

Hybrid Transition Density Approximation for Efficient Recursive Prediction of Nonlinear Dynamic Systems

Marco F. Huber and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory
Institute of Computer Science and Engineering
Universität Karlsruhe (TH), Germany
{marco.huber, uwe.hanebeck}@ieee.org

ABSTRACT

For several tasks in sensor networks, such as localization, information fusion, or sensor scheduling, Bayesian estimation is of paramount importance. Due to the limited computational and memory resources of the nodes in a sensor network, evaluation of the prediction step of the Bayesian estimator has to be performed very efficiently. An exact and closed-form representation of the predicted probability density function of the system state is typically impossible to obtain, since exactly solving the prediction step for nonlinear discrete-time dynamic systems in closed form is unfeasible. Assuming additive noise, we propose an accurate approximation of the predicted density, that can be calculated efficiently by optimally approximating the transition density using a hybrid density. A hybrid density consists of two different density types: Dirac delta functions that cover the domain of the current density of the system state, and another density type, e.g. Gaussian densities, that cover the domain of the predicted density. The freely selectable, second density type of the hybrid density depends strongly on the noise affecting the nonlinear system. So, the proposed approximation framework for nonlinear prediction is not restricted to a specific noise density. It further allows an analytical evaluation of the Chapman-Kolmogorov prediction equation and can be interpreted as a deterministic sampling estimation approach. In contrast to methods using random sampling like particle filters, a dramatic reduction in the number of components and a subsequent decrease in computation time for approximating the predicted density is gained.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: *Stochastic processes*;
G.1.2 [Numerical Analysis]: *Approximation—Nonlinear approximation, Special function approximations*

General Terms

Theory, Algorithms

Keywords

Recursive Bayesian Estimation, Nonlinear Prediction, Hybrid Density, Probability Density Approximation

1. INTRODUCTION

Noisy measurements and system properties that are unknown or too complex to model are typically sources of uncertainties a designer of a sensor network has to deal with. A very common representation of such uncertainties are random variables. Recursive processing of random variables requires the efficient application of the so-called prediction step, i.e., a recursive transformation of a random variable over time. Fig. 1 illustrates an example application from the field of sensor networks, where the prediction step is of great importance: The localization of a robot (for the robot description see [21]) via a sensor network requires continuous position propagations. Inaccuracies in the robot's kinematic model and the imprecise odometry leads to imprecisely known positions that have to be estimated by means



Figure 1: Localization of a mobile robot by means of a sensor network.

of a predictor. In Fig. 1, the estimated position is illustrated by means of covariance ellipses. Measurements of the surrounding sensor nodes can improve the position estimation. However, sensor nodes normally do have limited energy resources. To save energy only sensors that give the best improvement in position estimation should perform a measurement. This sensor scheduling task also requires the prediction of random variables [22].

The Bayesian estimator determines exact predictions of the probability density function of a random variable. In general, the density of the predicted random variable cannot be calculated in closed form and the complexity of the density representation increases with each time step. Especially due to the limited processing and memory resources of the sensor nodes this is an impractical computational effort.

While for linear systems with Gaussian random variables the Kalman filter provides exact solutions in an efficient manner [11], the nonlinear case requires the approximation of the true density. The well-known extended Kalman filter uses linearization to apply the Kalman filter equations to nonlinear systems [20], while the unscented Kalman filter offers increased higher-order accuracy by using a deterministic sampling approach [10]. The resulting single Gaussian density of both estimation methods is typically not a sufficient representation for the true complex density. Another possibility arises from the usage of more generic parameterized density functions such as the well-known Gaussian mixture densities. The bandwidth of estimators using Gaussian mixtures ranges from the efficient Gaussian sum filter [1] that allows only an individual updating of the mixture components up to computationally more expensive but precise methods [7]. Further generic parametric densities used for prediction are for example exponential densities [4], Edgeworth series [6], or Fourier densities [5].

Instead of a parametric representation one can use a sample representation of the density, like particle filters do [2]. Since these estimators apply Monte Carlo methods, i.e., random sampling of the underlying continuous density, the prediction results are not deterministic and a large number of samples is required in order to get satisfactory results. As an alternative to random sampling, Quasi-Monte Carlo estimators use deterministically drawn samples [15]. The techniques used for generating these samples are often very complex. Thus, scalability particularly for an increasing number of dimensions is a critical problem [14]. A likewise computationally demanding approximation of arbitrary prior densities with deterministically drawn samples is proposed in [16, 17]. In contrast to Monte Carlo sampling very few samples are sufficient to achieve precise prediction results, since the samples are optimally placed with respect to a given distance measure.

Generally, there are two possible ways for approximating the resulting density functions of nonlinear stochastic systems: 1. Approximating the system directly, like the extended Kalman filter does. 2. Approximating the prior density without affecting the system, e.g. particle filters. In this paper we introduce an efficient prediction approach for nonlinear dynamic discrete-time systems, which can be considered member of both classes. It is based on optimally approximating the transition density by means of a hybrid density, that consists of Dirac delta functions and another density type, e.g. Gaussian densities, that depends on the noise corrupting the system. Since we only assume additive

noise and no specific noise representation, the proposed prediction method is very generic. For approximation purposes we formulate an optimization problem under boundary conditions for minimizing the squared integral distance measure between the true transition density and the approximate hybrid density. In contrast to our purely Gaussian mixture transition density approximation approach in [8], this optimization problem can be solved analytically and on-line, i.e., for every single prediction step (see Fig. 2), and is not restricted to time-invariant systems.

Due to approximating the transition density, which is the probabilistic representation of the underlying nonlinear system, the proposed approach can be considered member of the first class of approximating predicted density functions. But it can also be interpreted as a sampling approach, that deterministically samples the prior density. Thus, it can be also considered a member of the second class. However, the hybrid structure of the approximate transition density allows an analytical evaluation of the Chapman-Kolmogorov prediction equation, which results in very low prediction costs, while the deterministic sampling interpretation gives a straightforward way for implementation. Using the proposed method the number of components can be dramatically reduced compared to methods using Monte Carlo sampling.

In the following section, we will review the Bayesian estimator and the Chapman-Kolmogorov equation for discrete-time systems. Furthermore, the hybrid density is introduced. The rest of the paper is structured as follows: In Section 3 the optimization problem for approximation is formulated and the analytical solution is derived. The generality of the proposed prediction approach is also pointed out. Performing the efficient and closed-form prediction step, its interpretation as deterministic sampling approach and the incorporation of measurements are highlighted in Section 4. Throughout the paper, an example system with additive Gaussian noise is investigated. This widely used noise representation is chosen for illustrating several details of the novel approach, while in Section 5 the prediction results for this example system are compared to those of the exact Bayesian estimator, Monte Carlo sampling, and the unscented Kalman filter. The paper closes with conclusions and an outlook on future work.

2. PROBLEM FORMULATION

In this paper we only consider scalar random variables, denoted by boldface letters, e.g. \mathbf{x} . This restriction is made for brevity and clarity only. All results are directly applicable to vector-valued random variables. Furthermore, we consider nonlinear, discrete-time systems with system equation

$$\mathbf{x}_{k+1} = a_k(\mathbf{x}_k) + \mathbf{w}_k, \quad (1)$$

where $a_k(\cdot)$ is the nonlinear system function with at most a finite number of points of discontinuities, \mathbf{x}_k is the system state at time step k with density $f_k^x(x_k)$ and \mathbf{w}_k is the zero-mean noise representing the unknown disturbance affecting the system. It is assumed to be a white random process with density $f_k^w(w_k)$. No further restrictions on the type of the noise density are necessary.

2.1 Nonlinear Prediction

In this section we mainly focus on the system equation (1), while updating by means of measurements from sensor nodes is considered in Section 4.3. Starting from an initial density

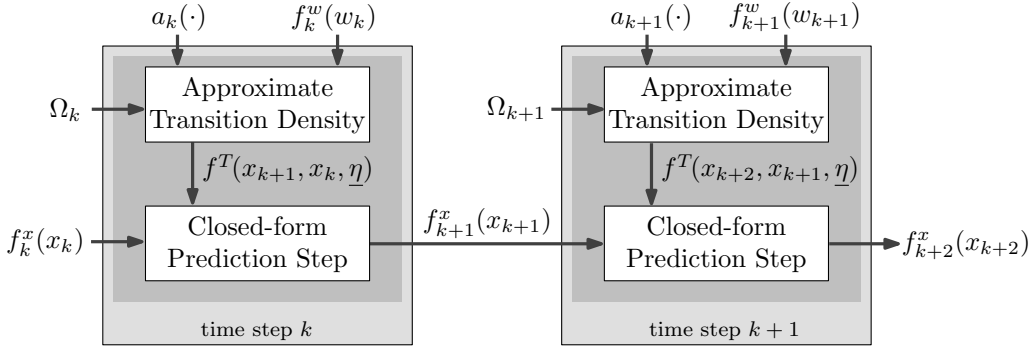


Figure 2: Recursive prediction. The transition density is approximated on-line. Afterward, a closed-form predictions step is performed.

$f_0^x(x_0)$ for \mathbf{x}_0 at $k = 0$, this equation is used in a Bayesian setting for a recursive system state propagation in time, e.g. estimating the position of an object tracked via a sensor network. This so-called *prediction step* of the Bayesian estimator is described by the Chapman-Kolmogorov equation (see for example [19]) and results in a *predicted density*

$$f_{k+1}^x(x_{k+1}) = \int_{\mathbb{R}} f^T(x_{k+1}|x_k) f_k^x(x_k) dx_k \quad (2)$$

for \mathbf{x}_{k+1} , where $f^T(x_{k+1}|x_k)$ is the *transition density*

$$f^T(x_{k+1}|x_k) = f_k^w(x_{k+1} - a_k(x_k)) , \quad (3)$$

which is the probabilistic representation of the system since it depends on the noise density of \mathbf{w}_k and the structure of the system equation. Since (1) is time-variant and the noise process \mathbf{w}_k is non-stationary, this conditional density is also time-variant, i.e., its shape changes over time k . Please note that the transition density has infinite extent, if the domain of $a_k(\cdot)$ is \mathbb{R} . However, as the prior $f_k^x(x_k)$ is a probability density, it is only necessary to consider the support

$$\Omega_k := [\alpha_k, \beta_k] \subset \mathbb{R} \quad (4)$$

of $f_k^x(x_k)$, that stimulates the transition density, where $f_k^x(x_k) > \epsilon$ for $\forall x_k \in \Omega_k$ with a small constant ϵ with $0 < \epsilon \ll 1$.

For nonlinear systems with arbitrarily distributed random variables, recursive prediction according to (2) is of conceptual value only, since the complex shape of the transition density generally prevents a closed-form and efficient solution. Furthermore, no analytical density that can be predicted without changing the type of representation is known in general. To overcome this problem, an appropriate approximation of the true predicted density $f_{k+1}^x(x_{k+1})$ is inevitable. From now on, true densities will be indicated by a tilde, e.g. $\tilde{f}(\cdot)$, the corresponding approximation will be denoted by $f(\cdot)$.

2.2 Hybrid Density

Instead of directly approximating the predicted density $\tilde{f}_{k+1}^x(x_{k+1})$ at every time step, which is computationally demanding, we propose an approximation of the transition

density $\tilde{f}^T(x_{k+1}|x_k)$ by means of the *hybrid density*

$$f^T(x_{k+1}, x_k, \underline{\eta}) = \sum_{i=1}^L \omega_i \cdot \delta(x_k - \mu_i^{(1)}) \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) , \quad (5)$$

with parameter vector

$$\underline{\eta} = [\underline{\eta}_1^T, \underline{\eta}_2^T, \dots, \underline{\eta}_L^T]^T ,$$

where

$$\underline{\eta}_i^T = [\omega_i, \mu_i^{(1)}, \mu_i^{(2)}, \underline{\nu}_i^T] .$$

Here, L is the number of components, ω_i are weighting coefficients with $\omega_i > 0$, $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are location parameters, and $\underline{\nu}_i$ are shape parameters of $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i)$. Each component of the hybrid density consists of two different types of analytical densities: the densities of \mathbf{x}_k are Dirac delta functions $\delta(x_k - \mu_i^{(1)})$, while the densities of \mathbf{x}_{k+1} are $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i)$, where Gaussian densities are a typical choice, i.e., $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) = \mathcal{N}(x_{k+1} - \mu_i^{(2)}, \sigma_i)$, with $\mu_i^{(2)}$ is the mean and σ_i is the standard deviation [18]. Thus, the marginal density of \mathbf{x}_k is a Dirac mixture, while the marginal density of \mathbf{x}_{k+1} is a Gaussian mixture.

Example 1 (Hybrid Density)

The system

$$\mathbf{x}_{k+1} = \sin(\mathbf{x}_k) + \mathbf{x}_k + \mathbf{w}_k \quad (6)$$

is similar to parts of the kinematic model of many vehicles, that can be tracked via sensor networks. The noise \mathbf{w}_k is Gaussian with density $f_k^w(w_k) = \mathcal{N}(w_k, \sigma_w)$, where $\sigma_w = 1$. Fig. 3(a) depicts the system function (black, dashed line) and the corresponding transition density. Furthermore, the red lines in Fig. 3(a) and Fig. 3(b) illustrate a hybrid density with $L = 4$ components. Due to the Dirac delta functions, a single hybrid density component can be interpreted as a vertical *slice* of the transition density.

Our goal is to minimize a certain distance measure $G(\underline{\eta})$ between the true transition density $\tilde{f}^T(x_{k+1}|x_k)$ and its approximation $f^T(x_{k+1}, x_k, \underline{\eta})$. Generally, the calculation of an appropriate parameter vector $\underline{\eta}$ for a high quality approximation of the transition density is computationally demanding. Since the transition density can be time-variant, these

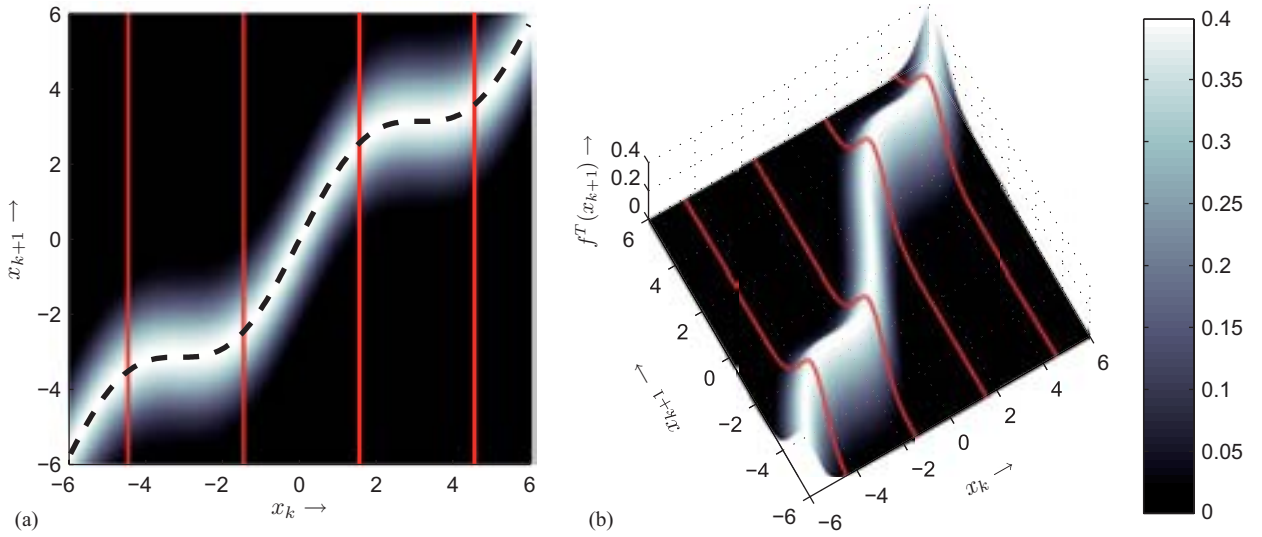


Figure 3: Top view (a) and perspective view (b) on the transition density of the system $x_{k+1} = \sin(x_k) + x_k + w_k$, where w_k is Gaussian. The shape of the transition density strongly depends on the system function (black, dashed line). An approximate hybrid density with $L = 4$ components slices the transition density (red, solid lines) in 5 parts. Each slice is a Gaussian density.

calculations emerge every time step. By selecting a hybrid density for approximation purposes, the required computational effort can be drastically reduced which allows on-line approximation at every time step. With the given transition density approximation the prediction step, which is depicted in Fig. 2, can be performed efficiently in closed form.

3. TRANSITION DENSITY APPROXIMATION

The approximation quality of $f_{k+1}^x(x_{k+1})$ strongly depends on the similarity between $\tilde{f}^T(x_{k+1}|x_k)$ and its hybrid density approximation $f^T(x_{k+1}, x_k, \underline{\eta})$. The key idea is now to reformulate the approximation problem as an optimization problem

$$\underline{\eta}_{min} = \arg \min_{\underline{\eta}} G(\underline{\eta}) , \quad (7)$$

by minimizing a certain distance measure $G(\underline{\eta})$. The result of this optimization problem yields the parameter vector for $f^T(x_{k+1}, x_k, \underline{\eta})$, which minimizes the distance to $\tilde{f}^T(x_{k+1}|x_k)$. As distance measure we employ the widely used squared integral distance measure [9]

$$G(\underline{\eta}) = \frac{1}{2} \int_{\mathbb{R} \Omega_k} (\tilde{f}^T(x_{k+1}|x_k) - f^T(x_{k+1}, x_k, \underline{\eta}))^2 dx_k dx_{k+1} , \quad (8)$$

where Ω_k comprises the support (4) of the prior density $f_k^x(x_k)$. Additionally, we have to formulate some boundary conditions, since the Dirac delta functions in $f^T(x_{k+1}, x_k, \underline{\eta})$ do not allow to compare the transition density and the hybrid density, i.e., there exist an infinite number of solutions for $\underline{\eta}_{min}$ with identical distance to the true transition

density. With

$$\begin{aligned} & 2 \int_{\mathbb{R} \alpha_k}^{\mu_1^{(1)}} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} \\ &= \int_{\mathbb{R} \mu_1^{(1)}}^{\mu_2^{(1)}} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} = \\ & \quad \vdots \\ &= \int_{\mathbb{R} \mu_{L-1}^{(1)}}^{\mu_L^{(1)}} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} \\ &= 2 \int_{\mathbb{R} \mu_L^{(1)}}^{\beta_k} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} , \quad (9) \end{aligned}$$

we ensure equal transition density probability mass enclosed between the several components (slices) of the hybrid density, when approximating the transition density. This is equivalent to employing the so-called *Cramér-von Mises distance* [3]

$$G(\underline{\eta}) = \frac{1}{2} \int_{\mathbb{R} \Omega_k} (\tilde{F}^T(x_{k+1}|x_k) - F^T(x_{k+1}, x_k, \underline{\eta}))^2 dx_k dx_{k+1}$$

between the cumulative distribution function $\tilde{F}^T(x_{k+1}|x_k)$ of the transition density and the cumulative distribution function $F^T(x_{k+1}, x_k, \underline{\eta})$ of the hybrid density.

3.1 Optimal Approximation

Normally, the underlying nonlinearity complicates solving (7), as pointed out for example in [8] for a pure Gaussian

mixture representation of the transition density approximation. Since we are using the hybrid density (5), the optimal solution can easily be derived in closed form.

Theorem 1 (Optimal Approximation)

Given the distance measure (8) with boundary conditions (9), the elements $\underline{\eta}_i^T = [\omega_i, \mu_i^{(1)}, \mu_i^{(2)}, \underline{\nu}_i^T]$ of the optimal solution $\underline{\eta}_{min}$ of the optimization problem (7) are

$$\begin{aligned} \omega_i &= 1, \\ \mu_i^{(1)} &= \alpha_k + \frac{2i-1}{2} \cdot \frac{\beta_k - \alpha_k}{L}, \\ \mu_i^{(2)} &= a_k(\mu_i^{(1)}). \end{aligned}$$

The shape parameter $\underline{\nu}_i$ coincides with the shape parameter of the noise density $f_k^w(w_k)$.

PROOF. Parameter $\mu_i^{(1)}$ can be derived by means of the boundary conditions. Solving one of the integrals in (9) by first integrating along x_{k+1} results in

$$\int_{\mathbb{R}} \int_{\mu_i^{(1)}}^{\mu_{i+1}^{(1)}} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} = \mu_{i+1}^{(1)} - \mu_i^{(1)},$$

for $i = 1, \dots, L-1$ and

$$\begin{aligned} 2 \int_{\mathbb{R}} \int_{\alpha_k}^{\mu_1^{(1)}} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} &= 2(\mu_1^{(1)} - \alpha_k), \\ 2 \int_{\mathbb{R}} \int_{\mu_L^{(1)}}^{\beta_k} \tilde{f}^T(x_{k+1}|x_k) dx_k dx_{k+1} &= 2(\beta_k - \mu_L^{(1)}), \end{aligned}$$

for the borders of the support.

Thus, the boundary conditions can only be met, if the location parameters $\mu_i^{(1)}$, $i = 1, \dots, L$ are chosen uniformly on Ω_k , according to

$$\mu_i^{(1)} = \alpha_k + \frac{2i-1}{2} \cdot \frac{\beta_k - \alpha_k}{L}.$$

Incorporating the necessary condition $\partial G(\underline{\eta})/\partial \underline{\eta} = \underline{0}$ for the existence of a minimum of $G(\underline{\eta})$ leads to the other optimal hybrid density parameters. The derivative with respect to the weighting coefficient ω_i , $i = 1, \dots, L$, leads to

$$\begin{aligned} \frac{\partial G(\underline{\eta})}{\partial \omega_i} &= - \int_{\mathbb{R}} \int_{\Omega_k} (\tilde{f}^T(x_{k+1}|x_k) - f^T(x_{k+1}, x_k, \underline{\eta})) \\ &\cdot \delta(x_k - \mu_i^{(1)}) \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) dx_k dx_{k+1}. \end{aligned} \quad (10)$$

By setting (10) equal to zero and utilizing (3), the sifting property of the Dirac delta function, and the derivations in [18], Section 4.2, we get

$$\begin{aligned} &\int_{\mathbb{R}} f_k^w(x_{k+1} - a_k(\mu_i^{(1)})) \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) dx_{k+1} \\ &= \int_{\mathbb{R}} \omega_i \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i) dx_{k+1}. \end{aligned} \quad (11)$$

The comparison of both sides of (11) yields

$$f_k^w(x_{k+1} - a_k(\mu_i^{(1)})) = \omega_i \cdot f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i).$$

By comparing parameters we get the optimal parameters for ω_i , $\mu_i^{(2)}$ and $\underline{\nu}_i$. \square

3.2 A General Framework

Summarizing the result of Theorem 1, optimally approximating the transition density is merely a uniform placement of the Dirac delta functions of the hybrid density. The components $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i)$ of the second density type are displaced duplicates of the noise density $f_k^w(w_k)$, that are placed along the nonlinear system function $a_k(\mu_i^{(1)})$. In doing so, no assumptions on the noise density's type are made, i.e., the result of Theorem 1 is not restricted to special noise types.

Typically, a parametric structure is used for representing the noise. So, the shape parameters $\underline{\nu}_i$ of $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i)$ can be directly set to the corresponding parameters of the noise.

Example 2 (Gaussian Noise)

Consider the system (6) of Example 1 and its transition density. Since the noise w_k is Gaussian, the hybrid density consists of Dirac delta functions and Gaussian densities. For $\Omega_k = [-6, 6]$ and $L = 4$ components the optimal parameters are listed in Table 1 and the hybrid density is illustrated in Fig. 3.

Table 1: Optimal parameters for Example 2.

i	ω_i	$\mu_i^{(1)}$	$\mu_i^{(2)}$	$\underline{\nu}_i = \sigma_i$
1	1	-4.5	-3.5225	1
2	1	-1.5	-2.4975	1
3	1	1.5	2.4975	1
4	1	4.5	3.5225	1

If a non-parametric noise density is available or the density type of the noise differs from the desired type for representing $f_{k+1}^x(x_{k+1})$, an appropriate selection of the shape parameters $\underline{\nu}_i$ is needed, such that $f(x_{k+1} - \mu_i^{(2)}, \underline{\nu}_i)$ optimally approximates $f_k^w(w_k)$, e.g. using the method described in [7].

4. EFFICIENT PREDICTION

Due to the simplicity of determining the optimal hybrid density, the transition density approximation is well-suited for sensor networks and can be performed on-line, i.e., at every time step k . The special structure of the hybrid transition density is very convenient for efficiently performing the prediction step.

4.1 Closed-Form Calculation

By means of a hybrid transition density, the Chapman-Kolmogorov integral (2) can be solved analytically, resulting in a closed-form solution of the prediction step.

Theorem 2 (Approximate Predicted Density)

Given the density $f_k^x(x_k)$ of the current system state x_k and the hybrid density (5) with parameter vector $\underline{\eta}$ according to Theorem 1, the approximate predicted density $f_{k+1}^x(x_{k+1})$ is a mixture density with L components that can be calculated analytically.

PROOF. With (2) we obtain

$$\begin{aligned}
f_{k+1}^x(x_{k+1}) &= \int_{\mathbb{R}} f^T(x_{k+1}, x_k, \eta) f_k^x(x_k) dx_k \\
&= \sum_{i=1}^L \omega_i \cdot f(x_{k+1} - \mu_i^{(2)}, \mathcal{L}_i) \\
&\quad \cdot \underbrace{\int_{\mathbb{R}} f_k^x(x_k) \delta(x_k - \mu_i^{(1)}) dx_k}_{=f_k^x(\mu_i^{(1)})} \\
&= \sum_{i=1}^L \omega_{k+1,i} \cdot f(x_{k+1} - \mu_i^{(2)}, \mathcal{L}_i) . \quad (12)
\end{aligned}$$

with $\omega_{k+1,i} = \omega_i \cdot f_k^x(\mu_i^{(1)})$. For $i = 1, \dots, L$, the weighting coefficients ω_i of the hybrid transition density have the same constant value, whose quantity is unessential for prediction. Thus, we can set $\omega_i = 1/\sum_{i=1}^L f_k^x(\mu_i^{(1)})$ to achieve a normalized predicted density. \square

The approximate predicted density is a mixture density, whose type of representation depends on the type $f(x_{k+1} - \mu_i^{(2)}, \mathcal{L}_i)$ of the hybrid density. The type of $f_{k+1}^x(x_{k+1})$ is unchanged during recursive prediction and additionally the complexity remains at a constant level, since the number of components representing $f_{k+1}^x(x_{k+1})$ only depends on the number L of components of the hybrid density. This number is dynamically adjustable if needed (see Section 4.4).

Example 3 (Prediction)

Consider again the nonlinear system (6) with Gaussian noise w_k . Let us further assume that the system state's current density $f_k^x(x_k)$ is a Gaussian mixture. Performing the prediction step according to (12) results in a Gaussian mixture representation

$$f_{k+1}^x(x_{k+1}) = \sum_{i=1}^L \omega_{k+1,i} \cdot \mathcal{N}(x_{k+1} - \mu_i^{(2)}, \sigma_i)$$

of the predicted density.

4.2 Relation to Sampling

As mentioned before, the density types of noise and system state are typically identical as well as represented in a parametric and continuous form. Thus, the predicted density is also parametric and continuous. This is worth mentioning, since the closed-form prediction framework proposed in this paper can be also interpreted as sampling the density $f_k^x(x_k)$ *deterministically*. Due to the Dirac delta functions used in the hybrid density and the boundary condition (9), samples are drawn deterministically from the interval Ω_k . The Dirac delta functions represent the samples that are always located uniformly in Ω_k . Therefore, the sampling is called deterministic. Except for a constant ω_i the weights $\omega_{k+1,i}$ in (12) coincide with the function values of $f_k^x(\mu_i^{(1)})$. Thus, we can replace $f_k^x(x_k)$ with

$$f_k^x(x_k) = \sum_{i=1}^L \omega_{k+1,i} \cdot \delta(x_k - \mu_i^{(1)}) .$$

Using the true transition density $\tilde{f}^T(x_{k+1}|x_k)$ in (2) leads to

$$\begin{aligned}
f_{k+1}^x(x_{k+1}) &= \int_{\mathbb{R}} \tilde{f}^T(x_{k+1}|x_k) f_k^x(x_k) dx_k \\
&= \int_{\mathbb{R}} \tilde{f}^T(x_{k+1}|x_k) \left(\sum_{i=1}^L \omega_{k+1,i} \cdot \delta(x_k - \mu_i^{(1)}) \right) dx_k \\
&= \sum_{i=1}^L \omega_{k+1,i} \underbrace{\int_{\mathbb{R}} \tilde{f}^T(x_{k+1}|x_k) \delta(x_k - \mu_i^{(1)}) dx_k}_{=f_k^w(x_{k+1} - a_k(\mu_i^{(1)}))} \\
&= \sum_{i=1}^L \omega_{k+1,i} \cdot f(x_{k+1} - \mu_i^{(2)}, \mathcal{L}_i) ,
\end{aligned}$$

that is identical to (12). Thus, the proposed approach combines both types of approximate nonlinear prediction: the approximation of the underlying nonlinear system as well as the approximation of the prior density (see Section 1). The sampling interpretation reveals a very straightforward way to implement the predictor. Determining the predicted density is reduced to using duplicated and shifted noise density functions as components for the predicted density and to evaluating the prior density at discrete points with subsequent normalization to obtain the weighting coefficients. The number of components of the predicted density coincides with the number of deterministically placed samples.

Sample-based predictors like the well known particle filter typically do not generate a continuous representation of $f_{k+1}^x(x_{k+1})$. Since they use Monte Carlo techniques, a sample representation is generated. Exceptions are the Gaussian (sum) particle filters described in [12, 13]. These filters use a Gaussian or a Gaussian mixture representation for the predicted density. But still random sampling and resampling is applied. The Dirac mixture estimator proposed in [16] draws samples in an optimal, deterministic way and calculates a continuous density representation. In contrast to the proposed approach, this estimator solves an optimization problem on the prior density. Therefore, numerical and thus computational demanding processing is required.

Example 4 (Prediction (cont'd))

Given the prior density $f_k^x(x_k) = \mathcal{N}(x_k - 2.5, 1)$, the nonlinear system (6), and the corresponding hybrid density with $L = 8$ components, the deterministically placed samples are shown in Fig. 4(a), while the predicted Gaussian mixture density is depicted in Fig. 4(b). The larger a sample weight is, the larger is its contribution to the predicted density. Thus, the first four components of the Gaussian mixture have an almost negligible contribution to the predicted density as the weights of the first four samples in Fig. 4(a) are nearly zero.

4.3 Incorporation of Measurements

The major task of a sensor network is to generate measurements. By including those measurements, estimates solely depending on the prediction result can be greatly improved. Therefore, Bayes' law [19]

$$f_k^e(x_k) = c_k f_k^L(\hat{y}_k|x_k) f_k^x(x_k) , \quad (13)$$

has to be applied to the predicted density $f_k^x(x_k)$ for obtaining the posterior density $f_k^e(x_k)$. In (13), $f_k^L(\hat{y}_k|x_k)$

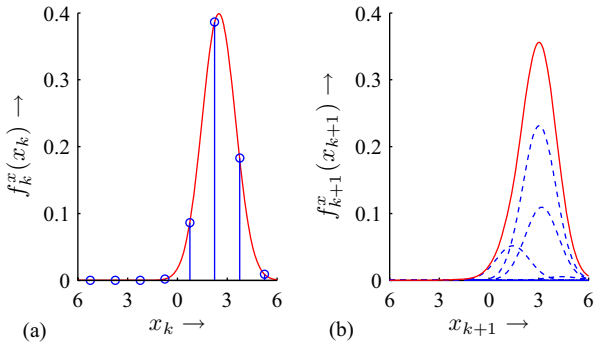


Figure 4: (a) A Gaussian prior density $f_k^x(x_k)$ (red, curve) is sampled at eight deterministically chosen positions (blue, stems) in $\Omega_k = [-6, 6]$. (b) The resulting predicted density (red, solid) and its eight Gaussian components (blue, dashed).

$= f_k^v(\hat{y}_k - h_k(x_k))$ is the so-called *likelihood*, that depends on the actual measurement \hat{y}_k and the measurement model

$$\mathbf{y}_k = h_k(\mathbf{x}_k) + \mathbf{v}_k ,$$

where \mathbf{v}_k is the zero-mean measurement noise with density $f_k^v(v_k)$ and $h_k(\cdot)$ is the nonlinear measurement function. An actual measurement \hat{y}_k is a realization of \mathbf{y}_k .

In a Bayesian framework, incorporating measurements is referred to as *measurement update*. This update is performed in a separate processing step. However, the hybrid density approximation of the transition density allows to perform a combined prediction and measurement update by inserting (13) into (2), i.e., replacing $f_k^x(x_k)$ by $f_k^e(x_k)$. Consequently we obtain

$$f_{k+1}^x(x_{k+1}) = c_k \int_{\mathbb{R}} \tilde{f}^T(x_{k+1}|x_k) f_k^L(\hat{y}_k|x_k) f_k^x(x_k) dx_k ,$$

where $c_k = 1 / \int_{\mathbb{R}} \tilde{f}(\hat{y}_k|x_k) f_k^x(x_k) dx_k$ is a normalization constant. Here, it is sufficient to approximate the transition density and to perform the prediction step as described before to additionally update the system state by means of an actual measurement. According to (12), a mixture representation for $f_{k+1}^x(x_{k+1})$ is obtained. However, by performing a combined prediction and update it is impossible to achieve a separate posterior density $f_k^e(x_k)$, which is not always preferable.

4.4 Number of Components

Due to the shape approximation of the true transition density, the approximation quality of $f_{k+1}^x(x_{k+1})$ increases with an increasing number L of components representing the approximate predicted density and the hybrid density, respectively. The systematic and efficient way the predicted density is generated allows dynamically adjusting the number L of required components for achieving a quantifiable accurate approximation if needed.

Following the description in [17], gradually evaluating an appropriate distance measure, which quantifies the deviation between the true predicted density and its approximation yields the optimal number of components. Due to the fact that the true predicted density is generally not available, we also use an approximate predicted density with L_{max} components, where L_{max} is enough large that the approximation

is very close to the truth. Now, a binary search starts with a minimum number L_{min} of components and stops, if the quality falls below a user-defined maximum error threshold.

In case of a Gaussian mixture representation for $f_{k+1}^x(x_{k+1})$, the squared integral distance measure (8) is suitable and analytically solvable.

5. SIMULATION RESULTS

To investigate the performance of the proposed prediction approach, we consider the nonlinear system

$$\mathbf{x}_{k+1} = \sin(\mathbf{x}_k) + \mathbf{x}_k + \mathbf{w}_k , \quad (14)$$

previously introduced in Example 1 for two noise density representations: At first, we treat the system state propagation under Gaussian noise. Afterward, noise modeled by means of a bimodal Gaussian mixture is considered.

5.1 Gaussian Noise

The Gaussian noise assumption is widely used since it approximates the reality in many scenarios very well. We investigate the prediction results for the noise $\mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k - 0, 0.6)$. Since the noise is Gaussian, the hybrid density consists of Dirac delta functions and Gaussian densities (see Example 2). Thus, we will acquire a Gaussian mixture representation for the predicted density. We approximate the transition density for $\Omega_k = [-6, 6]$, $\forall k \in \mathbb{N}$, with $L = 20$ components.

The simulation starts with the initial Gaussian density

$$f_0^x(x_0) = \mathcal{N}(x_0 + 1, 1.2)$$

of the system state \mathbf{x}_k at time step $k = 0$. Five consecutive prediction steps at time $k = 0, \dots, 4$ are executed. For comparison we take the results of the exact Bayesian estimator (Bayes), predictions based on Monte Carlo (MC) sampling with 200 components, and the predictions made by the unscented Kalman filter (UKF) [10]¹. The MC predictor uses a sample representation of the prior density with randomly drawn samples. Applying (2) on these samples yields a Gaussian mixture approximation of the predicted density. Fig. 5 depicts the predicted densities of the several estimators.

It is obvious that there is almost no shape difference between the exact predictions of the Bayesian estimator and the densities resulting from the proposed approach. Especially both modes are approximated almost exactly. Same is true for the means, see Table 2. Due to the fact that the number of components stays at a constant level of 20, the computation time is also constant and in the order of milliseconds, while exact Bayesian estimation has an exponentially growing time consumption as recursive numerical integration is applied.

The Gaussian assumption makes it impossible for the UKF to keep track of the multimodality of the true predicted densities. Due to random sampling, the MC approach performs predictions based on unsystematically placed samples. Thus, the resulting densities diverge relatively strongly from the ground truth, despite using one order of magnitude more samples as the proposed prediction approach. The same deviation can be determined for the approximated means as

¹The MC predictor and the unscented Kalman filter are both members of the second class of nonlinear approximate estimators as stated in Section 1.

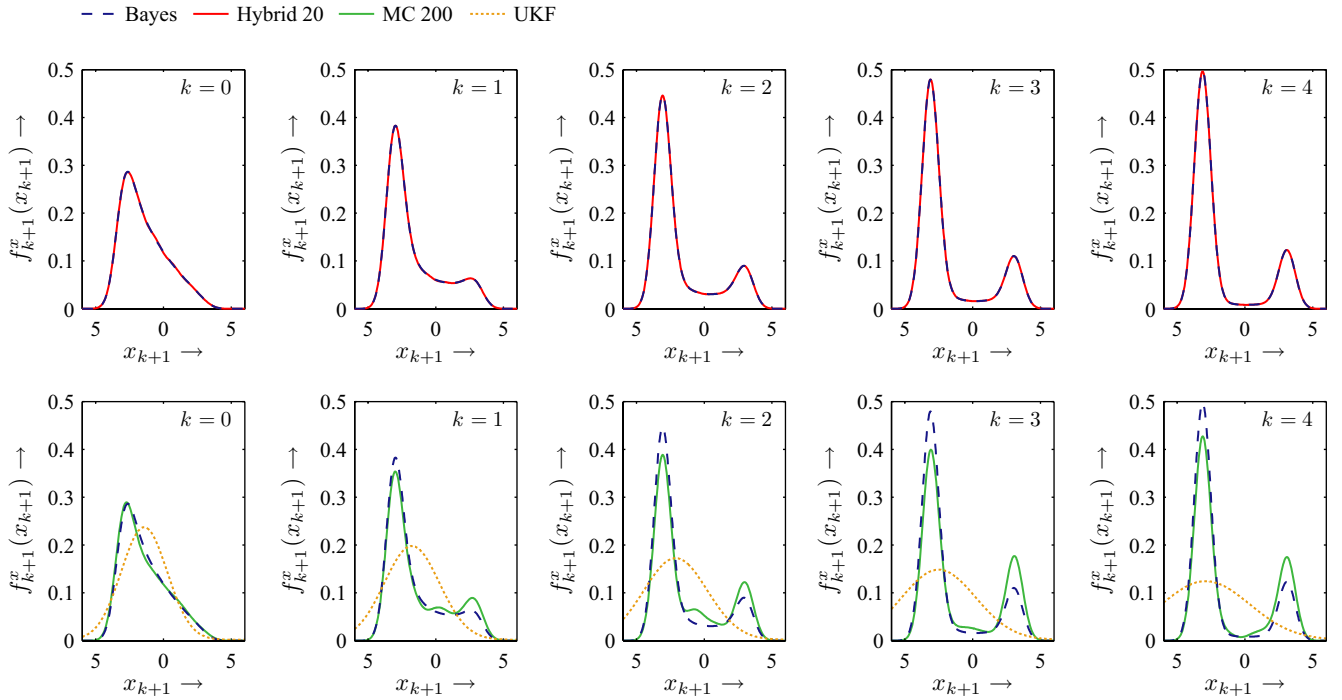


Figure 5: Five prediction steps of the proposed approach with 20 components (upper plots, red solid) in comparison with those of the exact Bayesian estimator (all plots, blue dashed), Monte Carlo sampling with 200 samples (lower plots, green solid) and the unscented Kalman filter (lower plots, orange, dotted).

Table 2: Means of the predicted densities (Gaussian noise).

k	mean: μ_{k+1}			
	Bayes	Hybrid	MC	UKF
0	-1.410	-1.409	-1.418	-1.425
1	-1.652	-1.651	-1.366	-1.761
2	-1.754	-1.753	-1.328	-2.113
3	-1.792	-1.790	-1.091	-2.508
4	-1.805	-1.802	-1.191	-2.903

shown in Table 2. Also higher-order moments cannot be tracked accurately by the UKF and MC approach. In contrast, the shape approximation provided by the proposed method allows to cover higher-order moments accurately.

5.2 Gaussian Mixture Noise

Consider again system (14), but now affected by the bimodal Gaussian mixture noise

$$f_k^w(w_k) = 0.5 \cdot \mathcal{N}(w_k - 1, 0.5) + 0.5 \cdot \mathcal{N}(w_k + 1, 0.5) .$$

All other simulation parameters are similar as in the previous simulation. The prediction results are depicted in Fig. 6. It is obvious that the bimodality of the noise has an effect on the modes of the predicted densities. Each of the formerly two modes under Gaussian noise is replaced by a pair of modes.

Again, the predicted densities determined by means of the approximate hybrid transition density almost coincide with the ground truth provided by the Bayesian estimator. These approximate densities are again represented by means of a Gaussian mixture, but now this Gaussian mixture has 40

Table 3: Means of the predicted densities (Gaussian mixture noise).

k	mean: μ_{k+1}			
	Bayes	Hybrid	MC	UKF
0	-1.410	-1.409	-1.513	-1.425
1	-1.547	-1.548	-1.830	-1.762
2	-1.596	-1.596	-2.136	-2.248
3	-1.615	-1.616	-2.272	-2.816
4	-1.621	-1.621	-2.218	-3.127

components due to the two components of the Gaussian mixture noise. For the i -th component of the hybrid density the relations

$$\begin{aligned} \omega_i &= 0.25 , \\ f(x_{k+1} - \mu_i^{(2)}, \mathcal{V}_i) &= \mathcal{N}(x_{k+1} - \mu_i^{(2)} - 1, 0.5) \\ &\quad + \mathcal{N}(x_{k+1} - \mu_i^{(2)} + 1, 0.5) \end{aligned}$$

hold.

Table 3 reveals a very accurate approximation of the true means by the proposed approach. The estimates from the MC predictor as well as the UKF strongly deviate.

6. CONCLUSIONS AND FUTURE WORK

The key idea of this paper is to approximate the transition density on-line at every prediction step by means of a hybrid density, in order to obtain an efficient recursive prediction. For achieving a high-quality approximation of the transition density and in turn of the predicted density, the approximation is formulated as an optimization problem. Due to

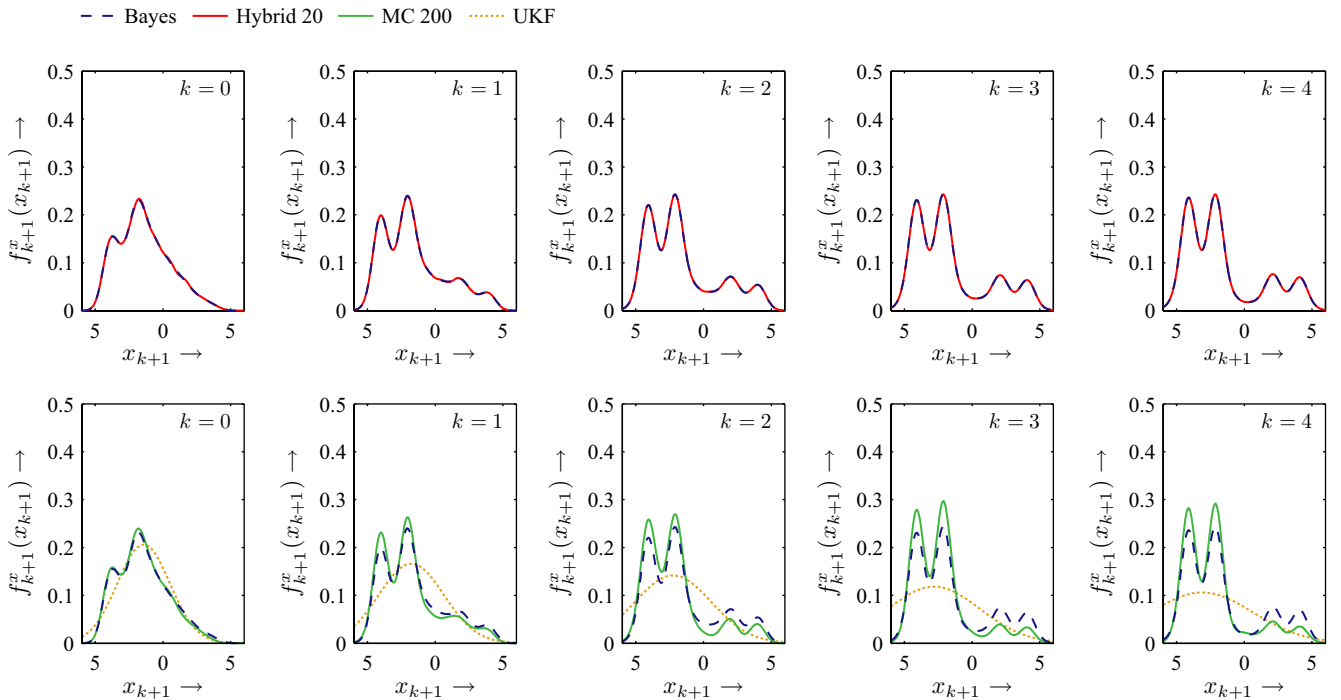


Figure 6: Five prediction steps with system (14) affected by Gaussian mixture noise. Results of the proposed approach with 20 components (upper plots, red solid), the exact Bayesian estimator (all plots, blue dashed), Monte Carlo sampling with 200 samples (lower plots, green solid), and the unscented Kalman filter (lower plots, orange, dotted).

the special structure of the hybrid density, this optimization problem can be solved analytically, permitting complex nonlinear system models and noise densities without demanding computations. Given the approximate transition density, the prediction step for nonlinear dynamic systems corrupted by arbitrary additive noise can be performed in closed form with low computational effort.

Predictions by means of the optimal hybrid density can be interpreted as a deterministic sampling approach. Compared to Monte Carlo methods like particle filters, which utilize random sampling, the proposed approach has several distinct advantages. The Dirac delta components of the hybrid density are systematically placed to solve the optimization problem. In doing so, very few components are needed for achieving predictions represented by means of a continuous parametric density. This offers the opportunity to efficiently derive predictions very close to the ground truth provided by the Bayesian estimator.

The described approach has been introduced for scalar random variables for the sake of brevity and clarity. Generalization to vector-valued random variables is straightforward. It is assumed, that the number of required components grows exponentially with the dimension. Investigating more elaborate distance measures used for the optimization could counter this exponential growth.

In the field of sensor networks the proposed prediction method can be applied in a wide variety of scenarios: sensor scheduling and placement, decentralized information fusion and reconstruction, sensor and object localization. The introduced combined prediction and measurement update approach offers the incorporation of noisy measurements from

the sensor nodes. However, performing separate measurement update steps especially when no system model is available is part of further research.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

- [1] D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian Estimation using Gaussian Sum Approximation. *IEEE Transactions on Automatic Control*, 17(4):439–448, August 1972.
- [2] 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.
- [3] D. D. Boos. Minimum Distance Estimators for Location and Goodness of Fit. *Journal of the American Statistical Association*, 76(375):663–670, 1981.
- [4] D. Brigo and F. LeGland. A Finite Dimensional Filter with Exponential Density. In *Proceedings of the 1997 IEEE Conference on Decision and Control (CDC’97), San Diego, CA*, volume 2, pages 1643–1644, 1997.
- [5] D. Brunn, F. Sawo, and U. D. Hanebeck. Efficient Nonlinear Bayesian Estimation based on Fourier

- Densities. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 312–322, September 2006.
- [6] 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, 2000.
- [7] U. D. Hanebeck, K. Briechle, and A. Rauh. Progressive Bayes: A New Framework for Nonlinear State Estimation. In *Proceedings of SPIE*, volume 5099, pages 256–267. AeroSense Symposium, 2003.
- [8] M. Huber, D. Brunn, and U. D. Hanebeck. Closed-Form Prediction of Nonlinear Dynamic Systems by Means of Gaussian Mixture Approximation of the Transition Density. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, pages 98–103, 2006.
- [9] A. J. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, March 1991.
- [10] S. J. Julier and J. K. Uhlmann. Unscented Filtering and Nonlinear Estimation. In *Proceedings of the IEEE*, volume 92, pages 401–422, 2004.
- [11] R. E. Kalman. A new Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME, Journal of Basic Engineering*, (82):35–45, 1960.
- [12] J. H. Kotecha and P. M. Djurić. Gaussian Particle Filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601, 2003.
- [13] J. H. Kotecha and P. M. Djurić. Gaussian Sum Particle Filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612, 2003.
- [14] A. B. Owen. Monte Carlo Extensions of Quasi-Monte Carlo. In *Winter Simulation Conference Proceedings*, pages 571–577, 1998.
- [15] V. Philomin, R. Duraiswami, and L. Davis. Quasi-Random Sampling for Condensation. In *ser. Lecture Notes in Computer Science*, volume 1843/2000, chapter Computer Vision, pages 134–149. Springer Berlin / Heidelberg, 2000.
- [16] O. C. Schrempf, D. Brunn, and U. D. Hanebeck. Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, pages 512–517, September 2006.
- [17] O. C. Schrempf and U. D. Hanebeck. Efficient Recursive Prediction of Stochastic Nonlinear Systems Based on Dirac Mixture Approximations. In *Proceedings of the American Control Conference*, 2007.
- [18] O. C. Schrempf, A. Hanselmann, and U. D. Hanebeck. Efficient Representation and Fusion of Hybrid Joint Densities for Clusters in Nonlinear Hybrid Bayesian Networks. In *The 9th International Conference on Information Fusion*, 2006.
- [19] F. C. Schweppe. *Uncertain Dynamic Systems*. Prentice-Hall, 1973.
- [20] H. W. Sorenson. *Kalman Filtering: Theory and Application*. Piscataway, NJ: IEEE, 1985.
- [21] F. Weissel and U. D. Hanebeck. A Test-Environment for Control Schemes in the Field of Collaborative Robots and Swarm Intelligence. In *Proceedings of the 7th Intl. Workshop on Computer Science and Information Technologies*, volume 1, September 2005.
- [22] F. Zhao, J. Shin, and J. Reich. Information-Driven Dynamic Sensor Collaboration. *IEEE Signal Processing Magazine*, 19:61–72, March 2002.