Efficient Nonlinear Measurement Updating based on Gaussian Mixture Approximation of Conditional Densities

Marco F. Huber, Dietrich Brunn, and Uwe D. Hanebeck

Abstract— Filtering or measurement updating for nonlinear stochastic dynamic systems requires approximate calculations, since an exact solution is impossible to obtain in general. We propose a Gaussian mixture approximation of the conditional density, which allows performing measurement updating in closed form. The conditional density is a probabilistic representation of the nonlinear system and depends on the random variable of the measurement given the system state. Unlike the likelihood, the conditional density is independent of actual measurements, which permits determining its approximation off-line. By treating the approximation task as an optimization problem, we use progressive processing to achieve high quality results. Once having calculated the conditional density, the likelihood can be determined on-line, which, in turn, offers an efficient approximate filter step. As result, a Gaussian mixture representation of the posterior density is obtained. The exponential growth of Gaussian mixture components resulting from repeated filtering is avoided implicitly by the prediction step using the proposed techniques.

I. INTRODUCTION

Fusing information that has been acquired by measurements is a common challenge in many technical applications like sensor-actuator-networks or robotics. Especially in the presence of uncertainties described by random variables, Bayesian filtering provides exact probability density determinations of the system state. In practical settings, a recursive processing of this so-called posterior density is needed. Since no exact density representation in closed-form and constant complexity is available, the filtering or measurement updating problem is computationally unfeasible in general.

While for linear systems with Gaussian random variables the Kalman filter provides exact solutions in an efficient manner [8], the nonlinear case requires the approximation of the true density. The well-known extended Kalman filter uses linearization to apply the Kalman filter equations on nonlinear systems [12], while the unscented Kalman filter offers increased higher-order accuracy by using a deterministic sampling approach [7]. The resulting single Gaussian density of both estimation methods is typically not sufficient for characterizing the true complex density. One possibility is using a sample representation of the density, like in particle filters [2]. Another possibility is to use generic parameterized density functions. Due to their universal approximation property [10], Gaussian mixtures are very convenient for that purpose. The bandwidth of estimators using Gaussian

Marco F. Huber, Dietrich Brunn, and Uwe D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. {marco.huber|uwe.hanebeck}@ieee.org brunn@ira.uka.de

mixtures ranges from the Gaussian sum filter [1], which is algorithmically straightforward, up to computationally more expensive but precise methods like the one presented in [4].

Unlike the previously mentioned estimation techniques, we introduce a new efficient measurement updating approach that approximates the conditional density, which is a probabilistic representation of the nonlinear measurement equation. For approximation purposes, Gaussian mixtures are used, whose parameters are calculated by means of progressively solving an optimization problem as proposed in [6]. Because of being independent of actual measurements, off-line optimization is possible. Given the approximate conditional density and an actual measurement, the likelihood is generated on-line. Since the likelihood is also represented by a Gaussian mixture, the filter step is reduced to simple multiplications of Gaussian densities resulting in a Gaussian mixture representation of the posterior density. To avoid an exponential growth of the number of Gaussian mixture components, we propose a simultaneous prediction and Gaussian mixture reduction. Hence, an efficient filter step with constant complexity is obtained.

In the following Section, we review Bayes' law for filtering discrete-time systems and point out the relation between conditional density and likelihood. The rest of the paper is structured as follows: In Section III, the progressive processing for approximating the conditional density is explained. An example application is also investigated. Approximating the likelihood and performing the filter step on-line is subject of Section IV, while in Section V the closed-form prediction step and Gaussian mixture reduction is derived. In Section VI, the interaction of both techniques with the efficient filter step is demonstrated and compared to the unscented Kalman filter, the particle filter, and the Bayesian estimator by means of the example application. 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. $\boldsymbol{x}$. Thus, we consider scalar nonlinear time-invariant systems, where scalar measurements $\hat{y}_k$ at time step k are related to the scalar system state $\boldsymbol{x}_k$ by means of the measurement equation

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k) + \boldsymbol{v}_k \ , \tag{1}$$

where the additive noise $\boldsymbol{v}_k$ is assumed as a white stationary Gaussian random process with density $f^v(v_k) = \mathcal{N}(v_k - \mu^v, \sigma^v)$, mean μ^v, and standard deviation σ^v. Note that an actual measurement $\hat{y}_k$ is a realization of (1).

Given a predicted density $f_k^p(x_k)$ for $\boldsymbol{x}_k$, a new measurement updates the system state via the *filter step* or *measurement update* according to Bayes' law [12]

$$f_k^e(x_k) = c_k f_k^p(x_k) f_k^L(x_k) \ , \qquad (2)$$

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

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

The likelihood depends on the noise density of $\boldsymbol{v}_k$, the structure of the measurement equation, and especially on the measurement $\hat{y}_k$. Hence, the likelihood's shape changes with every new measurement.

Recursively updating the predicted density $f_k^p(x_k)$ according to (2) is of conceptual value only, since the complex shape of the likelihood prevents a closed-form and efficient solution. Furthermore, for the case of nonlinear systems with arbitrarily distributed random variables, in general there exists no analytical density that can be updated in the filter step without changing the type of representation. To overcome these insufficiencies, an appropriate approximation of the true *posterior density* $f_k^e(x_k)$ 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 $f(\cdot)$.

The typically enormous computational effort when directly approximating $f_k^e(x_k)$ at every time step can be avoided, if we translate our approach for the prediction step proposed in [6] to the filter step. In doing so, we use a Gaussian mixture representation $f_k^L(x_k, \underline{\eta}_k)$ of $\tilde{f}_k^L(x_k)$ for approximation purposes, that depends on the parameter vector $\underline{\eta}_k$. The calculation of an appropriate parameter vector $\underline{\eta}_k$ for high quality approximations is computationally very demanding. Since the likelihood is time-variant, the demanding computations would also be necessary at every time step.

Instead, we can approximate the time-invariant *conditional density* $\tilde{f}(y_k|x_k) = f^v(y_k - h(x_k))$ by the Gaussian mixture density $f(y_k, x_k, \underline{\eta})$ with parameter vector $\underline{\eta}$. The conditional density can be interpreted as the aggregation of all possible likelihoods and thus is of higher dimensionality. In presence of a new measurement $\hat{y}_k$, we can easily obtain the corresponding likelihood with

$$\tilde{f}_k^L(x_k) = \tilde{f}(y_k|x_k)\Big|_{y_k=\hat{y}_k}$$

$$\Downarrow \text{ Approx. } \Downarrow$$

$$f_k^L(x_k, \underline{\eta}_k) = f(y_k, x_k, \underline{\eta})\Big|_{y_k=\hat{y}_k} \ .$$

Thus, the approximate likelihood $f_k^L(x_k, \underline{\eta}_k)$ can be determined on-line when needed by calculating its time-variant parameter vector $\underline{\eta}_k$ from $\underline{\eta}$ and $\hat{y}_k$ as shown in Section IV. The more extensive Gaussian mixture conditional density approximation depending on the parameter vector $\underline{\eta}$ can be solved off-line as illustrated in Fig. 1. However, Gaussian mixture approximations are considered a tough problem. In the following section an effective approximation scheme is presented.
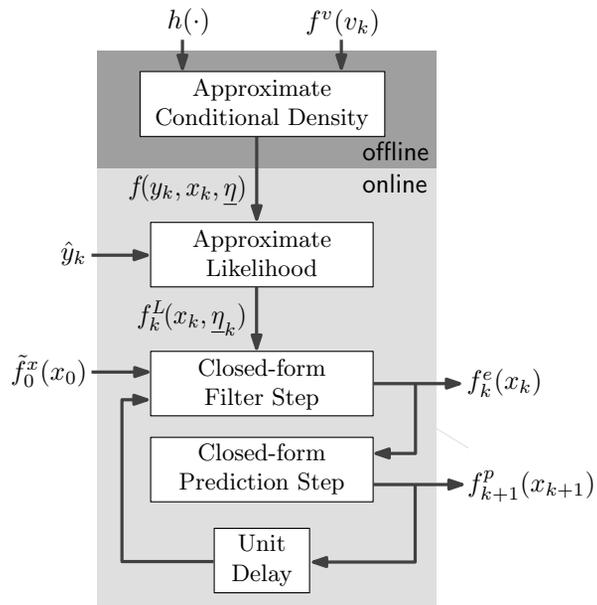

Fig. 1. Recursive, closed-form estimation. The conditional density approximation is performed off-line, while likelihood approximation and the filter step remain on-line tasks. A closed-form prediction step depending on transition density approximation as derived in [6] completes the efficient estimator for nonlinear systems.

III. APPROXIMATION OF THE CONDITIONAL DENSITY

The key idea is to reformulate the Gaussian mixture approximation problem as an optimization problem by minimizing a certain distance measure between $\tilde{f}(y_k|x_k)$ and $f(y_k, x_k, \underline{\eta})$. For solving this problem, we give a review on the progressive optimization scheme proposed in [6].

Since in real systems the system state is usually restricted to a finite interval, i.e.,

$$\forall k : x_k \in [a, b] =: \Omega \ ,$$

we are only interested in approximating the conditional density for $x_k \in \Omega$.

Furthermore, we use the special case of a Gaussian mixture with *axis-aligned* Gaussian components (short: *axis-aligned Gaussian mixture*) for representing the Gaussian mixture approximation $f(y_k, x_k, \underline{\eta})$. Here, each component is separable in every dimension, i.e.,

$$f(y_k, x_k, \underline{\eta}) = \sum_{i=1}^{L} \omega_i \cdot \mathcal{N}(y_k - \mu_i^y, \sigma_i^y)\mathcal{N}(x_k - \mu_i^x, \sigma_i^x) \ (3)$$

with the parameter vector

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

An axis-aligned Gaussian mixture has minor approximation capabilities compared to a non axis-aligned one. Hence, more components are required to achieve a comparable approximation quality. In exchange, the covariance matrices of the axis-aligned Gaussian mixture components are diagonal. Thus, less parameters for a single component have to be adjusted and the necessary determination of the gradient $\frac{\partial G}{\partial \eta}$ proves to be easier. Altogether, representing $f(y_k, x_k, \underline{\eta})$ as in (3) lowers the algorithmic complexity.

A. The Optimization Problem

The quality of the approximation $f_k^e(x_k)$ strongly depends on the similarity between $\tilde{f}(y_k|x_k)$ and its Gaussian mixture approximation $f(y_k, x_k, \underline{\eta})$ for $x_k \in \Omega$. Thus, this section is concerned with solving the optimization problem

$$\underline{\eta}_{min} = \arg \min_{\underline{\eta}} G(\underline{\eta}) \;, \qquad (4)$$

that yields the parameter vector for $f(y_k, x_k, \underline{\eta})$, which minimizes the distance to $\tilde{f}(y_k|x_k)$. The employed distance measure is the squared integral measure

$$G(\underline{\eta}) = \frac{1}{2} \iint_{\mathbb{R}\,\mathbb{R}} \left(\tilde{f}(y_k|x_k) - f(y_k, x_k, \underline{\eta}) \right)^2 \mathrm{d}x_k \, \mathrm{d}y_k \;. \quad (5)$$

Although this measure has been selected for its simplicity and convenience, it has been found to give excellent results.

Apart from the selected distance measure, the underlying nonlinearity complicates solving (4) significantly. In general, no closed-form solution can be derived. In addition, the high dimension of $\underline{\eta}$ makes the selection of an initial solution very difficult, so that the direct application of numerical minimization routines leads to insufficient local optima of $\underline{\eta}$.

B. Progressive Processing

Instead of attempting the direct approximation the conditional density, we pursue a progressive approach for finding $\underline{\eta}_{min}$ as shown in Fig. 2. This type of processing has been proposed in [4], [11]. In doing so, a *parameterized conditional density* $\tilde{f}(y_k|x_k, \gamma)$ with the *progression parameter* $\gamma \in [0,1]$ is introduced. Incrementing this progression parameter by small $\Delta\gamma$ ensures a continuous transformation of the solution of an initial, tractable optimization problem towards the desired true conditional density $\tilde{f}(y_k|x_k)$. In every single so-called progression step we achieve a gradually changed distance measure $G(\underline{\eta}, \gamma)$, for which the necessary condition for a minimum

$$\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}} = \underline{0}$$

has to be satisfied. For this purpose, the BFGS formula [3], a standard numerical optimization method, is employed. $G(\underline{\eta}, \gamma)$ results from using the progressive version $\tilde{f}(y_k|x_k, \gamma)$ of $\tilde{f}(y_k|x_k)$ in (5).

For introducing the parameterized conditional density, we use the *parameterized measurement function* $h(\boldsymbol{x}_k, \gamma)$,

$$h(\boldsymbol{x}_k, \gamma) = (1 - \gamma)H \cdot \boldsymbol{x}_k + \gamma h(\boldsymbol{x}_k) \;,$$

where in particular $H \in \mathbb{R}$ and

$$h(\boldsymbol{x}_k, \gamma = 0) = H \cdot \boldsymbol{x}_k \;, \qquad (6)$$
$$h(\boldsymbol{x}_k, \gamma = 1) = h(\boldsymbol{x}_k) \;.$$

This yields the modified measurement equation

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k, \gamma) + \boldsymbol{v}_k \;. \qquad (7)$$
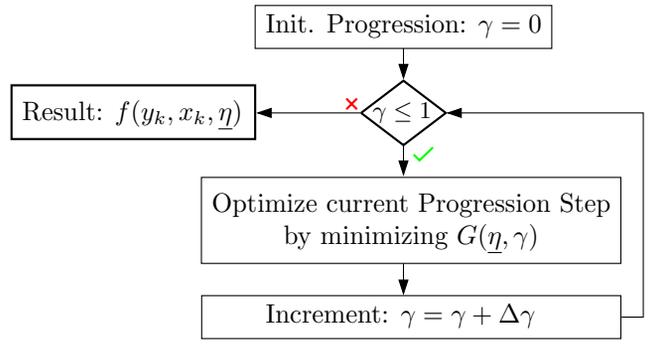

Fig. 2. Flow chart of the progressive processing to determine $\underline{\eta}_{min}$.

The dependence of $\tilde{f}(y_k|x_k, \gamma) = f^v(y_k - h(x_k, \gamma))$ on measurement equation (7) automatically causes its parameterization.

Example 1 (Quadratic Decay Measurement Function)
In a wireless communication scenario the measurement equation $\boldsymbol{y}_k = (1 + \boldsymbol{x}_k^2)^{-1} + \boldsymbol{v}_k$ relates the position $\boldsymbol{x}_k$ of a receiver to the relative signal strength (SNR) $\boldsymbol{y}_k$ according to the *free-space propagation model* [5]. Fig. 3 shows the progression of the corresponding parameterized measurement function $h(\boldsymbol{x}_k, \gamma) = (1 - \gamma)H \cdot \boldsymbol{x}_k + \gamma(1 + \boldsymbol{x}_k^2)^{-1}$ for $\Delta\gamma = 0.2$, $H = 0$ and $\Omega = [-3, 3]$. The parameterized conditional density $\tilde{f}(y_k|x_k, \gamma)$ performs the same transformation.

Initially, $\tilde{f}(y_k|x_k, \gamma = 0)$ corresponds to the conditional density of a linear system (6). For this optimization problem, there exists just one single optimum. Thus, the choice of insufficient starting parameters by the user and starting the progression with a local optimum is bypassed [6]. For $\gamma = 1$ the parameterized conditional density corresponds to the true conditional density, i.e., $\tilde{f}(y_k|x_k, \gamma = 1) = \tilde{f}(y_k|x_k)$.

IV. THE FILTER STEP

Together with the closed-form prediction step proposed in [6], the Gaussian mixture approximation of the conditional density and likelihood respectively allows on-line performing an efficient closed-form filter step as depicted in Fig. 1. For this purpose we assume that all involved densities are represented as Gaussian mixtures.

According to [6] we assume the predicted density $f_k^p(x_k)$ to be given by

$$f_k^p(x_k) = \sum_{j=1}^{L^p} \omega_{k,j}^p \cdot \mathcal{N}(x_k - \mu_{k,j}^p, \sigma_{k,j}^p) \;, \qquad (8)$$

where L^p is the constant number of Gaussian components, $\mathcal{N}(x_k - \mu_{k,j}^p, \sigma_{k,j}^p)$ is a Gaussian density with mean $\mu_{k,j}^p$ and standard deviation $\sigma_{k,j}^p$, and $\omega_{k,j}^p$ are weighting coefficients with $\omega_{k,j}^p > 0$ and $\sum_{j=1}^{L^p} \omega_{k,j}^p = 1$.

Given a Gaussian mixture approximation $f(y_k, x_k, \underline{\eta})$ according to (3), its axis-aligned structure allows the direct approximation of the likelihood $f_k^L(x_k, \underline{\eta}_k)$, if for time step
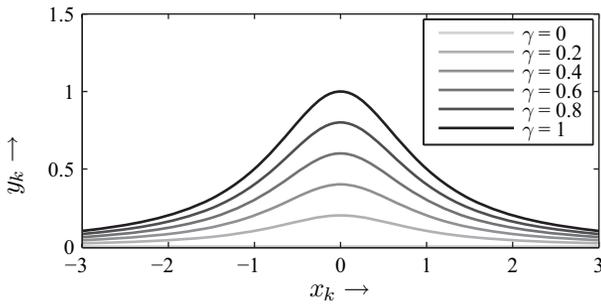
Fig. 3. Progression of the parameterized measurement function $h(\boldsymbol{x}_k, \gamma) = (1-\gamma)H \cdot \boldsymbol{x}_k + \gamma(1+\boldsymbol{x}_k^2)^{-1}$

k a measurement $\hat{y}_k$ is present,

$$f_k^L(x_k, \underline{\eta}_k) = f(y_k, x_k, \underline{\eta})\Big|_{y_k = \hat{y}_k}$$
$$= \sum_{i=1}^{L} \underbrace{\omega_i \cdot \mathcal{N}(\hat{y}_k - \mu_i^y, \sigma_i^y)}_{=:\omega_{k,i}} \cdot \mathcal{N}(x_k - \mu_i^x, \sigma_i^x)$$
$$= \sum_{i=1}^{L} \omega_{k,i} \cdot \mathcal{N}(x_k - \mu_i^x, \sigma_i^x) \qquad (9)$$

and

$$\underline{\eta}_k = [\underline{\eta}_{k,1}^{\mathrm{T}}, \underline{\eta}_{k,2}^{\mathrm{T}}, \ldots, \underline{\eta}_{k,L}^{\mathrm{T}}]^{\mathrm{T}} \text{, where } \underline{\eta}_{k,i} = [\omega_{k,i}, \mu_i^x, \sigma_i^x]^{\mathrm{T}} \text{ .}$$

Example 2 (Quadratic Decay Measurement (cont'd.))
We consider again the measurement equation of Example 1, where now $\boldsymbol{v}_k \sim \mathcal{N}(v_k - 0, 0.25)$ and $\Omega = [-3, 3]$. Using $L = 20$ Gaussian components leads to the conditional density approximation with quality $G(\eta) = 0.0039$ shown in Fig. 4. Applying a measurement $\hat{y}_k = 0.6$, we obtain the likelihood approximation depicted in Fig. 5. The little bumps at the interval borders of x_k result from sharply restricting x_k to the interval Ω. A more continuous windowing would alleviate this.

Incorporating an actual measurement and generating the likelihood corresponds to taking a slice parallel to the x_k-axis from the conditional density at position $y_k = \hat{y}_k$. The Gaussian mixture representation of $f_k^L(x_k, \underline{\eta}_k)$ itself is then very convenient for efficiently performing the filter step.

Theorem 1 (Approximate Posterior Density)
Given the Gaussian mixture representations (8) *and* (9) *for* $f_k^p(x_k)$ *and* $f_k^L(x_k, \underline{\eta}_k)$ *respectively, the approximate posterior density* $f_k^e(x_k)$ *is also a Gaussian mixture that can be calculated analytically.*

PROOF. Using Bayes' law (2) we obtain
$$f_k^e(x_k) = c_k f_k^p(x_k) f_k^L(x_k, \underline{\eta}_k)$$
$$= c_k \sum_{j=1}^{L^p} \sum_{i=1}^{L} \omega_{k,j}^p \omega_{k,i} \underbrace{\mathcal{N}(x_k - \mu_{k,j}^p, \sigma_{k,j}^p)\mathcal{N}(x_k - \mu_i^x, \sigma_i^x)}_{=z_{i,j} \cdot \mathcal{N}(x_k - \mu_{k,i,j}^e, \sigma_{k,i,j}^e)}$$
$$= c_k \sum_{j=1}^{L^p} \sum_{i=1}^{L} \omega_{k,i,j}^e \cdot \mathcal{N}(x_k - \mu_{k,i,j}^e, \sigma_{k,i,j}^e) \qquad (10)$$

where
$$z_{i,j} = \mathcal{N}\left(\mu_{k,j}^p - \mu_i^x, \sqrt{(\sigma_{k,j}^p)^2 + (\sigma_{k,i,j}^e)^2}\right)$$

and
$$\omega_{k,i,j}^e = z_{i,j} \cdot \omega_{k,j}^p \cdot \omega_{k,i} \text{ ,}$$
$$\mu_{k,i,j}^e = (\sigma_{k,i,j}^e)^2 \cdot \left(\frac{\mu_{k,j}^p}{(\sigma_{k,j}^p)^2} + \frac{\mu_i^x}{(\sigma_i^x)^2}\right) \text{ ,}$$
$$\sigma_{k,i,j}^e = \sqrt{\frac{(\sigma_{k,j}^p \cdot \sigma_i^x)^2}{(\sigma_{k,j}^p)^2 + (\sigma_i^x)^2}} \text{ .}$$

The normalization constant $c_k = 1/\sum_{j=1}^{L^p} \sum_{i=1}^{L} \omega_{k,i,j}^e$ results from integrating over both sums in (10). $\qquad \square$

For obtaining the result in (10), only multiplications of two Gaussian densities, denoted by $z_{i,j} \cdot \mathcal{N}(x_k - \mu_{k,i,j}^e, \sigma_{k,i,j}^e)$, have to be performed. Hence, (10) provides the closed-form and efficient solution for the filter step by means of the Gaussian mixture approximation of a likelihood. The accuracy of the approximation of $\tilde{f}_k^e(x_k)$ strongly depends on the number of components of $f(y_k, x_k, \underline{\eta})$ and $f_k^L(x_k, \underline{\eta}_k)$ respectively.

The obtained Gaussian mixture approximation for the posterior density $\tilde{f}_k^e(x_k)$ comprises $L^p \cdot L$ components. Thus, unlike the closed-form prediction step, the number of components in the approximation grows exponentially over time. To avoid this exponential growth it is standard practice to employ Gaussian mixture reduction techniques after the filter step. Instead, the subsequent closed-form prediction step, as depicted in Fig. 1, automatically leads to a Gaussian mixture reduction.

V. GAUSSIAN MIXTURE REDUCTION

Very popular Gaussian mixture reduction methods, Salmond's joining algorithm [13] or Maybeck's ISE based reduction algorithm [9], suffer from either poor reduction quality or a high computational burden. However, predictions with respect to a scalar nonlinear time-invariant system equation
$$\boldsymbol{x}_{k+1} = a(\boldsymbol{x}_k) + \boldsymbol{w}_k \text{ ,}$$
where $\boldsymbol{w}_k$ is white and stationary Gaussian noise, offer the opportunity for directly reducing the number of components to a constant value L^p. For this purpose, the closed-form prediction step
$$f_{k+1}^p(x_{k+1}) = \int_{\mathbb{R}} f^T(x_{k+1}, x_k, \underline{\eta}) f_k^e(x_k) \, dx_k \text{ ,} \quad (11)$$
proposed in [6] has to be performed. Here, $f^T(x_{k+1}, x_k, \underline{\eta})$ is the axis-aligned Gaussian mixture approximation with $\bar{L}^p$ components of the true transition density $\tilde{f}^T(x_{k+1}|x_k)$ for $x_k \in \Omega$. Since $\tilde{f}^T(x_{k+1}|x_k)$ is also a conditional density, its approximation is done similarly to $\tilde{f}(y_k|x_k)$.

Theorem 2 (Approximate Predicted Density)
Given the Gaussian mixture representation (10) *for* $f_k^e(x_k)$ *and an axis-aligned Gaussian mixture representation for* $f^T(x_{k+1}, x_k, \underline{\eta})$ *similar to* (3) *with* L^p *components, the approximate predicted density* $f_{k+1}^p(x_{k+1})$ *is also a Gaussian mixture with* L^p *components that can be calculated analytically.*
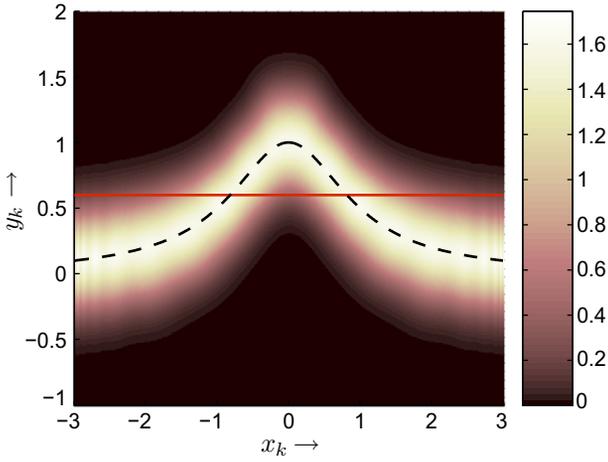
Fig. 4. Top view on the Gaussian mixture approximation of the conditional density $\tilde{f}(y_k|x_k) = \mathcal{N}(y_k - (1 + x_k^2)^{-1}, 0.25)$. The dashed black line displays the underlying measurement function $h(\boldsymbol{x}_k) = (1 + \boldsymbol{x}_k^2)^{-1}$. A measurement $\hat{y}_k = 0.6$ is indicated by the red line, leads to the likelihood depicted in Fig. 5.
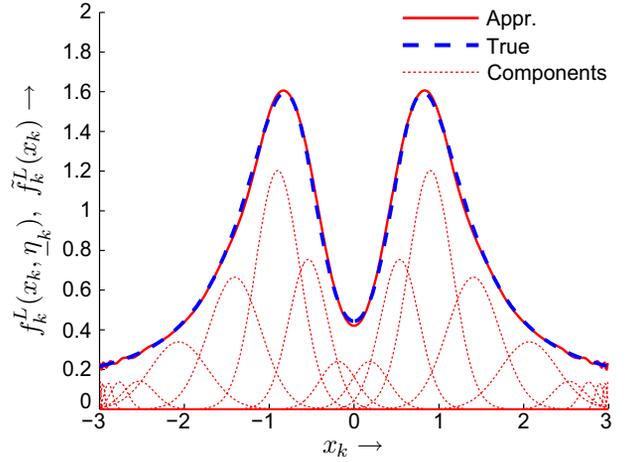

Fig. 5. The likelihood $\tilde{f}_k^L(x_k) = \mathcal{N}(\hat{y}_k - (1 + x_k^2)^{-1}, 0.25)$ with $\hat{y}_k = 0.6$ (blue, dashed) and its Gaussian mixture approximation (red, solid), generated from the conditional density approximation shown in Fig. 4. 20 Gaussian components (red, dotted) are used for representing the approximate likelihood $f_k^L(x_k, \underline{\eta}_k)$.

PROOF. See [6]. □

The number of components in $f_{k+1}^p(x_{k+1})$ depends only on the number of components of the approximate transition density. In return, performing the closed-form prediction step (11) automatically reduces the Gaussian mixture $f_k^e(x_k)$. Because of the possibility of calculating $f_{k+1}^p(x_{k+1})$ analytically, this simultaneous prediction and reduction is computationally very efficient.

Remark 1 (Generalization) *Until now we assumed, that $\boldsymbol{v}_k$ and $\boldsymbol{w}_k$ are Gaussian. Generalization of the introduced approximation techniques with respect to noise represented by a Gaussian mixture is straightforward. For general densities it is possible to first find a Gaussian mixture approximation of $f^v(v_k)$ and $f^w(w_k)$, e.g. using the method proposed in [4], and then to approximate the conditional and transition density afterwards.*

VI. SIMULATION: QUADRATIC DECAY

In this section we investigate the estimation results for the measurement equation

$$\boldsymbol{y}_k = \frac{1}{1 + \boldsymbol{x}_k^2} + \boldsymbol{v}_k \ , \tag{12}$$

introduced in Examples 1 and 2 with measurement noise $\boldsymbol{v}_k \sim \mathcal{N}(v_k, 0.1)$. We approximate the conditional density of (12) for $x_k \in \Omega = [-5, 5]$ with $L = 70$ Gaussian components, gaining a quality of $G(\underline{\eta}) = 0.225880$.

The prediction step is based on the system equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{w}_k \ , \tag{13}$$

where $\boldsymbol{w}_k \sim \mathcal{N}(w_k, 0.25)$. Although (13) is linear, we use a Gaussian mixture approximation of the transition density with $L^p = 50$ components for mixture reduction purposes. The linearity allows on-line approximation for each prediction step to dynamically cover the spread of the posterior density.

In the simulation, four consecutive filter and prediction steps are performed alternatingly, starting with the filter step and the density $f_0^x(x_0) = \mathcal{N}(x_0 + 0.5, 1)$ of the system state $\boldsymbol{x}_k$ at time step $k = 0$. The measurement sequence is

$$\hat{y}_0 = 0.4, \quad \hat{y}_1 = 0.75, \quad \hat{y}_2 = 0.5, \quad \hat{y}_3 = 0.9 \ .$$

We compare the posterior densities of our approach (denoted as Appr.) with those of the unscented Kalman filter (UKF) [7], a particle filter (PF) with 700 samples and systematic resampling [2], and the exact Bayesian estimator. Recursive estimation with the exact Bayesian estimator requires recursively applied numerical integration and is used as reference. Fig. 6 shows the resulting posterior densities of the system state $\boldsymbol{x}_k$ for the four consecutive filter and prediction step pairs at time $k = 0, \ldots, 3$. It is obvious that there is almost no shape difference between the estimates of the Bayesian estimator and our approach. Especially both modes are approximated almost exactly. The same is true for the means and standard deviations, as shown in Table I. Since the prediction simultaneously provides a Gaussian mixture reduction, the number of components in $f_{k+1}^p(x_{k+1})$ and $f_k^e(x_k)$ stays constant at 50 and 3500 respectively, without impairing the estimation results significantly.

Since the UKF provides a Gaussian density approximation, whose mean is accurate up to second-order, the estimated mean is relatively close to the true one. In contrast, the difference in shape and standard deviation is significant. Due to the shape approximation of the conditional density, our approach is also able to cover higher-order moments and the shape of the posterior density. Like the proposed approach the time consumption of the UKF is constant, but about two order of magnitudes less. However, our approach provides estimates with high accuracy, whose calculation need computing time that is close to real-time.

The density representation provided by the particle filter depends on randomly drawn samples. Thus, this representation is inappropriate for well-fitting density approximations,
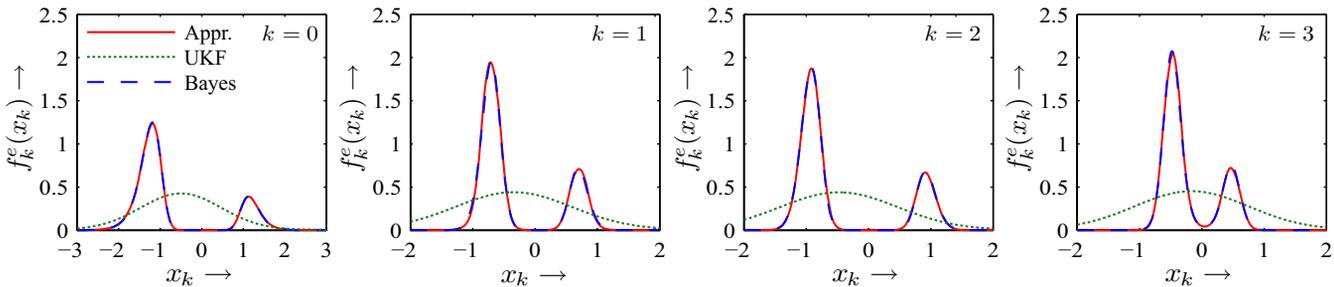
Fig. 6. The results of the approach of this paper (red, solid) in comparison with those of the Bayesian estimator (blue, dashed) and the unscented Kalman filter (green, dotted) are depicted. The particle filter provides only a sample representation and thus is omitted.

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE POSTERIOR DENSITIES.

	mean: μ_k^e				standard deviation: σ_k^e			
k	Bayes	Appr.	UKF	PF	Bayes	Appr.	UKF	PF
0	-0.72	-0.73	-0.50	-0.70	1.07	1.08	0.97	1.11
1	-0.33	-0.33	-0.35	-0.40	0.65	0.64	0.94	0.86
2	-0.44	-0.45	-0.47	-0.43	0.84	0.83	0.94	0.90
3	-0.22	-0.22	-0.16	-0.27	0.44	0.43	0.91	0.71

but very convenient for estimating moments. In Table I the average mean and standard deviation estimates of the PF over 50 simulation runs are recorded. The mean estimates are comparable to those provided by the UKF, while the standard deviations are more accurate. Only by drastically increasing the number of samples and thus the computation time, the PF results would get close to those of the proposed approach.

VII. CONCLUSIONS AND FUTURE WORK

The novel approach for closed-form measurement updating of dynamic time-invariant nonlinear systems introduced in this paper is based on the approximation of conditional densities by means of axis-aligned Gaussian mixtures. Given the Gaussian mixture approximation of the conditional density and an actual measurement, an on-line generation of the likelihood for performing the filter step is provided and results in an analytic calculation of the approximate posterior density. In contrast to the extended Kalman filter or unscented Kalman filter, the Gaussian mixture representation of our approach allows an accurate approximation of the posterior density, especially with regard to higher-order moments and a multimodal shape. Whereas particle filters only use a discrete approximation, the proposed estimation technique is able to give a continuous representation.

The exponential growth of the number of components of the approximate posterior density can be compensated by applying the closed-form prediction step derived in [6]. In doing so, performing predictions implicitly achieves Gaussian mixture reduction in an efficient manner.

Foundation of the proposed approach is an accurate conditional density approximation. To achieve approximation results in high quality and to avoid from getting trapped in local optima, a progressive optimization algorithm is proposed. Since the conditional density is time-invariant, the computationally demanding approximation can be executed off-line.

The described approach has been introduced for scalar random variables for the sake of brevity and clarity. Generalization to random vectors is straightforward. At the moment, the approach is restricted to time-invariant systems. Extension to time-variant systems is part of further research.

VIII. ACKNOWLEDGEMENTS

REFERENCES

[1] 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.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non–Gaussian Bayesian Tracking," *IEEE Transactions of Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[3] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. John Wiley and Sons Ltd, 2000.

[4] 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.

[5] S. Haykin, *Communication Systems*, 4th ed. John Wiley & Sons, Inc., 2001.

[6] 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*, 2006, pp. 98–103.

[7] 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.

[8] 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.

[9] P. S. Maybeck and B. D. Smith, "Multiple Model Tracker Based on Gaussian Mixture Reduction for Maneuvering Targets in Clutter," in *8th International Conference on Information Fusion*, vol. 1, 2005, pp. 40–47.

[10] V. Maz'ya and G. Schmidt, "On Approximate Approximations using Gaussian Kernels," *IMA Journal of Numerical Analysis*, vol. 16, no. 1, pp. 13–29, 1996.

[11] N. Oudjane and C. Musso, "Progressive Correction for Regularized Particle Filters," in *Proceedings of the 3rd International Conference on Information Fusion*, 2000, pp. THB2/10–THB2/17.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.

[13] D. J. Salmond, "Mixture Reduction Algorithms for Target Tracking in Clutter," in *SPIE Signal and Data Processing of Small Targets*, ser. 1305, April 1990, pp. 434–445.