 CMES, vol.111, no.1, pp.83-117, 2016

Multidirectional Gaussian Mixture Models for Nonlinear Uncertainty Propagation

V. Vittaldev[1] **and R. P. Russell**[2]

Abstract: Monte Carlo simulations are an accurate but computationally expensive procedure for approximating the resultant non-Gaussian probability density function (PDF) after propagation of an initial Gaussian PDF through a nonlinear function. Univariate splitting libraries for Gaussian Mixture Models (GMMs) exist with up to five elements in the literature. The number of splits are extended in the present work by generating three homoscedastic univariate splitting libraries with up to 39 elements. Mulitvariate GMMs are typically handled with splits along a single direction. Instead, we generate a regular multidirectional grid over the initial multivariate Gaussian distribution by recursively applying the splitting library along multiple directions. The splitting direction is arbitrary and no longer limited to directions parallel to the columns of the square-root of the covariance matrix. A second order Stirling's interpolation of the nonlinear function evaluated at the mean of the initial Gaussian distribution is used to quantify nonlinearity along candidate splitting directions. The directions with the highest nonlinearity benefit most from splitting. The Multidirectional GMM (MGMM) has applications for uncertainty quantification with computationally intensive nonlinear functions. The variable number of splits in each direction allows for a spectrum of models in the accuracy versus compute time design space, filling the gap between expensive Monte Carlos and fast linearized models. The multidirectional method is demonstrated with four test cases, including an orbit uncertainty propagation case, to illustrate the benefit of splitting along multiple directions and of ranking the splitting directions.

Keywords: Uncertainty Quantification, Gaussian Mixture Models.

[1] Ph.D. Candidate, The University of Texas at Austin, Austin, TX, USA.
E-mail: v.vittaldev@utexas.edu

[2] Associate Professor, The University of Texas at Austin, Austin, TX, USA.
E-mail: ryan.russell@utexas.edu

1 Introduction

A Gaussian distribution is a good assumption for the initial uncertainty in parameters for a large number of problems due to the central limit theorem [Feller (1945)]. Since a Gaussian distribution can completely be characterized by the first two statistical moments, the size of the uncertainty is encapsulated in the covariance. A nonlinear function maps the probability density function (PDF) of the uncertainty of the input parameters onto the uncertainty PDF of the output parameters. The resulting distribution is no longer Gaussian even if the initial distribution is completely Gaussian. Only for a linear function of the parameters will the output uncertainty distribution remain Gaussian. A Gaussian assumption may still be a good approximation for some cases but more information about the PDF is required for other cases.

An infinite number of statistical moments of the initial distribution would have to be propagated through the function to accurately represent the final PDF. In practice, an arbitrary number of moments can be captured, by using quadrature methods for example [Stroud and Secrest (1966)]. An approximation of the final non-Gaussian PDF can be found by using the principle of maximum entropy [Abramov (2007)]. For nonlinear dynamic systems the exact evolution of the PDF is the solution of the Fokker-Planck Equation (FPE) [Fuller (1969)]. Unfortunately, the FPE is a partial differential equation that is difficult to implement and computationally intensive to solve. There have been, however, recent efforts to solve the FPE in a computationally tractable manner [Sun and Kumar (2014); Kumar and Chakravorty (2012); Kumar, Chakravorty, Singla, and Junkins (2009); Terejanu, Singla, Singh, and Scott (2008, 2011); Vishwajeet, Singla, and Jah (2014)].

The easiest approach of approximating the resulting non-Gaussian distribution is Monte Carlo (MC) simulation. Random initial conditions are sampled from the distribution of the input parameters. The function is evaluated at the initial conditions to create a sample of the output. Information about the PDF of the output, such as the shape and statistical moments can then be extracted from the sample. With increasing dimension, MC simulation becomes computationally expensive because a large number of function evaluations are required due to the slow convergence rate [Cowles and Carlin (1996)]. Function evaluations can also be parallelized on multi-core processors or on Graphics Processing Units (GPUs) to reduce overall computation time. A reduction in the required number of MC samples can be achieved by importance sampling in cases when certain input parameters have a large impact on the function output [Siegmund (1976)]. Most importance sampling algorithms can unfortunately not be easily parallelized, but their parallelization is an active topic of research [Calderhead (2014)].

Surrogate models are an active area of research that seek to approximate the nonlinear function by a simpler and computationally less expensive model [Butler, Dawson, and Wildey (2013)]. The surrogate model can be evaluated in lieu of the full nonlinear function to sample the final distribution of the output parameters. There are many choices for the basis functions used for the surrogate model. One surrogate model approach for uncertainty propagation uses multivariate orthogonal polynomials i.e. Polynomial Chaos (PC) [(Jones, Doostan, and Born, 2013)]. The polynomial basis functions can be chosen based on the initial uncertainty according to the Weiner-Askey scheme [Wan and Karniadakis (2006)]. For an initial Gaussian distribution, the optimal choice is Hermite polynomials [Wiener (1938)]. The surrogate model forms a response surface and determining the coefficients of the polynomials is computationally less expensive than running a full MC simulation. PC methods can, however, be more computationally expensive than MC methods for some cases. The growth of the basis function is combinatorial in nature as the state dimension or the order of the polynomial is increased [Wiener (1938)]. Computation of the coefficients typically requires quadrature methods, which grow exponentially with state dimension. The computational load due to a quadrature can be reduced in some cases by using sparse grids [Jones, Parrish, and Doostan (2015)].

Assuming the uncertainty in the output parameters is a Gaussian distribution enables the use of analytical approximations because only the first two statistical moments of the initial PDF have to be propagated through the function. This Gaussian assumption is the basis for many filtering and state estimation applications. The initial covariance can be propagated using a Taylor series approximation of the function [Gelb (1974)], or by evaluating the function at deterministically chosen sigma points [Julier and Uhlmann (2004); Norgaard, Poulsen, and Ravn (2000); Adurthi, Singla, and Singh (2012)]. If these Gaussian propagation techniques are used, the maximum information from the resulting non-Gaussian distribution is an approximation of the first two statistical moments. Even if the mean and covariance are accurately captured by the Gaussian propagation techniques, there is no guarantee that the shape of the PDF is accurate since the isoprobability contours are assumed to be ellipsoids. Higher order sigma point methods such as the Conjugate Unscented Transform (CUT) [Adurthi, Singla, and Singh (2012); Adurthi and Singla (2015)] provide the mechanisms to capture higher order moments, and the isoprobability contours are no longer ellipsoids.

A Gaussian Mixture Model (GMM) works under the proposition that any PDF can be approximated in terms of the L_1-distance by using a weighted sum of Gaussian probability distribution functions [Alspach and Sorenson (1972)], where the sum of the weights is unity. GMMs are used to discretize the initial Gaussian PDF into

weighted Gaussian PDFs with a smaller covariance. Each element is propagated through the function using simplified methods such as the sigma point methods to enforce the Gaussian condition on the post-propagation PDF. The weighted sum of the individual propagated Gaussian elements better approximates the final non-Gaussian PDF as the number of elements is increased, while requiring fewer function evaluations than MC. Therefore, Gaussian Mixture Models form a compromise between the analytical approximations of the Gaussian propagation methods and the computationally intensive MC technique. An added benefit of GMMs is the availability of an analytical PDF for the GMM approximation, which is not the case for surrogate models such as PC.

Converting the initial Gaussian distribution into a GMM is typically achieved by using a univariate splitting library. The univariate splitting library is a GMM that approximates the standard normal distribution. The library is then applied along a column of the square-root (Spectral or Cholesky decomposition) of the initial covariance matrix to generate a multivariate GMM approximation of the original Gaussian distribution. The multivariate GMM elements are located on, and have a reduced variance along, the splitting direction. It is noted that the size of a univariate splitting library available in literature is limited to five [DeMars, Bishop, and Jah (2013)]. For multivariate cases, the univariate libraries are applied along all the n directions of freedom for an n-dimensional problem [Horwood, Aragon, and Poore (2011)]. Therein lies the motivation of the current work. Firstly, to increase the number of splits for the univariate case. Secondly to identify the directions that have the greatest effect on the non-Gaussianity of the propagated distribution and to split only along those directions.

Much of the active research in GMMs is in the applied fields of machine learning [Li, Prasad, and Fowler (2014); Baspinar, Varol, and Senyurek (2013)] where a GMM is fitted to a given sample from a non-Gaussian distribution using Expectation Maximization (EM) [Dempster, Laird, and Rubin (1977)]. The work in the current paper, however, deals with the more basic question of how to generalize the number and directions of the splits of the initial Gaussian distribution into a GMM. Splitting and merging of Gaussian elements for the approximation and propagation of an initial Gaussian distribution has also been the focus of some recent work. Gaussian distributions have been used as Radial Basis Functions in a Neural Network for nonlinear system identification in Li, Sundaragan, and Saratchandran (2000). GMMs have been used for nonlinear state estimation in Hanebeck, Briechle, and Rauh (2003). A four element univariate splitting library is provided by Huber, Bailey, Durrant-Whyte, and Hanebeck (2008). A procedure for approximating a GMM with another GMM using fewer elements is shown in Huber and Hanebeck (2008). An uncertainty propagation scheme utilizing both splitting and merging of GMM

elements for a dynamic system has been shown by Terejanu (2011). GMMs have been used for uncertainty propagation through dynamical systems with the number of elements adapted to minimize the FPE by Vishwajeet and Singla (2013, 2014).

The nonlinear orbital equations of motion cause an initial Gaussian state uncertainty distribution of space objects to become non-Gaussian with increasing flight time and perturbations. Capturing the evolution of the state uncertainty is an essential part of operations including orbit determination (OD) and conjunction assessment (CA). There exists a framework that efficiently, with varying degrees of accuracy, handles uncertainties that are assumed to remain Gaussian in OD [Tapley, Schutz, and Born (2004)] and in CA [Chan (2008)]. The same framework is readily extended to GMMs since each element remains Gaussian. A Gaussian Sum Filter (GSF) has been used to improve the performance of Gaussian filtering algorithms such as the EKF for nonlinear non-Gaussian problems [Alspach and Sorenson (1972)]. GMMs have also been used to propagate state uncertainties through nonlinear ODEs [Terejanu, Singla, Singh, and Scott (2008); Horwood, Aragon, and Poore (2011); DeMars, Bishop, and Jah (2013); Vishwajeet, Singla, and Jah (2014)]. Psiaki, Schoenberg, and Miller (2015) show a resampling strategy for GMMs to efficiently approximate one GMM with another so that the particles in a Particle Filter (PF) are replaced with a GMM. Most recently, GMMs have also been used to improve collision probability computations [Vittaldev and Russell (2013); Vittaldev and Russell (2016); DeMars and Jah (2014)]. Fujimoto and Scheeres (2015) use state transition tensors (STTs) as a computationally efficient surrogate model to the real dynamics to propagate the GMMs for a CA problem.

In Section 2, a univariate library for splitting the standard normal distribution into a GMM by minimizing the L_2 norm is presented. Solutions for predetermined multiple splitting rules are found via nonlinear optimization and archived for up to 39 univariate splits. The new solutions are the largest libraries in the literature to date. The same optimization procedure can also be used to compute more libraries by changing the constraints on the standard deviations of the elements. The splitting library is then applied along multiple directions of the covariance matrix to extend the theory to multivariate Gaussian distributions in Section 3. A second order Stirling's interpolation formula evaluated at the mean of the original Gaussian distribution is used to rank the candidate directions with respect to the nonlinearity. Also in Section 3, the number of splits along a given direction is chosen based on the relative degree of nonlinearity. The performance of the MGMMs is shown with some numerical test cases, including the uncertainty propagation of an object in Geostationary orbit (GEO), in Section 4.

2 Univariate Splitting Library

Before tackling multivariate scenarios, the univariate case is considered. To increase the applicability of the univariate GMM, the goal of this section is to develop splitting libraries. The libraries are carefully fit in advance and also have a larger number of elements than are currently available in the literature. A tool is created that generates a univariate splitting library based on the standard normal distribution. The library consists of the weights α_i, means μ_i, and the standard deviations σ_i for a desired number of elements that approximate the standard normal distribution. A GMM generated from the standard normal distribution can be shifted and scaled to fit any univariate Gaussian distribution with an arbitrary mean and standard deviation. The univariate splitting library is applied along a desired direction to approximate any multivariate Gaussian distribution as shown in more detail in Section 3.

2.1 Generating a Univariate Split

In developing the univariate library, a performance index is necessary to measure the difference between the exact Gaussian distribution and the approximated GMM. The weights, and location and magnitude of the means are optimization variables in a nonlinear programming problem. The performance index ideally should result in a value of zero if and only if the two distributions are identical. The result should be greater than or equal to zero for any two distributions [Hershey and Olsen (2007)]. There are many divergences that could be used to quantify the dissimilarity between two distributions such as the Kullback-Leibler [Kullback and Leibler (1951)] divergence, the Jensen-Shannon divergence [Endres and Schindelin (2003)]. A distance metric is the L_p distance [Horwood, Aragon, and Poore (2011)] between 2 distributions p_1 and p_2 is:

$$L_1(p_1, p_2) = \int_{\Omega} |p_1(\mathbf{x}) - p_1(\mathbf{x})|^p \, \mathrm{d}\mathbf{x} \tag{1}$$

2.2 Optimization

Although the GMMs approximate PDFs better with respect to the L_1 distance [Alspach and Sorenson (1972)], the L_2 distance is chosen as the dissimilarity metric because the solution is analytical and does not require solving a quadrature in the case of a GMM and a Gaussian distribution. The L_2 distance between a GMM PDF

p and a Gaussian distribution $p_g(\mathbf{x};\boldsymbol{\mu}_g,\mathbf{P}_g)$ is [DeMars, Bishop, and Jah (2013)]:

$$\begin{aligned} L_2(p,p_g) = &|4\pi\mathbf{P}_g|^{-1/2} + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K\left(\boldsymbol{\mu}_i,\boldsymbol{\mu}_j,\mathbf{P}_i,\mathbf{P}_j\right) \\ &- 2\sum_{i=1}^{N}\alpha_i K\left(\boldsymbol{\mu}_i,\boldsymbol{\mu}_g,\mathbf{P}_i,\mathbf{P}_g\right) \end{aligned} \tag{2}$$

where the operation defined by K is

$$\begin{aligned} K\left(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\mathbf{P}_1,\mathbf{P}_2\right) = &\left|2\pi\left(\mathbf{P}_1+\mathbf{P}_2\right)\right|^{-1/2} \\ &\times\exp\left(-\frac{1}{2}\left(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\right)^T\left(\mathbf{P}_1+\mathbf{P}_2\right)^{-1}\left(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\right)\right) \end{aligned} \tag{3}$$

α is a weight, $\boldsymbol{\mu}$ is a mean, and $\mathbf{P}$ is a covariance matrix. The weights are computed by minimizing the L_2 distance between p and p_g [Horwood, Aragon, and Poore (2011); DeMars, Bishop, and Jah (2013); Huber, Bailey, Durrant-Whyte, and Hanebeck (2008)]. Minimizing the L_2 distance becomes a constrained nonlinear optimization problem, with the cost function from Eq. 2 and the constraints from Eq. 5 and Eq. 6. In the existing literature, univariate libraries have been precomputed for $N = 4$ [Huber, Bailey, Durrant-Whyte, and Hanebeck (2008)] and $N = 3,4,5$ [DeMars, Bishop, and Jah (2013)]. To compute libraries with a higher N, a trust-region optimization [Conn, Gould, and Toint (1987)] algorithm is currently implemented, and solutions are found with odd N up to 39. There are N σ_i, N α_i, and N μ_i, totaling $3N$ variables to ultimately choose, where N is the desired number of splits.

Simplifications are made to reduce the number of free parameters in the optimization problem from $3N$ to $N-1$. All the components of the GMM are given the same standard deviation by specifying a rule for σ as a function of N. The homoscedastic assumption may seem restrictive, but assumptions are required to generate the univariate splitting libraries. Without any assumption, the optimization problem has an undesired trivial solution. The central element will have a weight of 1 and a standard deviation of 0. All the other elements will have a weight of 0 and the L_2 distance will also be 0. With the assumption of homoscedasticity, the cost function for minimization is a simplified version of the L_2 distance from Eq. 2:

$$\begin{aligned} J = &\frac{1}{2\sqrt{\pi}} + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\alpha_i\alpha_j}{2\sqrt{\pi\sigma^2}}\exp\left(\frac{-(\mu_i-\mu_j)^2}{4\sigma^2}\right) \\ &- 2\sum_{i=1}^{N}\frac{\alpha_i}{\sqrt{2\pi\left(\sigma^2+1\right)}}\exp\left(\frac{-\mu_i^2}{2\left(\sigma^2+1\right)}\right) \end{aligned} \tag{4}$$

After choosing a rule for σ, the remaining $2N$ free parameters are the weights α_i and the means μ_i. It should be noted that homoscedasticity assumption is only for the initial GMM. The variance of each element changes due to propagation through a nonlinear function. Assuming that the location and the weights of the elements are symmetric further reduces the number of free parameters to $N-1$. Details on the constraints and the remaining free parameters for odd N are presented here. The setup of the optimization for even N is similar but only odd N are computed to ensure that the location of the element with the largest weight coincides with the actual mean of the normal distribution. The $N+1$ equality constraints are:

$$0 = \mu_0 \tag{5a}$$

$$0 = \mu_i - \mu_{-i} \qquad i = 1,\ldots,(N-1)/2 \tag{5b}$$

$$0 = \alpha_i - \alpha_{-i} \qquad i = 1,\ldots,(N-1)/2 \tag{5c}$$

$$0 = 1 - \alpha_0 - 2\sum_{i=1}^{(N-1)/2} \alpha_i \tag{5d}$$

Inequality constraints are imposed to enforce a monotonic increase in the location of the elements and a monotonic decrease in the weights away from the center. The weights decrease monotonically as the distance from the center increases so that elements closer to the mean of the initial Gaussian distribution, i.e. areas with a high PDF value, have a higher importance, because the goal is to approximate the bell curve shown in Fig. 1.

$$\mu_{i-1} - \mu_i < 0 \qquad i = 1,\ldots,(N-1)/2 \tag{6a}$$

$$\alpha_{i-1} - \alpha_i < 0 \qquad i = 1,\ldots,(N-1)/2 \tag{6b}$$

Fig. 1 illustrates the effect of the optimization equality and inequality constraints on the PDFs of the individual elements for $N = 7$.

A trust-region optimizer is used due to its robustness in solving nonlinear optimization problems [Byrd, Schnabel, and Schultz (1987)]. Like a line search, the trust-region method enforces that each new iteration reduces the objective function. However, a maximum step distance (the trust-region) is first chosen, followed by the direction of the step. The trust-region is centered on the present iterate and the complicated nonlinear function is replaced by a simpler, quadratic approximation. Choosing the step direction translates to finding the minimum of the quadratic approximation. If the Hessian of the objective function is positive definite and the trust-region is large enough, then the step is identical to the Newton-Rhapson update. The details of computing the update step and the size of the regularly updated trust-region are beyond the scope of this paper. Trust-region optimization is an

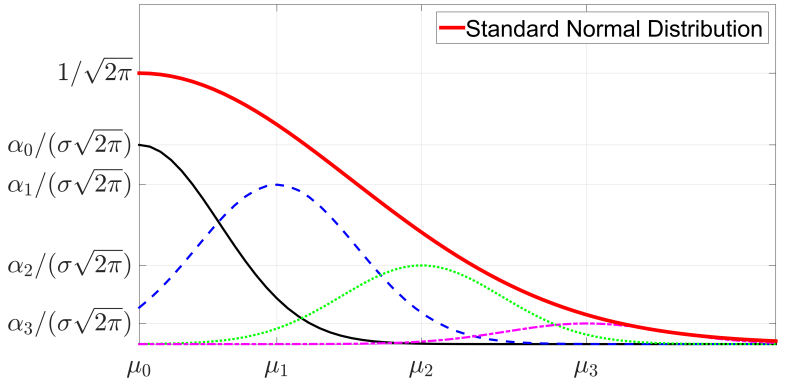


Figure 1: PDFs of the individual elements subject to the equality and inequality constraints for $N = 7$

active area of research and many variations can be found in literature [Moré and Sorensen (1983)].

The trust-region optimization algorithm requires the first and second derivatives of the objective function and the constraints with respect to the state variables. Analytical expressions for the Jacobians and Hessians are computed using the symbolic manipulator, Maple. The optimization routine is coded in Fortran due to the computational speed and the ability to use quadruple precision arithmetic, which is required for convergence when the number of splits, N is high. During the optimization, the GMM weights and means from the converged solution of the $N-2$ case are used as an initial guess for the current N case. Furthermore, a heuristic predictor step is employed to modestly increase the spread on the means, and lower the weight values. An extra element is added to the tail and all the weights are lowered so that the sum remains unity for the initial guess. The problem is highly nonlinear and convergence is increasingly difficult as N increases.

The computation of the trust-region optimization, takes approximately 3 minutes for N up to 39 on a 3.07 GHz Intel Xeon CPU. The trust-region and function routines are compiled using Intel Visual Fortran Composer XE 2011 and optimization settings -O3.

2.3 Resulting Univariate Libraries

Solutions for three different σ rules are found and archived online [1]:

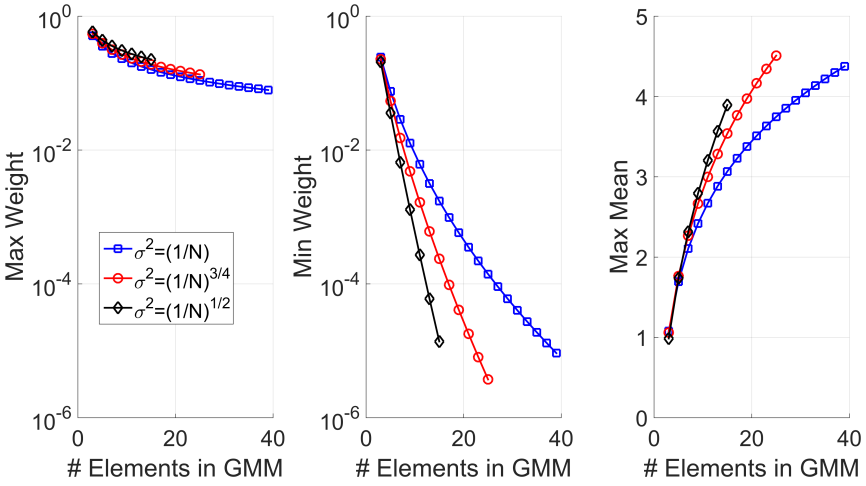
[1] `http://russell.ae.utexas.edu/code/GMMsplittingLibrary.txt/`

1. $\sigma^2 = (1/N)$ for odd-valued $N \leq 39$
2. $\sigma^2 = (1/N)^{3/4}$ for odd-valued $N \leq 25$
3. $\sigma^2 = (1/N)^{1/2}$ for odd-valued $N \leq 15$

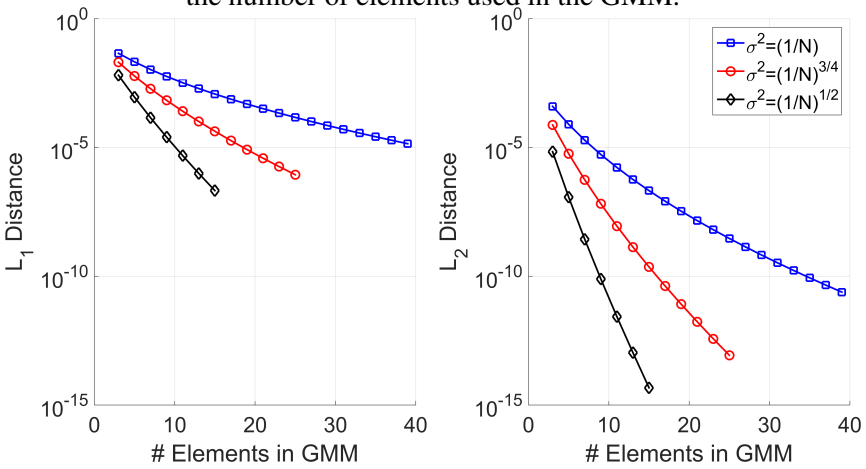
The rules for σ are selected arbitrarily and the optimization process can easily be repeated to generate libraries with different σ rules. An example of a 7 component univariate library is presented in Tab. 1. The L_1 and L_2 distances between the optimized solutions and the standard normal distribution, as a function of the number of the elements are found in Fig. 2b. Libraries with up to 39 elements have been created using an intensive optimization process that solves the highly nonlinear cost function with many local minima. The issue with increasing the number of elements in the library is mainly numeric sensitivity because the problem becomes too large and sensitive. It could be tackled with new formulations for the optimization problem that use better heuristics for initial guesses and decompose the problem into smaller, less sensitive, subproblems. A brute force approach is to solve the optimization using higher precision than quadruple. As the problem stands, however, the upper limit on the number of elements has been achieved using the current optimization strategy and quadruple precision.

It is difficult to view the splits when N is large because the size of individual PDFs in the tails becomes too small. Therefore, a good way to visualize the split is to plot the maximum and minimum weight, and the maximum value of the mean as in Fig. 2a. Since only an odd number of elements are used, the weights are symmetric about 0 and the means are antisymmetric. Also, the weights monotonically decrease from the center to the tails, and the absolute values of the means increase. The maximum mean value shows how far the furthest split is from the center, indicating how spread out the elements are. The maximum and minimum weights show the difference between the weight of the central element compared to the weight of the element at the tail end, which indicates the importance given to the central elements compared to the ends. For the various rules, Rule 1 with $\sigma^2 = (1/N)$ results in the smallest σ and Rule 3 with $\sigma^2 = (1/N)^{1/2}$ results in the largest σ for a given N. Fig. 2 shows that Rule 1 has the lowest weight of the different rules for the central mean compared to the other rules. The element with the lowest weight, i.e. the element that is furthest from the center, is closer to the center than for the other rules. For a given N, as the exponent on $(1/N)$ decreases for the various rules, the L_1 distance decreases. However, the distribution of the weights is sharper since the standard deviation of the computed weights for a given N increases.

The benefit of generating a complete high dimensional split is seen in Fig. 3, where a complete 9 element split is compared to a 9 element GMM created by recursively



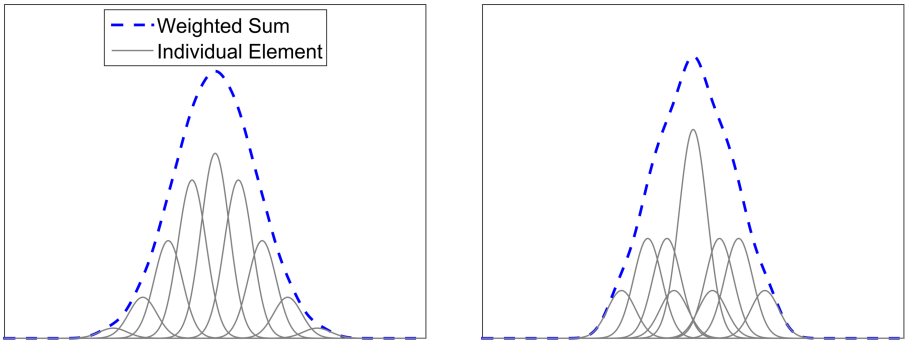
(a) Maximum and minimum weights, and the maximum mean as a function of the number of elements used in the GMM.



(b) L_1 and L_2 distances as a function of the number of elements used in the GMM.

Figure 2: Properties of the univariate splitting libraries for various rules for σ

splitting all the elements of a 3 element split [DeMars, Bishop, and Jah (2013)]. The standard normal distribution is better approximated, i.e. the L_2 distance is two orders of magnitude lower, and the weights and means are distributed more evenly.



(a) Complete 9 element split
$L_2 = 5.3 \times 10^{-6}$

(b) Recursive 3×3 split
$L_2 = 5.4 \times 10^{-4}$

Figure 3: PDFs of the univariate 9 element GMM of the standard normal distribution using a complete and a recursive splitting technique for $\sigma^2 = (1/N)$, and their L_2 distances from the standard normal distribution.

3 Multivariate GMMs

Uncertainty analysis typically requires a multivariate state. Therefore, it is necessary to split an initial multivariate Gaussian distribution. A univariate library can be applied along a specified direction in order to split a multivariate distribution. The direction choice is along a column of the square-root matrix, which is typically found using Cholesky or spectral decomposition (Huber, Bailey, Durrant-Whyte, and Hanebeck, 2008). The following subsection outlines the method for splitting along an arbitrary direction.

3.1 Splitting along an Arbitrary Direction

Spectral and Cholesky decomposition are two of the most common methods of computing the square-root $\mathbf{S}$ of the covariance matrix $\mathbf{P} = \mathbf{S}\mathbf{S}^T$. $\mathbf{S}$ enables a coordinate transformation to a new reference frame where the individual variables of the initial multivariate random variable are independent and identically distributed (i.i.d.) described by the standard normal distribution. The initial reference fame of the multivariate state and the square-root frame where the variables are i.i.d. are represented by $\mathcal{F}_I$ and $\mathcal{F}_S$, respectively. The transformation matrix from $\mathcal{F}_S$ to $\mathcal{F}_I$ is:

$$R_S^I = \mathbf{S} \tag{7}$$

This transformation matrix includes rotation and scaling.

Table 1: The 7-component univariate splitting library with $\sigma^2 = \left(\frac{1}{N}\right)$

Component	w	μ
1	0.028799777829539	−2.107361692265483
2	0.109875486136781	−1.329872113204359
3	0.222379075167735	−0.648762460764688
4	0.277891321731891	0
5	0.222379075167735	0.648762460764688
6	0.109875486136781	1.329872113204359
7	0.028799777829539	2.107361692265483

Let **a** be the direction along which the univariate splitting library is applied in $\mathscr{F}_I$. When **a** is parallel to the k^{th} column of **S** ($\mathbf{a} \parallel \mathbf{S}(:,k)$), and given a univariate splitting library mean μ_i and standard deviation σ_i, the multivariate mean and covariance are provided by Huber, Bailey, Durrant-Whyte, and Hanebeck (2008):

$$\boldsymbol{\mu}_i = \boldsymbol{\mu} + \mu_i \mathbf{S}(:,k) \tag{8a}$$

$$\mathbf{S}_k = \mathbf{S} \qquad \mathbf{S}_k(:,k) = \sigma_i \mathbf{S}(:,k) \tag{8b}$$

$$\mathbf{P}_i = \mathbf{S}_k \mathbf{S}_k^T \tag{8c}$$

where $\mathbf{S}_k$ is a copy of **S**, but with the k^{th} column multiplied by σ_i.

In case **a** is not parallel to any of the columns of the Cholesky or spectrally decomposed square-root matrix, a square-root matrix is constructed where $\mathbf{S}_\star(:,1) \parallel \mathbf{a}$ (Aristoff, Horwood, Singh, and Poore, 2014). A new reference frame $\mathscr{F}_A$ is defined such that the difference between $\mathscr{F}_A$ and $\mathscr{F}_S$ is a pure rotation and the first axis of $\mathscr{F}_A$ is parallel to **a**. Therefore, the variables in $\mathscr{F}_A$ are also i.i.d. Finally, the new square-root matrix and the mean and covariance matrix are:

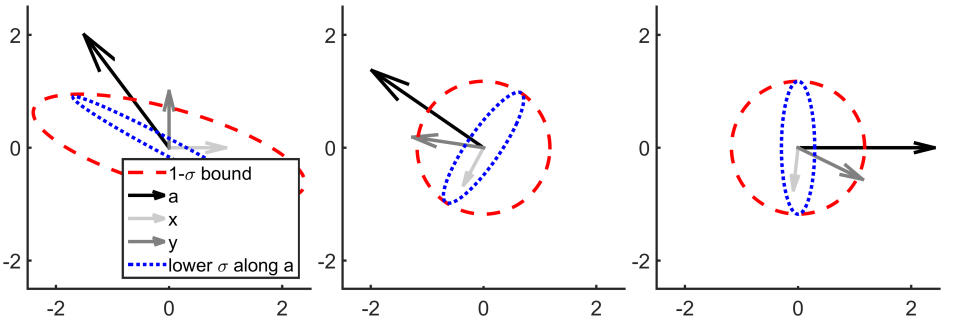
$$\mathbf{S}_\star = R_S^I R_A^S \tag{9a}$$

$$\boldsymbol{\mu}_i = \boldsymbol{\mu} + \mu_i \mathbf{S}_\star(:,1) \tag{9b}$$

$$\mathbf{P}_i = \mathbf{S}_\star \mathbf{S}_\star^T \tag{9c}$$

The rotation matrix R_S^A from $\mathscr{F}_S$ to $\mathscr{F}_A$ is found using Gram-Schmidt orthogonalization. The three reference frames with the confidence bounds for the covariance matrix and reduced variance along **a** are shown in Fig. 4 for a bivariate case.

A newly derived alternative method of computing the new covariance matrix is now shown where Gram-Schmidt orthogonalization is not required. The mean and



(a) $\mathscr{F}_I$: Original frame (b) $\mathscr{F}_S$: Intermediate Cholesky frame (c) $\mathscr{F}_A$: Desired square-root frame

Figure 4: The unit axes of $\mathscr{F}_I$, x and y, and the desired splitting direction, $\mathbf{a}$, in the three relevant reference frames for a 2-dimensional case.

covariance of the GMM elements are simple analytical equations. In $\mathscr{F}_S$, the state is i.i.d. Applying a univariate split along $\mathbf{a}$ in $\mathscr{F}_I$ is analogous to changing the standard deviation of the i.i.d. distribution in the $\hat{\mathbf{a}}^\star$ direction from 1 to σ in Fig. 4b. The unit vector $\hat{\mathbf{a}}^\star$ is simply the direction of $\mathbf{a}$ expressed in $\mathscr{F}_S$:

$$\hat{\mathbf{a}}^\star = \frac{\mathbf{S}^{-1}\mathbf{a}}{\|\mathbf{S}^{-1}\mathbf{a}\|_2} \tag{10}$$

Reducing the standard deviation along $\hat{\mathbf{a}}^\star$ is a linear transformation:

$$\mathbf{y} = \mathbf{x} - \left(\mathbf{x}^T\hat{\mathbf{a}}^\star\right)\hat{\mathbf{a}}^\star + \sigma\left(\mathbf{x}^T\hat{\mathbf{a}}^\star\right)\hat{\mathbf{a}}^\star = \mathbf{\Phi}\mathbf{x} = \left(\mathbf{I} + (\sigma-1)\hat{\mathbf{a}}^\star\hat{\mathbf{a}}^{\star T}\right)\mathbf{x} \tag{11}$$

where $\mathbf{y}$ is the variable with a standard deviation of σ in the $\hat{\mathbf{a}}^\star$ direction and $\mathbf{x}$ is the i.i.d. state in $\mathscr{F}_S$ with the covariance matrix $\mathbf{P}_{\mathbf{xx}} = \mathbf{I}$. The covariance matrix of $\mathbf{y}$ expressed in $\mathscr{F}_S$ is then:

$$\mathbf{P}_{\mathbf{yy}} = \mathbf{\Phi}\mathbf{P}_{\mathbf{xx}}\mathbf{\Phi}^T = \left(\mathbf{I} + (\sigma^2-1)\hat{\mathbf{a}}^\star\hat{\mathbf{a}}^{\star T}\right) \tag{12}$$

$\mathbf{P}_{\mathbf{yy}}$ expressed in $\mathscr{F}_I$ is the covariance matrix of the i^{th} element of the multivariate GMM due to applying the univariate splitting library mean μ_i and standard deviation σ_i along the splitting direction $\mathbf{a}$. The multivariate mean and covariance matrix are:

$$\boldsymbol{\mu}_i = \boldsymbol{\mu} + \mu_i\mathbf{S}\hat{\mathbf{a}}^\star \tag{13a}$$

$$\mathbf{P}_i = \mathbf{S}\left(\mathbf{I} + (\sigma_i^2-1)\hat{\mathbf{a}}^\star\hat{\mathbf{a}}^{\star T}\right)\mathbf{S}^T \tag{13b}$$

where $\hat{\mathbf{a}}^\star$ is computed from Eq. 10.

3.2 Importance of Splitting Direction using a 2D Example

It is possible to apply the univariate splitting library along any direction to generate a multivariate GMM. However, the benefit of splitting is highly sensitive to the choice of the direction. Using an example for visualization purposes, the sensitivity of choosing a splitting direction is investigated on a nonlinear transformation of a bivariate Gaussian distribution from [Haario, Laine, Mira, and Saksman (2006)]:

$$y_1 = ax_1 \tag{14a}$$

$$y_2 = x_2/a - b(a^2x_1^2 + a^2) \tag{14b}$$

In Eq. 14, a and b are parameters that control the nonlinearity of the target. The mean and initial covariance for (x_1, x_2) are chosen to be:

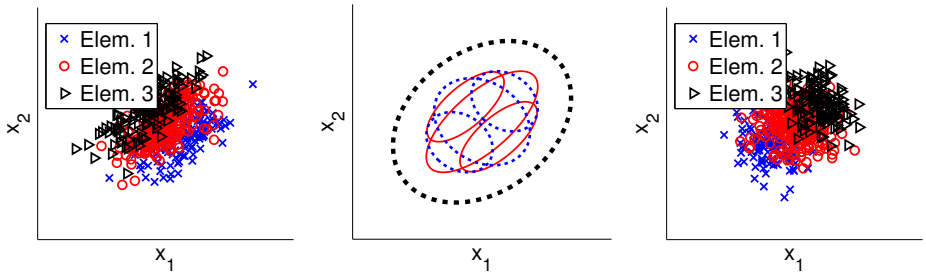
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mathbf{P_x} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix} \tag{15}$$

To limit the number of direction choices, spectral decomposition is used to create the splits along the two eigenvectors of the covariance matrix found in Eq. 15. Samples generated from the GMMs with splitting along the minor and major axes of the covariance matrix are seen in the point clouds found in Fig. 5a and Fig. 5c. The distributions are plotted as ellipses, as normal 2-dimensional covariance matrices are usually visualized, in Fig. 5b. To visualize a GMM as a set of ellipses, the covariance matrix of each ellipse element is scaled by its associated weight. Reducing the size of each element's covariance matrix is only a visualization technique. It is emphasized that mathematically, the weight is the probability that a random state is generated by that particular mixture element. Multiplying the weight, α_i, with the covariance matrix, $\mathbf{P}_i$ has the effect of reducing the standard deviations by a factor of $\sqrt{\alpha_i}$ and changing the correlation, which results in a different probability distribution.

The resulting distributions from the function in Eq. 14 are seen in Fig. 6. There is a discrepancy between the performance of splitting along the different spectral directions. By visual inspection of the covariance ellipses of the individual GMM elements, better accuracy is seen for this case, by splitting along the major-axis of the covariance ellipse.

3.3 Choosing the Splitting Direction

A Taylor series approximation improves as the interval $(\mathbf{x} - \mathbf{x}_0)$ reduces. The uncertainty of the state is a measure of the spread, so applying a GMM split decreases the size of the covariance matrix along the split direction and thus, each element



(a) Split along minor axis (b) Weighted covariance ellipses (c) Split along major axis

Figure 5: Sample point cloud from the covariance ellipses of the multivariate GMM

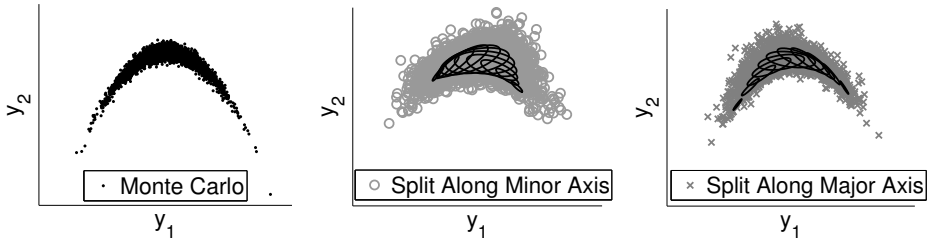


Figure 6: Point clouds of the resulting non-Gaussian distribution using MC and an 11 element GMM

covers a smaller domain $(\mathbf{x}-\mathbf{x}_0)$. However, the splits only decrease the domain along a single direction. The choice of which direction should consider the degree of nonlinearity each direction exhibits. However, the variance should also play a role in the direction choice, as it is a measure of the uncertainty along a splitting direction. Nonlinearity up to approximately the second order at the standard deviation locations can be quantified using the second-order divided difference form of Stirling's interpolation formula [Froberg (1969)]. Stirling's interpolation formula is used instead of Taylor series because the Taylor series approximation is less accurate further from the expansion location [Norgaard, Poulsen, and Ravn (2000)], i.e. the mean of the original Gaussian distribution. The second-order approximation of a univariate function $f(x)$ about $\bar{x}$ is:

$$f(x) \approx f(\bar{x}) + \frac{f(\bar{x}+h) - f(\bar{x}-h)}{2h}(x-\bar{x}) + \frac{f(\bar{x}+h) + f(\bar{x}-h) - 2f(\bar{x})}{2h^2}(x-\bar{x})^2 \tag{16}$$

where h is the step size used for the interpolation. The nonlinearity along any direction $\mathbf{a}$ of the multivariate function is quantified using the second term:

$$\boldsymbol{\phi} = \frac{f\left(\bar{\mathbf{x}} + \tilde{h}\sigma_{\|\hat{\mathbf{a}}}\hat{\mathbf{a}}\right) + f\left(\bar{\mathbf{x}} - \tilde{h}\sigma_{\|\hat{\mathbf{a}}}\hat{\mathbf{a}}\right) - 2f\left(\bar{\mathbf{x}}\right)}{2\tilde{h}^2} \tag{17}$$

where $\tilde{h} = \sqrt{3}$ is recommended so that the function evaluations are the sigma points for the DD2. The standard deviation of the cut along the arbitrary direction $\mathbf{a}$ passing through the mean is:

$$\sigma_{\|\hat{\mathbf{a}}} = \left\|\mathbf{S}^{-1}\hat{\mathbf{a}}\right\|_2^{-1} \tag{18}$$

where $\mathbf{S}$ is the square-root factor (Cholesky or Spectral decomposition) of the covariance matrix $\mathbf{P}$.

After evaluating $\boldsymbol{\phi}$ for all the desired directions, the maximum of some measure, such as the $p-$norm, is used to rank the directions in order of nonlinearity. The optimal direction of splitting is the direction which, considering the uncertainty in that direction, undergoes the most nonlinearity for a desired objective. Higher order divided differences of the function can also be used, at a higher computation cost, to include more information about the nonlinearity of the function and its sensitivity to the spectral directions. If the difference between two $\boldsymbol{\phi}_i$ is not large, both directions undergo similar amounts of nonlinearity up to the second order.

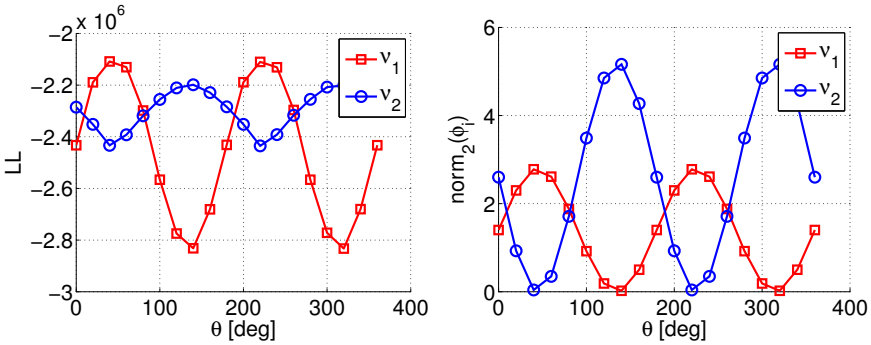
For the orbit propagation problem, the function of interest is a discrete function, which is the integration of the orbit dynamics for the desired time of flight. The different state vector elements have different units and therefore, normalized coordinates should be used such that the standard gravitational parameter of the central body is unity. The nonlinearity measure is a relative measure and allows the user to rank the splitting directions. Therefore, picking the number of splits and the number of directions to split along depends on the computational budget available.

The nonlinearity measure is illustrated on the test case from Section 3.2. The accuracy of the GMM approximated non-Gaussian distribution is quantified using the log-likelihood (LL) of the MC sample:

$$LL = \sum_{i=1}^{M} \log\left(\sum_{j=1}^{N} \alpha_j p_g\left(\mathbf{x}_i; \boldsymbol{\mu}_j, \mathbf{P}_j\right)\right) \tag{19}$$

where M is the total number of MC runs and $\mathbf{x}_i$ is the i^{th} MC sample point. The LL of the MC sample is a holistic measure of accuracy for the propagated PDF and has been used previously for demonstrating the accuracy of GMMs for propagation [Terejanu, Singla, Singh, and Scott (2008)]. Therefore, LL is used throughout

the work. The covariance matrix from Eq. 15 is first transformed into a new coordinate frame by using a two-dimensional rotation matrix with the angle θ. The intial state and uncertainty are then propagated through Eq. 14. The splitting directions are, however, the spectral directions of the original non-rotated covariance matrix. Fig. 7b shows the nonlinearity measure $|\boldsymbol{\phi}_i|_2$, as a function of the changing θ. Comparing Fig. 7a and Fig. 7b shows the importance of splitting along the spectral direction with the higher nonlinearity.



(a) Loglikelihood of distribution (b) Nonlinearity measure

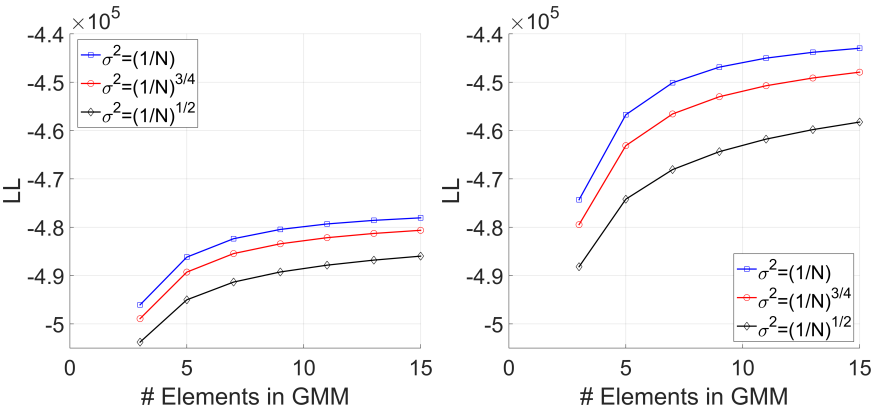
Figure 7: The loglikelihood of the MC simulation and the nonlinearity measure of the function with 5×10^5 sample points and 5 splits along $\boldsymbol{\nu}_1$ or $\boldsymbol{\nu}_2$ and increasing rotation angle θ

3.4 Sensitivity to Univariate Splitting Library

Three different splitting libraries are generated in Section 2. Since the three libraries produce different means and weights, the resulting accuracy after the propagation through the nonlinear function differs with the rule.

The test case from Eq. 14 in Section 3.2 is run with GMMs generated using the different univariate libraries. The accuracy with which the GMMs capture the non-Gaussian distribution is found in Fig. 8. The rule with $\sigma^2 = 1/N$ performs the best by consistently having a higher *LL* than the other splitting rules. A similar behavior is found for the test cases presented in Section 4. The rule with $\sigma^2 = 1/N$ gives each element more authority due to a more uniform weight distribution across all elements, and this leads to stronger performance for the highly nonlinear examples considered. The higher accuracy of the library with the smallest standard deviations per element is likely to be problem dependent, however, the rule with $\sigma^2 = 1/N$ is used exclusively in the following sections, and is recommended for future use.

Fig. 8 shows the expected trade-off between the accuracy and the number of required elements. Future work includes exploring other sigma rules not considered currently, which can be generated using the same optimization procedure outlined in Section 2.



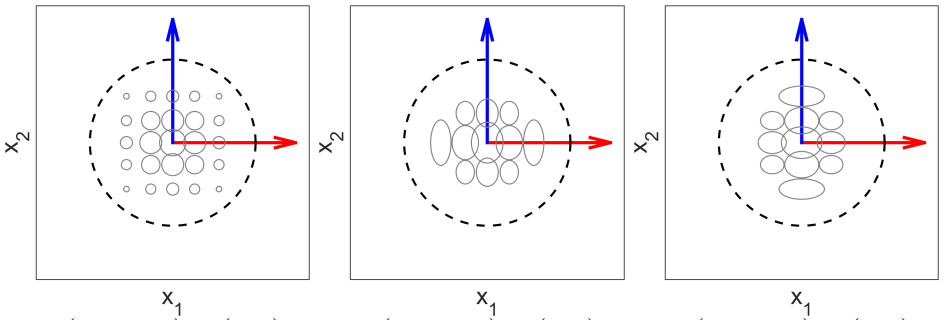
(a) Split along v_1 (b) Split along v_2

Figure 8: Performance of the various univariate splitting libraries for a nonlinear test case. Higher LL indicates a closer fit.

3.5 Multidirectional Gaussian Mixture Models

Splitting along only one direction cannot fully describe the resulting non-Gaussianity in the propagated distribution after a nonlinear transformation. Therefore, in cases where one direction is not sufficient, the natural extension is to apply the univariate split along multiple spectral directions. However, it is imperative to only apply the splits along directions that have been identified using the nonlinearity measure, and not along all degrees of freedom. The number of splits and the number of directions depends on the computational budget available. The resulting Multidirectional GMMs (MGMMs) are presented in this section.

Splitting along multiple directions is carried out in a recursive manner. After the initial multivariate Gaussian distribution is split along the first direction, each of the resulting multivariate mixture elements is split along the next specified direction. The elements are again all split as a full tensor product until all specified directions have been covered. For orthogonal directions, the order of the directions to be split along is not important and applying the univariate library only reduces the eigenvalue of the covariance matrix along the specified eigenvector direction.



(a) $(N_{\boldsymbol{\nu}_1}, N_{\boldsymbol{\nu}_2}) = (5,5)$, $\varepsilon = 0$, $N_{total} = 25$ (b) $(N_{\boldsymbol{\nu}_1}, N_{\boldsymbol{\nu}_2}) = (5,5)$, $\varepsilon = 0.05$, $N_{total} = 11$ (c) $(N_{\boldsymbol{\nu}_2}, N_{\boldsymbol{\nu}_1}) = (5,5)$, $\varepsilon = 0.05$, $N_{total} = 11$

Figure 9: Splitting a bivariate i.i.d. Gaussian into an MGMM with and without a minimum weight threshold and 5 elements along $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$. Note that the covariance matrices are multiplied by their weights for ease of visualization

The multidirectional multivariate splitting technique essentially forms a regular grid in probabilistic space. Since the initial conditions are stochastic, there is a probability associated with an initial condition. A nonlinear function increases the non-Gaussianity of the initial distribution based on the nonlinearity of the function and the size of the covariance matrix. By using the MGMM grid, the variance along the splitting directions is reduced. As illustrated in Fig. 9a, this grid is uniform since there is a symmetry about the mean of the initial multivariate Gaussian distribution.

The elements near the tails may have a very low weight when the number of elements is large along a certain direction, or due to the tensor product in multiple dimensions. Therefore, it may be beneficial to specify a minimum weight threshold ε such that the weights of all the elements in the MGMM are greater than ε. Since the computational burden of propagating an element through the nonlinear function is not dependent on the weight, increasing ε can reduce the total computation cost with only a slight degradation in accuracy. The multidirectional splitting with a threshold is illustrated for a bivariate case in Fig. 9b and Fig. 9c. The order of splitting is important when the threshold is specified because the weight budget decreases with each recursive split.

4 Numerical Test Cases

The MGMM splitting technique is now applied to two two-dimensional test problems, one six-dimensional orbit uncertainty propagation, and one ten-dimensional

problem. The two-dimensional problems are chosen in order to visualize the results. An initial bidirectional MGMM is expressed as (N_1, N_2), where N_1 and N_2 are the number of splits along the eigenvectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, respectively. For the bivariate cases, a GMM is simply an MGMM where either N_1 or N_2 are 1. The univariate library used to generate the MGMMs for all test cases in this section is Rule 1, $\sigma^2 = (1/N)$. The methodology and conclusions are applicable to higher dimensions as demonstrated by the orbit propagation and ten-dimensional problems.

The Gaussian distribution of the initial state is converted into an MGMM. Each element is then propagated through the nonlinear function. For the orbit uncertainty propagation case, the nonlinear function is the numerical integration of two-body dynamics over a given time period. The true final non-Gaussian distribution after the nonlinear transformation is assumed to be the same as the result of using MC runs. The log-likelihood of the MC samples generated by MGMMs from Eq. (19) is used as a holistic measure of accuracy. Each of the Gaussian elements is propagated through the nonlinear function using the DD2 sigma point method [Norgaard, Poulsen, and Ravn (2000)]. However, any other technique for propagating a Gaussian distribution, such as the UT or a Taylor series based method can be used. The weights of the GMM elements are not updated after propagating. Updating weights improves performance at the cost of decreasing ease of implementation when a first order Taylor series approximation is used for propagating the elements [Terejanu, Singla, Singh, and Scott (2008)]. However, no improvement is seen when higher order propagation methods such as the UT and DD2 are used [Horwood, Aragon, and Poore (2011)].

4.1 Conversion from Polar Coordinates to Cartesian Coordinates

The first test case is the conversion from Polar coordinates to Cartesian coordinates[Julier and Uhlmann (2004)]. The initial state, covariance of the uncertainty, and the eigenvalues and vectors of the covariance matrix are:

$$\begin{aligned} \begin{bmatrix} r \\ \theta \end{bmatrix} &= \begin{bmatrix} 70 \\ \pi/3 \end{bmatrix} \qquad \mathbf{P}_{\mathbf{x}} = \begin{bmatrix} 16 & 1.3 \\ 1.3 & (\pi/6)^2 \end{bmatrix} \\ \lambda_1 &= 0.1674 \qquad \boldsymbol{\nu}_1 = \begin{bmatrix} 0.0818 & -0.9966 \end{bmatrix}^T \\ \lambda_2 &= 16.1067 \qquad \boldsymbol{\nu}_2 = \begin{bmatrix} -0.9963 & -0.0818 \end{bmatrix}^T \end{aligned} \tag{20}$$

The initial conditions are chosen arbitrarily so that an MGMM is required. The resulting non-Gaussian distribution is illustrated in Fig. 10. When Eq. 17, is used along with the $2-$norm of $\boldsymbol{\phi}$, the values along the 2 spectral directions are: $|\boldsymbol{\phi}_1|_2 = 5.5823$, $|\boldsymbol{\phi}_2|_2 = 3.8793$. Since the nonlinearity is higher along $\boldsymbol{\nu}_1$ than along $\boldsymbol{\nu}_2$, a higher improvement is expected to be seen for splits along $\boldsymbol{\nu}_2$. However, since they

are still similar in value, splitting along both spectral directions is likely beneficial.

Distributions sampled from the propagated MGMMs are seen in Fig. 12. The likelihood of an MC run with respect to an MGMM is always smaller than 1 and therefore, the log-likelihood from Eq. 19 is a negative quantity. A new quantity is now defined as:

$$\Delta LL(n_1, n_2) = LL(n_{1max}, n_{2max}) - LL(n_1, n_2) \tag{21}$$

ΔLL is the difference between the LL of the most accurate GMM case $(39, 39)$, and the LL of an arbitrary case, and should always be positive. Figure 11 shows the ΔLL as a function of the number of splits per direction in the MGMM. As predicted by the nonlinearity metric, splitting along $\boldsymbol{\nu}_1$ is more effective than splitting along $\boldsymbol{\nu}_2$. However, the increase in performance due to the splitting along multiple spectral directions is clearly seen in Figure 11. Note that the performance of an MGMM (2,2) split, requiring 4 elements, is better than either the (1,39) or (39,1) GMMs that each require 39 elements.

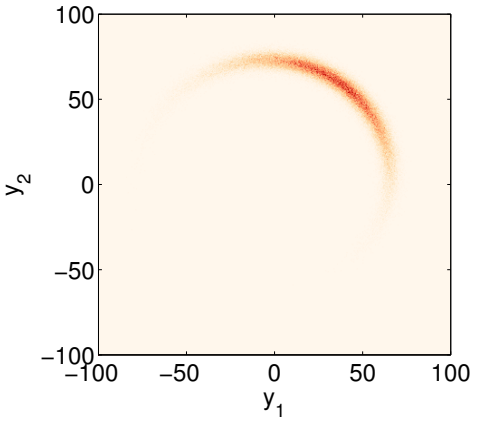


Figure 10: An MC run with 5×10^5 points for the conversion from Polar to Cartesian coordinates

4.2 Non-Linear ODE

A more complicated test function is now presented, where increasing the number of splits along only one direction does not suffice. The nonlinear function is the solution to the following differential equation:

$$\begin{aligned} \dot{x}_1 &= \cos(x_2)\sin(x_1) \\ \dot{x}_2 &= -\cos(x_1)\sin(x_2) \end{aligned} \tag{22}$$

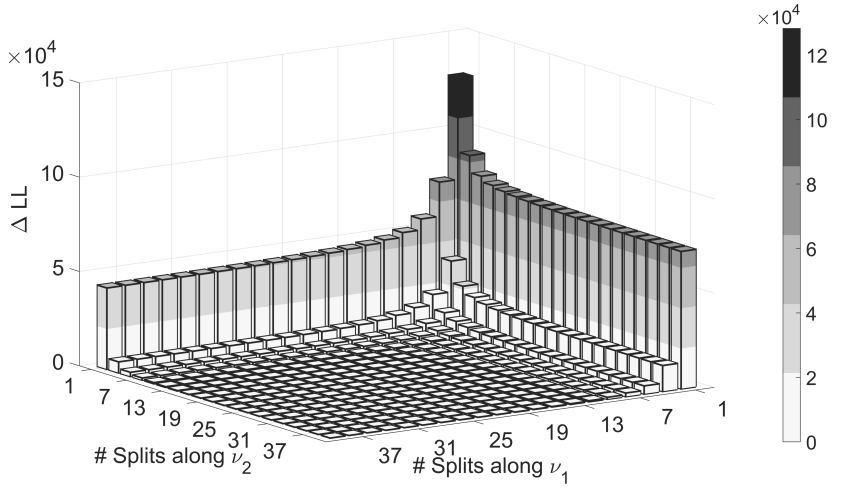


Figure 11: ΔLL of a 10^5 point MC distribution with respect to the MGMM with increasing number of elements and $LL(39,39) = -7.5789 \times 10^5$ for the conversion from Polar to Cartesian coordinates

Eq. 22 is numerically integrated from $t = 0$ to $t = 3$ and the resulting values of x_1 and x_2 are used as the final values after the nonlinear transformation. The integrated ODE solution is considered to be a discrete black box function. The initial state and covariance matrix are:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mathbf{P_x} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \tag{23}$$

The eigenvalues and eigenvectors of the covariance matrix are:

$$\begin{aligned} \lambda_1 &= 0.9 \quad & \boldsymbol{\nu}_1 &= \begin{bmatrix} -1 & 1 \end{bmatrix}^T \\ \lambda_2 &= 1.1 \quad & \boldsymbol{\nu}_2 &= \begin{bmatrix} 1 & 1 \end{bmatrix}^T \end{aligned} \tag{24}$$

The initial conditions are again chosen arbitrarily so that an MGMM is required. The result of an MC run with 10^4 sample points is shown in Fig. 13. The resulting distribution is highly non-Gaussian and bimodal, which favors the MGMM technique.

The nonlinearity metric using Eq. 17 results in $|\boldsymbol{\phi}_1|_2 = |\boldsymbol{\phi}_2|_2 = 0$. A $\mathbf{0}$-valued result using the second-order divided difference implies that the function is either linear, or it has higher-order nonlinearity at the mean of the initial Gaussian distribution. A conservative approach in this case is to split along both spectral directions.

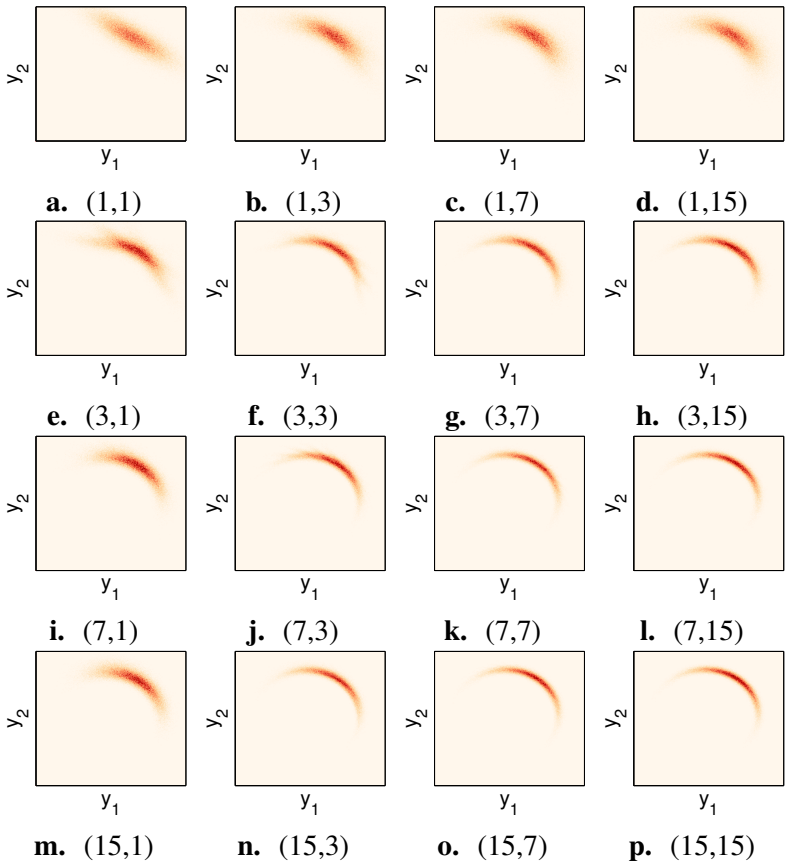


Figure 12: 5×10^5 points sampled from the resulting MGMMs with N_1 splits along $\boldsymbol{v}_1$ and N_2 splits along $\boldsymbol{v}_2$, (N_1, N_2) for the conversion from Polar to Cartesian coordinates

The samples from the propagated MGMMs are shown in Fig. 15. The ΔLL values for 10^5 MC sample points with respect to the MGMMs are seen in Fig. 14. The benefit of multiple splitting directions arises due to nonlinearity existing along both eigenvectors. Fig. 15 and Fig. 14 show that an MGMM with an equal number of splits in both directions performs better than a GMM with the total number of splits applied along only one spectral direction.

4.3 Orbit Uncertainty Propagation

The state uncertainty for a space object in GEO is propagated using two-body dynamics for a flight time of three days. The initial state and Gaussian uncertainty

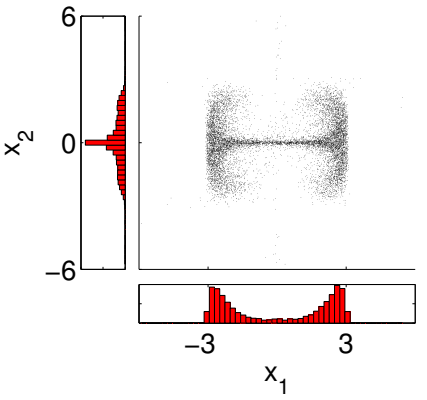


Figure 13: An MC run with 10^4 points for the bivariate ODE

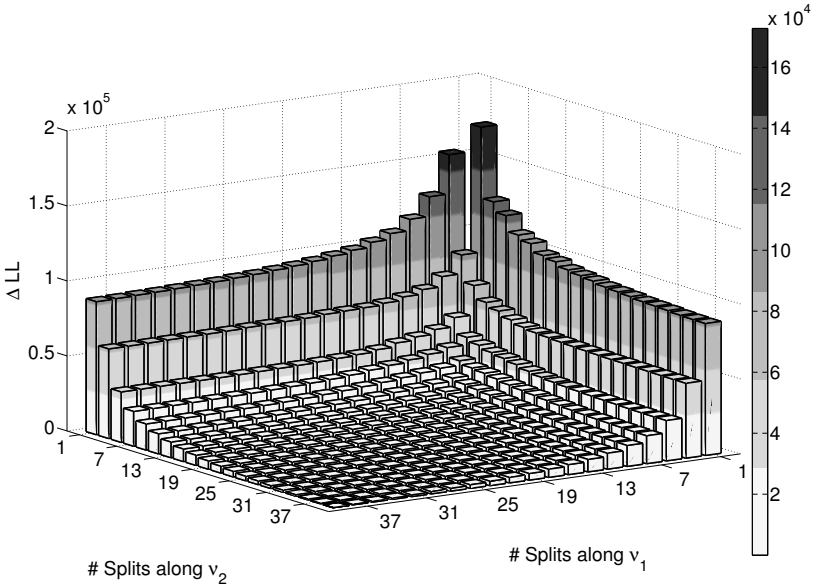


Figure 14: ΔLL of a 10^5 point MC distribution generated with respect to the MGMM with increasing number of elements and $LL(39,39) = -2.9264 \times 10^5$ for the bivariate ODE

are shown in Tab. 2. The orbit is assumed to be derived from optical observations, right ascension (RA) and declination (DEC), which in general create larger errors in the range direction for GEO objects. The nonlinear function considered here

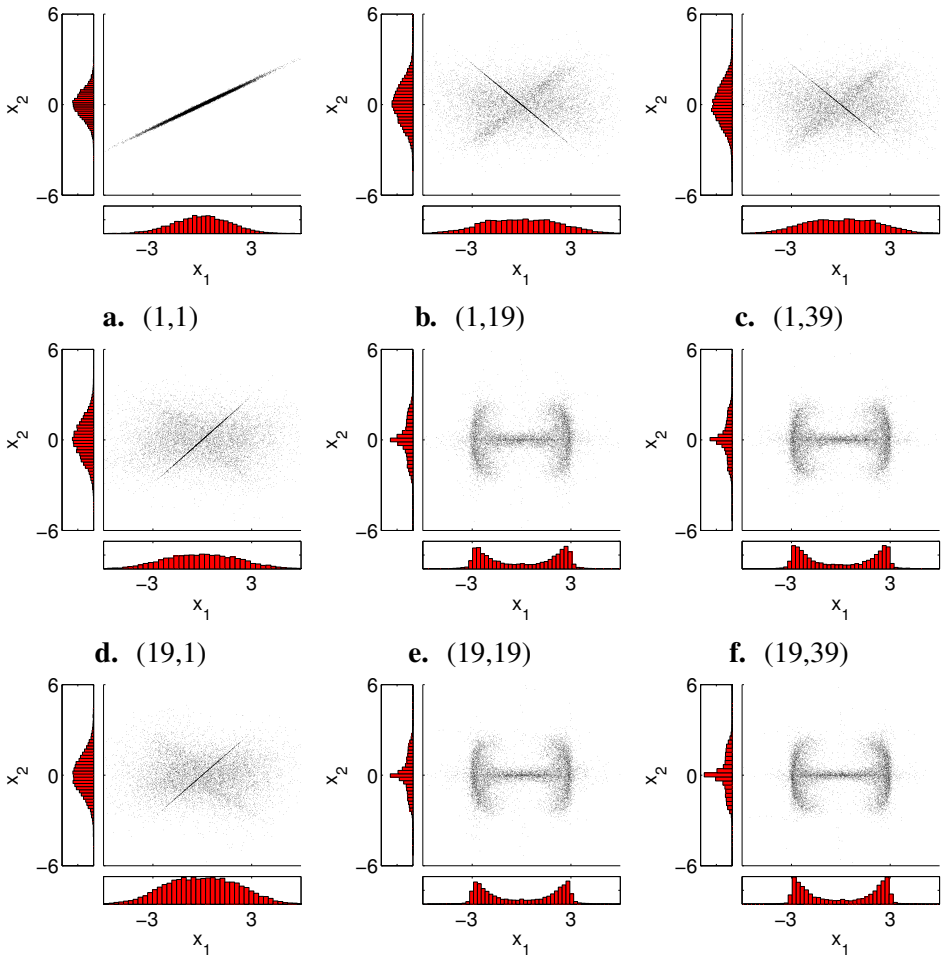


Figure 15: 10^4 points sampled from the resulting MGMMs with N_1 splits along $\boldsymbol{\nu}_1$ and N_2 splits along $\boldsymbol{\nu}_2$, (N_1, N_2) for the bivariate ODE

is the black box integration of the two-body orbit dynamics from epoch for the entire time of flight of three days, i.e. a discrete function. Therefore, the initial Gaussian distribution is first converted to an MGMM and all the elements are propagated for the entire time of flight. Cartesian coordinates are used to exacerbate the non-Gaussian behavior of the post propagation PDF.

The largest nonlinearity directions (expressed in normalized coordinates), found in Tab. 2, are along x and v_y, which are the radial and tangential directions, respective-

ly. The remaining directions have nonlinearity values that are at least three orders of magnitude lower and therefore, do not benefit from splitting. The number of splits along these two directions depends on the computational budget available.

Table 2: Initial state, Gussian uncorrelated uncertainty, and the nonlinearity measure in normalized coordinates for a space object in GEO

Variable	Mean	σ	$\|\boldsymbol{\phi}\|_2$
x [km]	42057.9	10.0	6.1×10^{-4}
y [km]	0	0.1	1.1×10^{-9}
z [km]	0	0.1	1.1×10^{-9}
v_x [km/s]	0	1.2×10^{-4}	2.8×10^{-7}
v_y [km/s]	3.0800809759824	0.6×10^{-4}	4.1×10^{-4}
v_z [km/s]	0	0.25×10^{-4}	1.2×10^{-8}

The ΔLL values for 10^5 MC sample points with respect to the MGMMs are seen in Fig. 16. As predicted by the nonlinearity measure, splitting along the x direction is more beneficial than splitting along the v_y direction. However, an MGMM along both x and v_y directions provides a more accurate result than a GMM along only one of the directions.

4.4 High Dimensional Problem

The MGMM splitting technique is now applied to the ten-dimensional Extended Freudenstein & Roth function [Andrei (2008)] to demonstrate the benefit of using MGMMs and the nonlinearity test from Eq. 17.

$$\begin{aligned} f(\mathbf{x}) = \sum_{i=1}^{5} &(-13 + x_{2i-1} + ((5 - x_{2i})x_{2i} - 2)x_{2i})^2 \\ &+ (-29 + x_{2i-1} + ((x_{2i} + 1)x_{2i} - 14)x_{2i})^2 \end{aligned} \tag{25}$$

The ten-dimensional state $\mathbf{x}$ has an initial mean and a diagonal covariance matrix for the uncertainty with the exact values found in Tab. 3. The initial means and variances are generated using a random number generator and then rounded to two decimal places. The absolute value of the nonlinearity measure, Eq. 17, along the various spectral directions is also found, and computed values are given in the last row of Tab. 3.

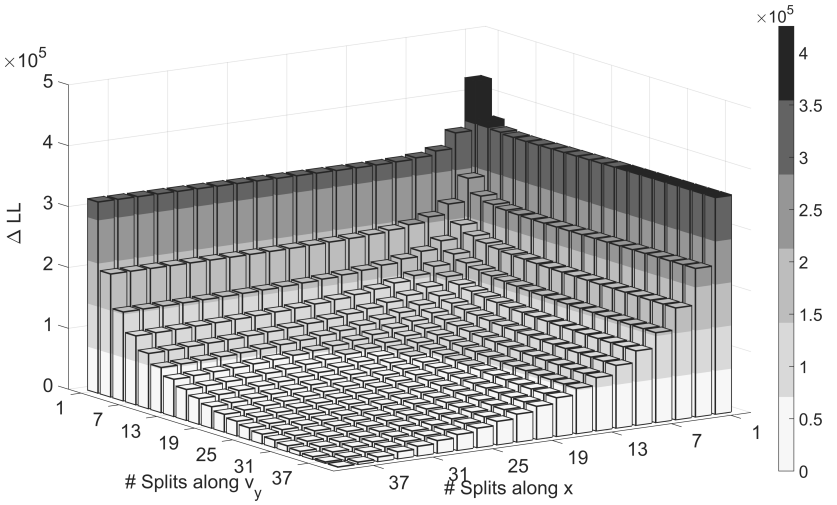
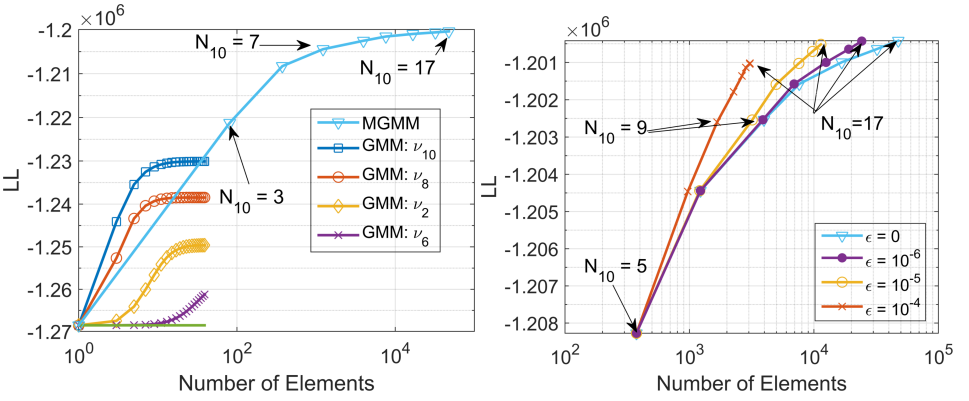


Figure 16: ΔLL of a 10^5 point MC distribution generated with respect to the MG-MM with increasing number of elements and $LL(39,39) = 4.8209 \times 10^6$ for the space object in GEO

Table 3: Initial mean, variance, and nonlinearity measure for the Extended Freudenstein & Roth function

Variable	Mean	σ^2	ϕ
x_1	6.19	1.67	3.34
x_2	3.76	1.81	4.80×10^3
x_3	1.94	1.27	2.54
x_4	0.21	1.01	12.42
x_5	1.53	1.67	3.34
x_6	3.36	1.08	904.84
x_7	6.67	2.40	4.80
x_8	4.93	1.67	1.68×10^4
x_9	2.33	1.35	2.70
x_{10}	5.73	1.06	1.99×10^4

Spectral decomposition of the covariance matrix results in the eigenvectors being the unit vectors of the individual univariate variables and $\lambda_i = \sigma_i^2$. The last row of Tab. 3 shows the largest nonlinearity directions are 10, 8, 2, and 6, in descending



(a) GMM and MGMM (b) MGMM with weight threshold

Figure 17: *LL* of a 10^5 point MC distribution with respect to GMMs and MGMMs for the 10-dimensional case

order. The nonlinearity along the other directions is considered to be insignificant. The values of the nonlinearity measure along a given direction correspond to the improvement in performance, shown in Fig. 17, gained by generating splits along that direction. Since nonlinearity is present along multiple spectral directions, an MGMM further improves accuracy compared to using a GMM. The nonlinearity measure is a useful ranking method to find the number of splits to implement along the spectral directions. The number of splits along direction 10 N_{10} is used to calibrate the number of splits along the other directions 8, 2, and 6 because the highest nonlinearity is along $i = 10$. The number of splits along the remaining spectral directions are computed using the following equation:

$$N_i = \text{odd}\left(\frac{\log(\phi_i)}{\log(\phi_{10})} N_{10}\right) \qquad i = 8,2,6 \tag{26}$$

where odd computes the closest odd integer to a number:

$$\text{odd}(x) = 2 \times \text{ceil}(x/2) - 1 \tag{27}$$

The performance of the GMMs along directions 10, 8, 2, and 6 are seen in Fig. 17, along with the performance of the 4-dimensional MGMM. The nonlinearity measure along a direction performs well as an indicator of increase in performance gained by splitting along that direction. Direction 10 has the highest value for the nonlinearity measure and therefore, splitting along this direction results in the largest increase in performance. Splitting only along any one direction has a plateau

in performance after a certain number of splits is reached. Using an MGMM, however, results in a higher value of the log-likelihood compared to a GMM along any of the directions.

The benefit of using a minimum threshold for the MGMM weights is seen in Fig. 17. When a threshold of $\varepsilon = 10^{-4}$ is used, approximately 6 times fewer elements are required to achieve the same accuracy ($LL \approx -1.2 \times 10^6$) for the MGMM without a threshold. For the highest accuracy presented here with $N_{10} = 17$, thresholds of $\varepsilon = 10^{-6}$ and $\varepsilon = 10^{-5}$ require approximately 2 times and 4 times fewer elements, respectively.

5 Conclusion

Nonlinear functions acting on an initial multivariate state with a Gaussian uncertainty are common scenarios. In the orbit uncertainty problem a Gaussian state uncertainty is the result of most Orbit Determination (OD) and state estimation algorithms for space objects. Accurately capturing the resulting non-Gaussian distribution after the nonlinear transformation without resorting to computationally expensive techniques such as Monte Carlo (MC) simulations remains an attractive ongoing area of research. The primary contributions of this paper are the extension of the univariate GMM library to 39 elements [2], a formal extension of the GMM concept to multi-directions, and a method to choose any split direction and introduction of a nonlinearity metric to rank the importance of splitting along a certain direction.

Univariate splitting libraries of up to 39 elements that are carefully fit in advance are generated by minimizing the L_2 distance with respect to a standard normal distribution. The library has an odd number of elements and standard deviation of all the elements is the same, which depends on the number of elements required. Instead of applying the univariate library along only a single direction, the univariate library is applied recursively along multiple directions forming a regular grid over multiple dimensions. These Multidimensional GMMs (MGMMs) approximate the propagated multivariate non-Gaussian distribution more accurately than if the split is made along only one direction. A second-order divided difference is used to measure nonlinearity in selected directions. Thus, the directions that benefit most from splitting are identified and can be exploited. This nonlinearity merit provides a ratio to relate the number of splits along each direction. Therefore, a practitioner can implement the MGMM technique with the selection of just one fidelity tuning parameter, as shown for a ten dimensional case. A user enforces a number of splits along a single direction, then the number of splits in the other directions

[2] `http://russell.ae.utexas.edu/code/GMMsplittingLibrary.txt/`

can be chosen according to their relative nonlinearity. A threshold for the weight further reduces the computational load by ensuring that the MGMM only contains elements above a certain weight.

Combining the univariate library with up to 39 elements with the multidirectional splitting results in a deterministic choice of initial states for propagation through the nonlinear function. The number of necessary function evaluations is ultimately a trade-off between accuracy and computational cost. The MGMM allows users to access all regions of this spectrum, filling in the gap that typically exists between Monte Carlos and simple linearized covariance analyses. The MGMM representation of the state uncertainty of space objects has the potential to improve the accuracy of commonly used estimation and conjunction assessment techniques with minor changes in the implementation.

References

Abramov, R. F. (2007): An improved algorithm for the multidimensional moment-constrained maximum entropy problem. *Journal of Computational Physics*, vol. 226, no. 1, pp. 621–644.

Adurthi, N.; Singla, P. (2015): Conjugate unscented transformation-based approach for accurate conjunction analysis. *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 9, pp. 1642–1658.

Adurthi, N.; Singla, P.; Singh, T. (2012): The conjugate unscented transform - an approach to evaluate multi-dimensional expectation integrals. In *Proceedings of the American Control Conference (ACC)*, pp. 5556–5561. IEEE.

Alspach, D. L.; Sorenson, H. W. (1972): Nonlinear bayesian estimation using gaussian sum approximations. *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448.

Andrei, N. (2008): An unconstrained optimization test functions collection. *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147–161.

Aristoff, J. M.; Horwood, J. T.; Singh, T.; Poore, A. B. (2014): Nonlinear uncertainty propagation in orbital elements and transformation to cartesian space without loss of realism. In *AIAA/AAS Astrodynamics Specialist Conference, San Diego, CA, Aug 4 - Aug 7*.

Baspinar, U.; Varol, H. S.; Senyurek, V. Y. (2013): Performance comparison of artificial neural network and gaussian mixture model in classifying hand motions by using semg signals. *Biocybernetics and Biomedical Engineering*, vol. 33, no. 1, pp. 33–45.

Butler, T.; Dawson, C.; Wildey, T. (2013): Propagation of uncertainties using improved surrogate models. *SIAM/ASA Journal on Uncertainty Quantification*, vol. 1, no. 1, pp. 164–191.

Byrd, R. H.; Schnabel, R. B.; Schultz, G. A. (1987): A trust region algorithm for nonlineraly constrained optimization. *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1152 – 1170.

Calderhead, B. (2014): A general construction for parallelizing metropolis-hastings algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 49, pp. 17408–17413.

Chan, F. K. (2008): *Spacecraft Collision Probability*. The Aerospace Press, American Institute of Aeronautics and Astronautics, Inc.

Conn, A. R.; Gould, N. I. M.; Toint, P. L. (1987): *Trust region methods*. No. 1 in MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics.

Cowles, M. K.; Carlin, B. P. (1996): Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904.

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

DeMars, K. J.; Jah, M. K. (2014): Collision probability with gaussian mixture orbit uncertainty. *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 3, pp. 979–985.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. (1977): Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–38.

Endres, D. M.; Schindelin, J. E. (2003): A new metric for probability distributions. *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1858 – 1860.

Feller, W. (1945): The fundamental limit theorems in probability. *Bulletin of the American Mathematical Society*, vol. 51, pp. 800–832.

Froberg, C. E. (1969): *Introduction to Numerical Analysis*. Addison Wesley Publishing Company, Boston, Massachusetts.

Fujimoto, K.; Scheeres, D. J. (2015): Tractable expressions for nononlinear propagated uncertainties. *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 6, pp. 1146–1151.

Fuller, A. T. (1969): Analysis of nonlinear stochastic systems by means of the fokker-planck equation. *International Journal of Control*, vol. 9, no. 6, pp. 603–655.

Gelb, A. (1974): *Applied Optimal Estimation*. The MIT Press, Cambridge, Massachusetts, 1 edition.

Haario, H.; Laine, M.; Mira, A.; Saksman, E. (2006): Dram: Efficient adaptive mcmc. *Statistics and Computing*, vol. 16, no. 4, pp. 339–354.

Hanebeck, U. D.; Briechle, K.; Rauh, A. (2003): Progressive bayes: A new framework for nonlinear state estimation. *Proc. SPIE*, vol. 5099, pp. 256–267.

Hershey, J. R.; Olsen, P. A. (2007): Approximating the kullback leibler divergence between gaussian mixture models. In *IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI*.

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

Huber, M. F.; Bailey, T.; Durrant-Whyte, H.; Hanebeck, U. D. (2008): On entropy approximation for gaussian mixture random vectors. *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pp. 181–188.

Huber, M. F.; Hanebeck, U. D. (2008): Progressive gaussian mixture reduction. In *11th Conference on Information Fusion*, pp. 1–8.

Jones, B. A.; Doostan, A.; Born, G. H. (2013): Nonlinear propagation of orbit uncertainty using non-intrusive polynomial chaos. *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 2, pp. 430–444.

Jones, B. A.; Parrish, N.; Doostan, A. (2015): Post-maneuver collision probability estimation using sparse polynomial chaos expansions. *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 8, pp. 1425–1437.

Julier, S.; Uhlmann, J. K. (2004): Unscented filtering and nonlinear estimation. In *Proceedings of the IEEE*, volume 92, pp. 401–422.

Kullback, S.; Leibler, R. A. (1951): On information and sufficiency. *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86.

Kumar, M.; Chakravorty, S. (2012): Nonlinear filter based on the fokker-planck equation. *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 1, pp. 68–79.

Kumar, M.; Chakravorty, S.; Singla, P.; Junkins, J. L. (2009): The partition of unitiy finite element approach with hp-refinement for the stationary fokker-planck equation. *Journal of Sound and Vibration*, vol. 327, no. 1-2, pp. 144–162.

Li, W.; Prasad, S.; Fowler, J. E. (2014): Hyperspectral image classification using gaussian mixture models and markov random fields. *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 1, pp. 153–157.

Li, Y.; Sundaragan, N.; Saratchandran, P. (2000): Analysis of minimal radial basis function network algorithm for real-time identification of nonlinear dynamic systems. *IEEE Proceedings of Control Theory and Applications*, vol. 147, no. 4, pp. 476–484.

Moré, J. J.; Sorensen, D. C. (1983): Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553–572.

Norgaard, M.; Poulsen, N. K.; Ravn, O. (2000): New developmeents in state estimation for nonlinear systems. *Automatica*, vol. 36, no. 11, pp. 1627–1638.

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

Siegmund, D. (1976): Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics*, vol. 4, no. 4, pp. 673–684.

Stroud, A. H.; Secrest, D. (1966): *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood Cliffs, New Jersey.

Sun, Y.; Kumar, M. (2014): 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.

Tapley, B. D.; Schutz, B. E.; Born, G. H. (2004): *Statistical Orbit Determination*. Elsevier Academic Press, 1 edition.

Terejanu, G. (2011): An adaptive split-merge scheme for uncertainty propagation using gaussian mixture models. In *49th AIAA Aerspace Sciences Meeting*.

Terejanu, G.; Singla, P.; Singh, T.; Scott, P. D. (2008): Uncertainty propagation for nonlinear dynamic systems using gaussian mixture models. *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 6, pp. 1623 – 1633.

Terejanu, G.; Singla, P.; Singh, T.; Scott, P. D. (2011): Adaptive gaussian sum filter for nonlinear bayesian estimation. *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156.

Vishwajeet, K.; Singla, P. (2013): Sparse approximation based gaussian mixture model approach for uncertainty propagation for nonlinear systems. In *American Control Conference*, pp. 1213 – 1218.

Vishwajeet, K.; Singla, P. (2014): Adaptive splitting technique for gaussian mixture models to solve kolmogorov equation. In *American Control Conference*, pp. 5186–5191.

Vishwajeet, K.; Singla, P.; Jah, M. (2014): Nonlinear uncertainty propagation for perturbed two-body orbits. *Journal of Guidance, Control, and Dynamics*, vol. Available Online.

Vittaldev, V.; Russell, R. P. (2013): Collision probability for resident space objects using gaussian mixture models, aas 13-351. In *23rd AAS/AIAA Spaceflight Mechanics Meeting, Kauai, Hawaii*.

Vittaldev, V.; Russell, R. P. (2016): Space object collision probability using multidirectional gaussian mixture models. In *Journal of Guidance, Control, and Dynamics, to appear*.

Wan, X.; Karniadakis, G. (2006): Beyond wiener-askey expansions: Handling arbitrary pdfs. *Journal of Scientific Computing*, vol. 27, no. 1-3, pp. 455–464.

Wiener, N. (1938): The homogeneous chaos. *Amer. J. Math.*, vol. 60, no. 4, pp. 897–936.