FURTHER IMPROVEMENT OF THE ADAPTIVE LEVEL OF DETAIL TRANSFORM: SPLITTING IN DIRECTION OF THE NONLINEARITY

Friedrich Faubel and Dietrich Klakow

Spoken Language Systems
Saarland University, Campus C7-1, D-66123 Saarbrücken, Germany
email: {friedrich.faubel, dietrich.klakow}@lsv.uni-saarland.de
web: www.lsv.uni-saarland.de

ABSTRACT

In earlier work, we have presented a novel approach to nonlinear, non-Gaussian tracking problems. The approach was based on keeping a bank of unscented Kalman filters, which were split and merged in order to adapt the level of detail of the filtering density according to the nonlinearity of the tracking problem. More recently, that approach has been refined and generalized to a general method for nonlinear transformations of Gaussian mixture random variables. Here, we further extend it by the following aspects: we consider splitting a Gaussian distribution into three components rather than two; and we show how splitting can be performed in direction of the nonlinearity, which in simulations gave a 25% reduction of the mean squared error, compared to the previous implementation of the split and merge unscented Gaussian mixture filter. In addition to that, we show how splitting can be implemented efficiently, through Cholesky downdates.

1. INTRODUCTION

We have recently proposed an adaptive Gaussian mixture filter [3] for nonlinear, non-Gaussian tracking problems. It is based on splitting Kalman filters in likely regions of state space and on merging them in unlikely ones, with the aim of adapting the level of detail of the filtering density according to the posterior probability of the modes. This approach was inspired by the resampling stage of particle filters [1] where a similar objective is achieved by multiplying and removing samples. Further, motivated by the fact that Kalman filters are optimal for linear, Gaussian problems, we introduced a split control technique [3] that prevents filters from being split if they operate in relatively linear regions of state space. It should be noted that splitting is beneficial in nonlinear regions as it decreases the variances and thereby the degree of nonlinearity to which the filters are subject. This idea was further refined in [2] where Gaussians are split based on both their weight in the mixture and the degree of nonlinearity. A related but theoretically more well-founded approach has been taken in [4] where the Gaussians to be split are determined based on the L^2 distance measure. That work, however, considered one-dimensional tracking problems only.

In [2, 3], multivariate Gaussian distributions are split in direction of the largest eigenvalue of the covariance matrix. That gives the greatest reduction in variance. However, it is not optimal as it disregards the fact that the reason for splitting is actually the nonlinearity, which the Gaussians are subject to during transformation. Hence, in this work, we investigate how Gaussians can be split in the direction of nonlinearity. In addition to that, we show how Gaussians can be split into mixtures of three components – rather than two [3] – and further give an efficient implementation of the splitting procedure, using Cholesky downdates. The latter is of interest, especially if the Kalman filters are implemented in their square root form [9], as in that case the covariance matrices are represented by Cholesky factors anyway.

The remaining part of this paper is organized as follows. Section 2 briefly reviews the unscented transform. Section 3 describes how Gaussian distributions can be split into mixtures of two and three components. In Section 4, we introduce the concept of splitting in direction of the nonlinearity, which is finally evaluated in experiments, in Section 5.

2. REVIEW OF THE UNSCENTED TRANSFORM

The unscented transform (UT) approximates a Gaussian probability distribution by a finite number of points, which are chosen in such a way that they have the same mean and covariance as the original distribution. This procedure was introduced by Julier and Uhlmann [6] in order to approximate the nonlinear transform $Y = f(X)$ of an n-dimensional Gaussian random variable X with distribution

$$p_X(x) = \mathcal{N}(x; \mu_X, \Sigma_X),$$

parametrized by the mean μ_X and the covariance matrix Σ_X. The point mass representation is based on the Cholesky decomposition $R^T R$ of the covariance matrix Σ_X and it is obtained as follows. Denoting the rows of R by R_i and defining $\lambda = n + \kappa$ for an arbitrary $\kappa \in \mathscr{R}$, the distribution of X can be represented by the weighted empirical distribution

$$\tilde{p}_X(x) = \sum_{i=0}^{2n} W_i \delta(x - \mathscr{X}_i) \tag{1}$$

where δ is the Dirac delta and where the points and weights, $\mathscr{X}_i$ and W_i, are given by

$$\begin{array}{rclcrcl} \mathscr{X}_0 & = & \mu_X & & W_0 & = & \kappa/\lambda \\ \mathscr{X}_{2i+1} & = & \mu_X + \sqrt{\lambda} R_i & & W_{2i+1} & = & 1/(2\lambda) \\ \mathscr{X}_{2i+2} & = & \mu_X - \sqrt{\lambda} R_i & & W_{2i+2} & = & 1/(2\lambda) \end{array} \tag{2}$$

$i = 0, \ldots, (n-1)$. Note that κ specifies how much weight is placed on the mean, $\mathscr{X}_0$. Setting κ to $1/2$ results in a

This work was supported by the German Research Foundation (DFG), under the research training network IRTG 715 "Language Technology and Cognitive Systems"; and by the Federal as well as State Governments of Germany, through the Cluster of Excellence for Multimodal Computing and Interaction (MMCI).

 ISSN 2076-1465

weight of $1/n$ for each of the points. Setting it to $3-n$ minimizes the error in the fourth moment [7]. Similar as in Monte Carlo methods, the weighted points $\mathscr{X}_i$ can be instantiated through the function f, $\mathscr{Y}_i = f(\mathscr{X}_i)$, which then in turn yields a weighted empirical distribution of Y:

$$\tilde{p}_Y(y) = \sum_{i=0}^{2n} W_i \delta(y - \mathscr{Y}_i). \tag{3}$$

Hence, a Gaussian approximation $\hat{p}_Y(y) = \mathcal{N}(Y; \hat{\mu}_Y, \hat{\Sigma}_Y)$ of the transformed distribution can be obtained by estimating the mean and covariance of Y in a maximum likelihood fashion:

$$\hat{\mu}_Y = \sum_{i=0}^{2n} W_i \mathscr{Y}_i, \quad \hat{\Sigma}_Y = \sum_{i=0}^{2n} W_i (\mathscr{Y}_i - \mu_Y)(\mathscr{Y}_i - \mu_Y)^T. \tag{4}$$

This is the unscented transform. For linear transforms it is exact – that is, the Gaussian fit is not an approximation but the true, transformed density. For nonlinear transforms its mean and covariance estimates are accurate up to the second order term of the Taylor series expansion [7]. In the latter case, the appropriateness of the Gaussian approximation can be determined by estimating the degree of nonlinearity as proposed in [2]:

$$\eta \triangleq \frac{1}{n} \sum_{i=0}^{n-1} \eta_i, \tag{5}$$

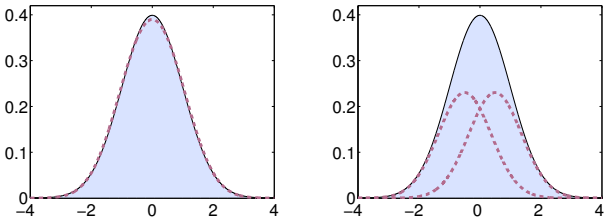
where η_i denotes the degree of nonlinearity of the i-th triple $\{\mathscr{X}_{2i+1}, \mathscr{X}_0, \mathscr{X}_{2i+2}\}$ of original points, calculated on the corresponding triple of transformed points $\{\mathscr{Y}_{2i+1}, \mathscr{Y}_0, \mathscr{Y}_{2i+2}\}$, as

$$\eta_i = \frac{1}{2} \|\mathscr{Y}_{2i+1} + \mathscr{Y}_{2i+2} - 2\mathscr{Y}_0\|^2. \tag{6}$$

This measure is based on the fact that each triple $\{\mathscr{X}_{2i+1}, \mathscr{X}_0, \mathscr{X}_{2i+2}\}$ forms a set of equidistant points on a line. Consequently, the degree of nonlinearity is calculated as the deviation from a linear fit of the transformed points (see [3] for a more thorough derivation).

3. SPLITTING GAUSSIAN DISTRIBUTIONS

By "splitting" a (multivariate) Gaussian distribution we mean approximating it by a mixture of Gaussian distributions with smaller variances. This can be achieved by using a splitting library, as in [4, 5], or by slightly displacing the means while adopting the covariance matrix of the original distribution [8, 4]. In this work, we take a different approach based on moment matching and symmetry. Moreover, we restrict ourselves to splitting in direction of eigenvectors. This avoids problems with – possibly resulting – indefinite covariance matrices; and it allows us to reduce the generally multivariate, n-dimensional case to splitting a standard normal distribution. The direction in which we split might be given by the eigenvector corresponding to the largest eigenvalue – that is, the direction of the largest variance – or by the eigenvector to which the direction of the nonlinearity is most similar (see Section 4 for details). In the following, we start with splitting the standard normal distribution into two and three components, respectively, in Sections 3.1 and 3.2. Then, in Section 3.3, we extend the splitting approach to the general, multivariate case.



(a) mixture *(b) components*

Figure 1: *Splitting into two Gaussians with a displacement of 0.5. The picture to the left shows the original distribution (solid line, highlighted area) along with the mixture of split components (dashed line). The picture to the right shows the individual components.*

3.1 Splitting into Two Components

In order to split the normal distribution $\mathcal{N}(x; 0, 1)$ into two components, $g_1(x)$ and $g_2(x)$, we first of all use its symmetry. The symmetry tells us that if we displace one of the Gaussians by ν from the origin the other Gaussian must be placed at $-\nu$. For the same reason, the two components must have the same mixture weight α and the same variance σ^2. This constrains the parameter optimization problem to finding the displacement as well as the variance of the two components,

$$g_1(x) = \mathcal{N}(x; \nu, \sigma^2), \quad g_2(x) = \mathcal{N}(x; -\nu, \sigma^2). \tag{7}$$

From the law of total probability it is clear that the mixture weights must be one half. Hence, splitting the normal distributions is tantamount to replacing it by the mixture

$$m(x) = 0.5 g_1(x) + 0.5 g_2(x). \tag{8}$$

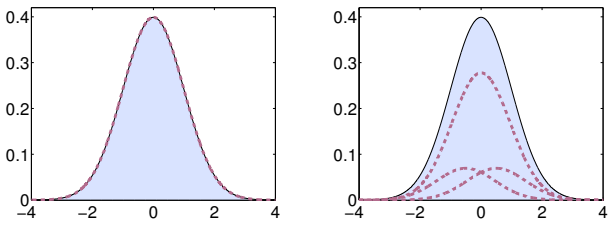
The second moment $\mathscr{E}_m\{x^2\} = \int x^2 m(x) dx$ of the mixture can be obtained by first using the linearity property of integration to get separate integrals over $g_1(x)$ and $g_2(x)$ and then performing a change of variables from x to $y = x - \nu$ and $y = x + \nu$, respectively, which yields: $\mathscr{E}_m\{x^2\} = \nu^2 + \sigma^2$. Subsequently matching the second moment of the mixture to that of the normal distribution, i.e. one, the variance can be expressed in dependency of the displacement:

$$\sigma^2 = 1 - \nu^2. \tag{9}$$

In order for this equation to be valid ν must be in the range $[-1, 1]$. Further, it can be shown that the absolute (L^1) error in the fourth moment is $2\nu^4$, which is clearly minimal for the trivial solution $\nu = 0$ and which monotonically increases with $|\nu|$ until it takes its maximum, 2, at $|\nu| = 1$. As a consequence of this result, we decided to keep ν as a parameter. A value of 0.5 seemed to give a good trade-off between displacement of components and accuracy of approximation, at least for the problems that we have been working on so far.

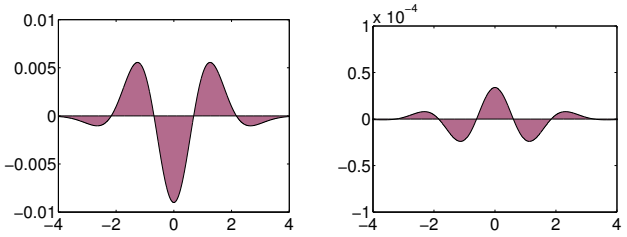
3.2 Splitting into Three Components

The splitting approach from the previous section can easily be extended to the case of splitting a Gaussian into three components, $g_1(x)$, $g_2(x)$ and $g_3(x)$. Making use of the symmetry of the normal distribution, we again displace the first Gaussian by ν, the second one by $-\nu$. The third Gaussian



(a) mixture *(b) components*

Figure 2: *Splitting into three Gaussians with a displacement of 0.5. The picture to the left shows the original distribution (solid line, highlighted area) along with the mixture of split components (dashed line). The picture to the right shows the individual components.*



(a) 2 Gaussians *(b) 3 Gaussians*

Figure 3: *Difference between original and split distributions for a displacement of 0.5. Not that the scale of the image to the right is one hundredth of that of the image to the left.*

is centered at zero as portrayed in Figure 2-b. Then, choosing a weight $\alpha \leq 0.5$ for each of the displaced components uniquely determines the weight of the center component, as $1-2\alpha$. Hence the mixture can be written

$$m(x) = \alpha g_1(x) + \alpha g_2(x) + (1-2\alpha) g_3(x). \tag{10}$$

The Gaussians in the mixture are further parameterized by the variance σ^2 of the displaced components, as well as the variance τ^2 of the center component. In the following we will assume that τ is equal to σ, which greatly simplifies the optimization problem in that it does not require matching the sixth moment of the distributions. After this simplification, the mixture components are:

$$\begin{aligned} &g_1(x) = \mathcal{N}(x; \nu, \sigma^2), \quad g_2(x) = \mathcal{N}(x; -\nu, \sigma^2), \\ &g_3(x) = \mathcal{N}(x; 0, \sigma^2). \end{aligned} \tag{11}$$

Similar to the case of two components, we will try to express α and σ^2 in dependence of the displacement ν. For that, we match the second and fourth moment of the mixture to those of the normal distribution:

$$\begin{aligned} \mathscr{E}_m\{x^2\} &= 2\alpha\nu^2 + \sigma^2 &&= 1 \\ \mathscr{E}_m\{x^4\} &= 2\alpha\nu^4 + 12\alpha\sigma^2\nu^2 + 3\sigma^4 &&= 3 \end{aligned}$$

Solving this system of equations and discarding the trivial solution $\sigma^2 = 1$ yields

$$\alpha = \frac{1}{6}, \quad \sigma^2 = 1 - \frac{1}{3}\nu^2. \tag{12}$$

These equations are valid for displacements ν in the range $[-\sqrt{3}, \sqrt{3}]$. The difference to the normal distribution is shown in Figure 3-b, for $\nu = 0.5$. Figure 3-a gives a comparison to the "splitting into two components" approach from the previous section. Furthermore, it can be shown that the absolute (L^1) error in the sixth moment is $\frac{2}{9}\nu^6$.

3.3 Multivariate Gaussian Distributions

Splitting a multivariate Gaussian distribution $\mathcal{N}(\mathbf{x}; \mu; \Sigma)$ in direction of an eigenvector can be reduced to splitting a standard normal distribution. For that, let $U^T \Lambda U$ be the eigendecomposition of the covariance matrix Σ, with a diagonal matrix Λ containing the eigenvalues λ_i and a unitary matrix U containing the corresponding eigenvectors $\mathbf{u}_i$:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}, \quad U = \begin{bmatrix} \boxed{\mathbf{u}_1^T} \\ \vdots \\ \boxed{\mathbf{u}_n^T} \end{bmatrix}.$$

Then using the exponentiation identity $\exp(x+y) = \exp(x) \cdot \exp(y)$ and the fact that the determinant of Σ can be factored as $\det(\Sigma) = \prod_i \lambda_i$, the probability density function can be written

$$p(\mathbf{x}) = \prod_{i=1}^{N} \underbrace{\mathcal{N}\left(\mathbf{u}_i^T \mathbf{x};\ \mathbf{u}_i^T \mu,\ \lambda_i\right)}_{\triangleq f_i(\mathbf{x})}. \tag{13}$$

Now, let $\tilde{g}_k(x) = \mathcal{N}(x; \tilde{\mu}_k, \tilde{\sigma}_k^2)$, $k = 1, \ldots, K$, be the components resulting from a split of the standard normal distribution. Then, splitting the multivariate distribution $p(\mathbf{x})$ in direction of the j-th eigenvector can be achieved by performing the following steps:

1. Scaling the components, $\tilde{g}_k(x)$, by $1/\sqrt{\lambda_j}$ in x-direction in order to match the variance λ_j of $p(\mathbf{x})$ in direction of the eigenvector $\mathbf{u}_j$.
2. Rotating the resulting, rescaled components with distribution $\bar{g}_k(x) = \mathcal{N}(x; \sqrt{\lambda_j}\tilde{\mu}_k, \lambda_j \tilde{\sigma}_k^2)$ into $\mathbf{u}_j$ and then adding the mean $\mathbf{u}_j^T \mu$, which gives:

$$\tilde{f}_{j,k}(\mathbf{x}) \triangleq \mathcal{N}\left(\mathbf{u}_j^T \mathbf{x};\ \mathbf{u}_j^T \mu + \sqrt{\lambda_j}\tilde{\mu}_k,\ \lambda_j \tilde{\sigma}_k^2\right).$$

3. Replacing $f_j(\mathbf{x})$ in (13) by $\tilde{f}_{j,k}(\mathbf{x})$ for $k = 1, \ldots, K$ in order to obtain the split components of the multivariate distribution:

$$g_k(\mathbf{x}) = \left(\prod_{\substack{i=1 \\ i \neq j}}^{n} f_i(\mathbf{x})\right) \tilde{f}_{j,k}(\mathbf{x}), \tag{14}$$

As $\mathbf{u_i}^T(\mu + \sqrt{\lambda_j}\tilde{\mu}_k \mathbf{u}_j)$ is $(\mathbf{u_i}^T \mu)$ for $i \neq j$ and $(\mathbf{u}_j^T \mu + \sqrt{\lambda_j}\tilde{\mu}_k)$ for $i = j$, the mean, μ_k, of the k-th component, g_k, can obviously be recovered as

$$\mu_k = \mu + \sqrt{\lambda_j}\tilde{\mu}_k \mathbf{u}_j. \tag{15}$$

Splitting into Two Gaussians		
$\omega_1 = \frac{1}{2}$	$\mu_1 = \mu + \nu\sqrt{\lambda}\mathbf{u}$	$\Sigma_1 = \Sigma - \nu^2\lambda\mathbf{u}\mathbf{u}^T$
$\omega_2 = \frac{1}{2}$	$\mu_2 = \mu - \nu\sqrt{\lambda}\mathbf{u}$	$\Sigma_2 = \Sigma - \nu^2\lambda\mathbf{u}\mathbf{u}^T$
Splitting into Three Gaussians		
$\omega_1 = \frac{1}{6}$	$\mu_1 = \mu + \nu\sqrt{\lambda}\mathbf{u}$	$\Sigma_1 = \Sigma - \frac{1}{3}\nu^2\lambda\mathbf{u}\mathbf{u}^T$
$\omega_2 = \frac{1}{6}$	$\mu_2 = \mu - \nu\sqrt{\lambda}\mathbf{u}$	$\Sigma_2 = \Sigma - \frac{1}{3}\nu^2\lambda\mathbf{u}\mathbf{u}^T$
$\omega_3 = \frac{4}{6}$	$\mu_3 = \mu$	$\Sigma_3 = \Sigma - \frac{1}{3}\nu^2\lambda\mathbf{u}\mathbf{u}^T$

Table 1: *Mixture parameters for splitting $\mathcal{N}(\mu,\Sigma)$ into two and three Gaussians with displacement ν in the direction of eigenvector $\mathbf{u}$ with corresponding eigenvalue λ.*

The corresponding covariance matrix Σ_k can be obtained by expressing it by means of its eigenvectors and eigenvalues:

$$\begin{aligned}\Sigma_k &= \sum_{\substack{i=1\\ i\neq j}}^{n} \lambda_i \mathbf{u}_i\mathbf{u}_i^T + \tilde{\sigma}_k^2\lambda_j\mathbf{u}_j\mathbf{u}_j^T \\ &= \underbrace{\sum_{i=1}^{n} \lambda_i \mathbf{u}_i\mathbf{u}_i^T}_{=\Sigma} - \left(1-\tilde{\sigma}_k^2\right)\lambda_j\mathbf{u}_j\mathbf{u}_j^T \end{aligned} \tag{16}$$

This is a simple rank-1 downdate. Hence, given the Cholesky factor of Σ is available – as is the case for the square root implementation of the unscented transform [10] – the Cholesky factor of Σ_k can efficiently be obtained through a Cholesky downdate. Table 1 concludes this section by explicitly giving the mixture weights ω_k, means μ_k and covariance matrices Σ_k for splitting into two and three Gaussians.

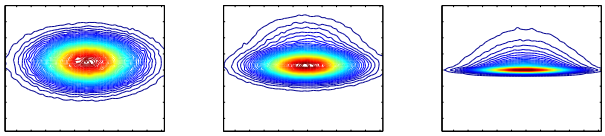
4. SPLITTING IN DIRECTION OF THE NONLINEARITY

In [5] and [3], splitting was performed in the direction of the largest variance, given by the eigenvector corresponding to the largest eigenvalue of the covariance matrix. Splitting in this direction obviously gives the greatest reduction in variance. However, it might not be optimal as the reason for splitting is the nonlinearity that the distribution is subject to during transformation. To illustrate this problem, consider the following example of a transformation $f(X)$ of a Gaussian random variable X with

$$\mu_X = \begin{bmatrix}0\\3\end{bmatrix}, \quad \Sigma_X = \begin{bmatrix}9 & 0\\ 0 & 1\end{bmatrix}, \quad f\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1\\x_2^2\end{bmatrix}.$$

For this example, the adaptive level of detail transform (ALoDT) from [2] cannot improve over the unscented transform – at least not in the first couple of iterations. That is because the distribution of X is split in x_1-direction, where the variance is largest but where f is linear. From Figure 4 it is obvious that the transformed distribution portrayed in 4-(a) is still "almost" Gaussian while the true distribution shown in Figure 4-(c) is strongly non-Gaussian. Therefore, we propose splitting in direction of the nonlinearity, which we perform as follows. Bearing in mind that, in the unscented transform, the i-th triple $\{\mathcal{X}_{2i+1}, \mathcal{X}_0, \mathcal{X}_{2i+2}\}$ of points forms a set of equidistant points on a line, the nonlinearity in direction

$$\phi_i = \frac{\mathcal{X}_{2i+1} - \mathcal{X}_0}{\|\mathcal{X}_{2i+1} - \mathcal{X}_0\|} \tag{17}$$



(a) ALoDT-4[v] *(b) ALoDT-4[n]* *(c) true density*

Figure 4: *Contour plots of the transformed distributions obtained with the ALoDT using 4 Gaussians, for splitting in direction of the largest variance [v] and splitting in direction of the nonlinearity [n].*

can obviously be approximated by the degree of nonlinearity, η_i, associated with these points. Consequently, the direction of nonlinearity can be defined as the eigenvector ψ corresponding to the largest eigenvalue of

$$\Psi = \sum_{i=0}^{n-1} \eta_i\phi_i\phi_i^T. \tag{18}$$

That is the direction in which the nonlinearity is strongest. As a computationally less demanding alternative, the direction of nonlinearity can be approximated as the average over the ϕ_i, weighted with the corresponding η_i:

$$\psi' = \frac{\sum_{i=0}^{n-1}\eta_i\phi_i}{\left\|\sum_{i=0}^{n-1}\eta_i\phi_i\right\|}. \tag{19}$$

For implementing a split in direction of the nonlinearity, it should be noted that splitting a Gaussian distribution in an arbitrary direction ψ turns out to be difficult unless ψ coincides with one of the principal axes of the covariance matrix. Hence, we split in the direction of that eigenvector $\mathbf{u}_i$, which ψ is most similar to, i.e. the one for which $\mathbf{u}_i^T\psi$ is maximal.

5. EXPERIMENTS

In order to evaluate the performance of the proposed extensions, we performed a series of simulations, in which a maneuvering object was tracked based on sensor measurements. In these simulations, the object moved along the synthetic trajectory portrayed in Figure 5,

$$x_t = 10\left[\sin(s_t)+1 \quad (\cos(s_t)+1)\sin(\tfrac{s_t}{2}) \quad \cos(2s_t)s_t\right]^T,$$

$s_t = \frac{4\pi t}{500}$, $t = 1,\ldots,500$, and was observed by virtual sensors located at $[0\ 0\ 0]^T$ providing measurements in polar coordinates. Additive measurement noise was simulated from a Gaussian, whose means and covariances were chosen at random – once for each of the 50 experiments performed. The average variance was 0.073 for the distance, 0.0031 and 0.0044 for the angles. As a process model we used a simple, zeroth-order linear dynamic model, $x_t = x_{t-1} + w_t$, with zero-mean Gaussian process noise, w_t, whose covariance matrix was estimated on the synthetic trajectory and further scaled by a factor of two in order to increase stability. At time $t = 1$ the filters were initialized with a Gaussian distribution around the true state, x_1, with process noise covariance.

Table 2 shows mean squared errors (MSE)s for the split and merge unscented Gaussian mixture filter [3] using the splitting priority from [2] as a splitting criterion. The numbers are averaged over 50 simulations and correspond to 500

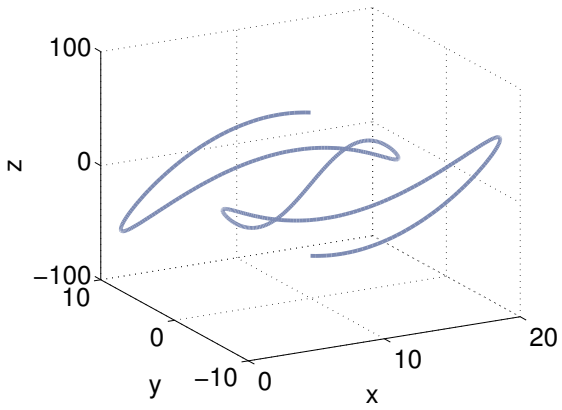


Figure 5: *Trajectory used in the simulations.*

splitting method	#Gaussians 2	4	8	16	32	64
2 components	779	774	768	765	764	762
3 components	733	738	739	736	735	734

Table 2: *MSE / track averaged over 50 runs of the split and merge unscented Gaussian mixture filter, for splitting into two and three Gaussians. Here, splitting was performed in direction of the largest eigenvalue of the covariance matrix.*

direction of split	#Gaussians 2	4	8	16	32	64
nonlinearity#1	935	770	688	632	621	608
nonlinearity#2	1104	825	703	628	595	573

Table 3: *MSE / track averaged over 50 runs of the split and merge unscented Gaussian mixture filter, for splitting in direction of the nonlinearity*

point trajectories. The first row of the table shows results for splitting into two components, in dependence of the number of filters (#Gaussians). With 2 filters, the MSE – being 927 for a single unscented Kalman filter – could be greatly reduced. Further increasing the number of filters just slightly improved the result. For splitting into three components (second row of the table), the MSE was lower in general but it behaved less consistently, which might have to do with the fact that, in the merge stage, we merge the filters successively in pairs. As a consequence of this, we only considered splitting into two components in the following.

Table 3 shows the mean squared errors we obtained by splitting in direction of the nonlinearity rather than in direction of the largest eigenvalue of the covariance matrix. In row one, the direction of nonlinearity was estimated as the eigenvector corresponding to the maximum eigenvalue of (18). In row two, it was estimated according to (19). The best result was obtained in row two, with 64 filters. In that case, the MSE was 573, which is almost 40% lower than that of the unscented Kalman filter and roughly 25% lower than the best result obtained with splitting in direction of the largest variance. For a low (≤ 4) number of filters, the proposed approach failed to give improvements. With just two filters it even performed worse than the UKF. This problem seems to be a consequence of the merge stage. The point is that splitting in direction of the nonlinearity can result in more distinct Gaussian components of the transformed distribution, especially if a lower number of filters is used. Then, if two distinct Gaussians are merged in the subsequent merge stage [3] the variances might be "blown up", which can have a very detrimental effect. As we have actually observed this problem with various Gaussian mixture filters – ones that deal with data association and ones that deal with nonlinearities – this might well be the next big thing to tackle.

6. CONCLUSIONS

We have demonstrated how Gaussians can be split in the direction of nonlinearity. In addition to that, we have shown how multivariate Gaussian distributions can be split into mixtures of two and three components, respectively, and analyzed the errors introduced by the resulting Gaussian mixture approximations. The effectiveness of the proposed extensions with respect to improving the accuracy of nonlinear tracking problems has been verified in simulations.

REFERENCES

[1] S. 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*, 50(2):174–188, Feb. 2002.

[2] F. Faubel and D. Klakow. An adaptive level of detail approach to nonlinear estimation. *Proc. ICASSP*, Mar. 2010.

[3] F. Faubel, J. McDonough, and D. Klakow. The split and merge unscented Gaussian mixture filter. *IEEE Signal Process. Lett.*, 16(9):786–789, Sept. 2009.

[4] U. Hanebeck, K. Briechle, and A. Rauh. Progressive Bayes: A new framework for nonlinear state estimation. *Proc. of SPIE, AeroSense Symposium*, 5099:256–267, May 2003.

[5] M. Huber, T. Bailey, H. Durrant-Whyte, and U. Hanebeck. On entropy approximation for Gaussian mixture random vectors. In *Proc. of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 2639–2644. IEEE, Aug. 2008.

[6] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. *Proc. AeroSense*, pages 182–193, Apr. 1997.

[7] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. IEEE*, 92(3):401–422, Mar. 2004.

[8] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. Split and merge EM algorithm for improving Gaussian mixture density estimates. *Journal of VLSI Signal Processing Systems*, 26(1-2):133–140, 2000.

[9] R. van der Merwe and E. Wan. Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. *Proc. ICASSP*, 6(IV):701–704, Apr. 2003.

[10] R. van der Merwe and E. A. Wan. The square root unscented Kalman filter for state and parameter-estimation. *Proc. ICASSP*, 6:3461 – 3464, May 2001.