

Recursive Prediction of Stochastic Nonlinear Systems Based on Optimal Dirac Mixture Approximations

Oliver C. Schrempf and Uwe D. Hanebeck

Abstract—This paper introduces a new approach to the recursive propagation of probability density functions through discrete-time stochastic nonlinear dynamic systems. An efficient recursive procedure is proposed that is based on the optimal approximation of the posterior densities after each prediction step by means of Dirac mixtures. The parameters of the individual components are selected by systematically minimizing a suitable distance measure in such a way that the future evolution of the approximate densities is as close to the exact densities as possible.

NOTATION

$\tilde{f}^x(x)$	probability density function of x
$f^x(x)$	approximation of $\tilde{f}^x(x)$
$\tilde{F}^x(x), F^x(x)$	corresponding distribution functions
$\delta(x)$	Dirac Delta function
$H(x)$	Heaviside step function
G	distance measure
η	parameter vector
γ	progression parameter
$\mathcal{N}(\cdot, m, \sigma)$	Gaussian density
k	time index

I. INTRODUCTION

In this paper, we consider the recursive propagation of probability density functions through a stochastic nonlinear dynamic system in discrete time. It is well known, that the resulting densities cannot be calculated exactly, as the type of density changes and the complexity increases. Hence, practical processing schemes call for an approximation of either the underlying system itself, which then allows the application of, e.g. linear predictors, or for an approximation of the densities resulting from the original system. Here, we focus on the latter approach, where two main types of generic density representations have been reported in the literature: i) sample-based representations [1] and ii) analytic density representations [2].

Sample-based predictors like the well known particle filter [3], [4] typically employ Monte Carlo methods in order to obtain the samples or particles required for representing the underlying continuous densities. For that purpose, random samples are typically generated by pseudo random number generators.

In their simplest form, particle filters propagate the randomly generated samples through the system and add randomly sampled system noise to it. A special type of particle filter is the so called regularized particle filter [5], [6], which

resamples from a continuous approximation of the posterior density and typically produces smoother results.

The Monte Carlo method suffers from a variety of problems caused by the randomness of the generated samples. Hence, the results are not deterministic and a large number of samples have to be applied in order to get satisfactory results. Furthermore, the required number of samples cannot be deduced a priori.

To remedy some of these problems, quasi random sequences have been proposed in [7], [8] as an alternative to random sampling. Quasi random sequences are also called low discrepancy sequences. A popular sequence is the Niederreiter series [9]. A filter based on quasi random sequences has been introduced in [10].

In this paper, we pursue a different approach, which is based on the systematic approximation of the true posteriors resulting from the prediction recursion and using this approximation as a system input for the next recursion step. For approximation purposes, Dirac mixtures are employed that, in contrast to Monte Carlo methods, are selected deterministically and are optimally close to the true densities. This type of approximation allows for exact computation of the output density of a stochastic nonlinear system for a single prediction step. Hence, this result is approximated again in order to come up with a recursion mechanism of constant complexity.

The approximations are selected in such a way that a certain measure of deviation between the true posterior and its Dirac mixture approximation is minimized. Several options for performing the approximation including the calculation of both the optimal parameters and the optimal number of components are pursued. In the case of comparing a Dirac mixture approximation to a given density, typical distance measures quantifying the distance between two densities are not well defined. Hence, we compare the corresponding cumulative distribution functions of the true density and its approximation in order to find optimal parameters for the Dirac mixture approximation as introduced in [11], [12].

The key contribution of this paper is the consideration of the system's future evolution, which is achieved by incorporating posterior densities in the optimization approach.

The paper is organized as follows: The prediction problem is formulated in Section II followed by a short sketch of the proposed solution procedure in Section III-A. We introduce Dirac mixture densities in Section III-B and give closed-form calculations for a single prediction step in Section III-C. Various approximation criteria of different computational complexity are introduced in Section IV in order to determine the optimal number of components.

Oliver C. Schrempf and Uwe D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. {schrempf|uwe.hanebeck}@ieee.org

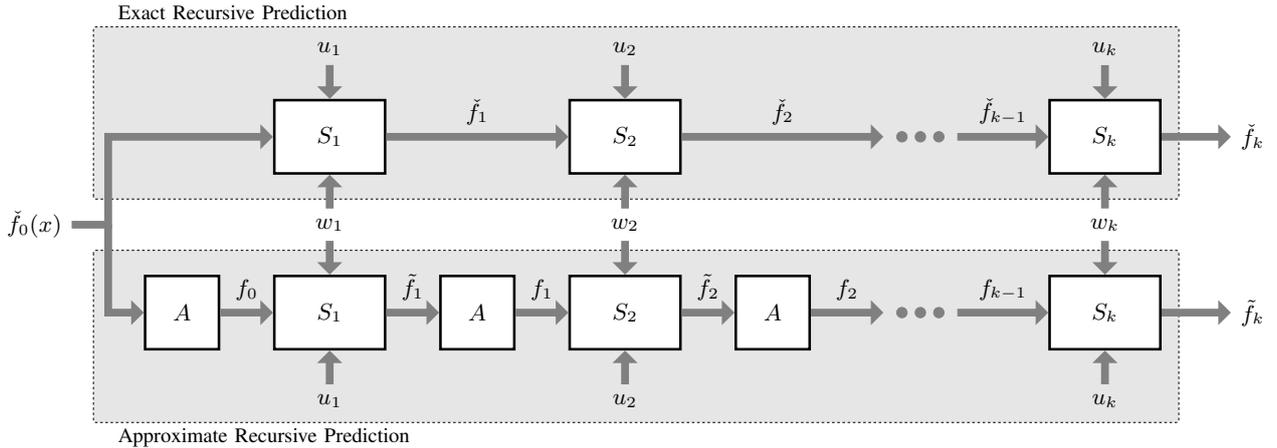


Fig. 1. Unrolled representation of recursive prediction. The top branch corresponds to successive exact prediction steps. Densities \tilde{f} denote the resulting exact densities. The bottom branch corresponds to approximate recursive prediction achieved by approximation of the resulting posterior after each prediction step. f denotes approximate densities. \tilde{f} denotes densities that result from a single exact prediction step based on the prior approximation.

Practical algorithms and distance measures are presented in Section V, followed by simulations in Section VI. Conclusions, possible extensions, and future work are given in Section VII.

II. PROBLEM FORMULATION

We consider a nonlinear dynamic system evolving in discrete time according to

$$x_{k+1} = a_k(x_k, u_k, w_k), \quad (1)$$

where x_k is the state, u_k a known input, and $a_k(\cdot)$ a time-varying nonlinear mapping. w_k represents both endogenous and exogenous noise sources acting upon the system and is described by means of a density function $f_k^w(w_k)$.

Starting from an initial density $f_0^x(x_0)$ over the initial state x_0 , our goal is to quantify the stochastic evolution of the system over a finite horizon $k = 1, \dots, N$ given the input values u_k , $k = 0, \dots, N - 1$.

This corresponds to sequentially calculating the state densities $f_k^x(x_k)$, $k = 1, \dots, N$ by means of a prediction step as depicted in the upper part of Figure 1.

Exact computation of these densities, however, is not feasible, as the complexity of the density increases in every step. In addition, the resulting densities cannot be calculated in analytic form in general.

The aim of this work is to provide a density representation that approximates the true density, in order to allow for closed-form calculation of the prediction step while maintaining a predefined quality of the approximation with respect to a given distance measure.

We restrict ourselves to scalar random variables in order to simplify the notation. In addition, the time index k is omitted in some cases without notice.

III. APPROXIMATION FOR NONLINEAR PREDICTION

A. Solution Sketch

As the exact recursive computation of the desired probability density functions is not feasible, an approximation

is inevitable. In this paper, we focus on approximating the posterior density by means of a Dirac mixture after each processing step, which serves as a prior for the next prediction step. This is depicted in the lower part of Figure 1.

In order to illustrate the basic prediction procedure and the advantages of using a Dirac mixture representation for all the occurring densities, we assume that a Dirac mixture density is already given at some time step k . This Dirac mixture is used as a prior density, i.e., as an input density, for the next prediction step from time step k to time step $k + 1$. Exploiting the properties of the Dirac delta function allows for closed-form calculation of the subsequent prediction step at time step $k + 1$. This will be discussed in Section III-C. Specifically, for the important case of nonlinear systems corrupted by additive Gaussian noise, where the resulting posterior is given by a Gaussian mixture density.

The posterior resulting from a Dirac mixture prior can easily be calculated in closed form. Due to the noise of the system it is not in general of the same type as the prior density. For recursive application of the proposed procedure, however, a Dirac mixture posterior is required as an input for the next prediction step. Hence, a systematic re-approximation of the exact posterior by means of a Dirac mixture is performed by minimizing a certain optimality criterion.

When implementing this type of approximate prediction procedure, a system designer is offered several degrees of freedom. The first choice concerns the type of Dirac mixture, which is discussed in Section III-B. Of course, there are several options for rating the quality of the approximation of the true posterior, where the most obvious is the best approximation according to some deviation measure. However, it also makes sense to include the quality of the future evolution of the approximate prediction when adjusting the parameters of the approximating function for the current time step. Four major types of approximation criteria will be introduced in Section IV. Given a certain approximation criterion, several

options for performing the actual calculation of appropriate density parameters are given in Section IV. This includes optimal optimization approaches and suboptimal but less computationally demanding approaches.

B. Dirac Mixture for Generic Density Representation

For representing arbitrary density functions, Dirac mixtures according to

$$f(x, \underline{\eta}) = \sum_{i=1}^L w_i \delta(x - x_i) \quad (2)$$

are employed according to [12], where

$$\underline{\eta} = [x_1, x_2, \dots, x_L, w_1, w_2, \dots, w_L]^T$$

is a parameter vector consisting of locations x_i , $i = 1, \dots, L$ and weighting coefficients w_i , $i = 1, \dots, L$. Since the properties of a density function $f(x)$ demand, that $f(x) \geq 0$ and $\int_{\mathbb{R}} f(x) dx = 1$, we require the weights w_i to be non-negative and to sum up to 1. Hence, we have $2L$ parameters with $2L - 1$ degrees of freedom. A simplified density representation can be given by equally weighted Dirac mixtures [11], where only L parameters and L degrees of freedom are used.

Dirac mixtures are a generic density representation for approximating complicated densities arising when propagating a given density through a nonlinear system suffering from noise. The properties of Dirac mixtures render them very useful for efficiently performing prediction of stochastic nonlinear dynamic systems, which will be shown in the next subsection.

C. Single Prediction Step

The state densities $f_{k+1}^x(x_{k+1})$, $k = 1, \dots, N$ are calculated by evaluating the Bayesian forward step, which can be performed in closed form for Dirac mixture priors. For simplicity, we will now focus on additive Gaussian noise without loss of generality.

The Bayesian forward step in this case is given by

$$f_{k+1}^x(x_{k+1}) = \int_{-\infty}^{\infty} f(x_{k+1}|x_k) f_k^x(x_k) dx_k, \quad (3)$$

where the transition density $f(x_{k+1}|x_k)$ is given by

$$f(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}, a_k(x_k, u_k), \sigma_v) .$$

For general prior densities $f_k^x(x_k)$, the integral involved in (3) cannot be solved analytically. However, for a given input point \hat{x}_k represented by the Dirac delta function $f_k^x(x_k) = \delta(x_k - \hat{x}_k)$, (3) can be solved in closed form according to

$$f_{k+1}^x(x_{k+1}) = \mathcal{N}(x_{k+1}, a_k(\hat{x}_k, u_k), \sigma_k^v) ,$$

which is a Gaussian density with a standard deviation σ_k^v corresponding to the system noise.

For a given Dirac mixture prior $f_k^x(x_k)$ according to (2) given by

$$f_k^x(x_k) = \sum_{i=1}^L w_k^i \delta(x_k - \hat{x}_k^i) , \quad (4)$$

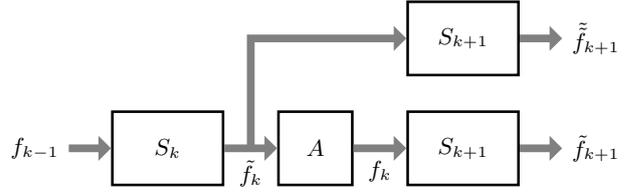


Fig. 2. Cutout of the unrolled prediction recursion shown in Figure 1 including two prediction steps and an intermediate density approximation. The input and noise values have been omitted for simplicity.

the posterior from (3) is a Gaussian mixture given by

$$f_{k+1}^x(x_{k+1}) = \sum_{i=1}^L w_i \mathcal{N}(x_{k+1}, a_k(\hat{x}_k^i, u_k), \sigma_k^v) ,$$

which is the closed form solution. Please note, that similar results can be derived for non-additive and non-Gaussian noise, see (1).

In order to perform closed-form predictions of this type recursively, the given posteriors are systematically approximated by means of Dirac mixtures with respect to a certain approximation criterion. The resulting Dirac mixtures are then used as priors for the subsequent prediction steps.

IV. APPROXIMATION CRITERIA

Given the exact posterior as described in the previous section, our goal is to find an optimal re-approximation by means of a Dirac mixture. More precisely, the goal is to calculate both the optimal number of components and their optimal parameters with respect to a certain optimality criterion.

In this section, several strategies for rating the effect of a given approximation of the prior density at time step k are introduced. The discussion focuses on the consideration of the future system evolution and is limited to the approximation structure, i.e., the location of suitable comparison points in time. Specific distance measures will be discussed in the next section.

For explaining the different approximation criteria, we consider a cutout of the unrolled prediction recursion comprising two prediction steps plus an intermediate density approximation shown in Figure 2.

A. Prior Optimization

The most obvious optimization approach is the direct comparison of the true density f_k and its approximation f_k . Although this is the most straightforward approach, a good approximation at time step k does not necessarily assure that the future evolution of the approximate state densities is close to the true states.

B. Lookahead Optimization

A more desirable approximation structure would be to find an approximation at time step k that minimizes deviation measures between true densities in the future and their respective approximations. Ideally, this would be performed for the desired, probably very large or even infinite, number of prediction steps.

Algorithm 1 Optimal # of components w.r.t. the prior density

```
1: Select max. error threshold  $G_{\max}$ 
2: Select initial number of components  $L = L_0$ 
3: Select search step  $\Delta L$ 
4: while  $G(\underline{\eta}) > G_{\max}$  do
5:    $\underline{\eta} = \text{approx}(L)$ 
6:    $\bar{L} = L + \Delta L$ 
7: end while
8:  $L_l = L - 2\Delta L$ 
9:  $L_u = L - \Delta L$ 
10: while  $L_u - L_l > 1$  do
11:    $L_t = L_l + \lfloor \frac{L_u - L_l}{2} \rfloor$ 
12:    $\underline{\eta} = \text{approx}(L_t)$ 
13:   if  $G(\underline{\eta}) > G_{\max}$  then
14:      $L_l = L_t$ 
15:   else
16:      $L_u = L_t$ 
17:   end if
18: end while
```

C. Posterior Optimization

A practical solution is limited to a few steps ahead into the future. For a single step into the future of the system's evolution, we compare the densities \tilde{f}_{k+1} and \hat{f}_{k+1} . \hat{f}_{k+1} is the "exact" density directly resulting from the approximation density f_k . \tilde{f}_{k+1} corresponds to the exact propagation of the exact density \tilde{f}_k of the previous prediction step *before* the considered approximation through the system. This approach is computationally more demanding than the first approach. However, it considers the system's future evolution while determining appropriate parameters of the approximation density at time step k .

D. Optimization of the Prior for Best Posterior Fit

An even more practical solution is the direct comparison of \tilde{f}_k and its approximation f_k as discussed in Section IV-A for the purpose of calculating optimal parameters for a given number of parameters. As an extension, the optimal number of parameters of this approximation is deduced from the comparison of the posterior densities \tilde{f}_{k+1} and \hat{f}_{k+1} one step ahead into the future.

So far, the discussion was limited to the approximation structure, i.e., to locating suitable densities used for assessing the quality of the selected approximation with respect to the future evolution of the system. The actual comparison of these densities and the calculation of appropriate approximation parameters will be systematically performed by means of distance measures. Two algorithms based on specific approximation structures and specific distance measures will be given in the next section.

V. OPTIMIZATION FOR SPECIFIC DISTANCE MEASURES

This section focuses on the two specific approximation structures given in Section IV-A and Section IV-D. Both approaches select the *parameters* of the Dirac mixture approximation of the prior density for a certain prediction step

based on comparing the prior densities. They differ, however, in the way the number of parameters are selected. The first approach according to Section IV-A that will be detailed in Section V-A also selects the *number of parameters* based on comparing the prior densities. Doing so is straightforward and just requires the evaluation of the given distance measure between the true prior and its Dirac mixture approximation. However, it is difficult for the designer to specify a measure for the approximation quality that is relevant for the future evolution of the system. Hence, the second approach according to Section IV-D that will be detailed in Section V-B considers the future evolution of the system while selecting the number of Dirac mixture components for approximating the prior. This is performed by comparing the posterior that results from the true prior and the posterior that results from the approximated prior depending upon the number of Dirac mixture components.

When comparing the continuous densities after the prediction step, standard distance measures can be applied. The comparison of the prior densities, however, requires measures of distance between a continuous density (the true prior) and its Dirac mixture approximation. Appropriate measures for this purpose will be introduced in Section V-C.

A. Optimizing for the Prior

We first present an approach for finding the optimal number of components for the approximation, considering only the prior densities as motivated in Section IV-A. Alg. 1 compares the approximation $f_k(x)$ to the true continuous density $\tilde{f}_k(x)$ with respect to a certain distance measure $G(\underline{\eta})$, where $\underline{\eta}$ is the parameter vector of the Dirac mixture approximation. A practical distance measure together with the optimal approximation method will be given in Section V-C.

For this algorithm a maximum acceptable error G_{\max} , an initial number of components L_0 , and an initial search step ΔL are selected. The rest of this algorithm is an application of a binary search algorithm.

B. Optimizing for the Posterior

The next algorithm presents the application of the approximation criterion given in Section IV-D. Alg. 2 also finds the optimal number of components, but in this case with respect to the predicted densities \tilde{f}_{k+1} and \hat{f}_{k+1} . The main difference to Alg. 1 is that first an initial approximation with a very large number of components is generated and passed through the prediction step, resulting in a continuous density representation with parameter vector $\underline{\kappa}_t$. Due to the high number of components we can assume this density to be very close to the true density \tilde{f}_{k+1} . The remainder of the algorithm is similar to Alg. 1 except for the fact that now a distance measure between two continuous density functions is applied. An efficient procedure for approximating arbitrary mixture densities with Dirac mixtures comprising a large number of components is given in Section V-D.

Algorithm 2 Optim. # of components w.r.t. the poster. dens.

```

1: Select max. Error Threshold  $G_{\max}$ 
2: Select initial number of Components  $L = L_0$ 
3: Select search step  $\Delta L$ 
4:  $\underline{\kappa}_t = \text{predict}(\text{approx}(L_{\text{large}}))$ 
5: while  $G > G_{\max}$  do
6:    $\underline{\kappa} = \text{predict}(\text{approx}(L))$ 
7:    $G = G(\underline{\kappa}_t, \underline{\kappa})$ 
8:    $L = L + \Delta L$ 
9: end while
10:  $L_l = L - 2\Delta L$ 
11:  $L_u = L - \Delta L$ 
12: while  $L_u - L_l > 1$  do
13:    $L_t = L_l + \lfloor \frac{L_u - L_l}{2} \rfloor$ 
14:    $\underline{\kappa} = \text{predict}(\text{approx}(L_t))$ 
15:    $G = G(\underline{\kappa}_t, \underline{\kappa})$ 
16:   if  $G > G_{\max}$  then
17:      $L_l = L_t$ 
18:   else
19:      $L_u = L_t$ 
20:   end if
21: end while

```

C. Optimal Approximation

Both algorithms presented in the previous sections depend on a distance measure and on an optimal approximation approach. As a distance measure for comparing continuous densities, measures like the Kullback–Leibler divergence [13] or integral quadratic measures are popular candidates. For comparing a continuous density to a Dirac mixture, however, they are not useful. Instead of comparing the densities directly, in our approach the corresponding (cumulative) distribution functions are employed for that purpose.

The distribution function corresponding to the true density $\tilde{f}(x)$ is given by

$$\tilde{F}(x) = \int_{-\infty}^x \tilde{f}(t) dt .$$

The distribution function corresponding to the Dirac mixture approximation can be written as

$$F(x, \underline{\eta}) = \int_{-\infty}^x f(t, \underline{\eta}) dt = \sum_{i=1}^L w_i H(x - x_i) , \quad (5)$$

where $H(\cdot)$ denotes the Heaviside function defined as

$$H(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & x = 0 \\ 1, & x > 0 \end{cases} .$$

A suitable distance measure is given by the weighted Cramér–von Mises distance [14]

$$G(\underline{\eta}) = \int_{-\infty}^{\infty} g(x) \left(\tilde{F}(x) - F(x, \underline{\eta}) \right)^2 dx , \quad (6)$$

where $g(x)$ is a non-negative weighting function. In the specific application, $g(x)$ is selected in such a way that only those portions of the considered probability density function

relevant for the future evolution are approximated with a high accuracy. This avoids to put much approximation effort into irrelevant regions of the state space. The goal is now to find a parameter vector $\underline{\eta}$ that minimizes (6) according to $\underline{\eta} = \arg \min_{\underline{\eta}} G(\underline{\eta})$.

Unfortunately it is not possible to solve this optimization problem directly in closed form. Hence, we introduce an approach to find the solution progressively by applying the homotopy continuation method [12]. In order to apply this method, we introduce a so called progression parameter γ into $\tilde{F}(x)$ that goes from 0...1. The purpose of this parameter is to find a simple and exact approximation of $\tilde{F}(x, \gamma)$ for $\gamma = 0$. Furthermore we must guarantee that $\tilde{F}(x, \gamma = 1) = \tilde{F}(x)$. By varying γ from 0 to 1 we track the parameter vector $\underline{\eta}$ that minimizes the distance measure.

In order to find the minimum of the distance measure, we have to find the root of the partial derivative with respect to $\underline{\eta}$ according to

$$\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}} = \begin{bmatrix} \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{x}} \\ \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{w}} \end{bmatrix} \stackrel{!}{=} \underline{0} . \quad (7)$$

To track the minimum of the distance measure we have to take the derivative of (7) with respect to γ . This results in a system of ordinary first order differential equations that can be written in vector–matrix–form as

$$\underline{\dot{b}} = \mathbf{P} \underline{\dot{\eta}} , \quad (8)$$

where

$$\underline{\dot{b}} = \begin{bmatrix} \frac{\partial \tilde{F}(x_1, \gamma)}{\partial \gamma} \\ \vdots \\ \frac{\partial \tilde{F}(x_L, \gamma)}{\partial \gamma} \\ \int_{x_0}^c \frac{\partial \tilde{F}(x, \gamma)}{\partial \gamma} dx \\ \int_{x_1}^c \frac{\partial \tilde{F}(x, \gamma)}{\partial \gamma} dx \\ \vdots \\ \int_{x_L}^c \frac{\partial \tilde{F}(x, \gamma)}{\partial \gamma} dx \end{bmatrix}$$

and

$$\underline{\dot{\eta}} = \frac{\partial \underline{\eta}}{\partial \gamma} = [\dot{x}_1, \dots, \dot{x}_L, \dot{w}_0, \dot{w}_1, \dots, \dot{w}_L]^T .$$

$\underline{\dot{\eta}}$ denotes the derivative of $\underline{\eta}$ with respect to γ . The \mathbf{P} matrix as well as the derivations are given in [12]. The approximation of $\tilde{f}(x)$ now boils down to solving (8) on the interval $[0, 1]$.

D. Suboptimal Approximation

For a growing number of Dirac mixture components, the approximation procedures become more and more difficult to implement. Of course, the computational demand of the optimal approximation also increases.

For a large number of samples, two alternative approaches are available. The first approach relies on Dirac mixtures with equally weighted components. This, however, might not yield the desired results. In addition, this approach also requires an optimization procedure. The second alternative is especially

interesting for the approximation of mixture densities, which are important in the context of this paper. It relies on the simple fact that a suboptimal approximation of a mixture of densities is obtained by a mixture of the approximations of the individual components. Here, another two options are available: The approximations of the individual components, e.g. Gaussians, could either be calculated on-line or calculated off-line and stored in some approximation library. The latter option requires some memory space but leads to extremely fast approximation procedures.

VI. SIMULATIONS

We now show some simulations using a stochastic nonlinear dynamic system that exhibits chaotic behavior, which was presented in [15]. An autonomous system

$$x_{k+1} = a(x_k) + w_k ,$$

with a time-invariant nonlinearity

$$a(x_k) = 0.230x_k(16 - x_k)$$

is assumed, where w_k is zero mean additive Gaussian noise with a standard deviation of $\sigma_w = 0.4$.

The simulation starts with $x_0 = 8$, which leads to a Gaussian density for $k = 1$ given by

$$\tilde{f}_1^x(x_1) = \mathcal{N}(x_1, 14.72, 0.4) .$$

We approximate the true prior density $f_1^x(x_1)$ and apply 5 prediction steps for calculating $f_2^x(x_2)$ to $f_6^x(x_6)$.

In order to have a ground truth to compare to, we calculate the true density functions $\tilde{f}_2^x(x_2)$ to $\tilde{f}_6^x(x_6)$, by means of brute force numerical integration. The resulting densities are shown with blue lines in Figure 3.

In the first simulation depicted in Figure 3 a), we show the performance of the proposed new approach presented in this paper for a fixed number of $L = 40$ components. We compare the Gaussian mixture densities resulting from each prediction step according to Section III-C to the ground truth. After each prediction step, the resulting Gaussian mixture is re-approximated with a constant number of components, here 40, in order to compute the next prediction step. The number of components is not optimized in this simulation run.

Figure 3 b) shows the proposed new approach with an optimal selection of the number of components based on the algorithm presented in Section V-B. The maximum distance measure was selected to $G_{\max} = 0.5\%$ for a normalized squared integral distance measure between the true posterior and its approximation. The minimum number of components for achieving a distance measure below this threshold and the actual error in each step are:

k	2	3	4	5	6
L	19	37	47	52	58
G	0.483%	0.444%	0.439%	0.425%	0.497%

In Figure 3 c) and d) we show the predictions based on Monte Carlo sampling with $n = 40$ and $n = 4000$ samples, respectively. In addition, Figure 3 e) shows the predictions made by the Unscented Kalman filter (UKF) [16].

A first observation here is that the Gaussian noise assumption of the UKF makes it impossible to keep track of the multimodality of the true density. Furthermore, we can see a low performance of the Monte Carlo approach for a small number of samples. Compared to the proposed new approach, two orders of magnitude more samples are required for achieving a similar approximation accuracy.

By comparing Figure 3 a) and b) we notice, that the number of components for an optimal approximation varies from prediction step to prediction step. This is due to the fact, that the complexity of the true density changes over time. The number of components does not necessarily increase over time as it can be observed here. There are also systems where the number of components decreases with time or stays constant.

VII. CONCLUSIONS AND FUTURE WORK

The key idea of this paper is to approximate the posterior density after every prediction step by means of a Dirac mixture in order to obtain a recursive prediction procedure.

Compared to particle filters that use random samples to represent the densities, the proposed approach has several distinct advantages. First, the Dirac components are systematically placed in order to minimize a given distance measure, which is selected in such a way that the future evolution of approximate densities is always close to the (unknown) true evolution. As a result, very few samples are sufficient to achieve an astonishing long-term prediction quality. Second, the optimization does not only include the parameters of the Dirac mixture approximation, i.e., weights and locations, but also the number of components. As a result, the number of components is automatically adjusted according to the complexity of the underlying true distribution. Third, as the approximation is fully deterministic, the proposed prediction procedure guarantees a dependable performance that can be specified beforehand.

The new approach is natural, mathematically rigorous, and based on efficient algorithms [11], [12] for the optimal approximation of arbitrary densities by Dirac mixtures with respect to a given distance measure. It is important to note that no assumptions about the nonlinearity of the dynamic system are made.

Implementation of the proposed approach is rather straightforward. In addition, the prototype implementation used for performing simulations turned out to be very stable, numerically uncritical, and produces repeatable results. Simulations with a chaotic system demonstrate the usefulness and practical relevance of the proposed new prediction approach in comparison to several standard filters.

The prediction mechanism is useful in its own respect and can be applied in a wide variety of scenarios: long-term prediction of the state of nonlinear dynamic systems corrupted by noise, prediction of time series, reachability analysis, and model-predictive control. However, a predictor is also an essential part of model-based estimation of the state of a stochastic nonlinear system from noisy measurements. Current research activities are focused on developing similar procedures for the filtering part of a complete estimator.

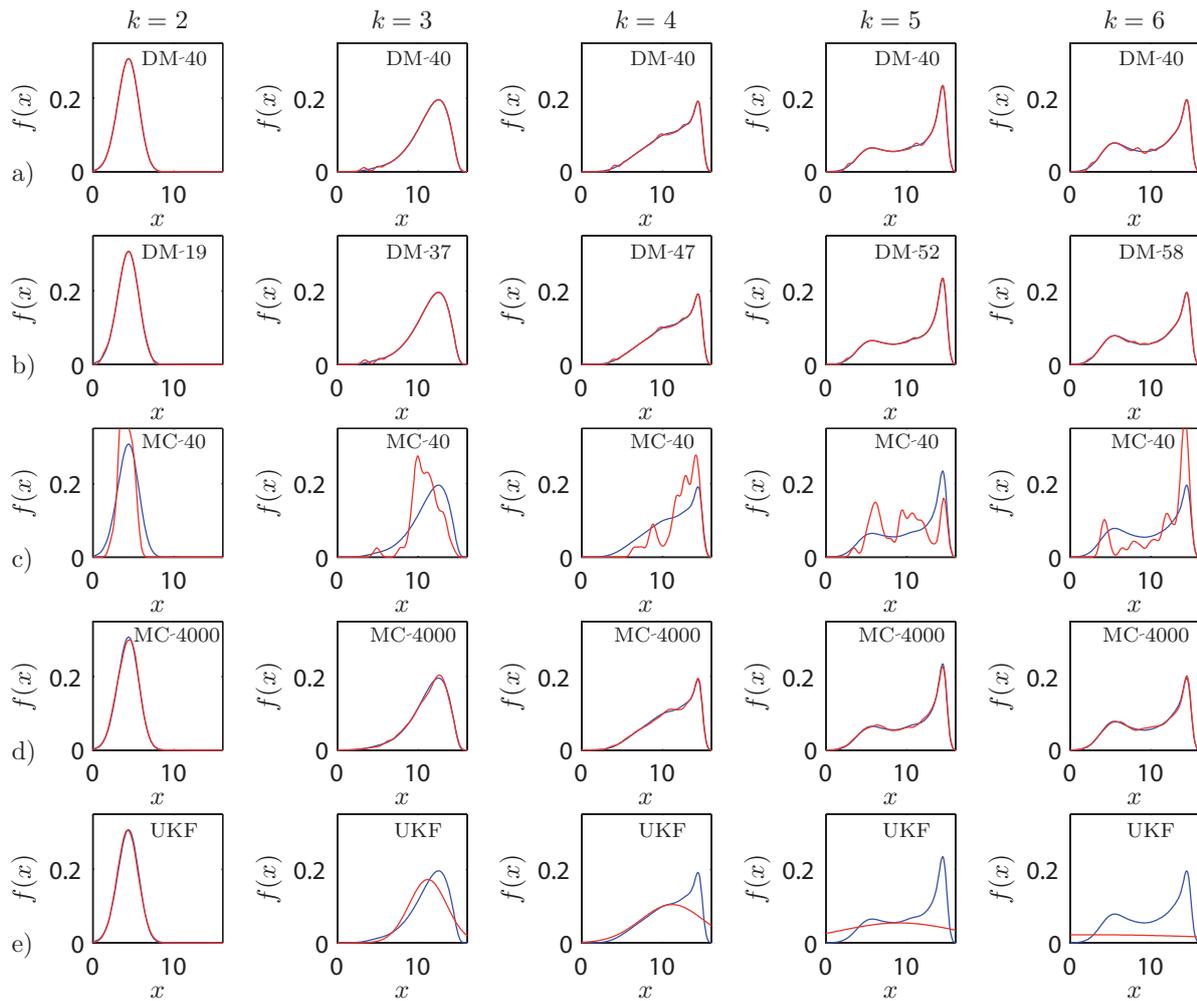


Fig. 3. Five prediction steps with Dirac mixture approximation a) 40 components and b) optimal number of components for a maximum posterior distance measure of 0.5%, Monte Carlo sampling with c) 40 and d) 4000 samples, and e) the Unscented Kalman filter (UKF), compared to the true densities (shown in blue) at each time step. The density for $k = 1$ is a Gaussian density for all filters given in (VI) and has been omitted.

REFERENCES

- [1] A. F. M. Smith and A. E. Gelfand, "Bayesian Statistics without Tears: A Sampling-Resampling Perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992.
- [2] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation Using Gaussian Sum Approximation," *IEEE Transactions on Automatic Control*, vol. AC-17, no. 4, pp. 439–448, 1972.
- [3] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [4] A. Doucet, N. D. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.
- [5] C. Musso, N. Oudjane, and F. LeGland, "Improving Regularized Particle Filters," Cambridge University Engineering Department, Tech. Rep. CUED/F-INFENG/TR 380, 2000.
- [6] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Transactions of Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [7] L. Kocis and W. J. Whiten, "Computational Investigations of Low-discrepancy Sequences," *ACM Transactions on Mathematical Software*, vol. 23, no. 2, pp. 266–294, June 1997.
- [8] W. J. Morokoff and R. E. Caflisch, "Quasi-Random Sequences and Their Discrepancies," *SIAM Journal on Scientific Computing*, vol. 15, no. 6, pp. 1251–1279, 1994.
- [9] G. Niederreiter, "Quasi-monte carlo methods and pseudo-random numbers," *Bulletin of the American Mathematical Society*, vol. 84, no. 6, pp. 957–1041, 1978.
- [10] V. Philomin, R. Duraiswami, and L. Davis, *Quasi-Random Sampling for Condensation*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2000, vol. 1843/2000, ch. Computer Vision - ECCV 2000, pp. 134–149.
- [11] 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'06), San Diego, California, USA, 2006*. [Online]. Available: http://isas.uka.de/downloads/06CDC_Schrempf.pdf
- [12] —, "Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance," in *Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006), Heidelberg, Germany, September 2006*, pp. 512–517. [Online]. Available: http://isas.uka.de/downloads/mfi_06schrempf.pdf
- [13] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 2, pp. 79–86, 1951.
- [14] 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.
- [15] Q. Yao and H. Tong, "On Prediction and Chaos in Stochastic Systems," *Philosophical Transactions: Physical Sciences and Engineering*, vol. 348, no. 1688, pp. 357–369, sep 1994.
- [16] S. Julier and J. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proceedings of SPIE AeroSense, 11th International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Orlando, FL, 1997*.