A Novel Gaussian Mixture Approximation for Nonlinear Estimation

Renato Zanetti and Kirsten Tuggle
Dept. of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin
Austin, Texas, 78712

***Abstract*—A novel adaptive nonlinear estimator is presented to accurately incorporate nonlinear/non-Gaussian measurement in a Bayesian framework. The underlying algorithm relies on a Gaussian Mixture Model (GMM) to approximate the probability density function (pdf) of the state conditioned on all current and past measurement. Automatic mixture components refining is performed to ensure that the posterior GMM approximation of the pdf accurately represents the true distribution.**

I. Introduction

Bayesian estimation is concerned with optimizing some performance index of the posterior probability density function (pdf) of an unknown random vector $\mathbf{x}$. The posterior distribution is the distribution of $\mathbf{x}$ conditioned on the value of all current and past available measurements $\mathbf{y}$. A widely used Bayesian estimator is the minimum mean square error estimator, where the estimate is given by the conditional mean of $\mathbf{x}$ given $\mathbf{y}$. For linear systems with Gaussian distributions, the posterior remains Gaussian with conditional mean and covariance matrix obtained through the Kalman filter equations. An analytical solution for non-Gaussian distributions is usually impossible to obtain, but Alspach and Sorenson [1] show that for linear systems where the noise is not Gaussian but exactly represented as a sum of Gaussians, a closed form representation of the conditional pdf is possible. Distributions are almost never exactly represented as sum of Gaussians and systems of interest are almost never linear; however every pdf can be approximated arbitrary well as Gaussian sum since the Dirac delta function is equivalent to a Gaussian distribution with zero variance; hence an arbitrary pdf $p_{\mathbf{x}}(\mathbf{x})$ over the support $\mathcal{S}$ can be written as

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}) &= \int_{\mathcal{S}} p_{\mathbf{x}}(\boldsymbol{\mu})\delta(\mathbf{x}-\boldsymbol{\mu})d\boldsymbol{\mu} \\ &= \lim_{M\to\infty}\sum_{i=1}^{M} p_{\mathbf{x}}(\mathbf{x})\ n(\mathbf{x};\boldsymbol{\mu}_i,\mathbf{P}/M) \\ &\simeq \sum_{i=1}^{M} w_i\ n(\mathbf{x};\boldsymbol{\mu}_i,\mathbf{P}_i) \end{aligned} \tag{1}$$

where $\mathbf{P}/M$ goes to zero as $M \to \infty$ and $n(\mathbf{x};\boldsymbol{\mu},\mathbf{P})$ represents a Gaussian distribution with argument $\mathbf{x}$, mean $\boldsymbol{\mu}$, and covariance matrix $\mathbf{P}$. The actual covariance matrices $\mathbf{P}_i$ of the GMM components are chosen to best approximate $p_{\mathbf{x}}(\mathbf{x})$. Expressing the pdf as the sum of many components with small covariance matrices has two benefits; first it allows us to better approximate it as in Eq. (1), secondly it divides the overall space where $\mathbf{x}$ is likely to be into many small regions. In a situation when these regions are small enough such that in each of them the measurement and dynamic functions are effectively linear, then the original algorithm of Alspach and Sorenson [1] proves to be an excellent mean to perform nonlinear filtering.

In practical situations, we often do not initially know where the dynamics and the measurements will take the conditional distribution. Therefore, it is often difficult or impossible to approximate the initial distribution of $\mathbf{x}$ as GMM such that the conditional distribution after a long time and many measurements is adequately represented by the components originally selected to represent the distribution at the initial time. Reference [2] first introduces a method to adaptably split GMM components in order to better represent the pdf during the propagation phase. In [2] the refinement is obtained by splitting components along the direction of maximum uncertainty (i.e. along the eigenvector corresponding to the largest eigenvalue of the component's covariance matrix). Other works that adopt the same philosophy are Refs. [3], [4], and [5]. Ref. [6] determines the splitting direction based on the second moment of a set of residual-weighted sigma points. Ref. [7] and our prior work [8] use directional splitting based on the second order contributions to a Taylor series centered at the estimate. Another approach is Ref. [9] that evaluates an integral cost over the eigendirections. In [10], the need to split is based on the uncertainty and the direction is chosen as the one that maximizes nonlinearity of the model dynamics. Ref. [11] uses a second-order divided difference term evaluated at a certain distance from the mean in order to quantify nonlinearity. This nonlinearity index 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 nonlinearity out of the elements of the finite set. In this work the direction to split is obtained integrating the full nonlinearity but a new scheme is devised to perform such an integration only once to obtain the maximum direction of nonlinearity out of all possible directions. As mentioned before, existing works perform calculations many times along several predetermined directions and choose the maximum nonlinearity out of this finite set.

Other approaches to solve a similar problem exist that rely

on studies of the validity of the Gaussian approximation [12]. Rather than looking at the direction of maximum non-linearity, a split can occur in the direction of maximum non-Gaussianity. Ref. [13] detects non-Gaussianity by checking if the absolute value of the third central moment is above a predetermined threshold, when this condition occurs, splitting occurs the splitting direction is chosen as one of the elements of the state vector, i.e. along one of the elements of the canonical base. The idea is that a Gaussian distribution undergoing a nonlinear transformation becomes non-Gaussian, and the non-Gaussian posterior is best represented by refining components in the direction of maximum non-Gaussianity. Another approach is to resample the entire distribution (Refs. [14] and [15]) i.e. the authors do not split individual components in one-dimensional directions, rather they resample the entire space subject to a matrix inequality constraint that insures the covariance of each of the components stays below a desired tunable value.

II. Problem Formulation

Assume a random vector $\mathbf{x}$ is distributed with a known Gaussian prior $\mathbf{x} \sim n(\mathbf{x}; \hat{\mathbf{x}}^-, P^-)$, assume also that a non-linear measurement $\mathbf{y} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta}$ is available where the additive noise $\boldsymbol{\eta}$ is zero-man and Gaussian with covariance matrix R. Using the properties of the Dirac delta function, the prior can be expressed as

$$n(\mathbf{x}; \hat{\mathbf{x}}^-, P^-) = \int ... \int n(\boldsymbol{\xi}; \hat{\mathbf{x}}^-, P^-) \delta(x_1 - \xi_1) ... \delta(x_n - \xi_n) d\xi_1 ... d\xi_n \tag{2}$$

which, with some abuse of notation, we rewrite as

$$n(\mathbf{x}; \hat{\mathbf{x}}^-, P^-) = \int n(\boldsymbol{\xi}; \hat{\mathbf{x}}^-, P^-) \delta(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3}$$

The posterior distribution $p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})$ can be calculated exactly as [1]

$$\begin{aligned} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) &= \kappa(\mathbf{y}) \int n(\mathbf{y} - \mathbf{h}(\boldsymbol{\xi}); \mathbf{0}, R) n(\boldsymbol{\xi}; \hat{\mathbf{x}}^-, P^-) \delta(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int w(\boldsymbol{\xi}, \mathbf{y}) \; n(\boldsymbol{\xi}; \hat{\mathbf{x}}^-, P^-) \; \delta(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \end{aligned} \tag{4}$$

where $\kappa(\mathbf{y})$ is a normalizing constant and

$$w(\boldsymbol{\xi}, \mathbf{y}) = \kappa(\mathbf{y}) \; n(\mathbf{y} - \mathbf{h}(\boldsymbol{\xi}); \mathbf{0}, \; R)$$

We have effectively expressed the prior as an infinite summation of infinitesimal Gaussians. Since they are infinitesimal, the measurement linearization holds exactly, and therefore we have an exact formula for the posterior distribution. The goal of this paper is to approximate this exact posterior distribution with a finite number of finite Gaussian components:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) \simeq \sum_{i=1}^{M} w_i^+(\mathbf{y}) \; n(\mathbf{x}; \hat{\mathbf{x}}_i^+, P_i^+) \tag{5}$$

where the subscripts $+$ refer to posterior quantities.

The strategy employed is to:

1) Detect when the spread of the prior is wide enough such that the nonlinearities of the measurement function are not negligible and regular extended Kalman filter update does not provide a satisfactory approximation of the posterior
2) Establish the direction where the nonlinearities are most pronounced given the spread of the prior in that direction
3) Approximate the prior as a sum of Gaussians with means placed along the direction of greatest nonlinearities
4) Utilize the Gaussian Mixture Model algorithm to obtain the posterior

This paper presents new methodologies to solve the first two points above, and utilizes existing techniques for the last two. For item 3 we use the library developed in Ref. [16]. For item 4 we use extended Kalman filter (EKF) based Gaussian sum filters. In the Gaussian, linear case the posterior distribution is exactly calculated with the Kalman filter equation. In the Gaussian, nonlinear measurement case, the EKF approximates the posterior when the prior covariance matrix is "small" enough such that the system is almost linear. Other linear estimators for nonlinear systems such as the unscented Kalman filter (UKF), the Quadrature Filter, the Gaussian Second Order Filter (GSOF), etc., are approximations of the Linear Minimum Mean Square Estimator, they are not approximations of the posterior density. They become approximations of the posterior density when the prior and posterior are almost Gaussian and the measurement is almost linear, e.g. when the prior covariance is "small". One could use one of these non-EKF algorithms in each of the Gaussian components, these algorithms are generally more robust to nonlinearities than the EKF and hence they would probably require less components. However, the goal of this paper is to approximate the true posterior; in order to do so we must calculate the posterior of each of the components, not the LMMSE estimate, but the actual conditional pdf. Given a Gaussian prior, a linear update will provide a good approximation of the posterior when the measurement is almost linear, e.g. when the covariance is small enough, which are the conditions under which the EKF performs well. In conditions under which the EKF would fail and other algorithms such as the UKF or GSOF would not, we know that the latter would produce a good approximation of the LMMSE estimate, not of the posterior distribution. Therefore, in this work there is no advantage in using any linear update scheme for the individual components other than the EKF, which we shall therefore use.

III. Proposed Solution

In the prior section we described the steps that form the proposed solution to the problem. We detailed the reasons why for Step 4 above we use an EKF-based GMM filter. For step 3 we utilize the 3-component split library proposed in [16] and detailed in Table I. Finally, steps 2 and 1 of the proposed algorithms are novel contributions of this work and are presented in the next two subsections.

A. Splitting Direction

Given a nonlinear function $\mathbf{h}(\mathbf{x}) \in \Re^m$ of a vector $\mathbf{x} \in \Re^n$ and its Jacobian $H(\mathbf{x}) \in \Re^{m\times n}$, the goal is to find the direction $\mathbf{u} \in \Re^n$ that most diverges from the linear approximation of $\mathbf{h}(\mathbf{x})$ centered at $\mathbf{x} = \mathbf{m}$ at distance of N standard deviations (σ's) from $\mathbf{m}$ ($N \in \Re^+$). That is, we want to find the maximum of

$$\max_{\mathbf{u}} \frac{1}{2}\big\|\mathbf{h}(N\mathbf{u}+\mathbf{m}) - \mathbf{h}(\mathbf{m}) - H(\mathbf{m})\ N\mathbf{u}\big\|^2 \tag{6}$$

subject to the contraint that $\mathbf{u}$ must lie on the unit-standard deviation ellipse described by covariance matrix P; that is $\mathbf{u}$ must satisfy

$$\mathbf{u}^{\mathrm{T}} P^{-1}\mathbf{u} = 1 \tag{7}$$

Hence, the augmented performance index is given by

$$\mathcal{J}(\mathbf{u}, N) = \max_{\mathbf{u}} \Big[\ \frac{1}{2}\big\|\mathbf{h}(N\mathbf{u}+\mathbf{m}) - \mathbf{h}(\mathbf{m}) - H(\mathbf{m})\ N\ \mathbf{u}\big\|^2 + \lambda\big(\mathbf{u}^{\mathrm{T}} P^{-1}\mathbf{u} - 1\big)\Big] \tag{8}$$

The optimal solution is obtained starting from the known optimal value of $\mathbf{u}$ for $N = 0$ and integrating an ordinary differential equation of $\mathbf{u}$ with N as the independent variable.

#	Weight	Mean	Std Dev
0	0.676659600665207	0	0.784394767127563
1	0.161670199667396	1.09080001170817	0.784394767127563
2	0.161670199667396	-1.09080001170817	0.784394767127563

TABLE I
UNIVARIATE STANDARD NORMAL SPLIT LIBRARY

Define

$$\mathbf{v}(\mathbf{u}, N) = \mathbf{h}(N\mathbf{u}+\mathbf{m}) - \mathbf{h}(\mathbf{m}) - H(\mathbf{m})\ N\mathbf{u} \tag{9}$$

the first order condition is given by

$$\mathcal{J}'(\mathbf{u}, N) = \frac{d}{d\mathbf{u}}\mathcal{J}(\mathbf{u}, N) = \mathbf{v}(\mathbf{u}, N)^{\mathrm{T}} A(\mathbf{u}, N) + \lambda\ \mathbf{u}^{\mathrm{T}} P^{-1} = 0 \tag{10}$$

where

$$A(\mathbf{u}, N) = \frac{d\mathbf{v}(\mathbf{u}, N)}{d\mathbf{u}} = [H(N\mathbf{u}+\mathbf{m}) - H(\mathbf{m})]\ N \tag{11}$$

Post-multiplying Eq. (10) by $\mathbf{u}$ we obtain

$$\mathbf{v}(\mathbf{u}, N)^{\mathrm{T}} A(\mathbf{u}, N)\ \mathbf{u} + \lambda\ \mathbf{u}^{\mathrm{T}} P^{-1}\mathbf{u} = 0 \tag{12}$$

substituting the constrain, we obtain

$$\lambda = -\mathbf{v}(\mathbf{u}, N)^{\mathrm{T}} A(\mathbf{u}, N)\ \mathbf{u} \tag{13}$$

and the first order condition reduces to

$$\mathcal{J}'(\mathbf{u}, N) = \mathbf{w}(\mathbf{u}, N)^{\mathrm{T}}\left(I - \mathbf{u}\mathbf{u}^{\mathrm{T}} P^{-1}\right) = 0 \tag{14}$$

where

$$\mathbf{w}(\mathbf{u}, N)^{\mathrm{T}} = \mathbf{v}(\mathbf{u}, N)^{\mathrm{T}} A(\mathbf{u}, N) \tag{15}$$

We can now take variations of this equation in the form

$$d\mathcal{J}'(\mathbf{u}, N) = \frac{d\mathcal{J}'(\mathbf{u}, N)}{d\mathbf{u}} d\mathbf{u} + \frac{d\mathcal{J}'(\mathbf{u}, N)}{dN} dN = 0 \tag{16}$$

where

$$\frac{d\mathcal{J}'(\mathbf{u}, N)}{d\mathbf{u}} = \left(I - P^{-1}\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)\frac{d\mathbf{w}(\mathbf{u}, N)}{d\mathbf{u}} - (\mathbf{w}(\mathbf{u}, N)^{\mathrm{T}}\mathbf{u})\ P^{-1}\ - P^{-1}\ \mathbf{u}\ \mathbf{w}(\mathbf{u}, N)^{\mathrm{T}} \tag{17}$$

$$\frac{d\mathbf{w}(\mathbf{u}, N)}{d\mathbf{u}} = A(\mathbf{u}, N)^{\mathrm{T}}\ \frac{d\mathbf{v}(\mathbf{u}, N)}{d\mathbf{u}} + N^2 \sum_{i=1}^{m} D_i(N\mathbf{u}+\mathbf{m})\ \mathbf{v}(\mathbf{u}, N) \tag{18}$$

$$\frac{d\mathbf{v}(\mathbf{u}, N)}{d\mathbf{u}} = A(\mathbf{u}, N) \tag{19}$$

$$D_i(\mathbf{x}) = \frac{d^2 h_i(\mathbf{x})}{d\mathbf{x}\ d\mathbf{x}^{\mathrm{T}}} \tag{20}$$

where h_i is the i-th component of vector $\mathbf{h}$ and D_i is the Hessian of h_i.

Similarly

$$\frac{d\mathcal{J}'(\mathbf{u}, N)}{dN} = \left(I - P^{-1}\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)\frac{d\mathbf{w}(\mathbf{u}, N)}{dN} \tag{21}$$

$$\frac{d\mathbf{w}(\mathbf{u}, N)}{dN} = A(\mathbf{u}, N)^{\mathrm{T}}\frac{d\mathbf{v}(\mathbf{u}, N)}{dN} + \frac{dA(\mathbf{u}, N)^{\mathrm{T}}}{dN}\ \mathbf{v}(\mathbf{u}, N) \tag{22}$$

$$\frac{d\mathbf{v}(\mathbf{u}, N)}{dN} = H(N\mathbf{u}+\mathbf{m})\ \mathbf{u} - H(\mathbf{m})\ \mathbf{u} \tag{23}$$

$$\frac{dA(\mathbf{u}, N)}{dN} = [H(N\mathbf{u}+\mathbf{m}) - H(\mathbf{m})] + N\mathbf{u}^{\mathrm{T}} \sum_{i=1}^{m} D_i(N\mathbf{u}+\mathbf{m}) \tag{24}$$

and we finally have that

$$\frac{d\mathbf{u}(N)}{dN} = -\left(\frac{d\mathcal{J}'(\mathbf{u}, N)}{d\mathbf{u}}\right)^{-1}\ \frac{d\mathcal{J}'(\mathbf{u}, N)}{dN} \tag{25}$$

This differential equation can be integrated in the domain $[0, N]$ initializing $\mathbf{u}(0)$ as the eigenvector corresponding to the maximum eigenvalue of $\sum_{i=1}^{m} D_i(\mathbf{m})$ scaled to satisfy $\mathbf{u}(0)^{\mathrm{T}}\ P^{-1}\mathbf{u}(0) = 1$. For infinitesimal values $N = dN$, the function $h_i(N\mathbf{u}+\mathbf{m})$ can be expressed as its second order Taylor series, therefore

$$h_i(dN\mathbf{u}+\mathbf{m}) = h_i(\mathbf{m}) + H_i(\mathbf{m})\ \mathbf{u}\ dN + \frac{1}{2}\mathbf{u}^{\mathrm{T}} D_i(\mathbf{m})\ \mathbf{u}\ dN^2 \tag{26}$$

$$\mathcal{J}(\mathbf{u}, dN) = \max_{\mathbf{u}} \frac{1}{2}\Big[\ \frac{1}{2}\mathbf{u}^{\mathrm{T}}\Big(\sum_{i=1}^{m} D_i(\mathbf{m})\Big)\mathbf{u}\ dN^2 + \lambda\big(\mathbf{u}^{\mathrm{T}} P^{-1}\mathbf{u} - 1\big)\Big] \tag{27}$$

therefore the optimizing value of $\mathcal{J}(\mathbf{u}, dN)$ is the maximum eigenvalue of $\sum_{i=1}^{m} D_i(\mathbf{m})$ scaled to satisfy $\mathbf{u}(dN)^{\mathrm{T}}\ P^{-1}\ \mathbf{u}(dN) = 1$.

Notice that it is necessary to check for the case where both $\frac{d\mathcal{J}'}{d\mathbf{u}}$ and $\frac{d\mathcal{J}'}{dN}$ are zero, which is a situation where the first order condition is stationary and $\frac{d\mathbf{u}}{dN}$ is zero.

As an example, consider the case when

$$h(\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\ \mathbf{x} + \mathbf{x}^{\mathrm{T}} T(\mathbf{x})^{\mathrm{T}} B\ T(\mathbf{x})\ \mathbf{x} \tag{28}$$

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \tag{29}$$

$$T(\mathbf{x}) = \begin{bmatrix} \cos(\mathbf{x}^{\mathrm{T}}\mathbf{x}) & \sin(\mathbf{x}^{\mathrm{T}}\mathbf{x}) \\ -\sin(\mathbf{x}^{\mathrm{T}}\mathbf{x}) & \cos(\mathbf{x}^{\mathrm{T}}\mathbf{x}) \end{bmatrix} \tag{30}$$

the center of the series is taken as $\mathbf{m} = \mathbf{0}$ and $P = I$, hence to satisfy $\mathbf{u}^{\mathrm{T}}\ P^{-1}\mathbf{u}$ we have that $\mathbf{u}$ must be of unit length. The partials are given by

$$H(\mathbf{x}) = \mathbf{a}^{\mathrm{T}} + 2\mathbf{x}^{\mathrm{T}} T(\mathbf{x})^{\mathrm{T}} B \left(T(\mathbf{x}) + 2T'(\mathbf{x})\ \mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \tag{31}$$

$$T'(\mathbf{x}) = \begin{bmatrix} -\sin(\mathbf{x}^{\mathrm{T}}\mathbf{x}) & \cos(\mathbf{x}^{\mathrm{T}}\mathbf{x}) \\ -\cos(\mathbf{x}^{\mathrm{T}}\mathbf{x}) & -\sin(\mathbf{x}^{\mathrm{T}}\mathbf{x}) \end{bmatrix} \tag{32}$$

$$H(\mathbf{0}) = \mathbf{a}^{\mathrm{T}} \tag{33}$$

$$\begin{aligned} D(\mathbf{x}) = {} & 2\left(T(\mathbf{x}) + 2T'(\mathbf{x})\ \mathbf{x}\mathbf{x}^{\mathrm{T}}\right)^{\mathrm{T}} B\ \left(T(\mathbf{x}) + 2T'(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \\ & + 4\left[\mathbf{x}^{\mathrm{T}} T'(\mathbf{x})^{\mathrm{T}} B\ T(\mathbf{x})\mathbf{x}\right]\ I + 4T'(\mathbf{x})^{\mathrm{T}} B\ T(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}} \\ & + 4\mathbf{x}\mathbf{x}^{\mathrm{T}} T(\mathbf{x})^{\mathrm{T}} B^{\mathrm{T}} T'(\mathbf{x}) - 8\mathbf{x}\mathbf{x}^{\mathrm{T}} T(\mathbf{x})^{\mathrm{T}} B\ T(\mathbf{x})\mathbf{x}\mathbf{x}^{\mathrm{T}} \end{aligned}$$

$$D(\mathbf{0}) = 2\ B \tag{34}$$

$$\mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{35}$$

so that,

$$h(N\mathbf{u}) - h(\mathbf{0}) - H(\mathbf{0})N\mathbf{u} = N^2\ \mathbf{u}^{\mathrm{T}} T(N\mathbf{u})^{\mathrm{T}} B\ T(N\mathbf{u})\ \mathbf{u} \tag{36}$$

$$T(N\mathbf{u}) = \begin{bmatrix} \cos(N^2\ \mathbf{u}^{\mathrm{T}}\mathbf{u}) & \sin(N^2\ \mathbf{u}^{\mathrm{T}}\mathbf{u}) \\ -\sin(N^2\ \mathbf{u}^{\mathrm{T}}\mathbf{u}) & \cos(N^2\ \mathbf{u}^{\mathrm{T}}\mathbf{u}) \end{bmatrix} \tag{37}$$

The maximum is attained when $\mathbf{u}$ is equal to $\begin{bmatrix}\cos(N^2) & \sin(N^2)\end{bmatrix}^{\mathrm{T}}$.

Choosing $N = 3$, the optimal value is given by $\mathbf{u}_M = [-0.9111 \quad 0.4121]^{\mathrm{T}}$ and numerical integration using the approach proposed here gives $[-0.9113 \quad 0.4118]^{\mathrm{T}}$.

B. Splitting Criteria

Consider a Kalman update from a linear measurement

$$\mathbf{y} = H\ \mathbf{x} + \boldsymbol{\eta} \tag{38}$$

where the prior state estimate and error covariance are denoted by $\hat{\mathbf{x}}^-$ and P^- while the posteriors are denoted with the superscript $+$. The measurement residual $\boldsymbol{\epsilon}$ and its covariance matrix W are given by

$$\boldsymbol{\epsilon} = \mathbf{y} - H\hat{\mathbf{x}}^- \tag{39}$$

$$W = HP^-H^{\mathrm{T}} + R \tag{40}$$

where R is the estimation error covariance matrix. The posterior residual and covariance are

$$\boldsymbol{\epsilon}^+ = \mathbf{y} - H\hat{\mathbf{x}}^+ \tag{41}$$

$$W^+ = RW^{-1}R \tag{42}$$

it is easy to show that in this linear scenario

$$\boldsymbol{\epsilon}^{\mathrm{T}} W^{-1}\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}^+)^{\mathrm{T}}(W^+)^{-1}\boldsymbol{\epsilon}^+ \tag{43}$$

Assume that the prior is split into several components with weights, mean, and covariance given by w_i, $\hat{\mathbf{x}}_i^-$ and P_i^-, respectively. The split is such that w_i and $\hat{\mathbf{x}}_i^-$ are symmetric about $\hat{\mathbf{x}}^-$, i.e. $\forall i\ \exists j : w_i = w_j \wedge \hat{\mathbf{x}}_i^- - \hat{\mathbf{x}}^- = \hat{\mathbf{x}}^- - \hat{\mathbf{x}}_j^-$. The split is also assumed to be covariance preserving

$$P^- + \hat{\mathbf{x}}^-\ (\hat{\mathbf{x}}^-)^{\mathrm{T}} = \sum_i w_i\ (P_i^- + \hat{\mathbf{x}}_i^-\ (\hat{\mathbf{x}}_i^-)^{\mathrm{T}}) \tag{44}$$

furthermore define $\mathbf{d}_i = \hat{\mathbf{x}}_i - \hat{\mathbf{x}}$ and assume all P_i^- are equal, hence

$$P^- = P_i^- + \sum_i w_i\ \mathbf{d}_i^-\ (\mathbf{d}_i^-)^{\mathrm{T}} \tag{45}$$

Because the measurement is linear, the posterior will be approximately the same whether the prior is processed directly with the Kalman update or as a Gaussian sum.

As an example, consider a Gaussian prior with mean and covariance given by

$$\hat{\mathbf{x}}^- = \begin{bmatrix}-10 & 0\end{bmatrix}^{\mathrm{T}} \tag{46}$$

$$P^- = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \tag{47}$$

and a linear measurement given by

$$y_L = H_L\mathbf{x} + \eta \tag{48}$$

$$H_L = \begin{bmatrix}-1 & 0\end{bmatrix} \tag{49}$$

$$\eta \simeq n(\xi; 0, .01) \tag{50}$$

assume a measurement value of $y_L = 8.5$ and that the prior is split in three along the y-axis, using DeMars' library values [2]. Fig. 1 shows the prior distributions (dashed lines), both the overall distribution and the three components of the GMM approximation are shown. The figure depicts 1σ ellipses, in the GMM case they are scaled by their component weight. Because the system is linear and Gaussian, the Kalman update is adequate and its posterior closely matches the GMM posterior. The top GMM component has a posterior mean of $[-8.5149 \quad 2.1150]$.

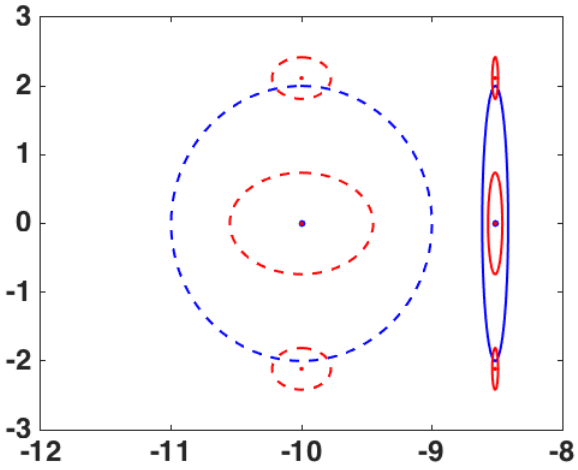

Fig. 1. Splitting Criteria Example for Linear Update

Going back to the theoretical development, under the symmetry and equal covariance assumptions, all components have the same residuals covariance $W_i = HP_i^- H^{\mathrm{T}} + R$ and we have that for each component of the GMM

$$(\boldsymbol{\epsilon}_i^+)^{\mathrm{T}}(W_i^+)^{-1}\boldsymbol{\epsilon}_i^+ = (\boldsymbol{\epsilon} - H\mathbf{d}_i)^{\mathrm{T}} W_i^{-1}(\boldsymbol{\epsilon} - H\mathbf{d}_i) \tag{51}$$

In the numerical example

$$\epsilon = y_L - H_L\hat{\mathbf{x}} = -1.5 \tag{52}$$

$$P_i^- = \begin{bmatrix} 1 & 0 \\ 0 & 1.804 \end{bmatrix} \tag{53}$$

$$\hat{\mathbf{x}}_0^- = \hat{\mathbf{x}} = [-10 \;\; 0]^{\mathrm{T}} \qquad \hat{\mathbf{x}}_{1,2}^- = [-10 \;\; \pm 2.115]^{\mathrm{T}} \tag{54}$$

$$\epsilon_{0,1,2} = \epsilon = -1.5 \tag{55}$$

The ratio of the square residual with its variance is $2.2277 = 1.4926^2$, about at the 1.5σ level.

Let's now assume a nonlinear measurement

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta} \tag{56}$$

the EKF performs an update assuming the Jacobian H can be approximated as constant in the domain of interest; its value evaluated at the prior mean is used

$$H(\hat{\mathbf{x}}^-) = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}^-} \tag{57}$$

To check the validity of the EKF assumption one could look at the difference in posterior estimates for a given component i when the Kalman gain is calculated with $H(\hat{\mathbf{x}}^-)$ and with $H(\hat{\mathbf{x}}_i^-)$, i.e. the mean of the prior and the center of the component. Let $\hat{\mathbf{x}}_i^+$ be the posterior estimate of the i-th component obtained calculating the Kalman gain with $H(\hat{\mathbf{x}}_i^-)$, when the linearization assumption holds, the following quantities should be approximately the same

$$\begin{aligned}(\mathbf{y}-\mathbf{h}(\hat{\mathbf{x}}_i^+))^{\mathrm{T}} R^{-1}\Big(H(\hat{\mathbf{x}}_i^-)P_i^- H(\hat{\mathbf{x}}_i^-)^{\mathrm{T}}+R\Big)R^{-1}(\mathbf{y}-\mathbf{h}(\hat{\mathbf{x}}_i^+)) \\ \approx (\mathbf{y}-\mathbf{h}(\hat{\mathbf{x}}^-) - H(\hat{\mathbf{x}}^-)\mathbf{d})^{\mathrm{T}}\Big(H(\hat{\mathbf{x}}^-)P_i^- H(\hat{\mathbf{x}}^-)^{\mathrm{T}}+R\Big)^{-1} \\ (\mathbf{y}-\mathbf{h}(\hat{\mathbf{x}}^-) - H(\hat{\mathbf{x}}^-)\mathbf{d})\end{aligned} \tag{58}$$

If these two quantities are significantly different, we conclude that the EKF approximation is not valid and a split should occur. A threshold is used to determine if the split should occur. For a scalar measurement the ratio is simply the square of the residual divided by its variance which provides an easy and intuitive meaning of the threshold for tuning purposes. When the measurement is an m-dimensional vector, the ratio is simply the sum of m scalar ratios.

Expanding on the numerical example, assume the same prior as before but consider now the nonlinear measurement case when

$$y_N = h(\mathbf{x}) + \eta = \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}} + \eta \tag{59}$$

$$H_N(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}/\|\mathbf{x}\| \tag{60}$$

$$H_N(\hat{\mathbf{x}}^-) = \begin{bmatrix} -1 & 0 \end{bmatrix} \tag{61}$$

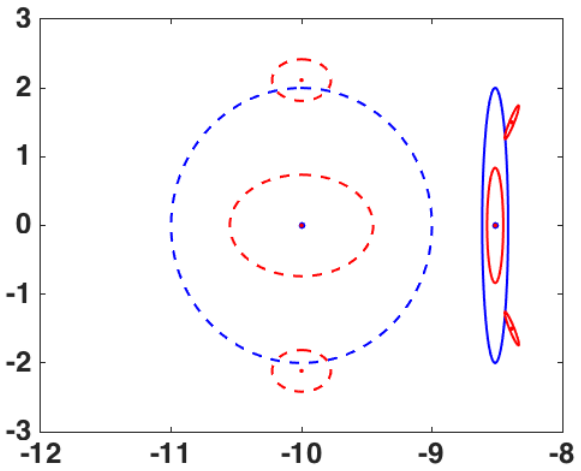

Fig. 2. Splitting Criteria Example for Non-Linear Update

once again $y_N = 8.5$, an EKF gives an update identical to the linear case shown in Fig. 1. Fig. 2 compares the EKF update with the GMM update.

Let's look at the component on the top and let's identify it as component 1. We have that

$$\hat{\mathbf{x}}_1^- = \begin{bmatrix} -10 & 2.115 \end{bmatrix}^{\mathrm{T}} \tag{62}$$

$$P_1^- = \begin{bmatrix} 1 & 0 \\ 0 & 1.804 \end{bmatrix} \tag{63}$$

$$w_1 = 0.2252 \tag{64}$$

$$\epsilon_1 = y_N - \|\hat{\mathbf{x}}_1^-\| = -1.7212 \tag{65}$$

$$W_1 = H(\hat{\mathbf{x}}_1^-)P_i H(\hat{\mathbf{x}}_1^-)^{\mathrm{T}} + R = 1.0444 \tag{66}$$

$$\epsilon_1^2/W_1 = 2.8366 = 1.6842^2 \tag{67}$$

$$\epsilon = y_N - \|\hat{\mathbf{x}}^-\| = -1.5 \tag{68}$$

$$W = H(\hat{\mathbf{x}}^-)P_i H(\hat{\mathbf{x}}^-)^{\mathrm{T}} + R = 1.01 \tag{69}$$

$$\Big(\epsilon^2 - H(\hat{\mathbf{x}}^-)(\hat{\mathbf{x}}_1^- - \hat{\mathbf{x}}^-)\Big)/W = 2.1543 = 1.4678^2 \tag{70}$$

The posterior values are

$$\hat{\mathbf{x}}_1^+ = \begin{bmatrix} -8.3877 & 1.4998 \end{bmatrix}^{\mathrm{T}} \tag{71}$$

$$P_1^+ = \begin{bmatrix} 0.0835 & 0.3497 \\ 0.3497 & 1.6706 \end{bmatrix} \tag{72}$$

$$w_1^+ = 0.1884 \tag{73}$$

$$\epsilon_1^+ = y_N - \|\hat{\mathbf{x}}_1^+\| = -0.0207 \tag{74}$$

$$(\epsilon_1^+)^2/(RW_1^{-1}R) = 4.4768 = 2.1158^2 \tag{75}$$

The linearization assumption puts the posterior residual at 1.4678σ, while re-evaluating the Jacobian puts it at 2.1158σ an increase of more than 33% indicating the linearization assumption is not adequate in this situation.

C. Algorithm Summary

In summary, our proposed algorithm executes the following steps. First the direction of maximum uncertainty is determined integrating Eq. (25). The integration is performed from

$N = 0$ to $N = 1.0908$, which is the number of standard deviations at which the mean of a Gaussian component is placed after the split, as shown in Table I. The rationale is that after determining where the components will be (in terms of number of standard deviations), the goal is to find the direction with the most nonlinearity at that given distance. Once the direction of maximum nonlinearity is established, we perform the candidate split in that direction; if the EKF update of the outer component is within a pre-established threshold from the update performed with the linearized assumption, i.e. the almost equal sign in Eq. (58) holds, no split occurs. Otherwise a split occurs. These steps are taken recursively.

IV. Numerical Example

A set of simulations was performed utilizing the algorithm outlined above. For each, a single range measurement was processed, and splitting of the prior was implemented automatically until the splitting criterion function for every component in the approximated prior was less than a threshold. The criterion used was that in Eq. (58) but weighted by the squared weight of the component in question. Such weighting is efficient in that it prevents splitting of components irrelevant to the overall mixture, and it was seen to improve performance. In automation, every split was taken into 3 components at a time.

A. Example I.

Example I. corresponds to a range measurement of 1, measurement noise variance of 0.01, and single Gaussian prior with the following moments:

$$\hat{\mathbf{x}}_i^- = \begin{bmatrix} -3 \\ 0 \end{bmatrix}, \; P^- = \begin{bmatrix} 1 & -0.1 \\ -0.1 & 0.4 \end{bmatrix} \tag{76}$$

A correlated prior covariance was used to avoid the specific case where the second direction maximizing the performance index is already known to be the one 180° from the first. Fig. 3 shows contours for the given prior and measurement where points along a contour of the same color correspond to equal values of the pdf and likelihood function respectively. Fig. 4 shows contours of the true posterior distribution that the GMM approximating solution seeks to replicate.

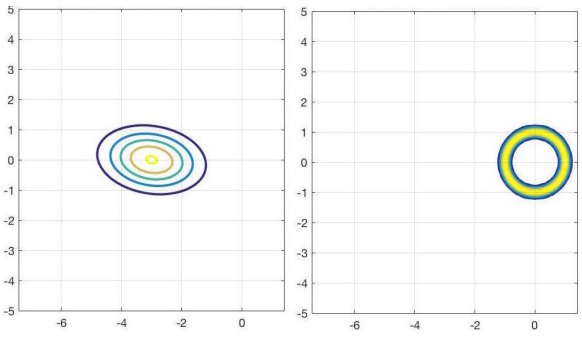

Fig. 3. Prior and Measurement Distribution for Example I.

The EKF and GMM solutions are provided in Fig. 5. Since the true posterior clearly cannot be well described by its first two moments, the EKF solution fails to capture the necessary curvature. However, the GMM solution is found to more accurately adapt to nonlinearities and therefore provides a better estimate of the truth. 67 components were used for this solution.

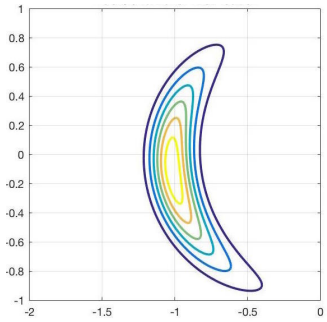

Fig. 4. True Posterior for Example I.

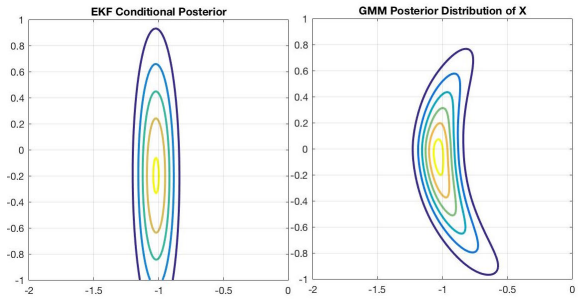

Fig. 5. EKF and GMM Posteriors for Example I.

B. Example II.

Example II. involves a larger range measurement of 4 with the same measurement noise variance but prior moments of:

$$\hat{\mathbf{x}}_i^- = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}, \; P^- = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 4 \end{bmatrix} \tag{77}$$

Contour plots for this new prior and measurement are provided in Fig. 6. The multiplicative treatment of the prior and likelihood in Bayesian updating produces here the different shape shown in Fig. 7.

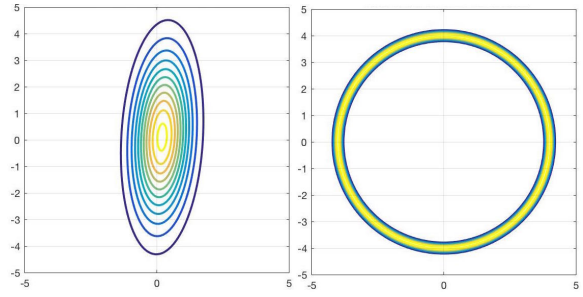

Fig. 6. Prior and Measurement Distribution for Example II.

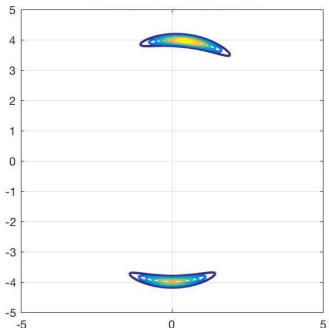

Fig. 7. True Posterior for Example II.

It is one for which the EKF solution is even less equipped to match, as seen in Fig. 8. The GMM solution, with 99 components, drastically better approximated the truth in this case. Doing so provides not only improved filtering of the current measurement but also a much better basis for prediction due to a fuller representation of the posterior pdf.

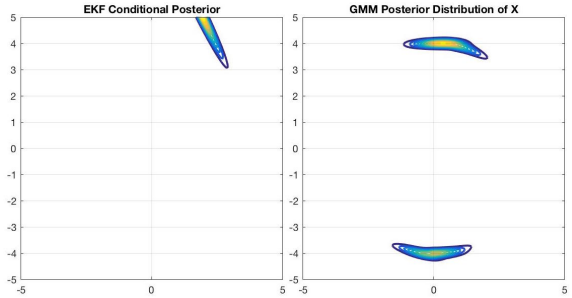

Fig. 8. EKF and GMM Posteriors for Example II.

V. Conclusions

In this work, a method was developed for Bayesian filtering of a nonlinear measurement using Gaussian mixture model representations of the prior. Adding component to a prior Gaussian distribution can be accomplished by the process of splitting, or placing new components along certain directions or dimensions of the state space. The splitting direction was found by examining differences betweens linear approximation (centered at the prior mean) of the measurement function and full nonlinear values. Knowing that the splitting library utilized places the highest weighted components (apart from the mean) at a certain number of standard deviations from the mean, the splitting direction was found to be the one along that particular covariance contour whose linear approximation error is greatest. This direction was found by integrating an equation describing evolution of the maximizing direction from 0 to the number of standard deviations prescribed by the splitting library. In contrast to other schemes, integration was performed only once per direction decision as opposed to over a set of candidates. The splitting scheme was implemented automatically, meaning all components of the prior are split until the linear approximation was acceptable enough to stop. This algorithm was simulated for a pair of range examples to demonstrate the effectiveness of GMM solutions in general and to illustrate the ability of this specific method in mitigating effects of linearization errors.

References

[1] D. Alspach and H. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972, doi: 10.1109/SAP.1970.270017.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156, 2011.

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

[12] M. R. Morelande and . F. Garca-Fernndez, "Analysis of kalman filter approximations for nonlinear measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5477–5484, Nov 2013.

[13] O. Straka, J. Dunk, and I. Puncochar, "Directional splitting for structure adaptation of bayesian filters," in *2016 American Control Conference (ACC)*, July 2016, pp. 2705–2710.

[14] 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.

[15] 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.

[16] R. Zanetti, K. J. DeMars, and K. Tuggle, "Orbital Uncertainty Quantification Using Directional Splitting and Gaussian Mixture Models," *Submitted for Publication*, 2018.