

Optimal Quantization of Circular Distributions

Igor Gilitschenski¹, Gerhard Kurz², Uwe D. Hanebeck², and Roland Siegwart¹

Abstract—In this work, a new method for approximating circular probability distributions by a mixture of weighted discrete samples is proposed. Particularly, the wrapped normal distribution, the von Mises distribution, and the Bingham distribution are considered. The approximation approach is based on formulating a quantizer and a global optimality measure, which can be optimized directly. Furthermore, a relationship between the Bingham distribution and the von Mises distribution are formulated showing that it is sufficient to approximate a von Mises distribution with suitably chosen parameters in order to obtain an optimal approximation of the Bingham distribution. The resulting approximation is of particular interest for filtering applications, because the involved optimality measure gives rise to a general error estimate in propagation of uncertainties through nontrivial functions in the circular domain.

I. INTRODUCTION

A broad range of applications in perception and sensing requires estimation of quantities that are defined on inherently nonlinear state-spaces. These quantities involve direction of the wind, poses of a robot, or direction of arrival measurements in signal processing. Linear approximation of such state-spaces may be employed in cases of low noise. This gives rise to using well-known linear estimation techniques such as the Kalman filter. However, in cases involving high levels of noise, neglecting the underlying geometric structure of the manifold becomes infeasible.

For the circle, which is considered in this work, a number of different filtering techniques have been proposed that take the underlying geometry into account. This is achieved by making use of probability distributions from directional statistics [1], [2], [3], which is a subfield of statistics that considers uncertainties defined on these manifolds. For example, circular equivalents of a Kalman Filter may be based on matching trigonometric moments and are used for consideration of simple system models, particularly the identity or a shift.

Whenever system models are more complex, the continuous distribution representing the current estimate is typically approximated by a discrete distribution in order to simplify propagation. For the linear case, there exists an extensive body of research discussing different types of such approximations. However, this is not the fact when approximating continuous densities defined on the circle as state-of-the-art is limited by only using a given number of discrete values (as in the UKF) or not providing any provable error bound for the estimation procedure.

In linear state-spaces, one of the approaches overcoming these limitations is based on stochastic quantization. Quantization has been well known for many years in the context of signal processing and information compression [4]. Recently, there has been an emphasis on the quantization of probability distributions [5]. For the Gaussian case, this has been proposed in [6], which is based on a quadratic quantization approach.

In this work, we investigate optimal quadratic quantization of circular probability distributions. The involved optimality measure gives rise to an error bound for computing expectations of nonlinear transformations of random variables. The proposed quantizers are of considerable interest for nonlinear filtering involving circular quantities because a discrete distribution with a sufficiently large number of samples can be chosen to maintain a predefined quality level for the prediction step. This work considers the wrapped normal, the von Mises, and the Bingham distributions because of their importance in many applications involving circular data. However, the proposed approach is easily adaptable to other circular distributions and other application scenarios besides stochastic filtering, such as analysis of time series involving periodic quantities.

In summary, this paper contains the following contributions

- A novel method for approximating probability densities on the circle with applications to wrapped normal (WN) and von Mises (VM) distributions.
- Derivation of a simple relationship between the VM and Bingham distributions for easily adapting the approximations of the VM distribution to the Bingham case.
- An error bound for scenarios where the resulting sample set is applied to numerical integration for Lipschitz continuous transforms of random variables.
- Evaluation comparing the proposed approach to other state-of-the-art approaches.

The remainder of this paper is structured as follows. The next section gives an overview of related work. Some preliminaries on circular distributions and stochastic quantization are given in Sec. III. The proposed method is presented in Sec. IV and evaluated in Sec. V. Finally, the work is concluded in Sec. VI.

II. RELATED WORK

There are numerous applications involving stochastic filtering based on directional statistics. In [7], a von Mises distribution based filter is used for azimuthal speaker tracking. This distribution is also used in [8] for GNSS signal tracking. Active speaker localization using von Mises and wrapped normal distributions is discussed in [9], [10]. A multivariate variant of the wrapped normal distribution is used in context of radar signal processing in [11]. Furthermore, a discussion of circular estimation using lattice theory is presented in [12]. Early literature considering

¹ I. Gilitschenski and R. Siegwart are with the Autonomous Systems Laboratory (ASL), Institute for Robotics and Intelligent Systems, ETH Zurich, Switzerland (e-mail: igilitschenski@ethz.ch, rsiegwart@ethz.ch).

² G. Kurz and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany (e-mail: gerhard.kurz@kit.edu, uwe.hanebeck@ieee.org).

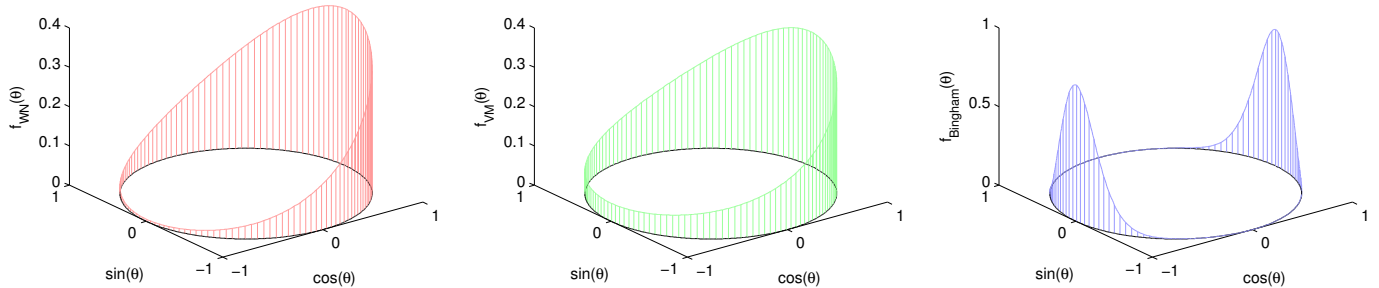


Figure 1: Density functions of the probability distributions considered in this work. All location parameters were set in a way ensuring the / one mode to be at 0. The wrapped normal density (red) has dispersion parameter $\sigma = 1$, for the von Mises density (green) $\kappa = 1$ was chosen, and the \mathbf{Z} matrix of the Bingham density (blue) was chosen as $\mathbf{Z} = \text{diag}(-10, 0)$.

filtering on nonlinear domains dates back to [13], [14] and is still an active area of research [15].

Most of the nonlinear filters discussed so far typically assume simple system models such as the identity or a mere shift. In order to tackle propagation through more complex functions sampling schemes reminiscent of the unscented Kalman filter (UKF) have been proposed in [16], [17]. Similar to the UKF, these schemes are based on matching moments for obtaining a discrete approximation of the underlying continuous distribution. The method presented in [18] considers a situation where the underlying discrete distribution has more Dirac components than required in order to satisfy the moment constraints. Then, a distance measure is used to exploit this redundancy in order to make the discrete distribution as similar as possible (with respect to this distance measure) to its original continuous counterpart.

In contrast to these approaches, the method presented in this work does not aim at satisfying moment constraints. Instead, a quantization based method is used in order to provide a discrete approximation of the continuous density. Unless precomputation and interpolation are used, this comes at the cost of a higher computation time compared to the UKF-like approaches. However, the main advantage over state of the art is an error bound for propagation of uncertainty through Lipschitz continuous functions. Therefore, in scenarios where a Lipschitz constant is known a priori, it is possible to approximate the result of an expectation computation up to a predefined level of precision.

III. PRELIMINARIES

This section first revisits some circular probability distributions and their properties. Then, the concept of optimal quadratic quantization is introduced.

A. Circular Distributions

Circular probability distributions are inherently defined on the circle S_1 . Thus, there are several ways to parametrize them and to derive their respective probability density functions (p.d.f.). The most widely used circular distributions arise from modifying non-circular counterparts in order to restrict them to a circular domain. All distributions considered in this work are

related to the normal distribution in some sense. Their density functions are shown in Fig. 1.

One of the main reasons for the wide-spread use of the normal distribution is the central limit theorem. Thus, it is desirable to obtain a circular distribution which also appears as a limit distribution for a wide class of circular densities. An obvious approach is wrapping a normal distributed random vector $x \sim \mathcal{N}(\mu, \sigma^2)$ around an interval of length 2π . This results in the wrapped normal distribution.

Definition 1. Let $X \in [-\pi, \pi)$ be a random variable with corresponding p.d.f.

$$f_{\text{WN}}(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right)$$

where $\sigma > 0$, $\mu, x \in [-\pi, \pi)$. Then, the distribution of X is called a wrapped normal distribution and we write $X \sim \text{WN}(\mu, \sigma)$.

It can be easily seen that this distribution naturally arises as a limit distribution for a circular summation scheme of other wrapped distributions (which originally converged to the normal distribution). Furthermore, the underlying density function appears as a solution of the heat equation giving further justification for considering this distribution [19]. Similar to the normal distribution, it can be shown that the sum of two wrapped normal distributed random variables is itself a wrapped normal distributed random variable (after a modulo 2π operation).

Unfortunately, the wrapped normal distribution is lacking one of the important properties of the normal distribution. The product of two wrapped normal densities is not itself a rescaled wrapped normal density. This is particularly a problem for a typical Bayesian measurement update step in stochastic filtering. Furthermore, handling of the infinite sum within the density might be computationally inconvenient in some applications, particularly for large σ .

The von Mises distribution arises when taking a $\mathcal{N}(\mu, \mathbf{C})$ distributed random vector in \mathbb{R}^2 with $\|\underline{\mu}\| = 1$ and conditioning it to the unit circle. The definition of this distribution is usually formulated as follows.

Definition 2. Let $X \in [-\pi, \pi)$ be a random variable with corresponding p.d.f.

$$f_{\text{VM}}(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu))$$

where $\kappa \geq 0$, $\mu, x \in [-\pi, \pi)$, and $I_0(\kappa)$ is the modified Bessel function of order 0. Then, the distribution of X is called a von Mises distribution and we write $X \sim \text{VM}(\mu, \kappa)$.

The von Mises distribution can be used for a closed form measurement update step, thus it is of particular interest for stochastic filtering. The same property is also maintained by the Bingham distribution [20]. It is an antipodally symmetric equivalent to the von Mises distribution and it is constructed in a similar way. Instead of requiring $|\underline{\mu}| = 1$, we require $\underline{\mu} = \underline{0}$. Thus, the resulting density has 180° symmetry.

Definition 3. Let $X \in [-\pi, \pi)$ be a random variable with corresponding p.d.f.

$$f_{\text{Bingham}}(\theta) = \frac{1}{N(\mathbf{Z})} \exp(v(\theta)^\top \mathbf{M} \mathbf{Z} \mathbf{M}^\top v(\theta))$$

where $v(\theta) = (\cos(\theta), \sin(\theta))^\top$, $\mathbf{M} \in \mathbb{R}^{2 \times 2}$ is orthogonal, and $\mathbf{Z} \in \mathbb{R}^{2 \times 2}$ is diagonal. Then, the distribution of X is called a two-dimensional Bingham distribution and we write $X \sim \text{Bingham}(\mathbf{Z})$.

The particular interest in the Bingham distribution stems from its universal applicability not only to the circular case but also to representing uncertain orientations using its four-dimensional variant. Besides the filtering applications mentioned in the introduction, the Bingham distribution was also applied in the analysis of textures [21] and shapes.

B. Optimal Quadratic Quantization

The basic idea of quantization in digital signal processing is to reduce the number of values (which is typically uncountable) of some signal to a countable set. Optimal quantization of a continuous probability distribution defined on a space $\Omega \subset \mathbb{R}^n$ addresses the same problem. The basic goal considered in this work is the reduction of the number of values describing a probability distribution to some finite set and thus approximation of a continuous probability distribution by a discrete distribution. Quantization of probability distributions is discussed in greater depth in [5].

Voronoi quantizers can be used for the considered approximation problem. Such a quantization can be thought of as a vector $\underline{x} \in \Omega^N$ consisting of means of Voronoi cells (in the scenario considered in this paper, it will always be assumed that this vector is sorted, i.e., $x_i < x_{i+1}$). The Voronoi cells defined by this vector are given by

$$C_i = \{x \in \Omega : d(x, x_i) < d(x, x_j), \forall j \neq i\} ,$$

where $d(\cdot, \cdot)$ is a suitable choice for the underlying distance. Thus, the following notation defining a quantizer is meaningful.

$$V_{\underline{x}}(X) = x_i \text{ for } x \in C_i . \quad (1)$$

For the purpose of this work, the boundary between two neighbouring sets C_i and C_j may be arbitrarily added to either C_i or C_j .

There are many ways to define a quantization error describing the quality of a given quantizer. Thus, a general definition is given here.

Definition 4. For an N -component Voronoi quantizer $V_{\underline{x}}$ (i.e., $\underline{x} \in \Omega^N$) of a probability distribution \mathbf{P} , the N -th quantization error for \mathbf{P} of order p is defined by $\mathbb{E}(\|X - V_{\underline{x}}(X)\|^p)$, where X is a \mathbf{P} -distributed random variable.

In the case where $\|\cdot\|$ is the Euclidean norm and $p = 2$, the resulting error measure is also known as (quadratic) distortion and denoted by $D_{\underline{x}}$. Using (1), this error measure can be formulated as

$$\mathbb{E}(\|X - V_{\underline{x}}(X)\|^p) = \int_{\Omega} \min_{1 \leq i \leq N} \|x_i - X\|^p d\mathbf{P} .$$

Finally, the problem of optimal quantization for \mathbf{P} of order p can be formulated as finding

$$\underline{x}^* = \arg \min_{\underline{x} \in \Omega^N} \mathbb{E}(\|X - V_{\underline{x}}(X)\|^p) .$$

A quantizer $V_{\underline{x}}$ can be used to define a discrete probability distribution where each component x_i is assigned probability weight

$$w_i = \mathbf{P}(X \in C_i) . \quad (2)$$

This can be applied to deriving error bounds for approximate numerical integration of some function $g : \Omega \rightarrow \Omega'$, e.g., by using an optimal quantizer with respect to the quadratic distortion error measure. The resulting approximate integral is given by

$$\begin{aligned} \mathbb{E}(g(X)) &\approx \mathbb{E}(g(V_{\underline{x}}(X))) \\ &= \sum_{i=1}^N \int_{C_i} g(x_i) d\mathbf{P} = \sum_{i=1}^N w_i g(x_i) . \end{aligned}$$

In this scenario, the following error bound [6], [22] can be obtained for a Lipschitz continuous function g with Lipschitz constant L_g

$$\begin{aligned} &|\mathbb{E}(g(X)) - \mathbb{E}(g(V_{\underline{x}}(X)))|_2 \\ &\leq \mathbb{E}(\|g(X) - g(V_{\underline{x}}(X))\|_2) \\ &\leq L_g \cdot \mathbb{E}(\|X - V_{\underline{x}}(X)\|_2) \\ &\leq L_g \cdot \sqrt{D_{\underline{x}}} , \end{aligned} \quad (3)$$

where $D_{\underline{x}}$ denotes the quadratic distortion of the quantizer $V_{\underline{x}}$.

IV. QUANTIZATION OF CIRCULAR DENSITIES

A circular equivalent to the quadratic distortion will be used as the considered error measure. It is defined by

$$D_{\underline{x}} := \mathbb{E}(d(X, V_{\underline{x}}(X))^2) , \quad (4)$$

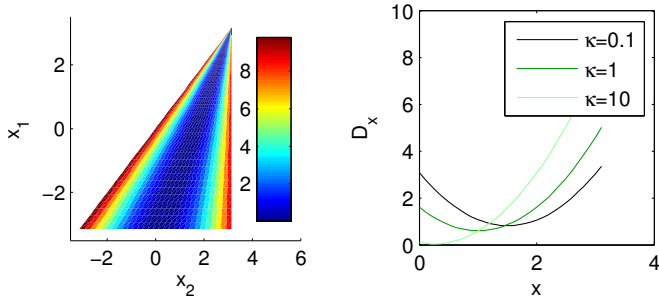


Figure 2: The first plot shows the the distortion for a 2-quantizer of a $WN(0, 1)$ distribution (where $x_1 < x_2$ is required). The second plot shows the distortion in a similar scenario for the $VM(0, \kappa)$ distribution, where the means of the Voronoi cells are symmetrically located at $-x$ and x .

where $d(\cdot, \cdot)$ is the circular equivalent of an Euclidean distance taking the periodicity of the circle into account. That is, $d(\cdot, \cdot)$ is given by

$$d(x, y) := \min\{|x - y|, 2\pi - |x - y|\}$$

for every $x, y \in [a, a + 2\pi)$ and an arbitrary domain a . This choice makes (3) adaptable to the circular case. An important application of this is the derivation of an upper bound for the error in an approximate propagation on the circular domain. It can be observed that the distortion of optimal quantizers with a fixed number of components is maximized when the considered densities approach/become the uniform distribution on the circle (this happens for $\sigma \rightarrow \infty$, $\kappa = 0$, and $\mathbf{Z} = \text{diag}(c, c)$ with arbitrary $c \in \mathbb{R}$). Thus, an optimal N -component quantizer of the uniform distribution yields an upper limit for the approximation error made by approximating the true continuous density by a discrete density obtained from quantization.

Using the 2π -periodicity of circular densities allows rewriting the error measure (4) as

$$\begin{aligned} D_{\underline{x}} &= \sum_{i=1}^N \int_{C_i} (x - x_i)^2 \cdot f(x) \, dx \\ &= \sum_{i=1}^N \int_{x_{i-1/2}}^{x_{i+1/2}} (x - x_i)^2 \cdot f(x) \, dx, \end{aligned}$$

where x_i are assumed to be in an increasing order, $x_0 = x_N - 2\pi$, $x_{N+1} = x_0 + 2\pi$ and $x_{i+1/2} = (x_i + x_{i+1})/2$.

Optimization procedures for computing an optimal quantizer can be usually speed up by providing the gradient. Thus, a general result is derived on the gradient of the distortion for continuous circular densities.

Some further notation will be needed for the formulation of the following results and their respective proofs. For a p.d.f. $f(y)$, let $F^{(1)}$ denote the antiderivative of $y^2 \cdot f(y)$, let $F^{(2)}(y)$ denote the antiderivative of $y \cdot f(y)$, and let $F^{(3)}(y)$ denote the antiderivative of the density $f(y)$ itself.

Proposition 1. Consider a circular random variable $X \in [-\pi, \pi)$ following a continuous distribution given by the density

f. Then, the gradient of the error measure (4) with respect to \underline{x} is given by

$$\frac{\partial D_{\underline{x}}}{\partial x_i} = -2 \left[F^{(2)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} + 2x_i \left[F^{(3)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}}.$$

A proof is given in Appendix A. In the following, a brief look will be taken at the considered circular distributions and some theoretical aspects of computing the distortion for a quantizer of this distributions.

A. Wrapped Normal Case

The antiderivative of a wrapped normal density can not be computed in closed form. However, due to its relationship to the Gaussian distribution it can be represented as an infinite sum involving the erf function.

Proposition 2. The distortion of a quantization $V_{\underline{x}}$ of the wrapped normal distribution $WN(0, \sigma)$ is given by

$$D_{\underline{x}} = \sum_{i=1}^L \sum_{k=-\infty}^{\infty} P_{i,k}^{(1)} - 2P_{i,k}^{(2)} + P_{i,k}^{(3)}.$$

The components $P_{i,k}^{(1)}$ are given by

$$\begin{aligned} P_{i,k}^{(1)} &= \left[\sigma^2(2k\pi - y) f_k(y) \right. \\ &\quad \left. + \left(2k^2\pi^2 + \frac{\sigma^2}{2} \right) \text{erf} \left(\frac{y + 2k\pi}{\sqrt{2}\sigma} \right) \right]_{x_{i-1/2}}^{x_{i+1/2}}, \\ P_{i,k}^{(2)} &= -x_i \left[\sigma^2 \cdot f_k(y) + k \cdot \pi \cdot \text{erf} \left(\frac{y + 2k\pi}{\sqrt{2}\sigma} \right) \right]_{x_{i-1/2}}^{x_{i+1/2}}, \end{aligned}$$

and

$$P_{i,k}^{(3)} = \left[\frac{x_i^2}{2} \text{erf} \left(\frac{y + 2k\pi}{\sqrt{2}\sigma} \right) \right]_{x_{i-1/2}}^{x_{i+1/2}}.$$

A proof is given in Appendix B. The resulting error measure for the case of $N = 2$ Voronoi cells is shown in the left plot Fig. 2. The triangular structure of the plot is due to the fact, that we assume the locations of the components to be given in an increasing order, i.e., $x_2 > x_1$.

B. Von Mises Case

A simplification occurs for the quantization of the von Mises distribution, because there is no infinite series involved in the actual density function. However, the antiderivative of the von Mises density needs to be computed numerically. It is given by

$$F_{VM}^{(3)}(x) = \frac{1}{2\pi} \left(x + \frac{2}{I_0(\kappa)} \sum_{i=1}^{\infty} I_i(\kappa) \frac{\sin(j(x - \mu))}{j} \right).$$

This series representation unfortunately involves the evaluation of a Bessel function in each summand. Furthermore, no similar representations for $F_{VM}^{(1)}$ and $F_{VM}^{(2)}$ are known so far.

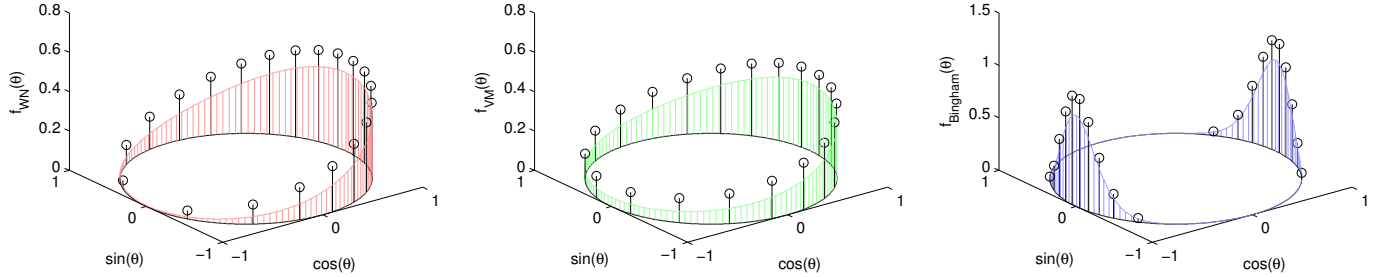


Figure 3: Optimal quantization of the $WN(0, 1)$, $VM(0, 1)$, and $Bingham(\mathbf{A}, \mathbf{Z})$ distributions, where \mathbf{A} is the same as in (6) and $\mathbf{Z} = \text{diag}(-10, 0)$. The quantizer is represented by the induced discrete probability density, where the height of the discrete components is proportional to their respective probability weights.

Algorithm 1: Optimal quantization of a Bingham distribution

Input: Number of Voronoi cells N ; Bingham distribution parameters $\mathbf{M} = (m_{i,j})$, $\mathbf{Z} = \text{diag}(z_1, z_2)$ with $z_2 > z_1$.

Output: Optimal $2N$ quantizer \underline{x}

```

/* Extract mode of Bingham */
 $\mu \leftarrow \text{atan2}(m_{2,2}, m_{1,2})$ 

/* von Mises parameter */
 $\kappa \leftarrow (z_2 - z_1)/2$ 

/* Optimal quantization of VM(0,  $\kappa$ ) */
 $\underline{y} \leftarrow \text{OptimalVonMisesQuantizer}(0, \kappa, N)$ 

/* Obtain optimal Bingham quantizer */
 $\underline{x} \leftarrow \left( \frac{y_1}{2}, \dots, \frac{y_N}{2}, \frac{y_1}{2} + \pi, \dots, \frac{y_N}{2} + \pi \right)$ 
 $\underline{x} \leftarrow ((x_1 + \mu) \bmod 2\pi, \dots, (x_{2N} + \mu) \bmod 2\pi)$ 

```

Consequently, using direct numerical integration is necessary to compute the distortion for a quantizer of the von Mises distribution. The same problem arises in the computation of its gradient using Proposition 1. However, a speed up can be achieved by leaving away the normalization constant (or precomputing it at the very beginning of the entire optimization procedure). This avoids numerous evaluations of the Bessel function. The utility function for approximating von Mises distribution using a quantizer with $N = 2$ and symmetric placement of the components $x_1 = -x_2$ is shown in the right plot of Fig. 2.

C. Bingham Case

Instead of deriving a quantizer particularly suited for the Bingham distribution, a relationship with the von Mises distribution can be used to obtain an optimal quantizer from an optimal quantizer of the von Mises distribution.

Proposition 3. *Let X be a $Bingham(\mathbf{M}, \mathbf{Z})$ distributed random variable, where $\mathbf{M} = (m_{i,j})$, $\mathbf{Z} = \text{diag}(z_1, z_2)$ with $z_2 > z_1$. Then, $2X \bmod 2\pi$ is a $VM(\theta, \kappa)$ distributed random*

variable with $\theta = (2 \cdot \text{atan2}(m_{2,2}, m_{1,2})) \bmod 2\pi$, and $\kappa = (z_2 - z_1)/2$.

This avoids the need for running an optimization method in a scenario involving multimodal densities, because an optimal $2N$ -component quantizer of a Bingham distribution can be derived from an N -component quantizer of a von Mises distribution. The whole process is described in Algorithm 1.

V. EVALUATION

Several problems need to be addressed when implementing a method to compute optimal quantizers. The starting value of the optimizer is of particular importance in finding the optimal quantizer because of the following three reasons. First, a bad starting value can result in the optimizer not finding the optimum, because components located sufficiently far away from the mean of a very peaked distribution might not change their position during the optimization process. This is a consequence of floating-point round-off errors making the computed derivative of the distortion (with respect to these components) become 0. Second, our utility measure assumes $x_i < x_{i+1}$ and thus, choosing the same value as starting point for all x_i might break this ordering, which would need to be fixed by introducing additional sort operations. Finally, a good heuristic for the starting value might significantly reduce the number of required iterations. However, in these evaluations, a very rough heuristic was used. For approximating the zero mean wrapped normal distribution the start values for the optimizer were chosen as

$$x_i = \frac{\pi(2i - 1 - N)}{N} \cdot \min(\sigma, 1), \quad i = 1, \dots, N,$$

and for the zero mean von Mises distribution the start values were chosen as

$$x_i = \frac{\pi(2i - 1 - N)}{N} \cdot \min\left(\frac{1}{\sqrt{\kappa}}, 1\right), \quad i = 1, \dots, N,$$

with $\min(\kappa^{-1/2}, 1) = 1$ for $\kappa = 0$. Approximations of distributions with a non-zero mean can be made by using a suitable shift.

Formulas from Proposition 2 were used for approximation of the distortion in the wrapped normal case. Numerical

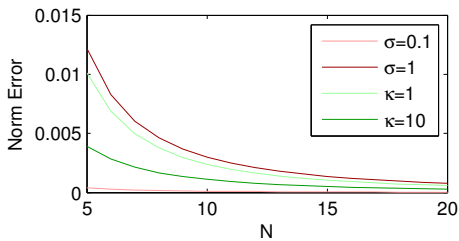


Figure 4: Norm error between the first circular moment of the discrete probability densities derived from optimal quantization (with N components) of wrapped normal (red) and von Mises (green) distributions and their true first circular moments.

integration was used for computing the distortion and its derivative in the von Mises case. The resulting optimal quantizers for some circular densities are shown in Fig. 3, where the quantizers are represented by the induced discrete probability distribution approximating the original continuous density.

The first circular moment was used to evaluate the approximation quality of the proposed quantizers. It is given by $\mathbb{E}(\exp(iX))$ and of particular importance for moment matching based parameter estimation of the involved distributions. The true first circular moment was compared with the first circular moment of the discrete distributions obtained from optimal quantization with respective probability weights as defined in (2). The results are shown in Fig. 4.

N	$\sigma = 0.1$		$\sigma = 1$	
	$D_{\underline{x}}$	iter	$D_{\underline{x}}$	iter
5	$8.0 \cdot 10^{-4}$	5	$7.7 \cdot 10^{-2}$	5
10	$2.3 \cdot 10^{-4}$	8	$2.1 \cdot 10^{-2}$	8
15	$1.1 \cdot 10^{-4}$	8	$9.8 \cdot 10^{-3}$	7
30	$2.8 \cdot 10^{-5}$	8	$2.5 \cdot 10^{-3}$	8

Table I: Simulation results for optimal quantization of wrapped normal distributions.

An overview of the optimization procedure for the wrapped normal case is given in Table I. These evaluations were computed using a straight forward implementation of the error measure (and its derivatives) in Matlab 2013b and a trust-region optimization procedure (which is the default for `fminunc`). All simulations were performed on a system with an Intel Core i7-4770 CPU. In the wrapped normal case, the computation time rises not only for a rising number of components but also for a rising σ . This is because the number of relevant terms in the infinite sum within the wrapped normal densities (and thus within the corresponding antiderivatives) is dependent on the parameter σ . The overall picture looks similar for the von Mises case (see Table II). Contrary to the wrapped normal distribution, the computation time rises for a very peaked density. This is due to the use of an adaptive integration method.

The presented quantization can be used directly in a scenario without real-time requirements, e.g., for analysis of certain types of periodical time-series. Furthermore, a better implementation (e.g., by generating compiled binary code and considering

N	$\kappa = 1$			$\kappa = 10$		
	$D_{\underline{x}}$	iter	time	$D_{\underline{x}}$	iter	time
5	$1.1 \cdot 10^{-1}$	4	0.1s	$8.8 \cdot 10^{-3}$	7	0.2s
10	$2.8 \cdot 10^{-2}$	5	0.5s	$2.5 \cdot 10^{-3}$	6	0.6s
15	$1.3 \cdot 10^{-2}$	6	1.3s	$1.2 \cdot 10^{-3}$	7	1.5s
30	$3.1 \cdot 10^{-3}$	6	5.0s	$3.2 \cdot 10^{-4}$	6	5.0s

Table II: Simulation results for optimal quantization of von Mises distributions.

symmetry) makes the proposed approach suitable for many real-time applications. Several strategies can be used to achieve a further speed up, where numerical optimization is avoided entirely. First, sub optimal quantizers can be used which are obtained by computationally efficient heuristics. In this case, the distortions of such quantizers need to be precomputed in advance in order to make a design choice about the proper number of components involved. Furthermore, computation of probability weights of the induced discrete probability distribution also involves numerical evaluations making a good heuristic choice necessary. Second, a combination of interpolation and precomputed look-up tables may be used in order to approximate an optimal quantization. In this case, error bounds for the involved interpolation method can be used to derive a bound on the suboptimality of the interpolated quantizers.

VI. CONCLUSION

In this work, Voronoi quantization of circular densities is proposed. These Voronoi quantizers give rise to an induced discrete probability distribution approximating the original continuous distribution. The proposed method can be used to guarantee a predefined degree of precision in computing transformations of circular random variables. This is of particular interest in the stochastic filtering. For future work, it is of interest to derive quantization based approximations of continuous probability densities on other manifolds, such as the manifold of orientations $SO(3)$ and the manifold of rigid-body motions $SE(3)$, because of their importance in many applications. Furthermore, use of different circular distributions and error measures might result in deriving more efficient approximation techniques.

APPENDIX

A. Proof of Proposition 1

The gradient can be decomposed into.

$$\frac{\partial D_{\underline{x}}}{\partial x_i} = G_i^{(1)} - 2G_i^{(2)} + G_i^{(3)}.$$

Furthermore, let $\tilde{x}_i := x_i$, $\tilde{x}_{i\pm 1/2} := x_{i\pm 1/2}$ for $2 < i < N$, $\tilde{x}_0 := x_L$, and $\tilde{x}_{N+1} := x_1$ (similarly $\tilde{x}_{N+1/2} := x_{1-1/2}$ and $\tilde{x}_{1-1/2} := x_{N+1/2}$). A first useful observation is that x and \tilde{x} have been chosen in a way such that

$$f(\tilde{x}_i) = f(x_i) \quad (5)$$

holds. In the next step, the $G_i^{(k)}$ are derived.

$$G_i^{(1)} = \frac{\delta}{\delta x_i} \left(\left[F^{(1)}(y) \right]_{\tilde{x}_{i-3/2}}^{\tilde{x}_{i-1/2}} + \left[F^{(1)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} + \left[F^{(1)}(y) \right]_{\tilde{x}_{i+1/2}}^{\tilde{x}_{i+3/2}} \right).$$

Use of \tilde{x} instead of x in the first and last summand is to make sure that for $i = 1$ or $i = L$ periodicity is taken into account correctly. Leaving away unnecessary terms yields

$$\begin{aligned} G_i^{(1)} &= \frac{\delta}{\delta x_i} F^{(1)}(\tilde{x}_{i-1/2}) \\ &\quad + \frac{\delta}{\delta x_i} \left(F^{(1)}(x_{i+1/2}) - F^{(1)}(x_{i-1/2}) \right) \\ &\quad - \frac{\delta}{\delta x_i} F^{(1)}(\tilde{x}_{i+1/2}) . \\ &\stackrel{(5)}{=} \frac{\tilde{x}_{i-1/2}^2 - x_{i-1/2}^2}{2} f(x_{i-1/2}) \\ &\quad + \frac{x_{i+1/2}^2 - \tilde{x}_{i+1/2}^2}{2} f(x_{i+1/2}) . \end{aligned}$$

Similar reasoning can be used for computation of $G_i^{(2)}$. This yields

$$\begin{aligned} G_i^{(2)} &= \frac{\tilde{x}_{i-1} \cdot \tilde{x}_{i-1/2} - x_i \cdot x_{i-1/2}}{2} \cdot f(x_{i-1/2}) \\ &\quad + \left[F^{(2)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} \\ &\quad + \frac{x_i \cdot x_{i+1/2} - \tilde{x}_{i+1} \cdot \tilde{x}_{i+1/2}}{2} \cdot f(x_{i+1/2}) \end{aligned}$$

And for $G_i^{(3)}$, we obtain

$$\begin{aligned} G_i^{(3)} &= \frac{\tilde{x}_{i-1}^2 - x_i^2}{2} \cdot f(x_{i-1/2}) + 2x_i \left[F^{(3)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} \\ &\quad + \frac{x_i^2 - \tilde{x}_{i+1}^2}{2} \cdot f(x_{i+1/2}) . \end{aligned}$$

Putting all of this together yields the desired result

$$\begin{aligned} \frac{\partial D_{\underline{x}}}{\partial x_i} &= G_i^{(1)} - 2 \cdot G_i^{(2)} + G_i^{(3)} \\ &= \frac{(\tilde{x}_{i-1/2} - \tilde{x}_{i-1})^2 - (x_{i-1/2} - x_i)^2}{2} f(x_{i-1/2}) \\ &\quad - 2 \left[F^{(2)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} + 2x_i \left[F^{(3)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} \\ &\quad + \frac{(x_{i+1/2} - x_i)^2 - (\tilde{x}_{i+1/2} - \tilde{x}_{i+1})^2}{2} f(x_{i+1/2}) \\ &= -2 \left[F^{(2)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} + 2x_i \left[F^{(3)}(y) \right]_{x_{i-1/2}}^{x_{i+1/2}} . \end{aligned}$$

□

B. Proof of Proposition 2

The expectation $\mathbb{E}[(X - q(X))^2]$ can be decomposed into

$$\mathbb{E}[(X - q(X))^2] = \sum_{i=1}^L \sum_{k=-\infty}^{\infty} P_{i,k}^{(1)} - 2P_{i,k}^{(2)} + P_{i,k}^{(3)}$$

Now, each component is computed separately. In each case, the computations come down to using the relationship to the Gaussian distribution.

$$\begin{aligned} P_{i,k}^{(1)} &= \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{y^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+2k\pi)^2}{2\sigma^2}\right) dy \\ &= \left[\frac{\sigma(2k\pi - y)}{\sqrt{2\pi}} \exp\left(-\frac{(y+2k\pi)^2}{2\sigma^2}\right) \right. \\ &\quad \left. + \left(2k^2\pi^2 + \frac{\sigma^2}{2}\right) \operatorname{erf}\left(\frac{y+2k\pi}{\sqrt{2}\sigma}\right) \right]_{x_{i-1/2}}^{x_{i+1/2}} , \\ &= \left[\sigma^2(2k\pi - y) f_k(y) \right. \\ &\quad \left. + \left(2k^2\pi^2 + \frac{\sigma^2}{2}\right) \operatorname{erf}\left(\frac{y+2k\pi}{\sqrt{2}\sigma}\right) \right]_{x_{i-1/2}}^{x_{i+1/2}} , \end{aligned}$$

$$\begin{aligned} P_{i,k}^{(2)} &= \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{x_i y}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+2k\pi)^2}{2\sigma^2}\right) dy \\ &= -x_i \left[\frac{\sigma \cdot \exp\left(-\frac{(y+2k\pi)^2}{2\sigma^2}\right)}{\sqrt{2\pi}} \right. \\ &\quad \left. + k \cdot \pi \cdot \operatorname{erf}\left(\frac{y+2k\pi}{\sqrt{2}\sigma}\right) \right]_{x_{i-1/2}}^{x_{i+1/2}} \\ &= -x_i \left[\sigma^2 \cdot f_k(y) + k \cdot \pi \cdot \operatorname{erf}\left(\frac{y+2k\pi}{\sqrt{2}\sigma}\right) \right]_{x_{i-1/2}}^{x_{i+1/2}} , \end{aligned}$$

$$\begin{aligned} P_{i,k}^{(3)} &= \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{x_i^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y+2k\pi)^2}{2\sigma^2}\right) dy \\ &= \left[\frac{x_i^2}{2} \operatorname{erf}\left(\frac{y+2k\pi}{\sqrt{2}\sigma}\right) \right]_{x_{i-1/2}}^{x_{i+1/2}} . \end{aligned}$$

□

C. Proof of Proposition 3

Due to the fact that $z_2 > z_1$, we observe that one of the modes of the Bingham distribution is located at μ with $(\cos(\mu), \sin(\mu)) = (m_{1,2}, m_{2,2})$. Furthermore, we note that \mathbf{M} is orthogonal. Consequently, the density can be rewritten as

$$f_{\text{Bingham}}(x; \mathbf{M}, \mathbf{Z}) = f_{\text{Bingham}}(x - \mu; \mathbf{A}, \mathbf{Z}) ,$$

with

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (6)$$

The combination of $z_1 < z_2$ and this particular choice of \mathbf{A} ensure $f_{\text{Bingham}}(x - \mu; \mathbf{A}, \mathbf{Z})$ to take its maximum values at $x = \mu$ and $x = \mu - \pi$. The mean μ is obtained by

$$\mu = \text{atan2}(m_{2,2}, m_{1,2}) .$$

Now, it is clearly sufficient to consider a Bingham(\mathbf{A}, \mathbf{Z}) distributed random variable X in the following.

In order to rewrite the density of Bingham(\mathbf{A}, \mathbf{Z}), we note that Bingham($\mathbf{A}, \mathbf{Z} + c \cdot \mathbf{I}$) for arbitrary $c \in \mathbb{R}$ (see [23]). Thus $\text{diag}(z_1, z_2)$ and $\text{diag}(0, z_2 - z_1)$ describe the same distribution and it can be assumed that the first entry of \mathbf{Z} is zero. Then, the p.d.f. of the Bingham distribution can be rewritten as

$$f(x) \propto \exp((z_2 - z_1) \cos(x)^2) .$$

Making use of

$$\cos(x)^2 = \frac{\cos(2x) + 1}{2}$$

yields the density representation

$$\begin{aligned} f(x) &\propto \exp\left((z_2 - z_1) \frac{\cos(2x) + 1}{2}\right) \\ &\propto \exp\left(\frac{z_2 - z_1}{2} \cos(2x)\right) . \end{aligned}$$

The density \tilde{f} of $2X$ can be obtained by applying the transformation theorem for probability densities. This procedure yields

$$\tilde{f}(x) \propto \exp\left(\frac{z_2 - z_1}{2} \cos(x)\right)$$

Because of its 2π periodicity, \tilde{f} is also proportional to the density of $2X \bmod 2\pi$. Thus, $2(X + \mu) \bmod 2\pi \sim \text{VM}(\theta, \kappa)$ with $\theta = (2\mu \bmod 2\pi)$ and $\kappa = (z_2 - z_1)/2$. \square

REFERENCES

- [1] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.
- [2] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Wiley, 1999.
- [3] G. S. Watson, *Statistics on Spheres*. Wiley, 1983.
- [4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer, 1992.
- [5] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Springer, 2000.
- [6] G. Pages and J. Printems, "Optimal Quadratic Quantization for Numerics: the Gaussian Case," *Monte Carlo Methods and Applications*, vol. 9, no. 2, 2003.
- [7] J. Traa and P. Smaragdis, "A Wrapped Kalman Filter for Azimuthal Speaker Tracking," *Signal Processing Letters, IEEE*, vol. 20, no. 12, pp. 1257–1260, dec 2013.
- [8] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun, "GNSS Dataless Signal Tracking with a Delay Semi-Open Loop and a Phase Open Loop," *Signal Processing*, vol. 93, no. 5, pp. 1192–1209, may 2013.
- [9] I. Markovic and I. Petrovic, "Bearing-only Tracking with a Mixture of von Mises Distributions," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, oct 2012, pp. 707–712.
- [10] I. Markovic, A. Portello, P. Danes, I. Petrovic, and S. Argentieri, "Active Speaker Localization with Circular Likelihoods and Bootstrap Filtering," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, nov 2013, pp. 2914–2920.
- [11] V. Carotenuto, A. Aubry, A. De Maio, and A. Farina, "Phase Noise Modeling and its Effects on the Performance of Some Radar Signal Processors," in *2015 IEEE Radar Conference (RadarCon)*. IEEE, may 2015, pp. 0274–0279.
- [12] R. G. McKilliam, "Lattice Theory, Circular Statistics and Polynomial Phase Signals," Ph.D. dissertation, University of Queensland, 2010.
- [13] J. Lo and A. Willsky, "Estimation for Rotational Processes with One Degree of Freedom—Part I: Introduction and Continuous-Time Processes," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 10–21, feb 1975.
- [14] A. Willsky and J. Lo, "Estimation for Rotational Processes with one Degree of Freedom—Part II: Discrete-Time Processes," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 22–30, feb 1975.
- [15] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A Recursive Fusion Filter for Angular Data," in *International Conference on Robotics and Biometrics (ROBIO)*, dec 2009, pp. 882–887.
- [16] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Bayesian Filtering in Circular State Spaces," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, 2016.
- [17] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, jan 2016.
- [18] U. D. Hanebeck and A. Lindquist, "Moment-based Dirac Mixture Approximation of Circular Densities," in *Proceedings of the 19th IFAC World Congress (IFAC 2014)*, Cape Town, South Africa, aug 2014.
- [19] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods*. Birkhäuser, 2009.
- [20] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, nov 1974.
- [21] K. Kunze and H. Schaeben, "The Bingham Distribution of Quaternions and Its Spherical Radon Transform in Texture Analysis," *Mathematical Geology*, vol. 36, no. 8, pp. 917–943, nov 2004.
- [22] G. Pagès, "A Space Quantization Method for Numerical Integration," *Journal of Computational and Applied Mathematics*, vol. 89, no. 1, pp. 1–38, mar 1998.
- [23] G. Kurz, I. Gilitschenski, S. J. Julier, and U. D. Hanebeck, "Recursive Estimation of Orientation Based on the Bingham Distribution," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.