The Hybrid Density Filter for Nonlinear Estimation based on Hybrid Conditional Density Approximation

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

Abstract—In nonlinear Bayesian estimation it is generally inevitable to incorporate approximate descriptions of the exact estimation algorithm. There are two possible ways to involve approximations: Approximating the nonlinear stochastic system model or approximating the prior probability density function. The key idea of the introduced novel estimator called Hybrid Density Filter relies on approximating the nonlinear system, thus approximating conditional densities. These densities nonlinearly relate the current system state to the future system state at predictions or to potential measurements at measurement updates. A hybrid density consisting of both Dirac delta functions and Gaussian densities is used for an optimal approximation. This paper addresses the optimization problem for treating the conditional density approximation. Furthermore, efficient estimation algorithms are derived based upon the special structure of the hybrid density, which yield a Gaussian mixture representation of the system state's density.

Keywords: Nonlinear estimation and filtering, conditional density, hybrid density, optimization.

I. INTRODUCTION

Many applications in technical fields like signal processing involve estimating the dynamic state of a nonlinear system. To comprise uncertainties that arise from imperfect knowledge or imprecise measurements, random variable are widely used for characterizing system states. Processing random variables requires the Bayesian estimator for exactly determining prediction and measurement update results. However, recursive Bayesian estimation is typically impractical due to the increasing complexity of the probability density functions, which typically cannot be calculated in closed-form.

While for linear systems with Gaussian random variables the Kalman filter provides exact solutions in an efficient manner [1], the nonlinear case requires the approximation of the underlying true density. The well-known extended Kalman filter uses linearization to apply the Kalman filter equations to nonlinear systems [2], [3], while the unscented Kalman filter offers higher-order accuracy by using a deterministic sampling approach [4]. 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. Gaussian mixtures are a convenient approach for parameterized density functions. The bandwidth of estimators using Gaussian mixtures ranges from the Gaussian sum filter [5] that allows only an individual updating of the mixture components up to computationally more expensive but precise methods [6].

Instead of a parametric density representation, particle filters use samples [7]. Since these estimators apply Monte Carlo methods, i.e., random sampling, the estimation 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 [8]. The techniques used for generating these samples are often very complex and thus scalability is a critical problem [9]. A likewise computationally demanding but optimal approximation of arbitrary prior densities with deterministically drawn samples is proposed in [10]. Due to the optimal placement, few samples are sufficient to achieve precise estimation results.

In this paper we introduce a novel estimator, the *Hybrid Density Filter* (HDF), for nonlinear dynamic discrete-time systems basing on the fundamentals derived in [11]. While most of the existing estimators focus on directly approximating the density function representing the system state, our filter approach is based on optimally approximating the conditional density that is a probabilistic representation of the underlying nonlinear system [12]. For approximating the conditional density and as inspiration for naming our estimation approach, a so-called hybrid density is used, that consists of Dirac delta functions and Gaussian densities. To optimally adapt the parameters of the hybrid density to the conditional density we reformulate the approximation problem into an optimization problem for minimizing the Cramér-von Mises distance. This special type of a squared integral measure is defined over the corresponding cumulative distribution functions of both densities and allows quantifying the distance especially in case of the used Dirac delta functions. In contrast to our purely Gaussian mixture conditional density approximation approach in [12], this optimization problem can be solved analytically and is not restricted to time-invariant systems.

The hybrid structure of the approximate conditional density allows an analytical and efficient evaluation of the prediction step as well as the measurement update step. While performing the prediction step is straightforward and results in a Gaussian mixture representation of the predicted density, measurement updating requires an additional, easy to compute interpolation step for retaining a Gaussian mixture representation. However,
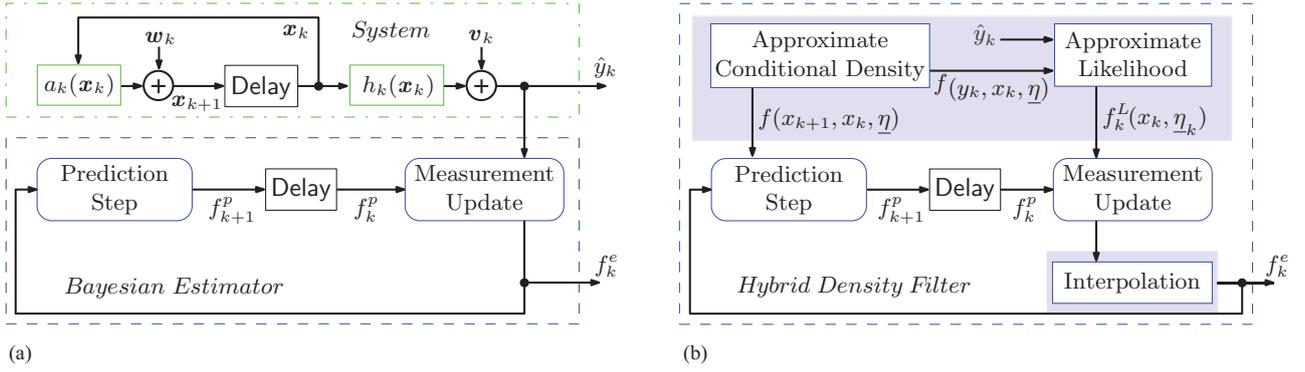
Fig. 1. Hybrid Density Filter (HDF). (a) Nonlinear discrete-time system with additive noise and the Bayesian estimator. (b) Structure of the HDF prediction and measurement update step. The shaded boxes highlight the approximations necessary for estimations by means of the HDF.

a very accurate approximation in shape and moments of the resulting complex density function of the system state is achieved. Predictions or measurement updates by means of the hybrid conditional density approximation can also be interpreted as sampling the prior density deterministically. This deterministic sampling interpretation gives a straightforward way for implementation. In contrast to methods using Monte Carlo sampling the number of components is dramatically reduced.

In the following section, we will give a review of conditional densities and their role in Bayesian estimation. The rest of the paper is structured as follows: Section III introduces the hybrid density, formulates the optimization problem for approximation, and derives the analytical solution. Performing the efficient and closed-form prediction step as well as its interpretation as a deterministic sampling approach is highlighted in Section IV. Section V addresses the measurement update of the HDF including its interpolation step. Section VI investigates an example system for demonstrating the performance by means of simulations. The paper closes with conclusions and an outlook on future work.

II. PROBLEM FORMULATION

In this paper, we only consider scalar random variables, denoted by boldface letters, e.g. $\boldsymbol{x}$. This restriction is made for brevity and clarity only. All results are directly applicable to vector-valued random variables (see e.g. [13]). Furthermore, we consider nonlinear, discrete-time systems affected by additive noise with a system equation

$$\boldsymbol{x}_{k+1} = a_k(\boldsymbol{x}_k) + \boldsymbol{w}_k \qquad (1)$$

and a measurement equation

$$\boldsymbol{y}_k = h_k(\boldsymbol{x}_k) + \boldsymbol{v}_k \quad , \qquad (2)$$

where $a_k(\cdot)$ and $h_k(\cdot)$ are nonlinear functions with at most a finite number of points of discontinuities, $\boldsymbol{x}_k \in \Omega_k$ is the scalar system state at time step k with density $f_k^x(x_k)$ and support

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

where $\forall x_k \in \Omega_k : f_k^x(x_k) > \epsilon$ for constant ϵ with $0 < \epsilon \ll 1$.

A measurement $\hat{y}_k$ is a realization of $\boldsymbol{y}_k$ in (2). $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are zero-mean noise processes representing the unknown disturbance affecting the system and the measurement, respectively. They are assumed as stochastically independent, white Gaussian processes. The noise densities are denoted as $f_k^w(w_k) = \mathcal{N}(w_k - \mu_k^w, \sigma_k^w)$ and $f_k^v(v_k) = \mathcal{N}(v_k - \mu_k^v, \sigma_k^v)$, respectively, where $\mathcal{N}(w_k - \mu_k^w, \sigma_k^w)$ is a Gaussian density with mean μ_k^w and standard deviation σ_k^w.

A. Nonlinear Bayesian Estimation

To determine the probability density function of $\boldsymbol{x}_k$, two steps have to be performed alternately in a Bayesian setting, namely the *prediction step* and the *filter step* or *measurement update* (see Fig. 1(a)).

1) Prediction step: Given a prior density $f_k^x(x_k)$ for $\boldsymbol{x}_k$ at time step k, (1) is used for recursively propagating the system state in time. The prediction step of the Bayesian estimator is described according to the Chapman-Kolmogorov equation (see for example [14]) and results in a *predicted density*

$$f_{k+1}^p(x_{k+1}) = \int_{\mathbb{R}} f^T(x_{k+1}|x_k) \, f_k^x(x_k) \, \mathrm{d}x_k \qquad (3)$$

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

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

in case of additive noise $\boldsymbol{w}_k$. Typically the posterior density is used in (3), i.e., $f_k^x(x_k) = f_k^e(x_k)$.

2) Measurement Update: The *posterior density* $f_k^e(x_k)$ itself is updated considering (2) and applying Bayes' law [14]

$$f_k^e(x_k) = c_k f_k^p(x_k) f_k^L(x_k) \quad , \qquad (4)$$

where $c_k = 1/\int_{\mathbb{R}} f_k^p(x_k) f_k^L(x_k) \, \mathrm{d}x_k$ is a normalization constant and $f_k^L(x_k)$ is the so-called *likelihood*

$$f_k^L(x_k) = f(\hat{y}_k|x_k) = f_k^v(\hat{y}_k - h(x_k)) \quad ,$$

for a given measurement $\hat{y}_k$. It is derived from the conditional density $f(y_k|x_k) = f_k^v(y_k - h(x_k))$ that gives the probability for the occurrence of a measurement $\boldsymbol{y}_k$ given the state $\boldsymbol{x}_k$, i.e., $f(y_k|x_k)$ can be interpreted as the aggregation of all possible likelihoods and thus is of higher dimensionality.
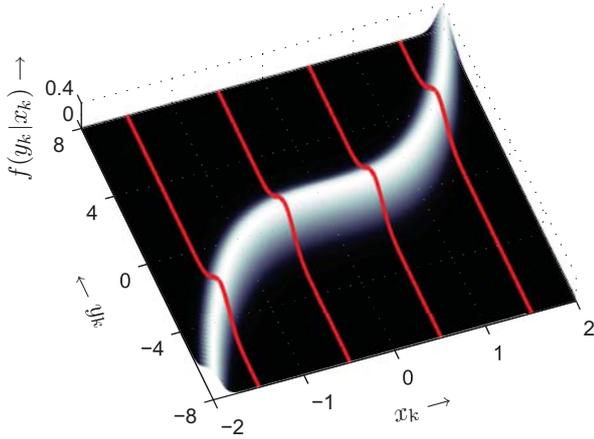
Fig. 2. Conditional density of the nonlinear measurement equation $\boldsymbol{y}_k = \boldsymbol{x}_k^3 + \boldsymbol{v}_k$. The shape of the conditional density strongly depends on the nonlinear function $h_k(\boldsymbol{x}_k) = \boldsymbol{x}_k^3$. A hybrid density with $L = 4$ components (red lines) slices the conditional density in 5 parts. Each slice is a Gaussian density.

Example 1 (Cubic Sensor Conditional Density)
The measurement equation $\boldsymbol{y}_k = \boldsymbol{x}_k^3 + \boldsymbol{v}_k$ is well-known as cubic sensor problem [15]. Fig. 2 depicts the conditional density $f(y_k|x_k) = \mathcal{N}(y_k - x_k^3, \sigma_k^v)$ for $\sigma_k^v = 1$ and $\Omega_k = [-2, 2]$. The likelihood $f_k^L(x_k)$ results from a horizontal cut through this conditional density at position $\hat{y}_k$.

B. Conditional Densities in Bayesian Estimation

Both, the likelihood and the transition density are based upon conditional densities, which in turn depend on the noise densities $f_k^v(v_k)$ and $f_k^w(w_k)$ as well as on the structure of the measurement and system equation, respectively. Additionally, the likelihood depends on the actual measurement $\hat{y}_k$. Since (1) and (2) are time-variant and both noise processes are non-stationary, transition density and likelihood are also time-variant, i.e., their shapes change with time index k.

Generally, recursive Bayesian estimation for nonlinear systems is of conceptual value only, since the complex shapes of the conditional densities prevent a closed-form and efficient solution. Furthermore, for the case of nonlinear systems with arbitrarily distributed random variables, there exists no analytic density that can be calculated without changing the type of representation in general. To overcome this problem, an appropriate approximation is inevitable. From now on, true densities will be indicated by a tilde, e.g. $\tilde{f}(\cdot)$, while the corresponding approximation will be denoted by $f(\cdot)$.

III. CONDITIONAL DENSITY APPROXIMATION

Instead of directly approximating the densities $\tilde{f}_{k+1}^p(x_{k+1})$ and $\tilde{f}_k^e(x_k)$ resulting from (3) and (4), respectively, which is computationally demanding, we propose to approximate the conditional density

$$\tilde{f}(y|x) = \mathcal{N}(y - \psi(x), \sigma) \ .$$

Here, $\psi(\cdot)$ represents a nonlinear function over the random variable $\boldsymbol{x} \in \Omega = [\alpha, \beta] \subset \mathbb{R}$. To obtain for example the

transition density, one has to substitute y, x, $\psi(\cdot)$ and σ such that

$$\tilde{f}^T(x_{k+1}|x_k) = \tilde{f}(y|x)\Big|_{y=x_{k+1},\, x=x_k,\, \psi(\cdot)=a_k(\cdot),\, \sigma=\sigma_k^w} \ .$$

A. Hybrid Density

As a novel type of density representation for approximating the conditional density $\tilde{f}(y|x)$, we propose the *hybrid density* [11]

$$f(y, x, \underline{\eta}) = \sum_{i=1}^{L} \omega_i \cdot \delta(x - \mu_i^x) \cdot \mathcal{N}(y - \mu_i^y, \sigma_i^y) \qquad (5)$$

with parameter vector

$$\underline{\eta} = [\underline{\eta}_1^\mathrm{T}, \underline{\eta}_2^\mathrm{T}, \ldots, \underline{\eta}_L^\mathrm{T}]^\mathrm{T}, \ \text{where} \ \underline{\eta}_i^\mathrm{T} = [\omega_i, \mu_i^x, \mu_i^y, \sigma_i^y] \ .$$

Here, L is the number of components, ω_i are weighting coefficients with $\omega_i > 0$, μ_i^x and μ_i^y are the means, and σ_i^y is the standard deviation of $\mathcal{N}(y - \mu_i^y, \sigma_i^y)$. The marginal densities of a hybrid density consist of two different types of analytic densities: the densities of $\boldsymbol{x}$ are Dirac delta functions $\delta(x_k - \mu_i^x)$ and the densities of $\boldsymbol{y}$ are Gaussians $\mathcal{N}(y - \mu_i^y, \sigma_i^y)$.

Our goal is the minimization of a certain distance measure $G(\underline{\eta})$ between the true conditional density $\tilde{f}(y|x)$ and its approximation $f(y, x, \underline{\eta})$. Generally, the calculation of an appropriate parameter vector $\underline{\eta}$ for a high quality approximation of the transition density is computationally demanding. Since the conditional densities can be time-variant, these calculations are required at every time step. By selecting a hybrid density for approximation purposes, the required computational effort can be drastically reduced and on-line approximation at every time step is possible. With the given conditional density approximation, the prediction and measurement update step can be performed efficiently in closed form, as illustrated in Fig. 1(b).

B. Optimal Approximation

The approximation quality of the HDF strongly depends on the similarity between $\tilde{f}(y|x)$ and its hybrid density approximation $f(y, x, \underline{\eta})$. The key idea is now to reformulate the approximation problem as an optimization problem

$$\underline{\eta}_{min} = \arg \min_{\underline{\eta}} G(\underline{\eta}) \ , \qquad (6)$$

by minimizing a certain distance measure $G(\underline{\eta})$. The result of this optimization problem yields the parameter vector $\underline{\eta}$ for $f(y, x, \underline{\eta})$, which minimizes the distance to $\tilde{f}(y|x)$.

Typical measures quantifying the distance between densities, like the Kullback-Leibler distance [16] or the squared integral measure [17], cannot be applied directly due to the used Dirac delta functions in (5). Thus, the corresponding cumulative distribution functions are employed instead. The distribution function of the true conditional density $\tilde{f}(y|x)$ for $x \in \Omega = [\alpha, \beta]$ can be written as

$$\tilde{F}(y|x) = \frac{1}{2} \int_\alpha^x 1 + \mathrm{erf}\left(\frac{y - \psi(s)}{\sqrt{2}\sigma}\right) \mathrm{d}s \ .$$

The *hybrid distribution function* of $f(y, x, \underline{\eta})$ is given by

$$F(y, x, \underline{\eta}) = \frac{1}{2} \sum_{i=1}^{L} \omega_i \, \mathrm{H}\left(x - \mu_i^x\right)\left(1 + \mathrm{erf}_i(y)\right) \quad , \quad (7)$$

where

$$\mathrm{erf}_i(y) = \mathrm{erf}\left(\frac{y - \mu_i^y}{\sqrt{2}\sigma_i^y}\right)$$

is the error function of component i and

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

is the Heaviside function at position μ.

As distance measure we employ the so-called *Cramér-von Mises distance* [18]

$$G(\underline{\eta}) = \frac{1}{2} \int_{\mathbb{R}} \int_{\Omega} \left(\tilde{F}(y|x) - F(y, x, \underline{\eta})\right)^2 \mathrm{d}x \, \mathrm{d}y \quad . \quad (8)$$

Normally, the underlying nonlinearity complicates solving (6), as pointed out for example in [12] for a pure Gaussian mixture representation of the conditional density approximation. Since we are using the hybrid distribution (7), the optimal solution can easily be derived in closed form.

Theorem 1 (Optimal Approximation)
Given the distance measure (8), the elements $\underline{\eta}_i^T = [\omega_i, \mu_i^x, \mu_i^y, \sigma_i^y]$ of the optimal solution $\underline{\eta}_{min}$ of the optimization problem (6) are

$$\omega_i = \frac{\beta - \alpha}{L} \, ,$$
$$\mu_i^x = \alpha + \omega_i \cdot \frac{2i - 1}{2} \, ,$$
$$\mu_i^y = \psi(\mu_i^x) \, ,$$
$$\sigma_i^y = \sigma \, .$$

At first, a Lemma is developed, which is used to prove Theorem 1.

Lemma 1 *Given the distance measure*

$$G^*(\underline{\eta}) = \frac{1}{2} \int_{\Omega} \left(\tilde{F}(x) - F(x, \underline{\eta})\right)^2 \mathrm{d}x \quad (9)$$

for $x \in \Omega$ over the distribution functions

$$\tilde{F}(x) = \int_{\alpha}^{x} \tilde{f}(s) \, \mathrm{d}s = x - \alpha \, , \quad (10)$$

$$F(x, \underline{\eta}) = \int_{\alpha}^{x} f(s, \underline{\eta}) \, \mathrm{d}s = \sum_{i=1}^{L} \omega_i \cdot \mathrm{H}(x - \mu_i^x) \quad$$

of the marginal densities

$$\tilde{f}(x) = 1 \, , \quad (11)$$

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

minimizing (9) results in

$$\mu_i^x = \alpha + \frac{1}{2}\omega_i + \sum_{j=1}^{i-1} \omega_j \, .$$

Proof: **[of Lemma 1]** Omitted due to space limitations. ∎

Proof: **[of Theorem 1]** The partial derivative of (8) with respect to μ_i^x, $i = 1, \ldots, L$ under incorporation of the necessary condition $\partial G(\underline{\eta})/\partial\mu_i^x = 0$ for a minimum leads to

$$\int_{\mathbb{R}} \int_{\Omega} \left(\tilde{F}(y|x) - F(y, x, \underline{\eta})\right) \cdot \frac{\partial F(y, x, \underline{\eta})}{\partial\mu_i^x} \, \mathrm{d}x \, \mathrm{d}y = 0 \, ,$$

with

$$\frac{\partial F(y, x, \underline{\eta})}{\partial\mu_i^x} = -\frac{1}{2}\omega_i \cdot \delta(x - \mu_i^x)\left(1 + \mathrm{erf}_i(y)\right) \, .$$

Utilizing the sifting property of the Dirac delta function we get

$$\int_{\mathbb{R}} \tilde{F}(y|\mu_i^x)\left(1 + \mathrm{erf}_i(y)\right) \mathrm{d}y$$
$$= \int_{\mathbb{R}} F(y, \mu_i^x, \underline{\eta})\left(1 + \mathrm{erf}_i(y)\right) \mathrm{d}y \, .$$

To allow further simplifications, we apply nonlinear shear, i.e., we set

$$\psi(\mu_i^x) = K \, , \quad K \in \mathbb{R} \text{ (constant)} \, . \quad (12)$$

This changes only the position of the probability mass along dimension y. The total probability mass and the total marginal probability mass of $\boldsymbol{y}$ stay unchanged. Thus, we get

$$\int_{\mathbb{R}} \frac{1}{2}(\mu_i^x - \alpha)\left(1 + \mathrm{erf}\left(\frac{y - K}{\sqrt{2}\sigma}\right)\right)\left(1 + \mathrm{erf}_i(y)\right) \mathrm{d}y$$
$$= \int_{\mathbb{R}} F(y, \mu_i^x, \underline{\eta})\left(1 + \mathrm{erf}_i(y)\right) \mathrm{d}y \, .$$

Resubstituting (12) and comparing coefficients leads to

$$\mu_i^x = \alpha + \frac{1}{2}\omega_i + \sum_{j=1}^{i-1} \omega_j \, , \quad (13)$$
$$\mu_i^y = \psi(\mu_i^x) \, ,$$
$$\sigma_i^y = \sigma \, .$$

The result of Lemma 1 coincides with (13). Thus, minimizing (9) is sufficient for obtaining ω_i and μ_i^x.

In consideration of (10) und (11) it is obvious that $\tilde{F}(x)$ represents the distribution function of an unnormalized uniform distribution on Ω. The optimal approximation of such a distribution by means of Dirac and Heaviside mixtures is well-known [19]. So, we obtain

$$\omega_i = \frac{\beta - \alpha}{L} \, ,$$
$$\mu_i^x = \alpha + \omega_i \cdot \frac{(2i - 1)}{2} \, .$$

∎

Summarizing the result of Theorem 1, optimally approximating the conditional density is merely a uniform placement of the Dirac delta functions of the hybrid density. The Gaussian elements $\mathcal{N}(y - \mu_i^y, \sigma_i^y)$ of the hybrid density are displaced duplicates of the noise density $f_k^w(w_k)$ or $f_k^v(v_k)$ that are placed along the nonlinear functions $a_k(\mu_i^x)$ or $h_k(\mu_i^x)$, respectively.

Remark 1 (Generalization)

Typically, a parametric structure is used for representing the noise, which allows directly setting σ_i^y of $\mathcal{N}(y - \mu_i^y, \sigma_i^y)$ to the corresponding parameter of the noise density. This can be generalized to other parametric noise density representations like Gaussian mixtures, exponential densities [20] or Edgeworth series [21]. There, the non-Dirac mixture density type of the hybrid density has to be chosen according to the current noise density representation [11].

If a non-parametric noise density is available or the density type of the noise differs from the desired type for representing the system state's density, it is possible to first find an appropriate noise density approximation, e.g. using the method described in [22] for non-parametric noise or the method described in [6] for differing parametric noise, and then to approximate the conditional density afterwards.

Example 2 (Hybrid Density)

Consider the cubic sensor measurement equation of Example 1 and its conditional density. Fig. 2 illustrates the approximate hybrid conditional density for $L = 4$ components, where the optimal parameters are $\omega_i = 1$, $\sigma_i^y = 1$, $\mu^x = \{-1.5, -0.5, 0.5, 1.5\}$, and $\mu^y = \{-3.375, -0.125, 0.125, 3.375\}$, for $i = 1, \ldots, 4$. Due to the Dirac delta functions, a single component can be interpreted as a *slice* of the conditional density.

IV. HDF PREDICTION STEP

Due to the simplicity of calculating the optimal hybrid density, the conditional density approximation can be performed on-line, i.e., at every time step k. This allows deriving an efficient and closed-form prediction algorithm.

A. Closed-Form Calculation

The special structure of the hybrid transition density approximation is very convenient for efficiently performing the prediction step, since it allows a closed-form solution of the Chapman-Kolmogorov integral (3).

Theorem 2 (Approximate Predicted Density)

Given the density $f_k^x(x_k)$ of the current system state x_k and the hybrid transition density approximation

$$f(x_{k+1}, x_k, \underline{\eta}) = \sum_{i=1}^{L} \omega_i \delta(x_k - \mu_i^x) \mathcal{N}(x_{k+1} - \mu_i^y, \sigma_i^y) \ , \ (14)$$

with parameter vector $\underline{\eta}$ according to Theorem 1, the approximate predicted density $f_{k+1}^p(x_{k+1})$ is a Gaussian mixture density with L components that can be calculated analytically.

Proof: With (3) and (14) we obtain

$$f_{k+1}^p(x_{k+1}) = \int_{\mathbb{R}} f(x_{k+1}, x_k, \underline{\eta}) f_k^x(x_k) \, dx_k$$

$$= \sum_{i=1}^{L} \omega_i \cdot \mathcal{N}(x_{k+1} - \mu_i^y, \sigma_i^y)$$

$$\cdot \underbrace{\int_{\mathbb{R}} f_k^x(x_k) \, \delta(x_k - \mu_i^x) \, dx_k}_{= f_k^x(\mu_i^x)}$$

$$= \sum_{i=1}^{L} \omega_{k+1,i} \cdot \mathcal{N}(x_{k+1} - \mu_i^y, \sigma_i^y) \ , \quad (15)$$

with $\omega_{k+1,i} = \omega_i \cdot f_k^x(\mu_i^x)$. For $i = 1, \ldots, L$, the weighting coefficients ω_i of the hybrid transition density have the same constant value, where the value of ω_i has no impact on the prediction. Thus, we can set $\omega_i = 1/\sum_{i=1}^{L} f_k^x(\mu_i^x)$ to achieve a normalized predicted density. ∎

In a Bayesian setting according to Fig. 1, the prior $f_k^x(x_k)$ can be any continuous density. The HDF prediction step preserves the density type representation subject to the condition that the prior is a Gaussian mixture. Furthermore, the complexity remains at a constant level, since the number of components representing $f_{k+1}^p(x_{k+1})$ only depends on the number L of components of the hybrid density. If desired, this number can also be adjusted dynamically.

B. Relation to Sampling

It is worth mentioning that the predicted density is represented in a parametric and continuous form, since the closed-form prediction of the HDF can also be interpreted as a *deterministic sampling* of the prior $f_k^x(x_k)$. The sampling is called deterministic due to the Dirac delta functions of the hybrid density that are always located uniformly in Ω_k. Except for the constant factor ω_i the weights $\omega_{k+1,i}$ in (15) coincide with the function values of $f_k^x(\mu_i^x)$. Thus, we can replace $f_k^x(x_k)$ with

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

Using the true transition density $\tilde{f}^T(x_{k+1}|x_k)$ in (3) leads to

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

$$= \sum_{i=1}^{L} \omega_{k+1,i} \underbrace{\int_{\mathbb{R}} \tilde{f}^T(x_{k+1}|x_k) \delta(x_k - \mu_i^x) \, dx_k}_{= f_k^w(x_{k+1} - a_k(\mu_i^x))}$$

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

which is identical to (14). Thus, the HDF prediction step can be interpreted as an approximation of the underlying nonlinear system as well as a sample approximation of the prior density. This interpretation holds also for the measurement update

step and offers a very convenient way for implementing the HDF [11].

Random sampling based estimators like the well-known particle filters typically do not generate a continuous representation of $f_{k+1}^p(x_{k+1})$. Since they use Monte Carlo techniques, a sample representation is generated. Exceptions are the Gaussian (sum) particle filters described in [23], [24]. These filters use a Gaussian or a Gaussian mixture representation. But still random sampling is applied. The Dirac mixture estimator proposed in [10] draws samples in an optimal, deterministic way and calculates an analytic continuous density representation. In contrast to the HDF, this estimator solves an optimization problem on the prior density. Therefore, numerical and thus computational demanding on-line processing is required.

V. HDF MEASUREMENT UPDATE

Performing measurement updates via the HDF differs in two aspects from the prediction step. First, instead of directly incorporating the hybrid conditional density, an approximate likelihood has to be generated. Second, the posterior density is not a Gaussian mixture due to the Dirac mixture representation of the likelihood. Thus, an additional interpolation step is needed in order to preserve a continuous density representation (see Fig. 1).

A. Likelihood Generation and Measurement Updating

For a given measurement $\hat{y}_k$ at time step k, generating the likelihood is straightforward. Plugging $\hat{y}_k$ into the hybrid conditional density approximation $f(y_k, x_k, \underline{\eta})$ yields the likelihood approximation

$$
\begin{aligned}
f_k^L(x_k, \underline{\eta}_k) &= f(y_k, x_k, \underline{\eta})\Big|_{y_k = \hat{y}_k} \\
&= \sum_{i=1}^{L} \underbrace{\omega_i \cdot \mathcal{N}(\hat{y}_k - \mu_i^y, \sigma_i^y)}_{=:\omega_{k,i}^x} \cdot \delta(x_k - \mu_i^x) \\
&= \sum_{i=1}^{L} \omega_{k,i}^x \cdot \delta(x_k - \mu_i^x) \ ,
\end{aligned}
\tag{16}
$$

with $\underline{\eta}_k = [\underline{\eta}_{k,1}^{\mathrm{T}}, \underline{\eta}_{k,2}^{\mathrm{T}}, \dots, \underline{\eta}_{k,L}^{\mathrm{T}}]^{\mathrm{T}}$, where $\underline{\eta}_{k,i}^{\mathrm{T}} = [\omega_{k,i}^x, \mu_i^x]$. Thus, the likelihood is represented by a Dirac mixture that in subsequent processing steps is very convenient to efficiently perform the measurement update.

Theorem 3 (Dirac Mixture Posterior Density)
Given the predicted density $f_k^p(x_k)$ of the current system state $\boldsymbol{x}_k$ and the approximate likelihood (16)*, the posterior density $\bar{f}_k^e(x_k)$ is a Dirac mixture.*

Proof: By using Bayes' law (4) we obtain

$$
\begin{aligned}
\bar{f}_k^e(x_k) &= c_k f_k^p(x_k) f_k^L(x_k, \underline{\eta}_k) \\
&= c_k \sum_{i=1}^{L} \omega_{k,i}^x \cdot \underbrace{\delta(x_k - \mu_i^x) \cdot f_k^p(x_k)}_{=f_k^p(\mu_i^x) \cdot \delta(x_k - \mu_i^x)} \\
&= c_k \sum_{i=1}^{L} \omega_{k,i} \cdot \delta(x_k - \mu_i^x) \ ,
\end{aligned}
\tag{17}
$$

where $\omega_{k,i} = \omega_{k,i}^x \cdot f_k^p(\mu_i^x)$. The normalization constant $c_k = 1/\sum_{i=1}^{L} \omega_{k,i}$ results from integrating over the sum in (17). ∎

Feeding this Dirac mixture posterior density into the prediction step leads to a degeneration of the predicted density, since $\bar{f}_k^e(x_k)$ has to be multiplied with the Dirac delta functions of the hybrid transition density. Instead, the Dirac delta functions of $\bar{f}_k^e(x_k)$ can be interpolated with arbitrary functions since the sampling property of the HDF prediction step allows to process any continuous density representation.

B. Interpolation Step

In the following, Gaussians are used for interpolation. This leads to a *Gaussian mixture posterior density*

$$
f_k^e(x_k) = \sum_{i=1}^{L} \omega_{k,i} \cdot \mathcal{N}(x_k - \mu_i^x, \sigma_i^x)
\tag{18}
$$

of $\bar{f}_k^e(x_k)$. This density type representation coincides with the result of the prediction step.

While the parameters $\omega_{k,i}$ and μ_i^x in (18) can be directly adopted from $\bar{f}_k^e(x_k)$, appropriate standard deviations σ_i^x have to be determined. In general, such an interpolation is computationally demanding. Keeping in mind that the conditional density $\tilde{f}(y_k|x_k)$ has the uniform distribution property when marginalizing along y_k (see Lemma 1), the required computational load can be drastically reduced and a suboptimal interpolation can be performed. We assume that all Gaussian components in (18) have the same standard deviation $\sigma_i^x = \sigma^x$, $i = 1, \dots, L$. Then, the interpolation problem can be reduced to an one-dimensional optimization problem with one single optimum. Thus, σ^x can be derived by means of, e.g. gradient descent. The required computational load can be further reduced, if we use the approximate solution [25]

$$
\sigma^x = \frac{\beta_k - \alpha_k}{\sqrt{2}L} \ ,
$$

that convergences to the correct solution for increasing L.

Remark 2 (Posterior Mean)
The mean of the Dirac mixture posterior density and the Gaussian mixture posterior density are equivalent, as calculating the mean of $f_k^e(x_k)$ is merely based on $\omega_{k,i}$ and μ_i^x.

VI. SIMULATION RESULTS

For simulation purposes, we consider the nonlinear system with system equation

$$
\boldsymbol{x}_{k+1} = \sin(\boldsymbol{x}_k) + \boldsymbol{x}_k + \boldsymbol{w}_k
$$

and measurement equation

$$
\boldsymbol{y}_k = \boldsymbol{x}_k^3 + \boldsymbol{v}_k
$$

previously introduced in Example 1. Monte Carlo simulations are performed with $\sigma_k^w = 0.8$, $\sigma_k^v = 0.3$, $L = 75$, and an initial Gaussian density $f_0^x(x_0) = \mathcal{N}(x_0 + 1.5, 1.2)$. Each of the 50 Monte Carlo simulation runs consists of 20 alternating prediction and measurement update steps, starting
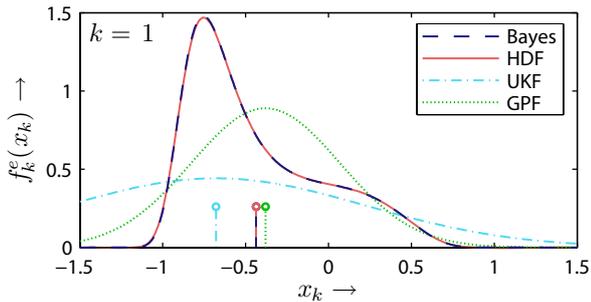
Fig. 3. Density and mean (stems) estimates after the first measurement update.

Fig. 4. Root mean square error over 50 Monte Carlo simulation runs.

TABLE I
AVERAGE RMS ERROR OVER 50 SIMULATION RUNS.

	HDF	UKF	GPF	PF
rms	0.029	0.490	0.094	0.167

with a prediction. For comparison, we take the results of the unscented Kalman filter (UKF), Gaussian particle filter (GPF), and the classical particle filter with systematic resampling (PF) [26][1]. Both particle filters employ 300 particles.

In Fig. 3, an exemplary result of the first measurement update is depicted.[2] We use the computationally demanding Bayesian estimator to obtain the exact density as reference. It is obvious that there is almost no shape difference between the exact density and the Gaussian mixture density approximation resulting from the HDF. Thus, the interpolation step of the HDF measurement update can provide an accurate approximation of the true continuous posterior deniy. In contrast, the Gaussian assumption of the UKF and GPF results in a significant difference in shape. This also appears for the mean estimates, while the HDF provides a mean estimate that is very exact. Also higher-order moments cannot be tracked that accurate by the UKF and GPF approach. In contrast, the shape approximation provided by the HDF allows to cover higher-order moments.

Focusing on the mean estimates, Fig. 4 depicts the root mean square (rms) error for each of the 50 simulations runs. The rms error of the HDF is the lowest in 45 simulation runs. For the remaining 5 runs the PF performs best. Consequently, as shown in Table I, the average rms error of the HDF over all simulation runs is the lowest and is three times lower as the average rms error of the GPF, although the number of particles of the GPF is four times larger than the number of components of the HDF, which also leads to a larger computation time of the GPF. Increasing the number of particles of the GPF does not improve the estimation accuracy significantly. Fig. 5(a)-(c) exemplarily shows the mean estimates of all estimators for one of the 50 Monte Carlo runs, while Fig. 5(d) illustrates for this single run the mean estimation difference of all estimators to the ground truth.

The relatively poor estimation results of the PF follows from the potential multimodal nature of the predicted density. So, the PF sometimes tracks the wrong modes and the mean estimates strongly differ from the true means. Unlike the PF, the HDF shows nearly no deviation from the true means.

---

[1]For UKF and PF we use the Matlab implementation available at [27].
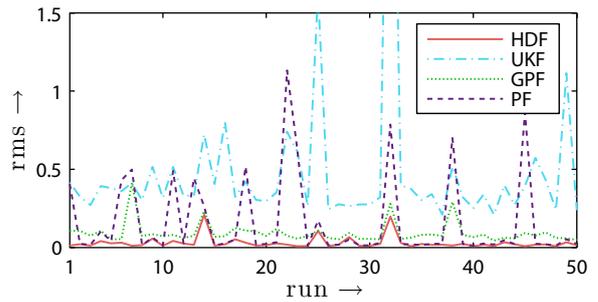[2]The PF provides just a particle representation of the density and thus is omitted here.

VII. CONCLUSIONS AND FUTURE WORK

The Hybrid Density Filter is based on optimally approximating the conditional densities involved in prediction and measurement update. A hybrid density is used for approximation purposes. For achieving a high quality approximation of the conditional density and thus of the estimation results, the approximation is formulated as an optimization problem. Due to the special structure of the hybrid density, this optimization problem can be solved analytically. Given the hybrid conditional density approximation, the prediction and measurement update for nonlinear dynamic systems can be performed in closed form with low computational effort.

Compared to the Gaussian assumption of the UKF, the Gaussian mixture density, which is the output of the HDF, offers an accurate approximation of the true complex density. HDF estimations can also be interpreted as deterministic sampling. Compared to particle filters, which utilize random sampling, the proposed approach has the advantage of a lower number of required samples as well as a continuous density representation. Also, a random number generator is not required, which leads to a simple implementation.

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 and has been done in numerous experiments. In a straightforward implementation the number of required components grows exponentially with the dimension. Investigating more elaborate distance measures used for the optimization could counter this exponential growth. Furthermore, the Gaussian noise assumption can be relaxed resulting in very similar estimation algorithms for the HDF. Improvements of the interpolation step for measurement updates are also intended.
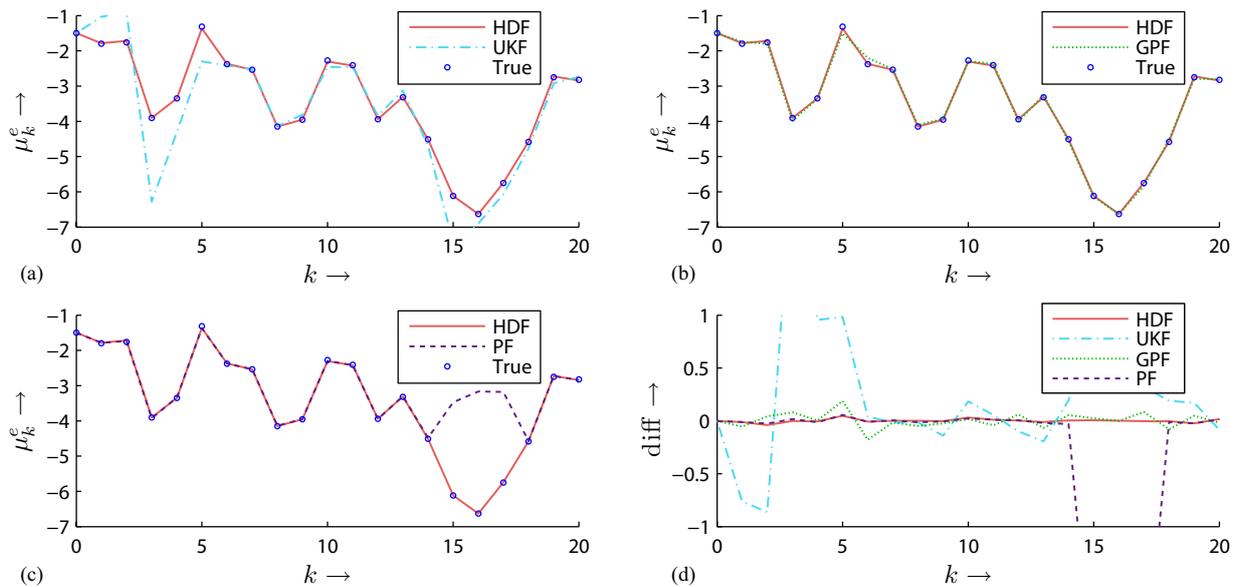
VIII. ACKNOWLEDGMENTS

Fig. 5. The true means (blue, circles) as well as the mean estimates of the HDF (red, solid) compared with those of the (a) UKF, (b) GPF and (c) PF. Difference of the four estimators to the ground truth (d).

REFERENCES

[1] R. E. Kalman, "A new Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, no. 82, pp. 35–45, 1960.

[2] D. Simon, *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*, 1st ed. Wiley & Sons, 2006.

[3] H. W. Sorenson, *Kalman Filtering: Theory and Application*. Piscataway, NJ: IEEE, 1985.

[4] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, vol. 92, no. 3, 2004, pp. 401–422.

[5] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation using Gaussian Sum Approximation," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972.

[6] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: A New Framework for Nonlinear State Estimation," in *Proceedings of SPIE*, vol. 5099. AeroSense Symposium, 2003, pp. 256–267.

[7] 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*, vol. 50, no. 2, pp. 174–188, February 2002.

[8] V. Philomin, R. Duraiswami, and L. Davis, "Quasi-Random Sampling for Condensation," in *ser. Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2000, vol. 1843/2000, ch. Computer Vision, pp. 134–149.

[9] A. B. Owen, "Monte Carlo Extensions of Quasi-Monte Carlo," in *Winter Simulation Conference Proceedings*, 1998, pp. 571–577. [Online]. Available: http://www-stat.stanford.edu/~owen/reports

[10] 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 (ACC)*, 2007.

[11] M. F. Huber and U. D. Hanebeck, "Hybrid Transition Density Approximation for Efficient Recursive Prediction of Nonlinear Dynamic Systems," in *International Conference on Information Processing in Sensor Networks (IPSN)*, 2007, pp. 283–292.

[12] 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)*, 2007.

[13] F. Sawo, M. F. Huber, and U. D. Hanebeck, "Parameter Identification and Reconstruction Based on Hybrid Density Filter for Distributed Phenomena," in *The 10th International Conference on Information Fusion*, 2007.

[14] F. C. Schweppe, *Uncertain Dynamic Systems*. Prentice-Hall, 1973.

[15] R. S. Bucy, "Bayes Theorem and Digital Realizations for Non-Linear Filters," *Journal of Astronautical Sciences*, vol. 17, pp. 80–94, 1969.

[16] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 2, pp. 79–86, 1951.

[17] A. J. Izenman, "Recent developments in nonparametric density estimation," *Journal of the American Statistical Association*, vol. 86, no. 413, pp. 205–224, March 1991.

[18] D. D. Boos, "Minimum Distance Estimators for Location and Goodness of Fit," *Journal of the American Statistical association*, vol. 76, no. 375, pp. 663–670, 1981.

[19] 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)*, December 2006, pp. 1709–1714.

[20] D. Brigo and F. LeGland, "A Finite Dimensional Filter with Exponential Density," in *Proceedings of the 1997 IEEE Conference on Decision and Control (CDC), San Diego, CA*, vol. 2, 1997, pp. 1643–1644.

[21] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy, "Nonlinear Filtering via Generalized Edgeworth Series and Gauss–Hermite Quadrature," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1816–1820, 2000.

[22] A. Hanselmann, O. C. Schrempf, and U. D. Hanebeck, "Optimal Parametric Density Estimation by Minimizing an Analytic Distance Measure," in *The 10th International Conference on Information Fusion*, 2007.

[23] J. H. Kotecha and P. M. Djurić, "Gaussian Particle Filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, 2003.

[24] ——, "Gaussian Sum Particle Filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, 2003.

[25] F. Beutler and U. D. Hanebeck, "The Probabilistic Instantaneous Matching Algorithm," in *International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, September 2006, pp. 311–316.

[26] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," in *IEE Proceedings Radar, Sonar and Navigation*, vol. 146, no. 1, February 1999, pp. 2–7.

[27] N. de Freitas, "Unscented particle filter." [Online]. Available: http://www.cs.ubc.ca/~nando/software.html