Deterministic Dirac Mixture Approximation of Gaussian Mixtures

Igor Gilitschenski*, **Jannik Steinbring***, **Uwe D. Hanebeck***, and **Miroslav Šimandl**†

*Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
gilitschenski@kit.edu, jannik.steinbring@kit.edu, uwe.hanebeck@ieee.org

† European Centre of Excellence – New Technologies for
Information Society and Department of Cybernetics
University of West Bohemia, Czech Republic
simandl@kky.zcu.cz

***Abstract*—In this work, we propose a novel way to approximating mixtures of Gaussian distributions by a set of deterministically chosen Dirac delta components. This approximation is performed by adapting a method for approximating single Gaussian distributions to the considered case. The proposed method turns the approximation problem into an optimization problem by minimizing a distance measure between the Gaussian mixture and its Dirac mixture approximation. Compared to the simple Gaussian case, the minimization criterion is much more complex as multiple, non-standard Gaussian distributions have to be considered.**

***Keywords*—Deterministic sampling, shape approximation, statistical distance, nonlinear propagation**

I. Introduction

Approximation of a continuous probability distribution using a discrete probability distribution is of particular interest in a wide area of applications. It is mainly motivated by the fact that handling discrete probability distributions is in some cases easier, particularly when nonlinear propagation of random variables needs to be performed. Furthermore, such approaches can be used for approximate numerical evaluation of integrals, for approximate inference, and for approximately solving uncertain optimization and feasibility problems. Thus, sampling appears in a broad area of real world applications involving state estimation in dynamic systems, optimal control, or parameter estimation in financial markets.

The considered problem of approximating continuous probability distributions by a discrete counterpart is usually addressed in one of two ways. First, random sampling can be used to approximate continuous probability distributions giving rise to a broad body of research in the context of Monte Carlo methods [1], [2]. Second, deterministic approximations are used, where some distance measure is minimized to obtain an optimal result. These approaches appear in the literature in the context of deterministic Dirac mixture approximation and they are related to optimal quantization approaches. We are particularly interested in the deterministic class of approaches, because they offer a homogeneous coverage of the underlying space and reproducibility of results.

The most wide-spread approach for considering nonlinearities in propagation of Gaussian uncertainties was proposed in the unscented Kalman filter (UKF) [3]. The approximation of an n-dimensional Gaussian density is based on placing $2n+1$ samples on the main axes of the corresponding covariance ellipsoid around the mean. In the Gaussian Filter (GF) [4], [5], this idea was generalized to using an arbitrary number of samples on each axis. Deterministic sampling can also be performed on nonlinear domains, such as the circle [6] or the hypersphere [7].

For propagation of Gaussian quantities involving strong nonlinearities, the use of an approximation which places all samples on the axes might be insufficient, i.e., it is of interest to achieve a better coverage of the state space. A semi-random approach was proposed in [8], where random sampling from the underlying Gaussian is replaced by sampling of several transformed UKF sample-sets. This results in a randomized UKF (RUKF). An entirely deterministic procedure is necessary in order to achieve a reproducible homogeneous coverage of the underlying space. This can be done by minimizing a suitable distance measure as in an approach based on Localized Cumulative Distributions (LCDs) [9], considering a numerical integration based approach as in the Gaussian-Hermite Kalman filter (GHKF) [10], or by using quantization techniques [11], [12]. In our previous work, the LCD based approach was used to approximate Gaussian densities [13], [14]. It is also possible to consider constraints, e.g., to maintain certain moments [15]. LCD based deterministic sampling has also been applied to linear regression Kalman filters in the Smart Sampling Kalman Filter (S^2KF) [16], and for a progressive Bayesian update step [17]. An overview of some approximations discussed here is shown in Fig. 1.

This paper contributes an approximation technique for Gaussian mixtures. Their use is of particular interest for at least two reasons. First, a mixture of Gaussian functions can be used to approximate arbitrary functions, and thus, Gaussian mixture distributions can be applied as an approximation [18] of arbitrary continuous probability densities (the degree of optimality can be predefined under certain regularity assumptions). Second, Gaussian mixtures naturally arise in stochastic filtering scenarios involving a prior given by some discrete distribution

and additive Gaussian system- or measurement noise. Thus, the proposed method can be used for handling stochastic filtering scenarios involving both strong nonlinearities in the system function and a complicated noise structure.

In the following, we extend our previous work [13] by proposing an approximation of Gaussian mixtures using equally weighted Dirac mixtures. For optimal approximation of a single Gaussian, it is sufficient to derive the distance measure approximating axis-aligned densities, because applying a suitable rotation does not break the optimality. However, this procedure cannot be used for approximating a Gaussian mixture with an arbitrary number of components because it is not possible to transform all of them into an axis-aligned form simultaneously. Thus, we compute a distance measure for arbitrary Gaussian mixtures. Furthermore, we provide its derivatives to speed-up numerical optimization.

The remainder of this paper is structured as follows. In the next section, we formulate the considered problem as a global optimization problem. The distance measure for this optimization problem is revisited in Sec. III. It is based on a generalization of the classical cumulative distribution function using Localized Cumulative Distributions and an adaption of the Cramér-von Mises criterion to obtain a new distance measure. The resulting distance measure and its derivatives are derived for the case of comparing Gaussian mixtures with Dirac mixtures in Sec. IV. The proposed approach is evaluated in Sec. V by discussing implementation issues and a comparison to a naïve approximation procedure. An outlook and possible future directions of research conclude the work.

II. Considered Problem

We consider a given n-dimensional Gaussian mixture density $f_{GM}(\underline{x})$ with M components parametrized by $\underline{\mu}_i$, $\mathbf{C}_i$ (where $i = 1, \ldots, M$), and respective positive weights w_i (with $w_1 + \ldots + w_L = 1$). That is, our density is given by

$$f_{GM}(\underline{x}) = \sum_{i=1}^{M} w_i \cdot \mathcal{N}(\underline{x} - \underline{\mu}_i, \mathbf{C}_i) \ ,$$

where $\mathcal{N}(\underline{x}, \mathbf{C})$ denotes the density of a zero-mean Gaussian distribution evaluated at $\underline{x}$ with covariance $\mathbf{C}$.

The problem considered in this paper is finding an approximation of the continuous density $f_{GM}(\underline{x})$ by a mixture of a given number L of equally weighted Dirac delta components. This can be thought of as finding a discrete distribution taking one of L different values $\underline{x}_i$ with equal probability $1/L$ such that it optimally approximates the shape of the Gaussian mixture. As the number of Dirac components is predefined, our task is the positioning of the $(\underline{x}_i)_{i=1,\ldots,L}$ in an optimal way. We represent this Dirac mixture distribution as

$$f_D(\underline{x}) = \sum_{i=1}^{L} \frac{1}{L} \cdot \delta(\underline{x} - \underline{x}_i)$$

and define a distance measure $D(\underline{\eta})$ between both considered distributions, where $\underline{\eta}$ is used to denote a parameter vector describing the location of the Dirac mixture components

$$\underline{\eta} = \left(\underline{x}_1^T, \ldots, \underline{x}_L^T\right)^T \ .$$

The resulting optimization problem is given as

$$\underline{\eta}^* = \arg\min_{\underline{\eta}} D(\underline{\eta}) \ . \tag{1}$$

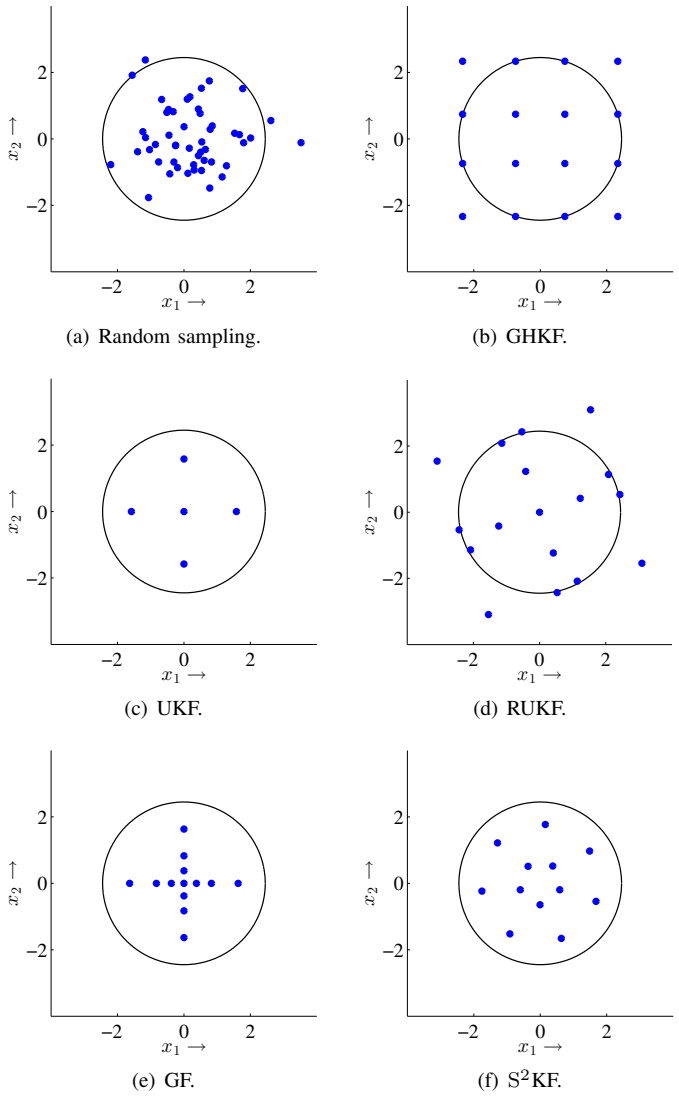


Figure 1. Different sample based approximation approaches for Gaussian densities.

Thus, a distance measure is needed for comparing continuous probability distributions and Dirac mixtures.

III. A Distance Measure for Dirac Mixture Approximation

The approach considered in this work yields a shape approximation of the underlying distribution, i.e., the positions of the individual components shall reflect the shape of the approximated distribution. Thus, the considered distance measure needs to compare local probability masses. Achieving this goal is rather difficult when using cumulative distribution functions (CDF), which are not symmetric and have a hard boundary for considering the probability mass. However, the CDF can be modified in order to put emphasis on local probability mass in a predefined way. This modification, denoted as the Localized Cumulative Distribution (LCD), can be used to define a suitable distance measure for probability distributions making a comparison between discrete and continuous distributions possible. Thus, in this section we revisit some definitions from [13].

Definition 1. *Consider a probability density function $f(\underline{x})$ defined on some n-dimensional domain $\mathbb{R}^n$. Then, its Localized Cumulative Distribution (LCD) is defined as*

$$F(\underline{m}, \underline{b}) = \int_{\mathbb{R}^n} f(\underline{x}) \cdot K(\underline{x} - \underline{m}, \underline{b}) \, \mathrm{d}\underline{x} \ ,$$

where $\underline{b} \in \mathbb{R}_+^n$ and $K(\cdot, \cdot)$ is symmetric and integrable.

The basic idea of this very general definition is best described by taking a look at its parameters and their natural meaning. Basically, $\underline{m}$ describes the location around which the probability mass is considered. The parameter $\underline{b}$ parametrizes the considered region around $\underline{m}$.

Gaussian kernels seem a natural choice in the considered scenario, because of several convenient properties. Particularly, the fact that the product of two Gaussian densities yields another (unnormalized) Gaussian density simplifies many computations. Thus, in this work isotropic Gaussian kernels are used. That is, the size of the kernel is described by a scalar value $b \in \mathbb{R}_+$. The resulting kernel is given as

$$K(\underline{x}, b) = \exp\left(-\frac{1}{2}\underline{x}^T(\tilde{\mathbf{C}}_b^{-1})\underline{x}\right) \ ,$$

where $\tilde{\mathbf{C}}_b = b^2 \, \mathbf{I}$.

In a univariate setting, the Cramér-von Mises criterion [19] can be used for comparing two probability distributions. A more general approach is desired for the purpose of our application. First, it is necessary to consider multivariate densities in order to approximate arbitrary Gaussian mixtures. Second, we require symmetry, because it follows the intuition for a distance measure. Using Localized Cumulative Distributions gives rise to a straightforward definition of the proposed distance measure.

Definition 2 (Modified Cramér-von Mises Criterion)**.** *The modified Cramér-von Mises criterion D between two LCDs $\tilde{F}(\underline{m}, b)$ and $F(\underline{m}, b)$ using a suitable weighting function $w(b)$ is given by*

$$D = \int_{\mathbb{R}^+} w(b) \int_{\mathbb{R}^n} \left(\tilde{F}(\underline{m}, b) - F(\underline{m}, b)\right)^2 \, \mathrm{d}\underline{m} \, \mathrm{d}b \ . \quad (2)$$

This distance measure can now be used directly for approximation of probability distributions. The use of LCDs makes a direct comparison between discrete and continuous distributions possible, which is infeasible for information theoretic measures, such as the Kulback-Leibler divergence [20], or the Rényi divergence [21].

IV. Approximation of Gaussian Mixtures

In this section, the distance measure (2) is derived for the case of comparing arbitrary Gaussian mixtures with an arbitrary discrete probability density defined on the same domain. We show that our choice of the Kernel function $K(\cdot, \cdot)$ results in a distance measure, which can be evaluated without use of multidimensional integration. Furthermore, we provide a similar derivation for the gradient of the proposed distance measure.

First, we take a look at the LCDs of the considered densities. The derivation of the LCD for a discrete probability distribution is straightforward. For the equally weighted case, it was already given in [13] by

$$F_D(\underline{m}, b, \underline{\eta}) = \frac{1}{L}\sum_{i=1}^{L} \exp\left(-\frac{1}{2}\frac{||\underline{x}_i - \underline{m}||_2^2}{b^2}\right) \ ,$$

where $\underline{x}_i$ denote the components of the considered discrete density.

The LCD for the Gaussian mixture involves multidimensional integration. It is given as

$$F_{GM}(\underline{m}, b) = \int_{R^n} f_{GM}(\underline{x}) \cdot K(\underline{x} - \underline{m}, b) \, \mathrm{d}\underline{x} \ .$$

In the following lemma, we obtain a more convenient representation for this LCD.

Lemma 1. *The LCD of a Gaussian mixture $f_{GM}(\underline{x})$ as defined above is given by*

$$F_{GM}(\underline{m}, b) = \sum_{i=1}^{M} (2\pi)^{n/2} b^n \, w_i \, \mathcal{N}(\underline{\mu}_i - \underline{m}, \mathbf{C}_i + \tilde{\mathbf{C}}_b) \ .$$

A proof is given in Appendix A.

Now, we can perform the computation of the actual distance measure

$$D(\underline{\eta}) = \int_{\mathbb{R}^+} w(b) \int_{\mathbb{R}^n} \left(F_{GM}(\underline{m}, b) - F_D(\underline{m}, b, \underline{\eta})\right)^2 \, \mathrm{d}\underline{m} \, \mathrm{d}b \ .$$

This, is where we extend our earlier work. A naïve approach would be avoiding this computation by performing componentwise approximation. As will be shown in the evaluation, this unfortunately yields suboptimal results. Thus, the following lemma is motivated by the need for a direct approximation in these cases. It is a generalization of Theorem III.1 from [13], where a similar result was formulated for an axis-aligned Gaussian.

Lemma 2. *Consider the LCD of a Gaussian mixture $F_{GM}(\underline{m}, b)$, the LCD of an equally weighted discrete probability distribution (Dirac mixture) $F_D(\underline{m}, b)$, and the weighting function*

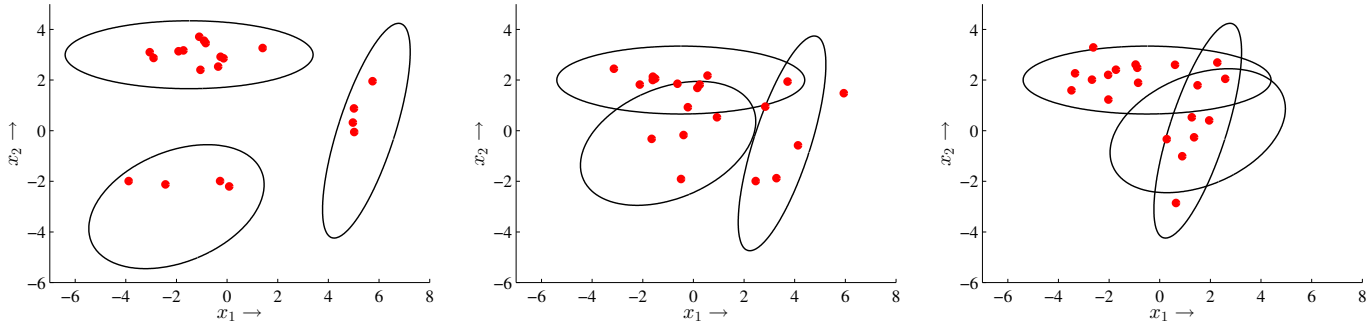
$$w(b) = \begin{cases} b^{n-1} & , \ \textit{if } 0 \leq b \leq b_{max} \\ 0 & , \ \textit{otherwise} \end{cases} \ .$$

Then, the corresponding modified Cramér-von Mises criterion D can be computed by

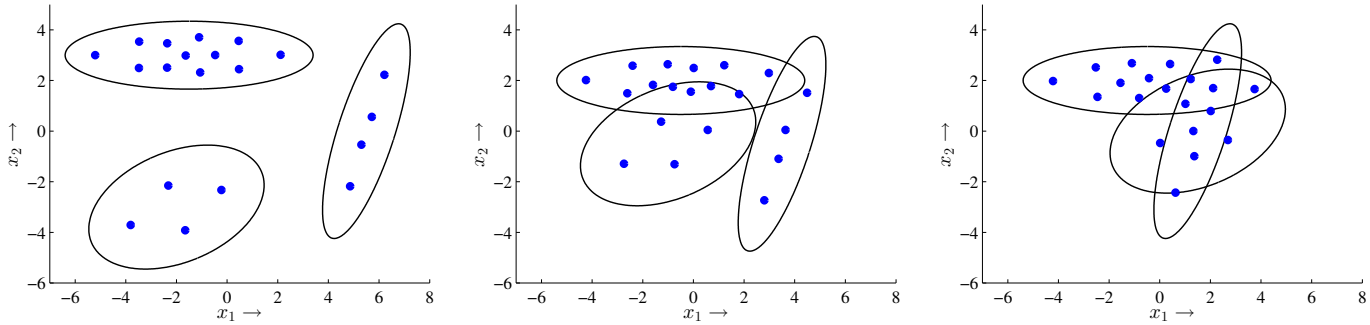
$$D(\underline{\eta}) = \int_0^{b_{max}} \left(P_1 - 2P_2(\underline{\eta}) + P_3(\underline{\eta})\right) \, \mathrm{d}b \ , \quad (3)$$

where

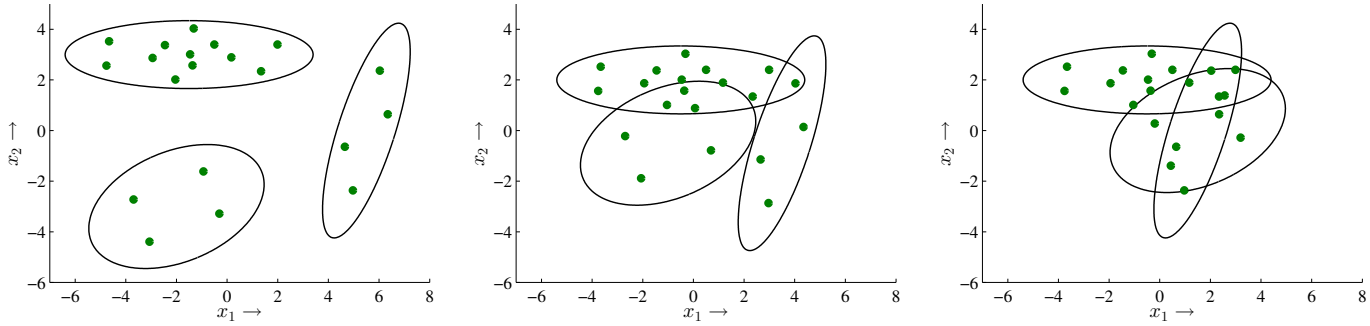
$$\begin{aligned} P_1 =& (2\pi)^n b^{n+1} \\ & \cdot \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \, \mathcal{N}(\underline{m}_i - \underline{m}_j, \mathbf{C}_i + 2\tilde{\mathbf{C}}_b + \mathbf{C}_j) \ , \\ P_2(\underline{\eta}) =& (2\pi)^n b^{n+1} \sum_{i=1}^{M}\sum_{j=1}^{L} \frac{w_i}{L} \mathcal{N}(\underline{\mu}_i - \underline{x}_j, \mathbf{C}_i + 2\tilde{\mathbf{C}}_b) \ , \\ P_3(\underline{\eta}) =& \frac{\pi^{n/2} b}{L^2} \sum_{i=1}^{L}\sum_{j=1}^{L} \exp\left(-\frac{1}{2}\frac{\left|\left|\underline{x}_i - \underline{x}_j\right|\right|_2^2}{2b^2}\right) \ . \end{aligned}$$



(a) Initial samples for the Optimization procedure obtained from sampling each Gaussian component individually using random sampling.



(b) Optimized samples obtained as a result of the Optimization procedure proposed in this paper.



(c) Suboptimal samples obtained from sampling each Gaussian component individually using the LCD approach.

Figure 2. Initial samples (red), optimized samples (blue), and suboptimal samples (green) for approximation of three different Gaussian mixtures (represented by the covariance ellipsoids of their components).

A proof is given in Appendix B.

Computation of the distance measure given in the previous lemma still involves an integral (over the kernel width b), which needs to be evaluated numerically. Even simplifications as discussed in [14] make use of numerical algorithms or approximations to compute the exponential integral.

This results in a high computational burden, when this distance measure is used as an optimality measure for approximating a Gaussian mixture by a Dirac mixture, because numerical integration happens in every iteration of the optimizer. Thus, it is of particular interest to speed-up the numerical optimization procedure, by reducing the number of required iterations. This can be done by providing a gradient of the proposed distance measure. The following lemma, once again, generalizes our earlier results from [13].

Lemma 3. *The derivative* $\frac{\partial D}{\partial x_i^{(j)}}$ *is given as*

$$G_{j,i} = -2G^a_{j,i} + G^b_{j,i} \ ,$$

where

$$G^a_{j,i} = \int_0^{b_{max}} \sum_{m=1}^{M} \frac{(2\pi)^n \, w_m}{L} \sum_{k=1}^{n} F^{b,m}_{j,k} (\mu_m^{(k)} - x_i^{(k)}) \cdot b^{n+1} \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b) \, \mathrm{d}b \ ,$$

and

$$G^b_{j,i} = \frac{-\pi^{n/2}}{L^2} \sum_{k=1}^{L} (x_i^{(j)} - x_k^{(j)}) \cdot \int_0^{b_{max}} \frac{1}{b} \exp\left(-\frac{1}{2} \frac{||\underline{x}_i - \underline{x}_k||_2^2}{2b^2} \right) \mathrm{d}b \ ,$$

where $F^{b,m}_{j,k}$ *is the entry in the* j*-th row and* k*-th column of the matrix* $(\mathbf{C}_m + 2\tilde{\mathbf{C}}_b)^{-1}$.

A proof is given in Appendix B3.

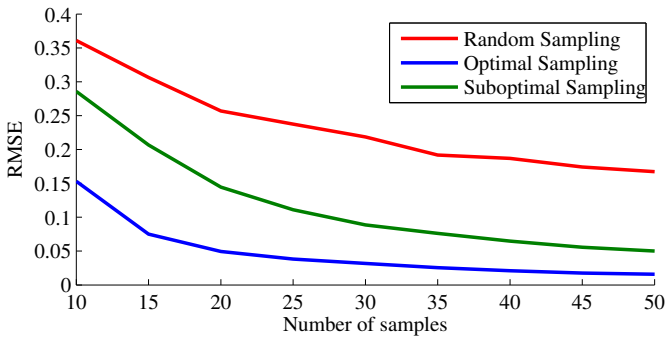


Figure 3. Evaluation of propagation through a nonlinear system function.

The optimization problem for finding an optimal approximation of Gaussian mixtures still suffers from strong nonlinearity involving local minima, because of the multiple modes in the Gaussian mixture. Consequently, a good choice of starting values is of particular importance. However, it is sufficient to ensure, that each Gaussian component is assigned a number of samples corresponding to its weight. Thus, an approximation of a Gaussian mixture with M components by a Dirac mixture with L components can be performed in three steps. First, each component of the Gaussian mixture is assigned a number L_i of Dirac components ($i = 1, \ldots, M$ and $L = L_1 + \ldots + L_M$) proportional to its probability weight w_i. Second, we sample L_i random samples from each Gaussian mixture component i. Finally, an optimizer is used minimizing (3) in order to obtain an approximation of the Gaussian mixture.

V. Evaluation

It is of particular interest to compare the proposed method to simply approximating each component of the Gaussian individually or using random sampling. In Fig. 2, we show three types of approximations for three different Gaussian mixtures. The first approximation is based on a random approximation of each component individually. The number of random samples is chosen according to the weight of each component. The second approximation uses the proposed method (in which the random samples serve as a starting value and $b_{max} = 10$). Finally, the third approximation uses a suboptimal approach, where each Gaussian mixture component is approximated individually. This is done by using a precomputed Dirac mixture approximation of the standard Gaussian, which is transformed (by using the matrix square-root of each components covariance) in order to match the covariance of each Gaussian component. The presented results follow the expectation, that Gaussian mixture components which do not have a strong overlap with other components are treated as if they were individual Gaussians. For the actual optimization step computing the placement of the Dirac delta components, we used the `fminunc` procedure from Matlab 2014a with a quasi-Newton algorithm and a termination tolerance on the function value of $1e-12$. That is, no moment constraints were imposed on the resulting approximation.

Probably the most common application of the proposed approximation is the approximate propagation of uncertainty through nonlinear transformations. Thus, it is of interest to investigate the proposed method in an uncertainty propagation scenario involving overlapping and closely located Gaussian mixture components. In our scenario, a two-dimensional Gaussian mixture will be propagated through the nonlinear function

$$g(\underline{x}) = x^{(1)} \cdot \cos(x^{(2)}) \ .$$

This example is motivated by the general fact that consideration of trigonometric functions is of particular interest in different scenarios (e.g., when transforming uncertain positions between different types of coordinate systems). The considered Gaussian mixture consists of three equally-weighted components with respective means $\underline{\mu}_1 = (1.5, 0)^T$, $\underline{\mu}_2 = (1.5, 0)^T$, and $\underline{\mu}_3 = (-0.5, 2)^T$. The covariance matrices in each component were given as

$$\mathbf{C}_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix} \ , \ \mathbf{C}_2 = \begin{pmatrix} 0.5 & 0.9 \\ 0.9 & 3 \end{pmatrix} \ , \ \mathbf{C}_3 = \begin{pmatrix} 4 & 0 \\ 0 & 0.3 \end{pmatrix} \ .$$

Furthermore, we used different initialization values for the actual optimization procedure in order to account for possible local minima. The suboptimal approximation discussed above and random sampling the Gaussian mixture were used for comparison. The entire simulation was performed using 1000 runs. Ground truth was obtained by random sampling using 10^8 samples. The results are shown in Fig. 3 for a different number of approximated / sampled Dirac delta components. As expected, random sampling has the worst performance. The proposed approximation yields superior results over a suboptimal componentwise approach. This result comes at the price of higher computational complexity. However, particularly for shape approximation involving Gaussian mixtures with strong overlap, the proposed approach yields a significantly better outcome.

VI. Conclusion

In this paper, we proposed a shape approximation for Gaussian mixture distributions by a Dirac mixture. This generalizes our earlier work [13] by providing a homogeneous coverage of the considered Gaussian mixture densities. The proposed method is superior over existing component-wise deterministic approximations, because these approaches yield poor results for strongly overlapping Gaussian mixture components.

Our method is based on deriving a distance measure ensuring shape approximation of the underlying Gaussian mixture distribution. The resulting method is of particular interest for a wide scope of applications, because Gaussian mixtures can be used to approximate other continuous distributions. Thus, whenever an efficient approximation of other densities by Gaussian mixtures is possible, this Gaussian mixture approximation can be used as an intermediate step in approximating the original distribution by a Dirac mixture. Furthermore, Gaussian mixtures arise naturally in stochastic filtering scenarios involving a discrete prior distribution and additive system noise.

The choice of an approximation technique in scenarios involving Gaussian mixtures is highly dependent on precision and performance requirements. A naïve approach approximating each mixture component using the unscented transform might be sufficient for some applications. Higher precision can be achieved by precomputing the proposed approximation for a standard Gaussian and applying the Mahalanobis transform for each component individually. For filtering scenarios where the noise term is distributed according to the same Gaussian mixture in each time step, the entire approximation of the Gaussian mixture can also be computed in advance. Componentwise approximations perform particularly bad in scenarios involving

strong overlap of the involved mixture components which serves as an additional motivation for using the proposed approach.

Thus, our future work will involve investigating more efficient approximation and optimization procedures and an in-depth analysis of different types of possible kernels. Furthermore, we are interested in deriving analytical results on approximation quality with respect to the number of involved Dirac delta components.

ACKNOWLEDGMENT

This work was jointly supported by the German Research Foundation (DFG) and the Czech Science Foundation (project no. P103-13-07058J) within the Collaborative Research Project "*Consistent Fusion in Networked Estimation Systems*", and the DFG Research Training Group RTG 1194 "*Self-organizing Sensor-Actuator-Networks*".

APPENDIX

A. Proof of Lemma 1

We can compute directly.

$$\begin{aligned} F_{GM}(\underline{m},b) &= \int_{\mathbb{R}^n} f_{GM}(\underline{x}) \cdot K(\underline{x}-\underline{m},b)\ \mathrm{d}\underline{x} \\ &= \sum_{i=0}^{M} \omega_i \int_{\mathbb{R}^n} f(\underline{x},\underline{\mu}_i,\mathbf{C}_i) \cdot K(\underline{x}-\underline{m},b)\ \mathrm{d}\underline{x} \\ &= \sum_{i=0}^{M} \omega_i \int_{\mathbb{R}^n} \mathcal{N}(\underline{x}-\underline{\mu}_i,\mathbf{C}_i) \\ &\qquad \cdot a_b\, \mathcal{N}(\underline{x}-\underline{m},\tilde{\mathbf{C}}_b)\ \mathrm{d}\underline{x}\ , \end{aligned}$$

where $a_b = \sqrt{det(2\pi\tilde{\mathbf{C}}_b)} = (2\pi)^{n/2}\, b^n$. Now, we can use the fact that the product of two Gaussian densities is itself a (rescaled) Gaussian density.

$$\begin{aligned} F_{GM}(\underline{m},b) &= \sum_{i=0}^{M} \omega_i \int_{\mathbb{R}^n} c_{i,b}\, \mathcal{N}(\underline{x}-\underline{\mu}_{i,b},\hat{\mathbf{C}}_{i,b})\ \mathrm{d}\underline{x} \\ &= \sum_{i=0}^{M} \omega_i\, c_{i,b}\ , \end{aligned}$$

where $\hat{\mathbf{C}}_{i,b} = (\mathbf{C}_i^{-1}+\mathbf{C}_b^{-1})^{-1}$, $\underline{\mu}_{i,b} = \hat{\mathbf{C}}_{i,b}(\mathbf{C}_i^{-1}\underline{\mu}_i + \mathbf{C}_b^{-1}\underline{m})$, and $c_{i,b} = a_b\, \mathcal{N}(\underline{\mu}_i - \underline{m}, \mathbf{C}_i + \mathbf{C}_b)$. □

B. Proof of Lemma 2

1) Computation of P_1:

$$\begin{aligned} P_1 &= \int_{\mathbb{R}^n} F_{GM}(\underline{m},b)^2\ \mathrm{d}\underline{m} \\ &= (2\pi)^n b^{2n} \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \int_{\mathbb{R}^n} \mathcal{N}(\underline{\mu}_i-\underline{m},\mathbf{C}_i+\tilde{\mathbf{C}}_b) \\ &\qquad \cdot \mathcal{N}(\underline{\mu}_j-\underline{m},\mathbf{C}_j+\tilde{\mathbf{C}}_b)\ \mathrm{d}\underline{m} \end{aligned}$$

Once again, we make use of the fact that multiplication of Gaussian densities yields a rescaled Gaussian density. This results into

$$P_1 = (2\pi)^n b^{2n} \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j\, \mathcal{N}(\underline{\mu}_i-\underline{\mu}_j,\mathbf{C}_i+2\tilde{\mathbf{C}}_b+\mathbf{C}_j)\ .$$

2) Computation of $P_2(\underline{\eta})$: First, we note that the LCD of an equal weighted discrete probability distribution can be rewritten as

$$F_D(\underline{m},b,\underline{\eta}) = \frac{1}{L}\sum_{i=1}^{L}(2\pi)^{n/2} b^n \mathcal{N}(\underline{x}_i-\underline{m},\tilde{\mathbf{C}}_b)\ .$$

Now, we compute directly

$$\begin{aligned} P_2(\underline{\eta}) &= \int_{\mathbb{R}^n} F_{GM}(\underline{m},b) F_D(\underline{m},b)\ \mathrm{d}\underline{m} \\ &= (2\pi)^n b^{2n} \sum_{i=1}^{M}\sum_{j=1}^{L} \frac{w_i}{L} \int_{\mathbb{R}^n} \mathcal{N}(\underline{\mu}_i-\underline{m},\mathbf{C}_i+\tilde{\mathbf{C}}_b) \\ &\qquad \cdot \mathcal{N}(\underline{x}_j-\underline{m},\tilde{\mathbf{C}}_b)\ \mathrm{d}\underline{m}\ . \end{aligned}$$

After applying the same argument as in the computation of P_1, we obtain

$$P_2(\underline{\eta}) = (2\pi)^n b^{2n} \sum_{i=1}^{M}\sum_{j=1}^{L} \frac{w_i}{L} \mathcal{N}(\underline{\mu}_i-\underline{x}_j,\mathbf{C}_i+2\tilde{\mathbf{C}}_b)\ .$$

3) Computation of $P_3(\underline{\eta})$: $P_3(\underline{\eta})$ does not depend on the LCD of the Gaussian Mixture. Thus, it is the same as in our previous work. Its computation is derived in the same way as in [13]. □

First, we note that no $\underline{x}_i$ appears in D_1. Thus, we have $\delta D_1/\delta\underline{x}_i^{(j)} = 0$. Thus $G_{i,j}^a$ and $G_{i,j}^b$ represent the derivatives $\delta - 2D_2/\delta x_i^{(j)}$ and $\delta D_3/\delta x_i^{(j)}$ respectively.

4) Computation of $G_{j,i}^a$:

$$\begin{aligned} G_{j,i}^a &= \frac{\delta}{\delta x_i^{(j)}} \int_{\mathbb{R}^+} w(b)\,(2\pi)^n b^{2n} \\ &\qquad \cdot \sum_{m=1}^{M}\sum_{l=1}^{L} \frac{w_m}{L} \mathcal{N}(\underline{\mu}_m-\underline{x}_l,\mathbf{C}_m+2\tilde{\mathbf{C}}_b)\ \mathrm{d}b \\ &= \sum_{m=1}^{M}\sum_{l=1}^{L} \frac{(2\pi)^n\, w_m}{L} \\ &\qquad \cdot \int_0^{b_{max}} b^{n+1} \frac{\delta\mathcal{N}(\underline{\mu}_m-\underline{x}_l,\mathbf{C}_m+2\tilde{\mathbf{C}}_b)}{\delta x_i^{(j)}}\ \mathrm{d}b \end{aligned}$$

Differentiation removes all terms, where $l \neq i$, and thus we have

$$\begin{aligned} G_{j,i}^a &= \sum_{m=1}^{M} \frac{(2\pi)^n\, w_m}{L} \\ &\qquad \cdot \int_0^{b_{max}} b^{n+1} \underbrace{\frac{\delta\mathcal{N}(\underline{\mu}_m-\underline{x}_i,\mathbf{C}_m+2\tilde{\mathbf{C}}_b)}{\delta x_i^{(j)}}}_{=:A_{j,i}^{b,m}}\ \mathrm{d}b\ . \end{aligned}$$

The last equality is due to the fact that $x_i^{(j)}$ does not appear for the summands where $l \neq i$. In order to compute the derivative of the Gaussian density $A_{j,i}^{b,m}$, we make use of the fact that the

covariance matrix (and its inverse) is symmetric. Furthermore, we define $\mathbf{F}^{b,m} = (\mathbf{C}_m + 2\tilde{\mathbf{C}}_b)^{-1}$. This yields

$$\begin{aligned} A_{j,i}^{b,m} =& \frac{\delta \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b)}{\delta x_i^{(j)}} \\ =& -\frac{1}{2} \cdot \left(\frac{\partial \mathrm{tr}\left((\underline{\mu}_m - \underline{x}_i)^T \mathbf{F}^{b,m} (\underline{\mu}_m - \underline{x}_i) \right)}{\partial x_i^{(j)}} \right) \\ & \cdot \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\mathbf{C}_b) \ . \end{aligned}$$

Now, we can use the fact that the trace is invariant under cyclic permutations. This gives us

$$\begin{aligned} A_{j,i}^{b,m} =& -\frac{1}{2} \cdot \left(\frac{\partial \mathrm{tr}\left(\mathbf{F}^{b,m} (\underline{\mu}_m - \underline{x}_i)(\underline{\mu}_m - \underline{x}_i)^T \right)}{\partial x_i^{(j)}} \right) \\ & \cdot \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b) \\ =& -\frac{1}{2} \cdot \left(\frac{\partial \left(\sum_{p,k=1}^{n} F_{p,k}^{b,m} (\underline{\mu}_m^{(k)} - \underline{x}_i^{(k)})(\underline{\mu}_m^{(p)} - \underline{x}_i^{(p)}) \right)}{\partial x_i^{(j)}} \right) \\ & \cdot \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b) \ , \end{aligned}$$

where $F_{p,k}^{b,m}$ denotes the entry in the p-th row and k-th column of $\mathbf{F}^{b,m}$. Performing the derivation and taking symmetry of $\mathbf{F}^{b,m}$ into account yields

$$\begin{aligned} A_{j,i}^{b,m} =& \sum_{k=1}^{n} F_{j,k}^{b,m} (\mu_m^{(k)} - x_i^{(k)}) \\ & \cdot \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b) \ . \end{aligned}$$

Using this result, we finally obtain

$$\begin{aligned} G_{j,i}^{a} = \int_0^{b_{max}} & \sum_{m=1}^{M} \frac{(2\pi)^n \, w_m}{L} \sum_{k=1}^{n} F_{j,k}^{b,m} (\mu_m^{(k)} - x_i^{(k)}) \\ & \cdot b^{n+1} \mathcal{N}(\underline{\mu}_m - \underline{x}_i, \mathbf{C}_m + 2\tilde{\mathbf{C}}_b) \ \mathrm{d}b \ . \end{aligned}$$

5) Computation of $G_{j,i}^{b}$: The value of $G_{j,i}^{b}$ depends only on the LCD of the discrete distribution. Thus, it is the same as in our earlier results on approximating axis-aligned or isotropic Gaussians. A proof is given in [13]. □

REFERENCES

[1] 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, Jul. 2000.

[2] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174 –188, 2002.

[3] S. Julier and J. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401 – 422, 2004.

[4] M. F. Huber and U. D. Hanebeck, "Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation," in *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, vol. 17, no. 2, Seoul, Republic of Korea, Jul. 2008.

[5] F. Beutler, M. F. Huber, and U. D. Hanebeck, "Gaussian Filtering using State Decomposition Methods," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, USA, Jul. 2009.

[6] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[7] I. Gilitschenski, G. Kurz, S. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *arXiv preprint: Systems and Control (cs.SY)*, 2013.

[8] O. Straka, J. Dunik, and M. Simandl, "Randomized Unscented Kalman filter in Target Tracking," in *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*, 2012, pp. 503–510.

[9] U. D. Hanebeck and V. Klumpp, "Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance," in *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, Seoul, Republic of Korea, Aug. 2008, pp. 33–39.

[10] K. Ito and K. Xiong, "Gaussian Filters for Nonlinear Filtering Problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[11] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Springer, 2000.

[12] G. Pages and J. Printems, "Optimal Quadratic Quantization for Numerics: the Gaussian Case," *Monte Carlo Methods and Applications*, vol. 9, no. 2, 2003.

[13] U. D. Hanebeck, M. F. Huber, and V. Klumpp, "Dirac Mixture Approximation of Multivariate Gaussian Densities," in *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, Dec. 2009.

[14] I. Gilitschenski and U. D. Hanebeck, "Efficient Deterministic Dirac Mixture Approximation," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[15] U. D. Hanebeck, "Kernel-based Deterministic Blue-noise Sampling of Arbitrary Probability Density Functions," in *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS 2014)*, Princeton, New Jersey, USA, Mar. 2014.

[16] J. Steinbring and U. D. Hanebeck, "S2KF: The Smart Sampling Kalman Filter," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.

[17] U. D. Hanebeck, "PGF 42: Progressive Gaussian Filtering with a Twist," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.

[18] V. Maz'ya, "On Approximate Approximations Using Gaussian Kernels," *IMA Journal of Numerical Analysis*, vol. 16, no. 1, pp. 13–29, Jan. 1996.

[19] T. W. Anderson, "On the Distribution of the Two-Sample Cramer-Von Mises Criterion," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1148–1159, 1962.

[20] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

[21] A. Rényi, "On Measures of Entropy and Information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.