

# Sample Set Design for Nonlinear Kalman Filters Viewed as a Moment Problem

Uwe D. Hanebeck

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

**Abstract**—For designing sample sets for nonlinear Kalman filters, i.e., Linear Regression Kalman Filters (LRKFs), a new method is introduced for approximating Gaussian densities by discrete densities, so called Dirac Mixtures (DMs). This approximating DM should maintain the mean and some higher-order moments and should homogeneously cover the support of the original density. Homogeneous approximations require redundancy, which means there are more Dirac components than necessary for fulfilling the moment constraints. Hence, some sort of regularization is required as the solution is no longer unique. Two types of regularizers are possible: The first type ensures smooth approximations, e.g., in a maximum entropy sense. The second type we pursue here ensures closeness of the approximating density to the given Gaussian. 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. As a result, the approximation problem is converted to an optimization problem as we now minimize the distance under the desired moment constraints.

**Keywords**—*Moment problem, Dirac mixture approximation, sampling, statistical distance measure.*

## I. INTRODUCTION

The goal of this paper is to generate deterministic Dirac mixture densities as approximations for Gaussian densities. These Dirac mixture densities are used to simplify typical operations occurring 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. A particularly interesting application is Gaussian filtering, e.g., the S2KF [1] and the PGF 42 [2].

The parameters of the approximating discrete density (the weights and locations of the Dirac components) are calculated in such a way that mean and some higher-order moments of the true density are maintained by its Dirac mixture approximation. Redundancy resulting from Dirac mixture approximations with more parameters than moment constraints is exploited by selecting the Dirac mixture approximation closest to the true density. An example of a Dirac mixture approximation with 20 equally weighted components that maintains moments up

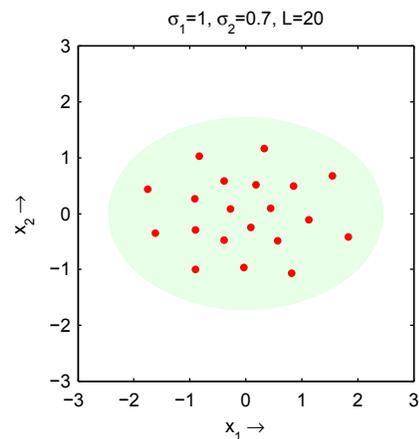


Fig. 1. Dirac mixture approximation of Gaussian density with  $L = 20$  components for  $\sigma_1 = 1.0$  and  $\sigma_2 = 0.7$ . Moments of up to third order are maintained.

to third order and homogeneously covers the given Gaussian density is shown in Fig. 1.

Measuring closeness can be performed in various ways. However, direct measures for shape differences are typically not applicable as they are not well defined for discrete densities on continuous domains. In a previous paper [3], we focused on comparing probability masses on all scales between the true continuous density and its discrete approximation with the goal of shifting the mass distribution of the approximating density as close to the true density as possible.

As comparing probability masses on all scales is computationally expensive, we devised a suboptimal derivative of this method in [4]. The key idea is to characterize individual Dirac components by so called repulsion kernels 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. 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.

This paper is based upon the idea in [4], but instead of arbitrary true densities focuses on the approximation of Gaussian densities. For the required distance measure between a

Gaussian density and its Dirac mixture approximation, a closed-form expression is derived. For performing the optimization, a randomized method is employed instead of a deterministic one as in [4].

The paper is structured as follows. A problem formulation is given in Sec. II. Sec. III gives an overview of the state of the art. The key idea of the paper is presented in Sec. IV. For comparing the true Gaussian density and its Dirac mixture approximation, an induced kernel density representing the Dirac mixture approximation is derived in Sec. V. Based on the induced kernel density, a distance measure is derived in Sec. VI. The randomized optimization method is introduced in Sec. VII. An evaluation is conducted in Sec. VIII. Conclusions are drawn in Sec. IX.

## II. PROBLEM FORMULATION

A random vector  $\underline{x} = [x_1, x_2, \dots, x_N]^T$  with realizations  $\underline{x} \in \mathbb{R}^N$  is characterized by a given *continuous* probability density function (pdf)  $\tilde{f}(\underline{x})$ . Here, we consider an arbitrary Gaussian density  $f(\underline{x})$ . For the derivations, we assume without loss of generality that it has been rotated to an axis-aligned form and shifted to the origin according to

$$\tilde{f}(\underline{x}) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2} \frac{x_k^2}{\sigma_k^2}\right), \quad (1)$$

where  $\sigma_k$  is the standard deviation along coordinate axis  $k$ .

### A. Dirac Mixture Approximation

The given continuous pdf  $\tilde{f}(\underline{x})$  is approximated 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

$$f(\underline{x}) = \sum_{i=1}^L w_i \cdot \delta(\underline{x} - \hat{\underline{x}}_i), \quad (2)$$

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 (2) 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 focus on adjusting the locations only. The component weights are all equal. In addition, we assume that a solution exists, i.e., the number of components  $L$  is selected to be large enough so that locations exist that fulfill the moment constraints defined in the next subsection.

We define a parameter vector  $\underline{\eta} \in \mathcal{S} = \mathbb{R}^{L \cdot N}$  containing the parameters as

$$\underline{\eta} = [\hat{x}_1^T, \hat{x}_2^T, \dots, \hat{x}_L^T]^T \quad (3)$$

and write  $f(\underline{x}) = f(\underline{x}, \underline{\eta})$ . Of course,  $\underline{\eta}$  can be obtained as  $\underline{\eta} = \hat{\mathbf{X}}(\cdot)$ . Please note that we use  $\underline{\eta}$  and  $\hat{\mathbf{X}}$  exchangeably also as an argument to functions such as the induced kernel density.

### B. Moment Constraints

The given Gaussian 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 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  $\underline{\eta}$ .

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

$$e_{\boldsymbol{\kappa}} = \int_{\mathbb{R}^N} \underline{x}^{\boldsymbol{\kappa}} f(\underline{x}) d\underline{x} \quad (4)$$

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

$$\mathbf{E}_M(\boldsymbol{\kappa} + 1) = \begin{cases} e_{\boldsymbol{\kappa}} & |\boldsymbol{\kappa}| \leq M \\ 0 & \text{elsewhere} \end{cases}, \quad (5)$$

with  $\boldsymbol{\kappa} \leq M$  and  $e_{\boldsymbol{\kappa}}$  from (4). 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 (14).

### C. Regularization

When the length  $L \cdot N$  of the parameter vector  $\underline{\eta}$  is larger than the number of moment constraints, the parameters of  $f(\underline{x}, \underline{\eta})$  are redundant and 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. This form of regularization does not consider the characteristics of the given continuous density. 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 *and* is close to the probability mass distribution of the given continuous density.

## III. STATE OF THE ART

Generating samples for a given probability density function is usually performed by random sampling. This is the simplest and fastest method and well-established algorithms are available for that purpose. However, with random sampling, samples are produced independently. As a result, convergence of its statistics, such as the moments, to the true values is slow so that many samples are required. The independent generation of samples also leads to a non-homogeneous coverage of the given density.

As an alternative to random sampling, a variety of deterministic sampling methods have been proposed. Moment-based

approximations of Gaussian densities are the basis for Linear Regression Kalman Filters (LRKFs), see [5]. Examples are the Unscented Kalman Filter (UKF) in [6] and its scaled version in [7], its higher-order generalization in [8], and a generalization to an arbitrary number of deterministic samples placed along the coordinate axes introduced in [9].

Deterministic sampling (Dirac mixture approximation) based on distance measures has been proposed for scalar continuous densities [10], [11] and a prespecified number of components. For sequentially increasing the number of components, an algorithm is given in [12] and applied to recursive nonlinear prediction in [13]. For arbitrary multi-dimensional Gaussian densities, Dirac mixture approximations are systematically calculated in [3]. A more efficient method for the case of standard normal distributions with a subsequent transformation is given in [14].

A faster but suboptimal version of [3] has been introduced in [4]. Instead of comparing the probability masses on all scales as in [3], repulsion kernels are introduced to assemble an induced kernel density and perform the comparison of the true density with its Dirac mixture approximation. This paper is based on [4] and specifically considers Gaussian densities for which a closed-form expression for the distance measure is derived. In addition, a randomized optimization method is employed instead of a quasi-Newton method. This has the advantage of giving almost anytime results with only a slightly worse final performance.

A first approach to Dirac mixture approximation of circular probability density functions analogous to the UKF for linear quantities is introduced in [15] for the von Mises distribution and the wrapped Normal distribution. Based on matching the first circular moment, three Dirac components are systematically placed. This Dirac mixture approximation of continuous circular probability density functions has already been applied to sensor scheduling based on bearings-only measurements [16]. The results have also been used to perform recursive circular filtering for tracking an object constrained to an arbitrary one-dimensional manifold in [17]. An approach similar to the one proposed in this paper has been derived for Dirac mixture approximation of circular probability density functions with an arbitrary number of Dirac components in [18].

In some cases, it makes sense to combine random and deterministic sampling. An example can be found in [19].

#### IV. KEY IDEA

We will now discuss the key idea for comparing a given continuous density with its Dirac mixture approximation.

Directly comparing continuous densities and Dirac mixture densities is not a trivial task and typical distance measures such as the Kullback-Leibler distance or squared integral distances are not suited for that purpose [20]. Thus, instead of using densities, cumulative distributions are compared. This, however, is limited to the scalar case. For multi-dimensional problems, the concept of localized cumulative distributions (LCDs) has been proposed [20]. In contrast to cumulative distributions, LCDs are unique and symmetric also in several dimensions.

Comparing the LCDs of the given continuous density and its Dirac mixture approximation corresponds to comparing

probability masses on all scales [3], i.e., the masses of both densities under certain kernels for all possible kernel locations and widths are compared. Integral measures can then be used for these mass functions. An appropriate scale-dependent weighting results in an error spectrum with “blue noise” characteristics [21]. The error is small for low frequencies and is allowed to be larger for higher frequencies.

In this paper, we also 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. However, instead of the LCD-based approach from [3], we use a simplified and computationally more efficient approach inspired by blue-noise sampling, e.g., see [22]. Instead of discs, however, we use kernels for describing the mass of individual Dirac components as in [23].

The key idea of this paper is to characterize the individual Dirac components by repulsion kernels, i.e., kernel functions representing the spread of probability mass that is appropriate at a given location. Only the true densities at the Dirac locations themselves are considered for calculating the kernel function parameters, which reduces computational complexity. By comparing the deviation between the true density and the induced kernel density, a distance measure is obtained. Here, a squared integral distance is used for that purpose. An optimization method is then used to find the Dirac locations that minimize this distance.

#### V. INDUCED KERNEL DENSITY

We now define repulsion kernels that characterize the amount of Dirac components required at a certain location in order to approximate the given continuous density well. The repulsion 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.

Each Dirac component  $i$  located at  $\hat{\underline{x}}_i$  is characterized by a Gaussian density normalized to the height of the true density and placed at the location of the Dirac component. We use 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 repulsion kernel

$$k_i(\underline{x}) = \tilde{f}(\hat{\underline{x}}_i) \exp\left(-\frac{1}{2} \frac{1}{\tau_i^2} (\underline{x} - \hat{\underline{x}}_i)^T (\underline{x} - \hat{\underline{x}}_i)\right) \quad (6)$$

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

The spread parameters  $\tau_i$  in (6) are determined by maintaining the mass represented by the Dirac component given by  $w_i$ , so we obtain

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

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

$$\tau_i = \frac{1}{\sqrt{2\pi}} \left( \frac{w_i}{\tilde{f}(\hat{\underline{x}}_i)} \right)^{\frac{1}{N}} . \quad (7)$$

The induced kernel density is obtained by inserting  $\tau_i$  from (7) into the respective repulsion kernels in (6) as

$$k(\underline{x}, \underline{\eta}) = \sum_{i=1}^L k_i(\underline{x}) \quad (8)$$

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

When  $k(\underline{x}, \underline{\eta})$  is close to  $\tilde{f}(\underline{x})$  on  $\mathbb{R}^N$ , the mass distributions of  $\tilde{f}(\underline{x})$  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 as a replacement for  $f(\underline{x}, \underline{\eta})$  to define a distance measure  $D(\underline{\eta})$  between  $\tilde{f}(\underline{x})$  and  $f(\underline{x}, \underline{\eta})$  on  $\mathbb{R}^N$ .

## VI. DISTANCE MEASURE

Given the induced kernel density  $k(\underline{x}, \underline{\eta})$  in (8) 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

$$D(\underline{\eta}) = D \left( \tilde{f}(\underline{x}), f(\underline{x}, \underline{\eta}) \right) \quad (9)$$

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

As specific distance measure, we employ the squared integral distance measure given by

$$D(\underline{\eta}) = \int_{\mathbb{R}^N} \left( \tilde{f}(\underline{x}) - k(\underline{x}, \underline{\eta}) \right)^2 d\underline{x} . \quad (10)$$

However, due to the definition of the repulsion kernel (6), the squared integral distance measure (10) places a higher emphasis on kernels placed at locations with large original density. To guarantee a homogeneous placement quality independent of the height of the original density, it is advisable to increase the 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 integral defining the distance  $D(\underline{\eta})$  can be calculated in closed form as

$$D(\underline{\eta}) = \sum_{i=1}^L \sum_{j=1}^L \tilde{f}(\hat{\underline{x}}_i) \tilde{f}(\hat{\underline{x}}_j) \prod_{k=1}^N I_k , \quad (12)$$

where  $I_k$  is given by

$$I_k = \frac{2\pi\sigma_k^2\tau_i\tau_j}{\sqrt{\Sigma_k}} \exp \left( -\frac{1}{2} \frac{\sigma_k^2(\hat{x}_{ki} - \hat{x}_{kj}) - \tau_j^2\hat{x}_{ki}^2 - \tau_i^2\hat{x}_{kj}^2}{\Sigma_k} \right)$$

and

$$\Sigma_k = \sigma_k^2(\tau_i^2 + \tau_j^2) - \tau_i^2\tau_j^2 .$$

The derivation is given in the appendix.

## VII. OPTIMIZATION

Minimizing the distance measure in (9) while maintaining the moment constraints in (5) of up to  $M$ -th order gives

$$\underline{\eta}_{\text{opt}} = \arg \min_{\underline{\eta} \in \mathcal{S}} D(\underline{\eta}) \quad \text{s.t.} \quad \mathbf{E}_M = \tilde{\mathbf{E}}_M . \quad (13)$$

As the optimization problem (13) is nonlinear and non-convex with nonlinear equality constraints, finding the optimal parameter vector  $\underline{\eta}_{\text{opt}}$  is not a trivial task. Standard optimization methods such as gradient descent will most likely end up in local minima.

To avoid local minima, a homotopy continuation method [24] was devised in [4]. This method starts with a density with known Dirac mixture approximation and then continuously approaches the desired density. A progression parameter  $\gamma \in [0, 1]$  is used to parametrize the true density according to  $\tilde{f}(\underline{x}, \gamma)$ , where  $\tilde{f}(\underline{x}, \gamma = 0)$  corresponds to the initial density with known Dirac mixture approximation and  $\tilde{f}(\underline{x}, \gamma = 1)$  corresponds to the given density for which a Dirac mixture approximation is desired. The parameter  $\gamma$  starts at zero and is progressively increased towards one. For every  $\gamma$ , the optimization problem is solved by just correcting the result from the previous step, which ensures tracking the desired solution. For  $\gamma = 1$ , the desired parameter vector  $\underline{\eta}_{\text{opt}}$  is reached.

In this paper, we pursue a different optimization approach, that is randomized optimization with modifying single Dirac component at a time. We use three ingredients: i) identification of critical components, ii) splitting one component into two, and iii) modification of a component.

We employ a modified distance measure that includes the moment constraints as an additional penalty term according to

$$D'(\underline{\eta}) = D(\underline{\eta}) + \gamma \cdot \|\tilde{\mathbf{E}}_M - \mathbf{E}_M\|_F^2 ,$$

where  $\gamma$  weights the penalty and  $\|\cdot\|_F$  is the Frobenius norm defined for a matrix  $\mathbf{H} \in \mathbb{R}^{N \times M}$  as

$$\|\mathbf{H}\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^M |h_{ij}|^2} .$$

The identification of critical components is performed based on the vector of absolute differences  $\underline{e} = |\tilde{f}(\underline{x}) - k(\underline{x}, \underline{\eta})|$  taken at the components' locations. A roulette wheel selection is used to determine a single component with a probability proportional to the value of the vector of absolute differences  $\underline{e}$ .

Critical components are split up into two components. Splitting is performed until the prespecified number of  $L$  components is reached.

Modification of critical components is performed by small random location changes. A location change is accepted when the distance measure is decreased. Otherwise, it is rejected. Components are modified until an equilibrium is achieved, i.e., changes in components do not decrease the distance measure anymore.

The proposed optimization method is shown as pseudo-code in Alg. 1. In the first step, the Dirac mixture approximation is initialized with two random components (line 2 to line 4).

While the actual number of Dirac components  $L'$  is smaller than the number of desired Dirac components  $L$ , splitting and subsequent modifications are performed. For splitting, a component is randomly selected from a discrete uniform density in line 8. A new component is split off the selected one  $i$  by first generating an  $N$ -dimensional realization  $\underline{r}$  of a Gaussian random variable with small covariances (line 9) that is then added to the selected component in line 10 so that the number of actual components is incremented by one, line 11.

For optimizing the current configuration given  $L'$  Dirac components, the corresponding distance measure is calculated and stored in line 12. In addition, the counter for tracking the consecutive number of unsuccessful modifications, Rejection-Counter, is reset to zero in line 13. Random modifications of the current  $L'$  Dirac components are now attempted as long as the consecutive number of unsuccessful modifications reaches its maximum, MaxRejections, see line 29.

The Dirac components are modified one at a time by first selecting a critical component  $i$  based on the error  $\underline{e}$  between the true density evaluated at the component locations  $\tilde{f}(\hat{\underline{X}})$  and the induced kernel density evaluated at the component locations  $k(\hat{\underline{X}}, \hat{\underline{X}})$  in line 15. A roulette wheel selection is used to produce a random index  $i$  with probabilities proportional to the error  $\underline{e}$  in line 16. The current Dirac mixture approximation is stored in a temporary matrix  $\hat{\underline{X}}^t$  in line 17, which is then modified in line 19 by adding a realization of a Gaussian random variable generated in line 18.

For accepting or rejecting a modification of a Dirac component, we prophylactically increment the rejection counter in line 20. The distance measure between the true density and the induced kernel density for the modified Dirac mixture approximation is calculated in line 21 and compared to the stored distance measure  $D_{\text{old}}$  in line 22. When the current distance measure is smaller, the modification is accepted and the temporary Dirac mixture approximation replaces the current one in line 23. The current distance measure is stored for the next comparison in line 27. The rejection counter is only reset, when the modification was significant, see line 24 to line 26.

## VIII. EVALUATION

In this section, some Dirac mixture approximations of Gaussian densities will be shown. The focus is on two-dimensional Gaussian, i.e.,  $N = 2$ , to allow for a visual inspection.  $N$ -dimensional approximations and a comparison with other methods for Dirac mixture approximation are omitted due to space limitations.

Gaussians with arbitrary means and different eigenvalues of the covariance matrix that are not aligned with the coordinate axes can be approximated. Before approximation, these densities are transformed, i.e., shifted and rotated, to the standard form in (1).

Please note that the arbitrary Gaussian densities are not scaled, especially not scaled with axis-dependent scaling factors, to obtain the standard form, as Dirac mixture approximations are not invariant to scaling!

The Dirac mixture approximation of (1) should maintain moments up to third order, so we consider a moment matrix

$\mathbf{E}_3$  with

$$\mathbf{E}_3 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ e_{00} & e_{01} & e_{02} & e_{03} \\ e_{10} & e_{11} & e_{12} & 0 \\ e_{20} & e_{21} & 0 & 0 \\ e_{30} & 0 & 0 & 0 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} . \quad (14)$$

For the considered true density, the Gaussian density in (1), 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  $\tilde{e}_{20} = \sigma_1^2$ ,  $\tilde{e}_{11} = 0$ ,  $\tilde{e}_{02} = \sigma_2^2$ , 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

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

The resulting Dirac mixture approximations for Gaussian densities with different standard deviations for three different numbers of components  $L = 15$ ,  $L = 30$ , and  $L = 50$  are given in Fig. 2, Fig. 3, and Fig. 4, respectively.

## IX. CONCLUSION

A randomized optimization approach for Dirac mixture approximation of Gaussian densities based on a closed-form distance measure is proposed. The resulting Dirac mixture approximations are homogeneously distributed and maintain a set of predefined moments.

Compared to the Dirac mixture approximation approach in [3] based on comparing probability masses at all scales, the approach based on repulsion kernels is slightly suboptimal, but allows a closed-form distance measure for Gaussian densities and is significantly faster. In contrast to the deterministic optimization approach in [4], the randomized optimization method proposed in this paper gives slightly worse results, but is by far simpler to implement, faster, and also provides useful results when interrupted prematurely (anytime algorithm).

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.

## APPENDIX

The induced kernel density in (8) can be written as

$$k(\underline{x}) = \sum_{i=1}^L \tilde{f}(\hat{\underline{x}}_i) \prod_{k=1}^N \exp\left(-\frac{1}{2} \frac{(x_k - \hat{x}_{ki})^2}{\tau_i^2}\right) .$$

As a result,  $k^2(\underline{x})$  is given by

$$k^2(\underline{x}) = \sum_{i=1}^L \sum_{j=1}^L \tilde{f}(\hat{\underline{x}}_i) \tilde{f}(\hat{\underline{x}}_j) \prod_{k=1}^N \exp\left(-\frac{1}{2} \frac{(x_k - \hat{x}_{ki})^2}{\tau_i^2}\right) \cdot \exp\left(-\frac{1}{2} \frac{(x_k - \hat{x}_{kj})^2}{\tau_j^2}\right) .$$

For  $k^2(\underline{x})/\tilde{f}(\underline{x})$ , we obtain

$$\frac{k^2(\underline{x})}{\tilde{f}(\underline{x})} = \sum_{i=1}^L \sum_{j=1}^L \tilde{f}(\hat{\underline{x}}_i) \tilde{f}(\hat{\underline{x}}_j) \prod_{k=1}^N \exp\left(-\frac{1}{2} \frac{(x_k - \hat{x}_{ki})^2}{\tau_i^2}\right) \cdot \exp\left(-\frac{1}{2} \frac{(x_k - \hat{x}_{kj})^2}{\tau_j^2}\right) \cdot \sqrt{2\pi}\sigma_k \exp\left(\frac{1}{2} \frac{x_k^2}{\sigma_k^2}\right).$$

Performing the integration in (11) dimension-wise in closed form gives the result in (12).

## 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] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] —, "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.
- [12] 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.
- [13] 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.
- [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] 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.
- [16] 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.
- [17] 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)*, Montbéliard, France, Oct. 2013.
- [18] U. D. Hanebeck and A. Lindquist, "Moment-based Dirac Mixture Approximation of Circular Densities (to appear)," in *Proceedings of the 19th IFAC World Congress (IFAC 2014)*, Cape Town, South Africa, Aug. 2014.
- [19] 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.
- [20] 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.
- [21] R. Ulichney, "Dithering with Blue Noise," *Proceedings of the IEEE*, vol. 76, no. 1, pp. 56–79, Jan. 1988.
- [22] 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>
- [23] 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>
- [24] E. L. Allgower and K. Georg, *Introduction to Numerical Continuation*

### 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 (3) and number of progression steps PC

```

1 // Initialize with two random components
2  $L' := 2$ ;
3  $\hat{x}_1 \sim \mathcal{N}(0, \epsilon \mathbf{I})$ ,  $\hat{x}_2 \sim \mathcal{N}(0, \epsilon \mathbf{I})$ ;
4  $\hat{X} := [\hat{x}_1, \hat{x}_2]$ ;

5 // Define maximum number of consecutive rejections
5 MaxRejections := 100;

6 // Define factor by which distance measures must differ
6  $\alpha := 0.9999$ ;

7 while  $L' < L$  do
8   // Select random component  $i$ 
8    $i \sim \text{DUD}(1, L')$ ;
9   // Split off new component from component  $i$ 
9    $r \sim \mathcal{N}(0, \epsilon \mathbf{I})$ ;
10   $\hat{X} := [\hat{X}, \hat{x}_i + r]$ ;
11   $L' := L' + 1$ ;

12 // Distance measure for current parameters
12  $D_{\text{old}} := D'(\tilde{f}(x), k(x, \hat{X}))$ ;
13 RejectionCounter := 0;
14 repeat
15   // Select critical component  $i$ 
15    $e := |\tilde{f}(\hat{X}) - k(\hat{X}, \hat{X})|$ ;
16    $i = \text{RouletteWheelSelection}(e)$ ;
17   // Modify selected component
17    $\hat{X}^t := \hat{X}$ ;
18    $r \sim \mathcal{N}\left(0, \epsilon \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)$ ;
19    $\hat{x}_i^t := \hat{x}_i + r$ ;
20   // Accept/reject modification
20   RejectionCounter := RejectionCounter + 1;
21    $D := D'(\tilde{f}(x), k(x, \hat{X}^t))$ ;
22   if  $D < D_{\text{old}}$  then
23      $\hat{X} := \hat{X}^t$ ;
24     if  $D < \alpha \cdot D_{\text{old}}$  then
25       | RejectionCounter := 0;
26     end
27      $D_{\text{old}} := D$ ;
28   end
29 until RejectionCounter == MaxRejections;
30 end
```

**Algorithm 1:** Approximation of a given continuous density  $\tilde{f}(x)$  by a Dirac mixture approximation  $f(x, \underline{\eta})$  with parameter vector  $\underline{\eta}$ .  $\text{DUD}(1, L')$  is the discrete uniform density over integers  $1, \dots, L'$ .  $\mathcal{N}(\underline{m}, \mathbf{C})$  is the normal distribution with mean  $\underline{m}$  and covariance matrix  $\mathbf{C}$ .  $\epsilon$  is a small positive real number.

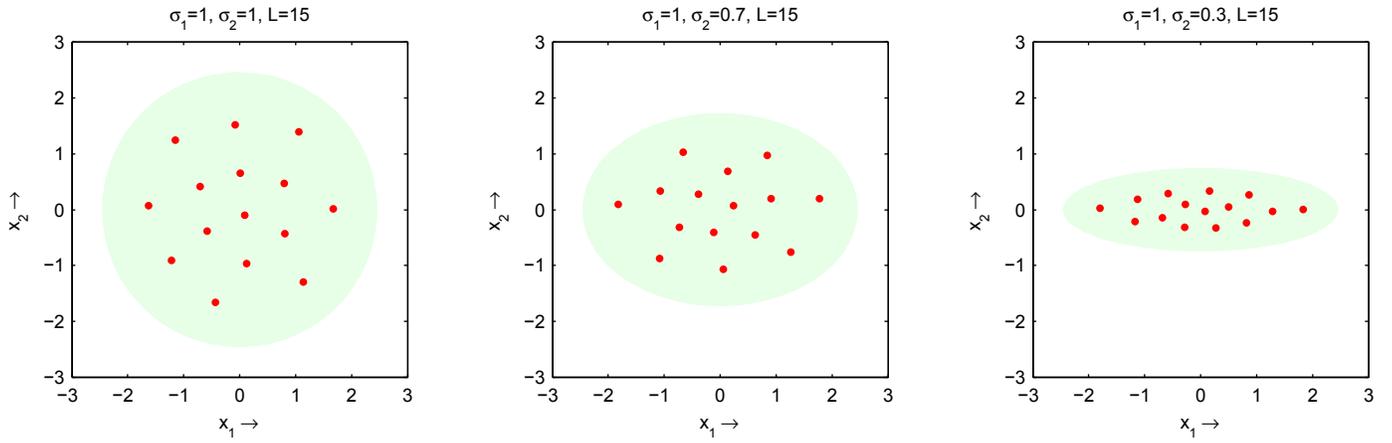


Fig. 2. Dirac mixture approximation of Gaussian density with  $L = 15$  components for  $\sigma_1 = 1.0$  and (left)  $\sigma_2 = 1.0$ , (middle)  $\sigma_2 = 0.7$ , (right)  $\sigma_2 = 0.3$ . Moments of up to third order are maintained.

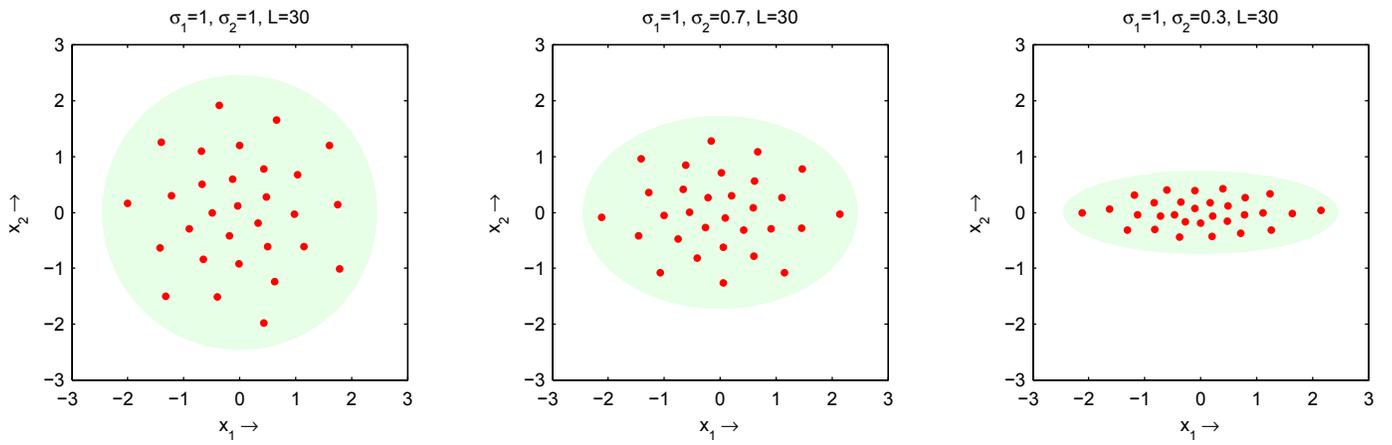


Fig. 3. Dirac mixture approximation of Gaussian density with  $L = 30$  components for  $\sigma_1 = 1.0$  and (left)  $\sigma_2 = 1.0$ , (middle)  $\sigma_2 = 0.7$ , (right)  $\sigma_2 = 0.3$ . Moments of up to third order are maintained.

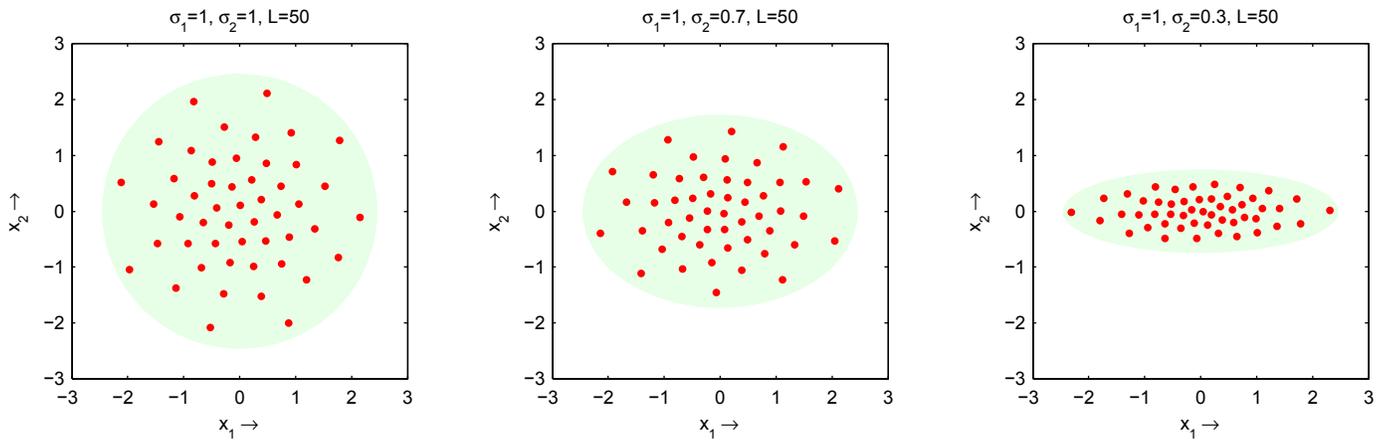


Fig. 4. Dirac mixture approximation of Gaussian density with  $L = 50$  components for  $\sigma_1 = 1.0$  and (left)  $\sigma_2 = 1.0$ , (middle)  $\sigma_2 = 0.7$ , (right)  $\sigma_2 = 0.3$ . Moments of up to third order are maintained.