

ROBUST GAUSSIAN SUM FILTERING WITH UNKNOWN NOISE STATISTICS: APPLICATION TO TARGET TRACKING

J. Vilà-Valls¹, Q. Wei², P. Closas¹, and C. Fernández-Prades¹

¹ Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 08860, Barcelona, Spain

² INP - ENSEEIHT, 2, rue Charles Camichel, 31071, Toulouse, France

Email: {jvila, pclosas, cfernandez}@cttc.cat, qi.wei@n7.fr

ABSTRACT

In many real-life Bayesian estimation problems, it is appropriate to consider non-Gaussian noise distributions to model the existence of outliers, impulsive behaviors or heavy-tailed physical phenomena in the measurements. Moreover, the complete knowledge of the system dynamics used to be limited, as well as for the process and measurement noise statistics. In this paper, we propose an adaptive recursive Gaussian sum filter that addresses the adaptive Bayesian filtering problem, tackling efficiently nonlinear behaviors while being robust to the weak knowledge of the system. The new method is based on the relationship between the measurement noise parameters and the innovations sequence, used to recursively infer the Gaussian mixture model noise parameters. Numerical results exhibit enhanced robustness against both non-Gaussian noise and unknown parameters, showing that good performances can be attained when compared to the standard known statistics case.

Index Terms— Adaptive Bayesian filtering, Gaussian sum filter, robustness, noise statistics estimation, innovations, tracking

1. INTRODUCTION

The problem under study concerns the derivation of efficient and robust methods to solve the recursive Bayesian filtering problem for nonlinear systems corrupted by non-Gaussian noise with unknown statistics. The classic filtering problem involves the recursive estimation of time-varying unknown states of a system using the incoming flow of information along some prior statistical knowledge about the variations of such states. The general discrete state-space model can be expressed as

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_k; \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k, \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the hidden state of the system at time k , $\mathbf{f}_{k-1}(\cdot)$ is a known, possibly nonlinear, function of the states; and \mathbf{v}_k is referred to as process noise; $\mathbf{z}_k \in \mathbb{R}^{n_z}$ is the measurement at time k , $\mathbf{h}_k(\cdot)$ is a known, possibly nonlinear, function, which relates measurements with states; and \mathbf{n}_k is referred to as measurement noise, independent of \mathbf{v}_k . The optimal Bayesian filtering solution is given by the marginal distribution $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, which gathers all the information about the system contained in the available observations, with $\mathbf{z}_{1:k} = \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$.

The Kalman filter (KF) provides the closed form solution to the optimal filtering problem in linear/Gaussian systems, assumptions

that not always hold, reason why suboptimal techniques have to be used. A plethora of alternatives have been proposed to solve the nonlinear estimation problem, among them, the Extended KF (EKF) and the family of Sigma-Point KFs (SPKF) [1, 2] within the Gaussian framework, and the family of Sequential Monte Carlo (SMC) methods [3], and Gaussian Sum Filters (GSFs) [4, 5], for arbitrary noise distributions. The latter constitutes an appealing alternative to SMC methods, which provide a powerful framework to deal with nonlinear/non-Gaussian systems at expense of a high computational load, being difficult to embed in digital light processors or in real-time applications. But the main limitation of all these methods is that they assume some *a priori* knowledge of the noise statistics affecting the system (*i.e.*, not only its distribution but its parameters). In many real-life systems, Gaussian noise models do not apply and the noise statistics are unknown. In these scenarios the methods based on the standard Gaussian Kalman framework give poor performance or even diverge, and we cannot directly apply SMC methods and GSFs because we need to estimate the states together with the noise statistics. Dropping these two classical assumptions, Gaussianity and known statistics, is the starting point of this work.

In this contribution, our attention focuses on the adaptive/robust filtering problem within the GSF framework, to deal with state-space models corrupted by non-Gaussian measurement noise with unknown statistics. Several improvements to the standard GSF and related methods have been proposed in the literature [2, 6, 7, 8], showing the promising applicability of GMMs, but all of them still considering known statistics. To the authors' knowledge, an adaptive filtering solution (*e.g.*, *adaptive filtering* referring to the unknown statistics estimation and robustness concepts) within the GSF framework to deal with systems corrupted by GMM non-Gaussian noise with unknown statistics is still an open problem.

The noise statistics estimation problem has been mainly studied for linear/Gaussian systems [9], and its extension to nonlinear systems usually makes use of linearization techniques. In order to deal with more general models, the problem has also been considered into the system parameters estimation framework [10]. In the case of GSFs, the noise statistics estimation reduces to the estimation of the GMM parameters. To solve this problem and provide a robust filtering solution, this paper proposes the Adaptive Gaussian Sum Filter (AGSF), which adjusts the parameters of the GMM distributions adaptively with no prior statistical information.

2. GAUSSIAN SUM FILTERING

In many applications, the Gaussian assumption does not hold [11, 12] and non-Gaussian distributions have to be used in order to capture the statistical behavior of the system dynamics and/or the prior

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness project TEC2012-39143 (SOSRAD), and by the European Commission in the Network of Excellence in Wireless COMMunications NEWCOM \sharp (contract n. 318306).

knowledge. On the basis of the Wiener approximation theory [13], any probability density function (pdf) can be expressed, or approximated with a given level of accuracy, using a finite sum of Gaussian densities, what is known as a Gaussian mixture model (GMM): $p_{\mathbf{x}}(\mathbf{x}) = \sum_{i=1}^L \alpha_i \mathcal{N}(\mathbf{x}; \mathbf{m}_i, \mathbf{B}_i)$, where $\sum_{i=1}^L \alpha_i = 1$, $\alpha_i \geq 0$ for $i = 1, \dots, L$, and $\mathcal{N}(\mathbf{x}; \mathbf{m}_i, \mathbf{B}_i)$ represents a Gaussian distribution with mean \mathbf{m}_i and covariance matrix \mathbf{B}_i (*i.e.*, for the sake of simplicity we use the notation $p(\mathbf{x}) \sim \{\alpha_i, \mathbf{m}_i, \mathbf{B}_i\}_{i=1}^L$). Therefore, any non-Gaussian distribution of interest can be expressed with a GMM, and considering GMM distributed additive noises, the predictive and posterior densities can also be expressed using Gaussian sums by using the following equalities. Assume that the posterior distribution, $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, can be represented with a GMM, $p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{i=1}^m \alpha_{i,k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{i,k}, \mathbf{B}_{i,k})$, then the mean and covariance matrix can be easily computed as,

$$\hat{\mathbf{x}}_{k|k} = \mathbb{E}(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{i=1}^m \alpha_{i,k} \mathbf{m}_{i,k}, \quad (2)$$

$$\Sigma_{k|k} = \sum_{i=1}^m \alpha_{i,k} [\mathbf{B}_{i,k} + (\hat{\mathbf{x}}_{k|k} - \mathbf{m}_{i,k})(\hat{\mathbf{x}}_{k|k} - \mathbf{m}_{i,k})^T]. \quad (3)$$

To obtain the recursive Bayesian filter in the GMM framework, the initial state (prior knowledge), and both process and measurement noise distributions, are assumed to follow a GMM distribution: $p(\mathbf{x}_0) \sim \{\alpha_{i,0}, \boldsymbol{\mu}_{i,0}, \Sigma_{x,i,0}\}_{i=1}^{l_0}$, $p(\mathbf{v}_k) \sim \{\beta_{j,k}, \boldsymbol{\nu}_{j,k}, \Sigma_{v,j,k}\}_{j=1}^{s_k}$ and $p(\mathbf{n}_k) \sim \{\gamma_{l,k}, \boldsymbol{\eta}_{l,k}, \Sigma_{n,l,k}\}_{l=1}^{m_k}$. Let us also assume that the posterior distribution at time $k-1$ is given by the following GMM, $p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1}) \sim \{\alpha_{i,k-1|k-1}, \hat{\mathbf{x}}_{i,k-1|k-1}, \Sigma_{x,i,k-1|k-1}\}_{i=1}^{l_{k-1}}$. The predictive density $p(\mathbf{x}_k | \mathbf{z}_{1:k-1})$ is obtained using the transition density and the Chapman-Kolmogorov equation, and approximated within the KF framework using a bank of $n_k = s_k l_{k-1}$ *time update* steps of the EKF,

$$\sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \beta_{j,k} \alpha_{i,k-1|k-1} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{ij,k|k-1}, \Sigma_{x,ij,k|k-1}), \quad (4)$$

where for each Gaussian term the state prediction and the corresponding error covariance matrix are

$$\hat{\mathbf{x}}_{ij,k|k-1} = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{i,k-1|k-1}) + \boldsymbol{\nu}_{j,k}, \quad (5)$$

$$\Sigma_{x,ij,k|k-1} = \mathbf{F}_{i,k-1} \Sigma_{x,i,k-1|k-1} \mathbf{F}_{i,k-1}^T + \Sigma_{v,j,k}, \quad (6)$$

with $\mathbf{F}_{i,k-1} = \left. \frac{\partial \mathbf{f}_{k-1}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{i,k-1|k-1}}$. At this point, note that

the predicted states are $\hat{\mathbf{x}}_{k|k-1} = \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \alpha_{ij,k|k-1} \hat{\mathbf{x}}_{ij,k|k-1}$ with $\alpha_{ij,k|k-1} = \beta_{j,k} \alpha_{i,k-1|k-1}$. The likelihood density can be written as $p(\mathbf{z}_k | \mathbf{x}_k) \sim \{\gamma_{l,k}, \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\eta}_{l,k}, \Sigma_{n,l,k}\}_{l=1}^{m_k}$ and the posterior distribution, $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, as done before with the predictive distribution, can be approximated using a GMM and a Kalman approach with a bank of $r_k = m_k s_k l_{k-1}$ *measurement update* steps of the EKF as $\sum_{l=1}^{m_k} \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \alpha_{ijl,k|k} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{ijl,k|k}, \Sigma_{x,ijl,k|k})$. For each term the state estimation and the corresponding error covariance matrix are $\hat{\mathbf{x}}_{ijl,k|k} = \hat{\mathbf{x}}_{ij,k|k-1} + \mathbf{K}_{ijl,k} (\mathbf{z}_k - \hat{\mathbf{z}}_{ijl,k|k-1})$ and $\Sigma_{x,ijl,k|k} = (\mathbf{I} - \mathbf{K}_{ijl,k}^T \mathbf{H}_{ij,k}) \Sigma_{x,ij,k|k-1}$, with $\mathbf{H}_{ij,k} = \left. \frac{\partial \mathbf{h}_k(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{ij,k|k-1}}$. Refer to Algorithm 2 for the computation of

the Kalman gain $\mathbf{K}_{ijl,k}$, the predicted measurement $\hat{\mathbf{z}}_{ijl,k|k-1}$, and the innovations covariance matrix $\Sigma_{zz,ijl,k|k-1}$. The weighting factor can be written as

$$\alpha_{ijl,k|k} = \frac{\gamma_{l,k} \beta_{j,k} \alpha_{i,k-1|k-1} \Omega_{ijl,k}}{\sum_{l=1}^{m_k} \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \gamma_{l,k} \beta_{j,k} \alpha_{i,k-1|k-1} \Omega_{ijl,k}}, \quad (7)$$

with $\Omega_{ijl,k} = \mathcal{N}(\mathbf{z}_k; \hat{\mathbf{z}}_{ijl,k|k-1}, \Sigma_{zz,ijl,k|k-1})$. The final state estimate is $\hat{\mathbf{x}}_{k|k} = \sum_{l=1}^{m_k} \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \alpha_{ijl,k|k} \hat{\mathbf{x}}_{ijl,k|k}$.

Notice that within one time step k , the number of Gaussian terms increased from l_{k-1} to $m_k s_k l_{k-1}$. In practice, Gaussian mixture reduction techniques [5, 2] have to be used to reduce the number of Gaussian terms after each iteration (in this work the implementation reported in [14] was used).

3. ADAPTIVE RECURSIVE BAYESIAN FILTER

This section proposes a new AGSF to solve the robust adaptive Bayesian filtering problem for nonlinear discrete state-space models corrupted by non-Gaussian measurement noise. This method estimates the statistical parameters of the GMM measurement noise using the innovations provided by the GSF.

3.1. Innovations and Measurement Noise

The innovations are denoted as $\tilde{\mathbf{z}}_k$ and the predicted state estimates are obtained as $\hat{\mathbf{x}}_{k|k-1} = \mathbb{E}(\mathbf{x}_k | \mathbf{z}_{1:k-1})$, which are available from the *Time Update* step of the GSF. The innovations sequence is defined as $\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) \approx \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{n}_k$, with $\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{H}_k = \left. \frac{\partial \mathbf{h}_k(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}$. If the measurement function \mathbf{h}_k is linear, the above approximation will turn into equality. Therefore, the pdf of innovations is equal to the convolution of the one-step prediction error's distribution and the measurement noise distribution,

$$p(\tilde{\mathbf{z}}_k | \mathbf{z}_{1:k-1}) \approx p(\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} | \mathbf{z}_{1:k-1}) \otimes p(\mathbf{n}_k), \quad (8)$$

where \otimes stands for the convolution operator. It is easy to see that both $p(\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} | \mathbf{z}_{1:k-1})$ and $p(\mathbf{n}_k)$ are Gaussian mixture distributed. Therefore, the essence of estimating $p(\mathbf{n}_k)$ from $\tilde{\mathbf{z}}_k$ is a deconvolution problem. Using (4), the pdf of $\tilde{\mathbf{x}}_{k|k-1}$ follows a GMM and is given by,

$$\begin{aligned} p(\tilde{\mathbf{x}}_{k|k-1} | \mathbf{z}_{1:k-1}) &= p(\mathbf{x}_k - \mathbb{E}(\mathbf{x}_k | \mathbf{z}_{1:k-1}) | \mathbf{z}_{1:k-1}) \\ &= \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \alpha_{ij,k|k-1} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{ij,k|k-1} - \hat{\mathbf{x}}_{k|k-1}, \Sigma_{x,ij,k|k-1}). \end{aligned} \quad (9)$$

where $\alpha_{ij,k|k-1} = \beta_{j,k} \alpha_{i,k-1|k-1}$. To simplify, we assume that the prediction error, $\tilde{\mathbf{x}}_{k|k-1}$ is Gaussian distributed, which means that we are only interested in the first and second moments of the distribution: $\mathbb{E}(\tilde{\mathbf{x}}_{k|k-1} | \mathbf{z}_{1:k-1}) = \mathbb{E}(\mathbf{x}_k | \mathbf{z}_{1:k-1}) - \hat{\mathbf{x}}_{k|k-1} = 0$ and $\Sigma_{k|k-1} = \sum_{j=1}^{s_k} \sum_{i=1}^{l_{k-1}} \alpha_{ij,k|k-1} [\Sigma_{x,ij,k|k-1} + (\hat{\mathbf{x}}_{ij,k|k-1} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{ij,k|k-1} - \hat{\mathbf{x}}_{k|k-1})^T]$.

Then, we conclude that

$$p(\mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} | \mathbf{z}_{1:k-1}) = \mathcal{N}(\tilde{\mathbf{x}}_{k|k-1}; \mathbf{0}, \mathbf{H}_k \Sigma_{k|k-1} \mathbf{H}_k^T). \quad (10)$$

Using (10) and (8), $p(\tilde{\mathbf{z}}_k | \mathbf{z}_{1:k-1}) = \sum_{l=1}^{m_k} \gamma_{l,k} \mathcal{N}(\tilde{\mathbf{z}}_k; \boldsymbol{\eta}_{l,k}, \mathbf{S}_{l,k})$, where $\mathbf{S}_{l,k} = \Sigma_{n,l,k} + \mathbf{H}_k \Sigma_{k|k-1} \mathbf{H}_k^T$. Therefore, the GMM innovations distribution has the same number of Gaussian terms than the measurement noise distribution.

3.2. Adaptive Estimation of the Noise Statistics

The adaptive filtering solution essentially refers to the estimation of the GMM parameters of the measurement noise based on the innovations sequence. The proposed method uses an Expectation Maximization (EM) solution [15], which is applied to process the

Algorithm 1 Recursive noise statistics estimation

1: Compute the auxiliar weights:

$$w_{l,k} = \frac{\hat{\gamma}_{l,k-1} \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_{l,k-1}, \hat{\mathbf{S}}_{l,k-1})}{\sum_{l=1}^{m_k} \hat{\gamma}_{l,k-1} \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_{l,k-1}, \hat{\mathbf{S}}_{l,k-1})} \text{ for } l = 1, \dots, m_k$$

2: Estimate the m_k weights: $\hat{\gamma}_{l,k} = \hat{\gamma}_{l,k-1} + \frac{1}{k}(w_{l,k} - \hat{\gamma}_{l,k-1})$.
 3: Estimate the m_k means: $\hat{\boldsymbol{\eta}}_{l,k} = \hat{\boldsymbol{\eta}}_{l,k-1} + G_{l,k}(\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,k-1})$.
 4: Estimate the m_k covariance matrices:

$$\hat{\mathbf{S}}_{l,k} = \hat{\mathbf{S}}_{l,k-1} + G_{l,k} \left[\frac{1}{1 - G_{l,k}} (\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,j})(\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,k})^T - \hat{\mathbf{S}}_{l,k-1} \right],$$

innovations, which obey a Gaussian mixture distribution, and finally get the GMM parameters estimates.

Assume that K innovations are collected from time instant 1 to K , and stacked in $\mathbf{Z}_K = (\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \dots, \tilde{\mathbf{z}}_K)$. The Gaussian mixture parameters for the l^{th} Gaussian term at time instant k are $\boldsymbol{\theta}_{l,k} = (\gamma_{l,k}, \boldsymbol{\eta}_{l,k}, \mathbf{S}_{l,k})$, and the full set is $\boldsymbol{\Theta} = (\boldsymbol{\theta}_{1,k}, \boldsymbol{\theta}_{2,k}, \dots, \boldsymbol{\theta}_{m_k,k})$. The log-likelihood of the innovations sequence is $\mathcal{L}(\mathbf{Z}_K | \boldsymbol{\Theta}_{1:K}) = \log \prod_{k=1}^K \sum_{l=1}^{m_k} \gamma_{l,k} \mathcal{N}(\tilde{\mathbf{z}}_k; \boldsymbol{\eta}_{l,k}, \mathbf{S}_{l,k})$. In this case, the parameters of each Gaussian term and the number of terms are assumed to be constant within the innovations sequence, so the time index is dropped to simplify. The parameters estimated in the previous iteration are denoted as $\boldsymbol{\theta}_l^g = \{\gamma_l^g, \boldsymbol{\eta}_l^g, \mathbf{S}_l^g\}_{l=1}^m$. From the likelihood expression, an EM method is applied to find the ML estimates of the GMM measurement noise parameters [15]. Two steps are performed recursively until convergence:

- E-step: $w_{l,k} = \mathbb{E}(\gamma_l | \tilde{\mathbf{z}}_k, \hat{\boldsymbol{\eta}}_l^g, \hat{\mathbf{S}}_l^g) = \frac{\gamma_l^g \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_l^g, \hat{\mathbf{S}}_l^g)}{\sum_{l=1}^m \gamma_l^g \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_l^g, \hat{\mathbf{S}}_l^g)}$ for $l = 1, \dots, m$, and $k = 1, \dots, K$.
- M-step: $\hat{\gamma}_l = \frac{1}{K} \sum_{k=1}^K w_{l,k}$, $\hat{\boldsymbol{\eta}}_l = \frac{\sum_{k=1}^K w_{l,k} \tilde{\mathbf{z}}_k}{\sum_{k=1}^K w_{l,k}}$ and $\hat{\mathbf{S}}_l = \frac{\sum_{k=1}^K w_{l,k} (\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_l)(\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_l)^T}{\sum_{k=1}^K w_{l,k}}$.

The final estimates are obtained as $\gamma_{l,k} = \hat{\gamma}_l$, $\boldsymbol{\eta}_{l,k} = \hat{\boldsymbol{\eta}}_l$ and $\mathbf{S}_{n,l,k} = \hat{\mathbf{S}}_l - \mathbf{H}_k \boldsymbol{\Sigma}_{k|k-1} \mathbf{H}_k^T$, for $k = 1, \dots, K$. This is a good solution if data can be stored and processed off-line, but in standard filtering applications the measurement noise parameters need to be estimated online. The sequential counterpart is presented hereafter. First, the auxiliar weights are obtained from the previous estimates and the current innovation, $w_{l,k} = \frac{\hat{\gamma}_{l,k-1} \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_{l,k-1}, \hat{\mathbf{S}}_{l,k-1})}{\sum_{l=1}^{m_k} \hat{\gamma}_{l,k-1} \mathcal{N}(\tilde{\mathbf{z}}_k; \hat{\boldsymbol{\eta}}_{l,k-1}, \hat{\mathbf{S}}_{l,k-1})}$, and then the new estimates are obtained as:
 $\hat{\gamma}_{l,k} = \hat{\gamma}_{l,k-1} + \frac{1}{k}(w_{l,k} - \hat{\gamma}_{l,k-1})$,
 $\hat{\boldsymbol{\eta}}_{l,k} = \hat{\boldsymbol{\eta}}_{l,k-1} + G_{l,k}(\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,k-1})$, and
 $\hat{\mathbf{S}}_{l,k} = \hat{\mathbf{S}}_{l,k-1} + G_{l,k} \left[\frac{1}{1 - G_{l,k}} (\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,k})(\tilde{\mathbf{z}}_k - \hat{\boldsymbol{\eta}}_{l,k})^T - \hat{\mathbf{S}}_{l,k-1} \right]$,
 where $G_{l,k} = \frac{w_{l,k}}{U_{l,k}}$ and $U_{l,k} = U_{l,k-1} + w_{l,k} (U_{l,1} = w_{l,1}, l = 1, \dots, m_k)$. One iteration (at time instant k) of the noise statistics estimation method is sketched in Algorithm 1 (*i.e.*, the previous estimates and the current innovation $\tilde{\mathbf{z}}_k$ are available), and marked in red in the complete AGSF sketched in Algorithm 2.

4. COMPUTER SIMULATIONS

In order to provide illustrative numerical results, the performance of the proposed method was evaluated in a 2-D radar target tracking application. The measurements were range and azimuth, $\mathbf{z}_k = [r_k \ \psi_k]^T$, and the states to be tracked were position and velocity of the target, respectively gathered in vector $\mathbf{x}_k =$

Algorithm 2 Adaptive Gaussian Sum Filter

Require: $\mathbf{z}_{1:\infty}$, $\{\alpha_{i,0}, \boldsymbol{\mu}_{i,0}, \boldsymbol{\Sigma}_{x,i,0}\}_{i=1}^{l_0}$, $\{\beta_{j,1}, \boldsymbol{\nu}_{j,1}, \boldsymbol{\Sigma}_{v,j,1}\}_{j=1}^{s_1}$, $\{\gamma_{l,1}, \boldsymbol{\eta}_{l,1}, \boldsymbol{\Sigma}_{m,l,1}\}_{l=1}^{m_1}$, l_0, s_1, m_1 .

1: Initialization: $\hat{\mathbf{x}}_{i,0} \sim \mathcal{N}(\boldsymbol{\mu}_{i,0}, \boldsymbol{\Sigma}_{x,i,0})$ for $i = 1, \dots, l_0$.
 2: **for** $k = 1$ to ∞ **do**

3: **Time update:**

4: **for** $j = 1$ to s_k **do**

5: **for** $i = 1$ to l_k **do**

6: Estimate the predicted states:

$$\hat{\mathbf{x}}_{i,j,k|k-1} = \mathbf{F}_{k-1}(\hat{\mathbf{x}}_{i,k-1|k-1}) + \boldsymbol{\nu}_{j,k}.$$

7: Estimate the error covariance matrices:

$$\boldsymbol{\Sigma}_{x,i,j,k|k-1} = \mathbf{F}_{i,k-1} \boldsymbol{\Sigma}_{x,i,k-1|k-1} \mathbf{F}_{i,k-1}^T + \boldsymbol{\Sigma}_{v,j,k}.$$

8: **end for**

9: **end for**

10: **Predicted states:** $\hat{\mathbf{x}}_{k|k-1} = \sum_{j=1}^{s_k} \sum_{i=1}^{l_k} \alpha_{i,j,k|k-1} \hat{\mathbf{x}}_{i,j,k|k-1}$.

11: **Innovation:** $\tilde{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$ with $\mathbf{H}_k = \left. \frac{\partial \mathbf{h}_k(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}$.

12: **Noise statistics estimation using Algorithm 1.**

13: **Measurement update:**

14: **for** $l = 1$ to m_k **do**

15: **for** $j = 1$ to s_k **do**

16: **for** $i = 1$ to l_k **do**

17: Estimate the predicted measurements:

$$\hat{\mathbf{z}}_{i,j,l,k|k-1} = \mathbf{h}_k(\hat{\mathbf{x}}_{i,j,k|k-1}) + \boldsymbol{\eta}_{l,k}.$$

18: Estimate the innovations covariance matrices:

$$\boldsymbol{\Sigma}_{z,z,i,j,l,k|k-1} = \mathbf{H}_{i,j,k} \boldsymbol{\Sigma}_{x,i,j,k|k-1} \mathbf{H}_{i,j,k}^T + \boldsymbol{\Sigma}_{n,l,k}.$$

19: Calculate the weights $\alpha_{i,j,l,k|k}$ using (7).

20: Estimate the Kalman gains:

$$\mathbf{K}_{i,j,l,k} = \boldsymbol{\Sigma}_{x,i,j,k|k-1} \mathbf{H}_{i,j,k}^T \boldsymbol{\Sigma}_{z,z,i,j,l,k|k-1}^{-1}.$$

21: Estimate the updated states:

$$\hat{\mathbf{x}}_{i,j,l,k|k} = \hat{\mathbf{x}}_{i,j,k|k-1} + \mathbf{K}_{i,j,l,k} (\mathbf{z}_k - \hat{\mathbf{z}}_{i,j,l,k|k-1}).$$

22: Estimate the corresponding error covariances:

$$\boldsymbol{\Sigma}_{x,i,j,l,k|k} = (\mathbf{I} - \mathbf{K}_{i,j,l,k}^T \mathbf{H}_{i,j,k}) \boldsymbol{\Sigma}_{x,i,j,k|k-1}.$$

23: **end for**

24: **end for**

25: **end for**

26: **Final state estimate:** $\hat{\mathbf{x}}_{k|k} = \sum_{l=1}^{m_k} \sum_{j=1}^{s_k} \sum_{i=1}^{l_k} \alpha_{i,j,l,k|k} \hat{\mathbf{x}}_{i,j,l,k|k}$,

27: **Gaussian mixture reduction:** pruning and merging.

28: **end for**

$[p_{x,k}, v_{x,k}, p_{y,k}, v_{y,k}]^T$. A two-terms GMM measurement noise was considered to model possible outliers in the measurement and the clutter effect in radar tracking systems. Notice that the standard GSF with known statistics was used as the benchmark, being the ultimate performance achievable for the proposed method.

The dynamics of the target were $\mathbf{x}_k = \boldsymbol{\Phi} \mathbf{x}_{k-1} + \boldsymbol{\Gamma} \mathbf{w}_k$, with

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \boldsymbol{\Gamma} = \begin{pmatrix} \frac{1}{2} T^2 & 0 \\ 0 & T \\ \frac{1}{2} T^2 & 0 \\ 0 & T \end{pmatrix}, T = 0.1s$$

the sampling period and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, 0.01 \cdot \mathbf{I}_2)$ a Gaussian process noise. The covariance of the resulting process noise can be expressed as $\mathbf{Q} = E[\boldsymbol{\Gamma} \mathbf{w}_k \mathbf{w}_k^T \boldsymbol{\Gamma}^T] = 0.01 \cdot \boldsymbol{\Gamma} \mathbf{I}_2 \boldsymbol{\Gamma}^T$. The initial state estimate was drawn from $\mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$ for each Monte Carlo trial ($L = 500$ independent Monte Carlo runs), with $\mathbf{x}_0 = [100 \ 1 \ 100 \ 1]^T$ being the true value, and $\mathbf{P}_0 = \text{diag}([100 \ 1 \ 100 \ 1])$.

The measurements were modeled as

$$\mathbf{z}_k = \begin{bmatrix} z_{k,1} \\ z_{k,2} \end{bmatrix} = \begin{pmatrix} \sqrt{p_{x,k}^2 + p_{y,k}^2} \\ \tan^{-1}(\frac{p_{y,k}}{p_{x,k}}) \end{pmatrix} + \mathbf{n}_k, \quad (11)$$

where \mathbf{n}_k followed a GMM distribution, $\mathbf{n}_k \sim \gamma_1 \mathcal{N}(\boldsymbol{\eta}_1, \mathbf{S}_1) + \gamma_2 \mathcal{N}(\boldsymbol{\eta}_2, \mathbf{S}_2)$, with $\gamma_1 = 0.3$, $\gamma_2 = 0.7$, $\boldsymbol{\eta}_1 = [-7 \ -0.35]^T$, $\boldsymbol{\eta}_2 = [3 \ 0.15]^T$ and $\mathbf{S}_1 = \mathbf{S}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0.01 \end{pmatrix}$. This noise was

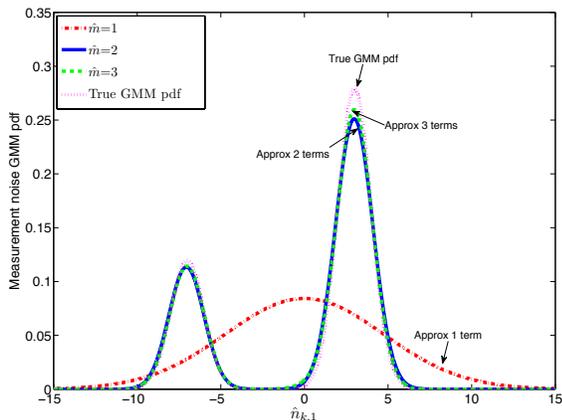
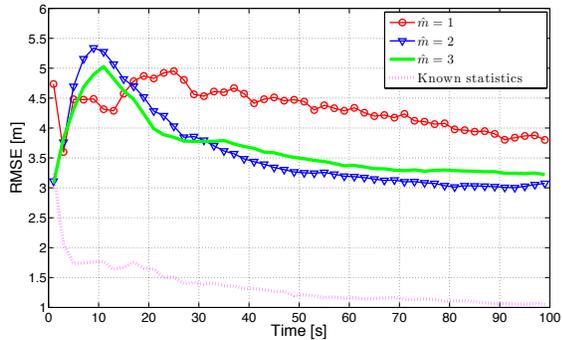


Fig. 1. (Top) RMSE for the AGSF ($\hat{m} = 1, 2, 3$) and the GSF with known statistics. (Bottom) Approximated pdf of the GMM measurement noise for $n_{k,1}$.

initialized as $\gamma_{i,0} = \gamma_i + 0.05\delta\gamma_i$, where $\delta\gamma_i \sim \mathcal{U}[0, 1]$, $\eta_{i,0} = \eta_i + \delta\eta_i \cdot [0.1 \ 0.01]^T$, where $\delta\eta_i \sim \mathcal{N}(0, 1)$, $S_{i,0} = 2S_i$, for $i = 1, 2$.

Figure 1 (top) shows the filtering results of the AGSF when $\hat{m} = 1, 2, 3$, and the GSF with known statistics. In this case, when considering 2 or 3 terms for the measurement noise approximation, good performances were obtained. The best results were obtained when the number of terms was exactly the correct one (2 in this case), but with a higher number of terms the loss in performance was not significant. In these cases the AGSF estimates the states of the system and the measurement noise statistics accurately, with limited performance degradation on the state estimation performance when compared to the benchmark, *i.e.*, the GSF with known statistics.

It seems obvious that the filtering results considering only one term were the worst, showing the improvement and therefore the interest of using the proposed method rather than an Adaptive KF when dealing with non-Gaussian measurement noise. Figure 1 (bottom) depicts the approximation of the true measurement noise pdf for three different approximations, $\hat{m} = 1, 2, 3$ terms, verifying the previous performance results because the approximation of the noises' pdf with 2 and 3 terms are almost the same, and much better than the approximation using only 1 Gaussian term.

5. CONCLUSIONS

This paper presented a solution to the robust Bayesian filtering problem. The new Adaptive Gaussian Sum Filter, based on an Expectation-Maximization type solution, is able to deal with non-

Gaussian measurement noise distributions with unknown statistics. This adaptive scheme can get rid of the limitations on the knowledge of the measurement noise, which means that the filter estimates the parameters of the Gaussian mixture model measurement noise exploiting information from the innovations. The method was validated by computer simulations, showing that the proposed algorithm can efficiently cope with the state estimation problem while correctly dealing with the unknown statistics of the non-Gaussian measurement noise. Future work goes towards formally proving the convergence of the algorithm.

6. REFERENCES

- [1] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Automat. Contr.*, May 2000.
- [2] I. Arasaratnam, S. Haykin, and R. J. Elliot, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proc. of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.
- [3] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo methods in practice*, Springer, 2001.
- [4] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, July 1971.
- [5] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Automat. Contr.*, vol. 17, no. 4, pp. 439–448, Aug. 1972.
- [6] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Trans. Automat. Contr.*, vol. 56, no. 9, pp. 2151–2156, 2011.
- [7] M. J. Caputi and R. L. Moose, "A modified Gaussian sum approach to estimation of non-Gaussian signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 2, pp. 446–451, 1993.
- [8] M. Kotecha and P. M. Djurić, "Gaussian Sum Particle Filtering," *IEEE Trans. Signal Processing*, October 2003.
- [9] M. Šimandl and J. Duňík, "Multi-step prediction and its application for estimation of state and measurement noise covariance matrices," Tech. Rep., Dep. of Cybernetics, University of West Bohemia, Pilsen, Czech Republic, 2006.
- [10] C. Andrieu, A. Doucet, S. S. Sumeetpal, and V. B. Tadić, "Particle methods for change detection, system identification, and control," *Proc. of the IEEE*, vol. 92, no. 3, pp. 423–438, 2004.
- [11] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods and results for Class A and Class B noise models," *IEEE Trans. on Information Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.
- [12] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley, "Statistics of co-channel interference in a field of poisson and poisson-poisson clustered interferers," *IEEE Trans. on Signal Processing*, vol. 58, no. 12, pp. 6207–6222, 2010.
- [13] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, USA, 1970.
- [14] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, nov. 2006.
- [15] J. A. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Tech. Rep. TR-97-021, 1997.