

Kernel-based Deterministic Blue-Noise Sampling of Arbitrary Probability Density Functions

Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
 Karlsruhe Institute of Technology (KIT), Germany.
 Email: uwe.hanebeck@ieee.org

Abstract—This paper provides an efficient method for approximating a given continuous probability density function (pdf) by a Dirac mixture density. Optimal parameters are determined by systematically minimizing a distance measure. As standard distance measures are typically not well defined for discrete densities on continuous domains, we focus on shifting the mass distribution of the approximating density as close to the true density as possible. Instead of globally comparing the masses as in a previous paper, the key idea is to characterize individual Dirac components by kernel functions representing the spread of probability mass that is appropriate at a given location. A distance measure is then obtained by comparing the deviation between the true density and the induced kernel density. This new method for Dirac mixture approximation provides high-quality approximation results, can handle arbitrary pdfs, allows considering constraints for, e.g., maintaining certain moments, and is fast enough for online processing.

Keywords—Dirac mixture approximation, sampling, statistical distance measure, moment problem, progressive processing, homotopy continuation.

I. INTRODUCTION

This paper provides an efficient method for approximating a given continuous probability density function (pdf) by a deterministic Dirac mixture density. The resulting Dirac mixture density is then used to simplify the typical operations in system identification, state estimation, filtering, data fusion, and control such as transforming random variables with nonlinear mappings or performing numerical integration for calculating nonlinear moments. State-of-the-art Dirac mixture approximations are currently used in diverse nonlinear filters such as Gaussian filters, e.g., the S2KF [1] and PGF42 [2], or in particle-based model-predictive control.

The problem is to find appropriate parameters of the approximating discrete density (the weights and locations of the Dirac components) so that it is in some way close to the true density. Various options for measuring closeness are available such as comparing certain moments. Direct measures for shape differences, however, are usually not applicable as they are not well-defined for discrete densities on continuous domains. In this paper, we focus on comparing probability masses between the two densities with the goal of shifting the mass distribution of the approximating density as close to the true density as possible.

The key idea is to characterize individual Dirac components by kernel functions representing the spread of probability mass that is appropriate at a given location. For reducing computational complexity, only the true densities at the Dirac locations themselves are considered for calculating the kernel function parameters leading to a heteroscedastic but isotropic induced kernel density. A distance measure is then obtained by comparing the deviation between the true density and the induced kernel density. Here, a squared integral measure is

used for that purpose. Finally, the weights and locations are optimized in such a way as to minimize this distance.

The required integrals can be solved in closed form for certain true densities such as Gaussian densities, Gaussian mixture densities, or rectangular densities in combination with adequate kernels, e.g., Gaussian or rectangular kernels. However, for arbitrary true densities, an analytic solution is not possible. To be able to come up with efficient solutions for arbitrary densities and Gaussian kernels, two integration schemes are used that are themselves based on Dirac mixture approximations. Discretization is either achieved by replacing the true density or the individual Gaussian kernel functions by Dirac mixture approximations that have been calculated using the closed-form approach.

This new method for Dirac mixture approximation provides high-quality approximation results, can handle arbitrary pdfs, which also includes the important case of products of densities, allows considering constraints for, e.g., maintaining certain moments, and is fast enough for online processing. In this paper, we will focus on generating Dirac mixture approximations for Gaussian densities.

II. PROBLEM FORMULATION

We consider a random vector $\underline{x} = [x_1, x_2, \dots, x_N]^T$ with realizations $\underline{x} \in \mathbb{R}^N$ characterized by a given *continuous* pdf $\tilde{f}(\underline{x})$. Our goal is to approximate $\tilde{f}(\underline{x})$ by a discrete pdf $f(\underline{x})$ on the continuous domain \mathbb{R}^N . Here, we use a so called Dirac mixture density $f(\underline{x})$ with L Dirac components given by

$$f(\underline{x}) = \sum_{i=1}^L w_i \cdot \delta(\underline{x} - \hat{\underline{x}}_i) , \quad (1)$$

with positive weights, i.e., $w_i > 0$ for $i = 1, \dots, L$, that sum up to one, i.e., $\sum_{i=1}^L w_i = 1$, and locations $\hat{\underline{x}}_i$ with components \hat{x}_{ki} for dimension k with $k = 1, \dots, N$. The locations are collected in a matrix $\hat{\mathbf{X}} = [\hat{\underline{x}}_1, \hat{\underline{x}}_2, \dots, \hat{\underline{x}}_L] \in \mathbb{R}^{N \times L}$.

Our goal is to systematically find a Dirac mixture density $f(\underline{x})$ in (1) that is in some sense as close as possible to the given density $\tilde{f}(\underline{x})$ by adjusting its parameters, i.e., its weights w_i and its locations $\hat{\underline{x}}_i$ for $i = 1, \dots, L$. In this paper, we assume solution existence and focus on adjusting the locations only. We define a parameter vector $\underline{\eta} \in \mathcal{S} = \mathbb{R}^{LN}$ containing the parameters as

$$\underline{\eta} = [\hat{\underline{x}}_1^T, \hat{\underline{x}}_2^T, \dots, \hat{\underline{x}}_L^T]^T \quad (2)$$

and write $f(\underline{x}) = f(\underline{x}, \underline{\eta})$. Of course, $\underline{\eta}$ can be obtained as $\underline{\eta} = \hat{\mathbf{X}}(\cdot)$.

For assuring closeness in a stochastic sense, the given density $\tilde{f}(\underline{x})$ and its Dirac mixture approximation $f(\underline{x}, \underline{\eta})$ should have certain moments in common, e.g., have the same mean and first central moments. This results in a special type of truncated moment problem: Given certain moments of the given density $\tilde{f}(\underline{x})$, these moments are to be matched by the corresponding moments of its Dirac mixture approximation $f(\underline{x}, \underline{\eta})$ by selecting a suitable parameter vector $\underline{\eta}$.

When the parameters of $f(\underline{x}, \underline{\eta})$ are redundant in the sense that the length LN of the parameter vector $\underline{\eta}$ is larger than the number of moment constraints, a regularizer for $f(\underline{x}, \underline{\eta})$ is required. Regularization could be performed by selecting the least informative Dirac mixture, e.g., the one having maximum entropy. In this paper, we go one step further and use redundancy to minimize a distance measure between the given density $\tilde{f}(\underline{x})$ and its Dirac mixture approximation $f(\underline{x}, \underline{\eta})$. This results in a constrained optimization problem, where the most regular Dirac mixture approximation $f(\underline{x}, \underline{\eta})$ is desired that fulfills the moment constraints.

III. STATE OF THE ART

Random sampling is the simplest and fastest method to discretize a given probability density function. However, as samples are produced independently, the convergence of its statistics, such as the moments, to the true values is slow so that lots of samples are required. In addition, the given density is not homogeneously covered.

For generating deterministic samples, moment-based approximations of Gaussian densities have been proposed in the context of the Linear Regression Kalman Filter (LRKF), see [3]. Examples are the Unscented Kalman Filter (UKF) in [4] and its scaled version in [5], its higher-order generalization in [6], and a generalization to an arbitrary number of deterministic samples placed along the coordinate axes introduced in [7]. Methods for deterministic Dirac mixture approximation based on distance measures have been proposed for the case of scalar continuous densities [8], [9]. An algorithm for sequentially increasing the number of components is given in [10] and applied to recursive nonlinear prediction in [11]. Systematic Dirac mixture approximations of arbitrary multi-dimensional Gaussian densities are calculated in [12]. A more efficient method for the case of standard normal distributions with a subsequent transformation is given in [13].

For circular probability density functions, a first approach to Dirac mixture approximation in the vein of the UKF is introduced in [14] for the von Mises distribution and the wrapped Normal distribution. Three components are systematically placed based on matching the first circular moment. This Dirac mixture approximation of continuous circular probability density functions has already been applied to sensor scheduling based on bearings-only measurements [15]. In [16], the results are used to perform recursive circular filtering for tracking an object constrained to an arbitrary one-dimensional manifold.

A combination of random and deterministic sampling can be found in [17].

IV. DIRAC MIXTURE APPROXIMATION

The goal of this section is to systematically find a Dirac mixture approximation of a given density in such a way that a certain distance measure is minimized. Typical distance measures such as the Kullback-Leibler distance or squared integral distances are not well defined [18] for directly comparing continuous densities and Dirac mixture densities, so that

cumulative distributions or localized cumulative distributions [18] are used instead. Here, we pursue a different approach that is similar to the mass-based comparison approach in [12], where we consider masses of both densities under certain kernels for all possible kernel locations and widths, which allows the use of integral measures for the mass functions. Appropriate weighting gives an error spectrum with “blue noise” characteristics [19], i.e., which is zero for low frequencies. In this paper, we use a simplified and computationally more efficient approach inspired by blue-noise sampling, e.g., see [20]. Instead of discs, however, we use kernels for describing the mass of individual Dirac components as in [21].

The key idea is to compare probability masses between the two densities with the goal of shifting the mass distribution of the approximating density as close to the true density as possible. For that purpose, the individual Dirac components are characterized by kernel functions representing the spread of probability mass that is appropriate at a given location. For reducing computational complexity, only the true densities at the Dirac locations themselves are considered for calculating the kernel function parameters. A distance measure is then obtained by comparing the deviation between the true density and the induced kernel density. Here, a squared integral distance is used for that purpose. Finally, the locations are optimized in such a way as to minimize this distance.

A. Induced Kernel Density

For characterizing each Dirac component i located at $\hat{\underline{x}}_i$, we use a Gaussian density normalized to the height of the true density at the location of the Dirac component. For that purpose, we take an axis-aligned *unnormalized* Gaussian density with mean located at $\hat{\underline{x}}_i$ and equal spread in every dimension. Multiplying it with the height of the true density at the Dirac component location $\tilde{f}(\hat{\underline{x}}_i)$ gives the desired kernel functions

$$k_i(\underline{x}) = \tilde{f}(\hat{\underline{x}}_i) \exp\left(-\frac{1}{2} \frac{1}{\tau_i^2} (\underline{x} - \hat{\underline{x}}_i)^T (\underline{x} - \hat{\underline{x}}_i)\right) \quad (3)$$

for $i = 1, \dots, L$, where the τ_i are individual spread parameters yet to be determined.

The spread parameters τ_i in (3) are determined by maintaining the mass in the Dirac component given by w_i , so we obtain

$$\int_{\mathbb{R}^N} k_i(\underline{x}) d\underline{x} = w_i ,$$

which gives $(\sqrt{2\pi} \tau_i)^N \tilde{f}(\hat{\underline{x}}_i) = w_i$ or

$$\tau_i = \frac{1}{\sqrt{2\pi}} \left(\frac{w_i}{\tilde{f}(\hat{\underline{x}}_i)} \right)^{\frac{1}{N}} . \quad (4)$$

The resulting kernels have a larger width in low-density regions of the given continuous density and these regions are characterized with few Dirac components per unit area. In high-density areas, the kernels have smaller widths, which results in more Dirac components per unit area.

Finally, by inserting τ_i from (4) into the respective kernels in (3), we obtain the induced kernel density as

$$k(\underline{x}, \underline{\eta}) = \sum_{i=1}^L k_i(\underline{x}) \quad (5)$$

$$= \sum_{i=1}^L \tilde{f}(\hat{\underline{x}}_i) \exp\left(-\pi \left(\frac{\tilde{f}(\hat{\underline{x}}_i)}{w_i}\right)^{\frac{2}{N}} (\underline{x} - \hat{\underline{x}}_i)^T (\underline{x} - \hat{\underline{x}}_i)\right) .$$

When $k(\underline{x}, \underline{\eta})$ is close to $\tilde{f}(\underline{x})$ on \mathbb{R}^N , the mass distributions of both $\tilde{f}(\underline{x})$ and its Dirac mixture approximation $f(\underline{x}, \underline{\eta})$ are similar and we consider the densities to be close. Hence, $k(\underline{x}, \underline{\eta})$ can be used to define a distance measure $D(\underline{\eta})$ between $\tilde{f}(\underline{x})$ and $f(\underline{x}, \underline{\eta})$ on \mathbb{R}^N .

B. Distance Measure

Given the induced kernel density $k(\underline{x}, \underline{\eta})$ in (5) corresponding to the Dirac mixture approximation $f(\underline{x}, \underline{\eta})$, we now define a distance measure between $\tilde{f}(\underline{x})$ and $f(\underline{x}, \underline{\eta})$ on \mathbb{R}^N as

$$D(\underline{\eta}) = D\left(\tilde{f}(\underline{x}), f(\underline{x}, \underline{\eta})\right) \quad (6)$$

by actually comparing $\tilde{f}(\underline{x})$ and the induced kernel density $k(\underline{x}, \underline{\eta})$.

As specific distance measures, we can now employ the Kullback-Leibler divergence given by

$$D(\underline{\eta}) = \int_{\mathbb{R}^N} \tilde{f}(\underline{x}) \log\left(\frac{\tilde{f}(\underline{x})}{k(\underline{x}, \underline{\eta})}\right) d\underline{x}$$

or the standard squared integral distance measure

$$D(\underline{\eta}) = \int_{\mathbb{R}^N} \left(\tilde{f}(\underline{x}) - k(\underline{x}, \underline{\eta})\right)^2 d\underline{x} . \quad (7)$$

C. Moment Constraints

For defining the moment constraints, we employ a multi-index notation with $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_N)$ containing non-negative integer indexes for every dimension. We define $|\boldsymbol{\kappa}| = \kappa_1 + \kappa_2 + \dots + \kappa_N$ and $\underline{x}^{\boldsymbol{\kappa}} = x_1^{\kappa_1} \cdot x_2^{\kappa_2} \cdot \dots \cdot x_N^{\kappa_N}$. For a scalar c , the expression $\boldsymbol{\kappa} \leq c$ is equivalent to $\kappa_k \leq c$ for $k = 1, 2, \dots, N$. The moments of a random vector \underline{x} with density $f(\underline{x})$ are given by

$$e_{\boldsymbol{\kappa}} = \int_{\mathbb{R}^N} \underline{x}^{\boldsymbol{\kappa}} f(\underline{x}) d\underline{x} \quad (8)$$

for $\boldsymbol{\kappa} \in \mathbb{N}_0^N$. For zero-mean random vectors \underline{x} , the moments coincide with the central moments.

Moments of a certain order m are given by $e_{\boldsymbol{\kappa}}$ for $|\boldsymbol{\kappa}| = m$. We define a multi-dimensional matrix \mathbf{E}_M of moments of up to order M as

$$\mathbf{E}_M(\boldsymbol{\kappa} + 1) = \begin{cases} e_{\boldsymbol{\kappa}} & |\boldsymbol{\kappa}| \leq M \\ 0 & \text{elsewhere} \end{cases} , \quad (9)$$

with $\boldsymbol{\kappa} \leq M$ and $e_{\boldsymbol{\kappa}}$ from (8). The multi-dimensional matrix $\mathbf{E}_M \in \mathbb{R}^{M+1 \times M+1 \times \dots \times M+1}$ is indexed from 1 to $M+1$ in every dimension. An example is given in (15).

D. Optimization

As the final optimization problem, we now desire to minimize the distance measure in (6) while maintaining the moment constraints in (9) of up to order M , i.e.,

$$\underline{\eta}_{\text{opt}} = \arg \min_{\underline{\eta} \in \mathcal{S}} D(\underline{\eta}) \quad \text{s.t.} \quad \mathbf{E}_M = \tilde{\mathbf{E}}_M . \quad (10)$$

V. IMPLEMENTATION ISSUES

For implementing the proposed method for Dirac mixture approximation, we have to select a specific distance measure such as the one in (7), calculate the required integrals, and perform an optimization for obtaining the desired parameters of the Dirac mixture approximation.

A. Variation of the Distance Measure

Instead of using the plain squared integral distance measure in (7), it is often advisable to place more emphasis on regions with a small original density. This is done by weighting the squared integral distance measure with one over the true density according to

$$D(\underline{\eta}) = \int_{\mathbb{R}^N} \left(\tilde{f}(\underline{x}) - k(\underline{x}, \underline{\eta})\right)^2 \frac{1}{\tilde{f}(\underline{x})} d\underline{x} ,$$

which after a simple expansion is seen to be equivalent to

$$D(\underline{\eta}) = \int_{\mathbb{R}^N} \frac{k^2(\underline{x}, \underline{\eta})}{\tilde{f}(\underline{x})} d\underline{x} , \quad (11)$$

where constant terms have been omitted. This is the distance measure we will consider from now on!

B. Performing the Integration

In some interesting cases, the integration in (11) can be performed analytically, which gives the most precise results. This includes the case of given Gaussian densities and Gaussian mixture densities for Gaussian kernels that we assumed for this paper anyway¹.

When analytic integration is not possible, several numerical integration schemes can be applied. Of course, one option is to use a standard integration method, which typically is not efficient as no problem-specific structure is exploited.

The first option for using the problem structure during integration is to discretize the given true density $\tilde{f}(\underline{x})$. When the discretized true density $\tilde{f}_d(\underline{x})$ is close to the true density, we can rewrite (11) as

$$D = \int_{\mathbb{R}^N} \frac{k^2(\underline{x}, \underline{\eta})}{\tilde{f}(\underline{x})} \frac{\tilde{f}_d(\underline{x})}{\tilde{f}(\underline{x})} d\underline{x} .$$

With $\tilde{f}_d(\underline{x})$ a Dirac mixture density with P components given by

$$\tilde{f}_d(\underline{x}) = \sum_{k=1}^P w_k \delta(\underline{x} - \underline{t}_k) ,$$

the integration becomes a summation

$$D = \sum_{k=1}^P w_k \frac{k^2(\underline{t}_k, \underline{\eta})}{\tilde{f}^2(\underline{t}_k)} .$$

Replacing the true density with a discretized version might look like a contradiction, as the goal of this paper is to derive such a discretization. However, a cheaper discretization such as random sampling might be applied. Of course, random sampling is only cheap for certain types of densities such as Gaussian densities, Gaussian mixture densities, or uniform densities over complicated support regions.

The second option for using the problem structure during integration is to discretize the induced kernel density $k(\underline{x}, \underline{\eta})$. Here, some discretization of the standard normal distribution is used and applied to every kernel of the induced kernel density by translation and scaling, which allows to convert the integration to summation similar to the first option. This is the most efficient but also most complicated integration method, as the discretization points depend upon the parameters and, hence, influence the gradient during the optimization process.

¹We even assume axis-aligned Gaussian kernels with equal widths in every dimension.

C. Optimization

The optimization problem (10) is nonlinear and nonconvex with nonlinear equality constraints so that numerical minimization can, depending upon the given starting values, get stuck in local minima. To avoid this problem, we start with a density with a known approximation and use a homotopy continuation method [22] to approach the desired density. For that purpose, we define a progression parameter $\gamma \in [0, 1]$ with $\tilde{f}(\underline{x}, \gamma = 0)$ corresponding to a density with known Dirac mixture approximation and $\tilde{f}(\underline{x}, \gamma = 1)$ corresponding to the given density for which a Dirac mixture approximation is desired.

A generic parametrization of the true density is given by

$$\tilde{f}(\underline{x}, \gamma) = \frac{\tilde{f}^\gamma(\underline{x})}{\int_{\mathbb{R}^N} \tilde{f}^\gamma(\underline{x}) d\underline{x}}. \quad (12)$$

For specific true densities, however, problem-specific and more natural parametrizations can be used as we will see in Sec. VI. We start with $\gamma = 0$ corresponding to the density $\tilde{f}(\underline{x}, \gamma = 0)$ with a known Dirac mixture approximation. Then, we progressively increase the parameter γ and for every γ solve the optimization problem by just correcting the result from the previous step, which ensures tracking the desired solution, see Alg. 1. No predictor is used. As a corrector in line 7, we use the function `fmincon` for constrained optimization in our MATLAB implementation. A maximum amount of iterations is prespecified for `fmincon` and success is reported when the optimum for an increased value of γ can be found within these iterations.

VI. EVALUATION

We will now evaluate the proposed Dirac mixture approximation method by discretizing a variety of Gaussian densities. It is important to note that arbitrary Gaussians (different eigenvalues of the covariance matrix, not aligned with the coordinate axes) are directly approximated. This is different from first transforming an arbitrary Gaussian to a standard normal distribution, performing the Dirac mixture approximation, and transforming the resulting Dirac mixture back. In addition, the mean and the central moments up to third order are explicitly maintained.

To also allow for a visual inspection, the approximation will be performed in two dimensions, i.e., $N = 2$. Without loss of generality, we consider true densities that are zero-mean Gaussian as

$$\tilde{f}(\underline{x}) = \frac{1}{2\pi\sqrt{|\mathbf{C}|}} \exp\left(-\frac{1}{2}\underline{x}^T \mathbf{C}^{-1} \underline{x}\right), \quad (13)$$

with a covariance matrix specified by the standard deviations σ_1, σ_2 of x_1, x_2 , respectively, and the correlation coefficient $\rho \in [-1, 1]$ according to

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}. \quad (14)$$

A closed-form expression for the distance measure in (11) is used in this case.

For progressive processing, we use a more problem-specific parametrization than (12). The covariance matrix in (14) is parametrized as

$$\mathbf{C}(\gamma) = \begin{bmatrix} \sigma_1^2(\gamma) & \gamma \rho \sigma_1(\gamma) \sigma_2(\gamma) \\ \gamma \rho \sigma_1(\gamma) \sigma_2(\gamma) & \sigma_2^2(\gamma) \end{bmatrix},$$

Dirac mixture approximation

Input : Given true density \tilde{f} , number of Dirac components L , number of moments M to maintain

Output: Dirac mixture approximation with parameter vector $\underline{\eta}$ according to (2) and number of progression steps PC

```

// Initialize progression step counter
PC := 0;
1
// Initialize progression parameter  $\gamma$ 
 $\gamma := 0$ ;
2
// Initialize  $\Delta\gamma$ 
 $\Delta\gamma := \epsilon$  (some small positive number);
3
// Parameters for in-/decreasing  $\Delta\gamma$ 
Down := 0.5, Up := 1.5;
4
// Initial parameter vector for  $\gamma = 0$ 
Initialize  $\underline{\eta}$ ;
5
while  $\gamma < 1$  do
6
    // Try correcting values for increased  $\gamma$ 
    [ $\underline{\eta}_{\text{tmp}}$ , success] := Corrector( $\underline{\eta}, \gamma + \Delta\gamma$ );
    7
    if Correction step successful? then
    8
        // Make trial update the temporary estimate
         $\underline{\eta} := \underline{\eta}_{\text{tmp}}$ ;
        9
        // Increment  $\gamma$ 
         $\gamma := \gamma + \Delta\gamma$ ;
        10
        // Increase step size
         $\Delta\gamma := \text{Up} * \Delta\gamma$ ;
        11
        // Increment progression step counter PC
        PC := PC + 1;
        12
    else
    13
        // Decrease step size
         $\Delta\gamma := \text{Down} * \Delta\gamma$ ;
        14
    end
    15
    // Limit  $\gamma$  to [0,1]
    if  $\gamma + \Delta\gamma > 1$  then
    16
        |  $\Delta\gamma := 1 - \gamma$ ;
        17
    end
    18
end
19

```

Algorithm 1: Dirac mixture approximation $f(\underline{x}, \underline{\eta})$ with parameter vector $\underline{\eta}$ of given continuous density $\tilde{f}(\underline{x})$.

with $\sigma_i(\gamma) = (1 - \gamma)(\sigma_1 + \sigma_2)/2 + \gamma \sigma_i$ for $i = 1, 2$. The parametrized true density

$$\tilde{f}(\underline{x}, \gamma) = \frac{1}{2\pi\sigma_1(\gamma)\sigma_2(\gamma)\sqrt{1 - \gamma^2\rho^2}} \exp\left(-\frac{1}{2}\underline{x}^T \mathbf{C}^{-1}(\gamma)\underline{x}\right)$$

now gives an axis-aligned Gaussian with equal widths in both dimensions for $\gamma = 0$ and the original true density in (13) for $\gamma = 1$.

For initializing the progressive optimization (line 5 in Alg. 1), random components are used. An alternative is to use a Dirac mixture approximation for a standard normal distribution calculated offline by, e.g., the method in [13], and stored, which is then scaled to the axis-aligned Gaussian with equal widths in both dimensions for $\gamma = 0$.

The Dirac mixture approximation of (13) should maintain moments up to third order, so we consider a moment matrix \mathbf{E}_3 with

$$\mathbf{E}_3 = \begin{bmatrix} e_{00} & e_{01} & e_{02} & e_{03} \\ e_{10} & e_{11} & e_{12} & 0 \\ e_{20} & e_{21} & 0 & 0 \\ e_{30} & 0 & 0 & 0 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix}. \quad (15)$$

For the considered true density, the parametrized Gaussian density, the normalization constant is one, i.e., $\tilde{e}_{00} = 1$, the means are zero, i.e., $\tilde{e}_{01} = 0$, $\tilde{e}_{10} = 0$, the three second-order moments are given by $\tilde{e}_{20} = \sigma_1^2(\gamma)$, $\tilde{e}_{11} = \gamma \rho \sigma_1(\gamma) \sigma_2(\gamma)$, $\tilde{e}_{02} = \sigma_2^2(\gamma)$, and the four third-order moments are given by $\tilde{e}_{30} = 0$, $\tilde{e}_{21} = 0$, $\tilde{e}_{12} = 0$, $\tilde{e}_{03} = 0$. Hence, the moment matrix of the parametrized true density is given by

$$\tilde{\mathbf{E}}_3(\gamma) = \begin{bmatrix} 1 & 0 & \sigma_2^2(\gamma) & 0 \\ 0 & \gamma \rho \sigma_1(\gamma) \sigma_2(\gamma) & 0 & 0 \\ \sigma_1^2(\gamma) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The resulting Dirac mixture approximations for Gaussian densities with $\sigma_1 = 1$, $\sigma_2 = 1$ and three different values of the correlation coefficient $\rho = 0.0$, $\rho = 0.4$, and $\rho = 0.8$ are given for three different numbers of components $L = 12$, $L = 20$, and $L = 30$ in Fig. 1, Fig. 2, and Fig. 3, respectively.

VII. CONCLUSION

The proposed method reliably provides well-distributed Dirac mixture approximations of arbitrary densities while exactly maintaining a set of predefined moments. The Dirac components are placed deterministically and homogeneously cover the support of the given density.

The resulting Dirac mixture approximations are used to replace continuous densities in order to make certain processing steps feasible that are required in system identification, state estimation, filtering, data fusion, and control. In contrast to random sampling, significantly fewer samples are required and reproducible results are generated.

As Gaussian densities play an important role, e.g., in Gaussian filtering, and many methods for sampling Gaussians have been proposed, we evaluated the proposed Dirac mixture approximation method by approximating Gaussians. Gaussian densities allow for an analytic integration of the distance measure, which speeds up the calculation even more. The resulting Dirac mixture approximations are well-distributed and also visually appealing.

The next step is to thoroughly compare the Dirac mixture approximations of Gaussian densities with the results of the method in [12], which is used in the generalization of the Unscented Kalman Filter (UKF) proposed in [1] and in the nonlinear Gaussian filter proposed in [2]. In these filters, the method in [12] is used to offline generate Dirac mixture approximations for standard normal distributions that are then online transformed to the desired arbitrary Gaussian. Compared to the method in [12], the new method proposed in this paper has three advantages with respect to approximating Gaussians: (i) It is faster and, hence, (ii) allows the online approximation of arbitrary Gaussians, and (iii) explicitly maintains moments up to a prespecified order.

Future generalizations include greedy sequential approximations that add new components until a prespecified approximation quality is achieved without changing existing components. In addition, densities over periodic quantities, i.e., densities on circles, spheres, and tori will be considered.

REFERENCES

- [1] 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.
- [2] 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.
- [3] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "The Linear Regression Kalman Filter," in *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*, ser. Springer Tracts in Advanced Robotics, 2005, vol. 19.
- [4] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.
- [5] S. J. Julier, "The Scaled Unscented Transformation," in *Proceedings of the 2002 IEEE American Control Conference (ACC 2002)*, vol. 6, Anchorage, Alaska, USA, May 2002, pp. 4555–4559.
- [6] D. Tenne and T. Singh, "The Higher Order Unscented Filter," in *Proceedings of the 2003 IEEE American Control Conference (ACC 2003)*, vol. 3, Denver, Colorado, USA, Jun. 2003, pp. 2441–2446.
- [7] 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.
- [8] 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 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, USA, Dec. 2006.
- [9] —, "Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance," in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sep. 2006, pp. 512–517.
- [10] U. D. Hanebeck and O. C. Schrempf, "Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions," in *Proceedings of the 2007 IEEE Conference on Decision and Control (CDC 2007)*, New Orleans, Louisiana, USA, Dec. 2007, pp. 3065–3071.
- [11] O. C. Schrempf and U. D. Hanebeck, "Recursive Prediction of Stochastic Nonlinear Systems Based on Optimal Dirac Mixture Approximations," in *Proceedings of the 2007 American Control Conference (ACC 2007)*, New York, New York, USA, Jul. 2007, pp. 1768–1774.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "Circular Statistics in Bearings-only Sensor Scheduling," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.
- [16] G. Kurz, F. Faion, and U. D. Hanebeck, "Constrained Object Tracking on Compact One-dimensional Manifolds Based on Directional Statistics," in *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)*, Montbeliard, France, Oct. 2013.
- [17] O. Straka, J. Dunik, and M. Simandl, "Randomized Unscented Kalman Filter in Target Tracking," in *15th International Conference on Information Fusion (FUSION 2012)*, 2012, pp. 503–510.
- [18] 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.
- [19] R. Ulichney, "Dithering with Blue Noise," *Proceedings of the IEEE*, vol. 76, no. 1, pp. 56–79, Jan. 1988.
- [20] T. Schlömer, D. Heck, and O. Deussen, "Farthest-Point Optimized Point Sets with Maximized Minimum Distance," in *Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics*, ser. HPG '11. New York, NY, USA: ACM, 2011, p. 135–142. [Online]. Available: <http://doi.acm.org/10.1145/2018323.2018345>
- [21] R. Fattal, "Blue-Noise Point Sampling Using Kernel Density Model," in *ACM SIGGRAPH 2011 papers*, ser. SIGGRAPH '11. New York, NY, USA: ACM, 2011, p. 48:1–48:12. [Online]. Available: <http://doi.acm.org/10.1145/1964921.1964943>
- [22] E. L. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*. SIAM, 2003.

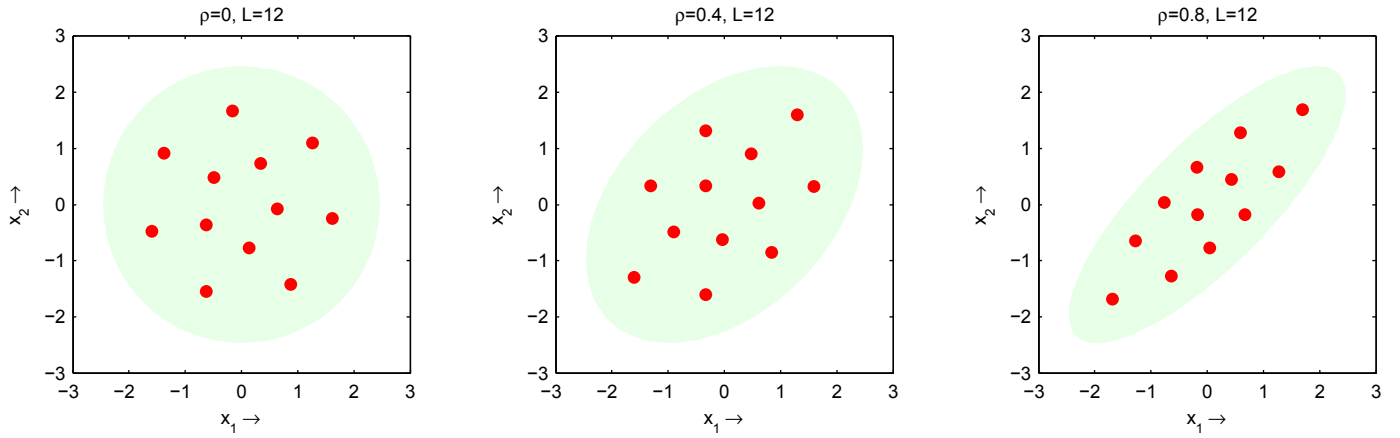


Fig. 1. Dirac mixture approximation of Gaussian density with $L = 12$ components for (left) $\rho = 0.0$, (middle) $\rho = 0.4$, (right) $\rho = 0.8$. All moments up to order three are exactly maintained.

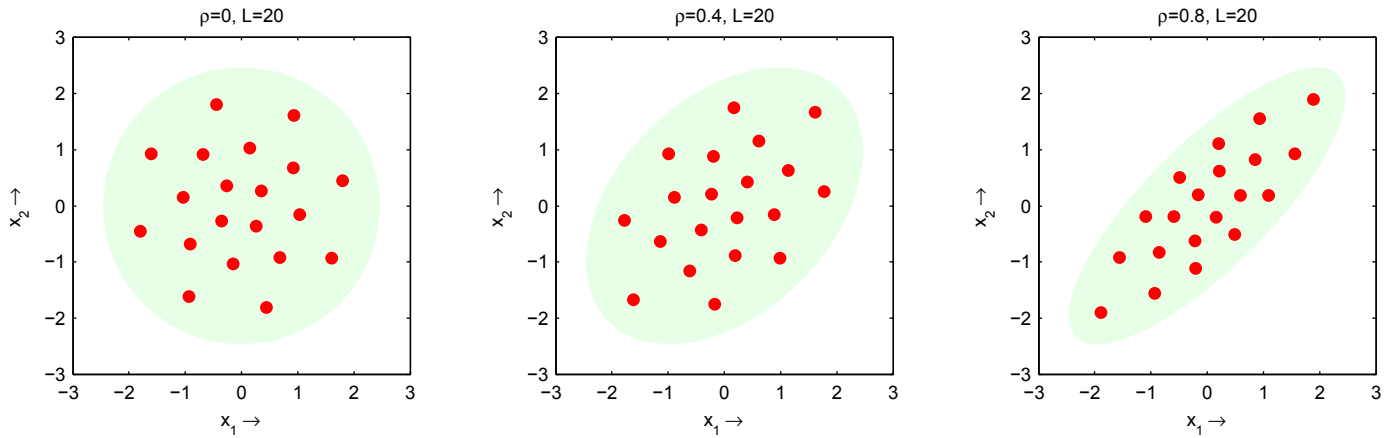


Fig. 2. Dirac mixture approximation of Gaussian density with $L = 20$ components for (left) $\rho = 0.0$, (middle) $\rho = 0.4$, (right) $\rho = 0.8$. All moments up to order three are exactly maintained.

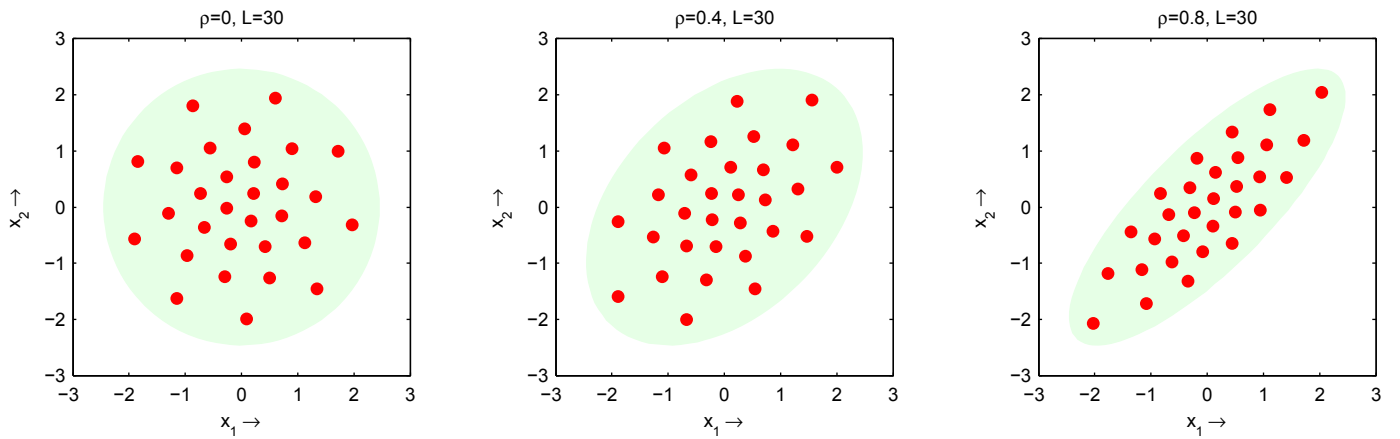


Fig. 3. Dirac mixture approximation of Gaussian density with $L = 30$ components for (left) $\rho = 0.0$, (middle) $\rho = 0.4$, (right) $\rho = 0.8$. All moments up to order three are exactly maintained.