

MULTI-MODAL FILTERING FOR NON-LINEAR ESTIMATION

*Sanket Kamthe*¹

Jan Peters^{1,2}

Marc Peter Deisenroth^{3,1}

¹Department of Computer Science
Technische Universität Darmstadt
Germany

²MPI for Intelligent Systems
Tübingen
Germany

³Department of Computing
Imperial College London
United Kingdom

ABSTRACT

Multi-modal densities appear frequently in time series and practical applications. However, they are not well represented by common state estimators, such as the Extended Kalman Filter and the Unscented Kalman Filter, which additionally suffer from the fact that uncertainty is often not captured sufficiently well. This can result in incoherent and divergent tracking performance. In this paper, we address these issues by devising a non-linear filtering algorithm where densities are represented by Gaussian mixture models, whose parameters are estimated in closed form. The resulting method exhibits a superior performance on nonlinear benchmarks.

Index Terms— State estimation, Non-linear dynamical systems, Non-Gaussian filtering, Gaussian sum

1. INTRODUCTION AND RELATED WORK

Inference of latent state variables from noisy observations in time series models has been extensively studied. For linear estimation in stationary and non-stationary time series models the Kalman filter [1] has been shown to be highly efficient, theoretically and practically. The Kalman filter is optimal for linear Gaussian systems [2]. In such systems, the Gaussianity allows us to derive the recursive filtering equations in closed-form. In contrast, in a non-linear system Gaussian uncertainties become non-Gaussian due to the non-linear transform. Typically, we approximate a non-Gaussian density by a Gaussian, e.g. by linearising the functions as in the Extended Kalman Filter (EKF) or by deterministic sampling as in the Unscented Kalman Filter (UKF) [3]. Such approximations implicitly assume that the true densities are uni-modal. Filters based on these approximations often severely under-perform when true densities are multi-modal. Hence, multi-modal approaches are frequently needed.

Particle filters [4] are a standard approach for representing multi-modal and non-Gaussian densities. They are computationally demanding since they often require a large number of particles for good performance, e.g. due to the curse of dimensionality. An insufficient number of particles may fail to

capture the tails of the density and lead to degenerate solutions. In practice, we have to compromise between the deterministic and fast (UKF/EKF) or the computationally demanding and more accurate Monte Carlo methods [4].

An ideal filter for a non-linear system should allow for multi-modal approximations, and at the same time its approximations should be consistent to avoid degenerate solutions. In this paper, we propose a filtering method that approximates a non-Gaussian density by a Gaussian mixture model (GMM). Such a GMM allows modelling multi-modality as well as representing any density with arbitrary accuracy given a sufficiently large number of Gaussians, see Section 8.4 in [2] for a proof. The GMM presents an elegant deterministic filtering solution in the form of the Gaussian Sum filter [5].

The Gaussian Sum filter (GSUM-F) was proposed as a solution to estimation problems with non-Gaussian noise or prior densities. The GSUM-F relies on linear dynamics and the assumption that the parameters of the Gaussian mixture approximation to the non-Gaussian noise or prior densities are known a priori. This linearity assumption can be relaxed, e.g. by linearisation (EKF GSUM-F) [2] or deterministic sampling (UKF GSUM-F) [6], but both solutions still require a priori knowledge of the GMM parameters. If, however, the prior and noise densities are Gaussian, the UKF GSUM-F and EKF GSUM-F equal the standard UKF and EKF, i.e. they become uni-modal filters. To account for a possible uni-modal to multi-modal transition in a non-linear system, we need to solve two problems: the propagation of the uncertainty and the parameter estimation of the GMM approximation. Kotecha and Duric [7], proposed random sampling for uncertainty propagation and Expectation-Maximisation (EM) to estimate the GMM parameters. In this paper, we propose to propagate uncertainty deterministically using the Unscented Transform, which also allows for a closed-form expression of the GMM parameters.

The main contributions of this paper are the derivation of the Multi-Modal-Filter (M-MF), a multi-modal approach to filtering in non-linear dynamical systems, where all densities are represented by Gaussian mixtures. Moreover, we present closed-form expressions for estimating the parameters of the Gaussian mixture model.

The research leading to these results has received funding from the EC's Seventh Framework Programme under grant agreement #270327.

2. SYSTEM MODEL

We consider nonlinear dynamical systems

$$x_n = f(x_{n-1}) + w_n, \quad w_n \sim \mathcal{N}(0, Q), \quad (1)$$

$$y_n = h(x_n) + v_n, \quad v_n \sim \mathcal{N}(0, R), \quad (2)$$

where f and h are the non-linear transition and measurement function, respectively. The noise processes w_n and v_n are i.i.d. zero mean Gaussian with covariances Q and R , respectively. If w_n and v_n are non-Gaussian we use We denote the D dimensional state by x_n , and y_n is the E dimensional observation. $Y_j = \{y_1, \dots, y_j\}$ represents all observations up to time step j . We define the state estimation problem as determining the density $p(x_n|Y_j)$. Filtering and prediction are defined for $j = n$ and $j < n$, respectively. We define a GMM representation of a state distributions as

$$p(x_n|Y_j) := \sum_{i=0}^M \delta_i \varphi_i(x_{n|j}), \quad (3)$$

$$\varphi_i(x_{n|j}) := \mathcal{N}(x_{n|j} | \mu_{n|j}^i, \Sigma_{n|j}^i), \quad (4)$$

with weights $\delta_i \in [0, 1]$ and requiring $\sum_i \delta_i = 1$, which ensures that $p(x_n|Y_j)$ is a valid probability distribution.

3. MULTI-MODAL FILTERING

In the following, we devise a closed-form filtering algorithm with multi-modal representations of the state distributions. Our algorithm is inspired by the following observation made by Julier and Uhlmann [3]: ‘‘Given only the mean and the variance of the underlying distribution, and, in absence of any a priori information, any distribution (with the same mean and variance) used to calculate the transformed mean and variance is trivially optimal.’’ This observation was the basis to derive the Unscented transform and the UKF. However, predictions based on the Unscented Transform often under-estimate the true predictive uncertainty, which can result in incoherent state estimation and divergent tracking performance.

To address this issue, we use a different (optimal) representation of the underlying distribution, which still matches the mean and variance: We propose to represent each sigma point in the Unscented transform by a Gaussian centred at this sigma point. This approximation of the original distribution is effectively a GMM with $2D + 1$ components.

In Section 3.1, we derive an optimal GMM representation of a state distribution $p(x_{n-1})$ of which only the mean and variance are known. In Section 3.2, we detail how to map this GMM through a non-linear function to obtain a predictive distribution $p(x_n)$, which is represented by a GMM. We generalise both uncertainty propagation and parameter estimation to the case where $p(x_{n-1})$ is given by a GMM. In Section 3.3, we propose a method for pruning the number of mixture components in a GMM to avoid their exponential increase in number of components. In Section 3.4, we propo-

se the resulting filtering algorithm, which exploits the results from Sections 3.1–3.3.

3.1. Estimation of the Gaussian Mixture Parameters

Let the mean and the variance of the state distribution $p(x_{n-1})$ be given by μ , Σ , respectively. Then, we can represent $p(x_{n-1})$ by a GMM $p(x_{n-1}) = \sum_{i=0}^{2D} \delta_i \varphi_i(x_{n-1})$, such that the mean and the variance of the approximate density $p(x_{n-1})$ equal the mean μ and variance Σ of $p(x_{n-1})$. This representation is achieved by the closed-form relations

$$\begin{aligned} \delta_i &= 1/(2D + 1), \\ \mu^0 &= \mu, \quad \mu^j = \mu + \sigma^j, \quad \mu^{j+D} = \mu - \sigma^j, \\ \Sigma^i &= \left(1 - \frac{2\alpha}{D+1}\right) \Sigma, \end{aligned} \quad (5)$$

where $i = 0, \dots, 2D$ and $j = 1, \dots, D$, where D is the dimensionality of the state variable x_{n-1} . The variable σ denotes D rows or columns from the matrix square root $\pm\sqrt{\alpha\Sigma}$. From (5), we can see that we need to calculate $\sqrt{\Sigma}$ only once for all $2D + 1$ Gaussians $\varphi_i(x_{n-1})$. To ensure that Σ^i is positive semi-definite, the scaling factor α should be chosen such that $2\alpha \leq (2D + 1)$, see (5). For $2\alpha = 2D + 1$ in (5), the equations above reduce to scaled sigma points. Hence, the GMM representation in (5) can be considered a generalisation of the classical sigma point representation of densities employed by the Unscented Transform, where each sigma point becomes an improper probability distribution.

3.2. Propagation of Uncertainty

A key step in filtering is the uncertainty propagation step, i.e. estimating the probability distribution of random variable, which has been transformed by means of the transition function f . Given $p(x_{n-1})$ and the system dynamics (1), we determine $p(x_n)$ by evaluating $\int p(x_n|x_{n-1})p(x_{n-1})dx_{n-1}$. For non-linear functions f , the integral above can only rarely be solved in closed form. Thus, approximate solutions are required.

Uncertainty propagation in non-linear systems can be achieved by approximate methods, employing linearisation or deterministic sampling as in the EKF and UKF. In such approaches, the state distribution $p(x_{n-1})$ and the approximate predictive density $p(x_n)$ are well represented by Gaussians. If the state distribution $p(x_{n-1})$ is a Gaussian mixture as in (5), we can estimate the predictive distribution $p(x_n)$ similarly, e.g. by applying such an approximate update to each mixture component in the GMM, see (4). In this paper, we propagate each mixture component $\varphi_i(x_{n-1})$ of the GMM through f and approximate $p(x_n)$ by

$$\begin{aligned} p(x_n) &= \int p(x_n|x_{n-1}) \sum_{i=0}^{2D} \delta_i \varphi_i(x_{n-1}) dx_{n-1} \\ &\approx \sum_{j=0}^{2D} \delta_j \varphi_j(x_n), \end{aligned} \quad (6)$$

where the mean and covariance of each $\varphi_j(x_n)$ are computed by means of the Unscented Transform.

If the prior density is a Gaussian mixture $p(x_{n-1}) = \sum_{j=0}^{M-1} \beta_j \varphi_j(x_{n-1})$, we repeat the procedure above for each mixture component in $p(x_{n-1})$, i.e. we split each mixture component φ_j into $2D+1$ components $\delta_i \varphi_{ji}$, $i = 0, \dots, 2D$, and propagate them forward using the Unscented Transform. For notational convenience, we define this operation on a Gaussian mixture as $\mathcal{F}_n(f, p(x_{n-1}))$, such that

$$\begin{aligned} p(x_n) &= \mathcal{F}_n(f, p(x_{n-1})) = \int p(x_n|x_{n-1})p(x_{n-1})dx_{n-1} \\ &= \sum_{j=0}^{M-1} \sum_{i=0}^{2D} \beta_j \delta_i \int p(x_n|x_{n-1})\varphi_{ij}(x_{n-1})dx_{n-1} \\ &\equiv \sum_{l=0}^{M(2D+1)-1} \gamma_l \varphi_l(x_n), \end{aligned} \quad (7)$$

where $\gamma_l = \delta_i \beta_j$. We compute the moments of the mixture components φ_l by means of the Unscented Transform.

3.3. Mixture Reduction

Up to this point, we have considered the case where a density with known mean and variance has been represented by a GMM, which could subsequently be used to estimate the predicted state distribution. Incorporating these steps into an recursive state estimator for time series, there is an exponential growth in the number of mixture components in (7). One way to mitigate this effect is to represent the estimated densities by a mixture model with a fixed number of components [2]. To keep the number of mixture components constant we can reduce them at each time step [2].

A straightforward and fast approach is to drop the Gaussian components with the lowest weights. Such omissions, however, can result in poor performance of the filter [8]. Kitagawa [8] suggested to repeatedly merge a pair Gaussian components. A pair is selected with lowest distance in terms of some distance metric. We evaluated multiple distance metrics, e.g. the L_2 distance [9], the KL divergence [10], and the Cauchy Schwarz divergence [11]. In this paper, we used the symmetric KL divergence [8], $D(p, q) = (\mathcal{KL}(p|q) + \mathcal{KL}(q|p))/2$, which outperformed aforementioned distance measures for mixture reduction in filtering.

3.4. Filtering

In the following, subsume all derivations in our multi-modal non-linear state estimator, whose time and measurement updates are summarised in the following.

3.4.1. Time Update

Assume that the filter distribution $p(x_{n-1|n-1})$ is represented by a GMM with M components. The time update, i.e. the

one-step ahead predictive distribution is given by

$$p(x_{n|n-1}) = \int p(x_n|x_{n-1})p(x_{n-1|n-1})dx_{n-1}. \quad (8)$$

This integral can be evaluated as $\mathcal{F}_n(f, p(x_{n-1|n-1}))$, such that we obtain a GMM representation of the time update

$$p(x_{n|n-1}) = \sum_{j=0}^{M(2D+1)-1} \gamma_j \varphi_j(x_{n|n-1}) \quad (9)$$

as detailed in (7).

3.4.2. Measurement Update

The measurement update can be approximated up to a normalisation constant by

$$p(x_{n|n}) \propto p(y_n|x_n)p(x_{n|n-1}), \quad (10)$$

where $p(x_{n|n-1})$ is the time update (9). We now apply a similar operation as in (7) with h as non-linear function and obtain

$$p(y_{n|n-1}) = \mathcal{F}_n(h, p(x_{n|n-1})). \quad (11)$$

Substituting (11) and (9) in (10) yields the measurement update, i.e. the filtered state distribution

$$\begin{aligned} p(x_{n|n}) &\propto \sum_{i=0}^{2D} \delta_i \varphi_i(y_{n|n-1}) \sum_{j=0}^{M(2D+1)-1} \gamma_j \varphi_j(x_{n|n-1}) \\ &\equiv \sum_{l=0}^{M(2D+1)^2-1} \beta_l \varphi_l(x_{n|n}). \end{aligned} \quad (12)$$

We calculate the measurement update for each pair φ_i and φ_j . Recalling that $\varphi_l(x_{n|n}) = \mathcal{N}(x|\mu_{n|n}^{ij}, \Sigma_{n|n}^{ij})$ for $i = 0, \dots, 2D$ and $j = 0, \dots, (2D+1)^2 - 1$, the measurement updates [12] and weight updates (Gaussian Sum [5]) can be derived by

$$\begin{aligned} K_n^j &= \Gamma_{n|n-1}^j (\Sigma_{n|n-1}^j)^{-1}, \\ \mu_{n|n}^{ij} &= \mu_{n|n-1}^i + K_n^j (y - \mu_{n|n-1}^j), \\ \Sigma_{n|n}^{ij} &= \Sigma_{n|n-1}^i - K_n^j (\Sigma_{n|n-1}^j) K_n^{jT}, \\ \beta_{i,j} &= \frac{\delta_i \gamma_j \mathcal{N}(x=y|\mu_{n|n-1}^j, \Sigma_{n|n-1}^j)}{\sum_{k,l} \delta_l \gamma_k \mathcal{N}(x=y|\mu_{n|n-1}^k, \Sigma_{n|n-1}^k)}, \end{aligned} \quad (13)$$

where $\Gamma_{n|n-1}^j$ is the cross covariance matrix $\text{cov}(x_{n-1}, x_n)$ determined via the Unscented Transform [12].

After the measurement update, we reduce the $M(2D+1)^2$ mixture components in the GMM, see (12), to M according to Section 3.3. The multi-modal filter can also handle non-Gaussian noise if we express the noise density as a Gaussian Mixture.

4. RESULTS

We evaluated our proposed filtering algorithm on data generated from one-dimensional non-linear benchmark dynamical systems. Both the UKF and the Multi-Modal Filter (M-MF) use the same parameters for the Unscented Transform, i.e. $\alpha = 1$, $\beta = 2$ and $\kappa = 2$. The prior $p(x_0)$ was a standard Gaussian $p(x_0) = \mathcal{N}(x|0, 1)$. Densities in the M-MF were represented by a Gaussian mixture with $M = 3$ components. The mean of the filtered state was estimated by the first moment of this Gaussian mixture. The employed particle filter (PF) was a standard particle filter with residual re-sampling scheme. The PF-UKF [13] on the other hand used the Unscented Transform as proposal distribution and the UKF for filtering. Unlike our M-MF the PF-UKF used random sampling and the UKF. We used the root mean square error (RMSE) and the predictive Negative Log-Likelihood (NLL) per observation as metrics to compare the performance of the different filters. Lower values indicate better performance. The results in Table 1 were obtained from 100 independent simulations with $T = 100$ time steps for each of the following system models.

4.1. Non-stationary Time Series

We tested the different filters on a standard system, the Uniform Non-stationary Growth Model (UNGM) from [4]

$$x_n = \frac{x_{n-1}}{2} + \frac{25x_{n-1}}{1+x_{n-1}^2} + 8 \cos(1.2(n-1)) + w, \quad (14)$$

$$y_n = \frac{x_n^2}{20} + v,$$

where $w \sim \mathcal{N}(0, 1)$, $v \sim \mathcal{N}(0, 1)$. The true state density of the non-stationary model stated above alternated between multi-modal and uni-modal distributions. The switch from uni-modal to a bi-modal density occurred when the mean was close to zero. The quadratic measurement function makes it difficult to distinguish between the two modes as they are symmetric around zero. This symmetry posed a substantial challenge for several filtering algorithms. The PF lost track of the state dynamics as $N = 500$ particles failed to capture true density especially in its tails, which led to degeneracy [4]. The particle filters in Table 1 are in their standard form and performance may improve if advanced techniques are used [4], as shown by the Gibbs filter [14]. The proposed M-MF could track both modes and, hence, led to more consistent estimates. The RMSE performance of both multi-modal filter (M-MF) and UKF are similar with M-MF performing better in terms of a lower mean error and standard deviation. The main advantage of the M-MF is its ability to capture the uncertainty. The NLL values of the M-MF were significantly better than the UKF even when the same parameters are used to calculate the Unscented Transform, see Table 1.

We tested the UNGM with an alternative measurement function $h(x) = 5 \sin(x)$. For this function, the performance of the EKF is best in terms of RMSE, since its estimates

	Stationary $h(x) = 5 \sin(x)$		Non-Stationary			
	RMSE	NLL	$h(x) = x^2/20$		$h(x) = 5 \sin(x)$	
			RMSE	NLL	RMSE	NLL
EKF	7.5 ± 0.4	340.7 ± 38.6	10.9 ± 1.3	103.6 ± 29.2	10.4 ± 0.4	662.2 ± 98.5
UKF	12.2 ± 2.3	64.2 ± 46.7	6.5 ± 1.9	15.1 ± 12.3	10.0 ± 2.4	42.0 ± 32.8
M-MF	0.97 ± 0.17	1.375 ± 0.1595	3.48 ± 0.56	2.03 ± 0.7	6.4 ± 1.73	8.65.7 ± 8.04
PF	2.4 ± 0.03	N/A	3.5 ± 0.86	N/A	11.2 ± 3.6	N/A
PF-UKF	16.8 ± 0.1	N/A	13.1 ± 1.8	N/A	20.0 ± 2.4	N/A
Gibbs Filter	3.62 ± 0.4	3.06 ± 0.03	3.97 ± 0.4	2.18 ± 0.1	8.64 ± 0.4	3.57 ± 0.08

Table 1. Average performances of the filters are shown along with standard deviation. Lower values are better. The M-MF filter performs better than the standard filtering methods.

are more stable. The proposed M-MF could track multiple modes, which resulted in significantly better performance in terms of NLL. Moreover, the filter performance was consistently stable, indicated by the small standard deviation values for the NLL measure.

4.2. Stationary Time Series

We modified the UNGM described above to

$$x_n = \frac{x_{n-1}}{2} + \frac{25x_{n-1}}{1+x_{n-1}^2} + w, \quad w \sim \mathcal{N}(0, 1),$$

by dropping the time dependant cosine term from (14) and use a sinusoidal measurement function

$$y_n = 5 \sin(x_n) + v, \quad v \sim \mathcal{N}(0, 1).$$

We see from the results in Table 1 that the UKF was outperformed by all other deterministic filters. The failures of the UKF and the PF-UKF are attributed to their overconfident predictions for sinusoidal functions, which confirms the results in [15]. The EKF approximates these sinusoidal functions better but both the UKF and EKF fail to capture the multi-modal nature of system dynamics. Thus, the EKF and UKF are inconsistent for the model and settings used in this experiment. The proposed M-MF on the other hand performed consistently better in terms of RMSE and NLL values. Moreover, the small standard deviation of the NLL suggests that our proposed M-MF is consistent and stable.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented the M-MF, a Gaussian mixture based multi-modal filter for state estimation in non-linear dynamical systems. Multi-modal densities are represented by Gaussian mixtures, whose parameters are computed in closed form. We demonstrated that the M-MF achieves superior performance compared to state-of-the-art state estimators and consistently captures the uncertainty in multi-modal densities.

In future, we will evaluate the significance of the scaling parameter α and its impact on higher moments of the approximations. The effect of the mixture reduction techniques also will be investigated to achieve a better filter performance. Moreover, we will extend the M-MF to a forward-backward smoothing algorithm.

6. REFERENCES

- [1] Thomas Kailath, Ali H Sayed, and Babak Hassibi, *Linear Estimation*, vol. 1, Prentice Hall, 2000.
- [2] Brian D O Anderson and John B Moore, *Optimal Filtering*, Dover Publications, 2005.
- [3] Simon J Julier and Jeffrey K Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [4] Arnaud Doucet, Nando de Freitas, and Neil Gordon, *Sequential Monte Carlo methods in practice*, vol. 1, Springer New York, 2001.
- [5] Daniel Alspach and Harold Sorenson, “Nonlinear Bayesian estimation using Gaussian sum approximations,” *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.
- [6] Xiaodong Luo, Irene M Moroz, and Ibrahim Hoteit, “Scaled unscented transform Gaussian sum filter: Theory and application,” *Physica D: Nonlinear Phenomena*, vol. 239, no. 10, pp. 684–701, 2010.
- [7] Jayesh H Kotecha and Petar M Djuric, “Gaussian sum particle filtering,” *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, 2003.
- [8] Genshiro Kitagawa, “The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother,” *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 4, pp. 605–623, 1994.
- [9] Jason L Williams and Peter S Maybeck, “Cost-function-based Gaussian mixture reduction for target tracking,” in *Proceedings of the Sixth International Conference of Information Fusion*, 2003, vol. 2, pp. 1047–1054.
- [10] Andrew R Runnalls, “Kullback-Leibler approach to Gaussian mixture reduction,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, 2007.
- [11] Kittipat Kampa, Erion Hasanbelliu, and Jose C Principe, “Closed-form Cauchy-Schwarz PDF divergence for mixture of Gaussians,” in *Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2011, pp. 2578–2585.
- [12] Simo Särkkä, “Unscented Rauch–Tung–Striebel smoother,” *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 845–849, 2008.
- [13] Rudolph van der Merwe, Arnaud Doucet, Nando de Freitas, and Eric Wan, “The unscented particle filter,” in *Advances in Neural Information Processing Systems*, 2000, pp. 584–590.
- [14] Marc P Deisenroth and Henrik Ohlsson, “A general perspective on Gaussian filtering and smoothing: Explaining current and deriving new algorithms,” in *Proceedings of the American Control Conference*. IEEE, 2011, pp. 1807–1812.
- [15] Marc P Deisenroth, Ryan Turner, Marco Huber, Uwe D Hanebeck, and Carl E Rasmussen, “Robust filtering and smoothing with Gaussian processes,” *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1865–1871, 2012.