A STATE ESTIMATOR FOR NONLINEAR STOCHASTIC SYSTEMS BASED ON DIRAC MIXTURE APPROXIMATIONS

Oliver C. Schrempf and Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory, Universität Karlsruhe (TH), Germany

Keywords: Nonlinear Dynamic Systems, Stochastic Filter, Dirac Mixture.

Abstract: This paper presents a filter approach for estimating the state of nonlinear dynamic systems based on recursive approximation of posterior densities by means of Dirac mixture functions. The filter consists of a prediction step and a filter step. The approximation approach is based on a systematic minimization of a distance measure and is hence optimal and deterministic. In contrast to non-deterministic methods we are able to determine the optimal number of components in the Dirac mixture. A further benefit of the proposed approach is the consideration of measurements during the approximation process in order to avoid parameter degradation.

NOTATION

k	time index
x_k	state variable
y_k	measurement variable
$\hat{y}_k$	actual measurement at time k
$\hat{x}_k$	point estimate at time k
$\tilde{f}^x(x)$	probability density function of $\boldsymbol{x}$
$f^x(x)$	approximation of $\tilde{f}^x(x)$
$f^p(x_{k+1})$	predicted density function
$f^e(x_k)$	filtered density function
$f^L(x_k, \hat{y}_k)$	Likelihood functon
$\delta(x)$	Dirac Delta function
$H(x)$	Heaviside step function
G	distance measure
η	parameter vector
γ	progression parameter
$\mathcal{N}(., m, \sigma)$	Gaussian density with mean m and standard deviation σ

1 INTRODUCTION

In this paper, we present a novel stochastic filter for nonlinear dynamic systems suffering from system as well as measurement noise. The uncertainty in the filter's estimate caused by the noise is described by means of probability density functions. The problem that arises with the application of stochastic filters to nonlinear systems is that the complexity of the density representation increases and the exact densities cannot be calculated directly in general. Common solutions to this problem in order to build practical estimators can be devided into two classes. The approaches of the first class approximate or modify the system and measurement functions and apply a filter. The idea of the second class is to approximate the resulting density functions themselves in order to calculate the filter steps in closed-form .

A common representative of the first class is the extended Kalman filter (EKF). It is based on linearization of the system and measurement functions and applying a standard Kalman filter to this modified system. This approach is applicable to systems with only negligible nonlinearities and additive noise, but fails in more general cases.

Another approach is to approximate the system together with its noise as a probabilistic model (Huber and Hanebeck, 2007). The application of adequate representations of the model like Gaussian mixtures with axis-aligned components (Huber et al., 2006), allows for efficient implementation of the filter steps.

Filters approximating the density functions instead of the system function can be divided into two

main approaches found in the literature: i) sample-based representations and ii) analytic density representations.

Sample-based filters like the popular particle filter (Doucet et al., 2000)(Doucet et al., 2001) apply Monte Carlo methods for obtaining a sample representation. Since these sample are usually produced by a random number generator, the resulting estimate is not deterministic. Furthermore, Markov Chain Monte Carlo Methods (MCMC) are iterative algorithms that are unsuited for recursive estimation, hence, importance sampling like in (Geweke, 1989) is often applied. The problem of sample degradation is usually tackled by bootstrap methods (Gordon, 1993).

Other methods describe the probability density functions by means of their moments. A popular approach is the so called Unscented Kalman filter (UKF) (Julier and Uhlmann, 1997) that uses the first moment and the second central moment for representing the densities. This allows for an efficient calculation of the update but fails in representing highly complex densities arising from nonlinear systems. Furthermore, the assumption of jointly Gaussian distributed states and measurements is made, which is not valid in general.

An approach that represents the densities by means of Gaussian mixture densities is the so called Gaussian sum filter (Alspach and Sorenson, 1972). The Gaussian mixture representation allows for approximating arbitrary density functions, but finding the appropriate parameters is a tough problem. A more recent approach is the Progressive Bayes filter (Hanebeck et al., 2003) which uses a distance measure for approximating the true densities. The key idea in this approach is to transform the approximation problem into an optimization problem. This is a major motivation for the approximation applied in the approach presented here.

The filter method we propose here follows the idea of approximating the density functions instead of the system itself, but the approximation is performed in a systematic manner. The general idea is to approximate the continuous density function by means of a Dirac mixture function that minimizes a certain distance measure to the true density. The approximation process itself is described in (Schrempf et al., 2006a)(Schrempf et al., 2006b) and will therefore only be discussed briefly in this paper. We will focus here on the complete filter consisting of approximation, prediction (Schrempf and Hanebeck, 2007) and filter step.

Since we make use of a distance measure, we are able to quantify the quality of our approximation.

Furthermore, it is possible to find an optimal number of components required for sufficient estimates. Following this idea we will extend our optimization method to a full estimation cycle by considering the measurement as well.

The paper is organized as follows: We will give a problem formulation in Section 2 followed by an overview of the complete filter in Section 3. The building blocks of the filter are described in Section 4 whereas Section 5 presents further optimization methods. Experimental results comparing the proposed filter to state-of-the-art filters are given in Section 6 followed by conclusions in Section 7.

2 PROBLEM FORMULATION

We consider discrete-time nonlinear dynamic systems according to

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

producing measurements according to the nonlinear function

$$y_k = h_k(x_k, v_k) \quad .$$

The state of the system that is not directly observable in general is represented by x_k. u_k is a known input, and y_k is an observable output of the system. $a_k(\cdot)$ is a time-varying nonlinear mapping describing the system's dynamic behavior. w_k represents both endogenous and exogenous noise sources acting upon the system and is described by means of a density function $f_k^w(w_k)$. $h_k(\cdot)$ maps the system state to an output value which suffers from noise v_k which is modeled by means of a density function $f_k^v(v_k)$.

Starting with an initial state x_0, our goal is to keep track of the system's state over time while maintaining a full continuous stochastic representation of the uncertainty involved caused by the system and measurement noise.

This corresponds to sequentially calculating the state densities $f_k^x(x_k), k = 1, \ldots, N$, by means of a prediction and a filter step where the system and measurement functions are applied.

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 an 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.

For reasons of brevity, we omit the input u_k. We further focus on additive noise, which results in the system equation

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

and a measurement equation

$$y_k = h_k(x_k) + v_k \quad.$$

In addition, the time index k is omitted in some cases without notice.

3 FILTER OUTLINE

In this section, we will give a brief overview of the recursive filtering scheme depicted as a block diagram in Figure 1. The left part of the figure shows the nonlinear system suffering from additive noise as described in Sec. 2. The right part shows the estimator. The input of the estimator is a measurement $\hat{y}_k$ coming from the system. The output of the estimator is a probability density function $f^e(x_k)$ from which a point estimate $\hat{x}_k$ can be derived. The estimator itself works recursively as can be seen from the loop in the diagram. Each recursion consists of a prediction step, an approximation step, and a filter step.

The prediction step receives a density $f^e(x_k)$ from the previous filter step. This density is an approximation represented by means of a Dirac mixture allowing for an analytically exact solution of the Bayesian prediction integral with respect to this approximation. The prediction yields a continuous mixture density representation (e.g. a Gaussian mixture) $\tilde{f}^p(x_{k+1})$. Details are given in Sec. 4.2.

The continuous mixture density $\tilde{f}^p(x_{k+1})$ resulting from the prediction step serves as input to the approximation step. The density is systematically approximated by means of a Dirac mixture $f^p(x_{k+1})$ minimizing a distance measure $G\left(\tilde{f}^p(x_{k+1}), f^p(x_{k+1})\right)$ as described in Sec. 4.1.

The approximated density $f^p(x_{k+1})$ is then fed to the filter step, where it gets fused with the likelihood function $f^L(x, \hat{y})$. This step is described in detail in Sec. 4.3.

4 FILTER COMPONENTS

4.1 Density Approximation

We will now introduce Dirac mixture functions and explain how they can be interpreted as parametric density functions. Subsequently, we briefly describe the systematic approximation scheme.

4.1.1 Dirac Mixture Density Representation

Dirac mixtures are a sum of weighted Dirac delta functions according to

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

where

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

is a parameter vector consisting of locations $x_i, i = 1, \ldots, L$ and weighting coefficients $w_i, i = 1, \ldots, L$. The Dirac delta function is an impulse representation with the properties

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ \text{not defined}, & x = 0 \end{cases}$$

and

$$\int_{\mathbf{R}} \delta(x)\, dx = 1 \quad.$$

This results in the fundamental property

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_i)\, dx = f(x_i) \quad.$$

A mixture of Dirac delta functions as given in (1) can be used for representing arbitrary density functions if the following requirements are considered. Since the properties of a density function $f(x)$ demand that $f(x) \geq 0$ and $\int_{\mathbf{R}} f(x)\, dx = 1$, we have

$$w_i \geq 0, i = 1, \ldots, L$$

and

$$\sum_{i=1}^{L} w_i = 1 \quad.$$

Hence, we require $2L$ parameters with $2L - 1$ degrees of freedom.

A simplified density representation is given by equally weighted Dirac mixtures, as

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

where only L parameters and L degrees of freedom are used. This results in a simpler, less memory consuming representation with less approximation capabilities.

Dirac mixtures are a generic density representation useful for approximating complicated densities arising in estimators for nonlinear dynamic systems.
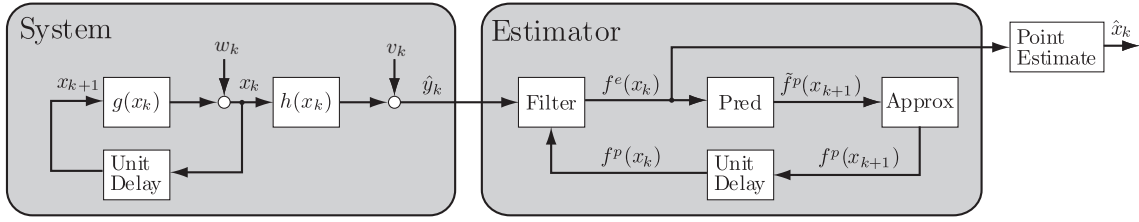
Figure 1: A block diagram of the recursive estimator. The estimator consists of a filter step, a prediction step and an approximation step.

4.1.2 Approximation Approach

A systematic approximation of continuous density by means of another density requires a distance measure between the two densities

$$G\left(\tilde{f}^p(x_{k+1}), f^p(x_{k+1}, \underline{\eta})\right) \ ,$$

where $\tilde{f}^p(\cdot)$ is an arbitrary continuous density function and $f^p(\cdot, \underline{\eta})$ is a Dirac mixture density. The approximation problem can then be reformulated as an optimization problem by finding a parameter vector $\underline{\eta}$ that minimizes this distance measure.

Popular distance measures for comparing continuous densities, measures are the Kullback–Leibler divergence (Kullback and Leibler, 1951) or the integral quadratic measure. For comparing a continuous density to a Dirac mixture, however, they are not very useful, since the Dirac mixture is zero between the Dirac pulses. Hence, instead of comparing the densities directly, the corresponding (cumulative) distribution functions are employed for that purpose. For the rest of this subsection we will omit the time index k and the p index in order to keep the formulae comprehensible.

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

$$F(x, \underline{\eta}) = \int_{-\infty}^{x} f(t, \underline{\eta})\,dt = \sum_{i=1}^{L} w_i H(x - x_i) \ , \quad (2)$$

where $H(.)$ 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 (Boos, 1981)

$$G(\underline{\eta}) = \int_{-\infty}^{\infty} r(x)\left(\tilde{F}(x) - F(x, \underline{\eta})\right)^2 dx \ , \quad (3)$$

where $r(x)$ is a nonnegative weighting function. $r(x)$ will later in the filter step be selected in such a way that only those portions of the predicted probability density function having support by the likelihood, are approximated with 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 (3) according to $\underline{\eta} = \arg\min_{\underline{\eta}} G(\underline{\eta})$. Unfortunately, it is not possible to solve this optimization problem directly. Hence, we apply a progressive method introduced in (Schrempf et al., 2006b). For this approach, we introduce a so called progression parameter γ into $\tilde{F}(x)$ that goes from $0\ldots1$. The purpose of this parameter is to find a very simple and exact approximation of $\tilde{F}(x, \gamma)$ for $\gamma = 0$. Further 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

$$\frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}} = \begin{bmatrix} \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{x}} \\ \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{w}} \end{bmatrix} \overset{!}{=} \underline{0} \ . \quad (4)$$

Together with (2) and (3) this results in the system of equations

$$\tilde{F}(x_i, \gamma) = \sum_{j=1}^{L} w_j H(x_i - x_j) \ ,$$

$$\int_{x_i}^{\infty} r(x)\tilde{F}(x, \gamma)\,dx = \sum_{j=1}^{L} w_j \int_{x_i}^{\infty} r(x)H(x - x_j)\,dx \ ,$$

where $i = 1, \ldots, L$.

In order to track the minimum of the distance measure we have to take the derivative of (4) with respect to γ.

This results in a system of ordinary first order differential equations that can be written in a vector–matrix–form as

$$\underline{b} = \mathbf{P}\dot{\underline{\eta}} \ , \quad (5)$$

where

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

and

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

$\dot{\eta}$ denotes the derivative of η with respect to γ.

The $\mathbf{P}$ matrix as well as the derivations are given in (Schrempf et al., 2006b). The approximation of $\tilde{f}(x)$ now boils down to solving (5).

4.2 Prediction Step

We now explain the Bayesian prediction step and show how the approximation introduced in the last subsection can be used for closed-form calculations.

Calculating the state densities $f^p(x_{k+1})$, $k = 1, \ldots, N$, is performed by evaluation the Bayesian forward step, which is given by

$$f^p(x_{k+1}) = \int_{-\infty}^{\infty} f(x_{k+1}|x_k) f^e(x_k) \, dx_k \ , \quad (6)$$

where the transition density $f(x_{k+1}|x_k)$ of the considered nonlinear system with additive noise is given by

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

where $f^w(\cdot)$ is the density of the system noise (e.g. Gaussian).

In general, the integral involved in (6) cannot be solved analytically for arbitrary prior densities $f^e(x_k)$ For a given input point $\bar{x}_k$, however, represented by the Dirac delta function $f^e(x_k) = \delta(x_k - \bar{x}_k)$, (6) can be solved in closed form according to

$$f_p(x_{k+1}) = f^w(x_{k+1} - g(\bar{x}_k)) \ .$$

In the case of zero mean Gaussian system noise with

$$f^w(w) = \mathcal{N}(w, 0, \sigma^w) \ ,$$

this yields

$$f_p(x_{k+1}) = \mathcal{N}(x_{k+1}, g(\bar{x}_k), \sigma^w) \ ,$$

which is a Gaussian Density with a standard deviation σ^w.

For a given Dirac mixture prior $f^e(x_k)$ according to (1) given by

$$f^e(x_k) = \sum_{i=1}^{L} w_k^{(i)} \delta(x_k - x_k^{(i)}) \ , \quad (7)$$

the posterior according to (6) is a Gaussian mixture given by

$$f^p(x_{k+1}) = \sum_{i=1}^{L} w_k^{(i)} \mathcal{N}\left(x_{k+1}, g(x_k^{(i)}), \sigma^w\right) \ ,$$

which is a closed-form solution.

Please note, that similar result can be derived for non-additive and non-Gaussian noise.

4.3 Filter Step

The filter step consists of fusing the predicted density $f^p(x_k)$ and the likelihood function $f^L(x_k, \hat{y}_k)$ governed by the measurement $\hat{y}_k$ according to

$$f^e(x_k) = c \cdot f^p(x_k) \cdot f^L(x_k, \hat{y}_k) \ , \quad (8)$$

where c is a normalizing constant. The likelihood function is given by

$$f^L(x_k, \hat{y}_k) = f(\hat{y}_k|x_k) \ .$$

For a nonlinear system with additive noise, the conditional density for the measurement $f(y_k|x_k)$ is given by

$$f(y_k|x_k) = f^v(y_k - h(x_k)) \ ,$$

where $f^v(\cdot)$ is the density of the measurement noise and $h(x_k)$ is the nonlinear measurement function. In the case of zero-mean Gaussian measurement noise the likelihood function can be written as

$$f^L(x_k, \hat{y}_k) = \mathcal{N}(\hat{y}_k, h(x_k), \sigma^v) \ .$$

We would like to emphasize, that in the general nonlinear case this likelihood function is no proper density function. Furthermore, a parametric representation of this function is not available in general. This is the reason, why the update equation (8) cannot be solved analytically, even if the prediction is given in a parametric representation.

Our solution to this problem is driven by the same observation made for solving the prediction step in Sec. 4.2. The likelihood can be evaluated at certain points $\bar{x}_k$, which yields constant values.

In order to calculate the product of a likelihood and a prediction, where the latter is already given as a Dirac mixture, it comes quite naturally to use the $x_k^{(i)}$ points of the Diracs to evaluate the likelihood. The obtained values of $f^L(\cdot)$ can then be used to reweight the predicted density according to

$$f^e(x_k) = \sum_{i=1}^{L} \bar{w}_k^{(i)} \delta(x_k - x_k^{(i)})$$

with

$$\bar{w}_k^{(i)} = c \cdot w_k^{(i)} \cdot f^v(\hat{y}_k - h(x_k^{(i)})) \ ,$$

where $w_k^{(i)}$ is the i'th weight and $x_k^{(i)}$ is the i'th position of the approximated prediction $f^p(x_k)$. The normalization constant can be calculated as

$$c = \left(\sum_{i=1}^{L} w_k^{(i)} \cdot f^v(\hat{y}_k - h(x_k^{(i)})) \right)^{-1} \ .$$

Naive approximation of the predicted density in a fixed interval may lead to many small weights, since not all regions of the state space supported by the prediction are as well supported by the likelihood. This phenomenon can be described as parameter degradation. To circumvent this problem, we make use of the weighting function $r(x)$ in (3). Details on this approach are presented in the next section.

5 OPTIMAL NUMBER OF PARAMETERS

In this section, we describe how to tackle the problem of parameter degradation that is inherent to all filter approaches considering only discrete points of the density. We further describe a method for finding an optimal number of components for the approximation taking into account the prediction and filter steps as well.

To fight the problem of parameter degradation described in the previous section we make use of the fact, that although the likelihood function is not a density it decreases to zero for $x \to \pm\infty$ in many cases. An example for this are polynomial systems suffering from additive noise. Therefore, we can define an area of support for which the likelihood is higher than a certain value. This area of support is an interval and can be represented by the weighting function $r(x)$ in (3). It guarantees, that all components of the approximation are located in this interval and are therefor not reweighed to zero the filter step. In other words, the complete mass of the approximation function accounts for the main area of interest.

In (Schrempf and Hanebeck, 2007), we introduced an algorithm for finding the optimal number of components required for the approximation with respect to the following prediction step. We will now extend this algorithm in order to account for the preceding filter step as well.

At the beginning of Algorithm 1 in line 6, 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

Algorithm 1 Optimal number of components w.r.t. the filter step and posterior density.

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

of components we can assume this density to be very close to the true density. An efficient procedure for approximating arbitrary mixture densities with Dirac mixtures comprising a large number of components is given in (Schrempf and Hanebeck, 2007).

In each search step of the algorithm, the distance measure of the approximated density at hand to the density defined by $\underline{\kappa}_t$ is calculated. In this way the smallest number of components for a prespecified error can be found.

6 EXPERIMENTAL RESULTS

In order to compare the performance of our filter to other state-of-the-art filters, we have simulated a nonlinear dynamic system according to the left part of Figure 1. We apply the filter to a strongly nonlinear cubic system and measurement function motivated by the cubic sensor problem introduced in (Bucy, 1969).

The simulated system function is

$$g(x_k) = 2x_k - 0.5x_k^3 + w$$

and the additive noise is Gaussian with $\sigma^w = 0.2$ standard deviation. The measurement function is

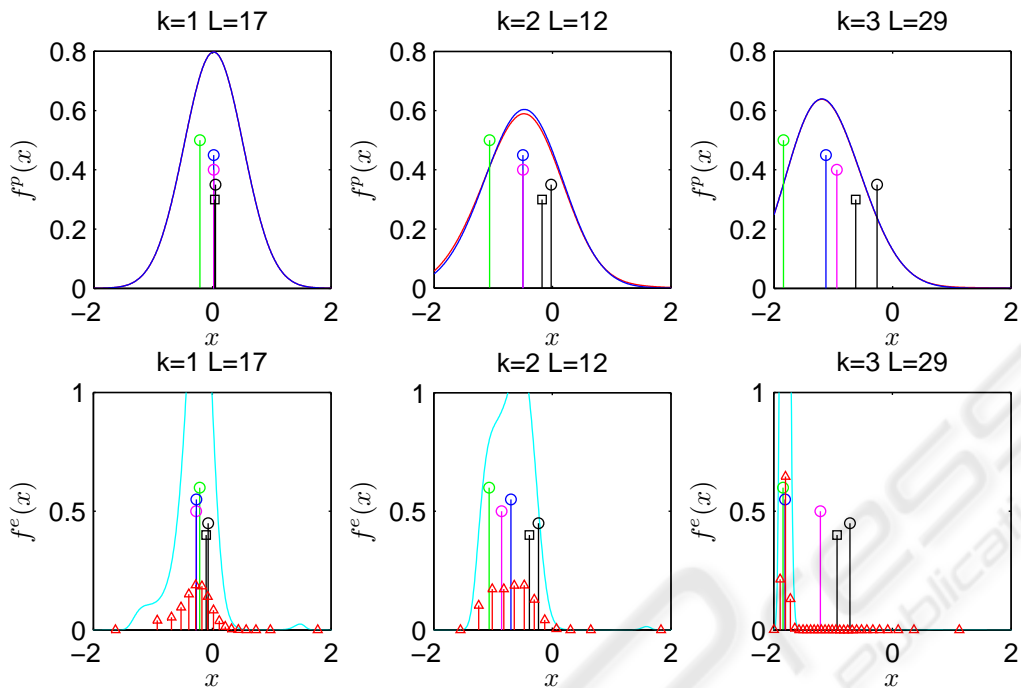$$h(x_k) = x_k - 0.5x_k^3 + v$$

Figure 2: The recursive filter for $T = 3$ steps. k indicates the step number and L the number of components for the Dirac mixture. The upper row shows the prediction steps, the lower row shows the filter steps. **Upper row:** The blue is the continuous density predicted by the DM filter, the red line underneath is the true density. The green marker depicts the true (simulated) system state, the other markers depict the predicted point estimates of the following filters: blue=DM, pink=UKF, black_circle=PF20, black_square=PF1000. **Lower row:** The cyan line shows the likelihood. The colors of the point estimates are similar to the upper line.

with additive Gaussian noise and $\sigma^v = 0.5$.

The generated measurements are used as input to our filter as well as to an unscented Kalman filter and a particle filter. The particle filter is applied in a variant with 20 particles and a variant with 1000 particles in order to compare the performance.

In a first run of the experiment we show $T = 3$ steps in Figure 2. The upper row shows the prediction steps, the lower row shows the corresponding filter steps. The continuous prediction $f^p(x_{k+1})$ of the Dirac mixture (DM) filter is depicted by the dark blue line. The red line underneath shows the true prediction computed numerically as a reference. The cyan plot in the lower line shows the likelihood given by the current measurement and the red arrows depict the Dirac mixture after the filter step.

Both rows also show the point estimates of the various applied filters in the current step. The green marker indicates the true (simulated) state, whereas blue stands for the Dirac mixture point estimate. Pink is the UKF estimate an black are the particle filter estimates. The particle filter indicated by the circle uses 20 particles, the one indicated by the square uses 1000 particles.

We simulated the system a further 10 times for $T = 7s$ in order to calculate the root means square

error e_{rms} of the 4 filters. The results are shown in Figure 3. The plot shows that the point estimates of
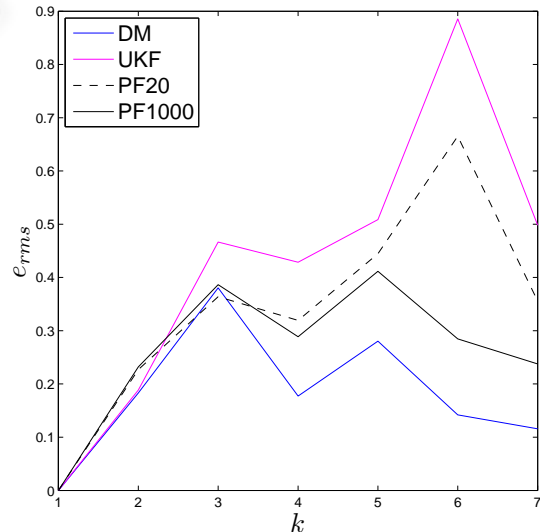

Figure 3: Root mean square error for 10 runs and $T = 7$ steps.

the Dirac mixture filter are much closer to the true state than the point estimates of the other filters.

7 CONCLUSION

In this paper, we presented a complete Dirac mixture filter that is based on the approximation of the posterior density. The filter makes use of the properties of the Dirac mixture approximation wherever they are required, but does not deny the continuous character of the true density. This can especially be seen after each prediction step, where the full continuous density representation is used.

The new approach is natural, mathematically rigorous, and based on an efficient algorithms (Schrempf et al., 2006a)(Schrempf et al., 2006b) for the optimal approximation of arbitrary densities by Dirac mixtures with respect to a given distance.

Compared to a particle filter, the proposed method has several 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 true density while also considering the actual measurements. As a result, very few samples are sufficient for achieving an excellent estimation 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 and the support area of a given likelihood. Third, as the approximation is fully deterministic, it guarantees repeatable results.

Compared to the Unscented Kalman Filter, the Dirac mixture filter has the advantage, that it is not restricted to first and second order moments. Hence, multi-modal densities, which cannot be described sufficiently by using only the first two moments, can be treated very efficiently. Such densties occur quite often in strongly nonlinear systems. Furthermore, no assumptions on the joint distribution of state and measurement have to be made.

REFERENCES

Alspach, D. L. and Sorenson, H. W. (1972). Nonlinear Bayesian Estimation Using Gaussian Sum Approximation. *IEEE Transactions on Automatic Control*, AC–17(4):439–448.

Boos, D. D. (1981). Minimum Distance Estimators for Location and Goodness of Fit. *Journal of the American Statistical association*, 76(375):663–670.

Bucy, R. S. (1969). Bayes Theorem and Digital Realizations for Non-Linear Filters. *Journal of Astronautical Sciences*, 17:80–94.

Doucet, A., Freitas, N. D., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistics and Computing*, 10(3):197–208.

Geweke, J. (1989). Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, 24:1317–1399.

Gordon, N. (1993). *Bayesian Methods for Tracking*. PhD thesis, University of London.

Hanebeck, U. D., Briechle, K., and Rauh, A. (2003). Progressive Bayes: A New Framework for Nonlinear State Estimation. In *Proceedings of SPIE*, volume 5099, pages 256–267, Orlando, Florida. AeroSense Symposium.

Huber, M., Brunn, D., and Hanebeck, U. D. (2006). Closed-Form Prediction of Nonlinear Dynamic Systems by Means of Gaussian Mixture Approximation of the Transition Density. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006), Heidelberg, Deutschland*, pages 98–103.

Huber, M. F. and Hanebeck, U. D. (2007). Hybrid Transition Density Approximation for Efficient Recursive Prediction of Nonlinear Dynamic Systems. In *International Conference on Information Processing in Sensor Networks (IPSN 2007), Cambridge, USA*.

Julier, S. and Uhlmann, J. (1997). 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*.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(2):79–86.

Schrempf, O. C., Brunn, D., and Hanebeck, U. D. (2006a). 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*.

Schrempf, O. C., Brunn, D., and Hanebeck, U. D. (2006b). 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*, pages 512–517.

Schrempf, O. C. and Hanebeck, U. D. (2007). Recursive Prediction of Stochastic Nonlinear Systems Based on Dirac Mixture Approximations. In *Proceedings of the American Control Conference (ACC '07), New York City, USA*.