Efficient Deterministic Dirac Mixture Approximation of Gaussian Distributions

Igor Gilitschenski[1] and Uwe D. Hanebeck[1]

***Abstract*— We propose an efficient method for approximating arbitrary Gaussian densities by a mixture of Dirac components. This approach is based on the modification of the classical Cramér-von Mises distance, which is adapted to the multivariate scenario by using Localized Cumulative Distributions (LCDs) as a replacement for the cumulative distribution function. LCDs consider the local probabilistic influence of a probability density around a given point. Our modification of the Cramér-von Mises distance can be approximated for certain special cases in closed-form. The created measure is minimized in order to compute the positions of the Dirac components for a standard normal distribution.**

I. Introduction

State estimation in dynamic systems is well-understood in the situation of a linear system equation and additive Gaussian noise. In this situation, the Kalman-Filter [11] is an optimal estimator of the system state. This task becomes much more difficult, when the system equation is nonlinear and the noise is either not additive or not Gaussian. Many results exist for special cases of this general nonlinear system.

When the effects making up the disturbance of a dynamic system are additive and independent identically distributed (i.i.d.), it is often a good approximation to assume the system noise as Gaussian. The popularity of the Gaussian distribution as a noise assumption for uncertain dynamic systems is due to the central limit theorem, which states that normalized sums of i.i.d. random variables with finite variance converge to a Gaussian distributed random variable [14]. Furthermore, in dynamic systems with Gaussian additive noise and mild nonlinearity, the posterior density can be approximated by a Gaussian distribution.

A well-known approach to handling nonlinear dynamic systems is by linearizing the system equation using its Taylor series expansion, which is the basic principle of the Extended Kalman Filter [15] (EKF). Only in a limited scope of applications, the results of the EKF are sufficiently satisfactory. Especially when there are strong nonlinearities, the EKF delivers poor results, because the linearized equation does not sufficiently capture the system behaviour.

One way to avoid such problems in estimation is the use of discrete probability distributions on a continuous domain as a description of system uncertainty. Usage of discrete densities offers the advantage of easy propagation through the system equation and thus, easier prediction of the true system state.

In order to use discrete probability distributions as an approximately correct description of system uncertainty, it is desirable to approximate continuous probability distributions by a mixture of Dirac components. Particularly, the approximation of a Gaussian distribution is of considerable interest. Existing methods to perform this approximation can be divided into two categories.

First, random sampling can be used to approximate probability distributions. These Sequential Monte Carlo methods, such as the particle filter [3], are easily implemented, but usually require a large number of random samples. Furthermore, the particle filter performs badly with a rising problem dimensionality. A related concept is used in the Ensemble Kalman Filter [6], where sampling is applied in high-dimensional problems to propagate a sample covariance instead of the full covariance matrix.

Second, the probability density function of a continuous distribution can be approximated by a direct placement of the Dirac mixture components preserving certain distribution characteristics. A well-known example for this approach is the Unscented Kalman Filter [10], which places $2N+1$ Dirac components into an N-dimensional state space.

Approximating probability distributions by Dirac mixtures can be done efficiently using classical distribution representations in the univariate case [13]. Our approach towards computing a Dirac Mixture approximation of the multivariate Gaussian distribution is based on minimizing a modification of the Cramér-von Mises distance. This modification incorporates Localized Cumulative Distributions [8] (LCDs) as a replacement for the classical cumulative distribution function and it can be applied to the multivariate case. This framework was already applied for reducing the number of dirac components for efficient control[5].

In this paper, we propose an efficient way of computing a Dirac mixture approximation of the Gaussian distribution. This is done by minimizing the modification of the Cramér-von Mises distance. This has originally been done by means of numerical quadrature [7]. This work presents an efficient way for computing a Dirac mixture approximation of the standard normal densities, which is later used for approximating an arbitrary Gaussian distribution. By using this strategy, applications in real-time environments are possible.

The remainder of this paper is structured as follows. In Sec. II, the Localized Cumulative Distribution is presented and applied within a modification of the classical Cramér-von Mises distance. Computation the modification of the Cramér-von Mises distance for the combination of a Dirac mixture and a Gaussian distribution is done in Sec. III. In Sec. IV, this computation is applied to approximating an Gaussian distribution using the Cholesky decomposition. The paper is concluded in Sec. V. Proofs of the new theorems are given in the appendix.

[1] I. Gilitschenski and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Germany (email: gilitschenski@kit.edu, uwe.hanebeck@ieee.org).

II. Localized Cumulative Distributions in a modified Cramér-von Mises Distance

A. Localized Cumulative Distribution

Using the classical cumulative distribution function in the multivariate case is often not meaningful. In one dimension, there are two meaningful definitions of a cumulative distribution function for a random variable X, which are $\mathbf{P}(X \leq a)$ or $\mathbf{P}(X > a)$ for a given $a \in [0, 1]$. They can obviously be simply computed from each other. Already in two dimensions this is not the case, where four possible definitions exist.

Localized Cumulative Distributions (LCDs) avoid this problem by describing a probability density in a localized way; that is, the probability mass around each point is considered with a certain degree of magnitude. In many applications an LCD presents a suitable replacement for the CDF. Here, LCDs are used for approximation of probability distributions. First, we give a definition of LCDs and discuss some of their properties. Then, we use them to introduce a modification of the Cramér-von Mises Distance. Further discussion of LCDs can be found in [8].

Definition II.1 (Localized Cumulative Distribution) *Let $g(\underline{x})$ be an n-dimensional density function. The corresponding Localized Cumulative Distribution is defined as*

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

where $\underline{b} \in \mathbb{R}^n_+$ and a suitable (symmetric & integrable) kernel $K(\underline{x} - \underline{m}, \underline{b})$. Here, $\underline{m}$ characterizes the location of the kernel and $\underline{b}$ its size.

The location $\underline{m}$ of the kernel describes the "center of interest" around which the probability mass shall be considered. The size $\underline{b}$ basically can be thought of as the size of the region around $\underline{m}$ having a strong influence on the LCD.

The kernel function $K(\underline{x} - \underline{m}, \underline{b})$ determines in what way the probability mass around $\underline{m}$ influences the value of the LCD. We focus our further attention on separable kernels

$$K(\underline{x} - \underline{m}, \underline{b}) = \prod_{k=1}^{n} K(x^{(k)} - m^{(k)}, b^{(k)}) \ .$$

Furthermore, it will be sufficient to consider the probability mass equally in each direction. Thus, the size can be described by the scalar $b \in \mathbb{R}_+$. The kernel used in the remainder of this paper is of Gaussian type. It is defined by

$$K(x^{(k)} - m^{(k)}, b) = \exp\left(-\frac{1}{2}\frac{x^{(k)} - m^{(k)}}{b^2}\right) \ .$$

B. A Modification of the Cramér-von Mises Criterion

The classical Cramér-von Mises criterion was originally proposed to test whether an empirical probability distribution F_1 fits an assumed cumulative distribution function F_2. It is obtained by computing

$$\int_{\mathbb{R}} (F_1(x) - F_2(x))^2 \, \mathrm{d}F_1(x) \ .$$

Many extensions were made to the original criterion. A major contribution in the field of testing was made by Anderson [2], who proposed using this distance measure for testing whether two samples were drawn from the same distribution. Several authors (see, e.g., [12], [4]) proposed the usage of weights in this distance measure. Our proposed modification of this distance shall enable the comparison of similarity of multivariate distributions, no matter whether these distributions are continuous or discrete.

Definition II.2 (Modified Cramér-von Mises Distance) *The modified Cramér-von Mises distance D between two LCDs $\tilde{F}(\underline{m}, b)$ and $F(\underline{m}, b)$ is given by*

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

where $w(b)$ is a suitable weighting function.

III. Efficient Representation for the Modified Cramér-von Mises Distance

Computing the Dirac mixture approximation of a Gaussian distribution is carried out by minimizing the modified Cramér-von Mises distance. In earlier works [7], numerical integration techniques were used for this computation.

We simplify the earlier computation for the standard normal case by deriving efficiently computable formulas for the modification of the Cramér-von Mises distance and its derivative. For the approximation of a standard normal distribution by a Dirac mixture, we need both the LCD representation of Dirac mixture distributions and the LCD representation of standard normal distributions.

The LCD of an N-dimensional Dirac mixture consisting of L components with respective weights w_i is given by

$$F(\underline{m}, b) = \sum_{i=1}^{L} w_i \prod_{k=1}^{N} \exp\left(-\frac{1}{2}\frac{x_i^{(k)} - m^{(k)}}{b^2}\right) \ .$$

The LCD of a N-dimensional standard normal distribution is given by

$$\tilde{F}(\underline{m}, b) = \frac{b^N}{\left(\sqrt{1+b^2}\right)^N} \prod_{k=1}^{N} \exp\left(-\frac{1}{2}\frac{(m^{(k)})^2}{1+b^2}\right) \ .$$

A. Computation of the Modified Cramér-von Mises Distance

We compute our modification of the Cramér-von Mises distance using the weighting function

$$w(b) = \begin{cases} b^{1-N} \ , & b \in [0, b_{\max}] \\ 0 \ , & \text{otherwise} \end{cases} \ .$$

Together with the above LCDs, this weighting function gives rise to an integral representation of the modification of the Cramér-von Mises distance, which is stated in the following theorem.

Theorem III.1 *For the LCDs of the Gaussian $\tilde{F}(\underline{m}, b)$ and the Dirac mixture $F(\underline{m}, b)$ and the weighting function*

$$w(b) = \begin{cases} b^{1-N} \ , & b \in [0, b_{max}] \\ 0 \ , & otherwise \end{cases} \ ,$$

the distance D can be represented as

$$D = D_1 - 2D_2 + D_3 \text{ with } D_i = \int_{\mathbb{R}_+} w(b) P_i \, \mathrm{d}b \ ,$$

where

$$P_1 = \pi^{\frac{N}{2}} b^{2N} \prod_{k=1}^{N} \frac{1}{\sqrt{\left(\sigma^{(k)}\right)^2 + b^2}} ,$$

$$P_2 = (2\pi)^{\frac{N}{2}} b^{2N} \left(\prod_{k=1}^{N} \frac{1}{\sqrt{\left(\sigma^{(k)}\right)^2 + 2b^2}} \right) \sum_{i=1}^{L} w_i \exp\left(-\frac{1}{2} \sum_{k=1}^{N} \frac{\left(x_i^{(k)}\right)^2}{\left(\sigma^{(k)}\right)^2 + 2b^2} \right) ,$$

$$P_3 = \pi^{\frac{N}{2}} b^{N} \sum_{t=1}^{L} \sum_{j=1}^{L} w_i w_j \exp\left(-\frac{1}{2} \frac{T_{i,j}}{2b^2} \right)$$

and

$$T_{i,j} = \sum_{k=1}^{N} \left(x_i^{(k)} - x_j^{(k)} \right)^2 .$$

A proof of this theorem is given in [7]. The variable $b_{\max}$ characterizes the area around each point, which shall be considered in the computation of the modification of the Cramér-von Mises distance.

Numerical integration techniques can be used for computing D. Our goal is to simplify this computation. The integral D_3 is given by

$$D_3 = \pi^{\frac{N}{2}} \sum_{i=1}^{L} \sum_{j=1}^{L} w_i w_j \times \frac{b_{\max}^2}{2} \exp\left(-\frac{1}{2} \frac{T_{i,j}}{2b_{\max}^2} \right) + T_{i,j} \operatorname{Ei}\left(-\frac{1}{2} \frac{T_{i,j}}{2b_{\max}^2} \right) .$$

Here, $\operatorname{Ei}(x)$ is the exponential integral defined as

$$\operatorname{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} \, \mathrm{d}t .$$

For the remaining computation, we consider the case of a standard normal distribution (i.e., $\sigma = 1$). In this situation, the integral D_1 can easily be computed by the following theorem.

Theorem III.2 *The integral D_1 is given by*

$$D_1 = \pi^{\frac{N}{2}} I_N(b_{max}) ,$$

where $I_n(t)$ is

$$I_1(t) = \frac{1}{2} \left(t\sqrt{1+t^2} - \operatorname{arcsinh}(t) \right) ,$$

$$I_2(t) = \frac{1}{2} \left(t^2 - \log(1+t^2) \right) ,$$

$$I_{2k}(t) = 2k \left(I_2(t) - \sum_{i=1}^{k-1} \frac{t^{2i} (\sqrt{1+t^2})^{2-2i}}{2i(2i-2)} \right) - \frac{t^{2k} (\sqrt{1+t^2})^{2-2k}}{(2k-2)} ,$$

$$I_{2l+1}(t) = (2l+1) \left(I_1(t) - \sum_{i=1}^{l-1} \frac{t^{2i+1} (\sqrt{1+t^2})^{1-2i}}{(2i+1)(2i-1)} \right) - \frac{t^{2l+1} (\sqrt{1+t^2})^{1-2l}}{(2l-1)} ,$$

for arbitrary $k, l \in \mathbb{N}$, where $k \geq 2$, $l \geq 1$.

PROOF. See Appendix A. ■

With this result, the integral D_1 can be computed directly. A similar, direct computation is not possible for D_2. The following result suggests how the computation of D_2 can be simplified when the dimension N is even. D_2 can be decomposed into its weighted summands, which shall be named $D_2^{(i)}$, that is $D_2 = (2\pi)^{\frac{N}{2}} \sum_{i=1}^{L} w_i D_2^{(i)}$. Using this notation, we can state our next result.

Theorem III.3 *If $N = 2\,m$ for $m \in \mathbb{N}$, the integral $D_2^{(i)}$ is given by*

$$D_2^{(i)} = J_{m,m}(b_{max}) - J_{m,m}(0)$$

with

$$J_{k,k}(b) = \frac{1}{2^k} \sum_{j=0}^{k} (-1)^j \binom{k}{j} J_{0,j}(b) ,$$

and the base cases

$$J_{0,0}(b) = \frac{1+2b^2}{4} \exp\left(-\frac{c}{2+4b^2} \right) + \frac{c}{8} \operatorname{Ei}\left(-\frac{c}{2+4b^2} \right) ,$$

$$J_{0,1}(b) = -\frac{1}{4} \operatorname{Ei}\left(-\frac{c}{2+4b^2} \right) ,$$

$$J_{0,l}(b) = \exp\left(-\frac{c}{2+4b^2} \right) \times \sum_{j=2}^{l} \frac{(l-2)!\; 2^{l-j-1}}{(l-2)!\; c^{l-j+1}\; (1+2b^2)^{j-2}} ,$$

for arbitrary $l \in \mathbb{N}$ with $l \geq 2$ and $c = \sum_{k=1}^{N} \left(x_i^{(k)} \right)^2$.

PROOF. See Appendix B. ■

Using this representation seems to be more complicated than using a quadrature algorithm on the original integral. This, however, is not the case. First, there are efficient algorithms for computating the exponential integral. Second, the exponential integral can be approximated for large values of $b_{\max}$.

B. Efficient Representation of the Gradient

Minimizing D with respect to the position of the Dirac components can be performed using a numerical optimization method. Usually, a finite difference method is used to approximate the derivative of D. In order to speed up the optimization procedure, our goal is to derive an efficiently computable representation of the gradient of D. The following theorem shows how the computation of the gradient can be decomposed into computing two integrals.

Theorem III.4 *The gradient for the general distance measure in Theorem III.1 with respect to the locations of the Dirac components is given by*

$$G_\xi^{(\eta)} = \frac{\delta D}{\delta x_\xi^{(\eta)}} = G_\xi^{(\eta,1)} + G_\xi^{(\eta,2)}$$

with

$$G_\xi^{(\eta,1)} = 2\,(2\,\pi)^{\frac{N}{2}}\,w_\xi\,x_\xi^{(\eta)} \int_0^{b_{max}} \frac{b^{N+1}}{\left(\sigma^{(\eta)}\right)^2 + 2\,b^2}$$
$$\left(\prod_{k=1}^{N} \frac{1}{\sqrt{\left(\sigma^{(k)}\right)^2 + 2\,b^2}}\right) \exp\left(-\frac{1}{2}\sum_{k=1}^{N} \frac{\left(x_\xi^{(k)}\right)^2}{\left(\sigma^{(k)}\right)^2 + 2\,b^2}\right) \mathrm{d}b$$

and

$$G_\xi^{(\eta,2)} = -\pi^{\frac{N}{2}}\,w_\xi \sum_{i=1}^{L} w_i \left(x_\xi^{(\eta)} - x_i^{(\eta)}\right) \int_0^{b_{max}} \frac{1}{b} \exp\left(-\frac{1}{2}\frac{T_{\xi,i}}{2\,b^2}\right) \mathrm{d}b$$

for component index $\xi = 1, \ldots, L$ and dimension index $\eta = 1, \ldots, N$ with

$$T_{\xi,i} = \sum_{k=1}^{N} \left(x_\xi^{(k)} - x_i^{(k)}\right)^2 \quad .$$

This theorem was also proven in [7]. $G_\xi^{(\eta,2)}$ was also computed there and is given by

$$G_\xi^{(\eta,2)} = \frac{\pi^{\frac{N}{2}}}{2} w_\xi \sum_{i=1}^{L} w_i (x_\xi^{(\eta)} - x_i^{(\eta)})\,\mathrm{Ei}\left(-\frac{1}{2}\frac{T_{\xi,i}}{2b_{max}^2}\right) \quad .$$

The integral $G_\xi^{(\eta,1)}$ can be computed in a similar way as Theorem III.3. Thus, a simple representation is also restricted to the even-dimensional case. Again, we decomose $G_\xi^{(\eta,1)}$ into its weighted summands $G_\xi^{(\eta,1,i)}$, that is $G_\xi^{(\eta,1)} = 2(2\pi)\sum_{i=1}^{L} w_\xi G_\xi^{(\eta,1,i)}$.

Theorem III.5 *If $N = 2\,k$ for $k \in \mathbb{N}$, the Integral $G_\xi^{(\eta,1)}$ is given by*

$$G_\xi^{(\eta,1,i)} = x_\xi^{(\eta)}(J_{k,k+1}(b_{max}) - J_{k,k+1}(0))$$

with

$$J_{k,k+1}(b) = \frac{1}{2^k}\sum_{j=0}^{k}(-1)^j \binom{k}{j} J_{0,j+1}(b) \ ,$$

where the base cases $J_{0,i}(b)$ are the same as in Theorem III.3.

Together with our previous result on the computation of $G_\xi^{(\eta,2)}$, this theorem gives us the ability to speed up the optimization. In a practical implementation, a further enhancement can be achieved by using precomputed exponential integral values at each optimization step, because the values of the exponential integrals computed during the computation of D also need to be computed for $G_\xi^{(\eta)}$.

IV. Dirac Mixture Approximation of the Gaussian Distribution

For approximating arbitrary Gaussian distributions, we make use of the results presented in the preceding section. For this generalization, the Mahalanobis transform [9] will be used. This will introduce some suboptimality and reduce the need for repeated numerical integration compared to a direct approximization as presented in [7]. As a first step, it is helpful to take a look at the Mahalanobis transform. A matrix square-root is taken in order to achieve the desired covariance.

Theorem IV.1 (Mahalanobis Transform) *Let $\mathbf{C} \in \mathbb{R}^{n\times n}$ be a covariance matrix (i.e., symmetric and positive definite), $\underline{\mu} \in \mathbb{R}^n$, and $\underline{Y} \sim \mathcal{N}(0, \mathbf{I}_n)$. Then $\mathbf{C}^{1/2}$ exists and*

$$\underline{X} := \mathbf{C}^{1/2}\underline{Y} + \underline{\mu}$$

is a $\mathcal{N}(\underline{\mu}, \mathbf{C})$ distributed random vector.

Proof. See [9], Theorem 4.5. ■

This transform can also be applied to each of our Dirac components. Thus, our method for computing a Dirac mixture approximation of $\mathcal{N}(\underline{\mu}, \mathbf{C})$ can be summarized in the following three steps.

1) Compute a Dirac mixture approximation of the standard normal distribution with the desired number of components in the desired dimensionality.
2) Compute the matrix square-root of $\mathbf{C}$.
3) Apply the Mahalanobis transform to each Dirac component and correct the mean of each component by adding $\underline{\mu}$.

We apply this sampling technique to the estimation problem in a nonlinear dynamic system. In this scenario, our deterministic particle positioning leads to a faster convergence rate of the mean compared to random sampling with the same amount of particles. This is due to the fact, that a deterministic sampling method is capturing the shape of a probability distribution better than simple random sampling.

The step involving numerical computation for the exponential integral has only to be done once in advance. As long as the number of used particles stays the same, there is no need for recomputing the whole approximation. The Mahalanobis transform is applied at each time step respectively. The Cholesky decomposition can be used for the computation of this transform.

The performance of the overall optimization procedure for minimizing D with respect to the positions of the Dirac components is highly dependent on the optimization algorithm. The computation of the integral D is in $\mathcal{O}(L^2N)$ for a fixed problem dimension, where L is the number of Dirac components and N is the number of dimensions. The driving force behind this computational complexity is D_3.

In the even-dimensional case, there are several ways to enhance performance. They are based on the fact, that all numerical computations in this situation come down to computing exponential integrals. Besides reducing the number of computations by reusing some results from the computation of D for the computation of its gradient, a more promising strategy can be applied. Numerical computations can be completely avoided by using a meaningful approximation [1] to the exponential integral, such as

$$\mathrm{Ei}(x) \approx \gamma + \ln|x| + x \ , \qquad x > 0 \ ,$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

Example IV.1 *Our goal is to approximate the two-dimensional Gaussian density $\mathcal{N}(\underline{0}, \mathbf{C})$, where $\mathbf{C}$ is the covariance matrix*

$$\mathbf{C} = \begin{pmatrix} 2.29 & 0.4 \\ 0.4 & 2.29 \end{pmatrix} \ .$$

In the first row of Figure 1 shows the approximation of the standard normal density for a mixture of 8, 12, 14, 16 Dirac components respectively. In the second row, the same approximations are shown after retransforming the Dirac Mixtures with a Cholesky matrix square-root of $\mathbf{C}$.

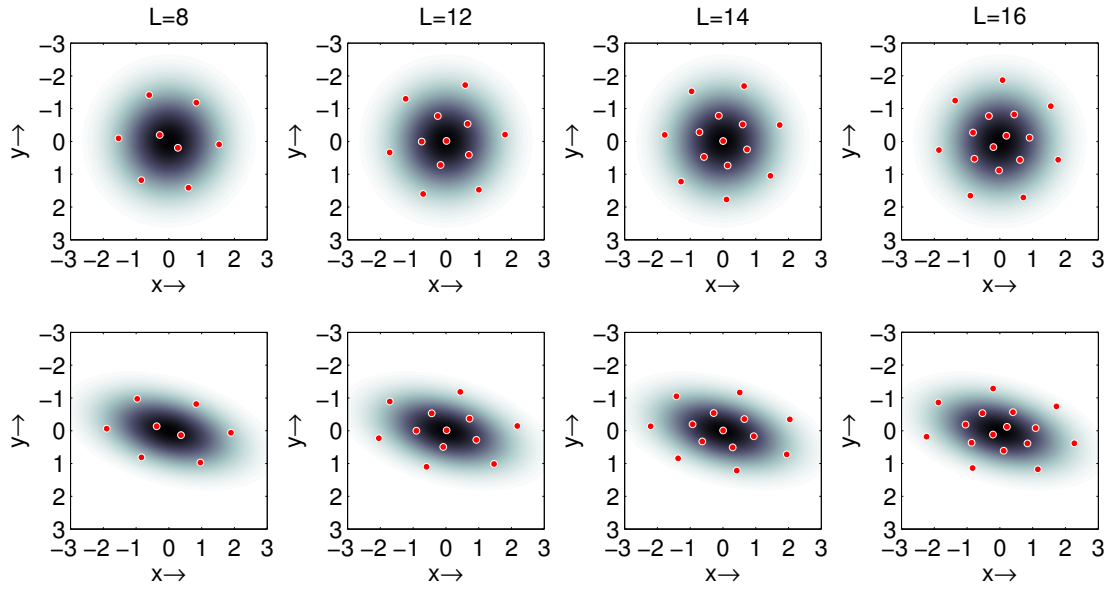

Fig. 1. Dirac Mixture approximations of a standard normal distribution (first line) and the respective approximations after considering the covariance matrix given in Example IV.1.

V. Conclusion

In this paper, an efficient way of approximating the standard normal distribution with a mixture of Dirac components was proposed. This was performed by efficiently computing a modification of the Cramér-von Mises distance, which is based on a localized replacement of the classical cumulative distribution function. We showed a way to approximate the modification of the Cramér-von Mises distance and its gradient in closed-form for the even-dimensional case and thus reduced the computational burden. The presented approach uses the Cholesky decomposition. It can also be modified to using other decompositions when generalizing the computation of the Dirac mixture approximation of a standard Gaussian distribution to the general Gaussian case.

Acknowledgment

This work was partially supported by a grant from the German Research Foundation (DFG) within the Research Training Group RTG 1194 “Self-organizing Sensor-Actuator-Networks”.

Appendix

A. Proof of Theorem III.2

For computing the integral $D1$, it is necessary to compute

$$I_n(t) := \int_0^t \frac{b^{n+1}}{\left(\sqrt{1+b^2}\right)^n} \,\mathrm{d}b \;,$$

where $n \in \mathbb{N}$. The base cases can be computed directly. For $I_1(t)$, we use integration by parts. The integrand is separated into $f(b) = b/2$ and $g(b) = \sqrt{1+b^2}$. It follows

$$\begin{aligned} I_1(t) =& \int_0^t \frac{b^2}{\sqrt{1+2b^2}} \,\mathrm{d}b = \int_0^t f(b)g'(b) \,\mathrm{d}b \\ =& \left[\frac{1}{2} b\sqrt{1+b^2}\right]_0^t - \int_0^t \frac{1}{2\sqrt{1+b^2}} \\ =& \frac{1}{2} t\sqrt{1+t^2} - \frac{\operatorname{arcsinh}(t)}{2} \;. \end{aligned}$$

For $I_2(t)$, we obtain

$$\begin{aligned} I_2(t) =& \int_0^t \frac{b^3}{1+b^2} \,\mathrm{d}b = \int_0^t \frac{b(b^2+1-1)}{1+b^2} \,\mathrm{d}b \\ =& \int_0^t b - \frac{b}{1+b^2} \,\mathrm{d}b = \frac{t}{2} - \frac{1}{2}\log(1+t^2) \;. \end{aligned}$$

For considering the case $n > 2$, let $f(b) := \sqrt{1+b^2}$ and $g(b) := b^n(1+b^2)^{(-n+1)/2}$. The derivatives of f and g are $f'(b) = b(1+b^2)^{-1/2}$ and

$$g'(b) = \frac{nb^{n-1}}{(\sqrt{1+b^2})^{n-1}} - \frac{(n-1)b^{n+1}}{(\sqrt{1+b^2})^{n-1}} \;.$$

The integrand can be separated into $f'(b)g(b)$, Using integration by parts, we obtain

$$\begin{aligned} I_n(t) =& \int_0^t \frac{b^{n+1}}{\left(\sqrt{1+b^2}\right)^n} \,\mathrm{d}b = \int_0^t f'(b)g(b) \,\mathrm{d}b \\ =& \left[\frac{b^n}{(1+b^2)^{n-2}}\right]_0^t \\ & - \int_0^t \frac{nb^{n-1}}{(\sqrt{1+b^2})^{n-2}} - \frac{(n-1)b^{n+1}}{(\sqrt{1+b^2})^n} \,\mathrm{d}b \\ =& \frac{t^n}{(1+t^2)^{n-2}} - nI_{n-2}(t) + (n-1)I_n(t) \;. \end{aligned}$$

This is used to compute a recursion formula

$$I_n(t) = \frac{1}{n-2}\left(nI_{n-2}(t) - \frac{t^n}{(1+t^2)^{n-2}}\right) \;.$$

Resolving this recursion yields the formulas for $I_{2k}(t)$ and $I_{2l+1}(t)$. ■

B. Proof of Theorem III.3

We define

$$J'_{j,k}(b) := \frac{b^{2j+1}}{(1+2b^2)^k} \exp\left(-\frac{1}{2}\frac{c}{1+2b^2}\right) \;.$$

In the above equation and the remainder of the proof, we assume $j, k \in \mathbb{N}_0$. Our goal is to compute the antiderivative of $J'_{k,k}(b)$. For $j, k \geq 1$, it holds

$$\frac{b^{2j+1}}{(1+2b^2)^k} = \frac{b^{2j-1}}{2(1+2b^2)^{k-1}} - \frac{b^{2j-1}}{2(1+2b^2)^k}$$

Thus, we observe the recursive relation

$$J'_{j,k}(b) = \frac{1}{2}\left(J'_{j-1,k-1}(b) - J'_{j-1,k}(b)\right) \ . \tag{1}$$

Representing this recursion in a triangular way leads to a geometric scheme based on Pascal's triangle. Thus, the equation for $J_{k,k}(b)$ is obtained.

For computing $J_{0,j}(b)$ with $j \geq 2$, we use integration by parts. Therefore, we define

$$f(b) := \frac{1}{2c(1+2b^2)^{j-2}} \quad , \quad g(b) := \exp\left(\frac{c}{2+4b^2}\right)$$

and obtain

$$\begin{aligned}\int_0^t J'_{0,j}(b)\ \mathrm{d}b &= \int_0^t f(b)g'(b)\ \mathrm{d}b \\ &= \left[\frac{\exp\left(-\frac{c}{2+4b^2}\right)}{2c(1+2b^2)^{j-2}}\right]_0^t \\ &\quad - \int_0^t \frac{4b(2-j)\exp\left(-\frac{c}{2+4b^2}\right)}{2c(1+2b^2)^{j-1}}\ \mathrm{d}b \ .\end{aligned}$$

This yields the recursive equation

$$J_{0,j}(b) = \left(\frac{\exp\left(-\frac{c}{2+4b^2}\right)}{2c(1+2b^2)^{j-2}}\right) + \frac{2(j-2)}{c} J_{0,j-1}(b) \ .$$

The base case $J_{0,2}(b)$ is straight forward

$$\begin{aligned}\int_0^t J'_{0,2}(b)\ \mathrm{d}b &= \int_0^t \frac{b}{2+4b^2}\exp\left(-\frac{c}{(1+2b^2)^2}\right)\ \mathrm{d}b \\ &= \frac{1}{2c}\int_0^t \frac{2cb}{(1+2b^2)^2}\exp\left(-\frac{c}{2+4b^2}\right) \\ &= \frac{1}{2c}\int_0^t \frac{8cb}{(2+4b^2)^2}\exp\left(-\frac{c}{2+4b^2}\right) \\ &= \left[\frac{1}{2c}\exp\left(-\frac{c}{2+4b^2}\right)\right]_0^t \ .\end{aligned}$$

Using this base case, the recursion above can be resolved, which yields our formula for $J_{0,j}(b)$ for $j \geq 2$.

For computing the antiderivative $J_{0,1}(b)$, substitution is used. We define $\varphi(b) := -c(2+4b^2)^{-1}$. Using this defintion, we obtain

$$\begin{aligned}\int_0^t J'_{0,j}(b)\ \mathrm{d}b &= \int_0^t \frac{b}{1+2b^2}\exp\left(-\frac{c}{2+4b^2}\right)\ \mathrm{d}b \\ &= -\frac{1}{4}\int_0^t \frac{\varphi'(b)}{\varphi(b)}\exp(\varphi(b))\ \mathrm{d}b \\ &= -\frac{1}{4}\int_{\varphi(0)}^{\varphi(t)} \frac{\mathrm{e}^x}{x}\ \mathrm{d}x \\ &= -\frac{1}{4}\left[\mathrm{Ei}\left(-\frac{c}{2+4b^2}\right)\right]_0^t \ .\end{aligned}$$

Handling $J'_{0,0}(b)$ is done in two steps. First, integration by parts is used. This is done by introducing $f(b) = b^2/2$ and $g(b) = \exp(-c(2+4b^2)^{-1})$. Second, the computation is reduced to computing $J_{0,1}(b)$ and $J_{0,2}(b)$. Putting it all together yields

$$\begin{aligned}\int_0^t J'_{0,0}(b)\ \mathrm{d}b &= \int_0^t f'(b)g(b)\ \mathrm{d}b \\ &= \left[\frac{b^2}{2}\exp\left(-\frac{c}{2+4b^2}\right)\right]_0^t \\ &\quad - \int_0^t \frac{8cb^3}{2(2+4b^2)^2}\exp\left(-\frac{c}{2+4b^2}\right)\ \mathrm{d}b \ .\end{aligned}$$

The integrand in the last integral is $c \cdot J'_{1,2}(b)$. Using (1), we get

$$\begin{aligned}J_{0,0}(b) &= \frac{b^2}{2}\exp\left(-\frac{c}{2+4b^2}\right) - \frac{c}{2}J_{0,1}(b) + \frac{c}{2}J_{0,2}(b) \\ &= \frac{1+2b^2}{4}\exp\left(-\frac{c}{2+4b^2}\right) + \frac{c}{8}\,\mathrm{Ei}\left(-\frac{c}{2+4b^2}\right) \ ,\end{aligned}$$

which completes the proof. ■

References

[1] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1965.

[2] T. W. Anderson, "On the Distribution of the Two-Sample Cramer-Von Mises Criterion," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1148–1159, 1962, mathematical Reviews number (MathSciNet): MR145607; Zentralblatt MATH identifier: 0116.37601.

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

[4] D. D. Boos, "Minimum Distance Estimators for Location and Goodness of Fit," *Journal of the American Statistical Association*, vol. 76, no. 375, pp. 663–670, Sep. 1981.

[5] C. Chlebek, A. Hekler, and U. D. Hanebeck, "Stochastic Nonlinear Model Predictive Control Based on Progressive Density Simplification," in *Proceedings of the 51st IEEE Conference on Decision and Control (CDC 2012)*, Maui, Hawaii, USA, Dec. 2012.

[6] G. Evensen, "The Ensemble Kalman Filter: Theoretical Formulation and practical Implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[7] 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.

[8] 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.

[9] W. Hardle and L. Simar, *Applied Multivariate Statistical Analysis*, 2nd ed. Springer Berlin Heidelberg, 2008.

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

[11] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME – Journal of Basic Engineering*, no. 82 (Series D), pp. 35–45, 1960.

[12] O. Oztürk and T. P. Hettmansperger, "Generalised Weighted Cramér-von Mises Distance Estimators," *Biometrika*, vol. 84, no. 2, pp. 283–294, Jun. 1997.

[13] 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.

[14] A. N. Shiryaev, *Probability*, 2nd ed. Springer, 1995.

[15] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, 1995.