Non-Gaussian Estimation and Observer-Based Feedback using the Gaussian Mixture Kalman and Extended Kalman Filters

Debdipta Goswami[1] and Derek A. Paley[2]

Abstract— **This paper considers the problem of non-Gaussian estimation and observer-based feedback in linear and nonlinear settings. Estimation in nonlinear systems with non-Gaussian process noise statistics is important for applications in atmospheric and oceanic sampling. Non-Gaussian filtering is, however, largely problem specific and mostly sub-optimal. This manuscript uses a Gaussian Mixture Model (GMM) to characterize the prior non-Gaussian distribution, and applies the Kalman filter update to estimate the state with uncertainty. The boundedness of error in both linear and nonlinear cases is analytically justified under various assumptions, and the resulting estimate is used for feedback control. To apply GMM in nonlinear settings, we utilize a common extension of the Kalman filter: the Extended Kalman Filter (EKF). The theoretical results are illustrated by numerical simulations.**

I. INTRODUCTION

Knowledge of state variables in dynamical systems is the key component for designing state-feedback control laws that can practically stabilize a plant. However in actual systems, lack of access to state variables motivates observer-based feedback control systems, where the current state is estimated from a set of observables. The situation worsens when the system dynamics and/or the observations are corrupted with noise, and the initial state estimate may only be given in terms of an *a priori* probability density. Hence, the need for a filter arises for state estimation. For a linear system with linear observations and Gaussian process and sensor noise, the optimal estimator is given by the celebrated Kalman filter [1]. But for a nonlinear system and systems with non-Gaussian noise, the moment propagation needs an infinite number of parameters and an optimal filter is difficult to formulate. As a compromise, several sub-optimal solutions have been developed, e.g., the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF), which rely on a finite number of parameters, thus sacrificing some of the system's details. However, these filters still rely on the Gaussian-process noise model, which may not hold in a nonlinear system.

The recursive Bayesian filter is the general representation of an optimal nonlinear filter that converts the prior (forecast) state probability density function (PDF) into the posterior (analysis) state PDF using the likelihood of the observation. However in many cases, the prior PDF is not explicitly known and therefore has to be estimated using an ensemble realization. For such cases, the recursive Bayesian estimation becomes computationally difficult and, therefore, a semi-parametric model must be realized.

Alspach and Sorenson [2] have shown that any probability density function can be approximated arbitrarily closely from an ensemble realization using a weighted sum of Gaussian PDFs. The so-called Gaussian Mixture Model (GMM) gives an approximate way to explicitly calculate the *posteriori* density of the states of a stochastic system, even in the nonlinear non-Gaussian case. There are several approaches for time and measurement updates using GMM as shown in the GMM-DO filter of Sondergaard and Lermusiaux [3]. GMM, equipped with Monte-Carlo data-fitting using the Expectation Maximization (EM) algorithm [4] and Bayesian Information Criteria (BIC) [5], provides an accurate estimate of the prior PDF. GMM represents the PDF using a weighted sum of Gaussian PDFs, each of which can be updated using individual Kalman filters if the measurement model is linear and the measurement noise profile is Gaussian.

Data assimilation, such as with atmospheric or oceanic sampling vehicles, often uses a linear (or linearized) observation model, and the measurement noise is assumed to be additive Gaussian [3]. With these assumptions, the GMM along with Kalman filter updates (i.e., GMM-KF) is sufficient to estimate the state. For nonlinear system dynamics, the Kalman update step may be replaced by an Extended Kalman Filter (EKF), which gives the prediction up to first-order precision [6]. The state estimate thus obtained may be used for feedback stabilization. A block diagram of such an observer-based feedback system is shown in Fig. 1. In this model, $\mathbf{x}_n$ and $\mathbf{y}_n$ are the state and observation, respectively at the n^{th} time step. The process is inherently noisy with an additive process-noise $\mathbf{W}_n$ and the observation is corrupted by the measurement-noise $\mathbf{V}_n$. The state-estimate $\hat{\mathbf{x}}_n$ is

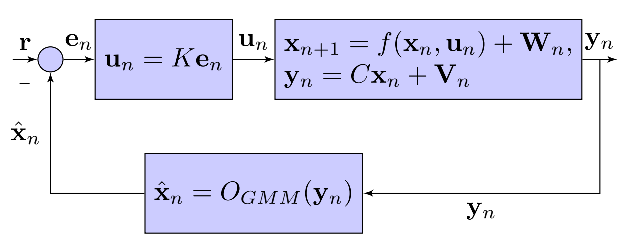


Fig. 1: Block-diagram of a discrete time closed-loop observer-based feedback control system with a Gaussian Mixture Model observer O_{GMM}.

[1]Debdipta Goswami is a graduate student in the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742. goswamid@umd.edu

[2]Derek A. Paley is the Willis H. Young Jr. Associate Professor of Aerospace Engineering Education in the Department of Aerospace Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 dpaley@umd.edu

obtained using GMM-KF observer (denoted as O_{GMM}) and this estimate is used to derive the feedback control signal $\mathbf{u}_n = K(\mathbf{r} - \hat{\mathbf{x}}_n)$.

The contributions of this work are (1) a theoretical guarantee of boundedness of the estimation error while using GMM-KF in a linear system with non-Gaussian noise; (2) ultimate boundedness of the norm of the state in a linear system with observer-based feedback; (3) extension of error bound analysis to a nonlinear system with non-Gaussian process noise; and (4) proof of the ultimate boundedness of the state in a nonlinear system using Lyapunov's method for observer-based feedback with GMM-EKF.

The manuscript is organized as follows. Section II provides a brief overview of Gaussian Mixture Model. Section III explores the analytical as well as numerical results for linear settings. Section IV extends the results to nonlinear systems using GMM-EKF. Section V summarizes the manuscript and discusses possible future work.

II. Gaussian Mixture Model

The GMM-KF works by representing the probability density function (PDF) using a Gaussian Mixture Model (GMM). Let w^j, $j=1,\ldots,M$, be scalar weights such that $\sum_{j=1}^M w^j = 1$. Let $\overline{\mathbf{x}}^j$ and P^j be the mean vector and covariance matrix respectively for a multivariate Gaussian $\mathcal{N}\left(\mathbf{x};\overline{\mathbf{x}}^j,P^j\right)$, $j=1,\ldots,M$. The weighted sum of the M Gaussian densities [3]

$$p_X\left(\mathbf{x};\{w^j,\overline{\mathbf{x}}^j,P^j\}_{j=1}^M\right) = \sum_{j=1}^M w^j \mathcal{N}\left(\mathbf{x};\overline{\mathbf{x}}^j,P^j\right) \tag{1}$$

is a valid probability density function (PDF) that integrates to unity and has an analytical representation. Through the selection of the weights, means, covariances, and number of mixture components, (1) can represent even highly non-Gaussian distributions.

Traditional ensemble/particle-based methods represent a PDF using sparse support of ensemble members (i.e., a Monte Carlo sampling of realizations) [7]. This representation enables nonlinear propagation of the uncertainty in the forecast step of the filter. Unfortunately, many particle filters suffer from degeneracy issues due to the sparsity of the PDF representation [7]. Kernel-based approaches address this issue by periodically creating a density estimate from the ensemble sample to mantain the full support of the state space and to facilitate resampling [7], [8], [9]. Unfortunately, such approaches require the arbitrary choice of fitting parameters such as kernel bandwidth [3]. For Gaussian mixtures, given a specific choice for mixture complexity, an Expectation Maximization algorithm automatically selects the weights, means, and covariances of the Gaussians to best fit the ensemble [10]. A key contribution of [3] is the use of the Bayesian Information Criterion (BIC) for the automatic selection of the mixture complexity. Let $\mathbf{x}_j$ be the j^{th} sample in an ensamble realization. The BIC may be (approximately) expressed as [3]

$$\text{BIC} = -2\sum_{j=1}^N \log p(\mathbf{x}_j|\Omega_{\text{ML}};M) + K\log N, \tag{2}$$

where K is the number of parameters in the model, Ω_{ML} is the maximum likelihood set of parameters (produced by the EM algorithm), and N is the number of ensemble members. For a multivariate Gaussian mixture where d is the dimension of the state vector, $K = M\left(2d+(d(d-1))/2+1\right)$ is the number of free parameters [3]. The BIC has two components: the first component evaluates the goodness-of-fit for the model of complexity M and the second component is a penalty on the overall model complexity [3]. By sequentially evaluating models of increasing complexity, one may identify a local minimum in the BIC. One seeks the best fit of a mixture of Gaussians to the data; the model-complexity term in the BIC ensures that a simpler model is preferred [3].

For many data-assimilation applications, the observation-operator C linearly extracts the measurement from the state vector, i.e.,

$$\mathbf{y} = C\mathbf{x} + \mathbf{V} \quad \text{with} \quad \mathbf{v}\sim\mathcal{N}(0,K_V), \tag{3}$$

where $\mathbf{V}$ is zero-mean measurement noise with covariance K_V. For a (single) Gaussian forecast PDF, Gaussian measurement noise, and a linear observation operator, the Kalman-analysis equations represent the optimal approach to Bayesian assimilation of a measurement. In the case of a mixture of Gaussians, the Kalman-analysis equations may be augmented with a weight-analysis equation to yield the proper application of Bayes' rule for each component in the mixture [3], as follows.

The Bayesian update of a Gaussian mixture prior with Gaussian observation model yields a Gaussian mixture posterior [3]. For a prior GMM

$$p_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^M w^j \mathcal{N}\left(\mathbf{x};\overline{\mathbf{x}}^j,P^j\right),$$

and a Gaussian observation model $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y};C\mathbf{x},R)$, the posterior PDF looks like [3]

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \sum_{j=1}^M \hat{w}^j \mathcal{N}\left(\mathbf{x};\hat{\mathbf{x}}^j,\hat{P}^j\right), \tag{4}$$

where

$$\begin{aligned}
\hat{\mathbf{x}}^j &= \overline{\mathbf{x}}^j + K^j(\mathbf{y} - C\overline{\mathbf{x}}_j),\\
K^j &= P^jC^T(CP^jC^T + K_V)^{-1},\\
\hat{P}^j &= (I - K^jC)P^j, \text{ and}\\
\hat{w}^j &= \frac{w^j\mathcal{N}(\mathbf{y};C\overline{\mathbf{x}}^j,CP^jC^T+K_V)}{\sum_{m=1}^M w^m\mathcal{N}(\mathbf{y};C\overline{\mathbf{x}}^m,CP^mC^T+K_V)}.
\end{aligned} \tag{5}$$

Combining the weighted Gaussians produces the posterior-mean estimate.

III. GMM-KF for Linear Systems

Usually GMM-KF filter works by invoking the Expectation Maximization algorithm at each step, and calculating the unknown parameters. But for the sake of theoretical tractability, and to reduce computational burdens, we are going to formulate a special case, where the GMM is calculated using EM and BIC once from the initial prior, and then updated using Kalman update steps.

First consider a linear Gaussian system

$$\mathbf{x}_{n+1} = A\mathbf{x}_n + \mathbf{W}_n \text{ and } \mathbf{y}_n = C\mathbf{x}_n + \mathbf{V}_n, n \geq 1, \quad (6)$$

where $\mathbf{x}_n \in \mathbb{R}^d$, $\mathbf{y}_n \in \mathbb{R}^q$, $\mathbf{W}_n, \mathbf{V}_n, n \geq 1$ are orthogonal and zero mean i.i.d. Gaussian random vectors with $cov(\mathbf{V}_n) = K_V$ and $cov(\mathbf{W}_n) = K_W$. The best estimate of the state is obtained by Kalman filter: i.e.,

$$\begin{aligned} \overline{\mathbf{x}}_n &= A\hat{\mathbf{x}}_{n-1}, \\ \hat{\mathbf{x}}_n &= \overline{\mathbf{x}}_n + K_n(\mathbf{y}_n - C\overline{\mathbf{x}}_n), \\ K_n &= S_n C^T (C S_n C^T + K_V)^{-1}, \\ S_n &= A\Sigma_{n-1}A^T + K_W, \\ \Sigma_n &= (I - K_n C) S_n, \end{aligned} \quad (7)$$

where $S_n = cov(\mathbf{x}_n - \overline{\mathbf{x}}_n)$ and $\Sigma_n = cov(\mathbf{x}_n - \hat{\mathbf{x}}_n)$. The error $\tilde{\mathbf{x}}_n$ for the Kalman filter is given by $\tilde{\mathbf{x}}_n = (I - K_n C)(A\tilde{\mathbf{x}}_{n-1} + \mathbf{W}_{n-1}) - K_n \mathbf{V}_n$.

A. Error Propagation in Linear GMM-KF

Theorem 1 [11]: Let $K_W = QQ^T$. Suppose that (A, Q) is reachable and (A, C) is observable. If $S_1 = 0$, then $\Sigma_n \to \Sigma, K_n \to K, S_n \to S$ as $n \to \infty$. The limiting matrices are the only solutions of the equations

$$\Sigma = (I - KC)S, \; K = SC^T (CSC^T + K_V)^{-1}, \\ \text{and } S = A\Sigma A^T + K_W.$$

Now we investigate the error dynamics of GMM-KF in the linear non-Gaussian case. In this scenario, the system remains same as Eq. (6), except $\mathbf{W}_n$ is non-Gaussian (may be assumed as zero mean, without loss of generality). The GMM-KF is, for $j = 1, \cdots, M$,

$$\begin{aligned} \overline{\mathbf{x}}_n^j &= A\hat{\mathbf{x}}_{n-1}^j, \\ \hat{\mathbf{x}}_n^j &= \overline{\mathbf{x}}_n^j + K_n^j(\mathbf{y}_n - C\overline{\mathbf{x}}_n^j), \\ K_n^j &= S_n^j C^T (C S_n^j C^T + K_V)^{-1}, \\ S_n^j &= A\Sigma_{n-1}^j A^T + K_W, \\ \Sigma_n^j &= (I - K_n C) S_n^j, \\ w_n^j &= \frac{w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, C S_n^j C^T + K_V)}{\sum_{m=1}^M w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, C S_n^m C^T + K_V)}, \\ \hat{\mathbf{x}}_n &= \sum_{j=1}^M w_n^j \hat{\mathbf{x}}_n^j, \end{aligned} \quad (8)$$

where $\sum_{j=1}^M w_n^j = 1 \; \forall n$, $S_n^j = cov(\mathbf{x}_n - \overline{\mathbf{x}}_n^j)$ and $\Sigma_n^j = cov(\mathbf{x}_n - \hat{\mathbf{x}}_n^j)$. Assume that the initial values of the parameters have been set by the EM algorithm and BIC.

To show that $\|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded, we proceed to the error analysis of this estimator, where the error $\tilde{\mathbf{x}}_n \triangleq \mathbf{x}_n - \hat{\mathbf{x}}_n$ can be written as

$$\tilde{\mathbf{x}}_n = \sum_{j=1}^M w_n^j \tilde{\mathbf{x}}_n^j = \sum_{j=1}^M w_n^j (\mathbf{x}_n - \hat{\mathbf{x}}_n^j). \quad (9)$$

Theorem 2: Consider the linear system (6) with possibly non-Gaussian noise $\mathbf{W}_n$. If the conditions of Theorem 1 are satisfied, the matrix C has full column-rank and the norm of the measurement noise $\|\mathbf{V}_n\|$ is bounded with probability P, then the norm of the error $\|\tilde{\mathbf{x}}_n\|$ is bounded with probability P for all n.

Proof: From Theorem 1, $\Sigma^j = (I - KC)S^j, K^j = S^j C^T (C S^j C^T + K_V)^{-1}, S^j = A\Sigma^j A^T + K_W, \forall j = 1, \cdots, M$ are the bounded limits of Σ_n^j, K_n^j and S_n^j as $n \to \infty$. Let $R_n^j = C S_n^j C^T + K_V$, d be the dimensionality of $\mathbf{x}_n$ and $\beta_n = \sum_{m=1}^M w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, C S_n^m C^T + K_V)$ be the finite normalization factor. From the Kalman update equation (8), $\tilde{\mathbf{x}}_n^j = \left(I - K_n^j C\right) \tilde{r}_n^j - K_n^j \mathbf{V}_n$, where $\tilde{r}_n^j \triangleq A\tilde{\mathbf{x}}_{n-1}^j + \mathbf{W}_{n-1}$. Using this result along with Eq. (8), Eq. (9) can be expanded as

$$\begin{aligned} \tilde{\mathbf{x}}_n &= \sum_{j=1}^M w_n^j \tilde{\mathbf{x}}_n^j \\ &= \frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, C S_n^j C^T + K_V) \tilde{\mathbf{x}}_n^j \\ &= \frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \\ &\quad \times \mathcal{N}(C(A\mathbf{x}_{n-1} + \mathbf{W}_{n-1}) + \mathbf{V}_n; CA\hat{\mathbf{x}}_{n-1}^j, R_n^j) \tilde{\mathbf{x}}_n^j \\ &= \frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \frac{(\sqrt{2\pi})^{-d}}{\sqrt{det(R_n^j)}} \\ &\quad \times e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1} \left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \\ &\quad \times \left(\left(I - K_n^j C\right) \tilde{r}_n^j - K_n^j \mathbf{V}_n\right). \\ &= \left[\frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \frac{(\sqrt{2\pi})^{-d}}{\sqrt{det(R_n^j)}} \right. \\ &\quad \left. \times e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1} \left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \tilde{r}_n^j \right] \\ &\quad - \left[\frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \frac{(\sqrt{2\pi})^{-d}}{\sqrt{det(R_n^j)}} \right. \\ &\quad \times e^{-\frac{1}{2}\left(C\tilde{r}_n^j + \mathbf{V}_n\right)^T (R_n^j)^{-1} \left(C\tilde{r}_n^j + \mathbf{V}_n\right)} \\ &\quad \left. \times K_n^j \left(C\tilde{r}_n^j + \mathbf{V}_n\right) \right]. \end{aligned} \quad (10)$$

From Theorem 1, $K_n^j \to K^j$ and $R_n^j \to R^j$ as $n \to \infty$ with finite-norm limit and $(R_n^j)^{-1}$ is positive definite. Because the norm of $e^{-\frac{1}{2}\mathbf{x}^T Q \mathbf{x}} x$ is bounded for any positive definite Q (see Appendix I), then the second term of the Eq. (10) is bounded. For the first term, let $q_n^j = C\tilde{r}_n^j + \mathbf{V}_n$. Since C has full column-rank, $\tilde{r}_n^j = (C^T C)^{-1} C^T (q_n^j - \mathbf{V}_n)$. Putting this in the first term of Eq. (10) we get

$$\frac{1}{\beta_n} \sum_{j=1}^M w_{n-1}^j \frac{(\sqrt{2\pi})^{-d}}{\sqrt{det(R_n^j)}}$$

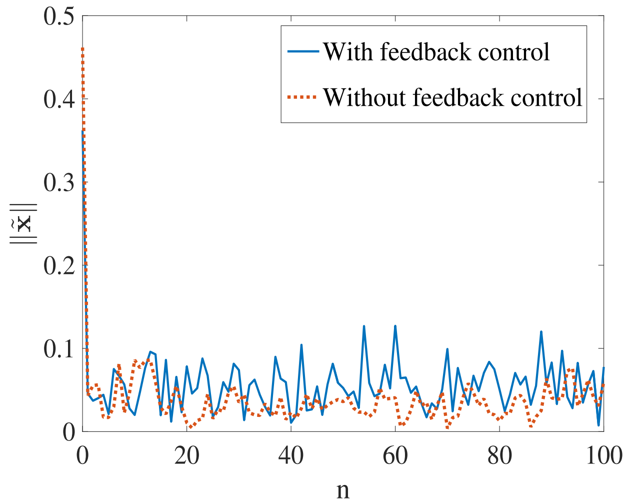


Fig. 2: L_2 error of GMM-KF estimate with and without feedback.

$$\times e^{-\frac{1}{2}(q_n^j)^T (R_n^j)^{-1} q_n^j} (C^T C)^{-1} C^T (q_n^j - \mathbf{V}_n)$$

which is bounded (see Appendix 1) with probability P if $\mathbf{V}_n$ is bounded with the same probability. ■

B. Observer-Based Feedback Control

To make use of the GMM-KF estimate in feedback control, consider a system

$$\mathbf{x}_{n+1} = A\mathbf{x}_n + B\mathbf{u}_n + \mathbf{W}_n.$$

where $\mathbf{W}_n$ is a zero mean possibly non-Gaussian process. With GMM-KF estimation and feedback, $\mathbf{u}_n = -K\hat{\mathbf{x}}_n$, K is such chosen that $|\lambda(A-BK)|_{max} < 1$, where $|\lambda|_{max}$ stands for maximum magnitude of the eigenvalues and, using the desired state $\mathbf{x}_{des} = 0$, we get

$$\begin{aligned} \mathbf{x}_{n+1} &= A\mathbf{x}_n - BK\hat{\mathbf{x}}_n + \mathbf{W}_n \\ &= (A - BK)\mathbf{x}_n + BK(\mathbf{x}_n - \hat{\mathbf{x}}_n) + \mathbf{W}_n \\ &\vdots \\ &= (A - BK)^n \mathbf{x}_0 \\ &\quad + (A - BK)^{n-1}(BK(\mathbf{x}_1 - \hat{\mathbf{x}}_1) + \mathbf{W}_0) \\ &\quad + \cdots + BK(\mathbf{x}_n - \hat{\mathbf{x}}_n) + \mathbf{W}_n \end{aligned} \tag{11}$$

Taking the expectation $\mathbb{E}$ of $\|\mathbf{x}_n - \mathbf{x}_{des}\| = \|\mathbf{x}_n\|$ and using the triangle inequality yields

$$\begin{aligned} \|\mathbf{x}_{n+1}\| &\le |\lambda|_{max}^n \|\mathbf{x}_0\| \\ &\quad + |\lambda|_{max}^{n-1} \|BK\| (\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\| + \|\mathbf{W}_0\|) \\ &\quad + \cdots + \|BK\| \|\mathbf{x}_n - \hat{\mathbf{x}}_n\| + \|\mathbf{W}_n\| \\ \Rightarrow \mathbb{E}\|\mathbf{x}_{n+1}\| &\le |\lambda|_{max}^n \|\mathbf{x}_0\| + |\lambda|_{max}^{n-1} \|BK\| \mathbb{E}\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\| \\ &\quad + \cdots + \|BK\| \mathbb{E}(\|\mathbf{x}_n - \hat{\mathbf{x}}_n\|). \end{aligned} \tag{12}$$

Since $|\lambda(A-BK)|_{max} < 1$, $\mathbb{E}\|\mathbf{x}_{n+1}\|$ remains bounded as $n \to \infty$ if $\mathbb{E}\|\mathbf{x}_n - \hat{\mathbf{x}}_n\|$ is bounded, which is indeed true by Theorem 2 with probability P depending on the measurement noise.

The validity of the boundedness of error and the practical stability [12] of observer-based feedback is illustrated by numerical simulation. The L_2 norm of the error for a linear system is given in Fig. 2. The effectiveness of the observer-based feedback is shown in Fig. 3.

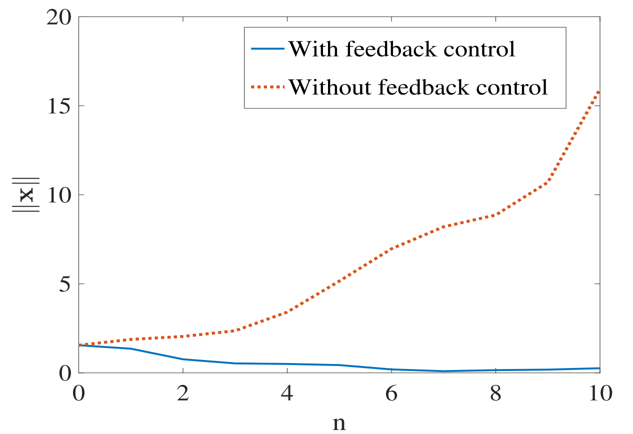


Fig. 3: Observer-based feedback effectively stabilizes the system.

Here we have used the system

$$\mathbf{x}_{n+1} = \begin{bmatrix} 1.0 & 0.9 \\ -0.5 & 1.2 \end{bmatrix} \mathbf{x}_n + \begin{bmatrix} 1.0 & 0 \\ 0 & 1.0 \end{bmatrix} \mathbf{u}_n + \mathbf{W}_n,$$

where $\mathbf{W}_n$ is a non-Gaussian zero mean i.i.d. random vector. This system is unstable and $\|\mathbf{x}_n\|$ increases exponentially with $\mathbf{u}_n = 0$ for all n. For feedback control, we use $\mathbf{u}_n = -\begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix} \hat{\mathbf{x}}_n$, so that the eigenvalues of $A - BK$ lie within the unit circle. The observer-based control effectively stabilizes the system as shown in Fig. 3.

IV. Extension to Nonlinear Systems

To extend the GMM-KF setting to nonlinear systems, we investigate the widely used filtering framework, the Extended Kalman Filter (EKF) [6]. We apply GMM to a nonlinear system with linear measurement model, i.e.,

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n) + \mathbf{W}_n \text{ and } \mathbf{y}_n = C\mathbf{x}_n + \mathbf{V}_n, n \ge 1, \tag{13}$$

where $\mathbf{W}_n$ is possibly non-Gaussian additive noise and $\mathbf{V}_n \sim \mathcal{N}(\mathbf{0}, K_V)$. The GMM framework in this case is, for $j = 1, \cdots, M$,

$$\begin{aligned} \overline{\mathbf{x}}_n^j &= f(\hat{\mathbf{x}}_{n-1}^j), \\ \hat{\mathbf{x}}_n^j &= \overline{\mathbf{x}}_n^j + K_n^j(\mathbf{y}_n - C\overline{\mathbf{x}}_n^j), \\ K_n^j &= S_n^j C^T (C S_n^j C^T + K_V)^{-1}, \\ S_n^j &= A(\hat{\mathbf{x}}_{n-1}) \Sigma_{n-1}^j A(\hat{\mathbf{x}}_{n-1})^T + K_W, \\ \Sigma_n^j &= (I - K_n C) S_n^j, \\ w_n^j &= \frac{w_{n-1}^j \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^j, C S_n^j C^T + K_V)}{\sum_{m=1}^M w_{n-1}^m \mathcal{N}(\mathbf{y}_n; C\overline{\mathbf{x}}_n^m, C S_n^m C^T + K_V)}, \\ \hat{\mathbf{x}}_n &= \sum_{j=1}^M w_n^j \hat{\mathbf{x}}_n^j, \end{aligned} \tag{14}$$

where $\sum_{j=1}^{M} w_n^j = 1 \; \forall n$, $S_n^j = cov(\mathbf{x}_n - \overline{\mathbf{x}}_n^j)$, $\Sigma_n^j = cov(\mathbf{x}_n - \hat{\mathbf{x}}_n^j)$ and $A(\mathbf{x}) = J_f(\mathbf{x})$ is the Jacobian of f evaluated at $\mathbf{x}$. The initial values of the parameters have been set by the Expectation Maximization algorithm and BIC as in the linear case.

A. Error Propagation

The error dynamics are

$$\begin{aligned} \tilde{\mathbf{x}}_n &= \sum_{j=1}^{M} w_n^j \tilde{\mathbf{x}}_n^j \\ &= \frac{1}{\beta_n} \sum_{j=1}^{M} w_{n-1}^j \frac{(\sqrt{2\pi})^{-d}}{\sqrt{det(R_n^j)}} \\ &\quad \times e^{-\frac{1}{2}(C\tilde{\mathbf{f}}_n^j + \mathbf{V}_n)^T (R_n^j)^{-1} (C\tilde{\mathbf{f}}_n^j + \mathbf{V}_n)} \\ &\quad \times \left((I - K_n^j C)\tilde{\mathbf{f}}_n^j - K_n^j \mathbf{V}_n \right), \end{aligned} \tag{15}$$

where $\tilde{\mathbf{f}}_n^j = f(\mathbf{x}_{n-1}) - f(\hat{\mathbf{x}}_{n-1}^j) + \mathbf{W}_{n-1}$ and $R_n^j = CS_n^j C^T + K_V$. To have $\|\tilde{\mathbf{x}}_n\|$ bounded in the nonlinear case, we need stricter assumptions than before because R_n^j and S_n^j no longer converge to finite-norm matrices when $n \to \infty$.

Assumption 1:
$\alpha_j I \leq \Sigma_n^j < \beta_j I$, $\alpha_j, \beta_j \geq 0$ for all $n \geq 0$ and $j = 1, \cdots, M$.

Under Assumption 1, S_n^j and R_n^j, being the positive-definiteness-preserving bilinear transformations of Σ_{n-1}^j and S_n^j (from Eq. 14), respectively, are always bounded above and below by positive definite matrices. Hence $K_n^j C = S_n^j C^T (R_n^j)^{-1} C$ is also positive definite and bounded above and below. Now proceeding in the exact same way as the proof of the Theorem 2, $\tilde{\mathbf{x}}_n$ will also be bounded with probability P if C has full column rank and $\mathbf{V}_n$ is bounded with probability P. Let the bound on $\tilde{\mathbf{x}}_n$ be $b_{\tilde{x}}(P)$.

B. Observer-Based Feedback Control

To control the nonlinear system (13), suppose we design a state-feedback controller $u(\mathbf{x})$, and drive it with the estimated state derived by the filter. The closed loop system looks like

$$\begin{aligned} \mathbf{x}_{n+1} &= f(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) \\ &= F(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{V}_n, \end{aligned} \tag{16}$$

where $\mathbf{W}_n$ is the non-Gaussian additive noise and $F(\mathbf{x}_n) = f(\mathbf{x}_n) + u(\mathbf{x}_n)$.

Assumption 2:
2.1: The nominal system $\mathbf{x}_{n+1} = F(\mathbf{x}_n)$ is uniformly asymptotically stable on a open ball $\mathcal{B}_r$ of radius r centered at 0 and $\exists$ a C^1 (i.e., differentiable) Lyapunov function $V : \mathbb{Z}_+ \times \mathcal{B}_r \to \mathbb{R}$ that satisfies

$$\alpha_1(\|\mathbf{x}_n\|) \leq V(n, \mathbf{x}_n) \leq \alpha_2(\|\mathbf{x}_n\|), \tag{17}$$

$$\Delta V_N(n, \mathbf{x}_n) \leq -\alpha_3(\|\mathbf{x}_n\|), \tag{18}$$

where $\alpha_i, i = 1, 2, 3$ are class $\mathcal{K}$ functions and $\Delta V_N(n, \mathbf{x}_n) = V(n+1, \mathbf{x}_{n+1}) - V(n, \mathbf{x}_n)$ under the nominal system [12], [13].

2.2: $\exists p \in \mathbb{R}_+$ and $M > 0$ such that $\left\| \frac{\partial u}{\partial \mathbf{x}} \right\| \leq M \|\mathbf{x}\|^{p-1}$.
With the application of the mean value inequality, we get $\|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n)\| \leq M b_{\tilde{x}}(P)$ from Assumption 1.

2.3: $\|\mathbf{W}_n\| < b_W(P)$ with probability P.
Assumption 2.3 characterizes the process noise, and $b_W(P)$ can be readily obtained from the PDF or cumulative disctribution of the process noise (only the latter if $\mathbf{W}_n$ is not absolutely continuous).

Theorem 3:
Consider the non-linear system (16) with non-Gaussian process noise and GMM-EKF state estimate $\hat{\mathbf{x}}_n$. If Assumptions 1–2 are satisfied, then $\|\mathbf{x}_n\|$ is bounded with an ultimate bound b_x with probability P, where b_x is a function of M, $b_{\tilde{x}}(P)$, and $b_W(P)$.

Proof: We use $V(n, \mathbf{x}_n)$ from Assumption 2 for the perturbed system. Thus,

$$\Delta V_P(n, \mathbf{x}_n) = \Delta V_N(n, \mathbf{x}_n) + \delta V_P(n, \mathbf{x}_n),$$

where

$$\begin{aligned} \delta V_P(n, \mathbf{x}_n) = V(n+1, F(\mathbf{x}_n) + u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n) \\ - V(n+1, F(\mathbf{x}_n)), \end{aligned}$$

and subscript P stands for the perturbed system. Since $V \in C^1$, $\exists \; l > 0$ such that

$$\begin{aligned} |\delta V_P(n, \mathbf{x}_n)| &\leq l \|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n) + \mathbf{W}_n\| \\ &\leq l(\|u(\hat{\mathbf{x}}_n) - u(\mathbf{x}_n)\| + \|\mathbf{W}_n\|). \end{aligned}$$

Now, from Assumptions 1–2,

$$\begin{aligned} \Delta V_P(n, \mathbf{x}_n) &= \Delta V_N(n, \mathbf{x}_n) + \delta V_P(n, \mathbf{x}_n) \\ &\leq -\alpha_3(\|\mathbf{x}_n\|) + lM(b_{\tilde{x}}(P) + b_W(P)), \text{ wp } P \\ &= -(1-\theta)\alpha_3(\|\mathbf{x}_n\|) - \theta\alpha_3(\|\mathbf{x}_n\|) \\ &\quad + lM(b_{\tilde{x}}(P) + b_W(P)), \\ &\quad \text{wp } P \text{ and } \theta \in (0,1) \\ &\leq -(1-\theta)\alpha_3(\|\mathbf{x}_n\|), \\ &\quad \forall \; \|\mathbf{x}_n\| \geq \alpha_3^{-1}\left(\frac{lM(b_{\tilde{x}}(P) + b_W(P))}{\theta} \right), \\ &\quad \text{wp } P. \end{aligned}$$

Hence the $\|\mathbf{x}_n\|$ for the system (16) is u.u.b, with ultimate bound

$$b_x(P) = \alpha_1^{-1} \circ \alpha_2 \circ \alpha_3^{-1}\left(\frac{lM(b_{\tilde{x}}(P) + b_W(P))}{\theta} \right), \text{ wp } P. \; \blacksquare \tag{19}$$

C. Simulation Results

We tested the GMM-EKF on a 2D nonlinear chaotic map attributed to the discrete-time version of the Duffing

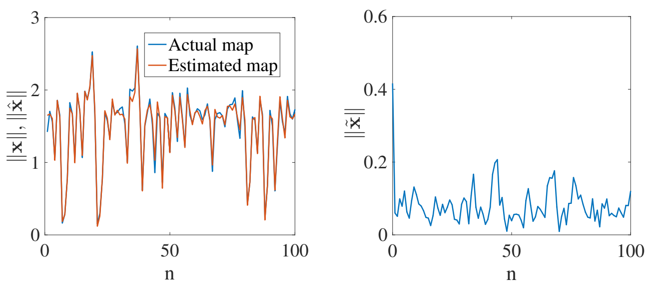
Oscillator [14]. The map is

$$\begin{aligned} x_{n+1}^1 &= x_n^2 \\ x_{n+1}^2 &= -bx_n^1 + ax_n^2 - (x_n^2)^3 + w_n. \end{aligned} \tag{20}$$

The map depends on the two constants a and b, usually set to $a = 2.75$ and $b = 0.2$ to produce chaotic behaviour [14]. We added zero-mean non-Gaussian i.i.d. random variable w_n as the process noise. The resultant estimated map and the error is shown in Fig. 4. The effectiveness of GMM-EKF in feedback control is illustrated in the Fig. 5 with the system

$$\begin{aligned} x_{n+1}^1 &= x_n^2 \\ x_{n+1}^2 &= -1.1x_n^1 x_n^2 + w_n + u_n, \end{aligned} \tag{21}$$

where w_n is non-Gaussian i.i.d. random variable. To stabilize the system, the feedback control is given by $u_n = -0.3\hat{x}_n^1 \hat{x}_n^2$. The observer-based feedback effectively stabilizes the system as shown in Fig. 5.



(a) (b)

Fig. 4: GMM-EKF estimate of Duffing map: (a) L_2 norm of the state, (b) L_2 error.

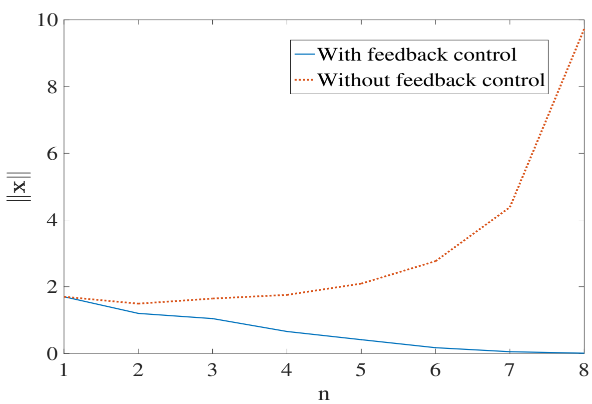


Fig. 5: Observer-based feedback (with GMM-EKF) effectively stabilizes the system.

V. CONCLUSIONS

This paper presents the Gaussian Mixture Model Kalman Filter (GMM-KF) as an effective tool to deal with non-Gaussian process noise for observer based feedback. The boundedness of estimation error is proven analytically and supported by simulation results. The framework is extended to nonlinear settings using the Extended Kalman Filter (EKF) in conjunction with GMM. GMM-KF is a dynamic Bayesian filtering technique that can be modified to less computational complexity, but works well with data-driven systems. The theoretical justification of the performance of GMM-KF is shown here. In ongoing work, we seek to extend this framework to nonlinear observations, as well as to a general Bayesian filtering technique.

APPENDIX I

Proposition: $\left\|\mathbf{x}\exp(-\mathbf{x}^T Q\mathbf{x})\right\|$ with $Q > 0$ is bounded.

Proof:

$$\begin{aligned} \left\|\mathbf{x}\exp(-\mathbf{x}^T Q\mathbf{x})\right\| &\leq \|\mathbf{x}\| \left|\exp(-\mathbf{x}^T Q\mathbf{x})\right| \\ &\leq \|\mathbf{x}\| \left|\exp(-\lambda_{min}(Q)\|\mathbf{x}\|^2)\right|, \\ &\quad \text{since } Q > 0, \lambda_{max} \geq \lambda_{min}(Q) > 0, \\ &\leq \frac{1}{\sqrt{2\lambda_{min}(Q)}}\sqrt{e}, \end{aligned} \tag{22}$$

from single-variable calculus.

ACKNOWLEDGMENTS

The authors thank Frank Lagor and Dipankar Maity for valuable discussions pertaining to GMM and BIC.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[2] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[3] T. Sondergaard and P. F. J. Lermusiaux, "Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme," *Mon. Weather Rev.*, vol. Volume 141, pp. 1737–1760, 2013.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[5] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 03 1978.

[6] M. I. Ribeiro, "Kalman and Extended Kalman Filters: concept, derivation and properties," *Technical Report*, 2004. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.5088

[7] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Proc.*, vol. 50, no. 2, pp. 174–188, 2002.

[8] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[9] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[10] C. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.

[11] J. Walrand, "Kalman filter: Convergence," *UC Berkley, EE 226A Lecture Note*. [Online]. Available: http://robotics.eecs.berkeley.edu/~wlr/226F06/226a-ch11.pdf

[12] H. K. Khalil, *Nonlinear systems*. Upper Saddle River, (N.J.): Prentice Hall, 1996.

[13] C. Cruz-Hernández, J. Alvarez-Gallegos, and R. Castro-Linjares, "Stability of discrete nonlinear systems under novanishing pertuabations: application to a nonlinear model-matching problem," *IMA Journal of Mathematical Control and Information*, vol. 16, no. 1, pp. 23–41, 1999.

[14] T. Kapitaniak, *Chaotic Oscillators: Theory and Applications*. World Scientific, 1992.