2020 54th Annual Conference on Information Sciences and Systems (CISS) 978-1-7281-4085-8/20/$31.00 ©2020 IEEE 10.1109/CISS48834.2020.1570617413

Deterministic Sampling of Multivariate Densities based on Projected Cumulative Distributions

Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
Uwe.Hanebeck@kit.edu

***Abstract*—We approximate general multivariate probability density functions by deterministic sample sets. For optimal sampling, the closeness to the given continuous density has to be assessed. This is a difficult challenge in multivariate settings. Simple solutions are restricted to the one-dimensional case. In this paper, we propose to employ one-dimensional density projections. These are the Radon transforms of the densities. For every projection, we compute their cumulative distribution function. These Projected Cumulative Distributions (PCDs) are compared for all possible projections (or a discrete set thereof). This leads to a tractable distance measure in multivariate space. The proposed approximation method is efficient as calculating the distance measure mainly entails sorting in one dimension. It is also surprisingly simple to implement.**

***Index Terms*—Multivariate probability density functions, density (re)approximation, Cramér-von Mises distance, Radon transform, deterministic sampling, Dirac mixtures, recursive Bayesian filtering.**

I. Introduction

Approximating a given probability density function by another one is a ubiquitous problem. It arises when the given density is too complex for further processing. This could be caused by an undesired and complicated functional representation. Even if the representation is of a desired form, the number of parameters could be unnecessarily large, e.g., Gaussian mixtures. Hence, the given density representation has to be replaced by a more convenient representation or one with less parameters.

Recursive Bayesian filtering is an important use case for density approximation. For nonlinear system models, the prediction step is usually infeasible for continuous prior densities. Hence, an approximation of the prior by a set of samples is often used instead. The filtering step entails the product of the prior and a likelihood function. In some cases, e.g., Gaussian mixture densities, this product leads to an undesired exponential increase in parameters and calls for regular reapproximation.

The given density and its approximation can both be either continuous or discrete (over continuous domains). This paper is on approximating a given continuous density by a deterministic discrete density, see Fig. 1. The discrete density is represented by a weighted set of Dirac delta functions at specific locations. The deterministic sampling procedure includes the calculation of appropriate locations and, if desired, also appropriate weights. In this paper, we focus on equally weighted samples or a least on samples with prescribed weights.

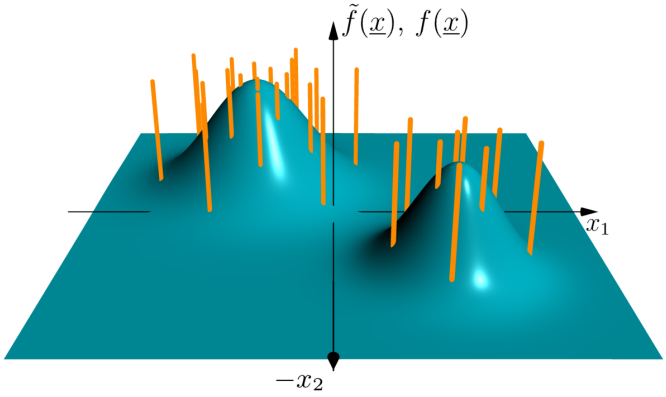

Fig. 1: Turquoise: Gaussian mixture with two components. Orange: Dirac mixture approximation.

II. Problem Formulation

The problem we solve in this paper is the systematic approximation of a given multivariate, non-Gaussian density $\tilde{f}(\underline{x})$ by a Dirac mixture density $f(\underline{x})$, i.e., by a set of L deterministic samples, given by

$$f(\underline{x}) = f(\underline{x} \,|\, \hat{\mathbf{X}}) = \sum_{i=1}^{L} w_i \, \delta(\underline{x} - \underline{\hat{x}}_i) \tag{1}$$

with given weights $w_i > 0$, $\sum_{i=1}^{L} w_i = 1$, and desired locations $\underline{\hat{x}}_i \in \mathbb{R}^N$. N is the number of dimensions. The locations are collected in an $N \times L$ matrix $\hat{\mathbf{X}} = \left[\underline{\hat{x}}_1, \underline{\hat{x}}_2, \ldots, \underline{\hat{x}}_L\right]$.

As we insist on a systematic approximation, we require a rigorous distance measure between the given density $\tilde{f}(\underline{x})$ and its approximation $f(\underline{x})$ to be minimized. As this entails a comparison between a continuous and a discrete density, finding an appropriate and feasible distance measure is a major challenge. An overview of the current literature is given in the next section.

978-1-7281-4085-8/20/$31.00 ©2020 IEEE

III. State of the Art

We will now give an overview of the current state of the literature on approximating given continuous probability density functions by a set of samples. Most results are available for Gaussian densities. Although we focus on the non-Gaussian case, we will also review some literature for Gaussians. We will focus on deterministic approximations.

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.

A. Deterministic Sampling: Moment-based Approaches

As an alternative to random sampling, deterministic sampling methods have been proposed. Moment-based approximations of Gaussian densities are the basis for Linear Regression Kalman Filters (LRKFs) [1]. Examples are the Unscented Kalman Filter (UKF) [2], its scaled version [3], and its higher-order generalization [4]. These methods are limited to a fixed number of samples per dimension. An arbitrary number of deterministic samples of a Gaussian density is provided by the spherical-radial rule [5] for the so called cubature filter.

B. Deterministic Sampling: Distance-based Approaches

Instead of relying on moment relations that quickly become intractable, distance-based approaches minimize an appropriate distance measure between the given continuous density and its discrete approximation. Distance-based approaches have several distinct advantages: (1) Arbitrary number of samples. (2) Homogeneous allocation of samples. (3) Prioritization of density regions especially important for further processing.

1) One-dimensional Densities: For one-dimensional (1D) densities, the vast literature on 1D statistical tests can be exploited. These are used for testing the hypothesis that a set of samples have a certain density. Inverting theses tests can be used to fit samples to a given density. We will focus on one-sample tests that compare the cumulative distribution function (CDF) of the continuous density with the empirical distribution function (EDF) of the samples. The Kolmogorov-Smirnov test [6] uses the maximum between the CDF and the EDF, which is difficult to optimize. Integral squared distances between CDF and EDF were first considered by Cramér [7] and von Mises leading to the Cramér-von Mises test. Several tests have been developed on that basis, differing only in the weighting function for the squared difference. One notable test is the Anderson-Darling test [8].

Dirac mixture approximation (with equally weighted samples) of continuous densities based on the Cramér-von Mises criterion for the univariate case is proposed in [9]. Optimizing both weights and locations is considered in [10]. Sequentially increasing the number of samples is the focus of [11] and is applied to recursive nonlinear prediction in [12].

2) Multivariate Densities: For N-dimensional (ND) densities with $N > 1$, systematic approximations are much harder to achieve. Cumulative distribution functions in higher dimensions are difficult to calculate. In addition, they are not unique [13]. This is in contrast to the univariate case, where there are only two directions of integration for obtaining the cumulative distribution function from the density. The results of the two directions are dependent, i.e., sum up to one. In higher dimensions, the number of orderings of possible integration directions increases exponentially with 2^N, where N is the number of dimensions. $2^N - 1$ of these distributions are independent [14, p. 617]. Selecting an arbitrary distribution from these possibilities for density approximation typically results in a bias of the estimated parameters [13].

In order to cope with the non-uniqueness, it has been proposed in the context of multivariate statistical tests, to consider all possible orderings [14, p. 617]. This can easily be done in two or three dimensions [15]. For higher-dimensional spaces, considering all ordering is impractical.

Instead of using the standard cumulative distribution with its non-uniqueness issues, an alternative cumulative transformation is introduced in [13]. It is called Localized Cumulative Distribution (LCD). A distance measure is then defined based on the LCDs of the two densities to be compared. The LCD-based approach is used for systematically approximating arbitrary multi-dimensional Gaussian densities [16]. An improved method with better numerical stability that exploits the symmetry of the Gaussian density with symmetric samples is proposed in [17]. For multivariate standard normal distributions, a more computationally efficient scheme is presented that relies on a subsequent transformation [18]. The LCD-based approach provides very good approximations. However, for the case of approximating continuous densities with Dirac mixtures, closed-form solutions for the distance measure are only available in special cases.

Comparing cumulative distribution functions is, of course, not the only way to compare discrete densities or continuous and discrete densities. Alternatives include comparing characteristic functions [19] and using kernel estimates. Kernel estimates of a discrete density are exceptionally simple in the considered case of finding the best-fitting discrete density for a given continuous density. In that case, we use the fact that the density at a certain sample location should be equal to the corresponding value of the continuous density [20]. So called repulsion kernels are employed at the sample locations. The induced kernel density is then compared to the given continuous density by means of, e.g., an integral squared distance [21]. For given Gaussian densities that are approximated by Dirac

mixtures, a closed-form expression for the distance measure is derived in [22]. A randomized optimization method is then used for finding the optimal locations instead of a quasi-Newton method in [21].

We have seen that the univariate case is significantly simpler than the multivariate one. So it comes as no surprise that many attempts have been made to reduce the multivariate case to the univariate one:

- *Approximation on Principal Axes:* The first idea that immediately springs to mind is to use 1D approximations on the principal axes of the given continuous density. This is, of course, limited to densities where principal axes can naturally be defined. This includes Gaussian densities [23] in $\mathbb{R}^N$ and the Bingham distribution [24] defined on $\mathbb{S}^{N-1} \subset \mathbb{R}^N$.
- *Cartesian Products:* For random vectors with independent components, e.g., axis-aligned Gaussian densities, the N marginal densities can be approximated independently. The N-dimensional Dirac mixture approximation can then be obtained by a Cartesian product of the individual marginal approximations. An obvious disadvantage is that the total number of samples is the product of the individual numbers of samples, i.e., scales exponentially with the number of dimensions N.
- *One-dimensional Projections:* General densities cannot be represented by their marginals. This only holds for densities that can be factorized. However, instead of limiting projections to the coordinate axes as done for the marginals, more projections onto different axes can be considered. For representing arbitrary densities, all possible one-dimensional projections have to be considered. This is called the Radon transform of the given density [25], [26], and [27] for the higher-dimensional case. In [28], a multivariate Wasserstein distance is constructed from Wasserstein distance between 1D projections (called slices therein).

IV. Radon Density Transform

In this section, we will represent N-dimensional probability density functions via the set of all one-dimensional projections, i.e., the Radon Transform [25], [26].

We consider the linear projection $\boldsymbol{r} = \underline{u}^\top \underline{\boldsymbol{x}}$ of a random vector $\underline{\boldsymbol{x}} \in \mathbb{R}^N$ to a scalar random variable $\boldsymbol{r} \in \mathbb{R}$ onto the line described by unit vector $\underline{u} \in \mathbb{S}^{N-1}$. Given the probability density function $f(\underline{x})$ of the random vector $\underline{\boldsymbol{x}}$, the density $f_r(r \,|\, \underline{u})$ of $\boldsymbol{r}$ is given by

$$f_r(r \,|\, \underline{u}) = \int_{\mathbb{R}^N} f(\underline{t})\, \delta(r - \underline{u}^\top \underline{t}) \,\mathrm{d}\underline{t} \ .$$

$f_r(r \,|\, \underline{u})$ is the Radon transform of $f(\underline{x})$ for all $\underline{u} \in \mathbb{S}^{N-1}$, see Fig. 2. For the Dirac mixture approximation in (1), the Radon transform is given by

$$f_r(r \,|\, \hat{\underline{r}}, \underline{u}) = \sum_{i=1}^{L} w_i\, \delta\left(\underline{u}^\top \underline{x} - \underline{u}^\top \hat{\underline{x}}_i\right) = \sum_{i=1}^{L} w_i\, \delta\big(r - \hat{r}_i(\underline{u})\big) \ ,$$

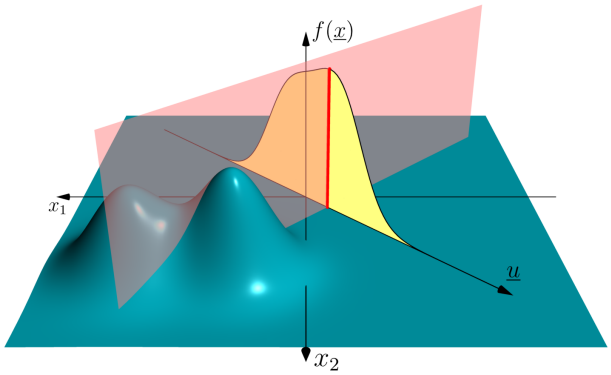

Fig. 2: Visualization of Radon transform.

where $\hat{r}_i(\underline{u}) = \underline{u}^\top \hat{\underline{x}}_i$, $i = 1, \ldots, L$ are the projected Dirac locations. We collect these locations in a vector $\hat{\underline{r}} = \left[\hat{r}_1(\underline{u}), \hat{r}_2(\underline{u}), \ldots, \hat{r}_L(\underline{u})\right]^\top$.

For Gaussian densities $f(\underline{x})$ with mean vector $\hat{\underline{x}}$ and covariance matrix $\mathbf{C}_x$, the density $f_r(r \,|\, \underline{u})$ resulting from the projection is also Gaussian. Its mean is given by $\hat{r}(\underline{u}) = \underline{u}^\top \hat{\underline{x}}$ and its standard deviation by $\sigma_r(\underline{u}) = \sqrt{\underline{u}^\top \mathbf{C}_x \underline{u}}$. For an N-dimensional Gaussian mixture density $f(\underline{x})$ with M components, means $\hat{\underline{x}}_i$, and covariance matrices $\mathbf{C}_{x,i}$

$$f(\underline{x}) = \sum_{i=1}^{M} w_i \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_{x,i}|}} \exp\left(-\frac{1}{2}(\underline{x} - \hat{\underline{x}}_i)^\top \mathbf{C}_{x,i}^{-1} (\underline{x} - \hat{\underline{x}}_i)\right),$$

the density $f_r(r, \underline{u})$ is also a Gaussian mixture. Due to the linearity of the projection operator, it is given by

$$f_r(r \,|\, \underline{u}) = \sum_{i=1}^{M} w_i \frac{1}{\sqrt{2\pi}\sigma_{r,i}(\underline{u})} \exp\left(-\frac{1}{2}\frac{\big(r - \hat{r}_i(\underline{u})\big)^2}{\sigma_{r,i}^2(\underline{u})}\right)$$

with $\hat{r}_i(\underline{u}) = \underline{u}^\top \hat{\underline{x}}_i$ and $\sigma_{r,i}(\underline{u}) = \sqrt{\underline{u}^\top \mathbf{C}_{x,i} \underline{u}}$, $i = 1, \ldots, M$.

V. Univariate Dirac Mixture Approximation

In this section, we consider one-dimensional Dirac mixture approximation. The one-dimensional approximation then serves as the basis for the N-dimensional approximation in the next section.

A. Distance Measure

The approximating density is given by the 1D Dirac mixture approximation

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

with weights $w_i > 0$ that sum up to one and locations $\hat{x}_i \in \mathbb{R}$. The locations are collected in a vector $\hat{\underline{x}} = \left[\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_L\right]^\top$. For rating the goodness of fit of $f(x \,|\, \hat{\underline{x}})$ to the given continuous density $\tilde{f}(x)$ we compare their cumulative distribution functions rather than their densities. For the Dirac mixture approximation $f(x \,|\, \hat{\underline{x}})$, the cumulative distribution function is given by

$$F(x \,|\, \hat{\underline{x}}) = \sum_{i=1}^{L} w_i\, \mathrm{H}(x - \hat{x}_i) \ ,$$

where H(.) is the Heaviside function given by

$$\mathrm{H}(x) = \int_{-\infty}^{x} \delta(t)\,\mathrm{d}t = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases} .$$

As a distance measure, we use the integral squared distance between the cumulative distribution function $\tilde{F}(x)$ of the standard normal distribution and the cumulative distribution function $F(x\,|\,\underline{\hat{x}})$ of its Dirac mixture approximation

$$D(\underline{\hat{x}}) = \int_{\mathbb{R}} \left[\tilde{F}(t) - F(t\,|\,\underline{\hat{x}})\right]^2 \mathrm{d}t \ .$$

B. Minimizing the Distance Measure

The gradient of the distance measure $D(\underline{\hat{x}})$ is given by

$$\begin{aligned} \underline{G}(\underline{\hat{x}}) = \nabla D(\underline{\hat{x}}) &= \frac{\partial D(\underline{\hat{x}})}{\partial \underline{\hat{x}}} \\ &= \left[\tfrac{\partial D(\underline{\hat{x}})}{\partial \hat{x}_1}, \quad \tfrac{\partial D(\underline{\hat{x}})}{\partial \hat{x}_2}, \quad \ldots, \quad \tfrac{\partial D(\underline{\hat{x}})}{\partial \hat{x}_L}\right]^\top \end{aligned}$$

with

$$\begin{aligned} \frac{\partial D(\underline{\hat{x}})}{\partial \hat{x}_i} &= 2w_i \int_{\mathbb{R}} \left[\tilde{F}(t) - F(t\,|\,\underline{\hat{x}})\right] \delta(t - \hat{x}_i)\,\mathrm{d}t \\ &= 2w_i \left[\tilde{F}(\hat{x}_i) - F(\hat{x}_i\,|\,\underline{\hat{x}})\right] \ . \end{aligned}$$

The empirical cumulative distribution function evaluated at location $\hat{x}_i$ is given by

$$F(\hat{x}_i\,|\,\underline{\hat{x}}) = \sum_{j=1}^{L} w_j\,\mathrm{H}(\hat{x}_i - \hat{x}_j) \ . \tag{2}$$

The Hesse matrix is given by

$$\mathbf{H}(\underline{\hat{x}}) = \mathrm{diag}\left(\left[\tfrac{\partial^2 D(\underline{\hat{x}})}{\partial \hat{x}_1^2}, \tfrac{\partial^2 D(\underline{\hat{x}})}{\partial \hat{x}_2^2}, \ldots, \tfrac{\partial^2 D(\underline{\hat{x}})}{\partial \hat{x}_L^2}\right]\right) \ ,$$

where

$$\frac{\partial^2 D(\underline{\hat{x}})}{\partial \hat{x}_i^2} = 2\,w_i\,\tilde{f}(\hat{x}_i) \ .$$

The minimum of $D(\underline{\hat{x}})$ is obtained iteratively using Newton's method

$$\Delta\underline{\hat{x}} = -\mathbf{H}^{-1}(\underline{\hat{x}})\,\underline{G}(\underline{\hat{x}})$$

starting with an initial location vector. When the location vector $\underline{\hat{x}}$ is sorted, i.e., $\hat{x}_1 < \hat{x}_2 < \ldots < \hat{x}_L$, the expression $H(\hat{x}_i - \hat{x}_j)$ in (2) gives

$$H(\hat{x}_i - \hat{x}_j) = \begin{cases} 0 & i < j \\ \frac{1}{2} & i = j \\ 1 & i > j \end{cases} .$$

As a result, $F(\hat{x}_i)$ can be written as

$$F(\hat{x}_i\,|\,\underline{\hat{x}}) = \frac{w_i}{2} + \sum_{j=1}^{i-1} w_j \ .$$

For equally weighted samples, i.e., $w_i = 1/L$, the expression can be simplified to

$$F(\hat{x}_i\,|\,\underline{\hat{x}}) = \frac{1}{2L} + (i-1)\frac{1}{L} = \frac{2i-1}{2L} \ .$$

VI. Multivariate Dirac Mixture Approximation

We are now given the three main ingredients: (1) We can represent a multivariate probability density function by its one-dimensional projections, see Sec. IV. (2) We have a mechanism for calculating a distance measure between a one-dimensional continuous density and its Dirac mixture approximation, see Subsec. V-A. (3) We can minimize this distance measure with a Newton-like method, see Subsec. V-B.

Based on these three ingredients, we will now assemble a multivariate distance measure between two continuous and/or discrete probability density functions. This entails the following seven steps.

Step 1 (1D Projections via Radon Transform): The given density $\tilde{f}(\underline{x})$ and its approximation $f(\underline{x})$ are represented by their Radon transforms $\tilde{f}(r\,|\,\underline{u})$ and $f(r\,|\,\underline{u})$, i.e., by their 1D projections onto unit vectors $\underline{u} \in \mathbb{S}^{N-1}$.

Step 2 (1D Cumulative Distributions): Based on the Radon transform $\tilde{f}(r\,|\,\underline{u})$, we calculate the one-dimensional cumulative distributions of the projected densities as

$$\tilde{F}(r\,|\,\underline{u}) = \int_{-\infty}^{r} \tilde{f}(t\,|\,\underline{u})\,\mathrm{d}t \ .$$

For a Dirac mixture approximation, the cumulative distribution function of its Radon transform is given by

$$F(r\,|\,\underline{\hat{r}}, \underline{u}) = \sum_{i=1}^{L} w_i\,\mathrm{H}\big(r - \hat{r}_i(\underline{u})\big) \ .$$

Step 3 (1D Distance): For comparing the one-dimensional projections, we compare their cumulative distributions $\tilde{F}(r\,|\,\underline{u})$ and $F(r\,|\,\underline{\hat{r}}, \underline{u})$ for all $\underline{u} \in \mathbb{S}^{N-1}$. As a distance measure we use the integral squared distance

$$D_1(\underline{\hat{r}}, \underline{u}) = \int_{\mathbb{R}} \left[\tilde{F}(r\,|\,\underline{u}) - F(r\,|\,\underline{\hat{r}}, \underline{u})\right]^2 \mathrm{d}r \ . \tag{3}$$

This gives us the distance between the projected densities in the direction of the unit vector $\underline{u}$ for all $\underline{u}$.

Step 4 (1D Newton Step): The Newton step from Subsec. V-B can now be written as

$$\Delta\underline{\hat{r}}(\underline{\hat{r}}, \underline{u}) = -\mathbf{H}^{-1}(\underline{\hat{r}}, \underline{u})\,\underline{G}(\underline{\hat{r}}, \underline{u})$$

with

$$\underline{G}(\underline{\hat{r}}, \underline{u}) = \left[\frac{\partial D_1(\underline{\hat{r}}, \underline{u})}{\partial \hat{r}_1}, \quad \frac{\partial D_1(\underline{\hat{r}}, \underline{u})}{\partial \hat{r}_2}, \quad \ldots, \quad \frac{\partial D_1(\underline{\hat{r}}, \underline{u})}{\partial \hat{r}_L}\right]^\top$$

and

$$\frac{\partial D_1(\underline{\hat{r}}, \underline{u})}{\partial \hat{r}_i} = 2w_i \left[\tilde{F}(\hat{r}_i\,|\,\underline{u}) - F(\hat{r}_i\,|\,\underline{\hat{r}}, \underline{u})\right] \ .$$

The Hessian $\mathbf{H}(\underline{\hat{r}}, \underline{u})$ is given by

$$\mathbf{H}(\underline{\hat{r}}, \underline{u}) = 2\,\mathrm{diag}\left(\left[w_1\tilde{f}(\hat{r}_1\,|\,\underline{u}), \quad w_2\tilde{f}(\hat{r}_2\,|\,\underline{u}), \quad \ldots, \quad w_L\tilde{f}(\hat{r}_L\,|\,\underline{u})\right]\right)$$

and the resulting Newton step is

$$\Delta\hat{\underline{r}}(\hat{\underline{r}},\underline{u}) = -\begin{bmatrix} \frac{\tilde{F}(\hat{r}_1\,|\,\underline{u}) - F(\hat{r}_1\,|\,\hat{\underline{r}},\underline{u})}{\tilde{f}(\hat{r}_1\,|\,\underline{u})} \\ \frac{\tilde{F}(\hat{r}_2\,|\,\underline{u}) - F(\hat{r}_2\,|\,\hat{\underline{r}},\underline{u})}{\tilde{f}(\hat{r}_2\,|\,\underline{u})} \\ \vdots \\ \frac{\tilde{F}(\hat{r}_L\,|\,\underline{u}) - F(\hat{r}_L\,|\,\hat{\underline{r}},\underline{u})}{\tilde{f}(\hat{r}_L\,|\,\underline{u})} \end{bmatrix} .$$

Step 5 (Backprojection to ND Space): For a specific projection vector $\underline{u}$, we obtain a Newton update $\Delta\hat{\underline{r}}(\hat{\underline{r}},\underline{u})$. By means of a backprojection into the original N-dimensional space, this update can be used to modify the original Dirac locations in the direction *along* the vector $\underline{u}$. For every location vector $\hat{\underline{x}}_i$ we obtain

$$\Delta\hat{\underline{x}}_i(\underline{u}) = \Delta\hat{\underline{r}}(\underline{r}_i,\underline{u}) \cdot \underline{u} \ .$$

Step 6 (Assemble Multivariate Distance): The individual 1D distances $D_1(\hat{\underline{r}},\underline{u})$ can be assembled to form a multivariate distance measure by integrating over all 1D distances depending on unit vector $\underline{u}$

$$D_N(\hat{\mathbf{X}}) = \frac{1}{A_N}\int_{\mathbb{S}^{N-1}} D_1(\hat{\underline{r}},\underline{u})\,\mathrm{d}\underline{u} \ .$$

A_N is the surface area of the (hyper)sphere $\mathbb{S}^{N-1}$ embedded in $\mathbb{R}^N$. Plugging in $D_1(\underline{r},\underline{u})$ from (3) gives

$$D_N(\hat{\mathbf{X}}) = \frac{1}{A_N}\int_{\mathbb{S}^{N-1}}\int_{\mathbb{R}} \left[\tilde{F}(r\,|\,\underline{u}) - F(r\,|\,\hat{\underline{r}},\underline{u})\right]^2 \mathrm{d}r\,\mathrm{d}\underline{u} \ ,$$

which is a generalized Cramér-von Mises distance.

Step 7 (Perform Full Newton Update): A full Newton update can now be performed by integrating over all partial updates along projection vectors $\underline{u}$

$$\Delta\hat{\underline{x}}_i = \frac{1}{A_N}\int_{\mathbb{S}^{N-1}} \Delta\hat{\underline{x}}_i(\underline{u})\,\mathrm{d}\underline{u} \ .$$

In a practical implementation, the space $\mathbb{S}^{N-1}$ containing the unit vectors $\underline{u}$ has to be discretized. Two options are available for performing the discretization: (1) Deterministic discretization, e.g., by calculating a grid or (2) random discretization by drawing uniform samples from the hypersphere. In both cases, we consider K samples $\hat{\underline{u}}_k$ and the integration reduces to a summation

$$\Delta\hat{\underline{x}}_i \approx \frac{1}{K}\sum_{i=1}^{K} \Delta\hat{\underline{x}}_i(\hat{\underline{u}}_k) \text{ for } i = 1,2,\ldots,L \ .$$

Given initial locations for the sample locations, full Newton updates are performed until the maximum change over all location vectors falls below a given threshold.

VII. Numerical Results

We focus on 2D examples for visualization purposes. The Dirac mixture approximation with $L = 50$ components of a Gaussian mixture with two components with equal weights, different means, and identity covariance matrices is shown in Fig. 3. The results for zero means and different covariance matrices are shown in Fig. 4.

$$w_1 = 0.5,\ w_2 = 0.5,\ m_1 = \begin{bmatrix} -1.4 \\ 0.0 \end{bmatrix},\ m_2 = \begin{bmatrix} 1.4 \\ 0.0 \end{bmatrix}$$
$$C_1 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix},\ C_2 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix},\ L = 50$$

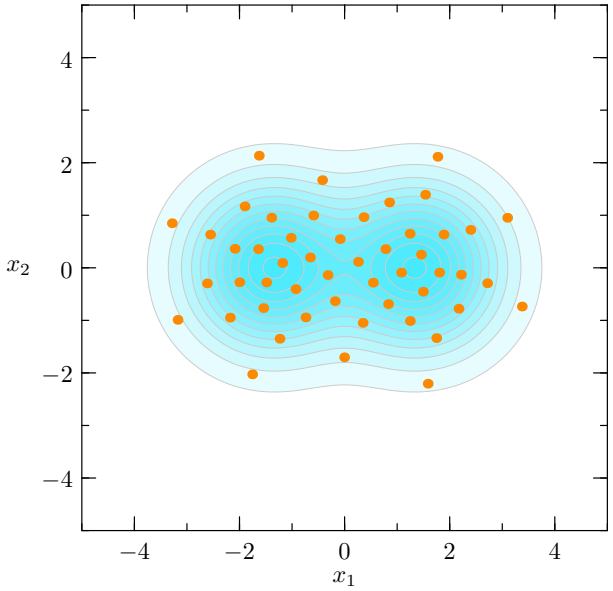

Fig. 3: Dirac mixture approximation with $L = 50$ components of Gaussian mixture with two components.

$$w_1 = 0.5,\ w_2 = 0.5,\ m_1 = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix},\ m_2 = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$$
$$C_1 = \begin{bmatrix} 3.0 & 2.8 \\ 2.8 & 3.0 \end{bmatrix},\ C_2 = \begin{bmatrix} 3.0 & -2.8 \\ -2.8 & 3.0 \end{bmatrix},\ L = 50$$

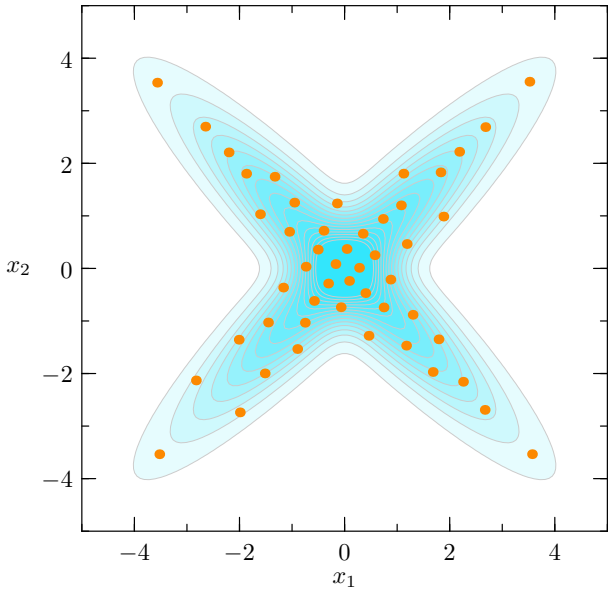

Fig. 4: Dirac mixture approximation with $L = 50$ components of Gaussian mixture with two components.

VIII. Conclusions

This paper proposes a new integral transformation of probability density functions (PDF). The so-called Projected Cumulative Distribution (PCD) is used for characterizing both continuous and discrete random variables. It is a convenient alternative to the standard cumulative distribution function (CDF). In multivariate settings, the CDF is non-unique, difficult to calculate, and asymmetric. This leads to various problems when used in density comparisons. The PCD does not suffer from these problems, can be efficiently calculated, and is simple to implement.

For comparing two densities, a generalized Cramér-von Mises distance between their two PCDs is derived. For a given continuous density $\tilde{f}(\underline{x})$, its Dirac mixture approximation $f(\underline{x})$ is then obtained by minimizing the distance measure with respect to its Dirac locations. Minimization is performed by combining Newton updates from all (discrete) projections.

The proposed density approximation is parallelizable. The combined Newton step can be performed in parallel for each projection.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] I. Arasaratnam and S. Haykin, "Cubature Kalman Filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[6] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[7] H. Cramér, "On the Composition of Elementary Errors: Second Paper: Statistical Applications," *Scandinavian Actuarial Journal*, vol. 1928, no. 1, pp. 141–180, Jan. 1928.

[8] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes," *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193–212, Jun. 1952.

[9] 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.

[10] ——, "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.

[11] 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.

[12] 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.

[13] 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.

[14] J. A. Peacock, "Two-Dimensional Goodness-of-Fit Testing in Astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 202, no. 3, pp. 615–627, Jan. 1983.

[15] G. Fasano and A. Franceschini, "A Multidimensional Version of the Kolmogorov-Smirnov Test," *Monthly Notices of the Royal Astronomical Society*, vol. 225, pp. 155–170, Mar. 1987.

[16] 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.

[17] J. Steinbring, M. Pander, and U. D. Hanebeck, "The Smart Sampling Kalman Filter with Symmetric Samples," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 71–90, Jun. 2016.

[18] 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.

[19] T. W. Epps and L. B. Pulley, "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika*, vol. 70, no. 3, pp. 723–726, 1983.

[20] R. Fattal, "Blue-Noise Point Sampling Using Kernel Density Model," in *ACM SIGGRAPH 2011 Papers*, ser. SIGGRAPH '11. New York, NY, USA: ACM, 2011, pp. 48:1–48:12.

[21] 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.

[22] ——, "Sample Set Design for Nonlinear Kalman Filters viewed as a Moment Problem," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, Jul. 2014.

[23] 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.

[24] K. Li, D. Frisch, B. Noack, and U. D. Hanebeck, "Geometry-Driven Deterministic Sampling for Nonlinear Bingham Filtering," in *Proceedings of the 2019 European Control Conference (ECC 2019)*, Naples, Italy, Jun. 2019.

[25] J. Radon, "Über Die Bestimmung von Funktionen Durch Ihre Integralwerte Längs Gewisser Mannigfaltigkeiten," *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse*, vol. 69, pp. 262–277, 1917.

[26] ——, "On the Determination of Functions from Their Integral Values Along Certain Manifolds," *IEEE Transactions on Medical Imaging*, vol. 5, no. 4, pp. 170–176, Dec. 1986.

[27] S. R. Deans, "The Radon Transform for Higher Dimensions," *Physics in Medicine and Biology*, vol. 23, no. 6, pp. 1173–1175, Nov. 1978.

[28] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister, "Sliced and Radon Wasserstein Barycenters of Measures," *Journal of Mathematical Imaging and Vision*, vol. 51, no. 1, pp. 22–45, Jan. 2015.