Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér–von Mises Distance

Oliver C. Schrempf, Dietrich Brunn, and Uwe D. Hanebeck

Abstract— **This paper proposes a systematic procedure for approximating arbitrary probability density functions by means of Dirac mixtures. For that purpose, a distance measure is required, which is in general not well defined for Dirac mixture densities. Hence, a distance measure comparing the corresponding cumulative distribution functions is employed. Here, we focus on the weighted Cramér–von Mises distance, a weighted integral quadratic distance measure, which is simple and intuitive. Since a closed–form solution of the given optimization problem is not possible in general, an efficient solution procedure based on a homotopy continuation approach is proposed. Compared to a standard particle approximation, the proposed procedure ensures an optimal approximation with respect to a given distance measure. Although useful in their own respect, the results also provide the basis for a recursive nonlinear filtering mechanism as an alternative to the popular particle filters.**

NOTATION

$\delta(x)$	Dirac Delta function
$H(x)$	Heaviside step function
G	distance measure
$\underline{\eta}$	parameter vector
γ	progression parameter
$N(.,m,\sigma)$	Gaussian density with mean m and standard deviation σ

I. INTRODUCTION

Processing density functions in nonlinear estimation procedures typically cannot be performed exactly. Especially in recursive processing, the type of density changes and the complexity increases. Hence, nonlinear estimation in general requires the approximation of the underlying true densities by means of generic density types.

Different types of generic approximation densities have been proposed in literature including Gaussian mixtures [1], Edgeworth series expansions [2], and exponential densities [3]. Another very popular approach is to represent the true density by means of a set of samples [4], which is used by the class of particle filters [5]. Typically, the appropriate locations and weights of the particles are calculated by means of Monte Carlo techniques [6], [7].

Here, we pursue a different interpretation. The particles are viewed as a mixture of weighted Dirac delta components used to systematically approximate the density at hand. This is different from the deterministic type of particle filters in [8] as a distance measure is employed to transform the approximation problem into an optimization problem. However, in the case of Dirac mixtures, typical distance measures quantifying the distance between two densities are not well defined. Examples are the Kullback–Leibler distance [9], its symmetrized version [10] or integral quadratic distances between the densities. Hence, the first key contribution of this paper is to compare the corresponding cumulative distribution functions of the true density and its approximation in order to find optimal parameters for the Dirac Mixture approximation. This can be viewed as a reversal of the procedure introduced in [11], where a distribution distance, in that case the Kolmogorv–Smirnov test statistic, is used to calculate optimal parameters of a density given observed samples.

In this paper, we focus on the weighted Cramér–von Mises distance [12] between the cumulative distributions. Other possible distribution distances would be the Kolmogorov–Smirnov distance [13], or the generalized Cramér–von Mises distance [14]. Since a closed–form solution of the given optimization problem is not possible in general, the second key contribution is an efficient solution procedure for arbitrary true densities based on a homotopy continuation approach similar to the approach introduced in [15].

The results of this paper can immediately be used for implementing a recursive nonlinear filter that could serve as an alternative to the popular particle filters. In contrast to a standard particle approximation, the proposed approach provides an optimal approximation with respect to a given distance measure. Furthermore, the proposed approximation procedure might be easier to understand than a real implementation of a particle approximation. The special case of equally weighted components is presented in [16].

The paper is organized as follows. After a formulation of the approximation problem in Section II, the conversion of the approximation problem into an equivalent optimization problem is described in Section III. A general solution approach for arbitrary densities is then given in Section IV. The algorithm of the solver applied is presented in Section V followed by some practical examples in Section VI. Conclusions and a few remarks about possible extensions and future work are given in Section VII.

It is important to note that this paper is restricted to the case of scalar random variables.

Oliver C. Schrempf, Dietrich Brunn, 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, brunn@ira.uka.de

II. Problem Formulation

We consider the problem of approximating a given density $\tilde{f}(x)$ by means of a Dirac Mixture given by

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

The parameters of this approximation are the centers of the individual Dirac functions x_i and the weighting coefficients w_i. These parameters are collected in a parameter vector $\underline{\eta}$ according to

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

For the remainder of this paper, we assume that the centers are ordered according to

$$x_1 < x_2 < \ldots < x_{L-1} < x_L \ .$$

Our goal is to minimize a certain distance measure G between the given density $\tilde{f}(x)$ and its approximation $f(x,\underline{\eta})$. For the purpose of nonlinear estimation, we are especially interested in approximations with an inhomogeneous approximation quality.

III. Approximation by Optimization

The first key idea is to reformulate the above approximation problem as an optimization problem by minimizing a certain distance between the true density $\tilde{f}(x)$ and its approximation $f(x,\underline{\eta})$. Instead of comparing the densities directly, which does not make sense for Dirac Delta functions, 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

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

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 [12]

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

where $g(x)$ is a nonnegative weighting function. This weighing function $g(x)$ has been introduced for achieving an approximation with an inhomogeneous approximation quality. In the specific application, $g(x)$ is selected in such a way that only those portions of the considered propability 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 next section is concerned with calculating an optimal parameter vector $\underline{\eta}$ minimizing the given distance.

IV. Solving the Optimization Problem by Homotopy Continuation

It is not possible to solve the optimization problem formulated in the previous section directly. Hence, we introduce an approach to find the solution progressively by applying the homotopy continuation method. In order to apply homotopy continuation, 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 or 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.

Example IV.1 An example of this continuation can be seen in Figure 1. Here we have a progression schedule for a standard normal density function. For $\gamma = 0$ we have a uniform density that can be approximated very easily. As the progression goes from $\gamma = 0\ldots1$ the function becomes the desired normal density. This behavior is achieved by parameterizing the standard deviation according to

$$\tilde{f}(x,\gamma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\frac{x^2}{\left(\frac{1+\epsilon}{\gamma+\epsilon}\sigma\right)^2}\right\} \ ,$$

where ϵ is a very small constant.

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

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

Applying this derivative we obtain

$$\frac{\partial G(\underline{\eta},\gamma)}{\partial x_i} = -w_i g(x_i)\left[\tilde{F}(x_i,\gamma) - F(x_i)\right]$$

$$\frac{\partial G(\underline{\eta},\gamma)}{\partial w_i} = \int_{-\infty}^{\infty} g(x)\left[\tilde{F}(x,\gamma) - F(x_i)\right] H(x-x_i)\,dx \ .$$

By setting $\frac{\partial G(\underline{\eta},\gamma)}{\partial \underline{\eta}}$ to zero and replacing $F(x,\underline{\eta})$ by $\sum_{j=1}^{L} w_j H(x-x_j)$ we obtain the system of equations

$$\begin{aligned} \tilde{F}(x_i,\gamma) &= \sum_{j=1}^{L} w_j H(x_i - x_j) \\ \int_{x_i}^{\infty} g(x)\tilde{F}(x,\gamma)\,dx &= \sum_{j=1}^{L} w_j \int_{x_i}^{\infty} g(x) H(x-x_j)\,dx \ . \end{aligned}$$

for $i = 1\ldots L$. By selecting the weighting function $g(x)$ as a piecewise constant function according to

$$g(x) = \begin{cases} 1 & -c < x < c \\ 0 & \text{else} \end{cases}$$

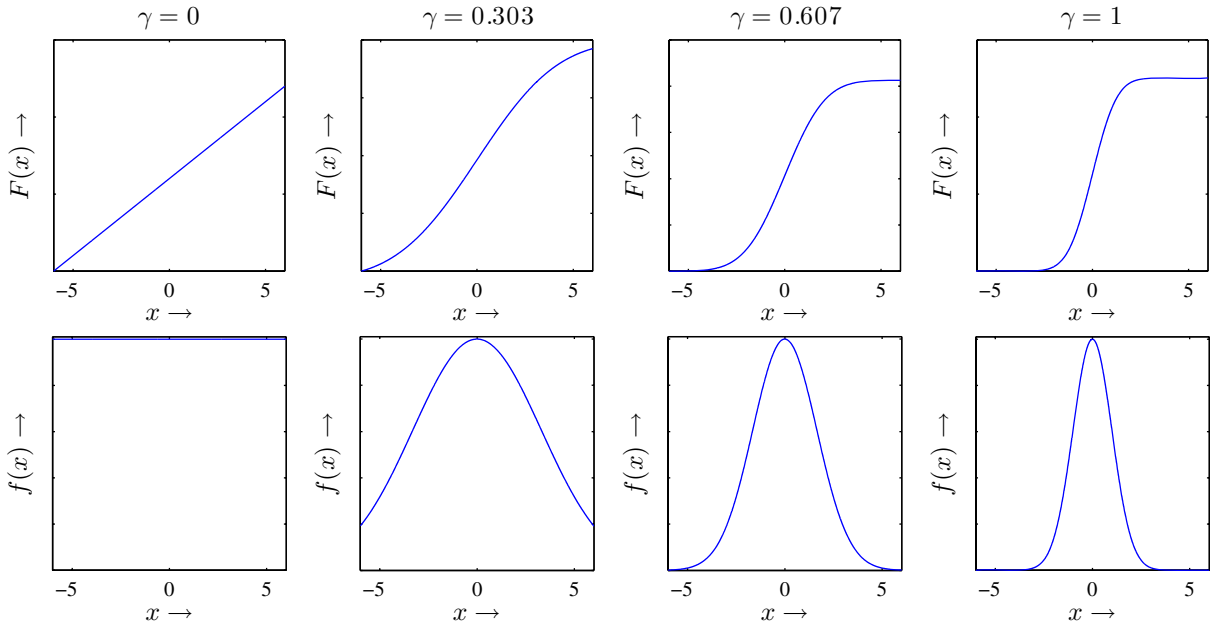

Fig. 1. Progression schedule for Gaussian a) distribution function b) density function. $\gamma = 0$ corresponds to the uniform density or distribution. $\gamma = 1$ corresponds to the original Gaussian density or distribution.

we can simplify the system of equations to

$$\tilde{F}(x_i, \gamma) = \sum_{j=1}^{i-1} w_j + \frac{w_i}{2} \tag{4}$$

and

$$\int_{x_i}^{c} \tilde{F}(x, \gamma)\ dx = c \sum_{j=1}^{L} w_j - x_i \sum_{j=1}^{i} w_j - \sum_{j=i+1}^{L} w_j x_j \ . \tag{5}$$

The weighting function $g(x)$ considers the interval $[-c, c]$ but equation (5) considers only the interval $[x_1, c]$. Since generally $x_1 > -c$, we have a small error. To compensate for this error in the interval $[-c, x_1]$, we introduce a further component $w_0 H(x - x_0)$ in the approximation for which we let $x_0 = -c$ constant. The weight w_0 automatically compensates for the error.

To track the minimum of the distance measure we have to take the derivative of (4) and (5) with respect to γ. Since $\tilde{F}(x_i, \gamma)$ is both an explicit and due to $x_i = x_i(\gamma)$ an implicit function of γ, we obtain

$$\frac{\partial \tilde{F}(x_i, \gamma)}{\partial \gamma} + \tilde{f}(x_i, \gamma)\dot{x}_i = \sum_{j=0}^{i-1} \dot{w}_j + \frac{\dot{w}_i}{2} \tag{6}$$

for $i = 1, \ldots, L$, and

$$\int_{x_i}^{c} \frac{\partial \tilde{F}(x, \gamma)}{\partial \gamma}\ dx = c \sum_{j=0}^{L} \dot{w}_j - x_i \sum_{j=0}^{i} \dot{w}_j - \sum_{j=i+1}^{L} x_j \dot{w}_j + \tilde{F}(x_i, \gamma)\dot{x}_i - \dot{x}_i \sum_{j=0}^{i} w_j - \sum_{j=i+1}^{L} w_j \dot{x}_j \tag{7}$$

for $i = 0, \ldots, L$ and $\dot{x}_0 = 0$.

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

$$\underline{b} = \mathbf{P}\dot{\underline{\eta}} \ , \tag{8}$$

where

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

and

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

The $\mathbf{P}$ matrix is given by (9).

Since we do not modify the x_0 parameter by defining $\dot{x}_0 = 0$, we can omit the rows and columns for $\dot{x}_0$.

V. Solver

Since (8) cannot be solved analytically we now give an algorithm that solves this ODE numerically. It is very efficient, due to adaptive stepsize control while keeping the error low. The error control is achieved by applying a predictor-corrector scheme. A pseudo code representation is given in Algorithm 1.

The algorithm starts with $\gamma = 0$. During the solution process we gradually increase γ while adjusting the parameter vector $\underline{\eta}$. Since we want to find the minimum of $G(\underline{\eta}, \gamma)$ for

$$\mathbf{P} = \begin{bmatrix} 1 & \frac{1}{2} & 0 & \dots & 0 & -\tilde{f}(x_1,\gamma) & 0 & \dots & 0 \\ 1 & 1 & \frac{1}{2} & \dots & 0 & 0 & -\tilde{f}(x_2,\gamma) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & \frac{1}{2} & 0 & 0 & \dots & -\tilde{f}(x_L,\gamma) \\ c-x_0 & c-x_1 & c-x_2 & \dots & c-x_L & -w_1 & -w_2 & \dots & -w_L \\ c-x_1 & c-x_1 & c-x_2 & \dots & c-x_L & \tilde{F}(x_1)-\sum_{i=0}^{1} w_i & -w_2 & \dots & -w_L \\ c-x_2 & c-x_2 & c-x_2 & \dots & c-x_L & 0 & \tilde{F}(x_2)-\sum_{i=0}^{2} w_i & \dots & -w_L \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c-x_L & c-x_L & c-x_L & \dots & c-x_L & 0 & 0 & \dots & \tilde{F}(x_L)-\sum_{i=0}^{L} w_i \end{bmatrix} \tag{9}$$

Algorithm 1 Predictor–corrector method with step–size control for parameter tracking

1: $\gamma := 0$
2: $\underline{\eta}_0 := \underline{\eta}(\gamma = 0)$ // $G(\underline{\eta}_0, \gamma = 0) = \underline{0}$
3: $\Delta\gamma := \gamma_{\texttt{step\_min}}$
4: **repeat**
5: $\quad \gamma := \gamma + \Delta\gamma$
6: $\quad \underline{\eta}_{\text{tmp}} := \text{Predictor}\left(\underline{\eta}, \gamma\right)$
7: $\quad \left[\underline{\eta}_{\text{tmp}}, \mathit{success}\right] := \text{Corrector}\left(\underline{\eta}_{\text{tmp}}\right)$
8: $\quad$ **if** *success* **then**
9: $\quad\quad \underline{\eta} := \underline{\eta}_{\text{tmp}}$
10: $\quad\quad \text{Increase}(\Delta\gamma, \gamma_{\texttt{step\_max}})$
11: $\quad$ **else**
12: $\quad\quad \gamma := \gamma - \Delta\gamma$
13: $\quad\quad \text{Decrease}(\Delta\gamma, \gamma_{\texttt{step\_min}})$
14: $\quad$ **end if**
15: **until** $\gamma = 1$

every γ, we initialize $\underline{\eta}$ so that $\underline{\eta}_0 = \arg\min G(\underline{\eta}, \gamma = 0)$ holds. We are able to do this, since we defined the parameterization of $G(\underline{\eta})$ that way in Section IV. We then initialize the step size $\Delta\gamma$ to the minimal step size.

The main part of the algorithm is a loop, in which γ is increased until $\gamma = 1$ is reached. In each loop iteration a parameter vector $\underline{\eta}_{\text{tmp}}$ for the increased γ is predicted. The corresponding predictor is described in subsection V-A.

Subsequently a corrector step is applied to minimize the error introduced by the predictor. The corrector is given in subsection V-B. If the corrector converges fast, we know, that the predicted parameter vector produced only a small error. Hence, we can increase the step size for the next iteration. If the convergence of the corrector fails, we know that the predicted parameter vector produces a large error. Hence, we revert the prediction step and decrease the step size before proceeding with the loop iteration.

A. Predictor

The predictor solves the system of equations given by (8) for $\dot{\underline{\eta}}$. We emphasize here, that we solve the system in this step as a linear system of equations for a given γ.

The parameter vector is then predicted as

$$\underline{\eta}_{\text{tmp}} = \underline{\eta} + \Delta\gamma\, \dot{\underline{\eta}} \ .$$

Since $\dot{\underline{\eta}}$ only gives a direction, this step causes an error governed by the step size $\Delta\gamma$. This error must be decreased by the corrector step.

B. Corrector

The corrector step is based on a Newton approach for determining roots. Following this approach, we apply the equation

$$\mathbf{J}\left(\underline{\eta}_k\right) \underbrace{\left[\underline{\eta}_{k+1} - \underline{\eta}_k\right]}_{\Delta\underline{\eta}} = -\underline{h}\left(\underline{\eta}_k\right) \tag{10}$$

where

$$\underline{h}\left(\underline{\eta}\right) = \frac{\partial G(\underline{\eta}, \gamma)}{\partial \underline{\eta}} = \left[h_1^{(1)}, \dots, h_1^{(L)}, h_2^{(0)}, \dots, h_2^{(L)}\right]^T$$

with

$$h_1^{(i)}\left(\underline{\eta}\right) = \tilde{F}(x_i, \gamma) - \sum_{j=0}^{i-1} w_j - \frac{w_i}{2}$$

and

$$h_2^{(i)}\left(\underline{\eta}\right) = \int_{x_i}^{c} \tilde{F}(x, \gamma)\, dx - c\sum_{j=1}^{L} w_j - x_i \sum_{j=1}^{i} w_j - \sum_{j=i+1}^{L} w_j x_j \ .$$

It is important to note, that γ is a constant in this step.

$\mathbf{J}\left(\underline{\eta}_k\right)$ is the Jacobian defined by

$$\mathbf{J}\left(\underline{\eta}_k\right) = \frac{\partial \underline{h}\left(\underline{\eta}\right)}{\partial \underline{\eta}^T} = \frac{\partial^2 G(\underline{\eta}, \gamma)}{\partial \underline{\eta} \partial \underline{\eta}^T} \ ,$$

which is in this case identical to the matrix $\mathbf{P}$ given by (9).

We solve (10) for $\Delta\underline{\eta}$ to obtain the recursion

$$\underline{\eta}_{k+1} = \underline{\eta}_k - \Delta\underline{\eta} \ .$$

If the initial parameter vector $\underline{\eta}_0$ is close to the true parameter vector, the method converges very fast, what can be detected by $\Delta\underline{\eta} \to \underline{0}$. In this case we return success and the converged $\underline{\eta}$. If the approach does not converge in few steps we abort and return failure.

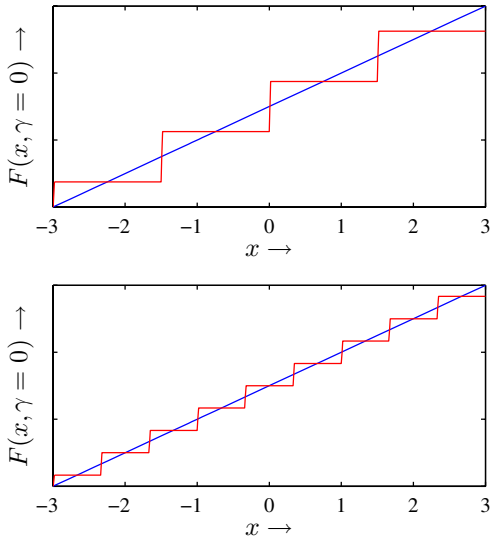

Fig. 2. Initial approximation of $\tilde{F}(x, \gamma = 0)$ with $L = 3$ and $L = 8$ components.

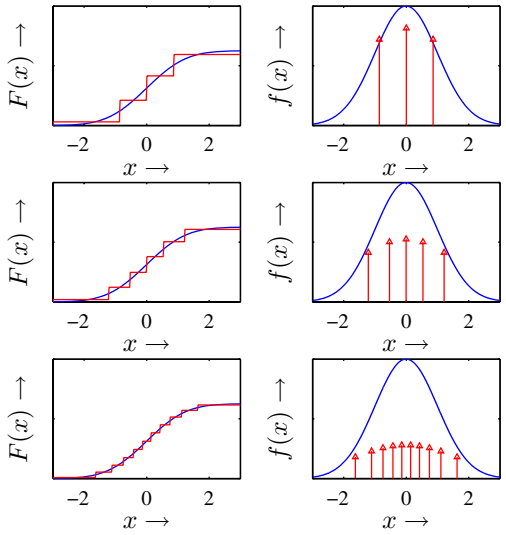

Fig. 3. Approximation of a standard normal distribution (left) and density (right) for a different number of components in the Dirac mixture approximation $L = 3$, $L = 5$, and $L = 10$.

VI. Examples

We now give two examples of approximating given parametric density functions by means of Dirac mixtures with different numbers of components.

A. Gaussian Density

First, we approximate a standard normal density. In order to guarantee that $\tilde{f}(x, \gamma = 0) = 1$, we use an unnormalized Gaussian and parameterize it according to

$$\tilde{f}(x, \gamma) = \exp\left\{ -\frac{1}{2} \frac{x^2}{\left(\frac{1+\epsilon}{\gamma+\epsilon}\sigma\right)^2} \right\} .$$

Since we solve the optimization problem in the distribution space, we have to consider

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

The lower boundary of the integral is due to the weighting function $g(x)$ of the weighted Cramér–von Mises distance we selected in section IV. The progression schedule for this function is depicted in Figure 1.

Since $\tilde{f}(x, \gamma = 0) = 1$, we have $\tilde{F}(x, \gamma = 0) = x$, which can be seen in the left most plots of Figure 1.

The optimal initial parameters of $F(x, \underline{\eta})$ with respect to the defined distance measure depend on the number of components. For L components we apply

$$x_i = \frac{2\,c\,i}{L+1} - c, \; w_0 = \frac{c}{L+1}, \; w_i = \frac{2\,c}{L+1} .$$

With these parameters we have $G\left(\underline{\eta}, \gamma = 0\right) = 0$, independent of the number of components. This can be seen in Figure 2 for $L = 3$ and $L = 8$ components.

The resulting approximation for $L = 3$, $L = 5$, and $L = 10$ is depicted in Figure 3. The solver needed 14, 25, and 49 steps respectively.

B. Gaussian Mixture Density

In the next example we approximate a Gaussian mixture density with three components. The parameters are given by

$$\begin{aligned} \underline{w} &= \left[0.45, 0.1, 0.45\right]^T \\ \underline{m} &= \left[-3, 0, 3\right]^T \\ \underline{\sigma} &= \left[0.5, 3, 0.5\right]^T \end{aligned}$$

as weight, mean and standard deviation values. The parameterization in each component of the Gaussian mixture is identical to the parameterization in the previous example. Hence, we start again with a uniform distribution and can reuse the initial approximation of the previous example.

The progression schedule for $L = 10$ components is depicted in Figure 4. The final approximation for $L = 3$, $L = 5$, and $L = 10$ components is shown in Figure 5. The solver needed 42, 66, and 111 steps respectively.

VII. Discussion and Future Work

This paper introduced a systematic procedure for approximating an arbitrary probability density function by means of a Dirac mixture by minimizing the weighted Cramér–von Mises distance. The results provide the basis for a recursive nonlinear filtering mechanism as an alternative to Monte Carlo based particle filters.

The proposed procedure has been introduced in the context of scalar random variables for the sake of simplicity. It can, however, be generalized to random vectors in a straightforward manner. In the multidimensional case, the efficient minimization procedure for obtaining optimal parameters is even more important than in the scalar case.

References

[1] 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.

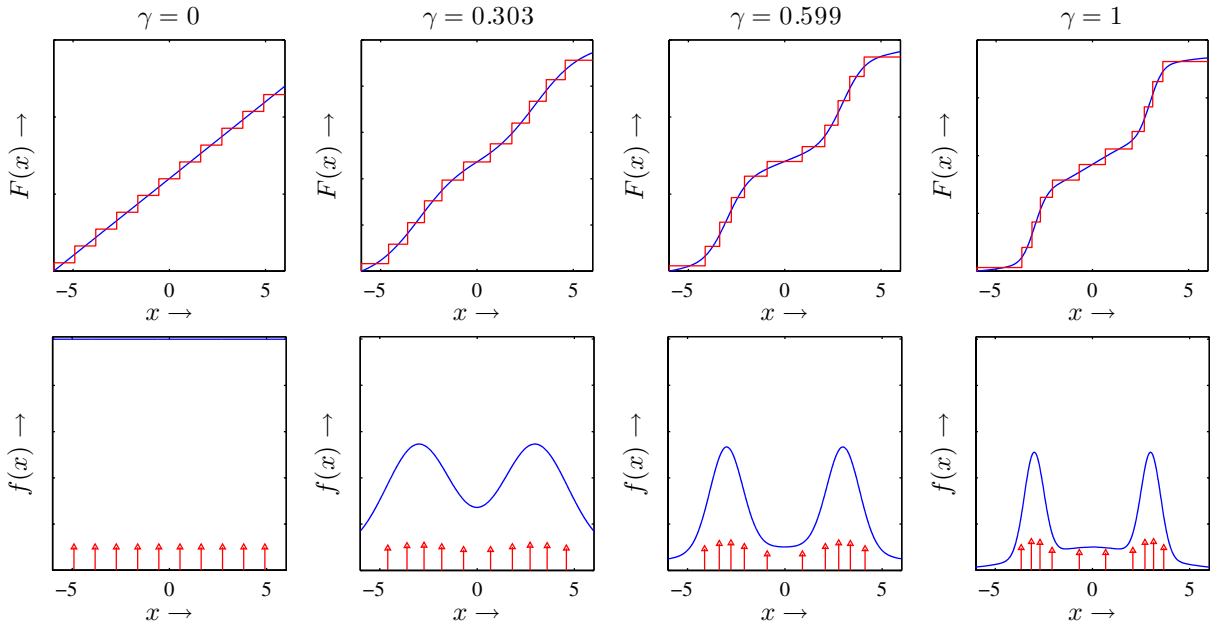

Fig. 4. Progressive approximation of a Gaussian mixture with three components for $\gamma = 0 \ldots 1$. The approximation density consists of $L = 10$ components. The height of each Dirac component in the density approximation corresponds to its weight value.

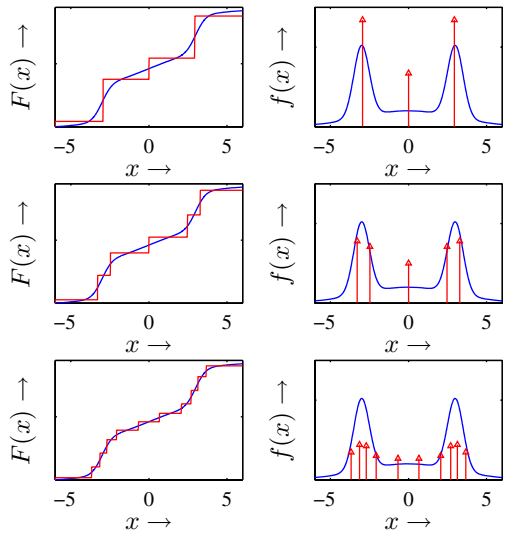

Fig. 5. Approximation of a Gaussian mixture distribution (left) and density (right) with 3 components for a different number of components in the Dirac mixture approximation $L = 3$, $L = 5$ and $L = 10$.

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

[3] D. Brigo, B. Hanzon, and F. LeGland, "A Differential Geometric Approach to Nonlinear Filtering: The Projection Filter," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 247–252, 1998.

[4] N. Oudjane and S. Rubenthaler, "Stability and Uniform Particle Approximation of Nonlinear Filters in Case of non–Ergodic Signal," Prépublication, Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VI, Tech. Rep. PMA786, January 2003. [Online]. Available: http://www.proba.jussieu.fr/mathdoc/textes/PMA-786.pdf

[5] 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.

[6] E. Bølviken, P. J. Acklam, N. Christophersen, and J.-M. Størdal, "Monte Carlo Filters for Non–Linear State Estimation," *Automatica*, vol. 37, pp. 177–183, 2001.

[7] 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.

[8] E. Bølviken and G. Storvik, "Deterministic and Stochastic Particle Filters in State Space Models," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer Series of "Statistics for Engineering and Information Science", 2001.

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

[10] D. H. Johnson and S. Sinanovic, "Symmetrizing the Kullback–Leibler Distance," submitted to IEEE Signal Processing Letters, 2002.

[11] M. D. Weber, L. M. Leemis, and R. K. Kincaid, "Minimum Kolmogorov-Smirnov Test Statistic Parameter Estimates," *Journal of Statistical Computation and Simulation*, vol. 76, no. 3, pp. 196–206, 2006.

[12] 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.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Kolmogorov–Smirnov Test," in *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992, pp. 623–628.

[14] O. Ozturk and T. P. Hettmansperger, "Generalized Weighted Cramér–von Mises Distance Estimators," *Biometrika*, vol. 84, no. 2, pp. 283–294, 1997.

[15] U. D. Hanebeck and O. Feiermann, "Progressive Bayesian Estimation for Nonlinear Discrete–Time Systems: The Filter Step for Scalar Measurements and Multidimensional States," in *Proceedings of the 2003 IEEE Conference on Decision and Control (CDC'03), Maui, Hawaii*, 2003, pp. 5366–5371.

[16] 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.