mathcal{FP}}(p_{g_i}) \tag{19}$$

where N_t is the number of components at time t, and e_{A_i} and e_{B_i} are the error contributions due to the ith Gaussian component as defined in Eqs. (12) and (13). The Kolmogorov equation error is expected to be large for components that are present in the nonlinear region of the actual PDF. These components are defined to be critical components and, hence, candidate components to be split. In this respect, one of the splitting criteria corresponds to split all such components that show errors greater than the allowed threshold [81]:

$$\left\{i_1^\star\right\} = \left\{j: \frac{\int_V (e_{A_j} - e_{B_j})^2 \,\mathrm{d}V}{\int_V (e_{A_j}^2 + e_{B_j}^2)\,\mathrm{d}V} > e_1^\star\right\} \tag{20}$$

where $j = 1, 2, \cdots, N_t$, and $e_1^\star$ is the maximum error allowed before splitting. However, the local errors do not take into account the errors due to other components. There should be a method to find the error in the approximation by a component relative to the errors by other components. If this error crosses a certain threshold for any component, such components also become the candidates for splitting. The following criteria are used to identify such components:

$$\left\{i_2^\star\right\} = \left\{j: \frac{\int_V (e_{A_j} - e_{B_j})^2 \,\mathrm{d}V}{\sum_{i=1}^{N_t} \int_V (e_{A_i} - e_{B_i})^2 \,\mathrm{d}V} > e_2^\star\right\} \tag{21}$$

where $j = 1, 2, \ldots, N_t$. One should keep track of the components common to $\{i_1^\star\}$ and $\{i_2^\star\}$ over time. A component should be split when the same component is common to both of these sets over a time window.

B. Splitting in All Directions

One of the most widely used methods for splitting is based upon the conservation of the first two moments of the Gaussian component being split into N_s number of components. The conservation equations can be summarized as follows:

$$\text{Conservation of weight: } w = \sum_{i=1}^{N_s} w^{(i)}$$

$$\text{Conservation of mean: } w\boldsymbol{\mu} = \sum_{i=1}^{N_s} w^{(i)} \boldsymbol{\mu}^{(i)}$$

$$\text{Conservation of covariance: } w\boldsymbol{P} = \sum_{i=1}^{N_s} w^{(i)} (\boldsymbol{P}^{(i)} + (\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu})(\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu})^T) \tag{22}$$

where $\boldsymbol{\mu}$, $\boldsymbol{P}$, and w are the mean, covariance, and weight of the component being split; and $(w^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{P}^{(i)})$ are the weight, mean, and covariance of the ith Gaussian component formed as a result of splitting, and they need to be determined. Because the number of unknowns is more than the number of equations, ideally, many solutions are possible. However, the constraint of positive definiteness on $\boldsymbol{P}^{(i)}$ makes the solutions of these equations more difficult. Richardson and Green [73] proposed a solution to these equations in one dimension by introducing three free parameters: α, β and u. Zhang et al. [71] extended the solutions to these equations in a multidimensional space where each Gaussian component was split into two components, as given in the following:

$$w^{(1)} = \alpha w \qquad w^{(2)} = (1-\alpha) w, \quad 0 < \alpha < 1$$

$$\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu} - \sqrt{\frac{w^{(2)}}{w^{(1)}}} u \boldsymbol{a}_l \quad \boldsymbol{\mu}^{(2)} = \boldsymbol{\mu} + \sqrt{\frac{w^{(2)}}{w^{(1)}}} u \boldsymbol{a}_l, \quad 0 < u < 1$$

$$\boldsymbol{P}^{(1)} = \frac{w^{(2)}}{w^{(1)}} \boldsymbol{P} + (\beta - \beta u^2 - 1) \frac{w}{w^{(1)}} \boldsymbol{a}_l \boldsymbol{a}_l^T + \boldsymbol{a}_l \boldsymbol{a}_l^T, \quad 0 < \beta < 1$$

$$\boldsymbol{P}^{(2)} = \frac{w^{(1)}}{w^{(2)}} \boldsymbol{P} + (\beta u^2 - \beta - u^2) \frac{w}{w^{(2)}} \boldsymbol{a}_l \boldsymbol{a}_l^T + \boldsymbol{a}_l \boldsymbol{a}_l^T \tag{23}$$

where $\boldsymbol{a}_l$ is the vector corresponding to the direction of maximum uncertainty in the states of the original Gaussian component, i.e., the eigenvector corresponding to the maximum eigenvalue of $\boldsymbol{P}$. Thus, one Gaussian component with amplitude w can be split into two components, with respective amplitudes being $w^{(1)}$ and $w^{(2)}$. Each of the resulting components can iteratively be split to generate more components. Varying these parameters, one gets different solutions, and it is difficult to find the optimal values of these parameters. It is also not obvious as to how the parameters are related to each other and how the solution changes with the variation in these parameters.

In this paper, a new method of splitting is introduced by reducing the number of free parameters, thus making the solution more controllable; i.e., one can control the accuracy in the approximation of the original PDF by changing the parameters. Let us assume that the splitting takes place at any time $t + \Delta t$. So, the equation for the conservation of the second moment can be written as follows:

$$w\boldsymbol{P}_{t+\Delta t} = \sum_{i=1}^{N_s} w^{(i)} \left(\boldsymbol{P}_{t+\Delta t}^{(i)} + \left(\boldsymbol{\mu}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)\left(\boldsymbol{\mu}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)^T\right) \tag{24}$$

If the unscented transform with zero weight for the central sigma point is used to propagate the moments of each Gaussian component, then the corresponding equation for $\boldsymbol{P}_{t+\Delta t}$ is given as follows:

$$\boldsymbol{P}_{t+\Delta t} = \sum_{i=1}^{2n} w'^{(i)} \left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)\left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)^T + \boldsymbol{Q}_k, \;\; w'^{(i)} = \frac{1}{2n}$$

$$\Rightarrow \sum_{i=1}^{2n} \left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)\left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)^T = 2n\boldsymbol{P}_{t+\Delta t} - 2n\boldsymbol{Q}_k \tag{25}$$

where $\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)}$ represents the ith sigma point according to the unscented transformation (see the Appendix), and $\boldsymbol{Q}_k$ is the first-order discrete approximation of the spectral density $\boldsymbol{Q}(t)$ and is given by $\boldsymbol{Q}_k = \Delta t \boldsymbol{Q}(t)$. It should be noted that the unscented transform with zero central weight is equivalent to the cubature transform introduced in [82]. Let us assume that the original component is being split into $2n$ components, and the mean of each of these components is written as follows:

$$\boldsymbol{\mu}_{t+\Delta t}^{(i)} = \boldsymbol{\mu}_{t+\Delta t} + \beta' \left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right); \quad i = 1, 2, \cdots, 2n \tag{26}$$

where β' is another free parameter and determines the spread of new components around the original component mean. Substituting this value in Eq. (24) leads to the following expression:

$$w\boldsymbol{P}_{t+\Delta t} = \sum_{i=1}^{2n} w^{(i)} \left(\boldsymbol{P}_{t+\Delta t}^{(i)} + \beta'^2 \left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)\left(\hat{\boldsymbol{X}}_{t+\Delta t}^{(i)} - \boldsymbol{\mu}_{t+\Delta t}\right)^T\right) \tag{27}$$

Assuming that all $w^{(i)}$ are equal and their magnitude is $w^{(1)}$, the substitution of Eq. (25) into Eq. (27) leads to the following expression:

$$wP_{t+\Delta t} = w^{(1)}\sum_{i=1}^{2n} P^{(i)}_{t+\Delta t} + w^{(1)}\sum_{i=1}^{2n}\beta'^2\left(\hat{X}^{(i)}_{t+\Delta t} - \mu_{t+\Delta t}\right)\left(\hat{X}^{(i)}_{t+\Delta t} - \mu_{t+\Delta t}\right)^T$$

$$\Rightarrow \frac{w}{w^{(1)}}P_{t+\Delta t} = \sum_{i=1}^{2n} P^{(i)}_{t+\Delta t} + 2n\beta'^2 P_{t+\Delta t} - 2n\beta'^2 Q_k \tag{28}$$

Let

$$P^{(i)}_{t+\Delta t} = \frac{P_{t+\Delta t}}{\alpha'} + \beta'^2 Q_k$$

So, the last equation modifies to the following:

$$\frac{w}{w^{(1)}} = 2n\frac{1}{\alpha'} + 2n\beta'^2 \tag{29}$$

The conservation of the weight leads to the following constraint:

$$2n\frac{w^{(1)}}{w} = 1 \quad \Rightarrow \beta'^2 = 1 - \frac{1}{\alpha'} \tag{30}$$

This is because α' is the scaling factor for the covariance of the new components, which should be smaller than the original covariance, and hence $\alpha\prime > 1$. From Eq. (30), one can conclude that $\beta' < 1$. β' determines the spread of split components around the original component, whereas α' determines the shrinkage of the original covariance matrix. The physical intuition behind these splitting parameters can be exploited to have a better control over splitting as compared to the method in [71].

An example of a two-dimensional dynamical system is shown in Fig. 1. The component to be split is shown in Fig. 1a, where $2n$ (i.e., four sigma point locations) are shown by stars. Figure 1b shows the approximation of this component with four smaller components located at

$$\mu^{(i)}_{t+\Delta t} = \mu_{t+\Delta t} + \beta'\left(\int_t^{t+\Delta t} f\left(t, \xi^{(i)}_t\right) dt - \mu_{t+\Delta t}\right)$$

The distance of the centers of the components introduced due to splitting from the center of the original Gaussian components [in this case, (0,0)] depends upon the value of β'. For $\beta' = 0$, the centers of the new components will collapse at the current center [i.e., (0,0)] itself. As the value of β' increases, the centers of the new components move away from the current center. The value of $\beta' = 0.5$ was used to show the splitting mechanism of Fig. 1. It should be noted that a very high value of α' will result in new components with very small covariance, and hence minimal overlap between components. In the limiting case, when $\alpha' \to \infty$, the resulting components will be delta functions located exactly at sigma points provided by the unscented transformation. Hence, the high value for α' will be beneficial to capture multimodal behavior; however, one would need to allow the Gaussian kernels to overlap to capture the highly skewed behavior of the state PDF.

C. Splitting in One Direction

Splitting in all directions may not be suitable when the non-Gaussian effects are more prevalent in a specific direction. Because the purpose of splitting is to approximate the state PDF by a Gaussian kernel in a local region, a better approximation can be obtained if the splitting is performed in the direction of maximum non-Gaussian effects, i.e., maximum nonlinearity. Notice that the direction of maximum nonlinearity may not be the same as the direction of maximum uncertainty provided by the eigenvector of the covariance matrix corresponding to the largest eigenvalue. Junkins and Singla introduced the nonlinearity index to measure the departure from linearity for a particular parameterization of a nonlinear system [16]. This nonlinearity index used the concept of the state transition matrix to measure the departure from linearity and is defined as follows [16,83]:

$$\nu = \max_i \frac{\|\Phi_i(t, t_0) - \bar{\Phi}(t, t_0)\|}{\|\bar{\Phi}(t, t_0)\|} \tag{31}$$

where $\bar{\Phi}(t, t_0)$ is the state transition matrix corresponding to the nominal or expected initial condition denoted by $\bar{x}(t_0) = E\{x_0\}$:

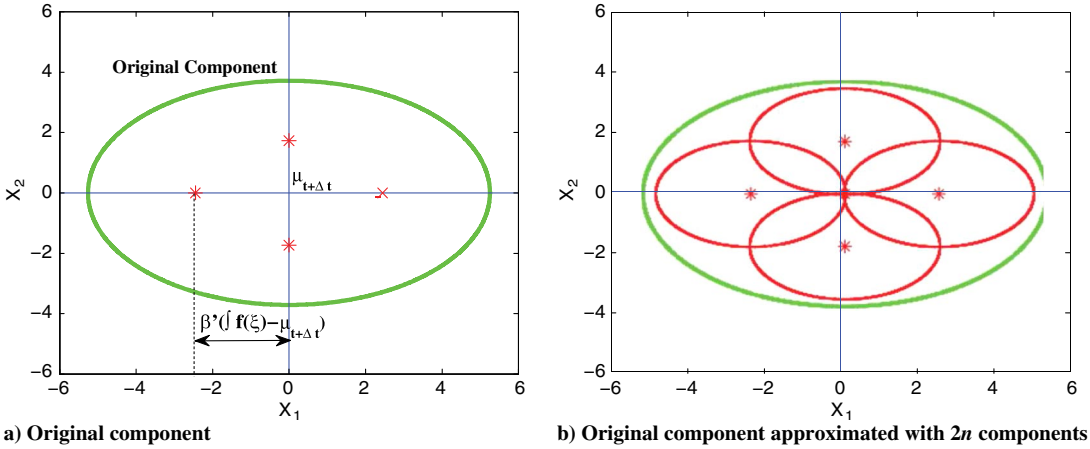
$$\bar{\Phi}(t, t_0) = \left[\frac{\partial \bar{x}(t)}{\partial \bar{x}(t_0)}\right] \triangleq \left[\frac{\partial x(t)}{\partial x(t_0)}\bigg|_{\bar{x}(t)}\right] \tag{32}$$

which satisfies the following equation:

$$\dot{\bar{\Phi}}(t, t_0) = \bar{A}\,\bar{\Phi}(t, t_0); \qquad \bar{A} = \left[\frac{\partial f(t, \bar{x}(t))}{\partial \bar{x}(t)}\right]\bar{\Phi}(t_0, t_0) = I \tag{33}$$

Similarly, the $\Phi_i(t, t_0)$ represents the state transition matrix corresponding to the initial condition in the domain of the initial state PDF. Notice that the nonlinearity index of Eq. (31) always lies between zero and one, and it is zero for linear systems as $\Phi_i(t, t_0) = \bar{\Phi}(t, t_0)$.

To find the direction of maximum nonlinearity for splitting purposes, the candidate directions are chosen among the directions of sigma points provided by the unscented transform for a candidate Gaussian kernel for splitting. If the square root of the covariance matrix is used to generate the sigma point, then sigma points lie along the eigenvectors of the covariance matrix. The nonlinearity index of Eq. (31) is approximately computed with the help of these sigma points rather than using the random points in [16]:



a) Original component b) Original component approximated with $2n$ components

Fig. 1 Splitting mechanism.

$$l = \underset{i=1,2,\cdots,2n}{\arg\max} \frac{\|\boldsymbol{\Phi}_i(t,t_0) - \bar{\boldsymbol{\Phi}}(t,t_0)\|}{\|\bar{\boldsymbol{\Phi}}(t,t_0)\|} \tag{34}$$

Once the nonlinearity index is computed, the direction of maximum nonlinearity is approximated by $\boldsymbol{a}_l$. If one uses the unscented transform to propagate the mean and covariance of each Gaussian component in the GMM, the state transition matrix along each sigma point can be used to approximate the nonlinearity index. Additional computation in finding the direction of maximum nonlinearity will then involve the calculation of the state transition matrix for each sigma point.

Assuming that each Gaussian component is split into $2q$, $q = 1, 2, \cdots$ components, the equations for splitting can be written as follows:

$$\begin{aligned} \boldsymbol{\mu}^{(i)}_{t+\Delta t} &= \boldsymbol{\mu}_{t+\Delta t} + u\boldsymbol{a}_l \quad i = 1, 2, \cdots, q \\ \boldsymbol{\mu}^{(i)}_{t+\Delta t} &= \boldsymbol{\mu}_{t+\Delta t} - u\boldsymbol{a}_l \quad i = q+1, 2, \cdots, 2q \end{aligned} \tag{35}$$

where u is a parameter that determines the spread of the components. Substitution of the aforementioned equations in the conservation equation for $\boldsymbol{P}$ in Eq. (22) and using Eq. (24) leads to the following expression:

$$w\boldsymbol{P}_{t+\Delta t} \quad = \sum_{i=1}^{2q} w^{(i)} \left(\boldsymbol{P}^{(i)}_{t+\Delta t} + u^2 \boldsymbol{a}_l \boldsymbol{a}_l^T \right) \tag{36}$$

Making use of the unscented transform approximation for the covariance matrix with the help of $2n+1$ sigma points, $\boldsymbol{P}_{t+\Delta t|k}$ can be written as follows:

$$\boldsymbol{P}_{t+\Delta t|k} = \sum_{i=0}^{p=2n} W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \tag{37}$$

where $W_c^{(i)}$ are the weights of the ith sigma point $\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t}$ (refer to the Appendix). The central weight $W_c^{(0)}$ can be written explicitly on the right-hand side as follows:

$$\begin{aligned} \boldsymbol{P}_{t+\Delta t|k} = &\sum_{i=1}^{2n} W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T \\ &+ W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \end{aligned} \tag{38}$$

The expression on the right-hand side can be rearranged as follows:

$$\begin{aligned} \boldsymbol{P}_{t+\Delta t|k} = &\frac{1}{2} \left(\sum_{i=1}^{n} 2W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T \right. \\ &\left. + W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \right) \\ &+ \frac{1}{2} \left(\sum_{i=n+1}^{2n} 2W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T \right. \\ &\left. + W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \right) \end{aligned} \tag{39}$$

The last equation represents the splitting of a Gaussian component into two components. If a single component is split into $2q$ ($q = 1, 2, \cdots$) components, the aforementioned equation can be rewritten as follows:

$$\begin{aligned} \boldsymbol{P}_{t+\Delta t|k} = &\sum_{j=1}^{q} \frac{1}{2q} \left(\sum_{i=1}^{n} 2W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \right. \\ &\left. + W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t|k} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T \right) \\ &+ \sum_{j=q+1}^{2q} \frac{1}{2q} \left(\sum_{i=n+1}^{2n} 2W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T + \boldsymbol{Q}_k \right. \\ &\left. + W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t|k} \right) \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t|k} - \boldsymbol{\mu}_{t+\Delta t|k} \right)^T \right) \end{aligned} \tag{40}$$

Comparing the last equation with Eqs. (36) and (37), we get the following:

$$w^{(i)} = \frac{w}{2q} \tag{41}$$

$$\begin{aligned} \boldsymbol{P}^{(i)}_{t+\Delta t} = &\sum_{i=1}^{n} 2W_c^{(i)} \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t} \right) \left(\hat{\boldsymbol{X}}^{(i)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t} \right)^T + \boldsymbol{Q}_k \\ &+ W_c^{(0)} \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t} \right) \left(\hat{\boldsymbol{X}}^{(0)}_{t+\Delta t} - \boldsymbol{\mu}_{t+\Delta t} \right)^T - u^2 \boldsymbol{a}_l \boldsymbol{a}_l^T \end{aligned} \tag{42}$$

$\boldsymbol{P}^{(i)}_{t+\Delta t}$ is guaranteed to be at least positive semidefinite for $0 < u \leq 1$. One can easily verify that Eqs. (35), (41), and (42) satisfy Eq. (22). The criteria for selecting the critical components to be split are the same as mentioned in Eqs. (20) and (21).

D. Reoptimizing the Weights

Every time new Gaussian components are introduced in the mixture model due to splitting of a Gaussian kernel, one needs to recompute the weights of mixture components to minimize the Kolmogorov equation error. The recomputation of the mixture component weight is necessary because the splitting of a mixture component is performed to conserve the first two moments in the local domain represented by the mixture component being split. The recomputation of mixture weights ensures that the Kolmogorov equation error is minimized over the whole domain and provides the best approximation of the PDF. Hence, the weights of all components at $t + \Delta t$ must be calculated by solving the optimization problem mentioned in Eq. (14) while taking into account the change in the number of components formed due to splitting.

The following algorithm summarizes the steps involved in splitting:

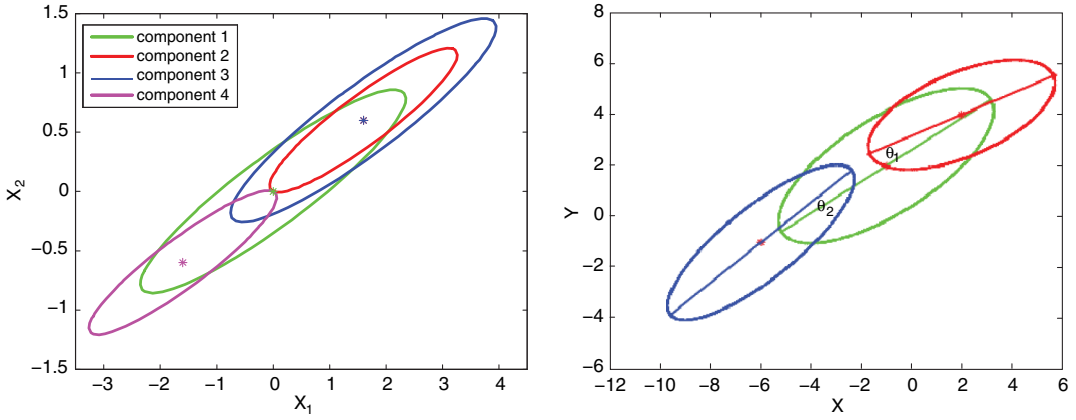
1) Propagate the $\boldsymbol{\mu}^i_t$ and $\boldsymbol{P}^i_t$ of each component using unscented transformation to compute $\boldsymbol{\mu}^i_{t+\Delta t}$ and $\boldsymbol{P}^i_{t+\Delta t}$.
2) Find the error criteria mentioned in Eqs. (20) and (21).
3) Perform splitting, if any.
4) Solve Eq. (14) to get $\boldsymbol{w}_{t+\Delta t}$.
5) Go back to step 1.

These steps are repeated until the final time.

IV. Merging/Pruning of Components

The continuous splitting of the Gaussian components in a mixture model can result in unbounded growth of GMM components. For proper tradeoff between accuracy and the number of components, a pruning process is required to merge similar Gaussian components in a mixture model. At time t, if m out of N_t components are split, $N_t - m$ components are considered for merging. The similarity between two components can be decided on the basis of the Kullback–Leibler divergence between the components. The KL divergence between two PDFs, p_1 and p_2, corresponding to a continuous random variable $\boldsymbol{x} \in \mathbb{R}^n$ is defined as follows [84]:

$$D_{\mathrm{KL}}(p_1 \| p_2) = \int_{-\infty}^{\infty} p_1(\boldsymbol{x}) \log \frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \tag{43}$$



a) Candidates for merging b) Angle between the largesteigenvectors

Fig. 2 Typical components of a Gaussian mixture.

When p_1 and p_2 are Gaussian distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and covariances $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, respectively, an analytical expression for the KL divergence can be obtained as follows [84]:

$$D_{\mathrm{KL}}(p_1\|p_2)=\frac{1}{2}\left(\mathrm{tr}(\boldsymbol{P}_2^{-1}\boldsymbol{P}_1)-n+(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{P}_2^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)-\log\left(\frac{\det(v)}{\det(\boldsymbol{P}_2)}\right)\right) \tag{44}$$

One can see that the KL divergence is invariant under translation of the PDFs, i.e., if the centers of PDFs p_1 and p_2 are moved in the same direction by the same amount, the KL divergence stays the same. This is because this divergence does not depend upon the absolute values of their mean (i.e., center) but rather the difference between the means. However, the KL divergence may not be an appropriate choice for the selection of components for merging. To illustrate this fact, let us consider a GMM depicted in Fig. 2a. In this illustration, components 1 and 3 have a smaller KL distance between themselves than between components 1 and 2. However, merging components 1 and 3 will have a higher increase in covariance than after merging components 1 and 2. Keeping in view the basic principle of Gaussian sum approximation, each component should have a small covariance and, as such, components 1 and 2 should be merged. Furthermore, merging of two Gaussian components, for which the principal axes are misaligned, can result in errors in approximating the skewness of the state PDF. Hence, an ideal merging criteria should take into consideration the angle between eigenvectors of the candidate Gaussian components for the merging. In this work, the angle between the eigenvectors corresponding to the maximum eigenvalues of two candidate Gaussian components for merging is considered, as shown in Fig. 2b.

The candidate components for merging are chosen from $N_t - m$ candidates. The first cutoff for merging corresponds to selecting components that contribute the least toward the Kolmogorov equation error. In this work, this cutoff is chosen to be 10% of $e_2^\star$. For each of these candidates, their closest neighbor is chosen based on the Euclidean distance between the means of the candidate Gaussian components. If the least Euclidean distance (E_{jk}) between any two components with centers $\boldsymbol{\mu}^j$ and $\boldsymbol{\mu}^k$ is less than the minimum standard deviation along any axis of these two components, then they are further considered a candidate for merging. The final cutoff for the merging corresponds to satisfying the angle criteria. In other words, if $\boldsymbol{P}^j$ and $\boldsymbol{P}^k$ denote the covariance matrices of the candidate pair of Gaussian components for merging and θ_{jk} is the angle between the eigenvectors corresponding to the maximum eigenvalues of $\boldsymbol{P}^j$ and $\boldsymbol{P}^k$, then these two components are merged if θ_{jk} is less than a prescribed tolerance. The equations for merging two Gaussian kernels are similar to the equations for splitting, and they can be summarized as follows:

$$\begin{aligned}
&\text{Conservation of weight: } \sum_{i=1}^{2} w^{(i)} = w\\
&\text{Conservation of mean: } \sum_{i=1}^{2} w^{(i)}\boldsymbol{\mu}^{(i)} = w\boldsymbol{\mu}\\
&\text{Conservation of covariance: } \sum_{i=1}^{2} w^{(i)}\left(\boldsymbol{P}^{(i)}+(\boldsymbol{\mu}^{(i)}-\boldsymbol{\mu})(\boldsymbol{\mu}^{(i)}-\boldsymbol{\mu})^T\right) = w\boldsymbol{P}
\end{aligned} \tag{45}$$

where $(w, \boldsymbol{\mu}, \boldsymbol{P})$ are the weight, mean, and covariance of the component after merging. It should be mentioned that, although a Gaussian kernel can be split into multiple kernels, the merging process merges only two Gaussian kernels to a single Gaussian kernel.

V. Sparse AGMM

It is quite possible that adaptive splitting and merging can result in undesirable growth of the GMM. Ideally, one wants the least number of Gaussian components in a GMM while minimizing the Kolmogorov equation error. The optimization of the GMM size (i.e., number of Gaussian kernels) along with the Gaussian kernel parameters (center, covariance, and weights) is a computationally intractable task. However, one can exploit sparse approximation tools [85] to select a set of mixture components from an overcomplete dictionary of Gaussian kernels. This overcomplete dictionary of basis functions can be provided by the splitting and merging process described in the previous sections.

Given a dictionary of Gaussian kernels, the component selection process corresponds to minimizing the cardinality of the amplitude/weight vector of Gaussian kernels in a dictionary along with the Kolmogorov equation error. Ideally, this requires the minimization of the l_0 norm of the weight vector; hence, the cost function of the optimization problem of Eq. (14) can be rewritten as follows:

$$\min_{\boldsymbol{w}_{t+\Delta t}} J_1 = \theta_1 J + \theta_2 \|\boldsymbol{w}_{t+\Delta t}\|_0 \tag{46}$$

where J is the original Kolmogorov equation error performance index as defined in Eq. (14). Also, θ_1 and θ_2 represent the relative weight of the Kolmogorov equation error and the sparse solution in the cost function. However, the l_0-norm optimization leads to a nonconvex optimization problem. On the other hand, the l_1 norm is convex and provides a close approximation to the l_0-norm cost function by making the component weights close to zero. This property of the l_1 norm has been exploited to devise sequential l_1-norm minimization algorithms to approximate the solution of the ideal l_0-norm problem [85,86]. The sequential process involves giving weightage to Gaussian kernels in inverse proportion to their amplitude. This makes sure that coefficients, which are close to zero, are driven to zero in the

subsequent iterations. The sequential l_1-norm minimization is used in this work to achieve an optimal tradeoff between number of components in the GMM and the Kolmogorov equation error as discussed in [79]. The resulting optimization problem can be described as follows, and the resulting process is termed the sparse AGMM:

$$\min_{w^i_{t+\Delta t|k}} \theta_1 J + \theta_2 \sum_{i=1}^{N} \frac{w^i_{t+\Delta t}}{\alpha^i + \epsilon}$$

subject to

$$\mathbf{1}^T_{N\times 1} \boldsymbol{w}_{t+\Delta t|k} = 1$$
$$w^i_{t+\Delta t|k} \geq 0, \qquad i = 1, 2, \ldots, N \tag{47}$$

where ϵ is a very small number to ensure that the denominator does not approach zero. Notice that the α^i represents the penalty for the ith Gaussian kernel. Hence, the sequential sparse optimization process is initialized with $\alpha^i = w^i_t$ so that a Gaussian kernel with small weight can be penalized more in the optimization process. Any Gaussian component weight $w^i_{t+\Delta t}$ less than ϵ is considered to be zero. If the resulting solution is not sparse enough, then the optimization problem is solved again with α^i equal to the previous solution for $w^i_{t+\Delta t}$. Cost functions, as in Eq. (47), are popular in the machine learning community and are known as weighted optimization problems. One such formulation was given in Huber et al. [87]. However, their formulation did not involve any l_1 minimization, and the penalty term added to the original cost function (i.e., J) was nonlinear.

VI. Numerical Simulation

In this section, the problem of uncertainty propagation through the two-body problem is considered to show the effectiveness of the proposed ideas.

A. Two-Body Problem

In particular, the motion of a low-Earth-orbit satellite is considered with following equations of motion:

$$\ddot{\boldsymbol{x}} = -\frac{\mu}{r^3}\boldsymbol{x} \tag{48}$$

where $\boldsymbol{x} = [x \quad y \quad z]$ is the position vector in three-dimensional Cartesian space, $r = \sqrt{x^2 + y^2 + z^2}$ is the Euclidean distance between the center of the Earth and the satellite, and $\mu = 398,601.2\ \text{km}^3/\text{s}^2$ is the gravitational constant for the Earth. For simulation purposes, the effects of any nonconservative forces are ignored. At time $t = 0$, the states (i.e., position and velocity of the satellite) are assumed to have Gaussian distribution with the following mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{P}_0$:

$$\boldsymbol{\mu}_0 = [7000 \quad 0 \quad 0 \quad 0 \quad -1.0374 \quad 7.4771]$$
$$\boldsymbol{P}_0 = \text{diag}(0.01, 0.01, 0.01, 10^{-6}, 10^{-6}, 10^{-6}) \tag{49}$$

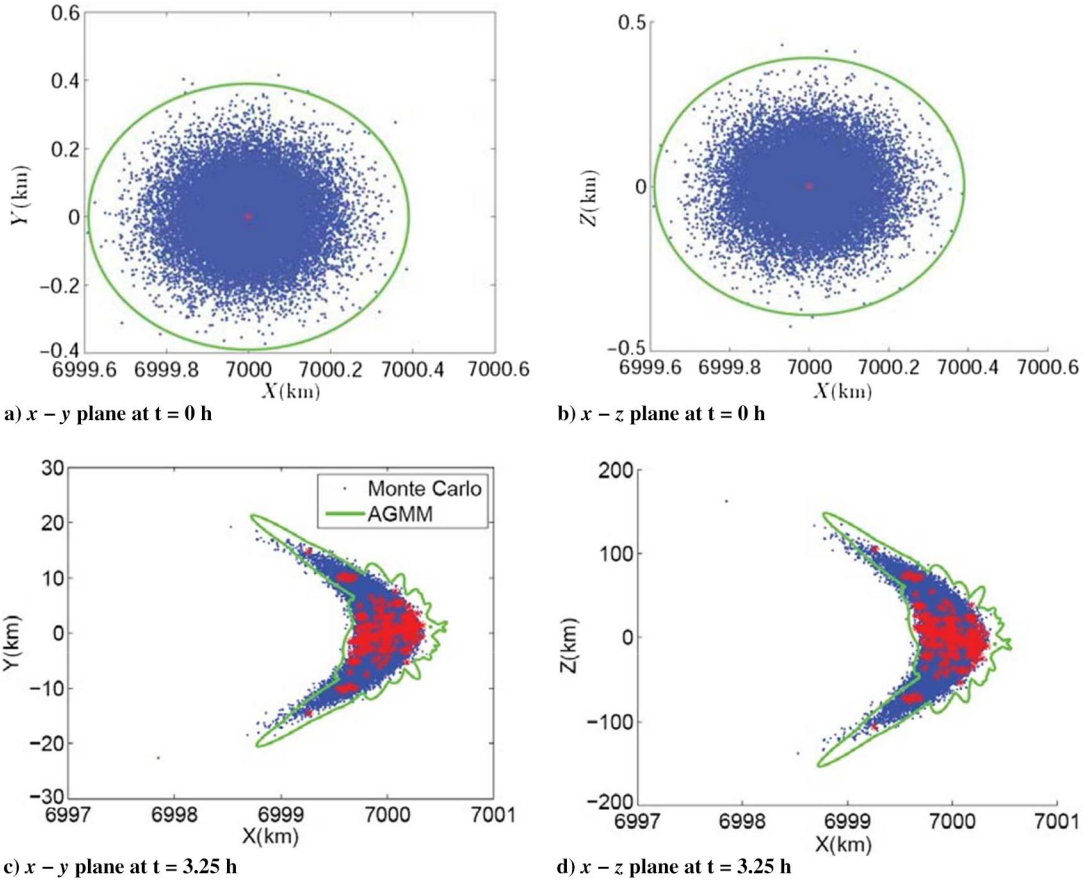
The first three components of $\boldsymbol{\mu}_0$ corresponding to the position of the satellite are in kilometers, whereas the remaining three velocity components are in kilometers per second. Similarly, the first three components of $\boldsymbol{P}_0$ corresponding to variance in position variables are given in squared kilometers, whereas the remaining three components corresponding to the variance of velocity variables are listed in squared kilometers per squared seconds. The motion of the satellite is studied for the next 3.25 h (approximately two orbits). For simulation purposes, the value of splitting criteria parameters $e^\star_1$ and $e^\star_2$ are chosen to be 0.01. Notice that $e^\star_1$ and $e^\star_2$ represent the maximum relative Kolmogorov equation error contribution for a specific Gaussian component; hence, the mixture model accuracy improves with the smaller values for $e^\star_1$ and $e^\star_2$. However, too small of values for $e^\star_1$ and $e^\star_2$ can lead to more frequent splitting of many Gaussian kernels and result in a large size of the mixture model. In this work, the values of these parameters are chosen to maintain a balance between PDF accuracy and the mixture model size. One can even adapt the values of these parameters over time to control the growth of the mixture model to achieve desired PDF accuracy. Figure 3 shows the contours of the state PDF in x–y and x–z planes at the initial and final times. The stars in the figure represent the centers of different Gaussian kernels in the GMM, whereas dots correspond to 50,000 Monte Carlo simulations. It should be noted that the state PDF at the final time is highly skewed, although the initial state PDF is Gaussian in nature. The high skewness in the state PDF is due to the fact that the motion of the satellite represents a conservative system, and hence each realization from the initial state PDF corresponds to an orbit with a different time period. Some of these realizations correspond to faster orbits and some to the slower orbits, resulting in a highly skewed state PDF. Each Gaussian kernel was split into 12 components whenever the splitting criteria were satisfied, resulting in a total of 144 components at the final time while following the development in Sec. III.B. These results clearly illustrate the effectiveness of the AGMM with splitting in all directions to capture the non-Gaussian effects accurately.

B. Splitting in Single Direction

Furthermore, the procedure outlined in Sec. III.C was used to split the candidate Gaussian kernels in a GMM in the direction of maximum nonlinearity. F and G solutions as outlined in [88] were used to compute the nonlinearity index of Eq. (31) to identify the direction of maximum nonlinearity at each time.

The values for splitting criteria $e^\star_1$ and $e^\star_2$ are chosen to be $1e-7$ and $1.5(1/N_t)$. Because the growth of the Gaussian mixture model is moderate in the case of splitting in one direction as compared to the case of splitting in all directions, the splitting criteria parameters are chosen to be more stringent. A single Gaussian kernel is used to capture the initial Gaussian PDF exactly. As time progresses, the PDF becomes non-Gaussian and splitting needs to be performed. At any time, it is assumed that the relative residual errors given in Eq. (21) are uniformly distributed over all the Gaussian kernels in a GMM. Once a candidate for splitting is identified, it is tracked for the next few time steps to make sure that it consistently satisfies the splitting criteria. If the splitting condition of Eq. (21) is satisfied for the next six time steps, the component is split into two components according to Eqs. (35), (41), and (42). If the splitting criteria are first satisfied at time t_2 during the window of six consecutive time steps from t_2 to $t_2 + 6\Delta t$, then the actual splitting is performed at previous splitting time t_1 to minimize the error accumulated between two splitting times t_1 and t_2. Figure 4a shows the schematic of splitting. Once the component to be split is identified at t_2, it is traced back at previous time t_1, which is the last time the splitting of a component was performed. Hence, the splitting is performed using $t = t_1$ in Eqs. (35), (41), and (42). It is noted that, even if the component is split at $t_1 + \Delta t$, the new components are included in the mixture model only at t_2. Weights of the components are updated once every minute ($\Delta t = 1$ min) to minimize the Kolmogorov equation error. Figure 4b shows the evolution of the number of components over a period of 3.25 h. The number of Gaussian kernels in the GMM grow from one at an initial time of $t = 0$ to about 150 at a final time of $t = 3.25$ h. It indicates the growth in the non-Gaussian nature of the PDF of the states. Figure 5 shows the PDF contours in the x–y and x–z planes at the beginning and end of the simulation. Once again, 50,000 MC runs were used to validate the results. Form these plots, it is clear that, although the GMM with adaptive splitting in one direction is able to capture the overall non-Gaussian effect, the accuracy of the GMM deteriorates in capturing the tails of the state PDF as compared to the splitting in all directions.

To further improve the accuracy of the GMM approximation, each Gaussian component is split into 6, 8, and 10 components rather than two while making use of Eqs. (41) and (42). Figure 6 shows the contour of the state PDF at the final time in the x–y, x–z, $\dot{x}$–$\dot{y}$, and $\dot{x}$–$\dot{z}$ planes when a Gaussian component is split into six components. One can see that the accuracy in this case is better than the results corresponding to splitting into two components, as shown in Fig. 5. In particular, the contour plots in the x–y and x–z planes clearly show that the splitting in six components results in better approximation of the ground truth PDF represented by MC runs without increasing the



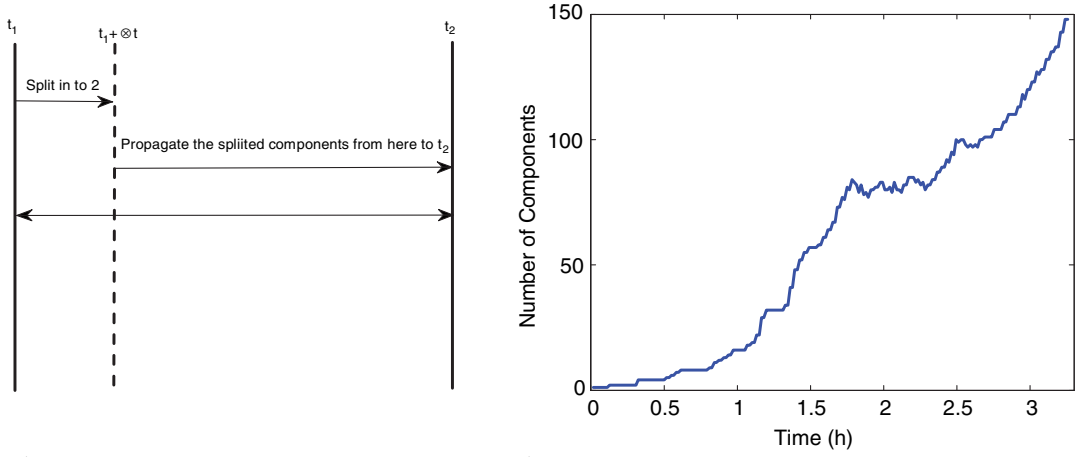
a) $x - y$ plane at t = 0 h

b) $x - z$ plane at t = 0 h

c) $x - y$ plane at t = 3.25 h

d) $x - z$ plane at t = 3.25 h

Fig. 3 Split in 12 components in all directions: state PDF contours: MC (dots), AGMM (thick line), and Gaussian kernel centers (stars).



a) Splitting mechanism

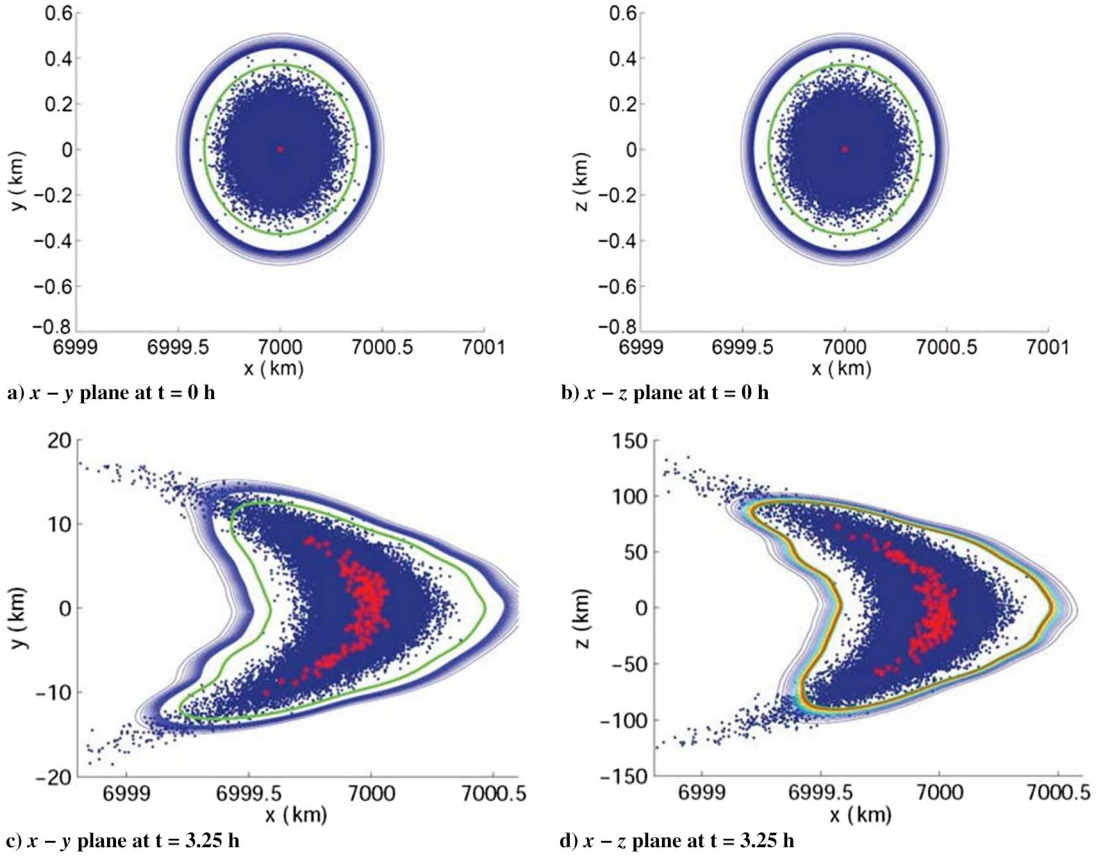
b) Growth in components overtime

Fig. 4 Splitting of components over time.

variance around the MC points. Figures 7 and 8 show the contours of the state PDF at the final time when a Gaussian component is split into 8 and 10 Gaussian components, respectively. Figure 9 shows the evolution of Gaussian components over time when a Gaussian component is split into 6, 8, or 10 components in the direction of maximum nonlinearity. It is clear from these results that state PDF approximation improves with the increase in the number of splitting components without the increase in size of the Gaussian mixture model. The decrease in Gaussian components from splitting in six components to splitting in 8 and 10 components with regard to Fig. 9 can be attributed to infrequent splitting due to better maintenance of the Kolmogorov equation error. From Figs. 5–8, it is clear that the centers of the Gaussian kernels have more tendency to align themselves with the skewness of the state PDF when splitting is performed in the direction of maximum nonlinearity, as compared to the center locations when splitting is performed in all directions (see Fig. 3). This is the indication that we are able to compute the direction of maximum nonlinearity with good accuracy. It should also be noted that the uncertainty in $\dot{y}$ and $\dot{z}$ is almost negligible as compared to uncertainty in the $\dot{x}$ direction. This leads to the Gaussian kernels having minimal spread in the $\dot{y}$ and $\dot{z}$ directions.

C. Merging/Pruning of Components

Furthermore, the merging procedure outlined in Sec. IV was used in conjunction with adaptive splitting in one direction to judiciously



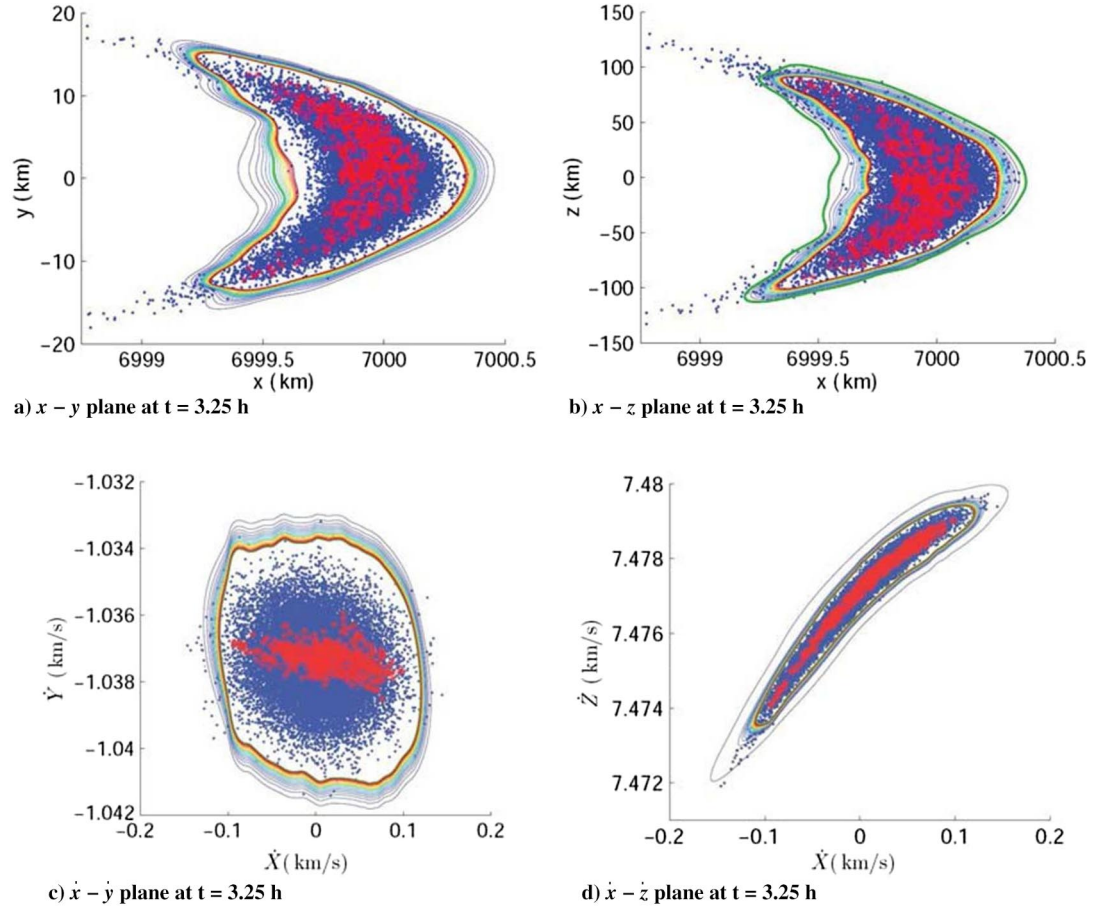
a) $x - y$ plane at t = 0 h

b) $x - z$ plane at t = 0 h

c) $x - y$ plane at t = 3.25 h

d) $x - z$ plane at t = 3.25 h

Fig. 5 Split in two components in one direction: state PDF contours [MC (dots), AGMM (thick line), and Gaussian kernel centers (stars)].



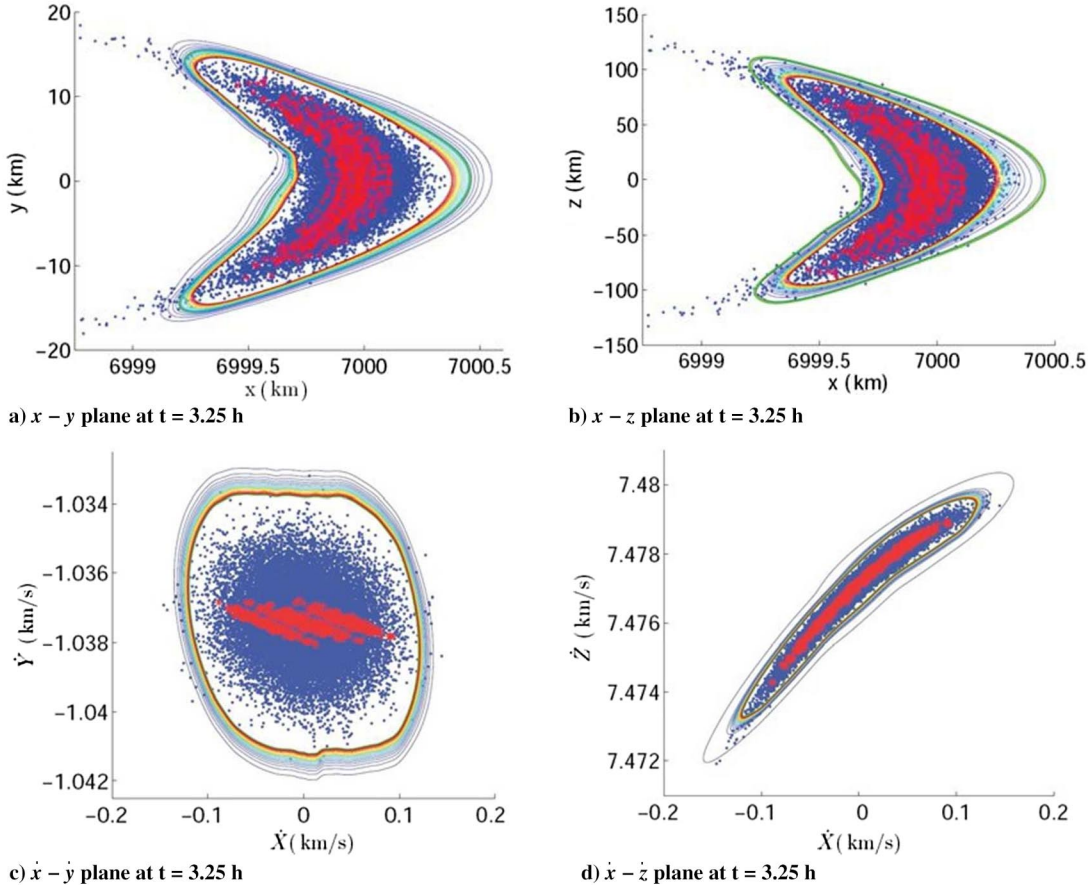
a) $x - y$ plane at t = 3.25 h

b) $x - z$ plane at t = 3.25 h

c) $\dot{x} - \dot{y}$ plane at t = 3.25 h

d) $\dot{x} - \dot{z}$ plane at t = 3.25 h

Fig. 6 Split in six components in one direction: state PDF contours [MC (dots), AGMM (thick line), and Gaussian kernel centers (stars)].



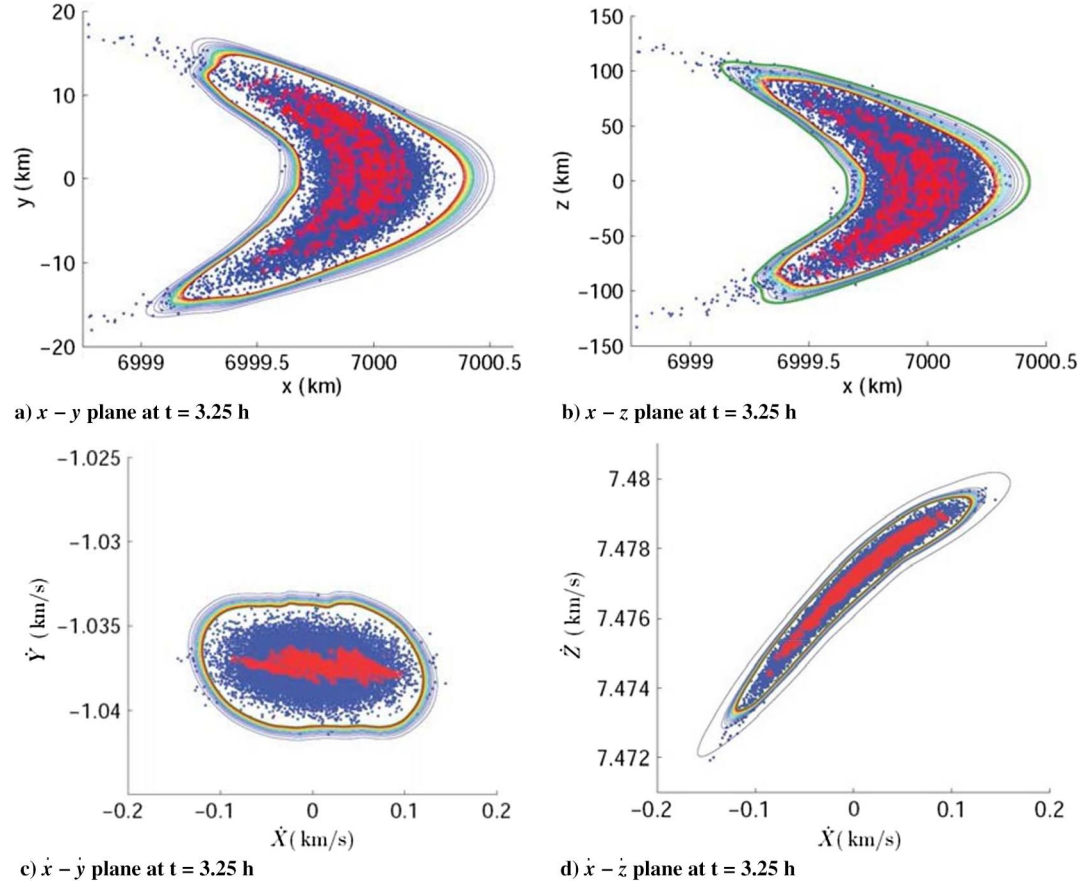
a) $x - y$ plane at t = 3.25 h

b) $x - z$ plane at t = 3.25 h

c) $\dot{x} - \dot{y}$ plane at t = 3.25 h

d) $\dot{x} - \dot{z}$ plane at t = 3.25 h

Fig. 7 Split in eight components in one direction: state PDF contours [MC (dots), AGMM (thick line), and Gaussian kernel centers (stars)].



a) $x - y$ plane at t = 3.25 h

b) $x - z$ plane at t = 3.25 h

c) $\dot{x} - \dot{y}$ plane at t = 3.25 h

d) $\dot{x} - \dot{z}$ plane at t = 3.25 h

Fig. 8 Split in 10 components in one direction: state PDF contours [MC (dots), AGMM (thick line) and Gaussian kernel centers (stars)].

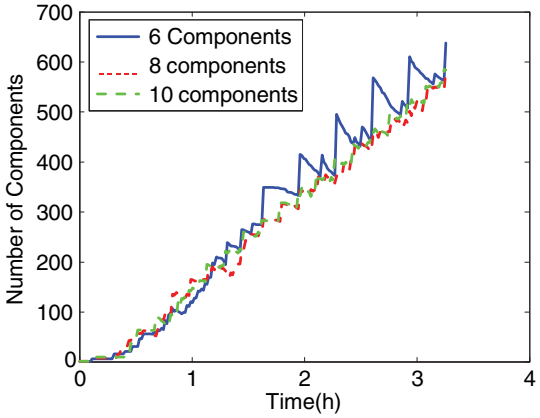


Fig. 9 Evolution of GMM over time.

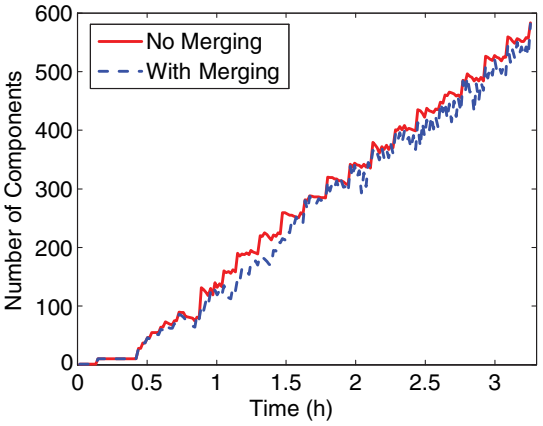


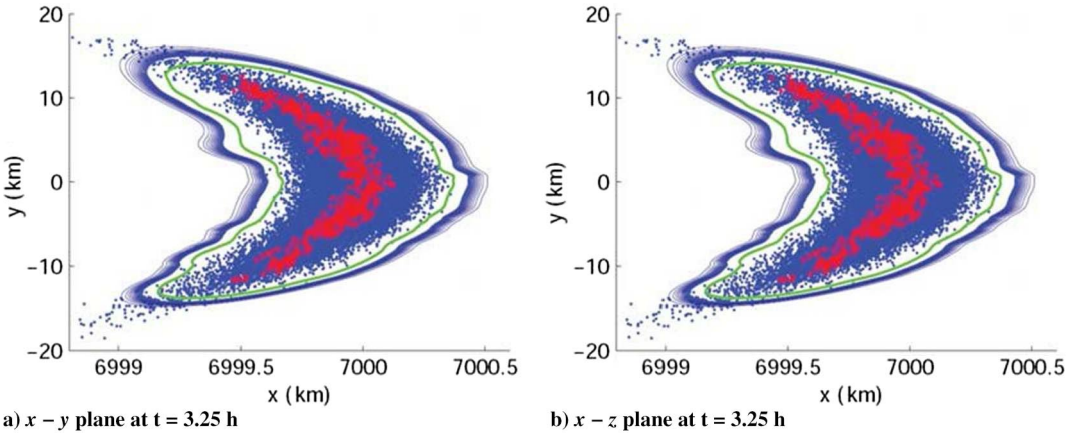
Fig. 10 GMM evolution with split/merge (10 components split).

select the number of components in the GMM. The cutoff angle between two eigenvectors corresponding to the covariance matrices of Gaussian kernels being considered for merging was chosen to be 0.1 rad. Figure 10 shows the evolution of Gaussian components in the GMM over time and Fig. 11 shows the contour plots of the state PDF when both splitting and merging were performed. As expected, the merging procedure did reduce the number of Gaussian components in the mixture model as compared to the splitting alone without the loss in accuracy. However, the reduction in the number of components was minimal as compared to the number of components when merging was not performed. This was due to the fact that, most of the time, the merging was immediately followed by splitting to maintain the desired Kolmogorov equation error.

In particular, it is noticed that the merging and pruning of components with very small weights (i.e., components with weights less than 10^{-8}) is not sufficient to lower the number of Gaussian kernels over time. Splitting of components takes place over time to account for the increase in the non-Gaussian nature of the PDF and, more importantly, to capture the tail of the state PDF. Because the weights of the components are calculated by minimizing the quadratic norm in Eq. (14), the result is sensitive to outliers; and many small weights, which capture the tail of the PDF, are carried over time. The number of components in the GMM can further be optimized by following the sparse AGMM development in Sec. V.

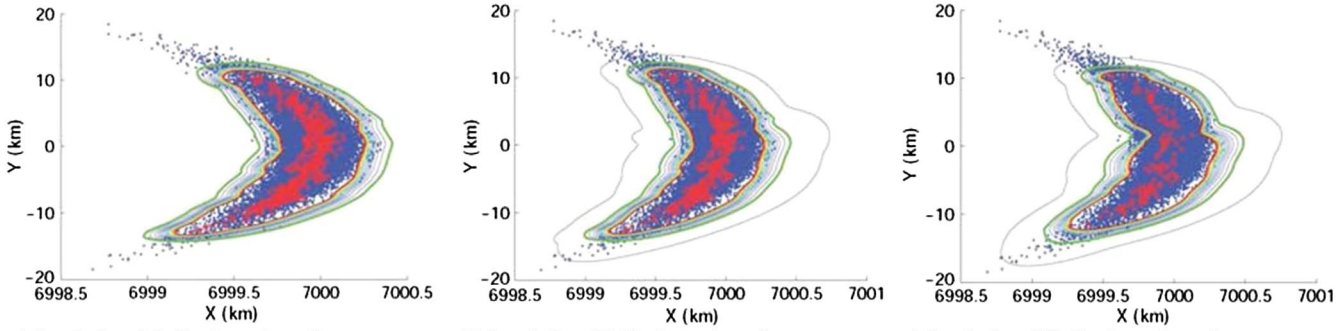
D. Pruning of Components Using l_1 Minimization

Figure 12 shows the performance of the sparse AGMM for different values of θ_1 and θ_2 at $t = 3.24$ h. Weights are computed from $t = 0$ to $t = 3.24$ h by solving the optimization problem in Eq. (14), whereas from $t = 3.24$ h to $t = 3.25$ h, they are computed by solving the sparse approximation problem of Eq. (47). As expected, the number of components is reduced and the approximation of the tail region of the PDF becomes less accurate with the increase in the value of θ_2. However, the sparse AGMM provides an efficient mechanism to trade off between the state PDF



a) $x - y$ plane at t = 3.25 h b) $x - z$ plane at t = 3.25 h

Fig. 11 Split and merge: state PDF contours [MC (dots), AGMM (thick line), and Gaussian kernel centers (stars)].



a) $\theta_1 = 1, \theta_2 = 0.1$, final number of components = 503 b) $\theta_1 = 1, \theta_2 = 10$, final number of components = 347 c) $\theta_1 = 1, \theta_2 = 100$, final number of components = 132

Fig. 12 Sparse AGMM: state PDF contours [MC (dots), AGMM (thick line), and Gaussian kernel centers (stars)].

approximation error and the number of components in a mixture model. For example, one may want to use θ_2 close to 10 to reduce the number of components in the GMM with an acceptable accuracy.

Finally, it should be mentioned that the simulation results presented in this paper were generated on a MATLAB platform on an IBM server with 24 cores. The processor time was almost 2 h for 3.25 h of simulation time, including the pruning of components with l_1 minimization. Although the time propagation of the means and covariances of different GMM components was parallelized, no parallelization was performed for splitting, merging, and l_1-minimization-based pruning. One can further optimize the computational performance of the proposed methodology by parallelization of the splitting, merging, and l_1-minimization-based pruning of the GMM components. Open-source convex optimization tools such as CMSoft's optimization resources [89,90] can be used to solve the convex optimization problems, resulting from the Kolmogorov equation error minimization as well as l_1 minimization, in a computationally efficient manner.

VII. Conclusions

An adaptive splitting and merging process is developed for dynamic selection of Gaussian kernels in a Gaussian mixture model to accurately propagate the state probability density function through a nonlinear system while solving the Kolmogorov equation. The splitting and merging of Gaussian kernels in a GMM are competing processes because splitting increases the number of components, whereas merging may increase the state probability density function (PDF) approximation error. The process for splitting and merging exploits the Kolmogorov equation error to define criteria to identify the critical components in the mixture model to be split or merged. Two different splitting mechanisms are discussed in this work. The first splitting mechanism corresponds to splitting one Gaussian kernel in all directions, whereas the second splitting mechanism corresponds to splitting in only the direction of maximum nonlinearity. The state transition matrix in conjunction with unscented transformation is used to compute the departure from linearity, and hence the direction of maximum nonlinearity. The merging mechanism exploits the angle between eigenvectors corresponding to the maximum eigenvalue of covariance matrices corresponding to two different Gaussian kernels to find candidate components for merging. The Kolmogorov equation error feedback is used to optimize the weights of the GMM after the splitting or merging of Gaussian kernels. Finally, a sparse optimization-based framework is used to trade off between the GMM accuracy and the number of Gaussian components in a mixture model. The sparse optimization approach solves an l_1-norm convex optimization problem to carry out the selection of Gaussian kernels while simultaneously minimizing the Kolmogorov equation error. Numerical results corresponding to initial state uncertainty propagation through the two-body model are presented to provide evidence in support of the efficacy of the proposed approach in accurately capturing the non-Gaussian state PDF.

Appendix: Unscented Transform

Consider the discrete-time nonlinear dynamical system described by the following equation:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{\phi}(k, \boldsymbol{x}_k) + \boldsymbol{\eta}_k \\ \boldsymbol{\eta}_k &= \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k) \end{aligned} \tag{A1}$$

UT involves propagation of a set of carefully selected points, called sigma points, through the system model. The transformed points can then be used to estimate the first two moments of the system. In general, for an n-dimensional system, the following $2n + 1$ points are used to exactly match moments of the states up to the second order:

$$\mathcal{X}_k = \left[\boldsymbol{\mu}_k \quad \boldsymbol{\mu}_k + \sqrt{(n+\kappa)\boldsymbol{P}_k} \quad \boldsymbol{\mu}_k - \sqrt{(n+\kappa)\boldsymbol{P}_k}\right]$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{P}_k$ are the mean and covariance of the states at time instant k. They can be calculated using the following equations of the unscented Kalman Filter [39]:

Prediction stage:

$$\hat{\boldsymbol{X}}_{k+1}^{(i)} = \boldsymbol{\phi}\left(t, \mathcal{X}_k^{(i)}\right) \tag{A2}$$

$$\boldsymbol{\mu}_{k+1|k} = \sum_{i=0}^{2n} W_m^{(i)} \hat{\boldsymbol{X}}_{k+1}^{(i)} \tag{A3}$$

$$\boldsymbol{P}_{k+1|k} = \sum_{i=0}^{2n} W_c^{(i)} \left(\hat{\boldsymbol{X}}_{k+1}^{(i)} - \boldsymbol{\mu}_{k+1|k}\right)\left(\hat{\boldsymbol{X}}_{k+1}^{(i)} - \boldsymbol{\mu}_{k+1|k}\right)^T + \boldsymbol{Q}_k \tag{A4}$$

Calculation of weights:

$$\begin{aligned} W_m^{(0)} &= \frac{\lambda}{n+\lambda} \qquad W_c^{(0)} = \frac{\lambda}{(n+\lambda)} + (1 - \alpha^2 + \beta) \\ W_m^{(i)} &= \frac{1}{2(n+\lambda)} \qquad W_c^{(i)} \frac{1}{2(n+\lambda)}; \qquad i = 1, 2, \ldots, 2n \end{aligned} \tag{A5}$$

where $\lambda = \alpha^2(n + \kappa) - n$. Note that α determines the spread of sigma points and is a small positive value, i.e., $1 \times 10^{-4} \leq \alpha \leq 1$. β is used to incorporate prior knowledge of the distribution and provides optimal estimation for a value of two in the case of Gaussian distribution. Also, κ exploits the value of higher-order moments and minimizes the mean-squared error up to the fourth order for $\kappa = 3 - n$.

Acknowledgments

This material is based upon the work jointly supported by the National Science Foundation under award no. CMMI-1054759 and the U.S. Air Force Office of Scientific Research grant FA9550-15-1-0313.

References

[1] Risken, H., *The Fokker–Planck Equation: Methods of Solution and Applications*, Springer, New York, 1989, pp. 1–12, 32–62.

[2] Chodas, P. W., and Yeomans, D. K., "Orbit Determination and Estimation of Impact Probability for Near Earth Objects," *Proceedings of the Guidance and Control Conference*, Vol. 101, Advances in the Astronautical Sciences, Univelt, San Diego, CA, 1999, pp. 21–40.

[3] Vittaldev, V., and Russell, R. P., "Space Object Collision Probability via Monte Carlo on the Graphics Processing Unit," *Journal of the Astronautical Sciences*, Vol. 64, No. 3, 2017, pp. 285–309. doi:10.1007/s40295-017-0113-9

[4] Adurthi, N., and Singla, P., "A Conjugate Unscented Transformation Based Approach for Accurate Conjunction Analysis," *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 9, Sept. 2015, pp. 1642–1658. doi:10.2514/1.G001027

[5] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "A Decision-Centric Framework for Density Forecasting," *Journal of Advances in Information Fusion*, Vol. 55, No. 2, Dec. 2010, pp. 73–87.

[6] Madankan, R., Pouget, S., Singla, P., Bursik, M., Dehn, J., Jones, M., Patra, A., Pavolonis, M., Pitman, E., and Singh, T., and Webley, P., "Computation of Probabilistic Hazard Maps and Source Parameter Estimation for Volcanic Ash Transport and Dispersion," *Journal of Computational Physics*, Vol. 271, Aug. 2014, pp. 39–59. doi:10.1016/j.jcp.2013.11.032

[7] Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics*, MIT Press, Cambridge, MA, 2005, pp. 91–119.

[8] Micheal, F., and Johnson, M. D., "Derivative Pricing with Non-Linear Fokker–Planck Dynamics," *Physica A*, Vol. 324, Nos. 1–2, 2003, pp. 359–365. doi:10.1016/S0378-4371(02)01906-4

[9] Polidori, D. C., and Beck, J. L., "Approximate Solutions for Nonlinear Vibration Problems," *Probabilistic Engineering Mechanics*, Vol. 11, No. 3, 1996, pp. 179–185. doi:10.1016/0266-8920(96)00011-2

[10] Singh, T., Singla, P., and Konda, U., "Polynomial Chaos Based Design of Robust Input Shapers," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 132, No. 5, 2010, Paper 051010. doi:10.1115/1.4001793

[11] Wang, M. C., and Uhlenbeck, G., "On the Theory of Brownian Motion II," *Reviews of Modern Physics*, Vol. 17, Nos. 2–3, 1945, pp. 323–342. doi:10.1103/RevModPhys.17.323

[12] Sjöberg, P., Lötstedt, P., and Elf, J., "Fokker–Planck Approximation of the Master Equation in Molecular Biology," *Computing and Visualization in Science*, Vol. 12, No. 1, 2009, pp. 37–50.

[13] Daum, F. E., "A New Nonlinear Filtering Formula for Discrete Time Measurements," *24th IEEE Conference on Decision and Control, 1985*, Vol. 24, IEEE Publ., Piscataway, NJ, Dec. 1985, pp. 1957–1958.

[14] Daum, F. E., "A New Nonlinear Filtering Formula Non-Gaussian Discrete Time Measurements," *25th IEEE Conference on Decision and Control, 1986*, Vol. 25, IEEE Publ., Piscataway, NJ, Dec. 1986, pp. 1030–1031.

[15] Junkins, J. L., "Adventures in the Interface of Dynamics and Control," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 6, 1997, pp. 1058–1071. doi:10.2514/2.4176

[16] Junkins, J. L., and Singla, P., "How Nonlinear Is It? A Tutorial on Nonlinearity of Orbit and Attitude Dynamics," *Journal of the Astronautical Sciences*, Vol. 52, No. 1, 2004, pp. 7–60.

[17] Junkins, J. L., Akella, M. R., and Alfriend, K. T., "Non-Gaussian Error Propagation in Orbital Mechanics," *Journal of the Astronautical Sciences*, Vol. 44, No. 4, Oct.–Dec. 1996, pp. 541–563.

[18] Fuller, A. T., "Analysis of Nonlinear Stochastic Systems by Means of the Fokker–Planck Equation," *International Journal of Control*, Vol. 9, No. 6, 1969, pp. 603–655. doi:10.1080/00207176908905786

[19] Benes, V. E., "Exact Finite-Dimensional Filters for Certain Diffusions with Nonlinear Drift," *Stochastics*, Vol. 5, Nos. 1–2, 1981, pp. 65–92. doi:10.1080/17442508108833174

[20] Daum, F. E., "New Nonlinear Filters and Exact Solutions of the Fokker–Planck Equation," *American Control Conference, 1986*, IEEE Publ., Piscataway, NJ, 1986, pp. 884–888.

[21] Wei-Qui, Z., "Exact Solutions for Stationary Responses of Several Classes of Nonlinear Systems to Parametric and/or External White Noise Excitations," *Applied Mathematics and Mechanics*, Vol. 11, No. 2, 1990, pp. 165–175. doi:10.1007/BF02014541

[22] Cai, G., and Lin, Y. K., "On Exact Stationary Solutions of Equivalent Non-Linear Stochastic Systems," *International Journal of Non-Linear Mechanics*, Vol. 23, No. 4, 1988, pp. 315–325. doi:10.1016/0020-7462(88)90028-5

[23] Chen, L., and Sun, J.-Q., "The Closed-Form Solution of the Reduced Fokker–Planck–Kolmogorov Equation for Nonlinear Systems," *Communications in Nonlinear Science and Numerical Simulation*, Vol. 41, Dec. 2016, pp. 1–10. doi:10.1016/j.cnsns.2016.03.015

[24] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Series D, Vol. 82, No. 1, 1960, pp. 35–45. doi:10.1115/1.3662552

[25] Kolmogorov, A. N., "On the Shannon Theory of Information Transmission in the Case of Continuous Signals," *IRE Transactions on Information Theory*, Vol. 2, No. 4, 1956, pp. 102–108. doi:10.1109/TIT.1956.1056823

[26] Lambert, H. C., Daum, F. E., and Weatherwax, J. L., "A Split-Step Solution of the Fokker–Planck Equation for the Conditional Density," *Fourtieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC'06*, IEEE Publ., Piscataway, NJ, Oct.–Nov. 2006, pp. 2014–2018.

[27] Kumar, M., Singla, P., Chakravorty, S., and Junkins, J. L., "A Multi-Resolution Approach for Steady State Uncertainty Determination in Nonlinear Dynamical Systems," *38th Southeastern Symposium on System Theory*, IEEE Publ., Piscataway, NJ, 2006, pp. 344–348.

[28] Kumar, M., Singla, P., Chakravorty, S., and Junkins, J. L., "The Partition of Unity Finite Element Approach to the Stationary Fokker–Planck Equation," *2006 AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, AIAA Paper 2006-6285, Aug. 2006.

[29] Muscolino, G., Ricciardi, G., and Vasta, M., "Stationary and Non-Stationary Probability Density Function for Non-Linear Oscillators," *International Journal of Non-Linear Mechanics*, Vol. 32, No. 6, 1997, pp. 1051–1064. doi:10.1016/S0020-7462(96)00134-5

[30] Paola, M. D., and Sofi, A., "Approximate Solution of the Fokker–Planck–Kolmogorov Equation," *Probabilistic Engineering Mechanics*, Vol. 17, No. 4, 2002, pp. 369–384. doi:10.1016/S0266-8920(02)00034-6

[31] Kumar, M., Chakravorty, S., Singla, P., and Junkins, J. L., "The Partition of Unity Finite Element Approach with hp-Refinement for the Stationary Fokker–Planck Equation," *Journal of Sound and Vibration*, Vol. 327, Nos. 1–2, 2009, pp. 144–162. doi:10.1016/j.jsv.2009.05.033

[32] Sun, Y., and Kumar, M., "Numerical Solution of High Dimensional Stationary Fokker–Planck Equations via Tensor Decomposition and Chebyshev Spectral Differentiation," *Computers and Mathematics with Applications*, Vol. 67, No. 10, 2014, pp. 1960–1977. doi:10.1016/j.camwa.2014.04.017

[33] Daum, F. E., "Bounds on Performance for Multiple Target Tracking," *IEEE Transactions on Automatic Control*, Vol. 35, No. 4, April 1990, pp. 443–446. doi:10.1109/9.52299

[34] Daum, F., Yau, S.-T., and Yau, S., "Comments on "Finite-Dimensional Filters with Nonlinear Drift" [and Addendum]," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 34, No. 2, April 1998, pp. 689–692. doi:10.1109/7.670361

[35] Daum, F. E., "Exact Solution to the Zakai Equation for Certain Diffusions," *24th IEEE Conference on Decision and Control, 1985*, Vol. 24, IEEE Publ., Piscataway, NJ, Dec. 1985, pp. 1964–1965.

[36] Daum, F. E., "Exact Finite Dimensional Nonlinear Filters," *24th IEEE Conference on Decision and Control, 1985*, Vol. 24, IEEE Publ., Piscataway, NJ, Dec. 1985, pp. 1938–1945.

[37] Daum, F. E., "Exact Finite Dimensional Nonlinear Filters for Continuous Time Processes with Discrete Time Measurements," *23rd IEEE Conference on Decision and Control, 1984*, Vol. 23, IEEE Publ., Piscataway, NJ, Dec. 1984, pp. 16–22.

[38] Ito, K., and Xiong, K., "Gaussian Filters for Nonlinear Filtering Problems," *IEEE Transactions on Automatic Control*, Vol. 45, No. 5, 2000, pp. 910–927. doi:10.1109/9.855552

[39] Julier, S., and Uhlmann, J., "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, Vol. 92, No. 3, March 2004, pp. 401–422. doi:10.1109/JPROC.2003.823141

[40] Sloan, I. H., and Woniakowski, H., "When are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?" *Journal of Complexity*, Vol. 14, No. 1, 1998, pp. 1–33. doi:10.1006/jcom.1997.0463

[41] Doucet, A., de Freitas, N., and Gordon, N., *Sequential Monte-Carlo Methods in Practice*, Springer–Verlag, New York, April 2001, pp. 77–93.

[42] Iyengar, R. N., and Dash, P. K., "Study of the Random Vibration of Nonlinear Systems by the Gaussian Closure Technique," *Journal of Applied Mechanics*, Vol. 45, No. 2, 1978, pp. 393–399. doi:10.1115/1.3424308

[43] Roberts, J. B., and Spanos, P. D., *Random Vibration and Statistical Linearization*, Wiley, New York, 1990, pp. 122–176.

[44] Lefebvre, T., Bruyninckx, H., and Schutter, J. D., "Kalman Filters of Non-Linear Systems: A Comparison of Performance," *International Journal of Control*, Vol. 77, No. 7, 2004, pp. 639–653. doi:10.1080/00207170410001704998

[45] Lefebvre, T., Bruyninckx, H., and Schutter, J. D., "Comment on a New Method for the Nonlinear Transformations of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, Vol. 47, No. 8, 2002, pp. 1406–1409. doi:10.1109/TAC.2002.800742

[46] Mitter, S. K., "Lectures on Nonlinear Filtering and Stochastic Control," *Proceedings of the C.I.M.E.*, Springer, Berlin, 1983, pp. 170–207.

[47] Jin, C., Singla, P., and Singh, T., "Uncertainty Propagation in Dynamic Systems Using Polynomial Chaos Based Multiresolution Approach," *2012 AIAA Guidance, Navigation and Control Conference*, AIAA Paper 2012-4933, 2012.

[48] Adurthi, N., Singla, P., and Singh, T., "The Conjugate Unscented Transform—An Approach to Evaluate Multi-Dimensional Expectation Integrals," *Proceedings of the American Control Conference*, IEEE Publ., Piscataway, NJ, 2012, pp. 5556–5561.

[49] Adurthi, N., Singla, P., and Singh, T., "Conjugate Unscented Transform Rules for Uniform Probability Density Functions," *Proceedings of the American Control Conference*, IEEE Publ., Piscataway, NJ, June 2013, pp. 2454–2459.

[50] Adurthi, N., Singla, P., and Singh, T., "Conjugate Unscented Transform and its Application to Filtering and Stochastic Integral Calculation," *AIAA Guidance, Navigation, and Control Conference*, AIAA Paper 2012-4934, Aug. 2012.

[51] Madankan, R., Singla, P., Singh, T., and Scott, P., "Polynomial Chaos Based Bayesian Approach for State and Parameter Estimation," *Journal of Guidance, Navigation, and Control*, Vol. 36, No. 4, 2013, pp. 1058–1074. doi:10.2514/1.58377

[52] Lambert, H., Daum, F., and Weatherwax, J., "A Split-Step Solution of the Fokker–Planck Equation for the Conditional Density," *Fourtieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC'06*, IEEE Publ., Piscataway, NJ, 2006, pp. 2014–2018.

[53] Daum, F., "Nonlinear Filters: Beyond the Kalman Filter," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 20, No. 8, 2005, pp. 57–69. doi:10.1109/MAES.2005.1499276

[54] Giles, M. B., Duta, M. C., Muller, J.-D., and Pierce, N. A., "Algorithm Developments for Discrete Adjoint Methods," *AIAA Journal*, Vol. 41, No. 2, 2003, pp. 198–205. doi:10.2514/2.1961

[55] Musso, C., Oudjane, N., and Le Gland, F., "Improving Regularised Particle Filters," *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001, pp. 247–271.

[56] LeGland, F., Musso, C., and Oudjane, N., "An Analysis of Regularized Interacting Particle Methods for Nonlinear Filtering," *Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*, IEEE Publ., Piscataway, NJ, 1998, pp. 167–174.

[57] Budd, C. J., Cullen, M., and Walsh, E., "Monge–Ampére Based Moving Mesh Methods for Numerical Weather Prediction, with Applications to the Eady Problem," *Journal of Computational Physics*, Vol. 236, March 2013, pp. 247–270. doi:10.1016/j.jcp.2012.11.014

[58] Yang, T., Mehta, P. G., and Meyn, S. P., "A Mean-Field Control-Oriented Approach to Particle Filtering," *American Control Conference (ACC), 2011*, IEEE Publ., Piscataway, NJ, 2011, pp. 2037–2043.

[59] Alspach, D., and Sorenson, H., "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, Vol. 17, No. 4, Aug. 1972, pp. 439–448. doi:10.1109/TAC.1972.1100034

[60] Horwood, J. T., Aragon, N. D., and Poore, A. B., "Gaussian Sum Filters for Space Surveillance: Theory and Simulations," *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 6, 2011, pp. 1839–1851. doi:10.2514/1.53793

[61] Vishwajeet, K., Singla, P., and Jah, M., "Nonlinear Uncertainty Propagation for Perturbed Two-Body Orbits," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 5, 2014, pp. 1415–1425. doi:10.2514/1.G000472

[62] DeMars, K. J., and Jah, M. K., "Probabilistic Initial Orbit Determination Using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 5, 2013, pp. 1324–1335. doi:10.2514/1.59844

[63] Vittaldev, V., and Russell, R. P., "Collision Probability for Space Objects Using Gaussian Mixture Models," *Proceedings of the 23rd AAS/AIAA Space Flight Mechanics Meeting*, Vol. 148, Univelt, AAS Paper 13-351, San Diego, CA, 2013.

[64] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "Uncertainty Propagation for Nonlinear Dynamic Systems Using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 6, 2008, pp. 1623–1633. doi:10.2514/1.36247

[65] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation," *IEEE Transactions on Automatic Control*, Vol. 56, No. 9, 2011, pp. 2151–2156. doi:10.1109/TAC.2011.2141550

[66] Terejanu, G., George, J., and Singla, P., "Spacecraft Attitude Estimation Using Adaptive Gaussian Sum Filter," *Journal of the Astronautical Sciences*, Vol. 57, Nos. 1–2, 2009, pp. 31–45.

[67] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "Uncertainty Propagation for Nonlinear Dynamical Systems Using Gaussian Mixture Models," *AIAA Guidance, Navigation and Control Conference and Exhibit*, AIAA Paper 2008-7472, 2008.

[68] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "Decision Based Uncertainty Propagation Using Adaptive Gaussian Mixtures," *2009 International Conference on Information Fusion*, IEEE Publ., Piscataway, NJ, 2009, pp. 702–709.

[69] Giza, D., and Singla, P., "An Approach for Nonlinear Uncertainty Propagation: Application to Orbital Mechanics," *2009 AIAA Guidance Navigation and Control Conference*, AIAA Paper 2009-6082, 2009.

[70] Hanebeck, U. D., Briechle, K., and Rauh, A., "Progressive Bayes: A New Framework for Nonlinear State Estimation," *Proceedings of SPIE, AeroSense. Symposium 5099*, 2003, pp. 256–267. doi:10:1117/12:487806

[71] Zhang, Z., Chen, C., Sun, J., and Chan, K., "EM Algorithms for Gaussian Mixtures with Split-and-Merge Operation," *Pattern Recognition*, Vol. 36, No. 9, 2003, pp. 1973–1983. doi:10.1016/S0031-3203(03)00059-1

[72] Ueda, N., Nakano, R., Ghahramani, Z., and Hinton, G., "SMEM Algorithm for Mixture Models," *Neural Computation*, Vol. 12, No. 9, 2000, pp. 2109–2128. doi:10.1162/089976600300015088

[73] Richardson, S., and Green, P. J., "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 59, No. 4, 1997, pp. 731–792. doi:10.1111/rssb.1997.59.issue-4

[74] Runnalls, A. R., "Kullback–Leibler Approach to Gaussian Mixture Reduction," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, No. 3, 2007, pp. 989–999. doi:10.1109/TAES.2007.4383588

[75] Vittaldev, V., and Russell, R., "Multidirectional Gaussian Mixture Models for Nonlinear Uncertainty Propagation," *CMES-Computer Modeling in Engineering and Sciences*, Vol. 111, No. 1, 2016, pp. 83–117.

[76] Clark, E., and Quinn, A., "A Data-Driven Bayesian Sampling Scheme for Unsupervised Image Segmentation," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings*, Vol. 6, IEEE Publ., Piscataway, NJ, 1999, pp. 3497–3500.

[77] DeMars, K. J., Bishop, R. H., and Jah, M. K., "Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 4, 2013, pp. 1047–1057. doi:10.2514/1.58987

[78] Terejanu, G., "An Adaptive Split-Merge Scheme for Uncertainty Propagation Using Gaussian Mixture Models," *49th AIAA Aerospace Science Meeting*, AIAA Paper 2011-0890, 2011.

[79] Vishwajeet, K., and Singla, P., "Sparse Approximation Based Gaussian Mixture Model Approach for Uncertainty Propagation for Nonlinear Systems," *2013 American Control Conference*, IEEE Publ., Piscataway, NJ, 2013, pp. 1213–1218.

[80] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Dover, New York, 2007, pp. 126–140.

[81] Vishwajeet, K., and Singla, P., "Adaptive Splitting Technique for Gaussian Mixture Models to Solve Kolmogorov Equation," *2014 American Control Conference*, IEEE Publ., Piscataway, NJ, 2014, pp. 5186–5191.

[82] Arasaratnam, I., and Haykin, S., "Cubature Kalman Filter," *IEEE Transactions on Automatic Control*, Vol. 54, No. 6, June 2009, pp. 1254–1269. doi:10.1109/TAC.2009.2019800

[83] Junkins, J. L., "von Karman Lecture: Adventures on the Interface of Dynamics and Control," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 6, 1997, pp. 1058–1071. doi:10.2514/2.4176

[84] Kullback, S., *Information Theory and Statistics*, Dover, New York, 1997, pp. 6–7.

[85] Candes, E. J., Wakin, M. B., and Boyd, S., "Enhancing Sparsity by Reweighted L1 Minimization," *Journal of Fourier Analysis and Applications*, Vol. 14, Nos. 5–6, 2008, pp. 877–905. doi:10.1007/s00041-008-9045-x

[86] Boyd, S., and Vandenberghe, L., *Convex Optimization*, Cambridge Univ. Press, New York, 2004, pp. 67–78, 152–159.

[87] Huber, M. F., Peter, K., and Hanebeck, U. D., "Superficial Gaussian Mixture Reduction," *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion Trends Solutions Applications SDF*, 2011, http://www.sciencedirect.com/science/article/pii/S1665642313715724#bibl0005.

[88] Battin, R. H., *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA, Reston, VA, 1999, pp. 110–114.

[89] Grant, M., and Boyd, S., "CMSoft: Sharing Computational and Mathematics Software Solutions," Feb. 2016, https://goo.gl/prA1NK [retrieved 22 Nov. 2017].

[90] "CVX: Matlab Software for Disciplined Convex Programming, Version 2.1," CVX Research, March 2014, http://cvxr.com/ [retrieved 22 Nov. 2017].