UNCERTAINTY QUANTIFICATION USING DIRECTIONAL SPLITTING AND GAUSSIAN MIXTURE MODELS WITH APPLICATIONS TO ORBITAL DYNAMICS

Renato Zanetti*, Kyle J. DeMars†, Derek Tuggle*, Kristen Michaelson*, and Maaninee Gupta†

A novel adaptive algorithm is proposed to propagate forward in time an arbitrary distribution through nonlinear dynamics forced by white noise. The new scheme employs Gaussian Mixture Models and automatically refines the number of components to better encompass the effects of the nonlinear dynamics. The refinement is done in directions that suffer from both high nonlinearity and high uncertainty. An orbital uncertainty quantification numerical example is provided to validate the effectiveness of the proposed methodology.

INTRODUCTION

Uncertainty quantification refers to determining the evolution through time of either the probability distribution function (PDF) or of some key moments of the distribution, such as mean and covariance. The first methods for quantifying the uncertainty when dealing with orbital motion began with applying techniques like the Kalman filter [1] to the problem of estimating the trajectories of satellites [2] and to the circumlunar navigation problem [3, 4]. A Bayesian connection to the Kalman filter was established shortly thereafter [5], which paved the way for key advancements in non-Gaussian filtering [6, 7]. Recent efforts in orbital uncertainty quantification have focused on determining the non-Gaussian evolution of the uncertainty in terms of achieving tractable descriptions of the probability density function (pdf) for the uncertainty of the position and velocity of an orbiting object, as opposed to moments (e.g. mean and covariance), of the pdf. Several methods for quantifying the pdf have been developed, such as using Gaussian mixture models (GMMs) [8, 9, 10, 11], differential algebra [12], or transformation of variables [13]. A great many other methods have been developed; an excellent survey of uncertainty propagation in orbital mechanics is given in Ref. [14].

The motivation for using GMMs in nonlinear estimation is twofold. First, local or linearized solutions such as an extended Kalman filter (EKF) are potentially unsuitable for approximating moments of a pdf generated by nonlinear transformations. If a prior pdf is represented as a GMM (a process known as splitting the prior), then smaller eigenvalues of each of the component's covariance matrices indicate that the mixture components are better candidates for local methods. Second, any pdf can be exactly represented by a GMM if an unlimited number of components is allowed such that each component has infinitesimally small covariance [6, 7], and thus finitely parameterized GMMs

*R. Zanetti, K. Michaelson, and D. Tuggle are with the Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas 78712

†K. J. DeMars and M. Gupta are with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843

are often perceived as universal approximators [15, 16]. This property has beneficial implications for use in representing both prior and posterior pdfs. Better suitability of linearized solutions is irrelevant if the GMM representation is a significant departure from what is considered to be the true prior of the problem, so the ability to reasonably approximate the prior is crucial. Nonlinear transformation of a single Gaussian results in a non-Gaussian pdf that can be better approximated by a GMM than by a single Gaussian constructed with the transformed mean and covariance matrix.

The current work provides a novel technique to propagate uncertainty through nonlinear orbital dynamics utilizing splitting GMMs and builds on the foundations from References [11] and [17]. A key contribution for this work is the expansion of the splitting philosophy in [17], which only treated the measurement update portion of the filter, to the continuous time propagation phase, although the approach is applicable to discrete time systems as well. Ref. [17] develops a scheme using a scalar measurement and checks nonlinearity of a single transformation, while the current work deals with vector-based time propagation and necessitates to detect not only if nonlinearities are present, but also when. The proposed splitting scheme has a distinct advantage of selecting a splitting direction based on the combined notions of nonlinearity and uncertainty while much of prior work only considers eigendirections of the prior covariance matrix [11, 18, 19, 20]. Ref. [21] utilizes splitting as a means to ensure validity of Taylor series expansions when approximating differential entropy of a multivariate Gaussian mixture. Implicitly then, [21] similarly recognizes that splitting is warranted in regions of the state-space where both nonlinearities and uncertainty are pronounced. Though this is a guiding principle, both concepts of nonlinearity and uncertainty are not jointly considered in designing when splitting should occur, along which direction it should occur, and when it is no longer warranted. The current work seeks to do so. Other methods addressing at least one of these design points exist, but they consider only discrete transformations rather than the continuous time propagation problem [22, 23, 24]. In [25], the decision to split is based on uncertainty while the direction taken for splitting was one that maximized nonlinearity of the model dynamics. Ref. [26] uses a second-order divided difference term evaluated at a certain distance from the mean in order to quantify nonlinearity. This quantity is evaluated for a set of directions (including spectral directions for the prior covariance as well as some that are specific to the underlying dynamics model), and the direction for splitting is chosen as the one that maximizes the quantity for the set.

Implicit in discussing direction, the splitting method is 1-D, meaning a univariate splitting technique is used to place multivariate components along a certain direction in the state-space. Many 1-D methods involve iteratively applying the scheme so that multidirectional splitting may be achieved. Another approach to multidirectional splitting is offered in Ref. [27] where pdfs are not traditionally split but rather resampled in an importance-type manner while bounding component covariances via linear matrix inequality constraints. A full filter for this approach was provided in [28] for discrete-discrete estimation. Ref. [29] introduces two splitting techniques, one multidirectional split, and one unidirectional split. The unidirectional splitting technique uses the state transition matrix in conjunction with unscented transformation to compute the departure from linearity, which is different from the approach proposed here that uses the Kullback-Leibler divergence (KLD).

Among 1-D methods, the current work includes the following two additional contributions. First, a univariate splitting library is provided that results from optimization of the Kullback-Leibler divergence (KLD) under the constraints of preserving the first two moments of the original pdf. The Kullback-Leibler divergence, also known as relative entropy, is extensively used for comparison among pdfs for purposes of density estimation [30]. Second, a novel procedure for expanding the

univariate results to multidirectional splitting for an arbitrary direction is offered. This procedure is both variance-preserving and avoids the explicit use of matrix square root for the prior covariance as previously done. Since the initial dissemination of this work [31], a plethora of nonlinearity- and non-Gaussianity-based splitting methods have been developed. Numerous methods for splitting the initial uncertainty distribution are discussed in [32]. An adaptive splitting method based on [32] is proposed in [33].

PRELIMINARY NOTIONS

The ultimate goal of this work is to approximate the time evolution of the pdf, $\mathbf{p}_{\mathbf{x}(t)}(\mathbf{x}, t)$, of a stochastic process $\mathbf{x}(t)$ whose evolution is governed by a nonlinear stochastic differential equation

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) \; + \mathbf{L}(\mathbf{x}, t) \; \boldsymbol{\nu}(t) \tag{1}$$

where $\mathbf{x}(t)$, $\mathbf{f}(\mathbf{x}, t)$, and $\dot{\mathbf{x}}(t)$ are n-dimensional column vectors ($n = 6$ for the orbit propagation problem considered here). The process noise $\boldsymbol{\nu}(t)$ is an m-dimensional column vector ($m = 3$ for the orbit propagation problem considered here). Matrix $\mathbf{L}(\mathbf{x}, t)$ is $n \times m$ and for the orbit propagation problem is given by $\mathbf{L}(\mathbf{x}, t) = \begin{bmatrix} \mathbf{O}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix}^{\mathrm{T}}$.

We are not limiting our study to the approximation of the first (mean) and second (covariance) moments; rather, we are interested in the full pdf, which results in knowledge of moments of all orders. Full knowledge of the distribution is generally needed to calculate collision probability or to obtain the Minimum Mean Square Error (MMSE) estimate of the state; it is necessary to fully quantify the uncertainty of the system.

For a generic stochastic dynamic system, the evolution of the probability density is governed by the Fokker-Planck-Kolmogorov (FPK) equation [34]. For the nonlinear dynamic system in Eq. (1) driven by zero-mean white noise $\boldsymbol{\nu}$ with power spectral density $\mathbf{Q}$ the FPK equation is

$$\frac{\partial p_{\mathbf{x}(t)}(\mathbf{x}, t)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial \Big(f_i(\mathbf{x}, t) \; p_{\mathbf{x}(t)}(\mathbf{x}, t) \Big)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 \Big((\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}(\mathbf{x}, t)^{\mathrm{T}})_{ij} \; p_{\mathbf{x}(t)}(\mathbf{x}, t) \Big)}{\partial x_i \; \partial x_j}$$

where $f_i(\mathbf{x}, t)$ is the i-th component of vector $\mathbf{f}(\mathbf{x}, t)$ and $\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}(\mathbf{x}, t)^{\mathrm{T}})_{ij}$ is the ij-th matrix component. The FPK equation is linear with respect to the distribution $p_{\mathbf{x}(t)}(\mathbf{x}, t)$, but it is not trivial to solve. In fact, exact analytical solutions are very scarce and numerical solutions are computationally demanding and difficult to obtain. Since this is a linear PDE, when the initial distribution is equal to the weighted sum of simpler functions, $p_i(\mathbf{x})$, via

$$p_{\mathbf{x}(t_0)}(\mathbf{x}, t_0) = \sum_{i=1}^{N} w_i \; p_i(\mathbf{x})$$

then the evolution of the pdf can be computed as the sum of the evolution of the simpler components. It is assumed that each of the components $p_i(\mathbf{x})$ satisfies the conditions of being itself a pdf (non-negative and integrate to one); therefore the weights w_i need to add to one in order for the sum to also integrate to one. It is also assumed that all weights are positive; this assumption trivially guarantees that the total pdf remains non-negative for all values of $\mathbf{x}$.

There are very few situations in which the FPK equation is solvable, but one particularly notable solution is for the linear Gaussian case in which the solution stays Gaussian at all times with the

evolution of the mean and covariance matrix exactly expressed by the Kalman filter propagation equations [35]. Furthermore, if: (1) the dynamic system is linear, and (2) the initial pdf is exactly representable as a sum of Gaussians, there is an exact, closed-form solution of the evolution of the non-Gaussian pdf [6]. Unfortunately, these two conditions are rarely encountered in practice. Therefore it is usually necessary to approximate the initial distribution as a sum of Gaussians in which all the Gaussian components have a covariance matrix "small" enough such that the nonlinear dynamics is appropriately approximated by its linearization. GMM-based nonlinear uncertainty quantification techniques operate exactly under this basic working assumption: the initial distribution is well approximated by a GMM in which components are "small" enough such that for each of them linearization holds.

In the GMM framework, the total initial probability density function is approximated by a sum of Gaussian probability density functions; hence, we have that

$$p_{\mathbf{x}(t_0)}(\mathbf{x}) = \sum_{i=1}^{N} w_i \; p_g(\mathbf{x}; \hat{\mathbf{x}}_i(t_0), \mathbf{P}_i(t_0))$$

where t_0 represents the integration start time and $p_g(\mathbf{x}; \mathbf{m}, \mathbf{P})$ represents a Gaussian pdf in $\mathbf{x}$ with mean $\mathbf{m}$ and covariance $\mathbf{P}$. Starting from this initial pdf, if for each of the Gaussian components the nonlinear dynamics $\mathbf{f}(\mathbf{x}, t)$ is accurately approximately by its first order Taylor series centered at the component's mean, the propagated probability density function is accurately approximated by

$$p_{\mathbf{x}(t_f)}(\mathbf{x}) = \sum_{i=1}^{N} w_i \; p_g(\mathbf{x}; \hat{\mathbf{x}}_i(t_f), \mathbf{P}_i(t_f))$$

where $\hat{\mathbf{x}}_i(t_f)$ and $\mathbf{P}_i(t_f)$ are the solution of integrating the following ODEs from t_0 to t_f [7]

$$\dot{\hat{\mathbf{x}}}_i(t) = \mathbf{f}(\hat{\mathbf{x}}_i, t) \tag{2}$$

$$\dot{\mathbf{P}}_i(t) = \mathbf{F}\big(\hat{\mathbf{x}}_i(t), t\big)\mathbf{P}_i(t) + \mathbf{P}_i(t)\mathbf{F}\big(\hat{\mathbf{x}}_i(t), t\big)^{\mathrm{T}} + \mathbf{L}\big(\hat{\mathbf{x}}_i(t), t\big)\; \mathbf{Q}\; \mathbf{L}\big(\hat{\mathbf{x}}_i(t), t\big)^{\mathrm{T}} \tag{3}$$

$$\mathbf{F}\big(\hat{\mathbf{x}}_i(t), t\big) = \left.\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_i(t)} \tag{4}$$

The total mean $\hat{\mathbf{x}}(t)$ and covariance $\mathbf{P}(t)$ of a GMM can be obtained from the mean and covariance of the individual components as

$$\hat{\mathbf{x}}(t) = \sum_i w_i \; \hat{\mathbf{x}}_i(t) \tag{5}$$

$$\mathbf{P}(t) + \hat{\mathbf{x}}(t)\; \hat{\mathbf{x}}(t)^{\mathrm{T}} = \sum_i w_i \; (\mathbf{P}_i(t) + \hat{\mathbf{x}}_i(t)\; \hat{\mathbf{x}}_i(t)^{\mathrm{T}}) \tag{6}$$

DEVELOPMENT OF PROPOSED SCHEME

As previously stated, the goal of the proposed algorithm is to represent the uncertainty during propagation through nonlinear dynamics by using GMMs. Techniques are developed to split GMM components at the appropriate time when the linearization assumption ceases to be accurate and in the appropriate direction where nonlinear effects are more pronounced. The proposed approach is

to: (1) develop an automated technique to determine when linearized propagation is not adequate and hence splitting of components should occur; (2) develop a technique to determine the direction of maximum nonlinearity during nonlinear propagation, the split should occur in this direction; (3) develop a covariance-preserving multivariate splitting technique that refines the components along a predefined direction; and (4) develop a univariate, variance preserving, splitting library. These steps will be introduced in the following subsections.

Splitting Criteria

The key contribution of this work is the directional splitting of a GMM during orbital dynamics propagation to better represent to evolution of the PDF. Most of the existing GMM splitting literature assumes discretized dynamics with a constant discretization time step. Discretization is not used here; the use of nonlinear integration is a key aspect of this work. The goal is to detect the need of a component split as soon as the linearization assumption ceases to be satisfactory. The nonlinear flow of the orbital dynamics can be expanded in a Taylor series; for small time steps linearization of the solution flow is a very good approximation of it; as the time step increases, the linearization assumptions ceases to be valid. A key characteristics of the orbital flow is that it is not an odd function and that when a first order approximation of the Taylor series ceases to be satisfactory the next terms to dominate the solution are the second order terms.

A variable step integrator with error control is used to propagate the orbital dynamics:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$$

The numerical integration scheme employed includes an exit condition, which, when triggered, terminates the integration and forces a component split. After the split integration resumes until the following split condition. The goal is therefore to find an exit condition that is triggered when the linearity assumption ceases to be satisfactory and the second order terms contribute enough to the solution. This exit condition is the splitting criteria presented in this section.

The automatic criteria to determine when a split is needed is chosen as [36, 17]. The GMM propagation is accurate as long as the propagation of the individual components also accurately represent the time evolution of the pdf. Starting from a Gaussian distribution, a propagation of only the mean and covariance accurately represents the entire pdf when it remains Gaussian, i.e. when the system is linear. Putting these two concepts together, the GMM propagation is accurate when each of the components is operating in a regime when the linearization assumption holds. Our strategy is to detect when a linearized solution (EKF) differs from a higher order solution (HOF), such as the Gaussian Second Order Filter or the Unscented Kalman Filter. This difference is quantified using the Kullback-Leibler divergence assuming a Gaussian posterior approximation by both filters; τ is a threshold.

$$D_{KL}[p_g(\mathbf{x}; \hat{\mathbf{x}}_{HOF}, P^+_{HOF}) \,||\, p_g(\mathbf{x}; \hat{\mathbf{x}}_{EKF}, P^+_{EKF})] \geq \tau \tag{7}$$

where

$$\begin{aligned} D_{KL}[p_g(\mathbf{x}; \hat{\mathbf{x}}_{HOF}, P^+_{HOF}) \,||\, p_g(\mathbf{x}; \hat{\mathbf{x}}_{EKF}, P^+_{EKF})] = \frac{1}{2}(&\log\{|P^+_{EKF}|/|P^+_{HOF}|\} - n \\ &+ \operatorname{tr}\left\{P^+_{HOF}(P^+_{EKF})^{-1}\right\} + (\hat{\mathbf{x}}_{EKF} - \hat{\mathbf{x}}_{HOF})^{\mathrm{T}}(P^+_{EKF})^{-1}(\hat{\mathbf{x}}_{EKF} - \hat{\mathbf{x}}_{HOF})) \end{aligned} \tag{8}$$

This approach provides an immediate quantification of the loss in information associated with approximating the higher order solution with an EKF. Eq. (8) provides the difference between including and ignoring nonlinearities in Gaussian propagation. A KL divergence that is above the desired

threshold for splitting indicates a possible inadequacy in approximating the propagated solution as Gaussian. As this approximation fails, the component is re-approximated itself as a GMM via a one dimensional split along the direction of highest nonlinearity weighted by the directional uncertainty.

The threshold τ is chosen to limit the maximum allowable divergence between the two distributions. Once again we consider two Gaussian random vectors with mean and covariance matrix represented in Table 1, where $\mathbf{v}$ is some unit vector and scalars $c > 0,\ k > 1$. Distribution 1 repre-

Table 1. Normal Distributions for Threshold Selection

Normal Distribution 1:	$\boldsymbol{\mu}$,	Σ
Normal Distribution 2:	$\boldsymbol{\mu} + c\Sigma^{\frac{1}{2}}\mathbf{v}$,	$\frac{1}{k}\Sigma$

sents the true distribution (in our case approximated by the higher order filter) while Distribution 2 represents some departure from it (in our case the EKF) parameterized by scalars c and k. The KL divergence for this comparison reduces to

$$\tau = D_{KL}[p_1||p_2] = \frac{1}{2}(n(k - \log k - 1) + c^2 k) \tag{9}$$

We know that an EKF will always have covariance matrix smaller than a Gaussian Second Order Filter. With this perspective, constant k is chosen as the maximum divergence that we allow the two covariances to have, while scalar c is the number of standard deviations by which we allow the means to diverge. Once the designers have chosen the maximum allowable discrepancy for their application, the algorithm detects when a component split is necessary using Eq. (7). As the threshold does not depend on the specific solutions at each step, it can be precomputed and stored.

Splitting Direction

After determining that linearized uncertainty propagation is no longer acceptable and that a component split is needed, the next step is to determine the direction of the split, i.e. the direction of maximum nonlinearity. This section expands prior work [36] to split nonlinear vector functions of the state. Prior work was only applicable to scalar functions and applied to the measurement update. Let $\mathbf{z} \in \Re^m$ be a nonlinear function of the state $\mathbf{x} \in \Re^n$ with additive error $\boldsymbol{\eta}$, i.e.

$$\mathbf{z} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\eta}$$

The $m \times n$ Jacobian is denoted by $\mathbf{F}$, such that

$$\mathbf{F} = \begin{bmatrix}\mathbf{F}_1^{\mathrm{T}} & \mathbf{F}_2^{\mathrm{T}} & \cdots & \mathbf{F}_m^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$$

where $\mathbf{F}_i$ is a row-vector representing the Jacobian of the i-th scalar component of the nonlinear function. A linear system posseses constant first-order derivatives; the directional derivative of the Jacobian evaluated at the prior mean, or the rate of change of the Jacobian at $\bar{\mathbf{x}}$ in direction $\mathbf{u}$, is therefore a reasonable quantification of nonlinearity. The directional derivative is given by

$$\nabla_{\mathbf{u}}\mathbf{F}(\mathbf{x})\Big|_{\bar{\mathbf{x}}} = \lim_{\alpha \to 0} \frac{\mathbf{F}(\bar{\mathbf{x}} + \alpha\mathbf{u}) - \mathbf{F}(\bar{\mathbf{x}})}{\alpha} \tag{10}$$

The Taylor series for the Jacobian, which is

$$\mathbf{F}_i(\mathbf{x}) = \mathbf{F}_i(\bar{\mathbf{x}}) + \left(\frac{\partial \mathbf{F}_i(\mathbf{x})^{\mathrm{T}}}{\partial \mathbf{x}}\Big|_{\bar{\mathbf{x}}} [\mathbf{x} - \bar{\mathbf{x}}] \right)^{\mathrm{T}} + \cdots$$

can be substituted into Eq. (10) to provide the directional derivative of the Jacobian of the i-th scalar component of $\mathbf{f}(\mathbf{x})$ as

$$\begin{aligned} \nabla_{\mathbf{u}} \mathbf{F}_i(\mathbf{x})\Big|_{\bar{\mathbf{x}}} &= \lim_{\alpha \to 0} \frac{\mathbf{F}_i(\bar{\mathbf{x}}) + \alpha \left(\frac{\partial \mathbf{F}_i(\mathbf{x})}{\partial \mathbf{x}}\Big|_{\bar{\mathbf{x}}} \mathbf{u} \right)^{\mathrm{T}} + \cdots - \mathbf{F}_i(\bar{\mathbf{x}})}{\alpha} \\ &= \mathbf{u}^{\mathrm{T}} \frac{\partial \mathbf{F}_i(\mathbf{x})^{\mathrm{T}}}{\partial \mathbf{x}}\Big|_{\bar{\mathbf{x}}} \\ &= \mathbf{u}^{\mathrm{T}} \mathbf{D}_i(\bar{\mathbf{x}})^{\mathrm{T}} \end{aligned}$$

where $\mathbf{D}_i(\mathbf{x}) \in \Re^{n \times n}$ is the Hessian of the i-th component of $\mathbf{f}(\mathbf{x})$. Assembling the components of the directional derivative, it can be shown that the squared Frobenius norm is

$$\begin{aligned} \left\| \nabla_{\mathbf{u}} \mathbf{F}(\mathbf{x})\Big|_{\bar{\mathbf{x}}} \right\|_F^2 &= \mathrm{trace} \left(\nabla_{\mathbf{u}} \mathbf{F}(\mathbf{x}) \, \nabla_{\mathbf{u}} \mathbf{F}(\mathbf{x})^{\mathrm{T}} \right) \\ &= \sum_{i=1}^{m} \mathbf{u}^{\mathrm{T}} \mathbf{D}_i^{\mathrm{T}}(\bar{\mathbf{x}}) \, \mathbf{D}_i(\bar{\mathbf{x}}) \mathbf{u} \\ &= \mathbf{u}^{\mathrm{T}} \left(\sum_{i=1}^{m} \mathbf{D}_i^{\mathrm{T}}(\bar{\mathbf{x}}) \, \mathbf{D}_i(\bar{\mathbf{x}}) \right) \mathbf{u} \\ &= \mathbf{u}^{\mathrm{T}} \, \mathbf{E}(\bar{\mathbf{x}}) \, \mathbf{u} \end{aligned} \tag{11}$$

Eq. (11) provides a measure of the nonlinearity in a particular direction, $\mathbf{u}$. Our choice of cost function weights Eq. (11) by the uncertainty in that direction, e.g. if a direction is very nonlinear but with no uncertainty, a split in that direction is not needed, leading to

$$J(\mathbf{u}) = \left\| \nabla_{\mathbf{u}} \mathbf{F}(\mathbf{x})\Big|_{\bar{\mathbf{x}}} \right\|_F^2 \sigma_u^2 = \frac{\mathbf{u}^{\mathrm{T}} \, \mathbf{E}(\bar{\mathbf{x}}) \, \mathbf{u}}{\mathbf{u}^{\mathrm{T}} \, \mathbf{P}^{-1} \, \mathbf{u}} \tag{12}$$

The direction $\mathbf{u}^*$ that maximizes Eq. (12) can be easily computed after a change of variables. Let $\mathbf{v} = \mathbf{S}^{-1}\mathbf{u}$ where $\mathbf{P}^{-1} = \mathbf{S}^{-\mathrm{T}}\mathbf{S}^{-1}$ so that the following equivalent cost function can be examined:

$$\tilde{J}(\mathbf{v}) = \frac{\mathbf{v}^{\mathrm{T}} \, \mathbf{S}^{\mathrm{T}} \, \mathbf{E}(\bar{\mathbf{x}}) \, \mathbf{S} \, \mathbf{v}}{\mathbf{v}^{\mathrm{T}} \, \mathbf{v}} \tag{13}$$

The Rayleigh-Ritz inequality provides an upper bound for Eq. (13) that is the maximum eigenvalue of matrix $\mathbf{S}^{\mathrm{T}} \, \mathbf{E}(\bar{\mathbf{x}}) \, \mathbf{S}$ with the associated eigenvector $\mathbf{v}^*$ being the maximizing argument. Changing variables back to the original problem, the chosen direction for splitting is $\mathbf{u}^* = \mathbf{S}\mathbf{v}^*$.

Arbitrary Direction Splitting

After detecting the nonlinearities and the need to split (Splitting Criteria), and deriving the direction $\mathbf{u}$ of maximum nonlinearity along which the split is necessary (Splitting Direction), the next step is to develop an algorithm able to apply the univariate split in an arbitrary direction $\mathbf{u}$ of a

multivariate Gaussian distribution. The univariate splitting library is presented in Section , and this section shows how to apply the splitting library along $\mathbf{u}$.

Assume the prior random vector $\mathbf{x}$ is Gaussian with mean $\mathbf{m}$ and covariance matrix $\mathbf{P}$; the objective is to determine a GMM that is parameterized by w_i, $\mathbf{m}_i$, and $\mathbf{P}_i$, $i \in \mathcal{I}$ that preserves mean and covariance, where $\mathcal{I}$ denotes the set of indices describing the components that the prior mean and covariance are being split into. The weights w_i are determined by the univariate splitting library. Let the variance of $\mathbf{x}$ along the unit vector $\mathbf{u}$ be denoted as σ_u^2, such that

$$\mathbf{u}^{\mathrm{T}} \, \mathbf{P}^{-1} \, \mathbf{u} = 1/\sigma_u^2 \tag{14}$$

where $\mathbf{u}$ is the direction along which the splitting is performed. Since the splitting occurs along the unit vector $\mathbf{u}$, it follows that

$$\mathbf{m}_i = \mathbf{m} + m_i \, \sigma_u \, \mathbf{u} \tag{15}$$

where the scalar means, m_i, are also determined by the univariate splitting library.

What is left to find are the component covariances, $\mathbf{P}_i$, such that the posterior GMM covariance matches the covariance prior to the split. As commonly done, all components of the GMM are assigned the same covariance matrix. Let $\mathbf{P}$ and $\mathbf{m}$ be the covariance and the mean prior to the split, respectively. Preserving the covariance between the original covariance and the covariance of the GMM implies that

$$\mathbf{P} + \mathbf{m} \, \mathbf{m}^{\mathrm{T}} = \sum_{i \in \mathcal{I}} w_i \, (\mathbf{P}_i + \mathbf{m}_i \, \mathbf{m}_i^{\mathrm{T}}) = \mathbf{P}_i + \sum_{i \in \mathcal{I}} w_i \, \mathbf{m}_i \, \mathbf{m}_i^{\mathrm{T}}$$

where it is reminded that $\mathbf{P}_i$ is the same for all of the GMM components. Substituting for the GMM means from Eq. (15),

$$\mathbf{P} + \mathbf{m} \, \mathbf{m}^{\mathrm{T}} = \mathbf{P}_i + \mathbf{m} \, \mathbf{m}^{\mathrm{T}} + \left[\sum_{i \in \mathcal{I}} w_i m_i^2 \right] \sigma_u^2 \, \mathbf{u} \, \mathbf{u}^{\mathrm{T}} + \left[\sum_{i \in \mathcal{I}} w_i m_i \right] \sigma_u \left[\mathbf{u} \, \mathbf{m}^{\mathrm{T}} + \mathbf{m} \, \mathbf{u}^{\mathrm{T}} \right] \tag{16}$$

Choosing a symmetric univariate split results in

$$\sum_{i \in \mathcal{I}} w_i m_i = 0$$

Hence, Eq. (16) can be solved for $\mathbf{P}_i$ to yield

$$\mathbf{P}_i = \mathbf{P} - \left[\sum_{i \in \mathcal{I}} w_i m_i^2 \right] \sigma_u^2 \, \mathbf{u} \, \mathbf{u}^{\mathrm{T}} \tag{17}$$

Eq. (17) is the general formula used by this algorithm. However, as shown in the next section, if the univariate library is itself variance preserving, it follows that (see Eq. (21))

$$\sum_{i \in \mathcal{I}} w_i m_i^2 = 1 - \sigma^2$$

meaning that in the case of variance-preserving univariate splitting, the covariance of each component should be

$$\mathbf{P}_i = \mathbf{P} + \left(\sigma^2 - 1 \right) \sigma_u^2 \, \mathbf{u} \, \mathbf{u}^{\mathrm{T}} \tag{18}$$

which is only true when the univariate splitting is variance preserving. If this is the case, then this formulation is equivalent to the solution by Vittaldev and Russell [26]. Notice, however, that the univariate splitting libraries of Ref. [26] and the corresponding multivariate directional splitting are not variance preserving. Eq. (18) has the advantage over Ref. [26] that it avoids the need of computing the matrix square root of the covariance.

To verify that the proper uncertainty is obtained by choosing $\mathbf{P}_i$ in accordance with Eq. (18), consider the uncertainty of $\mathbf{x}_i$ along the direction $\mathbf{u}$, which is

$$1/\sigma_{i,u}^2 = \mathbf{u}^{\mathrm{T}}\mathbf{P}_i^{-1}\mathbf{u} = \mathbf{u}^{\mathrm{T}}\Big[\mathbf{P} + (\sigma^2 - 1)\sigma_u^2\,\mathbf{u}\,\mathbf{u}^{\mathrm{T}}\Big]^{-1}\mathbf{u}$$

From the matrix inversion lemma, it follows that

$$1/\sigma_{i,u}^2 = \mathbf{u}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{u} - \mathbf{u}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{u}\bigg(\frac{1}{(\sigma^2 - 1)\sigma_u^2} + \mathbf{u}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{u}\bigg)^{-1}\mathbf{u}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{u}$$

Applying Eq. (14) and reducing gives the final result that

$$\frac{1}{\sigma_{i,u}^2} = \frac{1}{\sigma_u^2}\left[1 - \frac{\sigma^2 - 1}{\sigma^2}\right] = \frac{1}{\sigma_u^2}\,\frac{1}{\sigma^2} \tag{19}$$

That is, $\sigma_{i,u}^2 = \sigma_u^2\,\sigma^2$, which is the desired result for the uncertainty of $\mathbf{P}_i$ along the direction $\mathbf{u}$. Since Eqs. (16) and (17) reduce the uncertainty in the $\mathbf{u}$ direction, Eq. (19) proves $\mathbf{P}_i$ remains positive definite.

Variance-Preserving Univariate Splitting

All is left to complete the algorithm is the development of a univariate splitting scheme for Gaussian distributions. This is the topic of this section.

Starting from a univariate Gaussian distribution, denoted by $p(x)$, the goal is to approximate it by a GMM, denoted by $q(x)$, where the reason for the approximation is to replace the original Gaussian distribution by a combination of Gaussian distributions with smaller variance. The associated regions with non-negligible probability are smaller for each of the Gaussians that replace $p(x)$, which allows linear transformations of uncertainty to be carried out with less approximation error than would occur with the original distribution. Approximating $p(x)$ by $q(x)$ in this manner is referred to as splitting, where the objective is to determine the parameters of $q(x)$.

There are a variety of ways in which the splitting process can be formulated, such as minimizing the L_2 norm between $p(x)$ and $q(x)$ [11], minimizing the divergence between $p(x)$ and $q(x)$ [37], matching the moments of $q(x)$ to those of $p(x)$ [37], or employing a power law in conjunction with the L_2 norm [38]. In this work the difference between probability distributions is consistently evaluated with the Kullback-Leibler divergence (KLD); therefore the splitting library is formulated by minimizing the KLD between $p(x)$ and $q(x)$; that is, given $p(x) = p_g(x; 0, 1)$, the objective is to minimize

$$D_{KL}[p||q] = \int p(x) \log \frac{p(x)}{q(x;\Theta)}\,\mathrm{d}x$$

to obtain the parameters of $q(x)$, which is the GMM

$$q(x;\Theta) = w_0 p_g(x; m_0, P_0) + \sum_{\ell=-1}^{-M} w_\ell p_g(x; m_\ell, P_\ell) + \sum_{\ell=1}^{M} w_\ell p_g(x; m_\ell, P_\ell)\,, \tag{20}$$

where M is known and Θ is the set of parameters defining the GMM, i.e. Θ is the collection of the weights, means, and variances appearing in Eq. (20). It is common to add constraints to the GMM parameters, such as requiring that the weights are all positive, that they sum to one (which ensures that $q(x)$ is a pdf), or that the mean and variance of $q(x)$ match those of $p(x)$. Including these constraints, the optimization problem may be stated as

$$\min D_{KL}[p||q] \quad \text{subject to} \quad w_\ell > 0 \ \ \forall \ \ \ell \in \{-M, \ldots, M\}\,,$$
$$\sum_{\ell=-M}^{M} w_\ell = 1\,, \ \sum_{\ell=-M}^{M} w_\ell m_\ell = 0, \quad \text{and} \quad \sum_{\ell=-M}^{M} w_\ell (P_\ell + m_\ell^2) = 1\,.$$

To reduce the parameter space for the optimization problem, the GMM is assumed to be symmetric, i.e. $m_0 = 0$ and $w_{-\ell} = w_\ell$ and $m_{-\ell} = -m_\ell \ \forall \ \ell \neq 0$, and the GMM is assumed to be homoscedastic, such that $P_\ell = \sigma^2 \ \forall \ \ell$. Under this restricted GMM, the optimization problem becomes

$$\min D_{KL}[p||q] \quad \text{subject to} \quad w_\ell > 0 \ \ \forall \ \ \ell \in \{-M, \ldots, M\} \quad \text{and} \quad \sum_{\ell=-M}^{M} w_\ell = 1\,,$$

in order for the variance of $q(x)$ and $p(x)$ to be the same, the common variance parameter of the GMM must be chosen as

$$\sigma^2 = 1 - 2\sum_{\ell=1}^{M} w_\ell m_\ell^2\,. \tag{21}$$

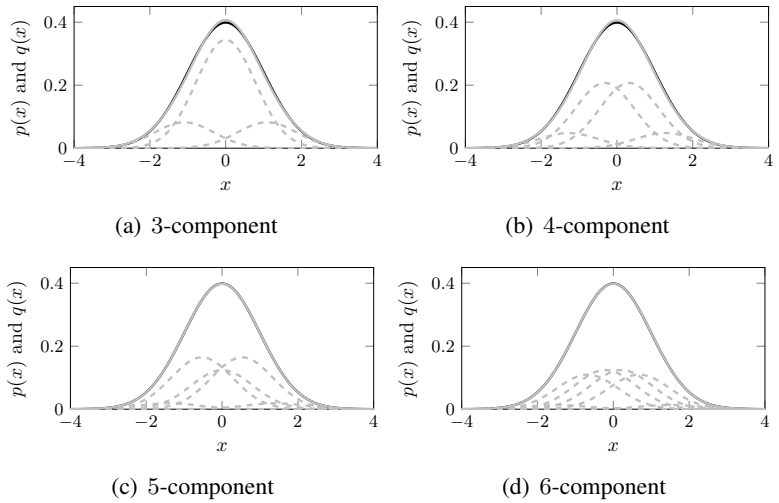
Thus, the first moment of $q(x)$ is matched by symmetry, and the second moment is matched by the selection of the common variance parameter via Eq. (21). Whenever an even-component mixture is desired, the central weight is set to zero, i.e. $w_0 = 0$, which effectively nullifies the central component's appearance in Eq. (20). Thus, w_ℓ and m_ℓ for $\ell > 0$ and w_0 (for odd-component mixtures) fully define $q(x)$.

In determining the weight and mean parameters, it is desired to ensure good separation between the components of the GMM. The method proposed here is to divide the support of the distribution into equiprobable bins and center the components at the means of these bins. Suppose it is desired to split the distribution into three components. The three equiprobable regions, each possessing $1/3$ of the probability, are $(-\infty, -0.43)$, $[-0.43, 0.43]$, and $(0.43, \infty)$, where the regions are straightforwardly determined from the cumulative distribution function. For the general case of splitting the Gaussian into L components and hence dividing the support into L regions, if each region is represented by an interval described by a_i and b_i, where $a_i < b_i$, then the mean of the region is

$$m_i = \frac{L}{\sqrt{2\pi}} \int_{a_i}^{b_i} x e^{-x^2/2}\, \mathrm{d}x$$

where the factor of L scales the distribution in this region to be a valid pdf. The means for each of the L regions are then used as the means for the L components in the GMM. Only $\lceil (L-1)/2 \rceil$ means need to be calculated; the remaining means are found from the symmetry requirement. Once the means are defined in this manner, w_ℓ for $\ell > 0$ fully define the variable parameters that can be chosen to minimize the $D_{KL}[p||q]$. Accounting for the symmetry constraint and the constraint that the weights must sum to unity, there are $\lfloor (L-1)/2 \rfloor$ free parameters in the optimization problem.

The constrained KL divergence minimization problem is used to find 3-, 4-, 5-, and 6-component splitting libraries (i.e. GMM approximations of a standard normal distribution). The resulting distributions are illustrated in Figure 1, which graphically compares the target distribution, $p(x)$, to the determined GMM distribution, including the individual components of $q(x)$. It is clearly



(a) 3-component (b) 4-component

(c) 5-component (d) 6-component

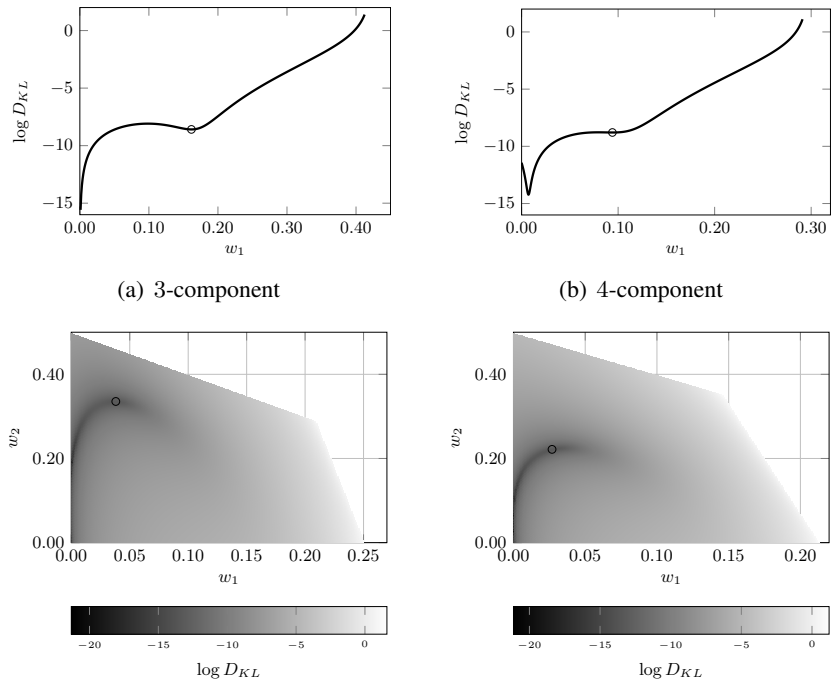
Figure 1. Splitting library solutions found by minimizing the KL divergence.

observed that, while increasing the number of components improves the matching of $q(x)$ to $p(x)$, the 5- and 6-component GMMs have two components that have significantly lower weights than the remaining components of the GMM.

To assess the solution quality, the KL divergence is computed over the permissible range of free parameters for each of the 3-, 4-, 5-, and 6-component cases. In particular, the only free parameter of the 3- and 4-component cases is taken to be w_1, and the free parameters of the 5- and 6-component cases are taken to be w_1 and w_2. The resulting curves and surfaces of the KL divergence are illustrated in Figure 2, and the minimizing solutions represented in Figure 1 are depicted by the marker along the curve or surface. From Figure 2(a), it is seen that as $w_1 \downarrow 0$, $D_{KL}[p||q] \downarrow 0$, which is the case where the 3-component GMM is being reduced to a 1-component GMM (i.e., a single Gaussian) that is exactly $p(x)$. As this eliminates the ability to split $p(x)$ into separate components, it is not of interest. Therefore, the local minimum that is found is the solution that is of interest. In Figures 2(b)–2(d), it is seen that several local minima exist. In each case, however, the local minimum with the largest weights is found, leading to a GMM with weights spread more completely across the components.

SIMULATION RESULTS

The performance of the developed directional splitting procedure is demonstrated in three numerical examples. The first example is a Keplerian orbit. This example is chosen to assess relative performance of the proposed algorithm against other, existing, techniques. The second numerical example is a more realistic example with perturbations added. The final numerical example addresses uncertainty propagation in cislunar space.



(a) 3-component (b) 4-component

(c) 5-component (d) 6-component

Figure 2. KL divergence as a function of library parameters.

Example I. Benefits of Directional Splitting

The purpose of this example is to show the benefits of the proposed directional splitting over splitting in the direction of maximum uncertainty. In order to show the advantages of the proposed directional splitting scheme, a direct comparison with AEGIS, the prior work from Ref. [11], is shown. Since this example focuses on relative performance with respect to AEGIS, a simple baseline Keplerian orbit is used. The absolute performance of the proposed algorithm against a more realistic scenario is left to the next example.

In [11] the GMM components splitting is done in the direction of maximum uncertainty, i.e. the nonlinearity of the dynamics is not taken into account, and splitting occurs in the direction of the eigenvector associated with the maximum eigenvalue of the covariance matrix. For the comparison to be meaningful, this example does not use the KL divergence criteria proposed in this work. Rather the same split criteria used in [11] is employed, which is based on the differential entropy difference between an EKF and a higher order filter. Much of the prior work in the literature performs splitting of GMM components along the direction of maximum uncertainty. While this approach has been shown to work well in practice, tailoring the added GMM components to the most needed directions reduces the number of total components.

The differential entropy of a random vector gives an indication of the average amount of information content and it can be used to evaluate the difference in information content between two distinct random vectors. The differential entropy of a random vector with probability density function $p(\mathbf{x})$

is defined as

$$H = -\int_{\mathcal{S}} p(\mathbf{x}) \log p(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \mathbb{E}\{-\log p(\mathbf{x})\} \tag{22}$$

where $\mathcal{S}$ is the support of the random vector. For a Gaussian distribution with covariance matrix $\mathbf{P}$, Eq. (22) reduces to

$$H = \frac{1}{2} \log |2\pi e \mathbf{P}|$$

It can be noticed that this quantity is only a function of the covariance of the Gaussian random variable, so it is not a function of the mean. The strategy employed in [11] is to monitor the difference in differential entropy between the linearized (EKF) solution and a higher order (UKF) solution. If the difference exceeds a predetermined threshold, the linearization assumption is deemed insufficient and a component split is performed.

For Hamiltonian systems [39] the differential entropy is constant across EKF time propagation [11]. Therefore, in this Keplerian orbital propagation scenario, it is not actually necessary to carry both the EKF solution and a higher order solution. Only the higher order solution is necessary, as its differential entropy variation is sufficient to determine its divergence from the linearized (constant entropy) solution. The differential entropy approach from [11] has therefore attractive computational savings in Hamiltonian propagation cases. Our proposed approach with the KL divergence, on the other hand, requires carrying both the EKF and the higher order solution even for Hamiltonian systems. However, in most practical engineering situations, diffusion terms are an important part of the dynamics evolution, and hence the EKF solution needs to be carried explicitly in the computations in order to implement the algorithm from [11].

An advantage of using the KL divergence over differential entropy is the fact that it evaluates differences in both the mean and covariance of the two solutions. Consider the continuous propagation equations of the mean and covariance matrix of Eq. (1) using the Gaussian Second Order Filter approximation [34, p. 340]

$$\dot{\hat{x}}_i = f_i(\hat{\mathbf{x}}, t) + \frac{1}{2} \operatorname{tr} \left\{ \mathbf{P} \left. \frac{\partial^2 f_i(\mathbf{x}, t)}{\partial \mathbf{x} \, \partial \mathbf{x}^{\mathrm{T}}} \right|_{\hat{\mathbf{x}}} \right\} \tag{23}$$

$$\dot{\mathbf{P}} = \left. \frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}} \mathbf{P} + \mathbf{P} \left. \frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} \tag{24}$$

The evolution of the covariance matrix does not depend directly on the second order effects, rather they enter indirectly through evaluating the Jacobian of the dynamics at the mean, whose evolution contain the Hessian of the dynamics directly. It follows that situations could naturally arise where the nonlinearity of the system is more effectively monitored with a splitting criteria that evaluates both means and covariances. A well-tuned criterion based on the KL divergence, therefore, would have the opportunity to identify the nonlinearity more rapidly. Nevertheless, a differential entropy approach has been successfully used in the past and has clear computational advantages in the traceless, diffusion-free, propagation case.

Since the orbit is Keplerian, without loss of generality, the orbital plane is chosen as the equatorial plane and the z coordinates are omitted (since they are identically zero at all times). Denoting the two-dimensional position by $\mathbf{r}$ and the two-dimensional velocity by $\mathbf{v}$, the state vector and equations

of motion that describe the nonlinear dynamical system are

$$\mathbf{x} = \begin{bmatrix} \mathbf{r} \\ \mathbf{v} \end{bmatrix} \quad \text{and} \quad \dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{r}} \\ \dot{\mathbf{v}} \end{bmatrix} = \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{v} \\ \mathbf{g}(\mathbf{r}) \end{bmatrix}$$

The acceleration due to gravity is denoted by $\mathbf{g}(\mathbf{r})$. For a two-body, point-mass formulation,

$$\mathbf{g}(\mathbf{r}) = -\frac{\mu}{r^3}\mathbf{r}$$

where $\mu = 3.986004415 \times 10^5$ km^3/s^2 is the gravitational parameter of the Earth and $r = \|\mathbf{r}\| = \sqrt{x^2 + y^2}$. Given the dynamical system, the Jacobian is readily obtained as

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \mathbf{O} & \mathbf{I} \\ \mathbf{G}(\mathbf{r}) & \mathbf{O} \end{bmatrix}$$

where $\mathbf{G}(\mathbf{r})$ is

$$\mathbf{G}(\mathbf{r}) = \frac{\partial \mathbf{g}(\mathbf{r})}{\partial \mathbf{r}} = -\frac{\mu}{r^3}\mathbf{I} + 3\frac{\mu}{r^5}\mathbf{r}\,\mathbf{r}^{\mathrm{T}}$$

Recalling that $\mathbf{D}_i(\mathbf{x})$ is the Hessian of the i-th component of $\mathbf{f}(\mathbf{x})$, it follows that $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{O}$ for the nonlinear dynamical system considered here; additionally, for $i = 1, 2$,

$$\mathbf{D}_{i+2}(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 g_i(\mathbf{r})}{\partial \mathbf{r}\,\partial \mathbf{r}^{\mathrm{T}}} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

where

$$\frac{\partial^2 g_i(\mathbf{r})}{\partial \mathbf{r}\,\partial \mathbf{r}^{\mathrm{T}}} = 3\frac{\mu}{r^5}\mathbf{e}_i\,\mathbf{r}^{\mathrm{T}} - 15\frac{\mu\, r_i}{r^7}\mathbf{r}\,\mathbf{r}^{\mathrm{T}} + 3\frac{\mu}{r^5}\mathbf{r}\,\mathbf{e}_i^{\mathrm{T}} + 3\frac{\mu\, r_i}{r^5}\,\mathbf{I} \tag{25}$$

and $\mathbf{e}_i$ is an element of the canonical base of $\Re^n$, $r_1 = x$, and $r_2 = y$.

For the orbital uncertainty propagation problems considered here, the initial state distribution is taken to be Gaussian. The initial mean is represented by orbital elements given by a semi-major axis of 35000 km, an eccentricity of 0.2, an argument of periapse of $0°$, and a mean anomaly of $0°$. The true initial mean is given by converting the specified orbital elements into Cartesian coordinates, yielding

$$\mathbf{m}_r = \begin{bmatrix} 28000 \\ 0 \end{bmatrix} \text{ km} \qquad \text{and} \qquad \mathbf{m}_v = \begin{bmatrix} 0 \\ 4.133144 \end{bmatrix} \text{ km/s}$$

where $\mathbf{m}_r$ is the position part of the initial mean and $\mathbf{m}_v$ is the velocity part of the initial mean. Furthermore, the initial covariance is taken to be diagonal with standard deviations in the position elements of 1 km and standard deviations in the velocity elements of 1 m/s. Given the prescribed initial mean and covariance in Cartesian coordinates, the objective is to propagate the uncertainty for one orbital period, which is approximately 0.754 days or 65,165 seconds. To be consistent with each other, both schemes perform the splitting into three components using the univariate library shown in Table 2.

Using a differential entropy threshold of 0.0081 nats the AEGIS scheme produces a GMM with 217 components. By splitting along the most nonlinear direction as proposed in this work, we are

Table 2. Univariate Splitting Library from Ref. [11]

Component #	Weight	Mean	Std. Dev.
-1	0.22522462491	-1.0575154614	0.67156628866
0	0.5495507501	0	0.67156628866
1	0.22522462491	1.0575154614	0.67156628866

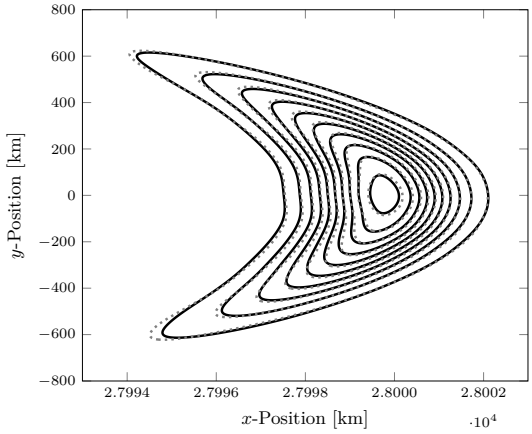


Figure 3. Comparison between AEGIS (solid line) and directional splitting (dotted line)

able to obtain the same level of performance with a saving of roughly 25% of components; the directional splitting scheme performs the propagation with 153 components. More importantly, computing a kernel density estimate of the position's distribution from the propagated Monte Carlo points, and comparing it with the position PDF produce by AEGIS the resulting Integrated Square Error (ISE) [16] is $2.44 \cdot 10^{-6}$, while the newly proposed approach has an ISE of $1.73 \cdot 10^{-6}$ a 30% improvement. ISE (error calculated from one set of Monte Carlo points) and MISE (mean of many ISEs each computed from a different Monte Carlo set of points) are the standard statistical metrics to evaluate PDF approximations.

Fig. 3 shows the distribution of the position uncertainty at the end of propagation. The solid lines represent equal probability contours of the AEGIS scheme, while the dashed lines depict the directional splitting results. It can be seen that the two approaches return effectively the same distribution, but directional splitting allows the same results at a fraction of the cost.

Example II. Propagation of Orbital Uncertainty

We next utilize the newly proposed splitting library in Table 3 with the threshold from Eq. (9) taking parameters $k = 1.01^2$ and $c = 0.35$. Furthermore, performance is demonstrated for a more realistic orbit determination problem: tracking of 3D position and velocity for a satellite in an Earth orbit. The satellite has a box-wing body equipped with a solar panel and maintains a certain reference attitude with respect to the Earth and Sun. Surface areas and reflectivity properties for each face are listed in Table 4. The true and modeled filter state dynamics are summarized in Table 5. Notable discrepancies between the true and filter dynamics are: fidelity of the Earth gravity field, attitude-based atmospheric drag, attitude-based solar radiation pressure (SRP), and material-based SRP reflectivity coefficients. Dynamic mismodeling was addressed via the inclusion

of process noise as in Eq. (1) with power spectral density value $10^{-20} I_{3\times 3}\ \frac{1}{s}$.

Table 3. New Univariate Splitting Library used in this example

Component #	Weight	Mean	Std. Dev.
-1	0.1616701997	-1.0908000117	0.78439476713
0	0.6766596007	0	0.78439476713
1	0.1616701997	1.0908000117	0.78439476713

Table 4. Satellite Geometry and Reflectivity Coefficients

Face (Body Axes)	Surface Area	C_{diffuse}	C_{specular}
$\pm X$	$6m^2$	0.04	0.59
$\pm Y$	$8m^2$	0.04	0.59
$+Z$	$12m^2$	0.80	0.04
$-Z$	$12m^2$	0.28	0.18
Solar Panel	$15m^2$	0.04	0.04

Table 5. Propagation Models

	Simulation	Filter
Earth Gravity	20×20 EGM96	point mass + J_2
Point Mass Gravity	Moon, Sun	Moon, Sun
Drag	attitude-based, Table 4 $C_D = 1.88$ m = 2000kg	cannonball, A=$15m^2$ $C_D = 1.88$ m = 2000kg
SRP	attitude-based, Table 4 material-based with Table m = 2000kg	cannonball, A=$15m^2$ $C_R = 1.0$ m = 2000kg

The proposed algorithm was applied to propagate the same initial distribution from Example I. using the filter dynamics over one orbital period. Additionally, 100,000 Monte Carlo samples were sampled from the initial distribution and propagated via the true simulation dynamics. Results are illustrated by Fig. 4 which shows a visual comparison between the proposed algorithm (solid line) and a Gaussian kernel density estimator (dashed line) for different position dimensions. The first 1000 Monte Carlo samples are shown in the plot (dots). The proposed algorithm required 633 components at the final time. The kernel density estimator is based on a 100×100 square grid of bins utilizing the Monte Carlo samples. The first two moments were computed for both the Monte Carlo and proposed filter solutions at the final time for quantitative comparison. The resulting means and state element standard deviations are provided in Eqs. (27) – (29). Euclidean distances are 3.68 km, 0.45 m/s between mean solutions and 1.0159 km, 0.1320 m/s between standard deviation solutions for position and velocity, respectively.

Monte Carlo:

$$\mathbf{m}_r = \begin{bmatrix} 28000.07 \\ 31.35 \\ -0.66 \end{bmatrix} km \qquad \sigma_{rr} = \begin{bmatrix} 2.1801 \\ 296.3976 \\ 0.9981 \end{bmatrix} km \tag{26}$$

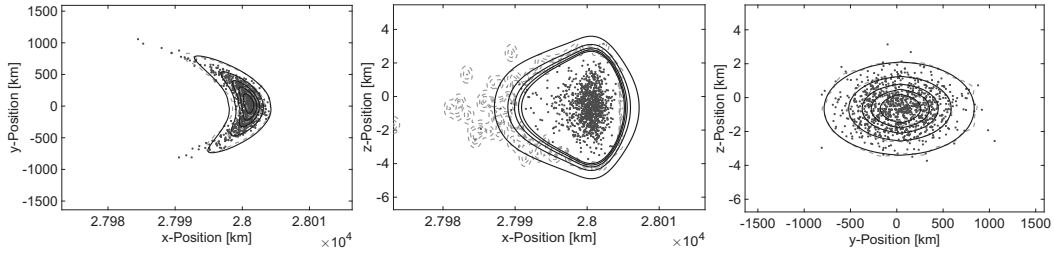
$$\mathbf{m}_v = \begin{bmatrix} -4.10 \\ 4132.79 \\ 0.063 \end{bmatrix} m/s \qquad \sigma_{vv} = \begin{bmatrix} 36.4843 \\ 1.0838 \\ 1.0099 \end{bmatrix} m/s \tag{27}$$

Proposed Algorithm:

$$\mathbf{m}_r = \begin{bmatrix} 28000.05 \\ 35.03 \\ -0.65 \end{bmatrix} km \qquad \sigma_{rr} = \begin{bmatrix} 1.8800 \\ 295.4271 \\ 1.0015 \end{bmatrix} km \tag{28}$$

$$\mathbf{m}_v = \begin{bmatrix} -4.55 \\ 4132.78 \\ 0.058 \end{bmatrix} m/s \qquad \sigma_{vv} = \begin{bmatrix} 36.3528 \\ 1.0881 \\ 0.9999 \end{bmatrix} m/s \tag{29}$$

It can be seen that the proposed algorithm compares favorably with the density estimator, but the results are obtained at a fraction of the computational cost of propagating $100,000$ samples.



(a) x vs. y (b) x vs. z (c) y vs. z

Figure 4. Comparison between proposed algorithm (solid line) and kernel density estimator (dashed line) for orbital uncertainty propagation example.

Example III. Cislunar Uncertainty Propagation

Consider the NRHO with apolune initial condition,

$$r_0 = \begin{bmatrix} 1.021340029542164 \\ 7.427823161812455e \times 10^{-8} \\ -1.81619990249204 \times 10^{-1} \end{bmatrix} \text{LU}, \quad v_0 = \begin{bmatrix} -8.944373429458848 \times 10^{-9} \\ 1.01759618312383 \times 10^{-1} \\ 4.569746377746858 \times 10^{-7} \end{bmatrix} \text{LU/TU}$$

Each value in r_0 and v_0 is defined in terms of the non-dimensionalized length units (LU) and time units (TU) in the circular-restricted three-body dynamics:

$$\begin{gathered} \dot{r}_x = v_x, \quad \dot{r}_y = v_y, \quad \dot{r}_z = v_z \\ \dot{v}_x = x + 2v_y - (1-\mu)(x+\mu)/r_e^3 - \mu(x-1+\mu)/r_m^3 \\ \dot{v}_y = y - 2v_x - (1-\mu)y/r_e^3 - \mu y/r_m^3 \\ \dot{v}_z = -(1-\mu)z/r_e^3 - \mu z/r_m^3 \end{gathered} \tag{30}$$

where

$$r_e = [(x+\mu)^2 + y^2 + z^2]^{\frac{1}{2}}$$
$$r_m = [(x-1+\mu)^2 + y^2 + z^2]^{\frac{1}{2}} \tag{31}$$

and $\mu = 1.2150584269940 \times 10^{-2}$ is the ratio of the moon's mass to the sum of the masses of the Earth and the moon. The period of the orbit is 1.50206 TU. We consider the uncertainty distribution at perilune, after 0.5 revolutions.

The initial distribution is Gaussian with covariance matrix,

$$P_0 = \text{diag}([10 \text{ km}, \quad 10 \text{ km}, \quad 10 \text{ km}, \quad 0.1 \text{ m/s}, \quad 0.1 \text{ m/s}, \quad 0.1 \text{ m/s}])^2 \tag{32}$$

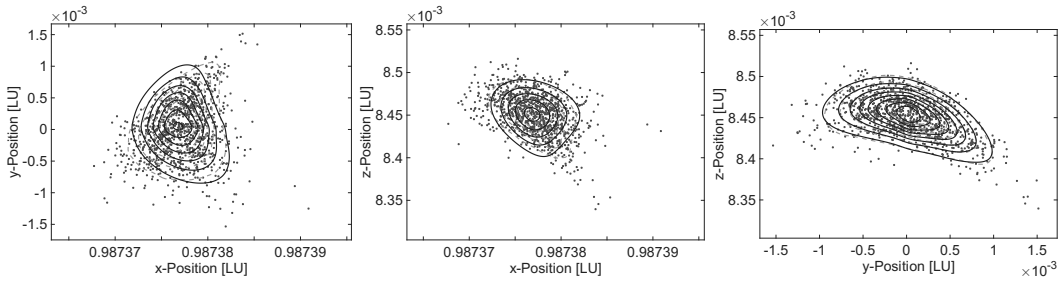
The process noise has power spectral density $10^{-20} I_{3\times 3} \frac{1}{TU}$. As in the previous example, the splitting library is shown in Table 3. The splitting threshold is shown in Eq. (9) with parameters $k = 1.5^2$ and $c = 0.35$. Figure 5 shows the results of the proposed algorithm compared to a kernel density estimate computed from 100,000 Monte Carlo points propagated at high fidelity, including gravitational accelerations relative to the the moon, the Earth, and the Sun, as well as solar radiation pressure [40]. Because the circular restricted three-body dynamics drift significantly from the high-fidelity dynamics, the mean of the GM distribution was adjusted to match the mean of the high-fidelity distribution. The proposed algorithm produced 117 components.

Monte Carlo:

$$\sigma_{rr} = \begin{bmatrix} 2.8277 \times 10^{-6} \\ 4.6662 \times 10^{-4} \\ 2.5260 \times 10^{-5} \end{bmatrix} LU \qquad \sigma_{vv} = \begin{bmatrix} 0.0049 \\ 0.0030 \\ 0.0472 \end{bmatrix} LU/TU \tag{33}$$

Proposed Algorithm:

$$\sigma_{rr} = \begin{bmatrix} 2.7112 \times 10^{-6} \\ 4.7399 \times 10^{-4} \\ 2.4558 \times 10^{-5} \end{bmatrix} LU \qquad \sigma_{vv} = \begin{bmatrix} 0.0051 \\ 0.0035 \\ 0.0506 \end{bmatrix} LU/TU \tag{34}$$



(a) x vs. y (b) x vs. z (c) y vs. z

Figure 5. Comparison between proposed algorithm (solid line) and kernel density estimator (dashed line) for cislunar example

CONCLUSIONS

The objective of this work is to provide accurate uncertainty quantification for a random process subject to nonlinear dynamics utilizing Gaussian mixture models. A new univariate splitting library was developed to minimize a measure of information loss (Kullback-Leibler divergence) while preserving the first two moments of the original Gaussian component before splitting. This was extended to multivariate components by applying the 1-D library along a single direction such that the first two moments are still preserved. The splitting direction chosen was one found to be associated with both large prior uncertainty and nonlinearity in the system dynamics. Throughout propagation, splitting was triggered automatically by an indication of unacceptable information loss incurred by the use of an linearized solution instead of a higher-order solution. This work focused on the propagation phase but could be augmented with existing schemes for the measurement phase.

ACKNOWLEDGMENT

This material is based on research sponsored by the Air Force Office of Scientific Research (AFOSR) under agreement number FA9550-23-1-0646, *Create the Future Independent Research Effort (CFIRE)*.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME – Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] P. Swerling, "First order error propagation in a stagewise smoothing procedure for satellite observations," *Journal of the Astronautical Sciences*, vol. 6, pp. 46–52, 1959.

[3] G. L. Smith, S. F. Schmidt, and L. A. McGee, "Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle," NASA, Technical Report R-135, 1962.

[4] S. F. Schmidt, "Applications of state space methods to navigation problems," *in C. T. Leondes, Editor, Advanced Control Systems*, vol. 3, pp. 293–340, 1966.

[5] Y. C. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.

[6] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, Jul. 1971.

[7] D. Alspach and H. Sorenson, "Nonlinear bayesian estimation using Gaussian sum approximations," *IEEE transactions on automatic control*, vol. 17, no. 4, pp. 439–448, 1972.

[8] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "A novel Gaussian sum filter method for accurate solution to the nonlinear filtering problem," in *Information Fusion, 2008 11th International Conference on*. IEEE, 2008, pp. 1–8.

[9] ——, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156, 2011.

[10] J. T. Horwood, N. D. Aragon, and A. B. Poore, "Gaussian sum filters for space surveillance: Theory and simulations," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 6, pp. 1839–1851, 2011.

[11] K. J. DeMars, R. H. Bishop, and M. K. Jah, "Entropy-based approach for uncertainty propagation of nonlinear dynamical systems," *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 4, pp. 1047–1057, 2013.

[12] M. Valli, R. Armellin, P. D. Lizia, and M. R. Lavagna, "Nonlinear mapping of uncertainties in celestial mechanics," *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 1, pp. 48–63, 2013.

[13] R. M. Weisman, M. Majji, and K. T. Alfriend, "Application of the transformation of variables technique for uncertainty mapping in nonlinear filtering," *Celestial Mechanics and Dynamical Astronomy*, vol. 118, no. 2, pp. 129–164, 2014.

[14] Y. Luo and Z. Yang, "A review of uncertainty propagation in orbital mechanics," *Progress in Aerospace Sciences*, vol. 89, pp. 23–39, 2017.

[15] D. M. Titterington, A. F. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions*. Wiley, 1985.

[16] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 1992.

[17] K. Tuggle and R. Zanetti, "Automated Splitting Gaussian Mixture Nonlinear Measurement Update," *Journal of Guidance, Control, and Dynamics*, vol. 41, no. 3, pp. 725–734, March 2018, doi: 10.2514/1.G003109.

[18] F. Faubel, J. McDonough, and D. Klakow, "The split and merge unscented Gaussian mixture filter," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 786–789, 2009.

[19] F. Faubel and D. Klakow, "Further improvement of the adaptive level of detail transform: Splitting in direction of the nonlinearity," in *Signal Processing Conference, 2010 18th European*. IEEE, 2010, pp. 850–854.

[20] V. Vittaldev and R. P. Russell, "Collision probability for space objects using Gaussian mixture models," in *Proceedings of the 23rd AAS/AIAA Space Flight Mechanics Meeting*, vol. 148. Advances in the Astronautical Sciences, Univelt San Diego, CA, 2013.

[21] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE, 2008, pp. 181–188.

[22] F. Havlak and M. Campbell, "Discrete and continuous, probabilistic anticipation for autonomous robots in urban environments," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 461–474, 2014.

[23] M. Raitoharju and S. Ali-Loytty, "An adaptive derivative free method for bayesian posterior approximation," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 87–90, 2012.

[24] M. F. Huber, "Adaptive Gaussian mixture filter based on statistical linearization," in *Information fusion (FUSION), 2011 proceedings of the 14th international conference on*. IEEE, 2011, pp. 1–8.

[25] G. Terejanu, "An adaptive split-merge scheme for uncertainty propagation using Gaussian mixture models," in *49th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, 2011, p. 890.

[26] V. Vittaldev and R. P. Russell, "Space object collision probability using multidirectional Gaussian mixture models," *Journal of Guidance, Control, and Dynamics*, 2016.

[27] M. L. Psiaki, J. R. Schoenberg, and I. T. Miller, "Gaussian sum reapproximation for use in a nonlinear filter," *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 2, pp. 292–303, 2015.

[28] M. L. Psiaki, "Gaussian mixture nonlinear filtering with resampling for mixand narrowing," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5499–5512, 2016.

[29] K. Vishwajeet and P. Singla, "Adaptive Split/Merge–Based Gaussian Mixture Model Approach for Uncertainty Propagation," *Journal of Guidance, Control, and Dynamics*, vol. 41, no. 3, pp. 603–617, 2018.

[30] F. Pérez-Cruz, "Kullback-leibler divergence estimation of continuous distributions," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1666–1670.

[31] K. E. Tuggle, "Model selection for gaussian mixture model filtering and sensor scheduling," Ph.D. dissertation, 2020.

[32] J. Kulik and K. A. LeGrand, "Nonlinearity and uncertainty informed moment-matching Gaussian mixture splitting," *arXiv preprint arXiv:2412.00343*, 2024.

[33] G. A. Siciliano, K. A. LeGrand, and J. Kulik, "Higher-order tensor-based deferral of gaussian splitting for orbit uncertainty propagation," *arXiv preprint arXiv:2507.01771*, 2025.

[34] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Sciences and Engineering. New York, New York 10003: Academic Press, 1970, vol. 64.

[35] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction," *Journal of Basic Engineering*, vol. 83, no. Series D, pp. 95–108, March 1961, doi:10.1115/1.3658902.

[36] K. Tuggle, R. Zanetti, and C. D'Souza, "A Splitting Gaussian Mixture Formulation for a Nonlinear Measurement Update," in *Proceedings of the AAS/AIAA Space–Flight Mechanics Meeting*, San Antonio, TX, Feb 5–9, 2017 2017, aAS 17-430.

[37] K. J. DeMars, Y. Cheng, R. H. Bishop, and M. K. Jah, "Methods for splitting Gaussian distributions and applications within the AEGIS filter," in *Advances in the Astronautical Sciences*, vol. 143, 2012, pp. 2379–2398.

[38] V. Vittaldev and R. P. Russell, "Multidirectional Gaussian mixture models for nonlinear uncertainty propagation," *Computer Modeling in Engineering & Sciences*, vol. 111, no. 1, pp. 83–117, 2016.

[39] J. D. Hand, L. N.; Finch, *Analytical Mechanics*. Cambridge University Press, 2008, ISBN 978-0-521-57572-0.

[40] K. J. DeMars, M. Gupta, R. Zanetti, and K. Michaelson, "Bifidelity uncertainty propagation with directional splitting for space domain awareness," in *International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE, 2025.