Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions

Uwe D. Hanebeck and Oliver C. Schrempf

Abstract— **Greedy procedures for suboptimal Dirac mixture approximation of an arbitrary probability density function are proposed, which approach the desired density by sequentially adding one component at a time. Similar to the batch solutions proposed earlier, a distance measure between the corresponding cumulative distributions is minimized by selecting the corresponding density parameters. This is due to the fact, that a distance between the densities is typically not well defined for Dirac mixtures. This paper focuses on the Cramér–von Mises distance, a weighted integral quadratic distance measure between the true distribution and its approximation. In contrast to the batch solutions, the computational complexity is much lower and grows only linearly with the number of components. Computational savings are especially severe, when the required number of components, e.g. the minimum number of components for achieving a given quality measure, is not a priori known and must be searched for as well. The performance of the proposed sequential approximation approaches is compared to the optimal batch solutions.**

NOTATION

$f(x)$	density function
$F(x)$	(cumulative) distribution function
$\delta(x)$	Dirac delta function
$H(x)$	Heaviside step function
G	distance measure
$\underline{\eta}$	parameter vector of Dirac mixture
$\underline{\kappa}$	parameter vector of Gaussian mixture
$\mathcal{N}(.,m,\sigma)$	Gaussian density with mean m and standard deviation σ

I. INTRODUCTION

Nonlinear processing of random quantities described by (prior) probability density functions in general requires the approximation of the resulting (posterior) densities. The true densities might not be explicitly available because the calculation is either too complex or even impossible. In most cases, the resulting densities are simply of a type that is not well suited for further treatment. Hence, the true posterior density, i.e., the result of the processing step, is approximated by means of a density that is more convenient for subsequent processing steps. Many types of generic analytic density representations are available for that purpose, including Gaussian mixtures [1], Edgeworth series expansions [2], and exponential densities [3].

As an alternative representation, we propose to use mixtures of Dirac delta functions (impulse functions), so called Dirac mixtures, for approximating the underlying true densities in analytic form. This is different from representing densities by means of random samples [4], which is used by the popular particle filters [5], where the appropriate density parameters, i.e., weights and locations of the particles, are typically calculated by means of Monte Carlo techniques [6], [7].

The proposed approach for calculating an optimal Dirac mixture approximation for a given density relies on the systematic minimization of a certain distance measure between the two densities. Standard distance measures operating in the density domain such as the Kullback–Leibler distance [8], its symmetrized version [9], or integral quadratic distances are obviously not well suited for Dirac mixtures. Hence, comparison is performed in the distribution domain. As a result, the fundamental property of the Dirac delta function in the density domain can be exploited to simplify processing while the continuity of the corresponding staircase function in the distribution domain is used for comparison purposes.

Several approaches that simultaneously calculate the locations and weights of all components, so-called batch approaches, have already been published. The case of components with equal weights is treated in [10]. A method for the simultaneous calculation of optimal weights and locations is given in [11]. For both cases, efficient solution procedures for arbitrary true densities based on a homotopy continuation technique have been proposed.

Typically, the number of components required for achieving a certain approximation quality is not known a priori and must be determined as well. Two approaches for doing so are possible on the basis of the available batch solutions. The first approach relies on trying different numbers of components suitable for the given approximation problem and employs certain heuristics in order to avoid an exhaustive search. For every number of components investigated, a full recalculation is performed [12]. The second approach adds new components one at a time and adjusts all components to guarantee optimality of the resulting approximation [13]. In both approaches, the complexity of the parameter adjustment grows at least quadratically with the number of Dirac mixture components.

The greedy approaches proposed in this paper also build up the final approximate density by sequentially adding one component at a time. However, the components are inserted locally without affecting the set of components already placed. This procedure yields suboptimal results at a much lower computational load compared to the batch approaches. The complexity grows linearly with the number of components. As a result, computational savings are especially severe, when an exhaustive search is used for finding the minimum number of components required to satisfy a certain predefined quality measure.

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

Potential applications of the greedy approximation include the calculation of the density of a function of random variables, long–term prediction [13], state estimation of nonlinear stochastic systems [11], reachability analysis, numerical integration, and even the generation of pseudo–random numbers.

The paper is structured as follows. The next section gives a mathematical formulation of the problem of approximating a given density by means of a Dirac mixture. Section III is concerned with several possible approaches for determining the parameters of a desired Dirac mixture approximation. Subsection III-A starts with the simplest problem of calculating the weights of a Dirac mixture approximation when all the locations are already known. Subsection III-B then considers the dual case of given weights and the calculation of appropriate locations. This of course also includes the interesting special case of equally weighted components. Finally, Subsection III-C discusses the most general case of calculating weights and locations simultaneously. In all three cases, batch approaches and their sequential counterparts are introduced. Section IV is concerned with the identification of critical locations for inserting new components. Several viable options are discussed for that purpose. Section V compares the two solution approaches, i.e., the batch approach and the sequential approach, introduced in Subsection III-C. The paper concludes with a discussion of the proposed new approximation approaches and an outlook to future work.

II. Problem Formulation

We consider a given true probability density function $\tilde{f}(x)$ that is approximated by means of a Dirac mixture density given by

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

with weighting factors w_i and components $\delta(x - x_i)$ at locations x_i. The components are given by Dirac delta functions according to

$$\delta(x - x_i) = \begin{cases} \text{undefined}, & x = x_i \\ 0, & \text{elsewhere} \end{cases} \ ,$$

with

$$\int_{-\infty}^{\infty} \delta(x - x_i) \, dx = \int_{x_i-\epsilon}^{x_i+\epsilon} \delta(x - x_i) \, dx = 1$$

for some $\epsilon > 0$.

The parameter vector $\underline{\eta}$ contains both the weighting factors and the locations of the individual Dirac functions according to

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

The weighting factors are positive

$$w_i > 0$$

for $i = 1, \ldots, L$ and sum up to one according to

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

The distribution function corresponding to the true density $\tilde{f}(x)$ is denoted by $\tilde{F}(x)$. The distribution function corresponding to the approximate density $f(x,\underline{\eta})$ is denoted by $F(x,\underline{\eta})$ and is given by a Heaviside mixture according to

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

$H(x - x_i)$ are Heaviside step functions for $i = 1, \ldots, L$ given by

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

Our goal is to perform a sequential approximation of $\tilde{f}(x)$ by starting with a single Dirac mixture component and subsequently adding one component at a time according to a certain criterion, until a certain distance measure G between the given density $\tilde{f}(x)$ and its approximation $f(x)$ is below a given threshold. As distance measures in the density domain are typically not well defined for Dirac mixture densities, distance measures between (cumulative) distributions are employed. In the following, we restrict our focus to quadratic distance measures evaluated on the interval $[x_{\min}, x_{\max}]$.

III. Possible Approaches

In this section, three types of approximation approaches will be introduced. In all cases, a batch solution, i.e., a solution considering the parameters of all components simultaneously, is introduced first. Subsequently, corresponding suboptimal sequential versions are derived that are all of greedy type and based on inserting components one at a time.

Remark III.1 *In all approaches considered, intervals between Dirac components are considered for inserting new component. Of course, inserting a new component only makes sense, when the true distribution $\tilde{F}(.)$ is strictly increasing over the considered interval.*

The first and most simple approach is based on given component locations so that only the weighting factors have to be adapted. The second approach is dual to the first approach in the sense that here the weights are fixed a priori and the locations remain to be calculated. In the third and most general approach, both weights and locations of the components are calculated.

A. First Approach: Calculation of Weights for Fixed Locations

The first approach assumes that the component locations have been fixed a priori and only the weighting factors remain to be calculated.

1) Batch Approach: The component locations are given at arbitrary, but fixed positions x_i for $i = 1, \ldots, L$. Although the calculation of the corresponding weights works for arbitrary locations, we assume that the components are placed at equidistant locations

$$x_i = x_{\min} + \frac{d}{2} + (i-1)d = x_{\min} + \frac{2i-1}{L} d \ ,$$

on the interval $[x_{\min}, x_{\max}]$ with $d = (x_{\max} - x_{\min})/L$. The locations of the first and the last component are given as

$$x_1 = x_{\min} + \frac{d}{2}$$

and

$$x_L = x_{\min} + \frac{2L-1}{2} d = x_{\max} - \frac{d}{2} \quad ,$$

respectively.

Of course, the corresponding weights w_i could be calculated according to some optimality criterion, e.g. by adapting the procedure given in Subsection III-C.

A simpler approach is to view the components as being the center of intervals of width d and to select the weighting factors in such a way that the distributions at the ends of these intervals are identical, i.e., $F(x_i - d/2) = \tilde{F}(x_i - d/2)$ and $F(x_i + d/2) = \tilde{F}(x_i + d/2)$.

This is achieved by selecting the corresponding weights of the approximating distribution as

$$w_i = \tilde{F}\left(x_i + \frac{d}{2}\right) - \tilde{F}\left(x_i - \frac{d}{2}\right) \quad ,$$

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

2) Sequential Approach: The sequential approach is based on replacing a single component by two components with smaller weights. This is achieved by splitting the interval containing the single component into two intervals of equal width. When splitting is performed in such a way that all the resulting intervals are of equal sizes, this yields the same approximation density as the batch approach. A more efficient solution is obtained when replacing a component by two new components only when the distance between the true distribution $\tilde{F}(x_i)$ and its approximation in the corresponding interval is large, see Section IV. By doing so, the sequential procedure automatically provides non–equally spaced components.

B. Second Approach: Calculation of Location for Fixed Weights

The second approach assumes that the weighting factors have been fixed a priori and only the component locations remain to be calculated.

For the case of equally weighted components, an optimal batch solution has been derived in [10], so that only the results will be given here.

1) Batch Approach: For L equally weighted components, the weighting factors are given by $w_i = 1/L$ for $i = 1, \ldots, L$. For an integral quadratic distance measure G between the true distribution $\tilde{F}(x_i)$ and its approximation according to

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

the optimal locations are given by

$$\tilde{F}(x_i) = \frac{2i-1}{2L} \tag{3}$$

for $i = 1, \ldots, L$ [10]. It is important to note that the locations x_i can be calculated independently of each other.

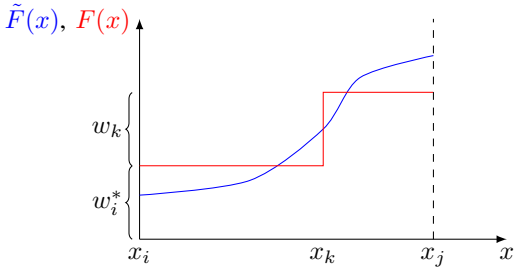

Fig. 1. Inserting a component in a regular interval at location x_k with weight w_k.

2) Sequential Approach: The sequential version successively divides the range of *values* of $\tilde{F}(x_i)$ between 0 and 1 into subintervals of equal size. For every new subinterval, a new component is inserted.

The approximation process is started by placing a single component with unit weighting factor at a location corresponding to $\tilde{F}(x_1) = 1/2$, which corresponds to setting $L = 1$ in (3) and yields the median of $\tilde{f}(x)$. Subsequent components are then placed by splitting up the resulting intervals recursively. As can easily be seen, the sequential solution provides the same result as the batch solution when every interval is split into equal sizes. However, a more efficient solution is achieved when inserting a new component in a certain interval only when the distance between the true distribution $\tilde{F}(x_i)$ and its approximation is large, see Sec. IV. In that case, the sequential procedure automatically provides different weighting factors from the set

$$\mathcal{W} = \left\{ \frac{1}{2^j}, j = 0, 1, 2, \ldots \right\}$$

depending upon the splitting level reached in the approximate density.

C. Third Approach: Calculation of Weights and Locations

The third approach is the most general as it calculates the weighting factors and the component locations simultaneously.

1) Batch Approach: The optimal batch solution for the simultaneous calculation of the weighting factors and the component locations has been derived in [11].

2) Sequential Approach: In this section, we assume that the critical location for inserting a new component is already known. The new component is then inserted in such a way — by selecting appropriate values for the weighting factor and the component location — that a distance measure between the true density and its approximation is minimized. The identification of critical locations will be discussed in the next section.

In the special case of scalar densities, it is sufficient to consider the intervals between Dirac components and the two border intervals, i.e., the interval from $x_{\min}$ to x_1 and the interval from x_L to $x_{\max}$. The border cases, i.e., adding a new component to the left or to the right of an existing Dirac mixture, require special attention and will be treated separately from the regular intervals.

For a regular interval, we assume that it is defined by two components indexed by i and j. A third component indexed by k is then inserted in between.

For determining the parameters of the new component according to Figure 1, a distance measure

$$G(w_i^*, w_k, x_k) = \frac{1}{2} \int_{x_i}^{x_j} \left(\tilde{F}(x) - w_i^* - w_k H(x - x_k) \right)^2 dx$$

is defined, where w_i^* is the absolute value of the approximate distribution at location x_i and w_k the weight at location x_k. Taking the partial derivatives with respect to w_i^*, w_k, and x_k gives three necessary conditions

$$\frac{\partial G}{\partial w_i^*} = \int_{x_i}^{x_j} \left(\tilde{F}(x) - w_i^* - w_k H(x - x_k) \right) dx \stackrel{!}{=} 0 \ ,$$

$$\frac{\partial G}{\partial w_k} = \int_{x_i}^{x_j} \left(\tilde{F}(x) - w_i^* - w_k H(x - x_k) \right) H(x - x_k)\, dx \stackrel{!}{=} 0 \ ,$$

and

$$\frac{\partial G}{\partial x_k} = \int_{x_i}^{x_j} \left(\tilde{F}(x) - w_i^* - w_k H(x - x_k) \right) w_k\, \delta(x - x_k)\, dx \stackrel{!}{=} 0 \ .$$

After simplification, we obtain three equations

$$\int_{x_i}^{x_j} \tilde{F}(x)\, d\, x = w_i^* (x_j - x_i) + w_k (x_j - x_k) \ ,$$

$$\int_{x_k}^{x_j} \tilde{F}(x)\, d\, x = (w_i^* + w_k)(x_j - x_k) \ ,$$

and

$$\tilde{F}(x_k) = w_i^* + \frac{w_k}{2} \ .$$

Solving this system of equations yields a nonlinear equation in x_k only, which is given by

$$\frac{\int_{x_i}^{x_k} \tilde{F}(x)\, dx}{x_k - x_i} + \frac{\int_{x_k}^{x_j} \tilde{F}(x)\, dx}{x_j - x_k} = 2\, \tilde{F}(x_k) \ .$$

Once the optimal location x_k is known, the weighting factors are obtained as

$$w_i^* = \frac{\int_{x_i}^{x_k} \tilde{F}(x)\, dx}{x_k - x_i}$$

and

$$w_k = \frac{\int_{x_k}^{x_j} \tilde{F}(x)\, dx}{x_j - x_k} - w_i^* \ .$$

Now, the cases of adding a component to the left or to the right of an existing Dirac mixture are derived. For the left border interval, i.e., the interval between $x_{\min}$ and the first component, the distance measure according to Figure 2 is given by

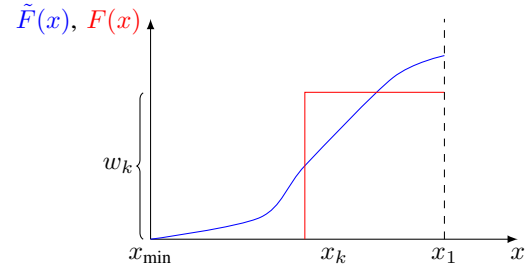

Fig. 2. Inserting a component in the left border interval at location x_k with weight w_k.

$$G(w_k, x_k) = \frac{1}{2} \int_{x_{\min}}^{x_1} \left(\tilde{F}(x) - w_k H(x - x_k) \right)^2 dx \ ,$$

and two necessary conditions can be obtained. The first condition is

$$\frac{\partial G}{\partial w_k} = \int_{x_{\min}}^{x_1} \left(\tilde{F}(x) - w_k H(x - x_k) \right) H(x - x_k)\, dx \stackrel{!}{=} 0 \ ,$$

resulting in

$$\int_{x_k}^{x_1} \tilde{F}(x)\, dx = w_k (x_1 - x_k) \ .$$

The second condition is obtained as

$$\frac{\partial G}{\partial x_k} = \int_{x_{\min}}^{x_1} \left(\tilde{F}(x) - w_k H(x - x_k) \right) w_k \delta(x - x_k)\, dx \stackrel{!}{=} 0 \ ,$$

which yields

$$\tilde{F}(x_k) = \frac{1}{2} w_k \ . \tag{4}$$

Eliminating w_k gives a single nonlinear integral equation for x_k

$$2\tilde{F}(x_k) = \frac{\int_{x_k}^{x_1} \tilde{F}(x)\, dx}{x_1 - x_k} \ .$$

Once x_k has been calculated, the weighting factor w_k is obtained from (4).

For the right border interval, i.e., the interval between the last component and $x_{\max}$, the distance measure according to Figure 3 is given by

$$G(w_L^*, x_k) = \frac{1}{2} \int_{x_L}^{x_{\max}} \Big(\tilde{F}(x) - w_L^* - \left(\tilde{F}(x_{\max}) - w_L^* \right) H(x - x_k) \Big)^2 dx \ ,$$

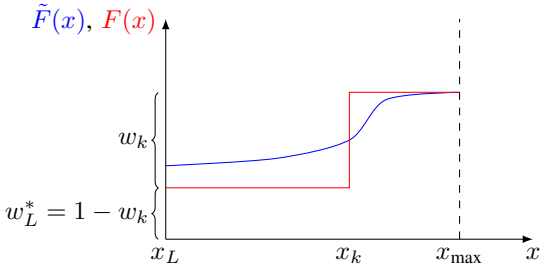

Fig. 3. Inserting a component in the right border interval at location x_k with weight w_k.

where $w_L^* = 1 - w_k$, which leads to

$$\frac{\partial G}{\partial w_L^*} = -\int_{x_L}^{x_{\max}} \left(\tilde{F}(x) - w_L^* - \left(\tilde{F}(x_{\max}) - w_L^*\right) \cdot H(x - x_k)\right)(1 - H(x - x_k))\, d\,x \stackrel{!}{=} 0 \ .$$

Simplification yields

$$\int_{x_L}^{x_{\max}} \left(\tilde{F}(x) - w_L^*\right)(1 - H(x - x_k))\ dx \stackrel{!}{=} 0$$

and finally

$$\int_{x_L}^{x_k} \tilde{F}(x)\, dx = w_L^*(x_k - x_L) \ .$$

The derivative of $G(w_L^*, x_k)$ with respect to x_k gives

$$\frac{1}{2}\int_{x_L}^{x_{\max}} \left(\tilde{F}(x) - w_L^* - (\tilde{F}(x_{\max}) - w_L^*) \cdot H(x - x_k)\right)\delta(x - x_k)\, dx \stackrel{!}{=} 0 \ ,$$

which leads to

$$w_L^* = 2\,\tilde{F}(x_k) - \tilde{F}(x_{\max}) \ . \tag{5}$$

The elimination of w_L^* again gives a single nonlinear integral equation for x_k

$$2\tilde{F}(x_k) - \tilde{F}(x_{\max}) = \frac{\int_{x_L}^{x_k} \tilde{F}(x)\, dx}{x_k - x_L} \ .$$

Once x_k has been calculated, the weighting factor w_L^* is obtained from (5) and $w_k = 1 - w_L^*$.

For starting the sequential approximation process, we have to perform an initialization, that is, we have to place the first component. The corresponding distance measure between the true distribution and its approximation is given by

$$G(x_k) = \frac{1}{2}\int_{x_{\min}}^{x_{\max}} \left(\tilde{F}(x) - H(x - x_k)\right)^2\ dx \ .$$

Taking the derivative with respect to x_k gives the necessary condition

$$\frac{\partial\, G}{\partial\, x_k} = \int_{x_{\min}}^{x_{\max}} \left(\tilde{F}(x) - H(x - x_k)\right)\delta(x - x_k)\, dx \stackrel{!}{=} 0 \ ,$$

which leads to

$$\tilde{F}(x_k) = \frac{1}{2} \ . \tag{6}$$

The solution x_k^s of this equation yields a minimum of $G(x_k)$, since the second derivative of $G(x_k)$ is larger than zero, i.e.,

$$\left.\frac{\partial^2\, G(x_k)}{\partial\, x_k^2}\right|_{x_k^s} = \left.\frac{\partial\, \tilde{F}(x_k)}{\partial\, x_k}\right|_{x_k^s} > 0 \ .$$

Remark III.2 *The condition in (6) is equivalent to placing the first component at the median of the given density $\tilde{f}(x)$.*

IV. Identification of Critical Locations

There are several options available for selecting the specific interval between Dirac mixture components used for inserting a new component.

When we consider the deviation between the true prior and its approximation, the first option is to determine the interval with the *maximum deviation*. Unfortunately, closed–form expressions for the deviation are not available for most types of true densities, not even for Gaussian mixtures. However, since distributions are always monotonically increasing, an interval containing the true value of the deviation can easily be calculated by means of upper/lower-bounding techniques to an arbitrary precision.

In the special case of a Gaussian mixture (true) prior, an analytic expression for the **absolute** deviation

$$G = \int_{x_i}^{x_j} \left|\tilde{F}(x) - F(x)\right| dx$$

between the true prior distribution $\tilde{F}(x)$ and the approximate prior $F(x)$ over the interval $[x_i, x_j]$ can be derived by adding up the areas between the true distributions. The second option is to determine the interval where inserting a new component leads to the maximum reduction of the deviation between the true prior and its approximation, which is typically different from the result according to the first option. The disadvantage of this option is the necessity of performing the insertion procedure for every interval considered, before the reduction can be calculated.

V. Comparison with Batch Optimization

In this section, we focus on the methods for calculating both weights and locations given in Subsection III-C. After a comparison of the features of the batch approach and its sequential counterpart in Subsection V-A, the performance of the two approaches will be compared by means of simulations in Subsection V-B.

A. Comparison of Features

The features of the two approaches, batch solution and sequential solution, for calculating both weights and locations given in Subsection III-C are compared in Table I.

TABLE I

COMPARISON OF THE FEATURES OF THE BATCH SOLUTION AND THE SEQUENTIAL SOLUTION GIVEN IN SUBSECTION III-C.

	Batch Solution	**Sequential Solution**
Optimality of approximation	Optimal	Suboptimal
Modification of number of components	Recalculation required	Addition of components one at a time
Computational complexity	Quadratic in number of components	Linear in number of components

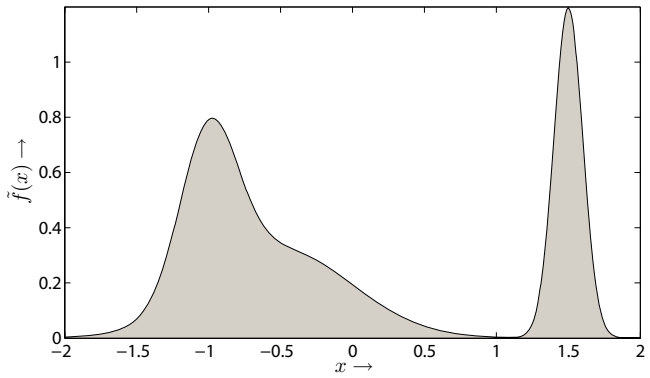

Fig. 4. Gaussian mixture density corresponding to the parameter vector $\underline{\kappa} = [0.3, 0.4, 0.3, -1.0, -0.5, 1.5, 0.2, 0.5, 0.1]^T$.

B. Comparison by Simulations

For comparing the performance of the two approaches, a Gaussian Mixture

$$\tilde{f}(x, \underline{\kappa}) = \sum_{i=1}^{K} c_i \, \mathcal{N}(x, m_i, \sigma_i)$$

is approximated by means of a Dirac mixture according to (1). $\underline{\kappa}$ is the parameter vector given by

$$\underline{\kappa} = \left[c_1, \ldots, c_K, m_1, \ldots, m_K, \sigma_1, \ldots, \sigma_K\right]^T \quad .$$

For evaluation purposes, we use $K = 3$ and

$$\begin{array}{ccc} c_1 = 0.3\,, & c_2 = 0.4\,, & c_3 = 0.3\,, \\ m_1 = -1.0\,, & m_2 = -0.5\,, & m_3 = 1.5\,, \\ \sigma_1 = 0.2\,, & \sigma_2 = 0.5\,, & \sigma_3 = 0.1\ . \end{array}$$

The corresponding Gaussian mixture density with three components is depicted in Figure 4.

Dirac mixture approximations of this Gaussian mixture are shown in Figure 5 for $L = 5$ (first row), $L = 10$ (second row), and $L = 15$ (third row) components. The left column corresponds to the Dirac mixtures calculated by means of the batch solution in Subsection III-C 1). The right column corresponds to the sequential solution in Subsection III-C 2).

The square root of the integral quadratic deviation according to (2) between the true distribution $\tilde{F}(x, \underline{\kappa})$ and the Dirac mixture approximation $F(x, \underline{\eta})$ provided by the two approaches is shown in Figure 6 for the number of components L varying between 5 and 25. The batch solution yields a better approximation quality, however, at a significantly higher computational cost and the need for a complete recalculation for every new number of components L.

VI. DISCUSSION AND FUTURE WORK

A procedure for the sequential enhancement of a Dirac mixture approximation by successive insertion of new components has been introduced. Compared to available batch solutions [10], [11], the proposed sequential solutions provide suboptimal results but have a much lower computational complexity. Furthermore, the complexity of the (optimal) insertion of a new component does not depend upon the number of components already available. Identification of an (optimal) location for inserting a new component into a given Dirac mixture approximation is at most linear in the number of components.

In many applications, a Dirac mixture approximation with a minimum number of components for achieving a certain task is required. When an exhaustive search is conducted, the proposed sequential approximation procedures are very useful as they allow to successively increase the number of components without extra effort. The successive application of the batch solutions yields better results. However, the higher computational complexity makes them unattractive in this case.

Of course, it is possible to combine the advantages of the sequential and the batch solution. The first combination uses both solutions separately and consists of using the sequential solution for estimating the appropriate number of components and the batch solution for calculating the optimal parameters of a mixture of this size. This is based on the assumption that the sequential solution is not far from the batch solution. The second combination is a bit tighter and uses the batch solution as a subalgorithm of a greedy scheme for calculating a local approximation that inserts more than one new component.

Future work is concerned with sequential solution approaches capable of providing optimal approximations. Of course, this requires the modification of the parameters of all the already existing components when inserting a new component. In addition, the order of component insertion at specific locations is important.

REFERENCES

[1] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, Aug. 1972.

[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, June 2000.

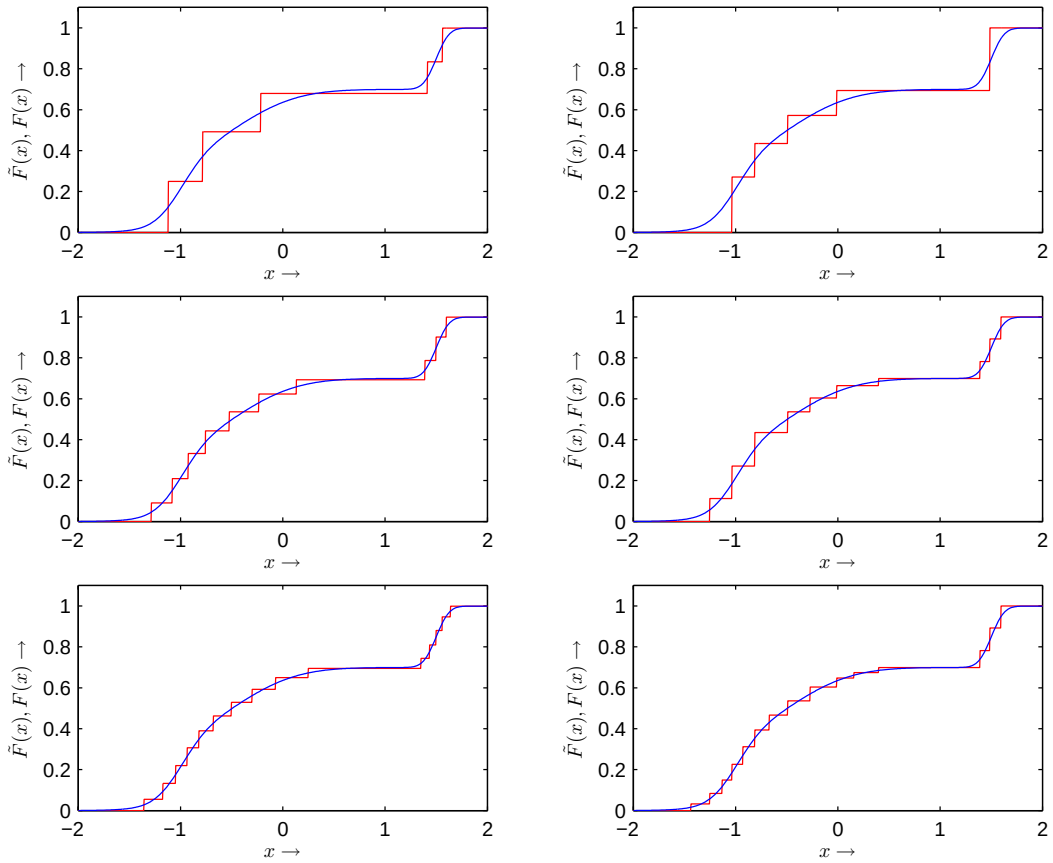

Fig. 5. Dirac mixture approximations of the Gaussian mixture from Figure 4 for $L = 5$ (first row), $L = 10$ (second row), and $L = 15$ (third row) components. The left column corresponds to the Dirac mixtures calculated by means of the batch solution in Subsection III-C 1). The right column corresponds to the sequential solution in Subsection III-C 2).

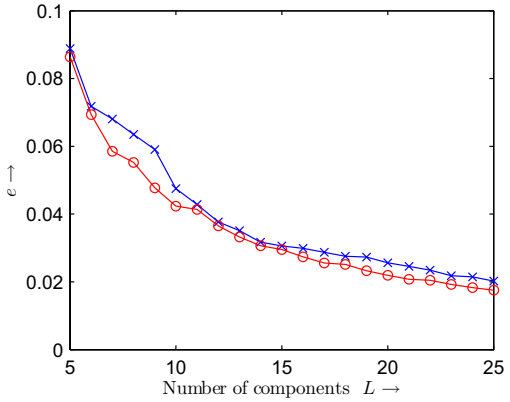

Fig. 6. Plot of the square root of the integral quadratic deviation between the true distribution $\tilde{F}(x, \underline{\kappa})$ and the Dirac mixture approximation $F(x, \underline{\eta})$ for different numbers of components L. The Dirac mixture approximation provided by the batch solution is marked by an 'o', the approximation provided by the sequential solution is marked by an 'x'.

[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, Jan. 2003. [Online]. Available: http://www.proba.jussieu.fr/mathdoc/textes/PMA-786.pdf

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

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

[10] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering," in *IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006.

[11] ——, "Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance," in *International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sept. 2006, pp. 512–517.

[12] O. C. Schrempf and U. D. Hanebeck, "Recursive Prediction of Stochastic Nonlinear Systems Based on Dirac Mixture Approximations," in *Proceedings of the American Control Conference (ACC 2007)*, New York, New York, July 2007.

[13] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: A New Framework for Nonlinear State Estimation," in *Proceedings of SPIE, AeroSense Symposium*, vol. 5099, Orlando, Florida, May 2003, pp. 256 – 267.