OPTIMAL RECURSIVE FILTERING USING GAUSSIAN MIXTURE MODEL

Igal Bilik and Joseph Tabrikian

Dept. of Electrical and Computer Engineering
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
E-mail: {bilik, joseph}@ee.bgu.ac.il

ABSTRACT

Kalman filter is an optimal recursive estimator of the system state in terms of minimum-mean-square error (MMSE) under linear Gaussian assumptions. The Gaussianity assumption is not satisfied in many applications, such as dynamic channel estimation in mobile communications, maneuvering target tracking and speech enhancement. In this paper, the MMSE estimator for linear, non-Gaussian problems is presented, where the Gaussian mixture model is used for non-Gaussian distributions. The resulting recursive algorithm, named as non-Gaussian Kalman filter (NGKF), is composed of several conventional Kalman filters combined in an optimal manner. The performance of the proposed NGKF, is compared to the Kalman and particle filters via simulations. It is shown that the proposed NGKF outperforms both the Kalman and particle filters.

Keywords – Kalman filter, particle filter, GMM, MMSE, sequential Bayesian estimation, dynamic state space.

1. INTRODUCTION

In many applications, time-varying system state estimation from a sequence of noisy observations is required [1]. Most of these problems can be analyzed using dynamic state-space (DSS) models for which the Bayesian approach can be efficiently applied. For implementation of the Bayesian approach the posterior probability density function (pdf) of the state is required. The optimal estimator in the sense of minimum-mean-square error (MMSE), can be found using the posterior distribution only. For most applications the state estimation is required at each time instance, leading to recursive algorithms. The optimal recursive estimator is usually tractable for linear, Gaussian systems [2]. However, the Gaussianity assumption is very limiting and extension of the Kalman filter (KF) to non-Gaussian models has been investigated in several works. For example, the unscented KF (UKF) uses a deterministic sample-based approximation to estimate the effect of non-Gaussian models [3]. The Gaussian sum filter (GSF) is a sub-optimal technique that approximates the state posterior pdf by a weighted sum of Gaussians, assuming Gaussian system noise [4], [5]. As shown in [4], GSF is implemented using a bank of KFs, for the mixture components. The main disadvantage of this method is the exponential growth of the number of the mixture components.

Recently, a new class of filtering methods was proposed based on the sequential Monte Carlo (MC) approach. The sequential importance sampling algorithm forms the base for most MC approaches [6]. Sequential MC techniques achieve the filtering task by recursively generating MC samples of the state variables. These methods are often very flexible in non-Gaussian problems due to the nature of the MC simulations [7]. Specifically, particle filters (PF) approximate the posterior distribution by a set of random samples with associated weights. The PF attempts to model the posterior distribution, using discrete random samples, rather than using an analytic model [8]. The Gaussian sum particle filter (GSPF), assuming Gaussian mixture distributions was introduced in [9]. This approach is sub-optimal, and its optimality is achieved only asymptotically when the number of particles (random samples) goes to infinity.

This work concerns with non-Gaussian linear DSS model. Thus, pdfs of the system initial state, system noise and posterior distributions are effectively modeled by GMM. Using the idea that almost any pdf can be effectively approximated by a mixture with a finite number of Gaussians [10], the non-Gaussian Kalman filter (NGKF) is developed. The use of Gaussian components in the GMM approximation allows effective implementation of the proposed NGKF. It can be shown that NGKF is the generalization of the original KF for the non-Gaussian case. It is proved that the proposed NGKF is optimal with respect to MMSE criterion. Additionally, the problem of exponential growth of the number of mixture components, discussed in [4], is solved using the expectation-maximization (EM) algorithm. The proposed NGKF algorithm differs from all algorithms developed after Kalman (for example [4]-[8]) in its optimality with respect to the MMSE criterion. The computational complexity of the NGKF is higher than the KF because of the necessity to estimate additional statistics of the state vector. Potential applications of the proposed NGKF include, non-Gaussian dynamic channel estimation in mobile communications, maneuvering target tracking, pitch tracking and speech enhancement.

2. PROBLEM FORMULATION

Consider a state sequence $\{\mathbf{s}[n],\ n = 0, 1, 2, \ldots\}$ whose time evolution is described by a non-Gaussian linear model, and observation equation as

$$\mathbf{s}[n] = \mathbf{A}[n]\mathbf{s}[n-1] + \mathbf{u}[n] \quad (1)$$

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n], \qquad n = 0, 1, 2, \ldots \quad (2)$$

where the transition matrices, $\mathbf{A}[n]$, and the observation matrices, $\mathbf{H}[n]$, are known. The i.i.d. random processes $\mathbf{u}[n]$ and $\mathbf{w}[n]$ denote the non-Gaussian driving and the measurement noise, respectively. These processes and the initial state, $\mathbf{s}[-1]$ are independent GMM-distributed:

$$\begin{aligned}\mathbf{s}[-1] &\sim GMM(\alpha_{\mathbf{s}j}[-1], \boldsymbol{\mu}_{\mathbf{s}j}[-1], \boldsymbol{\Gamma}_{\mathbf{s}j}[-1]; j=1,\ldots,N)\\ \mathbf{u}[n] &\sim GMM(\alpha_{\mathbf{u}k}[n], \boldsymbol{\mu}_{\mathbf{u}k}[n], \boldsymbol{\Gamma}_{\mathbf{u}k}[n]; k=1,\ldots,K)\\ \mathbf{w}[n] &\sim \mathcal{N}^c(\mathbf{0}, \boldsymbol{\Gamma}_{\mathbf{w}}[n]). \end{aligned} \quad (3)$$

where $GMM(\alpha_j, \boldsymbol{\mu}_j, \boldsymbol{\Gamma}_j, j=1,\ldots,N)$ denotes an N order Gaussian mixture distribution and can be expressed as

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \Phi_j(\mathbf{x}) \ , \quad (4)$$

where $\Phi_j(\mathbf{x})$ denotes a Gaussian pdf with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Gamma}_j$.

An additional possible generative model for an N-order GMM distributed random vector $\mathbf{x}$, consists of a hidden random vector $\boldsymbol{\eta}[n] \stackrel{\triangle}{=} [\eta_1[n], \ldots, \eta_N[n]]$, which indicates on the generating Gaussian. The pdf of $\boldsymbol{\eta}[n]$ is given by

$$f_{\boldsymbol{\eta}}(\boldsymbol{\eta}[n]) = \sum_{j=1}^{N} \alpha_j[n] \delta(\eta_j[n]-1) \ ,$$

where $\delta(\cdot)$ denotes the dirac delta function, and

$$\eta_j = \begin{cases} 1, & \text{if } \mathbf{x} \text{ is generated by the } j\text{th Gaussian} \\ 0, & \text{otherwise} \end{cases} . \quad (5)$$

The objective of this paper is to develop a recursive method for estimation of $\mathbf{s}[n]$ from the observed data

$$\mathcal{X}[n] = \{\mathbf{x}[0], \ldots, \mathbf{x}[n]\} \ .$$

To this end, we employ the MMSE criterion resulting in the conditional expectation estimator

$$\hat{\mathbf{s}}[n|m] = E(\mathbf{s}[n]|\mathcal{X}[m]) \ . \quad (6)$$

3. NGKF DERIVATION

Let the distribution of the data $\mathcal{X}[m]$ and the hidden states $\mathbf{s}[n]$ be jointly GMM of order N:

$$\begin{bmatrix} \mathbf{s}[n] \\ \mathcal{X}[m] \end{bmatrix} \sim GMM\left(\alpha_{\mathbf{s}\mathcal{X}j}[n], \boldsymbol{\mu}_{\mathbf{s}\mathcal{X}j}[n], \boldsymbol{\Gamma}_{\mathbf{s}\mathcal{X}j}[n]; j=1,\ldots,N\right),$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}\mathcal{X}j}[n] &= \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{s}j}[n] \\ \boldsymbol{\mu}_{\mathcal{X}j}[n] \end{bmatrix} \\ \boldsymbol{\Gamma}_{\mathbf{s}\mathcal{X}j}[n] &= \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{s}\mathbf{s}j}[n] & \boldsymbol{\Gamma}_{\mathbf{s}\mathcal{X}j}[n] \\ \boldsymbol{\Gamma}_{\mathcal{X}\mathbf{s}j}[n] & \boldsymbol{\Gamma}_{\mathcal{X}\mathcal{X}j}[n] \end{bmatrix} . \end{aligned}$$

The generative model for this process can be stated via the hidden random process $\{\eta_j[n]\}_{j=1}^N$, which indicates on the Gaussian from which the vector $\begin{bmatrix} \mathbf{s}[n] \\ \mathcal{X}[m] \end{bmatrix}$ is generated. In [11] it is shown that Eq. (6) can be rewritten as

$$\hat{\mathbf{s}}[n|m] = \sum_{j=1}^{N} P(\eta_j[n]|\mathcal{X}[m]) \boldsymbol{\mu}_{\mathbf{s}}[n|m, \eta_j[n]] \ , \quad (7)$$

where $P(\eta_j[n]|\mathcal{X}[m])$ is the conditional probability that the data $\mathcal{X}[m]$ are generated by the mixture component $\eta_j[n]$. In filtering we are interested in the case of $m=n$. The expectation in (6) can be calculated by noting that

$$E(\cdot|\mathcal{X}[n]) = E(\cdot|\mathcal{X}[n-1], \mathbf{x}[n]) = E(\cdot|\mathcal{X}[n-1], \tilde{\mathbf{x}}[n]) \ , \quad (8)$$

where $\tilde{\mathbf{x}}[n]$ denotes the innovation process

$$\tilde{\mathbf{x}}[n] \stackrel{\triangle}{=} \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] \ , \quad (9)$$

in which $\hat{\mathbf{x}}[n|n-1]$ is the conditional mean estimate of $\mathbf{x}[n]$ from $\mathcal{X}[n-1]$. Eq. (8) is satisfied if the transformation $\mathcal{X}[n] \leftrightarrow [\mathcal{X}^T[n-1], \tilde{\mathbf{x}}^T[n]]^T$ is a one-to-one mapping. Thus the following is obtained:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n, \eta_j[n]] &= E(\mathbf{s}[n]|\mathcal{X}[n-1], \mathbf{x}[n], \eta_j[n]) \\ &= E(\mathbf{s}[n]|\mathcal{X}[n-1], \tilde{\mathbf{x}}[n], \eta_j[n]) \ , \end{aligned} \quad (10)$$

where $\mathcal{X}[n-1]$ and $\tilde{\mathbf{x}}[n]$ are jointly GMM and orthogonal, therefore, they are independent, given any particular random mixture indicator $\eta_j[n]$.

In [11] it is proved that the mean estimation of the jth component of the mixture, at time instance n, updated with the innovation $\tilde{\mathbf{x}}[n]$ is:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n, \eta_j[n]] &= \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]] + E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n], \eta_j[n]) \\ &- \boldsymbol{\mu}_{\mathbf{s}j}[n] \ . \end{aligned} \quad (11)$$

In the prediction process, the GMM of $\mathbf{s}[n-1]$ with N components is mixed with GMM of $\mathbf{u}[n]$ with K components. The resulting GMM has $L = N \times K$ mixture components. Using the generative model of GMM, the prediction can be represented as the following mapping:

$$\begin{aligned}&\{\eta_j[n-1], \eta_{\mathbf{u}k}[n]\} \mapsto \tilde{\eta}_l[n] \\ &\qquad \forall j=1,\ldots,N; l=1,\ldots,L; k=1,\ldots,K \ , \end{aligned}$$

where the notation $\tilde{\eta}_l[n]$ is used for the random indicator in the resulting mixture with $L = N \times K$ components. Using this notation, the conditional mean $\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]]$ is interpreted as the estimate of $\mathbf{s}[n]$ from $\mathcal{X}[n-1]$ given the random mixture indicator $\eta_j[n]$. This term is evaluated as follows:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_l[n]] &= E(\mathbf{A}[n]\mathbf{s}[n-1] + \mathbf{u}[n]|\mathcal{X}[n-1], \tilde{\eta}_l[n]) \\ &= \mathbf{A}[n]\boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1, \eta_j[n-1]] \\ &+ \boldsymbol{\mu}_{\mathbf{u}k}[n] \end{aligned} \quad (12)$$

To calculate the covariance matrix of the lth component in the new mixture at the time instance n, the following expression is required

$$\begin{aligned}\varsigma_{\mathbf{s}j}[n] &= \mathbf{s}[n] - \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_l[n]] \\ &= \mathbf{A}[n]\mathbf{s}[n-1] + \mathbf{u}[n] - \mathbf{A}[n]\boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1, \eta_j[n-1]] \\ &- \boldsymbol{\mu}_{\mathbf{u}k}[n] + \mathbf{A}[n]\varsigma_{\mathbf{s}j}[n-1] + \varsigma_{\mathbf{u}}[n] \ , \end{aligned} \quad (13)$$

where $\varsigma_{\mathbf{u}}[n] = \mathbf{u}[n] - \boldsymbol{\mu}_{\mathbf{u}k}[n]$. The prediction covariance matrix for the lth mixture component is obtained using the last expression as follows:

$$\begin{aligned}
&\mathbf{M}[n|n-1,\tilde{\eta}_l[n]] = E(\varsigma[n]\varsigma^H[n]|\tilde{\eta}_l[n]) \\
&= \mathbf{A}[n]E(\varsigma_{\mathbf{s}j}[n-1]\varsigma^H_{\mathbf{s}j}[n-1]|\tilde{\eta}_l[n])\mathbf{A}^H[n] \\
&+\mathbf{A}[n]E(\varsigma_{\mathbf{s}j}[n-1]\varsigma^H_{\mathbf{u}}[n]\tilde{\eta}_l[n]) \\
&+E(\varsigma_{\mathbf{u}}[n]\varsigma^H_{\mathbf{s}j}[n-1]|\tilde{\eta}_l[n])\mathbf{A}^H[n] + E(\varsigma_{\mathbf{u}}[n]\varsigma^H_{\mathbf{u}}[n]|\tilde{\eta}_l[n]) \\
&= \mathbf{A}[n]\mathbf{M}[n-1|n-1,\eta_j[n-1]]\mathbf{A}^H[n] + \boldsymbol{\Gamma}_{\mathbf{u}k}[n],
\end{aligned} \tag{14}$$

where for last equality the following result is used

$$\begin{aligned}
&E(\varsigma_{\mathbf{u}}[n]\varsigma^H_{\mathbf{s}j}[n-1]|\tilde{\eta}_l[n]) = E((\mathbf{u}[n] - \boldsymbol{\mu}_{\mathbf{u}k}[n]|\tilde{\eta}_l[n])\cdot \\
&\cdot(\mathbf{s}[n-1] - \boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1,\eta_j[n-1]])^H|\tilde{\eta}_l[n]) \\
&= (\boldsymbol{\mu}_{\mathbf{u}k}[n] - \boldsymbol{\mu}_{\mathbf{u}k}[n]|\tilde{\eta}_l[n])\cdot \\
&\cdot E\left(\mathbf{s}[n-1] - \boldsymbol{\mu}_{\mathbf{s}}\left[n-1|n-1,\eta_j\left[n-1\right]\right]\right)^H|\tilde{\eta}_l[n]) = 0
\end{aligned} \tag{15}$$

considering that $\mathbf{s}[n-1]$ is independent of $\mathbf{u}[n]$.

The weights of the mixture components are calculated as follows:

$$\tilde{\alpha}_{\mathbf{s}l}[n] = \alpha_{\mathbf{s}j}[n-1]\cdot\alpha_{\mathbf{u}k}[n]\ . \tag{16}$$

Combining Eqs. (12), (14) and (16), the state distribution at time instance n is obtained as follows:

$$\begin{aligned}
\hat{\mathbf{s}}[n|n-1] \sim GMM\,(&\tilde{\alpha}_{\mathbf{s}l}[n], \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\tilde{\eta}_l[n]], \\
&\mathbf{M}[n|n-1,\tilde{\eta}_l[n]], \forall l = 1,2,\ldots,L)\ .
\end{aligned} \tag{17}$$

This mixture consists of $L = N \times K$ components. The Expectation Maximization (EM) algorithm is used to estimate this distribution with N mixture components. This algorithm prevents the exponential growth of the number of components in the mixture. For this reason an artificial set of data is randomized according to the GMM distribution stated in (17). A new reduced order GMM with N mixture components is estimated from this synthetic data set using the EM algorithm:

$$\begin{aligned}
\hat{\mathbf{s}}[n|n-1] \sim GMM\,(&\alpha_{\mathbf{s}j}[n], \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]], \\
&\mathbf{M}[n|n-1,\eta_j[n]], \forall j = 1,2,\ldots,N)\ .
\end{aligned} \tag{18}$$

Note, that in general the number of the mixture components may vary from state to state.

The state prediction is obtained from (7), choosing $m = n-1$:

$$\hat{\mathbf{s}}[n|n-1] = \sum_{j=1}^{N} P(\eta_j[n]|\mathcal{X}[n-1])\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]]\ . \tag{19}$$

Note, that $P(\eta_j[n]|\mathcal{X}[n-1])$, is the estimated mixture weights $\alpha_{\mathbf{s}j}[n]$, defined in (18). Therefore the prediction equation is rewritten as follows:

$$\hat{\mathbf{s}}[n|n-1] = \sum_{j=1}^{N} \alpha_{\mathbf{s}j}[n]\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]]\ . \tag{20}$$

The distribution of the innovation process $\tilde{\mathbf{x}}[n]$ is required to calculate the second term of (11). Using (2), and the fact that $\mathcal{X}[n-1]$ and $\mathbf{w}[n]$ are independent, the following is obtained:

$$\begin{aligned}
\hat{\mathbf{x}}[n|n-1] &= E(\mathbf{x}[n]|\mathcal{X}[n-1]) \\
&= E(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n]|\mathcal{X}[n-1]) = \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]\ .
\end{aligned} \tag{21}$$

Assuming that the measurement noise is zero-mean Gaussian, $\hat{\mathbf{x}}[n|n-1]$ is GMM-distributed. Note that Gaussian measurement noise model does not increase the number of mixture components. The prediction of the observation at time instance n is obtained as follows

$$\hat{\mathbf{x}}[n|n-1] = \sum_{j=1}^{N} \alpha_{\mathbf{s}j}[n]\boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_j[n]]\ , \tag{22}$$

where the mean estimation of $\mathbf{x}[n]$ based on $\mathcal{X}[n-1]$, given the jth Gaussian is:

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{x}}[n|n-1,\eta_j[n]] &= E(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n]|\mathcal{X}[n-1],\eta_j[n]) \\
&= \mathbf{H}[n]E(\mathbf{s}[n]|\mathcal{X}[n-1],\eta_j[n]) \\
&+ E(\mathbf{w}[n]|\mathcal{X}[n-1],\eta_j[n]) \\
&= \mathbf{H}[n]\boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]]\ .
\end{aligned}$$

It is clear from the (23), that the mean vectors in the distribution of $\hat{\mathbf{x}}[n|n-1]$ are linear transformation of the corresponding mean vectors in the distribution of $\hat{\mathbf{s}}[n|n-1]$.

From Eqs. (9) and (23), the innovation $\tilde{\mathbf{x}}[n]$ is a linear function of the measurement $\mathbf{x}[n]$ and the prediction $\hat{\mathbf{s}}[n|n-1]$. Note that $\tilde{\mathbf{x}}[n]$ and $\mathbf{x}[n]$ have jointly GMM distribution. Therefore, the distribution of the innovation process can be modeled by:

$$\tilde{\mathbf{x}} \sim GMM\left(\alpha_{\tilde{\mathbf{x}}_j}[n], \boldsymbol{\mu}_{\tilde{\mathbf{x}}_j}[n], \boldsymbol{\Gamma}_{\tilde{\mathbf{x}}_j}[n], \forall j = 1,2,\ldots,N\right)\ . \tag{23}$$

The mean vector of the jth component in this mixture is:

$$\begin{aligned}
\boldsymbol{\mu}_{\tilde{\mathbf{x}}_j}[n] &= E(\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1]|\eta_j[n]) \\
&= E(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]|\eta_j[n])) \\
&= \mathbf{H}[n]\left(E(\mathbf{s}[n]|\eta_j[n]) - E(\hat{\mathbf{s}}[n|n-1]|\eta_j[n])\right) = \\
&= \mathbf{H}[n]\left(\boldsymbol{\mu}_{\mathbf{s}_j}[n] - \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]]\right) = 0.
\end{aligned} \tag{24}$$

The matrix $\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n]$, can be calculated using the following expression:

$$\begin{aligned}
\tilde{\mathbf{x}}[n] &= \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1] \\
&= \mathbf{H}[n]\varsigma_{\mathbf{s}j}[n] + \mathbf{w}[n]\ .
\end{aligned} \tag{25}$$

This result holds because of (23) and (1). Using the last result $\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}}[n]$ is obtained as follows:

$$\begin{aligned}
\boldsymbol{\Gamma}_{\tilde{\mathbf{x}}j}[n] &= E(\tilde{\mathbf{x}}[n]\tilde{\mathbf{x}}^H[n]|\eta_j[n]) = \\
&= \mathbf{H}[n]E\left(\varsigma_{\mathbf{s}j}[n]\varsigma^H_{\mathbf{s}j}[n]|\eta_j[n]\right)\mathbf{H}^H[n] \\
&- \mathbf{H}[n]E\left(\varsigma_{\mathbf{s}j}[n]\mathbf{w}^H[n]|\eta_j[n]\right) - \\
&- E(\mathbf{w}[n]\varsigma^H_{\mathbf{s}j}[n]|\eta_j[n])\mathbf{H}^H[n] + E(\mathbf{w}[n]\mathbf{w}^H[n]|\eta_j[n]) \\
&= \mathbf{H}[n]\mathbf{M}[n|n-1|\eta_j[n]]\mathbf{H}^H[n] + \boldsymbol{\Gamma}_{\mathbf{w}}[n]\ .
\end{aligned} \tag{26}$$

The last equality holds because $\mathbf{s}[n]$ and $\mathbf{w}[n]$ are independent and $\mathbf{w}[n]$ is zero mean.

The measurement noise distribution is Gaussian and therefore, the mixture weights in (23) are

$$\alpha_{\tilde{\mathbf{x}}j}[n] = \alpha_{\mathbf{s}j}[n] \;\; \forall j = 1, \ldots, N \; . \tag{27}$$

Define the jth Kalman gain, corresponding to the jth mixture component, as follows:

$$\mathbf{K}_j[n] = \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}^{-1}_{\tilde{\mathbf{x}}j}[n] \; , \tag{28}$$

where

$$\begin{aligned}
&\mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n] = E(\mathbf{s}[n](\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1])^H|\eta_j[n]) \\
&= E((\mathbf{s}[n] - \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]] + \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]]) \\
&\cdot(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^H|\eta_j[n]) \\
&= E(\varsigma_{\mathbf{s}j}[n](\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^H|\eta_j[n]) \\
&+E(\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^H|\eta_j[n]) \\
&= E(\varsigma_{\mathbf{s}j}[n]\varsigma^H_{\mathbf{s}j}[n]|\eta_j[n])\mathbf{H}^H[n] \\
&+\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]]E(\tilde{\mathbf{x}}[n]|\eta_j[n]) = \mathbf{M}[n|n-1, \eta_j[n]]\mathbf{H}^H[n]
\end{aligned}$$

Using the definition of Kalman gain, the following is obtained:

$$\begin{aligned}
E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n], \eta_j[n]) &= \boldsymbol{\mu}_{\mathbf{s}j}[n] + \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n](\mathbf{\Gamma}^{-1}_{\tilde{\mathbf{x}}j}[n](\tilde{\mathbf{x}}[n] - \boldsymbol{\mu}_{\tilde{\mathbf{x}}j}[n]) \\
&= \boldsymbol{\mu}_{\mathbf{s}j}[n] + \left(\mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}^{-1}_{\tilde{\mathbf{x}}j}[n]\right)\tilde{\mathbf{x}}[n] \\
&= \boldsymbol{\mu}_{\mathbf{s}j}[n] + \mathbf{K}_j[n]\tilde{\mathbf{x}}[n].
\end{aligned} \tag{29}$$

Substitution of(29) into ?? yields Using the last expression, (11) can be rewriten as follows

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n, \eta_j[n]] = \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]] + \mathbf{K}_j[n]\tilde{\mathbf{x}}[n] \; . \tag{30}$$

The NGKF recursion requires the estimation of the covariance matrices. First the following expression is calculated:

$$\varrho[n] = \mathbf{s}[n] - \boldsymbol{\mu}_{\mathbf{s}}[n|n, \eta_j[n]] = \varsigma_{\mathbf{s}j}[n] - \mathbf{K}_j[n]\tilde{\mathbf{x}}[n] \; . \tag{31}$$

Finally, the covariance matrix for the jth mixture component is calculated as follows:

$$\begin{aligned}
&\mathbf{M}[n|n, \eta_j[n]] = E(\varrho[n]\varrho^H[n]|\eta_j[n]) \\
&= E(\varsigma_{\mathbf{s}j}[n]\varsigma^H_{\mathbf{s}j}[n]|\eta_j[n]) - E(\varsigma_{\mathbf{s}j}[n]\tilde{\mathbf{x}}^H[n]|\eta_j[n])\mathbf{K}^H_j[n] \\
&-\mathbf{K}_j[n]E(\tilde{\mathbf{x}}[n]\varsigma^H_{\mathbf{s}j}[n]|\eta_j[n]) \\
&+\mathbf{K}_j[n]\mathbf{\Gamma}_{\tilde{\mathbf{x}}j}[n]\mathbf{K}^H_j[n]\mathbf{M}[n|n-1, \eta_j[n]] - \mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n]\mathbf{K}^H_j[n] \\
&+\mathbf{\Gamma}_{\mathbf{s}\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}^{-1}_{\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}_{\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}^{-1}_{\tilde{\mathbf{x}}j}[n]\mathbf{\Gamma}^H_{\mathbf{s}\tilde{\mathbf{x}}j}[n] \\
&-\mathbf{K}_j[n]\mathbf{\Gamma}^H_{\mathbf{s}\tilde{\mathbf{x}}j}[n] = \mathbf{M}[n|n-1, \eta_j[n]] - \mathbf{K}_j[n]\mathbf{\Gamma}^H_{\mathbf{s}\tilde{\mathbf{x}}j}[n] \\
&= \mathbf{M}[n|n-1, \eta_j[n]] - \mathbf{K}_j[n]\mathbf{H}[n]\mathbf{M}[n|n-1, \eta_j[n]] \\
&= (\mathbf{I} - \mathbf{K}_j[n]\mathbf{H}[n])\mathbf{M}[n|n-1, \eta_j[n]] \; .
\end{aligned}$$

This completes the NGKF derivation.

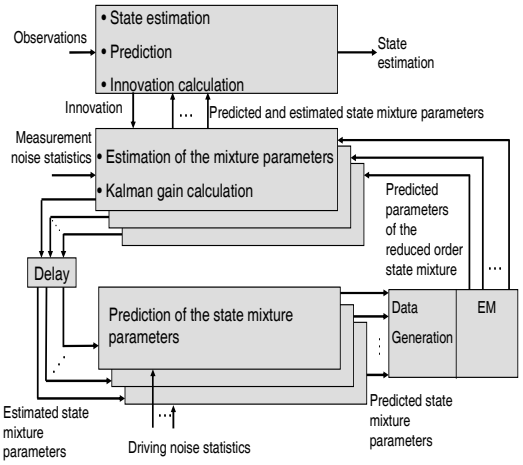

Fig. 1. NGKF conceptual scheme

4. NGKF ALGORITHM

A schematic description of the NGKF appears in Fig. 1, and its outline appears below.

• Prediction of the state mixture parameters:

$$\tilde{\alpha}_{\mathbf{s}_l}[n] = \alpha_{\mathbf{u}_k}[n] \cdot P(\eta_j[n-1]|\mathcal{X}[n-1])$$

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \tilde{\eta}_l[n]] = \mathbf{A}[n]\boldsymbol{\mu}_{\mathbf{s}}[n-1|n-1, \eta_j[n-1]] + \boldsymbol{\mu}_{\mathbf{u}k}[n],$$

$$\begin{aligned}
\mathbf{M}[n|n-1, \tilde{\eta}_l[n]] &= \mathbf{A}[n]\mathbf{M}[n-1|n-1, \eta_j[n-1]]\mathbf{A}^H[n]+ \\
&+\mathbf{\Gamma}_{\mathbf{u}k}[n], \\
\forall k = 1, 2, \ldots, K, \;\; &\forall j = 1, 2, \ldots, N, \;\; \forall l = 1, 2, \ldots, L \; .
\end{aligned}$$

Both the state and the driving noise pdfs are modeled by GMM, and therefore in the prediction stage the number of mixture components grows. The resulting mixture has $L = N \times K$ mixture components. The random indicator for the mixture with L components is denoted as $\tilde{\eta}_l[n], l = 1, \ldots, L$.

• Mixture order reduction:

$$\{\alpha_{\mathbf{s}_j}[n], \boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]], \mathbf{M}[n|n-1, \eta_j[n]]\}_{j=1}^N =$$

$$= EM_{L\to N}\left(\{\tilde{\alpha}_{\mathbf{s}_l}[n], \boldsymbol{\mu}_{\mathbf{s}}[n|n], \tilde{\eta}_l[n], \mathbf{M}[n|n-1, \tilde{\eta}_l[n]]\}_{l=1}^L\right) \; .$$

The $EM_{L\to N}$ operator means, that a synthetic data generated from L mixture components.

• Prediction:

$$\hat{\mathbf{s}}[n|n-1] = \sum_{j=1}^{N} \alpha_{\mathbf{s}_j}[n]\boldsymbol{\mu}_{\mathbf{s}}[n|n-1, \eta_j[n]] \tag{32}$$

• Kalman gain:

$$\begin{aligned}
\mathbf{K}_j[n] &= \mathbf{M}[n|n-1, \eta_j[n]]\mathbf{H}^H[n] \\
&\cdot\left(\mathbf{\Gamma}_{\mathbf{w}}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1, \eta_j[n]]\mathbf{H}^H[n]\right)
\end{aligned}$$

• Estimation of the mixture parameters:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_j[n]] &= \boldsymbol{\mu}_{\mathbf{s}}[n|n-1,\eta_j[n]] + \\ &+ \mathbf{K}_j[n]\left(\mathbf{x}[n]-\mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]\right) \quad (33)\\ \mathbf{M}[n|n,\eta_j[n]] &= (\mathbf{I}-\mathbf{K}_j[n]\mathbf{H}[n])\mathbf{M}[n|n-1,\eta_j[n]]\end{aligned}$$

The conditional probability that the data $\mathcal{X}[n]$ are generated by the mixture component $\eta_j[n]$ is calculated as follows

$$P(\eta_j[n]|\mathcal{X}[n]) = \frac{\alpha_{\mathbf{s}_j}[n]\mathcal{N}^c(\mathbf{0},\boldsymbol{\Gamma}_{\mathbf{x}j}[n])}{\sum_{i=1}^{N}\alpha_{\mathbf{s}_i}[n]\mathcal{N}^c(\mathbf{0},\boldsymbol{\Gamma}_{\mathbf{x}i}[n])},$$

• Estimation:

$$\hat{\mathbf{s}}[n|n] = \sum_{j=1}^{N} P(\eta_j[n]|\mathcal{X}[n])\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_j[n]] \quad (34)$$

where the conditional mean

$$\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_j[n]] \stackrel{\triangle}{=} E(\mathbf{s}[n]|\mathcal{X}[n],\eta_j[n]) \quad (35)$$

stands for the estimate of $\mathbf{s}[n]$ from $\mathcal{X}[n]$ given the random mixture indicator, $\eta_j[n]$.

The recursion is initialized with

$$\begin{aligned}\mathbf{s}[-1|-1] &= \sum_{j=1}^{N}\alpha_{\mathbf{s}j}[-1]\boldsymbol{\mu}_{\mathbf{s}j}[-1]\\ \mathbf{M}[-1|-1] &= \sum_{j=1}^{N}\alpha_{\mathbf{s}j}[-1]\boldsymbol{\Gamma}_{\mathbf{s}j}[-1]\,. \quad (36)\end{aligned}$$

The objective of this recursive algorithm is to update the estimation of $\mathbf{s}[n]$, given in (34), each time that a new measurement $\mathbf{x}[n]$ is obtained. This estimation is a weighted sum of $\boldsymbol{\mu}_{\mathbf{s}}[n|n,\eta_j[n]]$ - estimation of the mixture components means, based on the previous measurements. These mean estimations are recursively calculated for each mixture component in (34). In the investigated model, both the state and the driving noise are modeled by GMM. Therefore, the number of Gaussians in the pdf of $\hat{\mathbf{s}}[n|n-1]$ grows exponentially from iteration to iteration. In order to avoid the model order growth, synthetic data are generated, using GMM with $L = N \times K$ mixture components. Next, a new GMM of the predefined order-N is estimated from this synthetic data set, and used in Eq. (34). The NGKF can be interpreted as consisting of two parts. The first part consists of a bank of parallel algorithms for mixture statistics estimation. Each algorithm in this bank corresponds to the mixture component. In the second part of NGKF, the state prediction $\hat{\mathbf{s}}[n|n-1]$ and the innovation $\tilde{\mathbf{x}}[n]$, required for the Kalman gain calculation are obtained. Finally, the state estimation $\hat{\mathbf{s}}[n|n]$ is calculated.

5. SIMULATION RESULTS

In this section, two examples that demonstrate the estimation ability of the NGKF algorithm in a non-Gaussian problems, are presented. The estimation performance of NGKF is compared to the KF and to the PF. Although, both KF, and PF are sub-optimal algorithms in non-Gaussian problem, they serve as references for NGKF algorithm performance evaluation. The root-mean-square-error (RMSE) of the estimation was evaluated for all these algorithms. In these simulation examples, the standard sampling importance resampling (SIR) particle filter was used [1]. The number of particles were 10000. Different PF algorithms among many others could be simulated. There is no precise rule what PF should be used in each application. This implies that no rigorous PF algorithm exists. This is one of the disadvantages of PF algorithms comparing to the optimal NGKF algorithm. For estimation performance evaluation, 100 trials were performed. In both examples, the following NGKF algorithm was used:

- State vector $\mathbf{s}[n]$ of size 3.
- Measurement vector of size 1.
- The driving noise distribution is assumed to be GMM with 10 mixture components.
- The state distribution is GMM with 10 mixture components.

5.1. Example 1

In the first example, the following model was simulated:

- Transition matrix: $(\mathbf{A}[n])_{k,l} = 1,\ \ k,l = 1,2,3,\ ;\forall n$
- Measurement matrix $\mathbf{H}[n] = \left[\begin{array}{ccc} 1 & 0 & 0 \end{array}\right]\ \ \forall n$
- i.i.d. driving noise with uniform distribution $\mathbf{u}[n] \sim \mathbf{U}(-5,5)$.

The estimation performances of NGKF, KF and PF in terms of RMSE for this example are presented in Fig. 2. This figure shows that the RMSE, achieved by the NGKF, is lower than the PF and the KF.

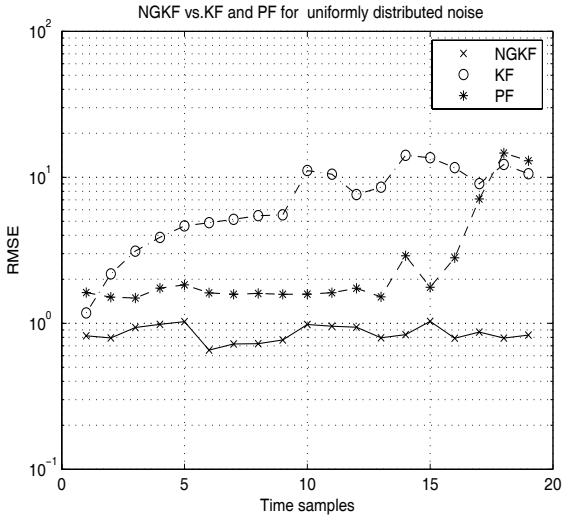

Fig. 2. NGKF vs. KF and PF estimation performance with uniformly distributed driving noise.

5.2. Example 2

This example represents a practical application for maneuvering radar target tracking. In this example, the state vector is composed of the target range, velocity and acceleration, where the target range is observed. The target is modeled by the state vector as follows:

$$\begin{bmatrix} r[n] \\ \dot{r}[n] \\ \ddot{r}[n] \end{bmatrix} = \begin{bmatrix} 1 & \Delta t_n & \frac{1}{2}\Delta t_n^2 \\ 0 & 1 & \Delta t_n \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r[n-1] \\ \dot{r}[n-1] \\ \ddot{r}[n-1] \end{bmatrix} + \mathbf{u}[n]$$

$$r[n] = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} r[n] \\ \dot{r}[n] \\ \ddot{r}[n] \end{bmatrix} + \mathbf{w}[n], \qquad (37)$$

where Δt_n is the sampling interval at time n. The target maneuvers (time-varying acceleration) are modeled by time-varying driving noise. Thus, in the first part of the simulation the target was assumed to have almost constant acceleration. This scenario is simulated by zero-mean Gaussian driving noise.

$$\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\rho) \quad \forall n = 1, 2, \ldots \qquad (38)$$

The ρ scales the covariance matrix of the driving noise. In the first part of the simulation ($n = 1, \ldots, 8$) this parameter is $\rho = 0.01$. In the second part of the simulation $n = 9, \ldots, 20$, the target's time-varying acceleration was modeled by the noise with "wider" Gaussian, defined by $\rho = 10$. The range estimation performance of NGKF, KF and PF are presented in Fig. 3. The RMSEs of the target

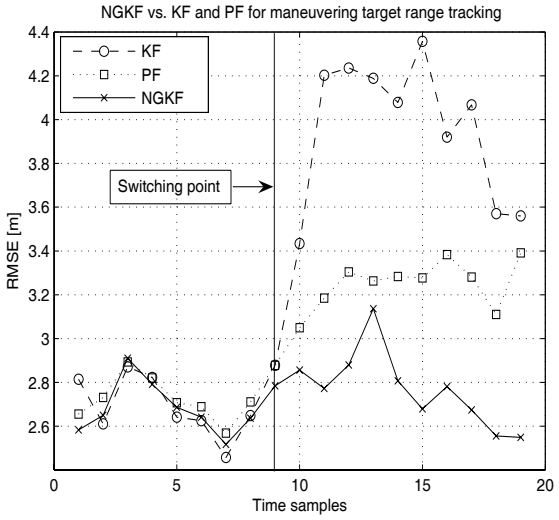

Fig. 3. NGKF vs. KF and PF for maneuvering target tracking.

range estimation by the NGKF, KF and PF are presented in this figure. The figure shows that in the first part of the simulation (fixed acceleration target), the RMSEs of the NGKF, PF and KF are very close. However, beyond the switching point (maneuvering target), both the NGKF and PF outperform the KF, and NGKF outperforms the PF.

6. CONCLUSION

A new recursive estimator, named NGKF, is presented in this work. This algorithm is optimal in the MMSE sense for a linear, non-Gaussian driving noise, that can be modeled by GMM. The system driving noise and state distributions are modeled by GMM. The proposed NGKF algorithm is the generalization of the original KF to the non-Gaussian model. The problem of exponential growth of the model order is solved via the mixture pdf re-estimation, using the EM algorithm. An implementation scheme of the NGKF algorithm was presented. Estimation performance of the NGKF is tested in non-Gaussian cases. As expected, the simulations show that the NGKF outperforms both the KF and PF. The proposed NGKF estimator is most useful in application, where the Gaussian assumption is not valid.

7. REFERENCES

[1] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," IEEE Trans. on Signal Processing, vol. 50, pp. 174-188, 2002.

[2] R. Kalman, "A new approach to linear filtering and prediction problem," Trans. ASME-J. of Basic Engineering, vol. 82D, pp. 35-45, 1960.

[3] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," In Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls, 1997.

[4] D. Alspach and H. Sorenson, "Non-linear Bayesian estimation using Gaussian sum approximation," IEEE Trans. Automat. Control, vol. AC-17, pp. 439-448, 1972.

[5] B. Anderson and J. Moore, Optimal filtering, Prentice-Hall, 1979.

[6] A. Doucet, J. de Freitas N. Gordon, "An introduction to sequential Monte Carlo methods," Sequential Monte Carlo methods in practice, Springer-Verlag: New York, 2001.

[7] R. Chen and J. Liu, "Mixture Kalman Filter," Journal of Royal Statistical Society, vol. 62, part 3, pp. 493-508, 2000.

[8] P. Djurić, J. Kotecha, J. Zhang, Y. Huang, T. Chirmai, M. Bugallo and J. Miguez, "Particle filtering," IEEE Signal Processing Magazine, pp. 19-38, 2003.

[9] J. Kotecha and P. Djurić, "Gaussian sum particle filtering," IEEE Trans. on Signal Processing, vol. 51, no. 10, pp. 2602-2612, 2003.

[10] J. Q. Li and A. R. Barron, "Mixture density estimation," In Advances in Neural Information Processing Systems 12, the MIT press, 2002.

[11] I. Bilik and J. Tabrikian, "Non-Gaussian Kalman Filter," Submitted to the IEEE Trans. on Signal Processing.