

# Closed-Form Prediction of Nonlinear Dynamic Systems by Means of Gaussian Mixture Approximation of the Transition Density

Marco Huber, Dietrich Brunn, and Uwe D. Hanebeck

**Abstract**—Recursive prediction of the state of a nonlinear stochastic dynamic system cannot be efficiently performed in general, since the complexity of the probability density function characterizing the system state increases with every prediction step. Thus, representing the density in an exact closed-form manner is too complex or even impossible. So, an appropriate approximation of the density is required. Instead of directly approximating the predicted density, we propose the approximation of the transition density by means of Gaussian mixtures. We treat the approximation task as an optimization problem that is solved offline via progressive processing to bypass initialization problems and to achieve high quality approximations. Once having calculated the transition density approximation offline, prediction can be performed efficiently resulting in a closed-form density representation with constant complexity.

## I. INTRODUCTION

Estimation of uncertain quantities is a typical challenge in many engineering applications like information processing in sensor-actuator-networks, localization of vehicles or robotics and machine learning. One aspect that arises is the inference of a given uncertain quantity through time. Particularly the recursive processing of this so-called prediction requires an efficient implementation for practical applications.

Typically, random variables are used to describe the quantities and their uncertainties. For such a representation the prediction problem is solved by the Bayesian estimator. In general, the probability density of the predicted quantity cannot be calculated in closed form and the complexity of the density representation increases with each time step. The consequence of this is an impractical computational effort. Only for some special cases full analytical solutions are available. For linear systems with Gaussian random variables the Kalman filter provides exact solutions in an efficient manner [7]. Versatile approximative techniques exist for the case of nonlinear systems: To overcome the problem of representing the whole predicted density, particle filters use samples instead [2]. They are easy to implement and to parallelise, but it is still a hard task to obtain adequate samples at every prediction step. Another possibility arises from the usage of generic parameterized density functions. The well known extended Kalman filter uses linearization to apply the Kalman filter equations on nonlinear systems [10], while the unscented Kalman filter offers in addition higher order accuracy by using a deterministic sampling approach [6]. The resulting single Gaussian density of both estimation

methods is typically not a sufficient representation for the true complex density. Due to their universal approximation property, Gaussian mixtures [8] are a much better approach for parameterized density functions. The bandwidth of estimators using Gaussian mixtures is wide. It 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 [5].

In this paper, we introduce a new closed-form prediction approach for nonlinear systems by means of a Gaussian mixture approximation of the transition density. The transition density is used to propagate the probability density of the current system state to the next time step. By approximating it, the prediction step can be solved analytically and results in a Gaussian mixture representation of the predicted density. To avoid getting trapped in local optima, and thus ensure a high accuracy, we approximate the transition density in a progressive way. For that purpose, a parameterized transition density is introduced, which starts from a simple density and continuously approaches the true transition density. The necessary demanding computations can be calculated offline, whereas the prediction still remains an online task, that is reduced to simple multiplications of Gaussian densities. Hence, an efficient prediction is available.

In the following section, we will review the Bayesian estimator for discrete-time systems and point out the relation between transition density and system model. Furthermore, the requirements for offline approximation are formulated. The rest of the paper is structured as follows: Section III derives the closed-form prediction of nonlinear systems using a special case of Gaussian mixtures for transition density approximation. The actual approximation and its progressive processing is explained in Section IV together with an example application comprising a cubic system model. Section V further investigates the cubic system model in order to compare and discuss the results of the new prediction approach with those of the extended Kalman filter and the Bayesian estimator. The paper closes with a conclusion and an outlook to future work.

## II. PROBLEM FORMULATION

For the sake of brevity and clarity, we only consider scalar random variables, denoted by boldface letters, e.g.  $\boldsymbol{x}$ . Furthermore, we regard nonlinear, time-invariant, discrete-time systems with a system equation

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

Marco 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. {mhuber|brunn}@ira.uka.de, uwe.hanebeck@ieee.org

where  $x_k$  is the scalar system state at time step  $k$  and  $w_k$  is additive noise representing the unknown disturbance acting upon the system. It is assumed as a white, stationary Gaussian random process with density  $f^w(w_k) = \mathcal{N}(w_k - \mu^w, \sigma^w)$ , where  $\mu^w$  is the mean and  $\sigma^w$  is the standard deviation.

To simplify matters, we just consider the state evolution in absence of a system input as in (1). All results of this paper also hold in the presence of a non-zero system input.

While dealing with the measurement or filter step depending on the measurement equation is subject of future work, we focus in this paper on the system equation (1). Given an estimate  $f_0^x(x_0)$  for  $x_0$  at  $k = 0$ , this equation is used in a Bayesian setting for a recursive system state propagation in time. According to [11] this so-called *prediction step* of the Bayesian estimator results in a density

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

for  $x_{k+1}$ , where  $f^T(x_{k+1})$  is the *transition density*

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

which depends upon the noise density of  $w_k$  and the structure of the system equation. Since (1) is time-invariant and  $w_k$  is stationary, this bivariate density is also time-invariant, i.e., its shape is constant for all time steps  $k$ .

In general, the recursive Bayesian estimator is computationally impractical. The complex shape of the transition density  $f^T(x_{k+1})$  prevents a closed-form and above all an efficient solution of (2). In general, no exact analytical density can be generated in the prediction step. Hence, for the general case of nonlinear systems with arbitrary distributed random variables an approximation of the true predicted density is inevitable. From now on true densities will be denoted by a tilde, e.g.  $\tilde{f}(\cdot)$ , while the corresponding approximation will be denoted by  $f(\cdot)$ .

Since directly approximating the true predicted density  $\tilde{f}_{k+1}^x(x_{k+1})$  is difficult, we use a Gaussian mixture representation  $f^T(x_{k+1}, \underline{\eta})$  of  $\tilde{f}^T(x_{k+1})$  for approximation purposes<sup>1</sup>, that depends upon the parameter vector  $\underline{\eta}$ . For high quality approximations, an appropriate parameter vector  $\underline{\eta}$  has to be calculated, that minimizes a given distance measure  $G(\underline{\eta})$  between the true transition density  $\tilde{f}^T(x_{k+1})$  and its approximation  $f^T(x_{k+1}, \underline{\eta})$ . This resulting optimization task can be solved offline and independent of the prediction as shown in Figure 1. For this we take advantage of the fact that in real systems the system state is usually restricted to a finite interval, i.e.,

$$\forall k : x_k \in [a, b] =: \Omega . \quad (3)$$

So, we are only interested in approximating the transition density for  $x_k \in \Omega$ . Together with the property of time-invariance of  $\tilde{f}^T(x_{k+1})$ , offline approximation is possible. Section IV explains the approximation in detail.

<sup>1</sup>The reader is reminded that  $\tilde{f}^T(x_{k+1})$  and  $f^T(x_{k+1}, \underline{\eta})$  should always be regarded as functions of  $(x_k, x_{k+1})$ .

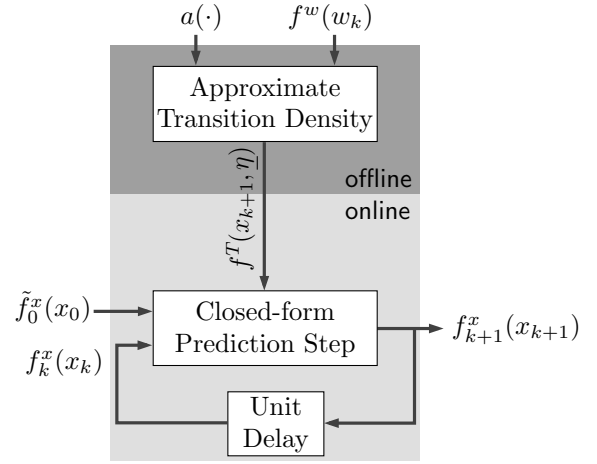


Fig. 1. Recursive, closed-form prediction. The necessary transition density approximation is performed offline, before the prediction. That one remains an online task.

### III. THE PREDICTION STEP

The transition density approximation allows to perform an efficient, closed-form prediction step online, depicted in Figure 1. For this purpose we assume that all involved densities are represented as Gaussian mixtures.

First we assume  $f_k^x(x_k)$  is given by

$$f_k^x(x_k) = \sum_{j=1}^{L_k} w_{k,j}^x \mathcal{N}(x_k - \mu_{k,j}^x, \sigma_{k,j}^x) , \quad (4)$$

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

For the given Gaussian mixture approximation  $f^T(x_{k+1}, \underline{\eta})$  we use the special case of a Gaussian mixture with *axis-aligned* Gaussian components (short: *axis-aligned Gaussian mixture*). Here, every component is separable in each dimension according to

$$f^T(x_{k+1}, \underline{\eta}) = \sum_{i=1}^{L_T} w_i \mathcal{N}(x_k - \mu_i^1, \sigma_i^1) \mathcal{N}(x_{k+1} - \mu_i^2, \sigma_i^2) , \quad (5)$$

with the parameter vector

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

where

$$\underline{\eta}_i = [w_i, \mu_i^1, \sigma_i^1, \mu_i^2, \sigma_i^2]^T .$$

Such a representation of  $f^T(x_{k+1}, \underline{\eta})$  is very convenient for efficiently performing the prediction step.

#### **Theorem 1 (Approximate Predicted Density)**

Given the Gaussian mixture representations (4) and (5) for  $f_k^x(x_k)$  and  $f^T(x_{k+1}, \underline{\eta})$  respectively, the approximate predicted density  $f_{k+1}^x(x_{k+1})$  is also a Gaussian mixture with  $L_T$  components that can be calculated analytically.

PROOF. Using the Bayesian prediction equation (2) we obtain

$$\begin{aligned}
f_{k+1}^x(x_{k+1}) &= \int_{\mathbb{R}} f^T(x_{k+1}, \underline{\eta}) f_k^x(x_k) dx_k \\
&= \sum_{i=1}^{L_T} w_i \mathcal{N}(x_{k+1} - \mu_i^2, \sigma_i^2) \left( \sum_{j=1}^{L_k} w_{k,j}^x \right. \\
&\quad \left. \int_{\mathbb{R}} \underbrace{\mathcal{N}(x_k - \mu_i^1, \sigma_i^1) \mathcal{N}(x_k - \mu_{k,j}^x, \sigma_{k,j}^x)}_{=: z_{i,j} \text{ (constant)}} dx_k \right) \\
&= \sum_{i=1}^{L_T} w_{k+1,i} \mathcal{N}(x_{k+1} - \mu_i^2, \sigma_i^2) \quad (6)
\end{aligned}$$

with  $w_{k+1,i} = w_i \sum_{j=1}^{L_k} w_{k,j}^x z_{i,j}$ .

The number of components in  $f_{k+1}^x(x_{k+1})$  depends only on the number of components in  $f^T(x_{k+1}, \underline{\eta})$ . Thus, the complexity of  $f_{k+1}^x(x_{k+1})$  remains constant over time.

For getting the result in (6), only the integral over a multiplication of two Gaussian densities, denoted as  $z_{i,j}$ , has to be solved. This corresponds exactly to the prediction step of a Kalman filter. Hence, (6) provides the closed-form and efficient solution for the prediction step by means of the Gaussian mixture approximation of a transition density.  $\square$

Generally, the more components  $f^T(x_{k+1}, \underline{\eta})$  contains, the more accurate the approximation of  $\tilde{f}_{k+1}^x(x_{k+1})$  is. Using a non axis-aligned Gaussian mixture for  $f^T(x_{k+1}, \underline{\eta})$  instead an exponential growth of components for  $f_{k+1}^x(x_{k+1})$  would be the consequence.

The two following remarks give attention to some consequences of Theorem 1.

**Remark 1 (Gaussian Mixture Reduction)** Several estimators using Gaussian mixture density representations suffer from the exponential growth of components. Theorem 1 offers a simple method for Gaussian mixture component reduction by applying additional prediction steps with the transition density approximation of the linear system

$$\mathbf{x}_{k+1} = \mathbf{A} \cdot \mathbf{x}_k + \mathbf{w}_k .$$

To keep the introduced error small,  $\mathbf{w}_k$  is Gaussian with zero-mean and standard deviation  $0 < \sigma^w \ll 1$ .

**Remark 2 (Normalization)** Typically  $f_{k+1}^x(x_{k+1})$  is not a valid probability density function, because  $\sum_{i=1}^{L_T} w_{k+1,i} \neq 1$ . This originates from the fact, that  $f^T(x_{k+1}, \underline{\eta})$  is just an approximation of the true transition density. To achieve a valid probability density for  $f_{k+1}^x(x_{k+1})$  we have to normalize it by multiplication with  $\frac{1}{\sum_{i=1}^{L_T} w_{k+1,i}}$ .

#### IV. APPROXIMATION OF THE TRANSITION DENSITY

The quality of the approximation  $f_{k+1}^x(x_{k+1})$  strongly depends on the similarity between  $\tilde{f}^T(x_{k+1})$  and its Gaussian mixture approximation  $f^T(x_{k+1}, \underline{\eta})$  for  $x_k \in \Omega$ . So, this section is concerned with solving the optimization problem

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

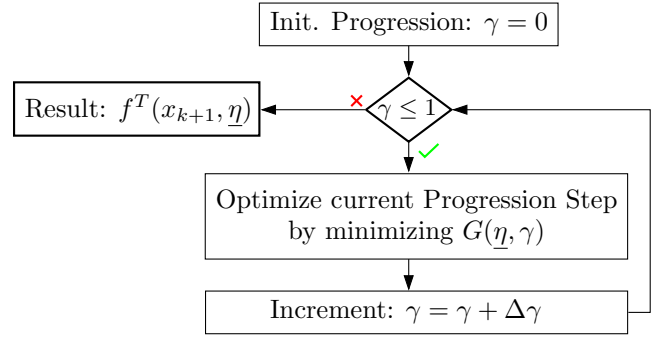


Fig. 2. Flow chart of the progressive processing of  $\underline{\eta}_{min}$ .

that yields the parameter vector for  $f^T(x_{k+1}, \underline{\eta})$  minimizing the distance to  $\tilde{f}^T(x_{k+1})$ . As distance measure we take the squared integral measure

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

Although this measure has been selected for its simplicity and convenience, it has been found to give excellent performance. Of course, the methods of this paper are not restricted to this measure.

Independent of the selected distance measure, Gaussian mixture approximations are considered a tough problem. In general, no closed-form solution of (7) can be derived. In addition, the high dimension of  $\underline{\eta}$  complicates the selection of an initial solution, so that the direct application of numerical minimization routines causes insufficient local optima for  $\underline{\eta}$ .

#### A. Approximation by Means of Progressive Processing

Instead of attempting to directly approximate the transition density, we pursue a progressive approach for finding  $\underline{\eta}_{min}$  as shown in Figure 2. This type of processing has been proposed in [5], [9]. In doing so, a parameterized transition density  $\tilde{f}^T(x_{k+1}, \gamma)$  with the progression parameter  $\gamma \in [0, 1]$  is introduced. This progression parameter ensures a continuous transformation of the solution of an initial, tractable optimization problem towards the desired true transition density  $\tilde{f}^T(x_{k+1})$  by tracking a gradually changing distance measure  $G(\underline{\eta}, \gamma)$ .

For  $\gamma = 0$  the initial optimization consists of approximating the transition density of the linear system

$$\mathbf{x}_{k+1} = \mathbf{A} \cdot \mathbf{x}_k + \mathbf{w}_k , \quad (9)$$

where  $\mathbf{A} \in \mathbb{R}$ . As discussed in Section IV-E, this problem easily allows calculating an optimum by numerical optimization without an initial parameter selection by the user. Starting from this optimum the progression parameter  $\gamma$  is gradually incremented by  $\Delta\gamma$ . In every single so-called progression step the distance measure  $G(\underline{\eta}, \gamma)$  between the parameterized transition density  $\tilde{f}^T(x_{k+1}, \gamma)$  and its approximation  $f^T(x_{k+1}, \underline{\eta})$  is minimized by employing of the BFGS formula [3], a well known optimization method.

The approximation  $f^T(x_{k+1}, \underline{\eta})$ , or more precisely the parameter vector  $\underline{\eta}$ , follows gradually  $\tilde{f}^T(x_{k+1}, \gamma)$  until the

desired true transition density  $\tilde{f}^T(x_{k+1})$  is finally reached for  $\gamma = 1$ . Hence, for  $\tilde{f}^T(x_{k+1}, \gamma)$  we obtain

$$\begin{aligned} \tilde{f}^T(x_{k+1}, \gamma = 0) &= f^w(x_{k+1} - A \cdot x_k) , \\ \tilde{f}^T(x_{k+1}, \gamma = 1) &= \tilde{f}^T(x_{k+1}) , \end{aligned} \quad (10)$$

if  $x_k \in \Omega$ , otherwise

$$\tilde{f}^T(x_{k+1}, \gamma) = 0 .$$

This progressive processing of (7) bypasses the choice of insufficient starting parameters. Hence, the typical problem of obtaining suboptimal solutions is attenuated or even prevented with a proper choice of  $\Delta\gamma$ .

### B. Parameterized System Function

To introduce the parameterized transition density, we use the *parameterized system function*  $a(\mathbf{x}_k, \gamma)$

$$a(\mathbf{x}_k, \gamma) = (1 - \gamma)A \cdot \mathbf{x}_k + \gamma a(\mathbf{x}_k) ,$$

where in particular

$$\begin{aligned} a(\mathbf{x}_k, \gamma = 0) &= A \cdot \mathbf{x}_k , \\ a(\mathbf{x}_k, \gamma = 1) &= a(\mathbf{x}_k) . \end{aligned}$$

This yields the modified system equation

$$\mathbf{x}_{k+1} = a(\mathbf{x}_k, \gamma) + \mathbf{w}_k . \quad (11)$$

The dependence of  $\tilde{f}^T(x_{k+1}, \gamma) = f^w(x_{k+1} - a(\mathbf{x}_k, \gamma))$  on system equation (11) automatically causes its parameterization according to (10).

**Example 1 (Cubic System Function)** Considering the system equation  $\mathbf{x}_{k+1} = a(\mathbf{x}_k) + \mathbf{w}_k$  with  $a(\mathbf{x}_k) = 2\mathbf{x}_k - 0.5\mathbf{x}_k^3$ , the corresponding parameterized system function is  $a(\mathbf{x}_k, \gamma) = (1 - \gamma)A \cdot \mathbf{x}_k + \gamma(2\mathbf{x}_k - 0.5\mathbf{x}_k^3)$ . Figure 3(a) shows the progression for  $\Delta\gamma = 0.2$ ,  $A = 0$  and  $\Omega = [-3, 3]$ . The parameterized transition density  $\tilde{f}^T(x_{k+1}, \gamma)$  performs the same transformation.

### C. Axis-Aligned Gaussian Mixture Approximation

Using axis-aligned Gaussian mixtures for approximating  $\tilde{f}^T(x_{k+1})$  is also advantageous for optimization purposes. An axis-aligned Gaussian mixture has minor approximation capabilities compared to a non axis-aligned one. Hence, more components are needed to achieve a comparable approximation quality. In exchange, the covariance matrix of axis-aligned Gaussian mixtures is diagonal. Thus, less parameters for a single component have to be adjusted and the necessary determination of the gradient  $\frac{\partial G}{\partial \underline{\eta}}$  prove to be easier. Altogether, representing  $f^T(x_{k+1}, \underline{\eta})$  as in (5) lowers the algorithmic complexity.

During the progression it is possible that the weights  $w_i$  of  $f^T(x_{k+1}, \underline{\eta})$  become negative. To ensure a valid density function, we use quadratic weights instead. Henceforth, we write in difference to (5)

$$f^T(x_{k+1}, \underline{\eta}) = \sum_{i=1}^{L_T} w_i^2 \mathcal{N}(x_k - \mu_i^1, \sigma_i^1) \mathcal{N}(x_{k+1} - \mu_i^2, \sigma_i^2) ,$$

without affecting the result of the prediction in principle. It must be pointed out that the probability mass of  $\tilde{f}^T(x_{k+1})$  is not equal to 1, as it is a conditional density.

### D. Squared Integral Distance Measure

The goal of the progression is to calculate the parameters  $\underline{\eta}$  minimizing  $G(\underline{\eta})$ . Thus, we use also the squared integral distance measure for  $G(\underline{\eta}, \gamma)$  by plugging the progressive version  $\tilde{f}^T(x_{k+1}, \gamma)$  of  $\tilde{f}^T(x_{k+1})$  in (8). Converting (8) and considering (10) results in

$$\begin{aligned} G(\underline{\eta}, \gamma) &= \frac{1}{2} \iint_{\mathbb{R} \Omega} \left( \tilde{f}^T(x_{k+1}, \gamma) \right)^2 dx_k dx_{k+1} \\ &\quad - \underbrace{\iint_{\mathbb{R} \Omega} \tilde{f}^T(x_{k+1}, \gamma) f^T(x_{k+1}, \underline{\eta}) dx_k dx_{k+1}}_{=: I} \\ &\quad + \frac{1}{2} \iint_{\mathbb{R} \mathbb{R}} \left( f^T(x_{k+1}, \underline{\eta}) \right)^2 dx_k dx_{k+1} , \end{aligned} \quad (12)$$

where merely the integral  $I$  cannot be solved analytically. Numerical integration methods like the adaptive Simpson quadrature [4] have to be applied.

The necessary condition for the existence of a minimum of  $G(\underline{\eta}, \gamma)$  for a given  $\gamma$  is

$$\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}} = \underline{0} . \quad (13)$$

Since (13) allows no closed-form solution we use the BFGS formula for minimization, which depends on calculating the gradient  $\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}}$ . To obtain the gradient it is sufficient to examine only the  $i$ -th component  $f_i^T(x_{k+1}, \underline{\eta}_i)$  of  $f^T(x_{k+1}, \underline{\eta})$

$$\begin{aligned} \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}_i} &= - \iint_{\mathbb{R} \Omega} \tilde{f}^T(x_{k+1}, \gamma) \frac{\partial f_i^T(x_{k+1}, \underline{\eta}_i)}{\partial \underline{\eta}_i} dx_k dx_{k+1} \\ &\quad + \iint_{\mathbb{R} \mathbb{R}} f^T(x_{k+1}, \underline{\eta}) \frac{\partial f_i^T(x_{k+1}, \underline{\eta}_i)}{\partial \underline{\eta}_i} dx_k dx_{k+1} , \end{aligned}$$

with

$$f_i^T(x_{k+1}, \underline{\eta}_i) = w_i^2 \mathcal{N}(x_k - \mu_i^1, \sigma_i^1) \mathcal{N}(x_{k+1} - \mu_i^2, \sigma_i^2) .$$

Stacking all  $i = 1, \dots, L_T$  partial derivatives leads to  $\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}}$ . Evaluating the gradient requires also numerical integration.

### E. Initialization

A complete analytical solution of (12) is given for  $\gamma = 0$ . Here, the parameterized transition density

$$\begin{aligned} \tilde{f}^T(x_{k+1}, \gamma = 0) &= f^w(x_{k+1} - A \cdot x_k) \\ &= \mathcal{N}(x_{k+1} - A \cdot x_k - \mu^w, \sigma^w) \end{aligned}$$

depends on the linear system equation (9). We can take advantage of the linearity as we initialize the progression by avoiding the selection of initial parameters  $\underline{\eta}$  for  $f^T(x_{k+1}, \underline{\eta})$  by the user. For this, fixed and equidistant means for the components of  $f^T(x_{k+1}, \underline{\eta})$  are given by

$$\begin{aligned} \mu_i^1 &= a + i \cdot \frac{b - a}{L_T + 1} , \\ \mu_i^2 &= A \cdot \mu_i^1 + \mu^w . \end{aligned}$$

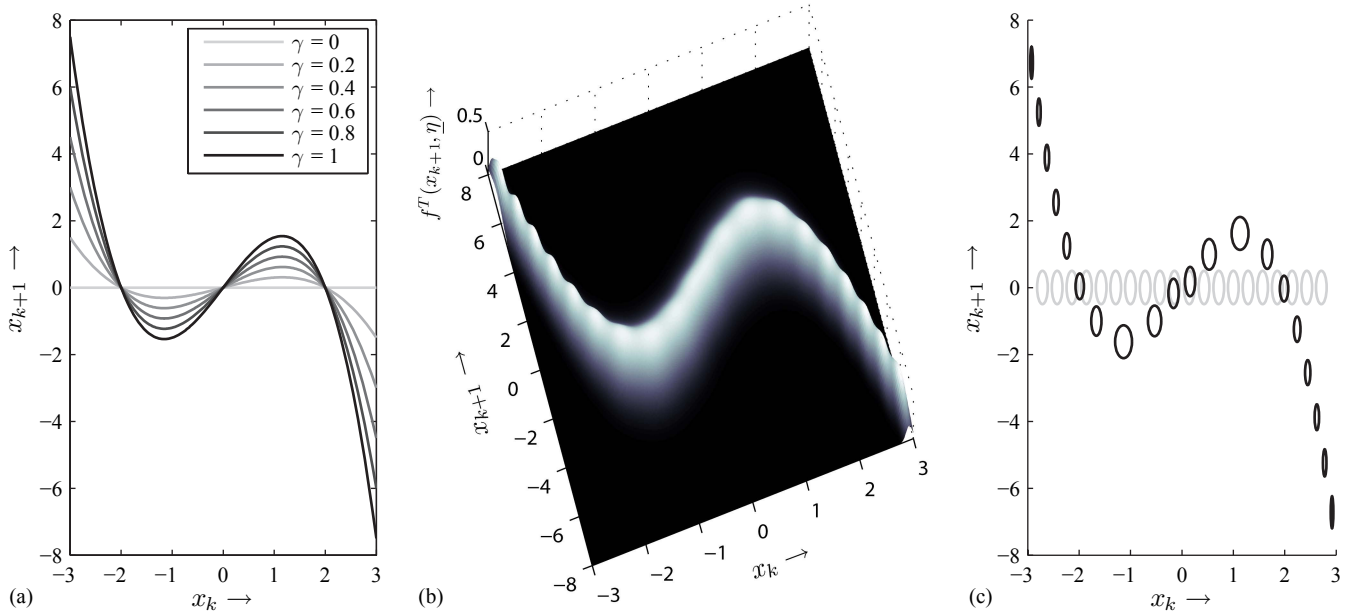


Fig. 3. (a) Progression of the parameterized system function  $a(\mathbf{x}_k, \gamma) = (1 - \gamma)A \cdot \mathbf{x}_k + \gamma(2\mathbf{x}_k - 0.5\mathbf{x}_k^3)$ . (b) Approximation of the transition density  $\tilde{f}^T(x_{k+1}) = \mathcal{N}(x_{k+1} - (2x_k - 0.5x_k^3), 1)$  with quality  $G(\underline{\eta}) = 0.0067$ . (c) Covariance ellipses of the components for  $\gamma = 0$  (gray) and  $\gamma = 1$  (black).

Thus, only the weighting coefficients  $w_i$  and the standard deviations  $\sigma_i^1, \sigma_i^2$  are adjustable. A further parameter reduction occurs by taking the same weights and standard deviations for all components. As a result, there are just the three parameters

$$w_i = w, \quad \sigma_i^1 = \sigma^1, \quad \sigma_i^2 = \sigma^2$$

that have to be optimized. This optimization problem has one single global optimum. So, no user initialization is required and as a consequence the risk of starting the progression with an insufficient local optimum is obviously avoided.

**Example 2 (Cubic System Function (cont'd.))** We consider again the cubic system equation of Example 1, now with system noise  $w_k = \mathcal{N}(w_k - 0, 1)$  and  $\Omega = [-3, 3]$ . Using the parameterized system function of Example 1 with  $A = 0$  and a Gaussian mixture with  $L_T = 20$  components leads to the transition density approximation shown in Figure 3(b). Figure 3(c) depicts the covariance ellipses of the single Gaussian components at the beginning and the end of the progression.

#### F. Generalization

Until now we assumed, that the system noise  $w_k$  is Gaussian. All the derivations of this paper can be directly generalized to noise that is represented by a Gaussian mixture. For general densities of  $w_k$  it is possible to first find a Gaussian mixture approximation of  $f^w(w_k)$  and then to approximate the transition density afterwards.

#### V. EXAMPLE: PREDICTION

In this section we investigate the prediction results for the system equation

$$\mathbf{x}_{k+1} = 2\mathbf{x}_k - 0.5\mathbf{x}_k^3 + \mathbf{w}_k,$$

introduced in Examples 1 and 2. The system noise  $w_k$  is white Gaussian with density  $f^w = \mathcal{N}(w_k - \mu^w, \sigma^w)$ , where

$\mu^w = 0$  and  $\sigma^w = 0.175$ . We approximate the transition density of this system according to Example 2 for  $x_k \in \Omega = [-3, 3]$ , but now with  $L_T = 50$  Gaussian components for  $f^T(x_{k+1}, \eta)$  since the standard deviation  $\sigma^w$  is now much smaller. This results in a quality  $G(\underline{\eta}) = 0.0207$ .

Starting with the density

$$f_0^x(x_0) = \mathcal{N}(x_0 - 0.4, 0.8)$$

of the system state  $\mathbf{x}_k$  at time step  $k = 0$  we compare the predictions of our approach (denoted as Appr.) with those of the extended Kalman filter (EKF) and the exact Bayesian estimator. Recursive prediction with the exact Bayesian estimator requires recursively applied numerical integration of (2) and is used as reference. Figure 4 shows the resulting densities of the predicted system state  $\mathbf{x}_{k+1}$  for four consecutive prediction steps at time  $k = 0, \dots, 3$ . It is obvious that there is almost no shape difference between the predictions of the Bayesian estimator and our approach. Especially both modes of the optimal predicted density  $\tilde{f}_{k+1}^x(x_{k+1})$  are approximated almost exactly. Same is true for the means, see Table I. Since the number of components in  $f_{k+1}^x(x_{k+1})$  stays constant at 50, our approach requires a constant and marginal amount of time<sup>2</sup> for prediction. For comparison see again Table I.

In contrast, the EKF is just able to provide a single Gaussian approximation of the true predicted density. Thus, the difference in shape and mean is significant. The time consumption of the EKF prediction is indeed constant and smaller. However, a Gaussian mixture approximation of the transition density allows explicit consideration of the influence of probability masses or components in the input

<sup>2</sup>The times base on measurements of a Matlab 7.1 implementation running on a PC with Microsoft Windows XP operating system and an Intel Pentium IV 3 GHz processor.

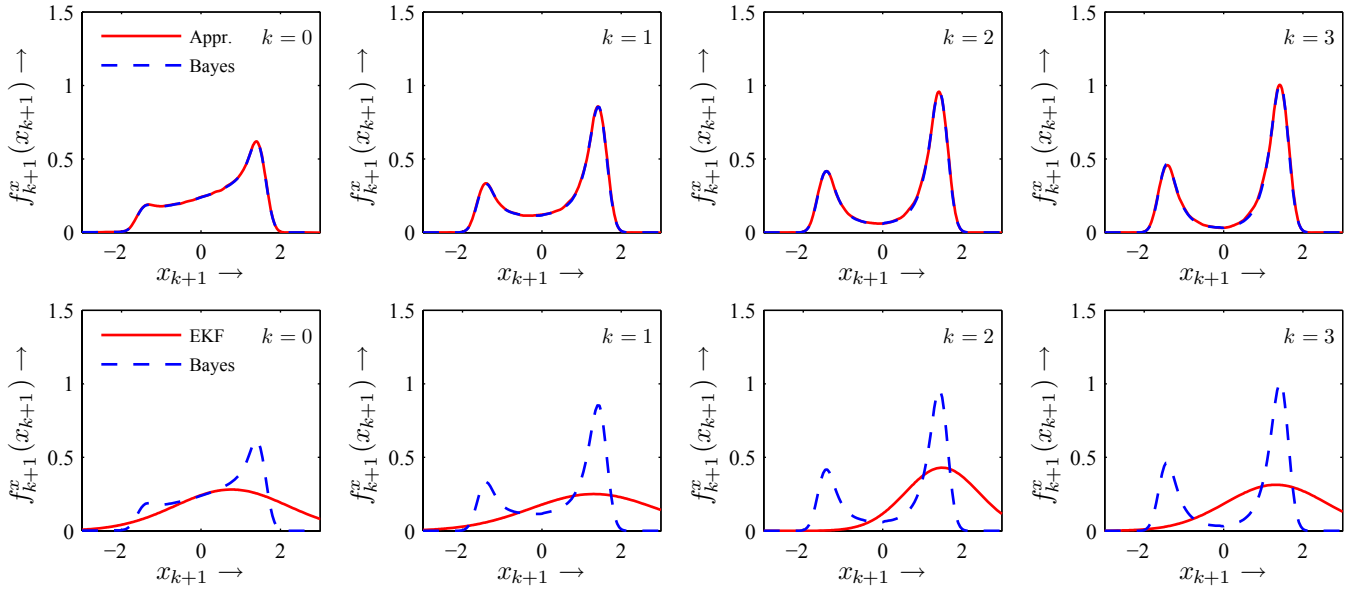


Fig. 4. The upper plots show the predicted densities using the approach of this paper (red, solid) and the Bayesian estimator (blue, dashed). The predictions of the Bayesian estimator in comparison with those of the extended Kalman filter (red, solid) are depicted on the lower plots.

TABLE I

MEANS OF THE PREDICTED DENSITIES AND THE PREDICTION TIME.

k	mean: $\mu_{k+1}^x$			time: $t/s$		
	Bayes	Appr.	EKF	Bayes	Appr.	EKF
0	0.404	0.408	0.768	6.016	0.166	0.003
1	0.447	0.454	1.31	223.8	0.172	0.005
2	0.453	0.462	1.496	14040	0.125	0.006
3	0.454	0.465	1.318	$6 \cdot 10^5$	0.125	0.002

density to the predicted density, by affecting the update of the weights  $w_{k+1,i}$  of  $f_{k+1}^x(x_{k+1})$  according to (6). So, higher approximation accuracy is available and enhanced with an increasing number of components  $L_T$ .

## VI. CONCLUSIONS AND FUTURE WORK

This paper introduced a novel approach for closed-form prediction of dynamic time-invariant nonlinear systems based on approximate transition densities. Approximating transition densities by means of Gaussian mixtures with axis-aligned components leads to an optimization problem. As most of the optimization methods suffer from getting trapped in local optima, a progressive processing is proposed that transforms the solution of an initial, tractable optimization problem continuously towards the desired transition density. As a result, high quality approximations are obtained. Since the optimization problem is solved offline, great effort can be spent without restricting the efficiency of the prediction step.

Due to the Gaussian mixture representation of the transition density, the prediction result is calculated analytically. Using axis-aligned Gaussian mixtures leads to a constant number of components describing the approximation of the predicted density. Thus, we obtain an efficient recursive prediction, whose accuracy depends on the adjustable transition density approximation. These properties were demonstrated by recursively predicting the system state of a cubic system.

The described approach has been introduced for scalar random variables for the sake of brevity and clarity. It

can be generalized to random vectors in a straightforward manner. Considering the filter step is also part of future work. Generally, the progressive processing offers room for improvement. For example, an adaptive progression parameter increment and adjustment of the number of Gaussian mixture components during the progression for additional approximation quality enhancement are possible.

## VII. ACKNOWLEDGEMENTS

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

## 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] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer-Verlag, 2001.
- [3] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. John Wiley and Sons Ltd, 2000.
- [4] W. Gander and W. Gautschi, “Adaptive Quadrature - Revisited,” *BIT*, vol. 40, no. 1, March 2000.
- [5] U. D. Hanebeck, K. Briechele, and A. Rauh, “Progressive Bayes: A New Framework for Nonlinear State Estimation,” in *Proceedings of SPIE*, vol. 5099. AeroSense Symposium, 2003.
- [6] S. J. Julier and J. K. Uhlmann, “Unscented Filtering and Nonlinear Estimation,” in *Proceedings of the IEEE*, vol. 92, no. 3, 2004.
- [7] 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.
- [8] 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.
- [9] N. Oudjane and C. Musso, “Progressive Correction for Regularized Particle Filters,” in *Proceedings of the 3rd International Conference on Information Fusion*, 2000.
- [10] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.
- [11] F. C. Scheweppe, *Uncertain Dynamic Systems*. Prentice-Hall, 1973.