

A SPLITTING GAUSSIAN MIXTURE FORMULATION FOR A NONLINEAR MEASUREMENT UPDATE

Kirsten Tuggle*, Renato Zanetti[†], Christopher D'Souza[‡]

Applications such as orbit determination and ground tracking necessitate accurate knowledge of state estimate uncertainty. Adaptive Gaussian mixture models are a favored approach with much work devoted to the prediction phase. Of the fewer splitting schemes provided for the filtering phase, nearly all choose to split components along the direction of maximum prior uncertainty. While this does address the goal of reducing uncertainty, it does not consider the role of the actual nonlinear measurement function in doing so. The proposed method offers a novel filtering step splitting scheme that addresses both contributors in a more computationally efficient manner than much of the literature.

INTRODUCTION

Aerospace estimation applications such as orbit determination, ground tracking, and ground filtering necessitate accurate knowledge of the uncertainty associated with state estimates. This most often must be achieved for systems governed by nonlinear dynamics and observed by nonlinear measurement sources. Since most operational nonlinear filtering techniques are approximates to the ideal Bayesian framework, a trade must be made between computational efficiency and estimator accuracy. A promising avenue that has garnered much attention since the 1970s is that of a Gaussian mixture formulation, in which conditional state distributions are modeled as weighted sums of individual Gaussian distributions.¹ This may be conceived as a generalization to the particle filter approach in which individual components have covariances that are finite rather than infinitesimal. A chosen nonlinear filtering scheme is then applied to each individual Gaussian component so that the final posterior distribution is a sum of alternatively weighted individual component posteriors. In doing so, many of the dimensionality and computational disadvantages of particle filters are mitigated. The number of mixture components can be adapted as system dynamics evolve and nonlinear measurements are accrued.² The number of components can be increased by splitting a subset of prior components into a larger number of components, and the number can be reduced by merging subsets of components. Splitting is the more typical operation due to the accumulation of the effect of nonlinearities over time. Due to this, and the fact that many splitting philosophies have similar counterparts in the merging process, many contributions, including this one, only develop splitting approaches.

*PhD Student, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, TX 78712.

[†]Assistant Professor, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, TX 78712.

[‡]Navigation Technical Discipline Lead, Aeroscience and Flight Mechanics Division, EG6, 2101 NASA Parkway. NASA Johnson Space Center, Houston, Texas 77058.

Many orbit determination problems are subject to sparse or infrequent measurement availability. The resulting long periods of propagation require great care in filter uncertainty prediction to avoid divergence issues. Accordingly, much consideration has been given to this phase with less work devoted to the filtering or measurement update phase. Accurate quantification of uncertainty after a Bayesian incorporation of information concerning the states using available observations significantly enhances performance of the approximate nonlinear estimator. Better consistency between the estimated filter uncertainty and the actual quality of the state estimates following a measurement update provides not only a better use of measurement information but also a better basis for prediction in the propagation phase.

A strategy for performing a measurement update on a Gaussian mixture model involves two main actions: designating criteria for splitting to determine if and which components warrant the computational cost of splitting as well as selection of a direction along which to split if splitting is to occur. The total process has been referred to as Gaussian sum reapproximation or resampling.³ The act of splitting a Gaussian component, effectively replacing it with its own mixture model representation, serves as a means to ensure the validity of the applied nonlinear estimator for all components. Each component within the substituting mixture will have more confined covariance bounds than that of the original Gaussian component so that after splitting, each component may be thought of as more local and therefore a better candidate for nonlinear filtering approaches which rely on local Gaussianity assumptions, i.e. the extended Kalman filter (EKF).⁴

Recognizing that the act of splitting reduces uncertainty along the splitting direction, nearly all approaches have selected this direction to be that of maximum variance for the a priori distribution.^{2,3} This corresponds to selecting the principal axis of the prior covariance ellipse associated with the maximum eigenvalue of the prior covariance matrix. However, this method is invariant to the actual nonlinear function at play, and the ultimate goal is reduction of uncertainty in the posterior rather than prior distribution. Another important point that has only been considered more recently is that the effects of nonlinearities in the measurement model are usually not confined to a single direction. One reapproximation method has been developed to effectively achieve splitting for the subspace of the parameter space associated with a designated significance in nonlinearity via the use of linear matrix inequalities enforced on component covariance matrices.³ Similarly, one-dimensional approaches have been modified to account for multiple directions by recursive implementations. A few methods have utilized statistical linearization.^{5,6} This provides a more direct quantification of the effect of nonlinearities on a solution but in some cases comes at a significant increase in computational cost with numerical rather than analytical solutions. Furthermore, some solutions are tailored only for use in an Unscented Kalman Filter framework.⁶ Computational efficiency becomes increasingly significant when the scheme used is applied recursively. For this reason, closed-form solutions are very desirable. Some analytical solutions have been provided but only for special cases of the nonlinear measurement model or rely upon intuition concerning the dynamics model in the case of the propagation phase.^{7,8} An approach similar to the proposed methodology provides an analytical expression for both the splitting criterion and selection of splitting direction.⁹ Substantial differences between the two approaches, however, occur in the optimization problems posed and in the thresholds chosen. A key contribution of Ref. 9 is the use of a numerical evaluation that avoids analytical differentiation; a feature common to both methods.

METHODOLOGY OVERVIEW

The current work focuses on the popular scenario in which an EKF is to be applied to Gaussian mixture components. Characterization of nonlinear effects is motivated by contributions of second-order Taylor series terms in updating means (estimates) and covariances (uncertainties) as well as the difference in information content between a linearized (EKF) solution and a higher-order, in this case second-order, solution. The selected criterion for splitting the prior GMM depends on the reduction in performance induced by the use of a linearized solution. The proposed method characterizes the loss of state information associated with the use of a standard EKF instead of a second-order EKF using Kullback-Leibler divergence, a popular information-theoretic quantity.¹⁰ Although the second-order EKF solution only involves the inclusion of a small number of additional terms, it does require computing the Hessian of the measurement model. In order to address this, an alternate criterion is provided that captures the significance of the additional term in the innovation covariance for the second-order EKF but avoids analytical differentiation via the use of an existing numerical method for a key term.⁹ To complete the scheme, a new choice of splitting direction is provided that exhibits desirable improvements with respect to the literature. The chosen direction corresponds to that which maximizes a cost function composed of the following terms: the directional derivative of the Jacobian evaluated at the prior estimate and the prior covariance conditioned such that all states fall along the direction at hand. A great benefit of this choice over much of the literature is the consideration of directions on the basis of significance of the nonlinearity of the measurement function rather than only prior covariance. Furthermore, this cost function reduces to a Rayleigh quotient expression which admits an analytical solution simply corresponding to an eigenvalue/eigenvector problem. A substantial difference from this work and Ref. 9 is that the present solution does not check the second order contributions to the covariance, but rather performs a test on the direction second order contributions of the nonlinearities.

Simulations are provided for measurement models commonly associated with Gaussian mixture model applications. Performance of the algorithm is evaluated for a variety of prior distributions and compared to proposed alternatives.

PROBLEM FORMULATION

In sequential state estimation, we may view any single measurement update as a parameter estimation problem at that instant with all previous filter knowledge provided by the prior state probability density function. The observation model accommodates a nonlinear transformation of the state.

$$\begin{aligned} y &= h(x) + \epsilon \in \mathbb{R} \\ \epsilon &\sim N(0, R) \\ R &\succeq 0 \\ \mathbb{E}\{\epsilon x^T\} &= 0_{1 \times n} \end{aligned} \tag{1}$$

Note that the current work assumes scalar measurements. A priori knowledge of the state is:

$$\begin{aligned} x &\sim N(\bar{x}, P) \\ x &\in \mathbb{R}^n \\ P &\succeq 0 \end{aligned} \tag{2}$$

The stated problem may represent treatment of the full filter in the case of a normally distributed prior or treatment of a single Gaussian mixture component (which is initially normally distributed

by formulation). The update will be performed by an EKF, the most ubiquitous nonlinear filtering scheme. With the filtering tool decided, the problem effectively becomes one of determining a representation or reapproximation of the prior pdf such that the EKF-solution of the modified problem best represents the true Bayesian solution of the full nonlinear problem under an acceptable computational cost. Furthermore, the reapproximation framework used is a general Gaussian mixture model rather than say, a collection of particles. As previously mentioned, determination of a reapproximation of the prior reduces to prescriptions for how and when splitting into a Gaussian mixture should occur. Before these ideas are developed, however, the EKF tool should be briefly reviewed.

EXTENDED KALMAN FILTER SUMMARY

Traditional (First-order) EKF

In the first-order EKF framework, the observation function is represented by its linearized form, resulting from truncating the Taylor series expansion about the prior mean after first-order terms.

$$\begin{aligned}
y &= h(\bar{x}) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}} [x - \bar{x}] + \mathcal{O}(\|x - \bar{x}\|_2^2) + \epsilon \\
&\approx h(\bar{x}) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}} [x - \bar{x}] + \epsilon \\
&= h(\bar{x}) + H[x - \bar{x}] + \epsilon \\
&\triangleq y_{EKF}
\end{aligned} \tag{3}$$

Terms H and y_{EKF} are defined to be the observation function Jacobian evaluated at the prior mean and the linearized observation model respectively. A Kalman filter may be applied for an optimal solution to the linearized problem.

$$\begin{aligned}
\hat{x} &= \bar{x} + K(y - \bar{y}_{EKF}) \\
P^+ &= P - K_{EKF} W_{EKF} K_{EKF}^T \\
K_{EKF} &= P H^T W_{EKF}^{-1}
\end{aligned} \tag{4}$$

where

$$\begin{aligned}
\bar{y}_{EKF} &= \mathbb{E}\{y_{EKF}\} = \mathbb{E}\{h(\bar{x}) + H[x - \bar{x}] + \epsilon\} = h(\bar{x}) \\
W_{EKF} &= \mathbb{E}\{[y_{EKF} - \bar{y}_{EKF}][y_{EKF} - \bar{y}_{EKF}]^T\} \\
&= \mathbb{E}\{H[x - \bar{x}][x - \bar{x}]^T H^T\} + \mathbb{E}\{H[x - \bar{x}]\epsilon^T\} + \mathbb{E}\{\epsilon[x - \bar{x}]^T H^T\} + \mathbb{E}\{\epsilon\epsilon^T\} \\
&= H P H^T + R
\end{aligned} \tag{5}$$

Second-order EKF

For a second-order EKF, the overall update is designated to take the same linear form.

$$\begin{aligned}
\hat{x} &= \bar{x} + K(y - \bar{y}_{SEKF}) \\
P^+ &= P - K_{SEKF} W_{SEKF} K_{SEKF}^T \\
K_{SEKF} &= P H^T W_{SEKF}^{-1}
\end{aligned} \tag{6}$$

However, the observation model is effectively replaced by its second-order Taylor series expansion rather than the linearized version, so estimated moments of the innovation process incorporate some

higher-order terms - specifically, those that are quadratic in the state.

$$\begin{aligned} y_{SEKF} &\triangleq h(\bar{x}) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}} [x - \bar{x}] + \frac{1}{2} [x - \bar{x}]^T \left. \frac{\partial^2 h}{\partial^2 x} \right|_{\bar{x}} [x - \bar{x}] + \epsilon \\ &= h(\bar{x}) + H[x - \bar{x}] + \frac{1}{2} [x - \bar{x}]^T D[x - \bar{x}] + \epsilon \end{aligned} \quad (7)$$

Term D is defined to be the Hessian of the observation function evaluated at the prior mean. It can be shown that innovation process terms reduce to the following.¹¹

$$\begin{aligned} \bar{y}_{SEKF} &= \mathbb{E}\{y_{SEKF}\} = \mathbb{E}\{h(\bar{x}) + H[x - \bar{x}] + \frac{1}{2} [x - \bar{x}]^T D[x - \bar{x}] + \epsilon\} = h(\bar{x}) + \frac{1}{2} \text{tr}(DP) \\ W_{SEKF} &= \mathbb{E}\{[y_{SEKF} - \bar{y}_{SEKF}][y_{SEKF} - \bar{y}_{SEKF}]^T\} \\ &\approx HPH^T + R + \frac{1}{2} \text{tr}(DPDP) \end{aligned} \quad (8)$$

DEVELOPMENT OF THE SPLITTING SCHEME

In developing a philosophy for splitting, the effects both of prior uncertainty and nonlinearity on the EKF solution must be somehow quantified and ideally minimized, or at the very least significantly mitigated.

Splitting Direction

The governing idea for capturing nonlinearity in the current work relies on behavior of the observation function Jacobian. In the case of a linear observation model, the Jacobian is constant over the entire state space. In other words, the rate of change of the Jacobian with respect to the state, meaning the Hessian of the observation function, is identically zero when h is linear in x. Large changes in the Jacobian then are associated with the significance of higher-order terms in the observation function. The directional derivative evaluated at the prior mean, or the rate of change of the Jacobian at \bar{x} in direction \hat{u} , is a reasonable quantification of nonlinearity.

$$\nabla_{\hat{u}} H(x) \Big|_{\bar{x}} = \lim_{\alpha \rightarrow 0} \frac{H(\bar{x} + \alpha \hat{u}) - H(\bar{x})}{\alpha} \quad (9)$$

The Taylor series for the Jacobian provided in Eq. (10) can be substituted into Eq. (9) to provide Eq. (11).

$$H(x) = H(\bar{x}) + \left. \frac{\partial H}{\partial x} \right|_{\bar{x}} [x - \bar{x}] + \dots \quad (10)$$

$$\begin{aligned} \nabla_{\hat{u}} H(x) \Big|_{\bar{x}} &= \lim_{\alpha \rightarrow 0} \frac{H(\bar{x}) + \alpha \left. \frac{\partial H}{\partial x} \right|_{\bar{x}} \hat{u} + \frac{1}{2} \alpha^2 \hat{u}^T \left. \frac{\partial^2 H}{\partial^2 x} \right|_{\bar{x}} \hat{u} + \dots - H(\bar{x})}{\alpha} \\ &= \left. \frac{\partial H}{\partial x} \right|_{\bar{x}} \hat{u} \\ &= D\hat{u} \end{aligned} \quad (11)$$

In order to quantify the magnitude of prior uncertainty for a specific direction, one may examine the standard deviation of the state pdf when the prior is conditioned on all directions orthogonal to that considered. It can be shown, as in the App., that this is the quantity given in Eq. (12).

$$\sigma_{\hat{u}} = \frac{1}{\sqrt{\hat{u}^T P^{-1} \hat{u}}} \quad (12)$$

With both contributors to degradation of the EKF-solution quantified, the following cost function is chosen for the current work.

$$\begin{aligned} J(\hat{u}) &= \|\nabla_{\hat{u}} H(x) \Big|_{\hat{x}}\|^2 \sigma_{\hat{u}}^2 \\ &= \frac{\hat{u}^T D^T D \hat{u}}{\hat{u}^T P^{-1} \hat{u}} \end{aligned} \quad (13)$$

As mentioned, a major benefit of this choice is the incorporation of both prior uncertainty and non-linearity of h in the choice of splitting direction. The form of $J(\hat{u})$ may be thought of a means for selecting directions with relevant prior uncertainty, in that a direction in which the filter is significantly uncertain but along which the observation function is nearly linear will be given less priority than perhaps a direction with less uncertainty but much more nonlinearity. $J(\hat{u})$ similarly may be thought of as capturing relevant nonlinearity.

Note that nonlinearity is evaluated only locally at the prior mean. Another scheme integrates nonlinearity along each direction but does not result in an exact, closed-form solution for the cost function like that which is provided here.⁵ Furthermore, only eigendirections of the prior are considered, a common trait in other methods.² Another advantage of the chosen cost function is that it is not restricted to a certain estimator by formulation, as opposed to other schemes which for example require terms used in an unscented Kalman filter.⁶

The maximizer of the cost function in Eq. (13) is easily attained once it is recognized as a Rayleigh quotient expression. The solution follows easily from a change of variables and the Rayleigh-Ritz inequality.

$$\begin{aligned} P^{-1} &= P^{-T/2} P^{-1/2} \\ v &= P^{-1/2} \hat{u} \end{aligned} \quad (14)$$

$$J(v) = \frac{v^T P^{T/2} D^T D P^{1/2} v}{v^T v} \quad (15)$$

The Rayleigh-Ritz inequality states:

$$\lambda_{\min}(P^{T/2} D^T D P^{1/2}) v^T v \leq v^T P^{T/2} D^T D P^{1/2} v \leq \lambda_{\max}(P^{T/2} D^T D P^{1/2}) v^T v \quad (16)$$

Substituting Eq. (16) into Eq. (15) gives the following result.

$$J(v) \leq \frac{\lambda_{\max}(P^{T/2} D^T D P^{1/2}) v^T v}{v^T v} = \lambda_{\max}(P^{T/2} D^T D P^{1/2}) \quad (17)$$

Clearly then, the cost function is maximized when v is taken to be the eigenvector associated with the maximum eigenvalue of matrix $D P^{1/2}$, so the choice of a splitting direction is equivalent to an eigenvalue/eigenvector problem. Note that a modification of Ref. 9 provides a means for avoiding analytical evaluation of the observation function Hessian.

Splitting Criteria

A natural choice for a splitting criterion given the previous discussion utilizes the maximum of the cost function associated with the potential splitting direction.

$$\lambda_{\max}(P^{T/2} D^T D P^{1/2}) \geq \tau \quad (18)$$

This would result in very few additional calculations in the event that splitting is deemed necessary. The term τ represents a threshold value.

The remaining options rely on comparison of the linearized EKF solution with a higher-order, chosen here to be the second-order EKF, solution. Knowing that incorporation of the quadratic terms in the SEKF appears within the innovation terms, one option for a criterion involves the difference in innovation covariances.

$$\text{tr}(DPDP) \geq \gamma HPH^T \quad (19)$$

In this scenario, if the additional term in the SEKF is deemed to be large enough relative to the term from the EKF (via some ratio γ), then splitting is chosen to occur. Again, Ref. 9 provides a numerical means for calculating matrix DP .

Another option that compares the EKF and SEKF solutions is the Kullback-Leibler divergence between Gaussian posteriors approximated by both filters. τ is used again to represent some threshold.

$$D_{KL} \left(N(\hat{x}_{SEKF}, P_{SEKF}^+) \parallel N(\hat{x}_{EKF}, P_{EKF}^+) \right) \geq \tau \quad (20)$$

where

$$D_{KL} \left(N(\hat{x}_{SEKF}, P_{SEKF}^+) \parallel N(\hat{x}_{EKF}, P_{EKF}^+) \right) = \frac{1}{2} \left(\ln \frac{|P_{EKF}^+|}{|P_{SEKF}^+|} - n \right) + [\hat{x}_{EKF} - \hat{x}_{SEKF}]^T P_{EKF}^{+^{-1}} [\hat{x}_{EKF} - \hat{x}_{SEKF}] + \text{tr}(P_{SEKF}^+ P_{EKF}^{+^{-1}}) \quad (21)$$

A drawback for this method is that it does essentially require carrying the SEKF along with the EKF. However, from Eq. (5) and Eq. (8), few additional terms are necessary, and this burden is much less than if the additional estimator were, for example, a sigma-point filter. The main advantage of this method is that it provides a clearer representation of the loss of information associated with a linearized solution. Starting from a Gaussian prior, the EKF framework results in a Gaussian posterior, although the true conditional state pdf will not be Gaussian for general nonlinear h . Eq. (20) is therefore a fitting quantity to represent the role that nonlinearities in h plays in making the update to a Gaussian component differ from a Gaussian posterior. A K-L divergence that is above the threshold for splitting, signals significant failure in approximating the posterior solution as Gaussian. When this occurs, the prior is reapproximated as Gaussian mixture model, components then of the reapproximated prior will be more local and more suitable for an EKF solution and will contribute to a Gaussian mixture approximation of the posterior.

RESULTS AND DISCUSSION

Results from a set of three simple examples will visually illustrate important aspects of the proposed splitting scheme and the utility of Gaussian mixture models in general. Finally, a quantitative performance summary of more extensive simulation will be provided.

All presented scenarios involve a single measurement update as detailed in the Problem Formulation section with range as the nonlinear measurement model. For all simulations, the measurement noise follows the standard normal distribution. All scenarios also utilize a single iteration of the scheme with distributions undergoing a single split into twenty-five components. Many other implementations are possible, including those which iterate the scheme until the criterion signals a stop when no further splitting is deemed necessary.

Example I.

A classic example for Gaussian mixture model filtering is presented first. It serves as motivation for Gaussian mixture representations in general whereas the next example highlights advantages of the particular scheme presented.

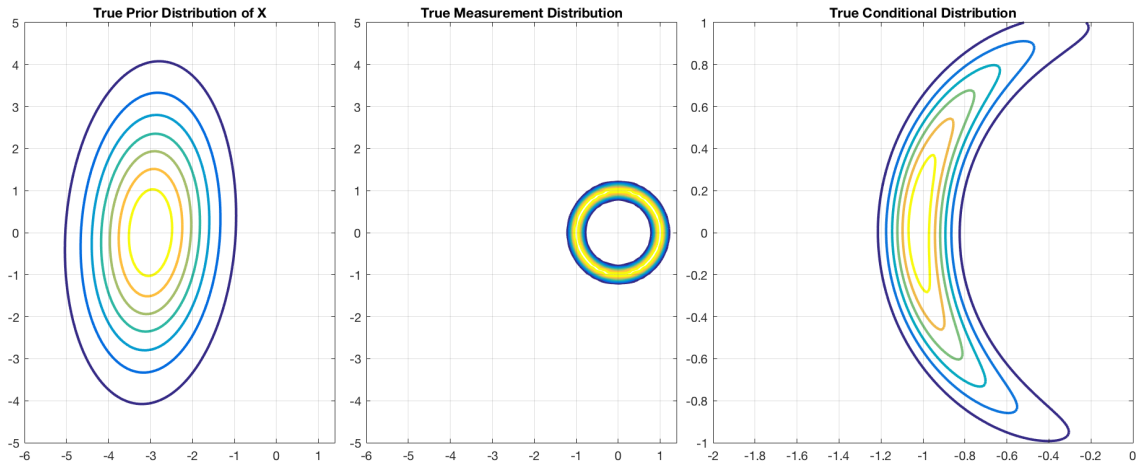


Figure 1. True Prior, Measurement Distribution, and Posterior for Example I.

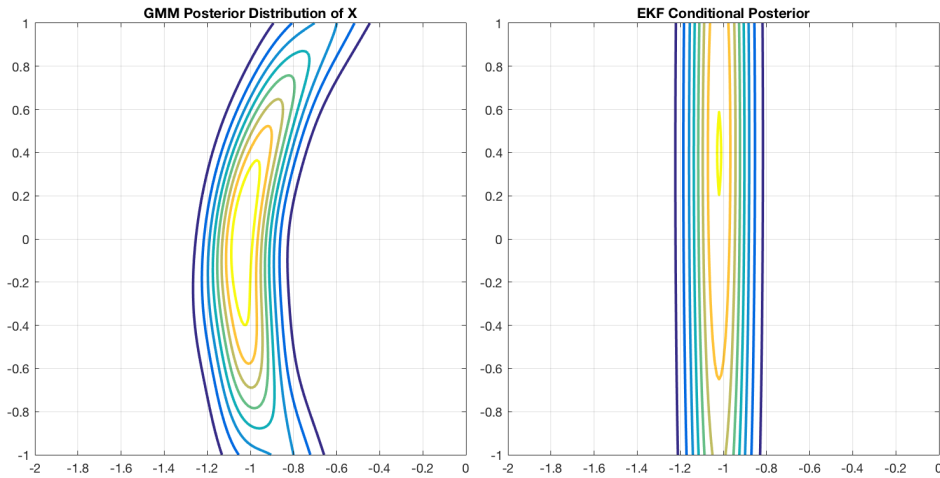


Figure 2. GMM and EKF Posteriors for Example I.

As seen in Figure 1, the true state posterior conditioned on the observed range appears in what is often called the “banana” shape. Starting from a Gaussian prior, the EKF formulation predicts a Gaussian posterior, although the EKF does not explicitly assume any distribution, it only estimates its first two moments. Figure 2 shows on the right a Gaussian distribution with mean and covariance coinciding with those of the EKF solution. This representation is chosen to show that the EKF is unable to correctly predict the mean and covariance of the posterior because it cannot possibly capture the banana shape. Notice that the prior covariance reflects a slight correlation in the state components and much more variance in the nearly vertical major axis than in the nearly horizontal minor axis of the ellipse. This results in the presented splitting scheme returning the principal axis

as the chosen splitting direction, even when nonlinear measurement effects are taken into account.

Example II.

This scenario follows the setup of Example I. but with an altered initial covariance. Here, no single direction displays dominant prior uncertainty, so traditional schemes fail to provide a single desired direction.

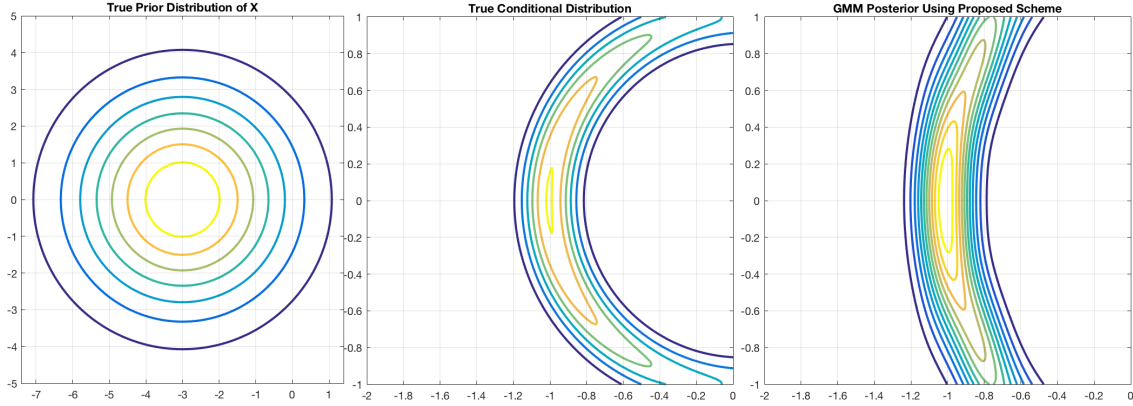


Figure 3. Prior, True Posterior, and GMM Posterior for Example II.

The results of Example II. highlight the significance of nonlinear effects in the true Bayes solution and the benefits of attempting to address them in the choice of a splitting direction. The measurement nonlinearity term in the numerator of the cost function in Eq. (13) is maximized by the coordinate frame vertical axis. With all directions having equal values in the denominator, this direction is selected for splitting by the proposed scheme. Visually, it is clear from the true conditional posterior in Figure 3. that the nonlinearities elongated the spread more vertically than horizontally, so placing and filtering more components along that axis better approximates the true nonlinear estimator.

Example III.

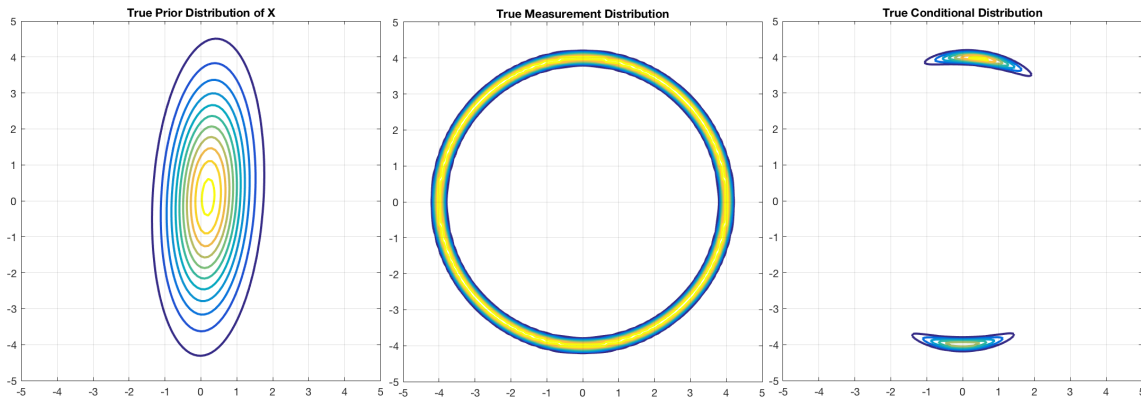


Figure 4. True Prior, Measurement Distribution, and Posterior for Example III.

As seen in Figure 4, Example III. involves a larger range measurement with respect to the size of the initial covariance ellipse. By the multiplicative nature between the likelihood function and prior distribution in Bayes theorem, the resulting posterior displays different significant equiprobability contours. The result is two distinct blobs or individual banana shapes.



Figure 5. EKF Prior, Measurement Distribution, and Posterior for Example III.

By formulation, the traditional EKF method cannot match the form of the true posterior in this scenario. The solution shown in Figure 5. does not even remotely resemble that of the truth in Figure 4., as opposed to those shown in Example I.

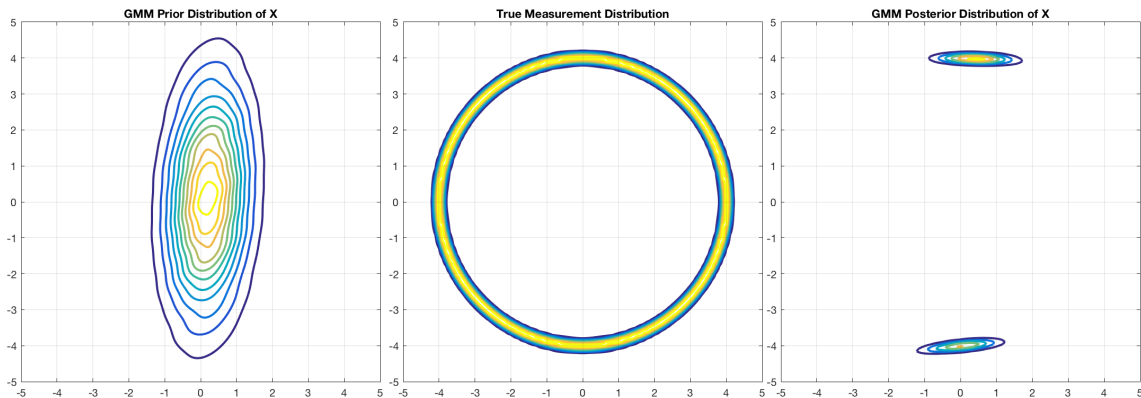


Figure 6. GMM Prior, Measurement Distribution, and Posterior for Example III.

Figure 6. again demonstrates the flexibility and benefits of Gaussian mixture formulations. The splitting scheme was much better able to reflect the effects that prior uncertainty and measurement nonlinearity had on the geometry of the full nonlinear solution.

Analysis for Multiple Runs

In order to provide a more quantitative evaluation of the current scheme, a series of randomized simulations inspired by Ref. 9 was performed. The following conditions were varied: initial mean with elements selected uniformly from 0 and 10, a cross-covariance term selected uniformly

between -10 and 10, and a perfect observation chosen uniformly between 0 and 10. Diagonal elements of the prior covariance were each set to be 10 for every run, and the observed range resulted from summing the perfect observation with a single realization of the measurement noise. Kullback-Leibler divergences were approximated between the Monte-Carlo-calculated true posterior and the GMM solution as well as the truth and the EKF solution. From Gibbs' inequality, D_{KL} is always non-negative. The ratio between these divergences was averaged across each set of runs and reported with that of the EKF in the numerator divided by that of the GMM solution.

Table 1. Relative Performance of Proposed Scheme

Number of Runs	$D_{KL}(truth, EKF)/D_{KL}(truth, GMM)$
100	5.079
1000	5.302

Results in Table 1. are nearly equal between scenarios with an order of magnitude of difference in the number of runs. Significantly, the resulting Kullback-Leibler divergences indicated an average of over five times better approximation of this form of information content present in the true solution. Again, these results were obtained after a single, twenty-five component split. The scheme could be implemented in the same way with more components or with less components and recursive evaluation and computation of the splitting scheme.

CONCLUSIONS

Among the necessary state-space discretization approaches for approximating Bayesian solutions to nonlinear filtering, Gaussian mixture models offer a promising balance between accuracy and computational demand. This work outlined a splitting philosophy and formulation with compelling benefits with respect to many of those present in the existing literature. Specifically, incorporation of the specific nonlinear nature of the system in a simple and practically feasible manner was provided. The rewards of such an approach were illustrated via visual and numerical results.

APPENDIX: DIRECTIONAL VARIANCE

The directional standard deviation given in Eq. (12) will be derived. First, suppose direction \hat{u} is an eigendirection of matrix P , that is, it is a principal axis of the covariance ellipsoid defined by the locus of points satisfying the following:

$$x^T P^{-1} x = 1 \quad (22)$$

Notice that the quantity in Eq. (12) corresponds to the length between the origin and the surface of the ellipsoid in the direction of \hat{u} . This may be verified by taking a point called x^* that lies on the ellipsoid in direction \hat{u} and solving for its magnitude using Eq. (22).

$$\|x^*\|^2 \hat{u}^T P^{-1} \hat{u} = 1 \implies \|x^*\|^2 = \frac{1}{\hat{u}^T P^{-1} \hat{u}} = \sigma_{\hat{u}}^2 \quad (23)$$

It must be shown that conditional covariance (conditioned on all remaining principal directions) results in Eq. (12). Since \hat{u} is a principal direction, which we can call direction 1 without loss of generality, the conditional covariance is given by the well-known formula below for a Gaussian distribution.

$$P_{1|2} = P_{11}^{-1} - P_{12} P_{22}^{-1} P_{21} \quad (24)$$

where the prior covariance is partitioned as

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \quad (25)$$

Using block matrix inversion, the squared-value given in Eq. (12) may be expressed as follows.

$$\begin{aligned} \frac{1}{\hat{u}^T P^{-1} \hat{u}} &= \frac{1}{[1 \ 0 \ \dots \ 0] \begin{bmatrix} (P_{11}^{-1} - P_{12} P_{22}^{-1} P_{21})^{-1} & * \\ * & * \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \\ &= \frac{1}{(P_{11}^{-1} - P_{12} P_{22}^{-1} P_{21})^{-1}} = P_{11}^{-1} - P_{12} P_{22}^{-1} P_{21} = P_{1|2} \end{aligned} \quad (26)$$

Now, it only remains to be shown that the relationship holds when \hat{u} is not a principal direction. For this, an orthogonal change of basis is performed such that the first axis of the transformed basis is aligned with \hat{u} .

$$\tilde{x} = Vx \text{ such that } V\hat{u} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } VV^T = I \quad (27)$$

It will be shown that the conditional variance for direction \hat{u} , which can be calculated using the traditional formula in the transformed frame, will match that provided by Eq. (12) in the original frame. Moments for the transformed state are:

$$\begin{aligned} \mathbb{E}\{\tilde{x}\} &= V\bar{x} \\ \mathbb{E}\{[\tilde{x} - \mathbb{E}\{\tilde{x}\}][\tilde{x} - \mathbb{E}\{\tilde{x}\}]^T\} &= V P V^T \end{aligned} \quad (28)$$

The result in Eq. (26) holds in the transformed frame as well, giving:

$$\begin{aligned} \tilde{P}_{1|2} &= \frac{1}{[1 \ 0 \ \dots \ 0] \tilde{P}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \\ &= \frac{1}{(V\hat{u})^T (V P V^T)^{-1} V\hat{u}} \\ &= \frac{1}{\hat{u}^T V^T V P^{-1} V^T V \hat{u}} = \frac{1}{\hat{u}^T P^{-1} \hat{u}} \end{aligned} \quad (29)$$

The result holds then for all directions in the state space.

REFERENCES

- [1] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE transactions on automatic control*, Vol. 17, No. 4, 1972, pp. 439–448.

- [2] K. J. DeMars, R. H. Bishop, and M. K. Jah, "Entropy-based approach for uncertainty propagation of nonlinear dynamical systems," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 4, 2013, pp. 1047–1057.
- [3] 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, 2015, pp. 292–303.
- [4] A. Gelb, *Applied optimal estimation*. MIT press, 1974.
- [5] M. F. Huber, "Adaptive Gaussian mixture filter based on statistical linearization," *Information fusion (FUSION), 2011 proceedings of the 14th international conference on*, IEEE, 2011, pp. 1–8.
- [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, 2014, pp. 461–474.
- [7] A. Rauh, K. Briechele, and U. D. Hanebeck, "Nonlinear measurement update and prediction: Prior density splitting mixture estimator," *2009 IEEE Control Applications,(CCA) & Intelligent Control,(ISIC)*, IEEE, 2009, pp. 1421–1426.
- [8] V. Vittaldev, R. Linares, and R. P. Russell, "Spacecraft Uncertainty Propagation Using Gaussian Mixture Models and Polynomial Chaos Expansions," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 9, 2016, pp. 2163–2169.
- [9] M. Raitoharju and S. Ali-Loytty, "An adaptive derivative free method for Bayesian posterior approximation," *IEEE Signal Processing Letters*, Vol. 19, No. 2, 2012, pp. 87–90.
- [10] J. M. Joyce, "Kullback-leibler divergence," *International Encyclopedia of Statistical Science*, pp. 720–722, Springer, 2011.
- [11] R. Zanetti, K. J. DeMars, and R. H. Bishop, "Underweighting nonlinear measurements," *Journal of guidance, control, and dynamics*, Vol. 33, No. 5, 2010, pp. 1670–1675.