Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation

Marco F. Huber * **Uwe D. Hanebeck** *

* *Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany; e-mail: marco.huber@ieee.org, uwe.hanebeck@ieee.org.*

Abstract: In this paper, a Gaussian filter for nonlinear Bayesian estimation is introduced that is based on a deterministic sample selection scheme. For an effective sample selection, a parametric density function representation of the sample points is employed, which allows approximating the cumulative distribution function of the prior Gaussian density. The computationally demanding parts of the optimization problem formulated for approximation are carried out off-line for obtaining an efficient filter, whose estimation quality can be altered by adjusting the number of used sample points. The improved performance of the proposed Gaussian filter compared to the well-known unscented Kalman filter is demonstrated by means of two examples.

1. INTRODUCTION

The Bayesian estimation framework generally allows recursively processing the state of a nonlinear dynamic system. However, the resulting probability density functions representing the system state cannot be calculated analytically for arbitrary systems and noise types. Furthermore, the type of density changes and the complexity increases over time. A well-known exception exists for linear systems corrupted by Gaussian noise. Here, the Kalman filter provides exact solutions (Kalman (1960)).

A large variety of estimators has been developed for calculating approximate solutions in case of nonlinear systems. Particle filters for example employ Monte Carlo methods in order to obtain a sample or particle representation of the true density (Arulampalam et al. (2002)). Besides that, estimators employing analytic density types like Gaussian mixtures (Huber et al. (2007)) or Edgeworth series (Challa et al. (2000)) often provide accurate results. However, their computational complexity increases exponentially with the dimension of the system state.

More practical algorithms like the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) focus on approximating mean and covariance of density functions. While the EKF uses linearization to apply the Kalman filter equations to nonlinear systems (Simon (2006)), the UKF offers increased higher-order accuracy by using a deterministic sampling approach (Julier and Uhlmann (2004), Wan and van der Merwe (2000)), which is comparable to stochastic linearization (Lefebvre et al. (2002)). Extensions of the UKF modify the selection of the sample set to further improve the estimation accuracy (see e.g. Wu et al. (2004), Tenne and Singh (2003)). However, adapting parameters gets more involved or scaling is not guaranteed.

While the previously mentioned Kalman filter derivatives only take the lower-order statistics of the state into account, the sample selection scheme of the Gaussian filter proposed in this paper also considers the shape in terms of the distribution function of the prior Gaussian density. In recursive state estimation this incorporation of shape information leads automatically to a more accurate consideration of higher-order moments, especially in cases of near Gaussian posterior densities. Together with a freely adjustable number of sample points an improved estimation accuracy is the consequence. To provide shape approximation under the constraints of exactly capturing mean and covariance, an optimization problem is formulated, where the sample points are interpreted as analytic density function. This is different from the unscented Kalman filter or particle filters, where the sample points are not chosen in order to explicitly incorporate shape information and higher-order moments, respectively. The solution of the optimization problem can be calculated off-line and is extended to multivariate densities without suffering from the exponential increase of complexity.

In the next section, the problem formulation is given. The remainder of the paper is structured as follows: Sec. 4 addresses the optimization problem and its solution, while in Sec. 5 its multivariate extension is treated. The Gaussian filter is derived in Sec. 6 and its performance is demonstrated by means of simulations in Sec. 7. The paper closes with conclusions and an outlook to future work.

2. PROBLEM FORMULATION

In this paper, nonlinear discrete-time dynamic systems given in explicit form

$$\begin{aligned} \underline{\boldsymbol{x}}_{k+1} &= \underline{a}_k(\underline{\boldsymbol{x}}_k, \underline{u}_k, \underline{\boldsymbol{w}}_k) \\ \underline{\boldsymbol{y}}_k &= \underline{h}_k(\underline{\boldsymbol{x}}_k, \underline{\boldsymbol{v}}_k) \end{aligned} \tag{1}$$

are considered, where $\underline{a}_k(\cdot)$ and $\underline{h}_k(\cdot)$ are vector-valued functions that are assumed to be known. Furthermore, $\underline{\boldsymbol{x}}_k$ is the N-dimensional system state at time step k, $\underline{u}_k$ is the

known system input, $\underline{\boldsymbol{y}}_k$ is the measurement vector, and $\underline{\boldsymbol{w}}_k$, $\underline{\boldsymbol{v}}_k$ are white noise vectors acting upon the system.

Predicting the system state $\underline{\boldsymbol{x}}_k$ and the measurement $\underline{\boldsymbol{y}}_k$ by means of the nonlinear functions $\underline{a}_k(\,\cdot\,)$ and $\underline{h}_k(\,\cdot\,)$ can generally be expressed as the nonlinear transformation

$$\underline{\boldsymbol{y}} = \underline{g}(\underline{\boldsymbol{x}}) \;, \tag{2}$$

where $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ are random vectors and $\underline{g}(\,\cdot\,)$ is an arbitrary nonlinear function.

With (2), the random vector $\underline{\boldsymbol{x}}$ is nonlinearly mapped to the random vector $\underline{\boldsymbol{y}}$. Generally, calculating the density function or the statistics of $\underline{\boldsymbol{y}}$ cannot be carried out in closed form. To avoid numerical solutions and thus for performing an efficient estimation of $\underline{\boldsymbol{y}}$, the exact density has to be approximated in an adequate manner. Furthermore, the number of parameters characterizing the approximation should be at a constant level and the user should be able to adjust the quality as well as the computational demand of the approximation.

Instead of directly processing the true density function $\tilde{f}_x(\underline{x})$ of $\underline{\boldsymbol{x}}$ or its moments, which is computationally demanding, imprecise or even impossible in cases where the nonlinear transformation is not given in an analytic form, the goal is now to efficiently determine an accurate sample representation of $\tilde{f}_x(\underline{x})$. Sample points can be easily propagated through the nonlinear transformation (2) and in turn allow efficiently approximating the true density function of $\underline{\boldsymbol{y}}$ by a Gaussian density $f_y(\underline{y}) = \mathcal{N}(\underline{y} - \underline{\hat{y}}, \mathbf{C}_y)$. Therefore, only the two moments mean $\underline{\hat{y}}$ and covariance matrix $\mathbf{C}_y$ need to be calculated, which can be done with polynomial complexity with respect to the dimension of $\underline{\boldsymbol{y}}$. Furthermore, given only these two moments, a Gaussian density is the Kullback-Leibler divergence minimizing or entropy maximizing density for approximating $\tilde{f}_y(\underline{y})$ (Catlin (1989)).

3. DETERMINISTIC SAMPLING

The achievable accuracy of the Gaussian approximation for $\underline{\boldsymbol{y}}$ strongly depends on the strategy of determining the sample representation of $\underline{\boldsymbol{x}}$. The unscented Kalman filter (UKF, see Julier and Uhlmann (2004)) for instance uses a minimal fixed-size set consisting of $2N + 1$ sample points with corresponding weights that exactly captures the mean $\underline{\hat{x}}$ and the covariance matrix $\mathbf{C}_x$ of $\underline{\boldsymbol{x}}$. For any nonlinear transformation, the propagated sample points capture the posterior mean and covariance matrix accurately up to the second order of the corresponding Taylor series expansion.

For further improving the accuracy, the set of sample points used for the Gaussian filter proposed in this paper is not restricted to a fixed size. Increasing the number of sample points generates several advantages. Since more sample points are propagated, more information of the nonlinear transformation is captured. This leads to improved and more robust estimates. Furthermore, the Gaussian filter is well applicable to a larger number of nonlinear transformations.

To gain these advantages, as much information about $\underline{\boldsymbol{x}}$ as possible has to be incorporated when determining the sample points. While the UKF only considers mean and covariance, the sample selection scheme derived in the following section is based on directly approximating the distribution function of the prior Gaussian. This is motivated by the fact that an accurate approximation of the distribution function automatically approximates higher-order moments. These moments in turn have an impact on higher-order terms of the Taylor series of the nonlinear function, which leads to improved estimation results.

For an accurate approximation, the key idea is now to reformulate the approximation problem corresponding to determining the sample points as an optimization problem by minimizing a certain distance measure $G(\,\cdot\,)$ between the Gaussian and a convenient analytic sample representation under the constraints that mean $\underline{\hat{x}}$ and covariance $\mathbf{C}_x$ of $\underline{\boldsymbol{x}}$ are captured exactly. This is different from the UKF or particle filters, where no distance measure is employed.

4. ONE-DIMENSIONAL APPROXIMATION

At first, only one-dimensional transformations $\boldsymbol{y} = g(\boldsymbol{x})$ are considered, where the random variable $\boldsymbol{x}$ is characterized by mean $\hat{x}$ and variance σ_x^2. Without loss of generality, it can be assumed that $\hat{x} = 0$ and $\sigma_x^2 = 1$, which leads to a *standard* Gaussian density $\tilde{f}_x(x) = \mathcal{N}(x, 1)$. This restriction is justified, since every Gaussian can be transformed into a standard Gaussian density.

4.1 Dirac Mixture

To provide an approximation in both density and moments of $\boldsymbol{x}$, an analytic and parametric form for representing the samples in terms of a so-called *Dirac mixture* density function

$$f_x(x, \underline{\eta}) = \sum_{i=1}^{L} \omega_i \cdot \delta(x - \mu_i) \;, \tag{3}$$

is employed, which is a weighted sum of L Dirac delta functions $\delta(x - \mu_i)$ located at sample positions μ_i. The parameter vector $\underline{\eta}$ comprises the weighting coefficients ω_i and the sample positions μ_i.

To reduce the number of parameters of $f_x(x, \underline{\eta})$ to be adjusted for approximation, in this paper equal weighting coefficients ω_i are assumed, i.e., $\omega_i = 1/L$. Assuming that the positions μ_i of the Dirac delta functions are sorted, i.e.,

$$\mu_1 < \mu_2 < \ldots < \mu_L \;,$$

capturing the mean $\hat{x}$ can easily be guaranteed by placing the Dirac delta functions symmetrically around $\hat{x}$, i.e.,

$$\mu_{L+1-i} = 2\hat{x} - \mu_i = -\mu_i$$

for $i = 1, 2, \ldots, \bar{L}$ with $\bar{L} := \lceil \frac{L-1}{2} \rceil$. This further reduces the length of $\underline{\eta}$. If L is odd, the center Dirac delta function is fixed at the mean $\hat{x}$, i.e., we set $\mu_{\bar{L}+1} = \hat{x} = 0$. Finally, the parameter vector $\underline{\eta}$ is given by

$$\underline{\eta} = [\mu_1, \mu_2, \ldots, \mu_{\bar{L}}]^{\mathrm{T}} \;.$$

4.2 Distance Measure

Typical measures quantifying the distance between densities, like the Kullback-Leibler divergence (Kullback and

Leibler (1951)) or the squared integral measure (Izenman (1991)), cannot be applied directly due to the used Dirac delta functions in (3). Thus, the corresponding cumulative distribution functions are employed instead. The distribution function of the true density $\tilde{f}_x(x)$ can be written as

$$\tilde{F}_x(x) = \int_{-\infty}^{x} \tilde{f}_x(t)\,\mathrm{d}t = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\hat{x}}{\sqrt{2}\sigma_x}\right)\right) ,$$

where $\operatorname{erf}(\,\cdot\,)$ is the error function, while the distribution function corresponding to the Dirac mixture $f_x(x,\underline{\eta})$ is given by

$$F_x(x,\underline{\eta}) = \sum_{i=1}^{L} \omega_i \cdot \mathrm{H}(x-\mu_i) ,$$

where $\mathrm{H}(\,\cdot\,)$ is the Heaviside step function

$$\mathrm{H}(x-\mu) = \begin{cases} 1 , & x > \mu \\ \frac{1}{2} , & x = \mu \\ 0 , & \text{otherwise} \end{cases}$$

at position μ.

As distance measure we employ

$$G(\underline{\eta},\lambda) = \frac{1}{2}\int_{\mathbb{R}} \Big(\tilde{F}_x(x) - F_x(x,\underline{\eta})\Big)^2 \mathrm{d}x + \lambda\Big(\frac{1}{L}\sum_{i=1}^{L}\mu_i^2 - \sigma_x^2\Big) . \tag{4}$$

The first term in (4) is the so-called *Cramér-von Mises distance* (Boos (1981), Schrempf et al. (2006)) quantifying the divergence between the distribution functions and λ in the second term is a *Lagrange multiplier*. By utilizing the Lagrange multiplier approach, exactly capturing the variance σ_x^2 is guaranteed if $L \geq 2$, while exactly capturing the mean is guaranteed by the symmetric positioning of the sample points.

4.3 Solution

To minimize the distance measure with respect to $\underline{\eta}$ and λ, the necessary conditions for a minimum $\partial G(\underline{\eta},\lambda)/\partial\underline{\eta} = \underline{0}$ and $\partial G(\underline{\eta},\lambda)/\partial\lambda = 0$ have to be satisfied. Utilizing the sifting property of the Dirac delta function, the partial derivative of $G(\,\cdot\,)$ with respect to sample position μ_i yields

$$\frac{\partial G(\underline{\eta},\lambda)}{\partial\mu_i} = \frac{1}{L}\Big(\tilde{F}_x(\mu_i) - F_x(\mu_i,\underline{\eta}) + 2\lambda\mu_i - \big(\tilde{F}_x(-\mu_i) - F_x(-\mu_i,\underline{\eta})\big)\Big) , \tag{5}$$

for $i = 1,\ldots,\bar{L}$. With the identities

$$\begin{aligned} F_x(\mu_i,\underline{\eta}) &= \tfrac{2i-1}{2L} , \\ \tilde{F}_x(-\mu_i) &= 1 - \tilde{F}_x(\mu_i) , \\ F_x(-\mu_i,\underline{\eta}) &= 1 - F_x(\mu_i,\underline{\eta}) , \end{aligned}$$

and by setting (5) equal zero we obtain

$$\tilde{F}_x(\mu_i) - \frac{2i-1}{2L} + \lambda\mu_i = 0 , \tag{6}$$

The partial derivative of $G(\,\cdot\,)$ with respect to λ yields

$$\sum_{i=1}^{\bar{L}} \mu_i^2 - \tfrac{L}{2}\sigma_x^2 = 0 . \tag{7}$$

The resulting system of nonlinear equations comprising (6) and (7) is square, i.e., the number of equations equals the number of unknowns. For determining a root, an iterative root finding algorithms can be applied, where we use the trust-region dogleg method (Powell (1970)). As initial solution we choose

$$\begin{aligned} \mu_i &= \sqrt{2}\,\mathrm{erf}^{-1}\left(\tfrac{2i-1-L}{L}\right) , \; i = 1,\ldots,\bar{L} , \\ \lambda &= 0 , \end{aligned}$$

which is the optimal solution of minimizing $G(\,\cdot\,)$ without considering the variance constraint (Schrempf et al. (2006)).

4.4 Off-line Approximation

The resulting sample positions μ_i that minimize (4) are valid for a standard Gaussian density. For arbitrary Gaussians it is beneficial to split the approximation task into an off-line and an on-line part, instead of solving a similar optimization problem on-line. In doing so, the sample approximation derived at the last paragraphs is performed off-line for a desired number of samples (see Tab. 1 for several approximations). Then, for on-line estimation, these samples have to be scaled and shifted according to

$$\hat{x} + \sigma_x \cdot \mu_i ,$$

where $\hat{x}$ and σ_x are now arbitrary means and standard deviations, respectively. This transformation leads to an on-line approximation of any Gaussian density without impairing the approximation quality. Furthermore, the on-line performance for state estimation is drastically increased.

Table 1. Sample Positions for several numbers of samples.

L	μ_1	μ_2	μ_3
3	-1.2247	–	–
5	-1.4795	-0.5578	–
7	-1.6346	-0.8275	-0.3788

5. EXTENSION TO MULTIDIMENSIONAL CASE

In the following, vector-valued nonlinear transformations as in (2) are considered, i.e., the multivariate random vector $\underline{\boldsymbol{x}} \in \mathbb{R}^N$ with mean vector $\underline{\hat{x}}$ and covariance matrix $\mathbf{C}_x$ is mapped to the random vector $\underline{\boldsymbol{y}}$. The goal is to approximate the Gaussian density $\tilde{f}_x(\underline{x}) = \mathcal{N}(\underline{x}-\underline{\hat{x}},\mathbf{C}_x)$ by utilizing the approximation approach proposed before.

5.1 Reduction to One-Dimensional Case

One way to do so, is to directly extend the optimization problem to multivariate Gaussians. Although this extension would work, it suffers from a computational load increasing with the dimension of $\underline{\boldsymbol{x}}$, where multidimensional integrals have to be evaluated numerically. Instead, a more efficient but suboptimal way is to reduce the N-dimensional optimization problem to N one-dimensional optimization problems.

For a multivariate standard Gaussian $\tilde{f}_x(\underline{x}) = \mathcal{N}(\underline{x},\mathbf{I})$ with zero mean and covariance matrix $\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, the univariate marginal density of dimension i is given by

$$\mathcal{N}(x_i,1) = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} \mathcal{N}(\underline{x},\mathbf{I})\,\mathrm{d}x_1\cdots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\cdots\mathrm{d}x_N .$$

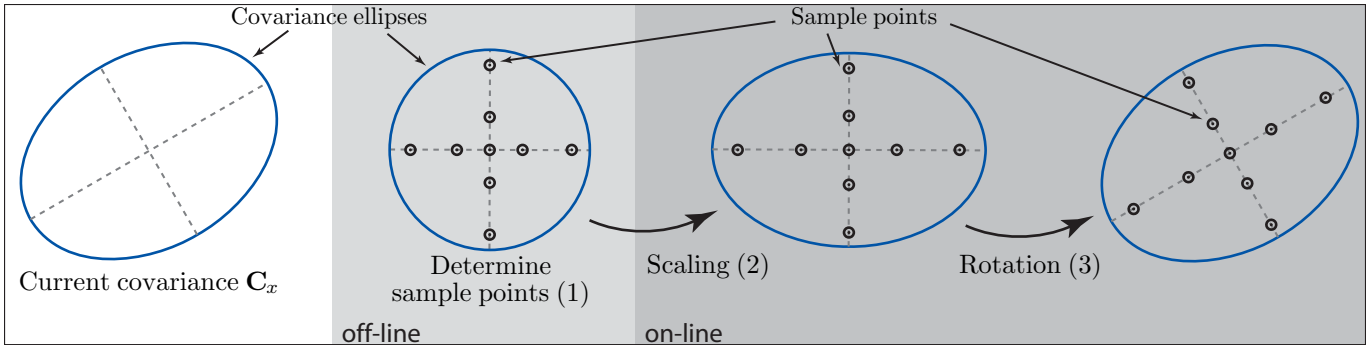

Fig. 1. For applying the one-dimensional Dirac mixture approximation to the multivariate Gaussian illustrated on the left by its covariance ellipses, at first the sample points for each dimension have to be determined (1). Performing the scaling (2) and rotation (3) operations completes the approximation.

Thus, we can apply the approach described in Sec. 4 dimension by dimension to each univariate Gaussian.

5.2 Calculating the Sample Points

This procedure corresponds to placing the samples points along the principal axis of $\mathcal{N}(\underline{x}, \mathbf{I})$ (see Fig. 1). To approximate any arbitrary Gaussian $\tilde{f}_x(\underline{x}) = \mathcal{N}(\underline{x} - \hat{\underline{x}}, \mathbf{C}_x)$, the resulting sample points are scaled and rotated by employing the eigenvalue decomposition

$$\mathbf{C}_x = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{T}} \ ,$$

where $\mathbf{V}$ is the orthogonal matrix of eigenvectors and $\mathbf{D}$ is a diagonal matrix of eigenvalues of $\mathbf{C}_x$. Thus, calculating the multivariate samples $\underline{\mu}_{i,j}$ by means of the component samples $\mu_{i,j}$ is carried out as

$$\begin{aligned} \underline{\mu}_{i,j} &= \hat{\underline{x}} + \mathbf{V}\sqrt{\mathbf{D}} \cdot \mu_{i,j} \cdot \underline{e}_j \ , \\ \omega_{i,j} &= \tfrac{1}{N \cdot L} \ , \end{aligned} \tag{8}$$

where index i, j indicates the i-th sample used for dimension j, where L samples are used for each dimension. Hence, $i = 1, \ldots, L$ and $j = 1, \ldots, N$. $\omega_{i,j}$ are the corresponding weighting coefficients and $\underline{e}_j = [0, \ldots, 0, 1, 0, 0, \ldots, 0]^{\mathrm{T}}$ is the canonical unit vector, where only element j is one. In case of L being odd, the mean vector $\hat{\underline{x}} = \underline{\mu}_{\bar{L}+1,j}$ appears N times in (8). To avoid overestimating the mean, the weights have to be adapted according to

$$\omega_{i,j} = \tfrac{1}{1+N\cdot(L-1)} \cdot \begin{cases} \frac{1}{N} \ , & i = \bar{L}+1 \\ 1 \ , & \text{otherwise} \end{cases} .$$

Altogether, in analogy to the scalar case approximating a multivariate standard Gaussian is performed off-line, while scaling and rotation are on-line operations. The Dirac mixture used for approximation can be written as

$$f_x(\underline{x}, \underline{\eta}) = \sum_{i,j} \omega_{i,j} \cdot \delta(\underline{x} - \underline{\mu}_{i,j}) \ .$$

For improved readability, from now on the index i, j is substituted by l, where $l = 1, 2, \ldots, L \cdot N$.

6. NONLINEAR ESTIMATOR

The sample points (8) are now propagated through the nonlinear transformation (2) according to

$$\underline{\mu}_l^y = \underline{g}(\underline{\mu}_l) \ , \quad \forall l \ .$$

With the resulting sample points $\underline{\mu}_l^y$ and the sample weights ω_l, the mean and covariance matrix representing $\underline{\boldsymbol{y}}$ are determined as

$$\begin{aligned} \hat{\underline{y}} &= \sum_l \omega_l \cdot \underline{\mu}_l^y \ , \\ \mathbf{C}_y &= \frac{1}{L} \sum_l \left(\underline{\mu}_l^y - \hat{\underline{y}}\right) \cdot \left(\underline{\mu}_l^y - \hat{\underline{y}}\right)^{\mathrm{T}} \ . \end{aligned} \tag{9}$$

It is important to note that (9) does not exactly correspond to the sample covariance as the factor $1/L$ is used instead of the sample weights ω_l. This provides an unbiased estimate of $\mathbf{C}_y$ for the employed sample representation.

For applying the proposed deterministic sampling approach to recursive nonlinear estimation for systems characterized according to (1), the system state has to be augmented with the noise variables. The resulting augmented system state is denoted by $\underline{\boldsymbol{X}}_k = [\underline{\boldsymbol{x}}_k^{\mathrm{T}}, \underline{\boldsymbol{w}}_k^{\mathrm{T}}, \underline{\boldsymbol{v}}_k^{\mathrm{T}}]^{\mathrm{T}}$. The sample points $\underline{\mu}_l$ are now determined for the augmented system state.

In the following, the equations of the Gaussian filter for nonlinear state estimation based on the novel approximation scheme are presented [1]. Initially, the augmented system state at time step $k = 0$ is given by $\underline{\boldsymbol{X}}_0$ with mean and covariance matrix [2]

$$\hat{\underline{X}}_0 = [(\hat{\underline{x}}_k^e)^{\mathrm{T}}, \underline{0}^{\mathrm{T}}, \underline{0}^{\mathrm{T}}]^{\mathrm{T}} \ , \qquad \mathbf{C}_0^X = \begin{bmatrix} \mathbf{C}_0^e & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_0^w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_0^v \end{bmatrix} .$$

For the time step $k = 1, 2, \ldots$ do:

(1) Determine sample points with weights: $\left\{\underline{\mu}_l \ , \ \omega_l\right\}$

(2) Prediction step:

$$\begin{aligned} \underline{\mu}_l^p &= \underline{a}_k(\underline{\mu}_l^x, \underline{u}_k, \underline{\mu}_l^w) \\ \hat{\underline{x}}_k^p &= \sum_l \omega_l \cdot \underline{\mu}_l^p \\ \mathbf{C}_k^p &= \frac{1}{L} \sum_l \left(\underline{\mu}_l^p - \hat{\underline{x}}_k^p\right) \cdot \left(\underline{\mu}_l^p - \hat{\underline{x}}_k^p\right)^{\mathrm{T}} \\ \underline{\mu}_l^y &= \underline{h}_k(\underline{\mu}_l^p, \underline{\mu}_l^v) \\ \hat{\underline{y}}_k^p &= \sum_l \omega_l \cdot \underline{\mu}_l^y \end{aligned}$$

[1] in analogy to (Wan and van der Merwe (2000))

[2] Superscripts p and e indicate predicted and filtered (estimated) system states.

(3) Measurement update step:

$$\mathbf{C}_k^y = \frac{1}{L}\sum_l \left(\underline{\mu}_l^y - \underline{\hat{y}}_k^p\right)\cdot\left(\underline{\mu}_l^y - \underline{\hat{y}}_k^p\right)^{\mathrm{T}}$$

$$\mathbf{C}_k^{xy} = \frac{1}{L}\sum_l \left(\underline{\mu}_l^p - \underline{\hat{x}}_k^p\right)\cdot\left(\underline{\mu}_l^y - \underline{\hat{y}}_k^p\right)^{\mathrm{T}}$$

$$\mathbf{K}_k = \mathbf{C}_k^{xy}\left(\mathbf{C}_k^y\right)^{-1}$$

$$\underline{\hat{x}}_k^e = \underline{\hat{x}}_k^p + \mathbf{K}_k\left(\underline{\hat{y}}_k - \underline{\hat{y}}_k^p\right)$$

$$\mathbf{C}_k^e = \mathbf{C}_k^p - \mathbf{K}_k\mathbf{C}_k^y\mathbf{K}_k^{\mathrm{T}} \ ,$$

where the samples points of the augmented state are $\underline{\mu}_l = [(\underline{\mu}_l^x)^{\mathrm{T}}, (\underline{\mu}_l^w)^{\mathrm{T}}, (\underline{\mu}_l^v)^{\mathrm{T}}]^{\mathrm{T}}$ and $\underline{\hat{y}}_k$ is the measurement vector at time step k.

For the measurement update step it is assumed that the predicted system state $\underline{\boldsymbol{x}}_k^p$ and the measurement $\underline{\boldsymbol{y}}_k$ are jointly Gaussian. Thus, the Kalman filter equations can be applied (last three equations), even in cases where the measurement function $\underline{h}_k(\,\cdot\,)$ is not given in analytic form.

It is important to note that the Gaussian filter described above has the same structure as the UKF. Furthermore, the computational complexity for calculating an estimate is comparable to the UKF for small L, since the main effort is spent for solving the optimization problem, which can be carried out off-line. Performing the eigenvalue decomposition of $\mathbf{C}_k^X$ for on-line scaling and rotation has the same complexity as the matrix square root required for the UKF.

7. SIMULATION RESULTS

The proposed extension to the UKF is applied to two different estimation problems.

7.1 Example I: Scalar Transformation

First, we consider the scalar transformation

$$\boldsymbol{y} = |\boldsymbol{x}| \ ,$$

where $\boldsymbol{x}$ is Gaussian with mean $\hat{x} = 0.5$ and variance $\sigma_x^2 = 1$, i.e., $\tilde{f}_x(x) = \mathcal{N}(x - 0.5, 1)$. The resulting random variable $\boldsymbol{y}$ is non-Gaussian and the odd moments are non-zero. A Monte Carlo (MC) simulation with 10 million samples is performed to determine the moments of $\boldsymbol{y}$. The results are shown in Tab. 2 together with the estimates of the UKF [3] and the proposed estimator (denoted as GF). For the GF, $L \in \{3, 7, 15\}$ sample points are used. It is obvious that the results of UKF and GF coincide for $L = 3$, while the moment estimates of the GF converge to the MC results when increasing the number of sample points. Especially the moments of order three and higher can be approached, which is not possible for the UKF. The only parameter to be adjusted for an increased estimation quality is the number of sample points L, while the UKF provides three parameters, whose tuning has to be carried out very carefully.

The quality of approximating higher-order moments with the GF can be increased, if information about higher-order moments is explicitly considered when calculating the sample points. By introducing further Lagrange multipliers in (4) for explicitly capturing higher-order moments of $\boldsymbol{x}$, the corresponding approximation quality drastically increases. For the last two rows in Tab. 2, indicated by GF*, we use one additional multiplier for capturing the forth moment of $\boldsymbol{x}$, which is 3 for the considered Gaussian. With $L = 7$ and $L = 15$ for instance, this approximation yields better estimates of the moments of $\boldsymbol{y}$ compared to the GF without the additional multiplier.

[3] The parameters of the UKF are set to $\alpha = 1$, $\beta = 0$ and $\kappa = 0.5$.

Table 2. Moments of the random variable $\boldsymbol{y}$.

	Moments				
	1	2	3	4	5
MC	0.8955	1.2500	2.2064	4.5616	10.6191
UKF	0.9832	1.2500	1.8788	3.0625	5.1646
GF(3)	0.9832	1.2500	1.8788	3.0625	5.1646
GF(7)	0.9177	1.2500	2.0523	3.7419	7.2679
GF(15)	0.9032	1.2500	2.1266	4.0948	8.5405
GF*(7)	0.8708	1.2500	2.2498	4.5625	9.8338
GF*(15)	0.8845	1.2500	2.2270	4.5625	10.2094

7.2 Example II: Vehicle Localization

The second example is from the field of vehicle localization, where a vehicle with bicycle kinematics is localized using dead-reckoning and absolute measurements of the positions of the vehicle. Dead-reckoning employs the kinematic model

$$\underline{\boldsymbol{x}}_{k+1} := \begin{bmatrix}\boldsymbol{a}_{k+1}\\ \boldsymbol{b}_{k+1}\\ \boldsymbol{\phi}_{k+1}\end{bmatrix} = \underline{\boldsymbol{x}}_k + (s_k + \boldsymbol{w}_k^s)\cdot\begin{bmatrix}\cos(\boldsymbol{\phi}_k)\\ \sin(\boldsymbol{\phi}_k)\\ \tan(\alpha_k + \boldsymbol{w}_k^\alpha)\end{bmatrix} \ ,$$

which is often also employed for modeling cars. The measurement model is given by

$$\underline{\boldsymbol{y}}_k = \begin{bmatrix}|\boldsymbol{a}_k|\\ |\boldsymbol{b}_k|\end{bmatrix} + \underline{\boldsymbol{v}}_k \ ,$$

where the system state $\underline{\boldsymbol{x}}_k$ comprises the position $[\boldsymbol{a}_k, \boldsymbol{b}_k]^{\mathrm{T}}$ and the orientation $\boldsymbol{\phi}_k$ of the bicycle. The noise $\underline{\boldsymbol{w}}_k = [\boldsymbol{w}_k^s, \boldsymbol{w}_k^\alpha]^{\mathrm{T}}$ and $\underline{\boldsymbol{v}}_k$ are zero-mean white Gaussian with covariance matrix $\mathbf{C}_k^w = \mathrm{diag}(0.1, 0.01)$ and $\mathbf{C}_k^v = \mathrm{diag}(0.1, 0.5)$, respectively.

For simulation purposes, constant inputs $s_k = 5$, which is the velocity, and $\alpha_k = 0.05$, which is the steering angle, are used. The initial system state $\underline{\boldsymbol{x}}_0$ has the mean $\underline{\hat{x}}_0 = [0, 0, 0]^{\mathrm{T}}$ and covariance matrix $\mathbf{C}_0^x = \mathrm{diag}(0.1, 0.1, 0.01)$. With this configuration 100 Monte Carlo simulation runs are performed. Each run consists of 50 alternating prediction and measurement update steps. We compare the performance of the GF with $L = 5$ sample points per dimension and the UKF with $\alpha = 1$, $\beta = 0$, $\kappa = 0$.

Table 3. Average rmse over 100 simulations.

	rmse_a	rmse_b	rmse_ϕ
UKF	0.643	1.307	0.587
GF(5)	0.410	0.817	0.521

In Tab. 3 the average root mean square error (rmse) over all runs is listed separately for each element of the state vector. The GF significantly outperforms the estimation results of the UKF, while the computation time is almost identical (GF needs 0.1259 s per run on average, UKF needs 0.1219 s). For instance in 90 out of the 100 runs the GF provides better estimates of position $\hat{b}_k$ than the

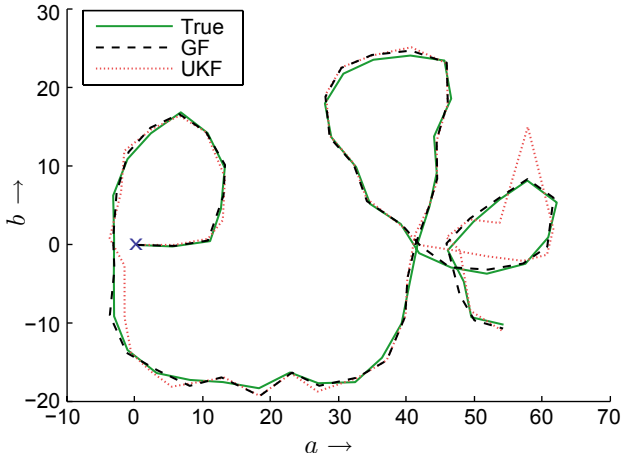

Fig. 2. Example simulation run. The true trajectory of the vehicle in $[\hat{a}_k, \hat{b}_k]^{\mathrm{T}}$ (green, solid) is depicted together with the estimates of the GF (black, dashed) and the unscented Kalman filter (red, dotted).

UKF. One reason why the results of the UKF are inferior is illustrated in Fig. 2. Here, the UKF provides imprecise estimates if one of the position coordinates $[\hat{a}_k, \hat{b}_k]$ is close to zero, e.g., at the most right loop of the trajectory. At those positions, the measurement model has distinct nonlinearities, which cannot be considered accurately by the UKF, while the GF provides a more sophisticated sample representation comprising a larger sample set that offers higher-order accuracy and thus, it behaves more robust at these positions.

8. CONCLUSIONS AND FUTURE WORK

In this paper, the idea of the unscented transform is extended by interpreting the sample points as analytic density function, namely a Dirac mixture. Through this, the sample points are used for directly approximating the distribution function of the prior Gaussian, where computationally demanding parts of the approximation are carried out off-line. In contrast to the unscented Kalman filter, the number of sample points is adjustable, which allows altering the approximation quality. No further tuning parameters are required, which eases the use of the proposed filter.

Altogether, higher-order information of the Gaussian density is implicitly incorporated and nonlinearities of the state transformation are captured more accurate. This consequently leads to improved estimation results, especially with respect to mean and covariance. Based on the proposed deterministic sample calculation scheme a Gaussian filter has been presented. Its structure and computational complexity is comparable to the UKF, whereas its superior estimation accuracy has been demonstrated by means of simulations.

The proposed sample calculation scheme is extendable into many ways, e.g., non-equal weighting coefficients and non-Gaussian densities can be considered.

9. ACKNOWLEDGMENTS

This work was partially supported by the German Research Foundation (DFG) within the Research Training Group GRK 1194 “Self-organizing Sensor-Actuator-Networks”.

REFERENCES

M. 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, February 2002.

D. D. Boos. Minimum Distance Estimators for Location and Goodness of Fit. *Journal of the American Statistical association*, 76(375):663–670, 1981.

D. E. Catlin. *Estimation, Control, and the Discrete Kalman Filter*, volume 71 of *Applied Mathematical Sciences*. New York: Springer-Verlag, 1. edition, 1989.

S. Challa, Y. Bar-Shalom, and V. Krishnamurthy. Nonlinear Filtering via Generalized Edgeworth Series and Gauss–Hermite Quadrature. *IEEE Transactions on Signal Processing*, 48(6):1816–1820, June 2000.

M. F. Huber, D. Brunn, and U. D. Hanebeck. Efficient Nonlinear Measurement Updating based on Gaussian Mixture Approximation of Conditional Densities. In *Proceedings of the American Control Conference (ACC)*, pages 4425–4430, July, 2007.

A. J. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, March 1991.

S. J. Julier and J. K. Uhlmann. Unscented Filtering and Nonlinear Estimation. In *Proceedings of the IEEE*, volume 92, pages 401–422, 2004.

R. E. Kalman. A new Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME, Journal of Basic Engineering*, (82):35–45, 1960.

S. Kullback and R. A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(2): 79–86, 1951.

T. Lefebvre, H. Bruyninckx, and J. De Schutter. Comments on “A New Method for the Nonlinear Trannsformation of Means and Covariances in Filters and Estimators”. *IEEE Transactions on Automatic Control*, 45 (8):1406–1408, 2002.

M. J. D. Powell. A Fortran Subroutine for Solving Systems of Nonlinear Algebraic Equations. In P. Rabinowitz, editor, *Numerical Methods for Nonlinear Algebraic Equations*, chapter 7. 1970.

O. C. Schrempf, D. Brunn, and U. D. Hanebeck. Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering. In *Proceedings of the 45th IEEE Conference on Decision and Control (CDC 2006)*, pages 1709–1714, December 2006.

D. Simon. *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Wiley & Sons, 1. edition, 2006.

D. Tenne and T. Singh. The Higher Order Unscented Filter. In *Proceedings of the American Control Conference*, pages 2441–2446, June 2003.

E. A. Wan and R. van der Merwe. The Unscented Kalman Filter for Nonlinear Estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, pages 153–158, 2000.

Y. Wu, M. Wu, D. Hu, and X. Hu. An Improvement to Unscented Transformation. In *AI 2004: Advances in Artifical Intelligence*, volume 3339/2004. Springer Berlin/Heidelberg, 2004.