Approximate MMSE Estimator for Linear Dynamic Systems with Gaussian Mixture Noise

Leila Pishdad, Fabrice Labeau

Abstract—**In this work we propose an approximate Minimum Mean-Square Error (MMSE) filter for linear dynamic systems with Gaussian Mixture noise. The proposed estimator tracks each component of the Gaussian Mixture (GM) posterior with an individual filter and minimizes the trace of the covariance matrix of the bank of filters, as opposed to minimizing the MSE of individual filters in the commonly used Gaussian sum filter (GSF). Hence, the spread of means in the proposed method is smaller than that of GSF which makes it more robust to removing components. Consequently, reduction schemes with lower computational complexity can be used with the proposed filter without losing estimation accuracy and precision. This is supported through simulations on synthetic data as well as experimental data related to an indoor localization system. Additionally, we show that in two limit cases the state estimation provided by our proposed method converges to that of GSF, and we provide simulation results supporting this in other cases.**

Index Terms—**Bayesian tracking, linear estimation, Gaussian mixture noise, Gaussian sum filter, minimum mean-square-error (MMSE) estimator**

I. Introduction

The problem of estimating the unobservable state of a dynamic system from its available noisy measurements is prevalent in numerous signal processing contexts. Bayesian tracking techniques have been used for this purpose by applying a probabilistic framework and approximating the posterior, i.e. the conditional probability distribution function of the state given the measurements. For the special case of Gaussian noise with linear dynamic and measurement models, this posterior is Gaussian and its sufficient statistics are optimally tracked by a Kalman filter [1], [2]. The mean of this pdf acts as the state estimate and it is proved to be the minimum mean-square error (MMSE) estimator [3]. However, in many applications the posterior distribution is not Gaussian. For instance, in event-based state estimation the likelihood distribution is not Gaussian, which leads to a non-Gaussian posterior distribution [4], [5]. For the case of non-Gaussian distributions, Kalman filter cannot optimally track the mean and covariance matrix of the posterior(e.g. see [6]) and approximations should be made to provide suboptimal solutions [2].

Gaussian sum approximation has been an attractive method for estimating non-Gaussian distributions, since it provides asymptotically unbiased estimations [7], with the desired precision[1] [3, Chapter8; Lemma 4.1]. Additionally, by using Gaussian mixtures (GM), the approximated pdf is represented as a conditionally Gaussian distribution and this enables the analytic evaluation of a closed-form expression for the belief function. This is possible since, with GM distributions, the *multiple model* approach can be used, where each component in the GM corresponds to a model in the system and can be tracked by a Kalman filter [8]. Hence, the partitioned posterior can be estimated by a bank of Kalman filters, i.e. the Gaussian sum filter (GSF). Consequently, GMs have been widely used to model the different non-Gaussian distributions in sequential Bayesian tracking, including the prior [3], [9], [10], likelihood [4], [11], [12], predictive distribution [12] and noise distributions [8]–[10], [13]–[17]. They have also been used to directly approximate the posterior distribution [16], [18]–[21].

With GM prior, likelihood, or predictive density, the posterior is also a GM and the number of its components remains constant over time, as long as the noise distributions are Gaussian. For instance, in [3], it is shown that starting with a GM prior with a finite number of components, and additive white Gaussian noise, the predictive and posterior distributions are GMs with the same number of components. However, for GM noise distributions, the number of models in the system and consequently the number of components in the posterior grow exponentially over time. Hence, suitable GM reduction algorithms should be used, to merge or remove some of the components in the posterior as time progresses.

The mixture reduction algorithms can be categorized into three classes. In the first group, Expectation maximization (EM) is used to simultaneously predict and reduce the GM [11]. The second class of reduction algorithms rely on merging a pair or a group of components, i.e. replacing them by their moment-matching Gaussian distribution. There are different criteria for selecting the components to be merged. For instance, in Gaussian pseudo Bayesian (GPB) estimators, the components with the same state history are merged and replaced by a single Gaussian distribution [8]. A less computationally complex solution, also approximating a GM by a single Gaussian, is interacting multiple models (IMM) [8] which is used commonly as it requires fewer filters. Alternatively, in [16] the components are merged in the unlikely regions of the distribution, and they are split in the likely regions. Optimization techniques can also be used to select the merging components such that a cost function quantifying the dissimilarity between the GM distribution and the reduced distribution is minimized. An example cost function used in the literature is the Kullback-Leibler divergence (KLD) [22], [23]. Finally, the last category of reduction schemes requires removing a group of components [9], [10]. This class has

The authors are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada (e-mail: leila.pishdad@mail.mcgill.ca; fabrice.labeau@mcgill.ca)

[1] The parameters of the approximated GM including the number of its components, can be chosen such that the integral of the approximation error over the sample space is as small as desired.

the lowest computational complexity, especially since the reduction can be done before tracking, making it possible to avoid the evaluation of unused parameters. In the extreme case, the active component of the posterior is determined and all the other components are removed [14], [17]. This is equivalent to making a hard decision about the model in effect and hence the accuracy and precision of estimation is dependent on the correct choice of this model. At the cost of increased computational complexity, the performance of these methods can be improved by applying *resampling* procedures [13]. Alternatively, in [15] a *forgetting and merging* algorithm is proposed, where the components with weights smaller than a given threshold are removed, and the components with close enough moments are merged. A comprehensive review of the reduction algorithms for GM distributions is provided in [24].

The MMSE estimator is the expected value of the posterior [8]. Hence, GSF is the MMSE estimator of state[2] [25], [26]. However, due to using parallel Kalman filters in GSF, the MSE of each individual filter is minimized irrespective of the location of its mean with respect to the other components. Yet, the total covariance matrix of the filter is a function of both the individual filters' state estimation covariance matrices and the spread of their means.

In this work, we propose an approximate MMSE state estimator called AMMSE, which, unlike GSF, minimizes the trace of the total covariance matrix of the filter rather than the traces of the covariance matrices of individual filters. For this purpose, we re-derive the gains of individual filters such that the trace of the total covariance matrix of the filter is minimized. Hence, the spread of the means of the posterior components in the proposed estimator is smaller than that of GSF, making it more robust to removing some of the components. Consequently, AMMSE estimator can be used with the simplest and least computationally complex reduction scheme and yet achieve better performance when compared with GSF with the same reduction scheme. In other words, rather than using more computationally complex reduction schemes to improve the performance, we adjust the means of the components of the posterior and simply use the component with the largest weight as the state estimator, thus avoiding the unnecessary evaluation of the parameters of the other components. Additionally, through simulations we show that the difference between the distributions of the estimated state in GSF and AMMSE, as well as the true state is not statistically significant. Hence, despite the fact that by changing the gains of individual filters we are deviating from the true posterior, the total state estimation of AMMSE converges to the MMSE state estimation provided by GSF, and the true state.

The rest of this paper is organized as follows: In Section II the system model is defined and the notation used throughout the paper is introduced. Next, in Section III, we provide the details of GSF. In Section IV, we present our proposed AMMSE filter: first, in Section IV-A the gains of individual filters in the proposed method are derived, and then in Section IV-B it is compared with GSF in terms of computational complexity and the convergence of the two filters is analyzed. The numerical results are provided in Section V, comparing AMMSE, GSF, Kalman and Matched filters in terms of estimation accuracy with synthetic data (Section V-A) and experimental data gathered from an indoor localization system (Section V-B), and the convergence[3] of the state estimations of the filters and the true state is tested with synthetic data. Finally, in Section VI we provide the concluding remarks.

II. System Model

Suppose a discrete-time linear dynamic system, in which the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$, evolves as a first-order Markov process with additive noise with the initial state pdf $p(\mathbf{x}_0)$. Hence, using state-space representation, the dynamics equation can be written as

$$\mathbf{x}_k = F_k \mathbf{x}_{k-1} + \mathbf{v}_k, \tag{1}$$

where the process noise, $\{\mathbf{v}_k, k \in \mathbb{N}\}$ are independent random vectors with the pdfs $\{p(\mathbf{v}_k), k \in \mathbb{N}\}$ and F_k is the matrix describing the linear relationship between the previous and current state. If n_x denotes the dimension of the state vector, the process noise is of dimension n_x and F_k is of size $n_x \times n_x$.

In many applications, the state of the system cannot be observed directly. Hence, it is desirable to estimate the unobservable state from the available measurements. If we denote the measurement sequence by $\{\mathbf{z}_k, k \in \mathbb{N}\}$, the relationship between $\mathbf{x}_k$ and $\mathbf{z}_k$ is described by the measurement equation,

$$\mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{w}_k, \tag{2}$$

where the measurement noise $\{\mathbf{w}_k, k \in \mathbb{N}\}$ are independent random vectors with the pdfs $\{p(\mathbf{w}_k), k \in \mathbb{N}\}$, and it is independent from the process noise. The matrix defining the linear relationship between the current state and measurement vectors, H_k, is a known matrix of size $n_z \times n_x$, where n_z is the dimension of the measurement vector $\mathbf{z}_k$, and the noise vector $\mathbf{w}_k$.

Having the above, Bayesian tracking techniques can be used to probabilistically estimate the current state of the system from the available measurements. This is done by recursively estimating the posterior, $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, where $\mathbf{z}_{1:k}$ represents the available measurements up to and including time k. The recursive estimation of the posterior comprises of two steps: prediction and update. In the first step, $\mathbf{x}_k$ is predicted using the previous measurements, i.e. $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is estimated using Chapman-Kolmogrov equation on the previous posterior, $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$, and $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. Next, in the update phase, Bayes rule is used for updating the prior and evaluating the posterior, $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. The posterior is then used for next iteration estimation. With Gaussian process and measurement noise, the posterior distribution will be Gaussian and its sufficient statistics, i.e. mean and covariance matrix, are optimally tracked by a Kalman filter [1]. Additionally, since the posterior is Gaussian, the filtered mean approximates the state and it is the MMSE estimator [8].

[2] If no reduction scheme is used and the distributions are GMs (not *approximated* by GMs).

[3] Convergence in distribution is tested using two-sample Kolmogrov-Smirnov non-parametric tests.

In this work we use GM models for the initial state, as well as the process and measurement noise processes, since they are mathematically tractable and they can be used to approximate non-Gaussian distributions. Hence, the process noise, $\mathbf{v}_k$ is represented by a GM with $C_{\mathbf{v}_k}$ components, with $\left\{\mathbf{u}_k^i, 1 \le i \le C_{\mathbf{v}_k}\right\}$ the component means, $\left\{Q_k^i, 1 \le i \le C_{\mathbf{v}_k}\right\}$ the component covariance matrices and $\left\{\mathcal{W}_k^i, 1 \le i \le C_{\mathbf{v}_k}\right\}$ the non-negative mixing coefficients. This can be written as:

$$p(\mathbf{v}_k)=\sum_{i=1}^{C_{\mathbf{v}_k}} \mathcal{W}_k^i \mathcal{N}\left(\mathbf{v}_k; \mathbf{u}_k^i, Q_k^i\right), \tag{3}$$

where $\sum_{i=1}^{C_{\mathbf{v}_k}} \mathcal{W}_k^i = 1$, and $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ represents a Gaussian distribution with argument $\mathbf{x}$, mean $\boldsymbol{\mu}$, and covariance matrix Σ. The measurement noise distribution is also represented by a GM in a similar manner, and written as:

$$p(\mathbf{w}_k)=\sum_{j=1}^{C_{\mathbf{w}_k}} \mathcal{P}_k^j \mathcal{N}\left(\mathbf{w}_k; \mathbf{b}_k^j, R_k^j\right), \tag{4}$$

where, $C_{\mathbf{w}_k}$ is the number of components of the GM distribution with the non-negative coefficients $\left\{\mathcal{P}_k^j, 1 \le j \le C_{\mathbf{w}_k}\right\}$, and $\sum_{i=1}^{C_{\mathbf{w}_k}} \mathcal{P}_k^j = 1$. The mean and covariance matrix of component $j, 1 \le j \le C_{\mathbf{w}_k}$ are $\mathbf{b}_k^j$ and R_k^j, respectively.

III. Gaussian Sum Filters

With the GM noise distributions, i.e. (3)–(4), the dynamic system defined in (1)–(2), can be described as a *Multiple Model* system, with models $\left\{M_k^{ij}; 1 \le i \le C_{\mathbf{v}_k}, 1 \le j \le C_{\mathbf{w}_k}\right\}$, corresponding to the different components or *modes* of the process and measurement noises.[4] Hence, the posterior can be partitioned as follows:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_{i,j} p\left(\mathbf{x}_k|\mathbf{z}_{1:k}, M_k^{ij}\right) p\left(M_k^{ij}|\mathbf{z}_{1:k}\right). \tag{5}$$

The mode-conditioned posterior, $p\left(\mathbf{x}_k|M_k^{ij}, \mathbf{z}_{1:k}\right)$, is a Gaussian distribution with the pdf,

$$p\left(\mathbf{x}_k|M_k^{ij}, \mathbf{z}_{1:k}\right) = \mathcal{N}\left(\mathbf{x}_k; \hat{\boldsymbol{x}}_{k|k}^{ij}, \mathsf{P}_{k|k}^{ij}\right). \tag{6}$$

where its parameters $\hat{\boldsymbol{x}}_{k|k}^{ij}$ and $\mathsf{P}_{k|k}^{ij}$ can be tracked using the *mode-matched* Kalman filter [1], [8] as follows:

$$\hat{\mathbf{x}}_{k|k-1}^i = F_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{u}_k^i, \tag{7}$$

$$P_{k|k-1}^i = Q_k^i + F_k P_{k-1|k-1} F_k^{\mathsf{T}}, \tag{8}$$

$$\hat{\mathbf{z}}_k^{ij} = H_k \hat{\mathbf{x}}_{k|k-1}^i + \mathbf{b}_k^j, \tag{9}$$

$$\boldsymbol{\nu}_k^{ij} = \mathbf{z}_k - \hat{\mathbf{z}}_k^{ij}, \tag{10}$$

$$S_k^{ij} = H_k P_{k|k-1}^i H_k^{\mathsf{T}} + R_k^j, \tag{11}$$

$$\mathsf{W}_k^{ij} = P_{k|k-1}^i H_k^{\mathsf{T}} {S_k^{ij}}^{-1}, \tag{12}$$

$$\hat{\boldsymbol{x}}_{k|k}^{ij} = \hat{\mathbf{x}}_{k|k-1}^i + \mathsf{W}_k^{ij} \boldsymbol{\nu}_k^{ij}, \tag{13}$$

$$\mathsf{P}_{k|k}^{ij} = P_{k|k-1}^i - \mathsf{W}_k^{ij} S_k^{ij} {\mathsf{W}_k^{ij}}^{\mathsf{T}}. \tag{14}$$

where $(.)^{\mathsf{T}}$ indicates the transpose of its argument.

Hence, defining

$$\mu_k^{ij} \triangleq p\left(M_k^{ij}|\mathbf{z}_{1:k}\right), \tag{15}$$

we can write (5) as a GM distribution, with $C_{\mathbf{v}_k} C_{\mathbf{w}_k}$ components:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_{i,j} \mu_k^{ij} \mathcal{N}\left(\mathbf{x}_k; \hat{\boldsymbol{x}}_{k|k}^{ij}, \mathsf{P}_{k|k}^{ij}\right). \tag{16}$$

The coefficients μ_k^{ij} can be evaluated as:

$$\mu_k^{ij} = p\left(M_k^{ij}|\mathbf{z}_k, \mathbf{z}_{1:k-1}\right) \tag{17}$$

$$= \frac{p\left(\mathbf{z}_k|M_k^{ij}, \mathbf{z}_{1:k-1}\right) p\left(M_k^{ij}|\mathbf{z}_{1:k-1}\right)}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{18}$$

$$= \frac{\mathcal{N}\left(\mathbf{z}_k; \hat{\mathbf{z}}_k^{ij}, S_k^{ij}\right) p\left(M_k^{ij}|\mathbf{z}_{1:k-1}\right)}{\sum_{lm} \mathcal{N}\left(\mathbf{z}_k; \hat{\mathbf{z}}_k^{lm}, S_k^{lm}\right) p\left(M_k^{lm}|\mathbf{z}_{1:k-1}\right)}. \tag{19}$$

Assuming that the current model is independent from the previous model we have[5]

$$p\left(M_k^{ij}|\mathbf{z}_{1:k-1}\right) = \mathcal{W}_k^i \mathcal{P}_k^j. \tag{20}$$

Hence, we can write:

$$\mu_k^{ij} = \frac{\mathcal{W}_k^i \mathcal{P}_k^j \mathcal{N}\left(\mathbf{z}_k; \hat{\mathbf{z}}_k^{ij}, S_k^{ij}\right)}{\sum_{lm} \mathcal{W}_k^l \mathcal{P}_k^m \mathcal{N}\left(\mathbf{z}_k; \hat{\mathbf{z}}_k^{lm}, S_k^{lm}\right)}. \tag{21}$$

A. Reduction Schemes

As mentioned earlier, to avoid an exponentially growing bank size, reduction schemes should be applied to the posterior. In our work we use the less computationally complex schemes: merging all components to their moment-matching Gaussian distribution, and removing the components with smaller weights.[6]

For the first method, the moment-matched Gaussian distribution will have the following mean and covariance matrix:

$$\hat{\boldsymbol{x}}_{k|k} = \sum_{ij} \mu_k^{ij} \hat{\boldsymbol{x}}_{k|k}^{ij}, \tag{22}$$

$$\mathsf{P}_{k|k} = \sum_{ij} \mu_k^{ij} \left(\mathsf{P}_{k|k}^{ij} + \left(\hat{\boldsymbol{x}}_{k|k}^{ij} - \hat{\boldsymbol{x}}_{k|k}\right)\left(\hat{\boldsymbol{x}}_{k|k}^{ij} - \hat{\boldsymbol{x}}_{k|k}\right)^{\mathsf{T}}\right) \tag{23}$$

$$= \sum_{ij} \mu_k^{ij} \left(\mathsf{P}_{k|k}^{ij} + \hat{\boldsymbol{x}}_{k|k}^{ij} {\hat{\boldsymbol{x}}_{k|k}^{ij}}^{\mathsf{T}}\right) - \hat{\boldsymbol{x}}_{k|k} \hat{\boldsymbol{x}}_{k|k}^{\mathsf{T}}. \tag{24}$$

Alternatively, rather than making a soft decision about the active model, it can be determined by a hard decision. This is the approach used in [9], [10], [14], [17]. A simple scheme to determine the active model is to choose the component with the largest weight. Using this approach, if

$$ij = \arg\max_{lm} \mu_k^{lm}, \tag{25}$$

[4]For simplicity we assume a system of order one, where all the models have the same history of states. This assumption can be easily relaxed.

[5]This assumption can be easily relaxed.

[6]Other metrics can be used to determine the active model.

we have

$$\hat{\boldsymbol{x}}_{k|k} =\hat{\boldsymbol{x}}^{ij}_{k|k}, \tag{26}$$

$$\mathsf{P}_{k|k} =\mathsf{P}^{ij}_{k|k}. \tag{27}$$

One of the advantages of using this method over replacing the GM with its moment-matched Gaussian distribution, is that by determining the active model before evaluating the parameters of all components, computational resources can be saved. Additionally, evaluating (22)–(24) is more computationally complex than (26)–(27). Besides the computational complexity, by using a soft decision approach, we could be getting drifted from the *matched* estimation. Specifically, at every iteration only one model is active corresponding to the *Matched* filter. By incorporating the outputs of the mismatched filters in (22)–(24), the estimation accuracy and precision are lost. However, if the active model, hence the Matched filter are not chosen correctly, the soft decision method will provide better estimations. For simplicity, we refer to the first method as *merge* and the second as *remove*, e.g. GSF with the first method as the reduction scheme is referred to as *GSF-merge* in this paper.

Additionally, throughout this paper we use the term *mode-matched filter* to refer to the filter estimating the mode-conditioned state, from $p\left(\mathbf{x}_k|M^{ij}_k,\mathbf{z}_{1:k}\right)$. The term *Matched filter*, is used to denote the mode-matched filter corresponding to the known active model, M^*_k. Having the information about the active model, (5) is simplified and the posterior of the Matched filter will be

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = p(\mathbf{x}_k|\mathbf{z}_{1:k}, M^*_k). \tag{28}$$

Matched filter cannot be implemented without prior knowledge about the active model and is only used for comparison purposes (see Section V-A).

IV. AMMSE Estimator for GM noise

In this section we evaluate the gains of individual filters, W^{ij}_k such that the trace of the covariance matrix of the bank of filters, (24) is minimized. Since this covariance matrix is conditional on the measurements sequence, the evaluated gains are functions of innovations.

A. Derivation of Filter Gains

For an arbitrary filter gain W^{ij}_k, the state estimation error covariance matrix for filter ij can be written as follows:[7]

$$\begin{aligned}P^{ij}_{k|k} =&P^{i}_{k|k-1} - W^{ij}_k H_k P^{i}_{k|k-1}\\ &- P^{i}_{k|k-1} H^{\mathsf{T}}_k {W^{ij}_k}^{\mathsf{T}} + W^{ij}_k S^{ij}_k {W^{ij}_k}^{\mathsf{T}}\end{aligned} \tag{29}$$

Additionally, using (7) and (13), we can write

$$\hat{\mathbf{x}}^{ij}_{k|k} =F_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{u}^i_k + W^{ij}_k\boldsymbol{\nu}^{ij}_k. \tag{30}$$

Using this in (22), we have:

$$\hat{\mathbf{x}}_{k|k} = F_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbb{U}_k + \mathbb{S}_k, \tag{31}$$

[7]For Kalman gain this equation is simplified as in (12).

where

$$\mathbb{S}_k \triangleq \sum_{ij} \mu^{ij}_k W^{ij}_k \boldsymbol{\nu}^{ij}_k, \tag{32}$$

$$\mathbb{U}_k \triangleq \sum_{ij} \mu^{ij}_k \mathbf{u}^{i}_k. \tag{33}$$

To find the filter gains minimizing the MSE, the element-wise partial derivatives of the trace of the covariance matrix with respect to W^{ij}_k is set zero, i.e.:

$$\frac{\partial \mathrm{tr}\left(P_{k|k}\right)}{\partial W^{ij}_k} =0. \tag{34}$$

Thus, using (29), (30), and (31) in (24), we can write:[8]

$$\begin{aligned}&-P^{i}_{k|k-1} H^{\mathsf{T}}_k + W^{ij}_k S^{ij}_k + \mathbf{u}^i_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} + W^{ij}_k\boldsymbol{\nu}^{ij}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\\ &- \mathbb{U}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} - \mathbb{S}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} = 0,\end{aligned} \tag{35}$$

and we have:

$$\begin{aligned}W^{ij}_k =& \left(P^{i}_{k|k-1} H^{\mathsf{T}}_k + \mathbb{U}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} - \mathbf{u}^i_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} + \mathbb{S}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\right)\\ &\times \left(S^{ij}_k + \boldsymbol{\nu}^{ij}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\right)^{-1} = \mathrm{A}^{ij}_k + \mathbb{S}_k\mathrm{B}^{ij}_k,\end{aligned} \tag{36}$$

where

$$\begin{aligned}\mathrm{A}^{ij}_k \triangleq& \left(P^{i}_{k|k-1} H^{\mathsf{T}}_k + \mathbb{U}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} - \mathbf{u}^i_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\right)\\ &\times \left(S^{ij}_k + \boldsymbol{\nu}^{ij}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\right)^{-1},\end{aligned} \tag{37}$$

$$\mathrm{B}^{ij}_k \triangleq{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}} \left(S^{ij}_k + \boldsymbol{\nu}^{ij}_k{\boldsymbol{\nu}^{ij}_k}^{\mathsf{T}}\right)^{-1}. \tag{38}$$

Using (32) and (36), $\mathbb{S}_k$ can be evaluated as follows:

$$\sum_{ij} \mu^{ij}_k W^{ij}_k \boldsymbol{\nu}^{ij}_k = \sum_{ij} \mu^{ij}_k \mathrm{A}^{ij}_k \boldsymbol{\nu}^{ij}_k + \mathbb{S}_k \sum_{ij} \mu^{ij}_k \mathrm{B}^{ij}_k \boldsymbol{\nu}^{ij}_k. \tag{39}$$

Hence we have

$$\mathbb{S}_k =\sum_{ij} \mu^{ij}_k \mathrm{A}^{ij}_k \boldsymbol{\nu}^{ij}_k \left(1 - \sum_{ij} \mu^{ij}_k \mathrm{B}^{ij}_k \boldsymbol{\nu}^{ij}_k\right)^{-1}. \tag{40}$$

Using (40) in (36) the the optimal gains for the individual filters are computed and the estimated state, $\hat{\mathbf{x}}_{k|k}$ minimizing the trace of the covariance matrix of the filter can be evaluated from (31). However, changing the parameters of the components in the posterior, leads to larger MSE when compared with GSF, the MMSE filter.[9] Specifically, the MSE of AMMSE filter is the trace of the covariance matrix of the AMMSE filter and it can be evaluated as follows:

$$\mathrm{MSE}^{\mathrm{AMMSE}} \triangleq \mathrm{tr}\left(\mathbb{E}_{\mathbf{x}_k}\left\{\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right)^{\mathsf{T}}|\mathbf{z}_{1:k}\right\}\right) \tag{41}$$

$$= \mathrm{tr}\left(\mathsf{P}_{k|k} + \left(\hat{\boldsymbol{x}}_{k|k} - \hat{\mathbf{x}}_{k|k}\right)\left(\hat{\boldsymbol{x}}_{k|k} - \hat{\mathbf{x}}_{k|k}\right)^{\mathsf{T}}\right) \tag{42}$$

[8]We assume $\mu^{ij}_k \neq 0$, since for $\mu^{ij}_k = 0$ there is no need to evaluate the parameters of filter ij.

[9]GSF is the MMSE estimator when no reduction scheme is used and the noise distributions are GMs, rather than being *approximated* by GMs. However, as shown in [25] even with mixture reduction, GSF-merge converges to the MMSE filter.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2016.2594378

where $\mathbb{E}_{\mathbf{x}}\{g(\mathbf{x})\}$ is the expected value of the function $g(\mathbf{x})$ with respect to the random variable $\mathbf{x}$ with pdf $p(\mathbf{x})$, and $\hat{\boldsymbol{x}}_{k|k}$, $\mathsf{P}_{k|k}$ are the parameters of the MMSE filter evaluated as in (22)–(24).

B. Comparison with GSF

In this section we compare GSF and the proposed AMMSE filter in terms of computational complexity for one iteration (Section IV-B1) and discuss the convergence of AMMSE estimator to GSF (Section IV-B2).

1) Computational Complexity: In the prediction stage of Bayesian tracking and the evaluation of the coefficients μ_k^{ij}, both GSF and AMMSE have the same steps, hence the same computational complexity regardless of the choice of the reduction scheme. However, the reduction scheme affects the computational complexity of estimating the state and covariance matrix as well as the update stage of Bayesian tracking. Specifically, merging the means and covariance matrices of the components in (22)–(24) is more computationally complex than simply removing the components with smaller weights in (26)–(27). The update stage can be divided into three main operations: evaluating the gains, the component means, and covariance matrices. Depending on the reduction scheme, the number of times that these parameters have to be evaluated changes. For instance, in GSF-merge and AMMSE-merge, the parameters of all $C_{\mathbf{v}_k}C_{\mathbf{w}_k}$ filters in the bank have to be evaluated. Hence, both GSF-merge and AMMSE-merge have the same computational complexity for evaluating the gains, component means and covariance matrices, as well as merging the moments of the individual filters. For GSF-remove, each parameter needs to be evaluated for the *correct* component only. However, this is not true for AMMSE-remove, as the gains are dependent. Thus, to compute $\mathbb{S}_k$, the parameters A_k^{ij} have to be evaluated for all filters. Having $\mathbb{S}_k$, the parameters of the component with the maximum weight can be computed with the same computational complexity as GSF-remove. Hence, in terms of number of operations GSF-merge and AMMSE-merge are similar, while AMMSE-remove requires less operations and GSF-remove has the lowest computational complexity. However, as shown in Section V this is at the cost of degrading performance. The other three filters, i.e. AMMSE-merge/-remove and GSF-merge show similar performance with AMMSE-remove requiring the least number of operations.

2) Convergence to MMSE Estimator: At iteration k, the AMMSE state estimation, $\hat{\mathbf{x}}_{k|k}$, converges in distribution to the MMSE state estimation provided by GSF, $\hat{\boldsymbol{x}}_{k|k}$. Using (41), the convergence of the state estimations yields the convergence of the MSEs. Hence, the MSE of the AMMSE filter converges to the MSE of GSF-merge. Since it is not easily feasible to provide an analytical proof for all cases due to the matrix inversions in the evaluation of gains, we prove the convergence for two limit cases in Appendix A, and use Kolmogrov-Smirnov test for the other cases in Section V.

In the analytical proof in Appendix A, the convergence of state estimations is proved by showing the convergence of the parameters of the GM posterior, namely the coefficients, μ_k^{ij}, the means, $\hat{\mathbf{x}}_{k|k}^{ij}$, and the covariance matrices, $P_{k|k}^{ij}$, of the components. Since these parameters are dependent on the innovations of individual filters, $\boldsymbol{\nu}_k^{ij}$, two limit cases are considered: when the distance between innovations approaches zero and when it goes to infinity. The first case applies to a posterior with highly overlapping components, where the likelihoods of all models are close to one. Contrarily, when the distance between innovations approaches infinity, the likelihood of the active model is close to one and all the other components have negligible likelihoods. In other words, the active model can be well determined by using the coefficients, μ_k^{ij}, of the GM posterior.

For general GM noise models, the analytical proof is not straightforward, due to the matrix inversions in the gains, $\left(S_k^{ij} + \boldsymbol{\nu}_k^{ij}{\boldsymbol{\nu}_k^{ij}}^{\mathsf{T}}\right)^{-1}$ and ${S_k^{ij}}^{-1}$. Hence, in Section V, we provide Kolmogrov-Smirnov (KS) statistic for the distributions, $p\left(\hat{\mathbf{x}}_{k|k}\right)$ and $p\left(\hat{\boldsymbol{x}}_{k|k}\right)$ under different types of GM noise parameters. Specifically, we use GM noise models with different separations between components, and by running simulations on synthetically generated data with these noise models, we generate samples from the distributions $p\left(\hat{\mathbf{x}}_{k|k}\right)$ and $p\left(\hat{\boldsymbol{x}}_{k|k}\right)$. Using the KS test on these data, the hypothesis that the two data samples belong to the same distribution, is accepted at 95% confidence level.

V. Numerical Results

We consider two scenarios: In the first scenario, we use synthetically generated process and measurement noise processes, whereas for the second scenario we gather experimental data from an indoor localization system with ultra-wideband (UWB) sensors. To be consistent, we use the same process and measurement equations for both scenarios and they only differ in the noise distributions.

For both scenarios, we consider an indoor localization problem, in a 2D setting, and track the position in each direction independently assuming noise distributions with time-invariant statistics. The state vector contains the position and the velocity, but only noisy information about the position is observable and measured. Hence, in each direction we have $n_x = 2$, $n_z = 1$, and

$$F_k = \begin{bmatrix} 1 & \Delta t_k \\ 0 & 1 \end{bmatrix}, H_k = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{43}$$

where Δt_k is the time interval between the measurements z_{k-1} and z_k. In our localization system, the time intervals between measurements are multiples of 0.1080 s. Hence, in our synthetic setting, we use $\Delta t = 0.1080$ s for all iterations.

We use a random walk velocity motion model,

$$\mathbf{v}_k = v_k \times \begin{bmatrix} \Delta t_k \\ 1 \end{bmatrix}, \tag{44}$$

where v_k is a univariate GM random variable.

For the experimental setup, the noise distributions are estimated using the data gathered from the UWB sensors. For the synthetically generated data, we assume the same distribution for process and measurement noise.

The following filtering schemes are used and compared in terms of root-mean-square error (RMSE):

TABLE I
THE PARAMETERS OF THE GM MODELS USED FOR GENERATING SYNTHETIC DATA

	coefficients	means
Model 1	$[0.2, 0.2, 0.2, 0.2, 0.2]$	$c\,[-50, -30, 0, 30, 50]$
Model 2	$[0.1, 0.1, 0.6, 0.1, 0.1]$	$c\,[-50, -30, 0, 30, 50]$
Model 3	$[0.5, 0.1, 0.1, 0.1, 0.2]$	$c\,[-50, 10, 30, 50, 80]$

1) Kalman filter (KF)
2) GSF-merge/-remove
3) AMMSE-merge/-remove
4) Matched filter (for synthetically generated data)

A. Synthetic data

For the synthetically generated data, the process and measurement noise are assumed to be i.i.d. samples from one of the following GM models:

Model 1: Symmetric distribution with the components all having the same coefficients.

Model 2: Symmetric distribution with components having different weights.

Model 3: Asymmetric distribution.

The coefficients of the components are chosen such that the distributions are zero mean. In our simulations, we assume GM distributions with 5 components, each having a variance of 1. The mixing coefficients and the means of the components are given in Table I. The parameter c is used to vary the multi-modality of the GM distributions.

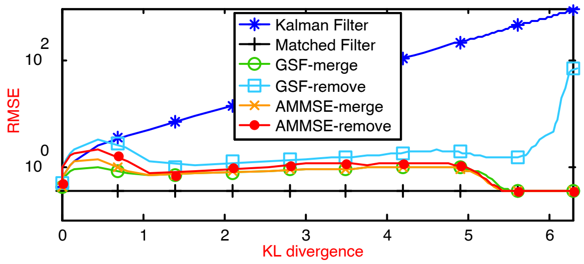
Using the GM noise models in Table I, we generate the measurements and estimate the state with Kalman filter, GSF-merge/-remove, and AMMSE-merge/-remove, as well as the Matched filter. To get the Matched filter estimations, we label the generated data by the active noise models in effect, and use this information to choose the correct filter in the bank to achieve the Matched filter. To further investigate the effect of multi-modality on the performance of these filters, we vary the parameter c for the noise models in Table I and evaluate the estimated state, and RMSE for all filters. For each value of c, 1000 Monte-Carlo runs are used to estimate the RMSE at 95% confidence level. We also approximate the KL divergence between the GM noise distributions and their corresponding moment-matched Gaussian density for each value of c. Fig. 1 shows the RMSE of the filters for different values of KL divergence between the used noise model and its moment-matched Gaussian distribution. As can be expected, as the divergence between the noise distribution and its corresponding fitted Gaussian increases, the performance of Kalman filter drops in terms of RMSE. This is due to the fact that Kalman filter is best suited for systems with Gaussian distributions, and it fails to provide good estimations for multi-modal noise models. Additionally, since the component variances of noise distributions are the same and by varying c only the separation between the components changes, the performance of the Matched filter remains the same for all values of KL divergence. However, for GSF-merge/-remove, and AMMSE-merge/-remove, the performance changes when the KL divergence increases.

With all noise models, the performance of all filters are close to the Matched filter for small KL divergences (as shown in Section IV-B2). However, as the KL divergence increases, RMSE increases for all filters, until it reaches its maximum. This is due to the fact that with an increase in the separation between the components, hence, the KL divergence with the corresponding moment-matched Gaussian distribution, the overlap between the components decreases. However, this decrease is not enough for the filters to correctly find the active model from the weights of components. In other words, the posterior is multi-modal, hence it cannot be well approximated by a single Gaussian. However, the overlap between the components of the GM posterior leads to drifting from the Matched filter estimation. Further increase in the KL divergence results in improved performance for AMMSE-merge/-remove and GSF-merge until they converge to the Matched filter (as shown in Section IV-B2). However, this is not the case for GSF-remove. Specifically, as the KL divergence increases, the RMSE of GSF-remove decreases until it reaches its minimum. But further increase in the KL divergence results in increased RMSE for this filter. This is because with GSF-remove the correct choice of the active component is of particular importance. Although, with the increase in KL divergence the component with the maximum weight represents the active model most of the times, when it fails and a wrong component is chosen, there is a larger error due to the increased distance between the component corresponding to the active model and the other components.

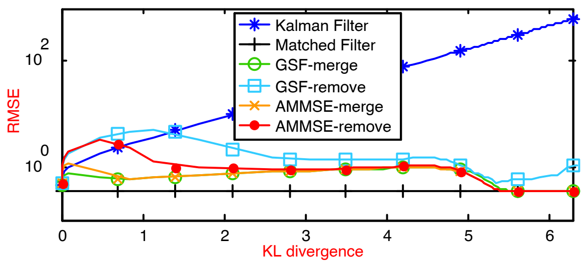
For all models, the performances of GSF-merge and AMMSE-merge are very close, especially, for KL divergence values greater than 1. To further investigate the relationship between these two filters, we used the state estimations, $\hat{\mathbf{x}}_{k|k}$, and $\hat{\boldsymbol{x}}_{k|k}$ to test the hypothesis that they come from the same distribution. Using Kolmogrov-Smirnov (KS) test, this hypothesis is accepted at 95% confidence level for all noise models and KL divergences. Additionally, since GSF-merge is not the MMSE filter due to the reduction of GM posteriors, we also carried out the two-sample KS hypothesis tests on the state estimations from AMMSE-merge and the true state. The hypotheses that the samples come from the same distributions are accepted at 95% confidence level for all noise models for KL divergences greater than 1. Moreover, to test the variances of the two filters we used Ansari-Bradley[10] test with the null hypothesis that the variances of these two sample sets are equal. The null hypotheses for Ansari-Bradley tests were also accepted for all noise models and KL divergences greater than 0.04 at 95% confidence level.

The RMSE of AMMSE-remove is always smaller or equal to the RMSE of GSF-remove. This is due to the fact that with AMMSE, the means of components are evaluated such that the trace of the total state estimation error covariance matrix including the spread of means in (23) is minimized. By contrast, in GSF the component means, $\hat{\boldsymbol{x}}^{ij}_{k|k}$ are minimizing the trace of the covariance matrix of each individual filter $\mathsf{P}^{ij}_{k|k}$. Hence, the component means are closer in AMMSE filter.

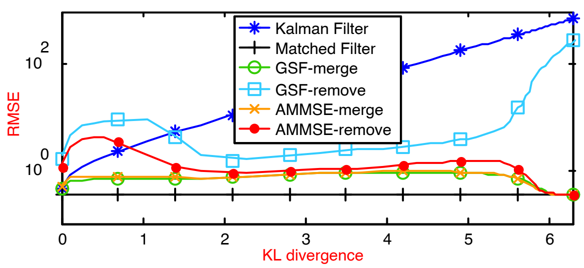
[10]Since the posteriors are not Gaussian the chi-square tests for normalized estimation error squares (NEES) cannot be applied.



(a) Model 1



(b) Model 2



(c) Model 3

Fig. 1. RMSE for the synthetic data generated using Models 1–3 vs. KL divergence between the noise distribution and the moment-matched Gaussian

Consequently, when removing the components with smaller weights as a reduction scheme, the performance is better.

Based on the results, it is evident that when the noise distributions are far from Gaussian (the KL divergence between the noise distribution and the corresponding moment-matched Gaussian density is large), GSF-merge, and AMMSE-merge/-remove perform similarly regardless of the shape of the noise model in terms of symmetry. This is particularly important, since among these filters which achieve comparable estimation accuracy, AMMSE-remove requires the least number of operations by avoiding the evaluation of $P^{ij}_{k|k}$ (as shown in Section IV-B1).

B. Experimental Data: Indoor Positioning System

In this section we provide the numerical results on the experimental data gathered from the indoor positioning system with off-the-shelf *Ubisense* UWB location sensors described in [27]. The noise distributions are approximated by GMs.

To compare the performance of different filtering schemes, in Table II we provide the RMSE in each direction. These values are the average of RMSE over all our experiments. Since the noise distributions are very close to their moment-matched Gaussian distribution[11], Kalman filter is similar to GSF-merge and AMMSE-merge in terms of RMSE. However, by minimizing the trace of the total state estimation error covariance matrix and decreasing the spread of means, AMMSE provides the best results in both directions and the performance of AMMSE-remove is better than GSF-remove.

TABLE II
RMSE FOR EXPERIMENTAL DATA

	x	y
Kalman	88.1850	73.0512
GSF-merge	85.7485	74.3982
GSF-remove	110.4006	75.1093
AMMSE-merge	**80.7388**	**71.3662**
AMMSE-remove	101.8026	72.0491

VI. CONCLUSION

In this paper, we propose an approximate MMSE (AMMSE) state estimator, for linear dynamic systems with Gaussian Mixture (GM) noise. For this purpose, we use a bank of Kalman filters with adjusted gains to track the models corresponding to the different components of the GM noise distributions. This is done by minimizing the trace of the total state estimation error covariance matrix, including the individual filters covariance matrices and the spread of their means. Hence, comparing with Gaussian Sum Filter (GSF) which minimizes the trace of the individual filters covariance matrices by using parallel Kalman filters, our proposed method has a smaller spread of means, and is more robust to removing components. Specifically, we have shown through simulations that unlike GSF, the performance of the proposed AMMSE filter does not change when instead of merging all components, we reduce the number of components by taking the component with the maximum weight. This is specifically important for applications which require lower computational complexities. We have also shown that the distributions of state estimations with GSF and AMMSE filter converge in two limit cases: when the distance between the GM components approaches zero and infinity. For the other cases, this is tested with Kolmogrov-Smirnov test on the state estimations of the two filters.

APPENDIX A
PROOF OF CONVERGENCE

Lemma 1: The estimated state of GSF and AMMSE filter converge in distribution when the distances between innovations approach zero.

Proof: Since the distances between innovations approach zero, they all converge to the same value, $\boldsymbol{\nu}^*_k$. In this case $\mathbb{S}_k$ approaches $\sum_{ij} \mu^{ij}_k W^{ij}_k \boldsymbol{\nu}^*_k$. Now, using (35), we can write:

$$-\sum_{ij} \mu^{ij}_k P^i_{k|k-1} H^\top_k + \sum_{ij} \mu^{ij}_k W^{ij}_k S^{ij}_k \to 0, \qquad (45)$$

[11]The KL divergences between the GM approximations of the process and measurement noise distributions and their corresponding Gaussian distributions are 0.4253 and 0.1759 in x direction and 1.1971 and 0.0200 in y direction, respectively [27].

which holds for GSF gains, W_k^{ij} in (12). But since there is only one unique solution for the set of equalities given in (35) for all i, j, in limit the gains of the individual filters in GSF and AMMSE become equal. Consequently, the parameters of the GM posteriors for GSF and AMMSE filter will converge in this limit case. ■

Lemma 2: The estimated state of GSF, AMMSE filter, and the Matched filter converge in distribution when the distances between innovations approach infinity.[12]

Proof: If we denote the active model by M_k^*, $\boldsymbol{\nu}_k^*$ approaches zero and its likelihood approaches 1. On the other hand, for all the other models $\{M_k^{lm}; 1 \leq l \leq C_{\mathbf{v}_k}, 1 \leq m \leq C_{\mathbf{w}_k}, M_k^{lm} \neq M_k^*\}$, the innovations $\boldsymbol{\nu}_k^{lm}$ increase and approach infinity and their likelihood approaches zero. Thus, using (19), we have

$$\mu_k^* \to 1; \mu_k^{lm} \to 0. \tag{46}$$

Hence, after merging[13]

$$\hat{\boldsymbol{x}}_{k|k} \to \hat{\boldsymbol{x}}_{k|k}^*; \hat{\mathbf{x}}_{k|k} \to \hat{\mathbf{x}}_{k|k}^*, \tag{47}$$

$$\mathsf{P}_{k|k} \to \mathsf{P}_{k|k}^*; P_{k|k} \to P_{k|k}^*. \tag{48}$$

Moreover, using (37)–(38), we can see that A_k^* approaches W_k^{ij} and B_k^* approaches zero, whereas for the non-matching models both A_k^* and B_k^* approach zero. Thus,

$$\hat{\mathbf{x}}_{k|k}^* \to \hat{\boldsymbol{x}}_{k|k}^*; P_{k|k}^* \to \mathsf{P}_{k|k}^*. \tag{49}$$

■

Acknowledgment

This work was partly supported by the Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners, through the Healthcare Support through Information Technology Enhancements (hSITE) Strategic Research Network.

References

[1] Y.-C. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.
[2] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
[3] B. D. O. Anderson and J. B. Moore, *Optimal filtering*. Englewood Cliffs: Prentice-Hall, 1979.
[4] D. Shi, T. Chen, and L. Shi, "An event-triggered approach to state estimation with multiple point- and set-valued measurements," *Automatica*, vol. 50, no. 6, pp. 1641–1648, 2014.
[5] J. Sijs and M. Lazar, "Event based state estimation with time synchronous updates," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2650–2655, Oct 2012.
[6] J. C. Spall and K. D. Wall, "Asymptotic distribution theory for the kalman filter state estimator," *Communications in Statistics - Theory and Methods*, vol. 13, no. 16, pp. 1981–2003, 1984.
[7] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
[8] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. Wiley-Interscience, 2001.
[9] 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.
[10] H. Sorenson and D. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, Jul. 1971.
[11] M. Huber, D. Brunn, and U. Hanebeck, "Efficient nonlinear measurement updating based on Gaussian mixture approximation of conditional densities," in *American Control Conference (ACC)*, 2007, pp. 4425–4430.
[12] J. Spall, "Estimation via markov chain monte carlo," *Control Systems, IEEE*, vol. 23, no. 2, pp. 34–45, Apr 2003.
[13] J. Kotecha and P. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.
[14] X. Sun, L. Munoz, and R. Horowitz, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *American Control Conference*, vol. 3, 2004, pp. 2098–2103.
[15] S. Ali-Löytty, "Box Gaussian mixture filter," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2165–2169, 2010.
[16] F. Faubel, J. McDonough, and D. Klakow, "The split and merge unscented Gaussian mixture filter," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 786–789, 2009.
[17] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.
[18] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.
[19] J. Kotecha and P. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.
[20] G. Terejanu, P. Singla, T. Singh, and P. Scott, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156, 2011.
[21] O. C. Schrempf and U. D. Hanebeck, "Efficient recursive prediction of stochastic nonlinear systems based on Dirac mixture approximations," in *In Proceedings of the American Control Conference*, 2007.
[22] J. R. Schoenberg, M. Campbell, and I. Miller, "Posterior representation with a multi-modal likelihood using the Gaussian sum filter for localization in a known map," *Journal of Field Robotics*, vol. 29, no. 2, pp. 240–257, 2012.
[23] A. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, 2007.
[24] D. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A look at Gaussian mixture reduction algorithms," in *14th International Conference on Information Fusion (FUSION)*, Jul. 2011, pp. 1–8.
[25] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.
[26] J. Tugnait and A. Haddad, "Adaptive estimation in linear systems with unknown Markovian noise statistics," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 66–78, 1980.
[27] L. Pishdad and F. Labeau, "A new reduction scheme for Gaussian Sum Filters," in *Asilomar Conference on Signals, Systems, and Computers*, 2014.

[12] We have assumed that the filters (GSF/AMMSE) are stable and have bounded error. Specifically, by choosing the wrong components, the state estimation error can increase for all individual filters in the bank leading to all the innovations approaching infinity. This is not the case considered here.

[13] If reduction is done by removing the components with smaller weights, as in (26)–(27), the two sides are equal.