Parameter Estimation for the Bivariate Wrapped Normal Distribution

Gerhard Kurz[1] and Uwe D. Hanebeck[1]

***Abstract*— Correlated uncertain angular quantities can be modeled using the bivariate wrapped normal distribution. In this paper, we focus on the problem of estimating the distribution's parameters from a given set of samples. For this purpose, we propose several new parameter estimation methods and compare them to estimation techniques found in literature. All methods are thoroughly evaluated in simulations. One of the novel methods is shown to combine the advantages of maximum likelihood estimation and moment-based methods, thus outperforming current state-of-the-art techniques.**

I. Introduction

Uncertain angular quantities appear in a variety of fields, such as robotics, control systems, meteorology, biology, geology, and signal processing. Typical examples include the direction a robot is facing, the wind direction, or the phase of a signal. Quantities of this type can be treated using directional statistics [1], a subfield of statistics that deals with nonlinear manifolds. However, many applications are not limited to a single uncertain angle, but consider multiple uncertain angles that may not be stochastically independent. In these cases, the circular–circular correlation between those angles, an adaptation of the concept of correlation to periodic quantities, has to be taken into account [2].

A single angle can be understood as a point on the unit circle. To be able to handle two angles, we have to consider the Cartesian product of two circles, i.e., the torus as the underlying manifold. This can be generalized to the n-torus for a larger number of angles. One of the most important probability distributions on the torus (and the n-torus) is the *bivariate* (or *toroidal*) wrapped normal (BWN) distribution. In this paper, we focus on the problem of parameter estimation for this distribution. The parameter estimation method could then be used in a variety of applications, for example in the context of a recursive filtering scheme [3]. In the following, we will restrict ourselves to the 2-torus parameterized by $[0, 2\pi)^2$.

Formally, the problem considered in this paper can be stated as follows. Given a set of n independent and identically distributed (i.i.d.) sample vectors $\underline{x}^{(1)}, \dots, \underline{x}^{(n)} \in [0, 2\pi)^2$, we seek to estimate the parameters $\underline{\mu} \in \mathbb{R}^2$ and symmetric positive definite $\mathbf{C} \in \mathbb{R}^{2\times 2}$ of a BWN distribution. This problem is visualized in Fig. 1.

A few approaches to parameter estimation for the BWN distribution have been proposed in literature. Jammalamadaka proposed a moment-based approach [2] that unfortunately fails in many practically relevant cases by returning a matrix $\mathbf{C}$ that is not positive definite. Recently, we proposed an improved approach based on moment matching that works in most practically relevant cases, but still does not guarantee positive definiteness [3]. The reason for these failures is not numerical inaccuracy but rather the fact that no BWN distribution with certain predefined moments may exist.

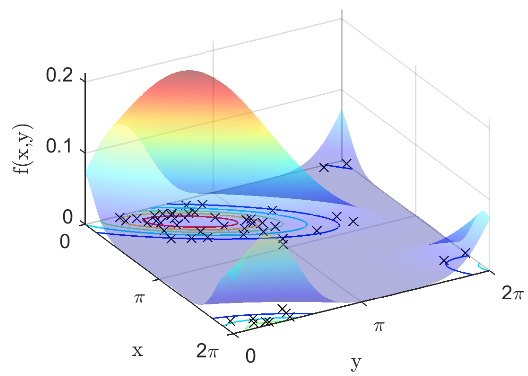

Fig. 1: A bivariate wrapped normal distribution and the corresponding samples. The goal is to estimate its parameters based on a given sample set. Note that both x and y are 2π-periodic, i.e., the left and right side of the plot as well as the front and back side of the plot are connected, thus creating the topology of a torus.

Some authors have also considered approaches based on maximum likelihood estimation [4]. However, no closed-form solution for the true maximum likelihood estimate is known. To obtain a closed-form solution, a bound based on Jensen's inequality has to be used, which introduces a certain amount of suboptimality.

Finally, there have been some approaches based on unwrapping, i.e., reversing the wrapping from $\mathbb{R}^2$ to the torus, and then fitting a regular normal distribution. Some authors also use the term *data augmentation* [5]. An unwrapping method based on expectation maximization was proposed by Fisher et al. [6], but it is not very efficient. To reduce the computational burden, other authors have considered Markov Chain Monte-Carlo (MCMC) unwrapping methods [7], [8], [9]. However, these techniques are still computationally quite intensive and have a number of tuning parameters that are difficult to choose.

The contribution of this paper can be summarized as follows. First, we introduce a number of approaches from literature and describe them in a unified manner. Second, we

[1] The authors 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`

propose several new parameter estimation algorithms, namely an algorithm based on Jupp's correlation coefficient [10], an algorithm based directly on certain entries of the covariance matrix, and an algorithm that tries to combine the benefits of maximum likelihood and moment-based approaches. Third, we perform a thorough evaluation of all these approaches with respect to accuracy, computation time, and chance of failure.

II. Toroidal Statistics

Before we address the problem of parameter estimation, we give a brief introduction to the topic of toroidal statistics. Consider a real-valued two-dimensional random vector $\underline{z} \in \mathbb{R}^2$, which is normally distributed according to $\underline{z} \sim \mathcal{N}(\underline{x}; \underline{\mu}, \mathbf{C})$ where $\underline{\mu} \in \mathbb{R}^2$ and $\mathbf{C} \in \mathbb{R}^{2\times 2}$. If we take $\underline{x} = \underline{z} \mod 2\pi$, where the modulo operation is performed componentwise, we obtain a random variable on $[0, 2\pi)$ that follows the *bivariate wrapped normal distribution*.

Definition 1 (Bivariate Wrapped Normal Distribution)**.** The bivariate (or toroidal) wrapped normal distribution is given by the probability density function

$$\mathcal{BWN}(\underline{x}; \underline{\mu}, \mathbf{C}) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \mathcal{N}(\underline{x} + [2\pi j, 2\pi k]^T; \underline{\mu}, \mathbf{C}) \ ,$$

where $\underline{x}, \underline{\mu} \in [0, 2\pi)^2$ and $\mathbf{C} \in \mathbb{R}^{2\times 2}$ is symmetric positive definite.

The BWN distribution can be seen as a bivariate generalization to the wrapped normal distribution. This distribution has previously been discussed by a number of authors, e.g., [2, Sec. 3.1], [11, Example 7.3], [7]. It has five degrees of freedom, two for the location $\underline{\mu}$, two for the uncertainty of each dimension encoded in $c_{1,1}$ and $c_{2,2}$, and one for the correlation encoded in $c_{1,2}$. Consequently, at least three two-dimensional samples (which have six degrees of freedom) are required to estimate the parameters of a BWN distribution[1]. The BWN distribution is a special case of the partially wrapped normal distribution discussed in [12]. For the partially wrapped normal distribution, some dimensions can be periodic whereas others can be linear.

A. Moments

In the following, we consider a toroidal analogon to the concept of (power) moments in the linear case. The moments discussed in this paper are a special case of the hybrid moments defined in [12]. Consider a random vector $\underline{x}$ following a Gaussian distribution. It is well known that this distribution is uniquely defined by its mean, i.e., the first moment $\mathbb{E}(\underline{x})$, and covariance, i.e., the second central moment $\mathbb{E}\left((\underline{x} - \mathbb{E}(\underline{x})) \cdot (\underline{x} - \mathbb{E}(\underline{x}))^T\right)$. These two moments directly coincide with the parameters $\underline{\mu}$ and $\mathbf{C}$ of the Gaussian distribution.

In the toroidal case, we consider a random vector $\underline{x} \in [0, 2\pi)^2$. Because of the periodicity involved, the moments of $\underline{x} = [x_1, x_2]^T$ are not particularly useful. Instead, we consider the transformed random variable

$$\tilde{\underline{x}} = \begin{bmatrix} \operatorname{Re}\exp(ix_1) \\ \operatorname{Im}\exp(ix_1) \\ \operatorname{Re}\exp(ix_2) \\ \operatorname{Im}\exp(ix_2) \end{bmatrix} = \begin{bmatrix} \cos(x_1) \\ \sin(x_1) \\ \cos(x_2) \\ \sin(x_2) \end{bmatrix} \ , \tag{1}$$

i.e., we apply both the sine and the cosine to each component of $\underline{x}$. The first moment of the transformed random vector $\tilde{x}$ is given by $\tilde{\underline{\mu}} = \mathbb{E}(\tilde{\underline{x}})$ and the second central moment is given by $\tilde{\mathbf{C}} = \mathbb{E}\left((\tilde{\underline{x}} - \tilde{\underline{\mu}})(\tilde{\underline{x}} - \tilde{\underline{\mu}})^T\right)$. The entries of $\tilde{\mathbf{C}}$ are shown in Fig. 2. By applying the identity

$$\operatorname{Cov}[x, y] = \mathbb{E}(xy) - \mathbb{E}(x)\,\mathbb{E}(y) \ ,$$

we can decompose $\tilde{\mathbf{C}}$ into $\tilde{\mathbf{C}} = \tilde{\mathbf{A}} - \tilde{\mathbf{B}}$, where $\tilde{\mathbf{A}} = \mathbb{E}\left(\tilde{\underline{x}}\tilde{\underline{x}}^T\right)$ and $\tilde{\mathbf{B}} = \mathbb{E}(\tilde{\underline{x}})\,\mathbb{E}(\tilde{\underline{x}})^T$. The first moment coincides with a vector representation of the first toroidal moment discussed in [3].

The moments $\tilde{\underline{\mu}}$ and $\tilde{\mathbf{C}}$ of a BWN distribution can be calculated in closed form. We only consider the case $\underline{\mu} = [0, 0]^T$ because we can shift any BWN distribution accordingly.

Theorem 1 (Moments of the BWN Distribution)**.** The moments of a BWN distribution with parameters $\underline{\mu} = [0, 0]^T$ and $\mathbf{C}$ are given by

$$\tilde{\underline{\mu}} = \begin{bmatrix} \exp(-c_{1,1}/2) \\ 0 \\ \exp(-c_{2,2}/2) \\ 0 \end{bmatrix} \ , \ \tilde{\mathbf{C}} = \begin{bmatrix} \tilde{c}_{1,1} & 0 & \tilde{c}_{1,3} & 0 \\ 0 & \tilde{c}_{2,2} & 0 & \tilde{c}_{2,4} \\ \tilde{c}_{1,3} & 0 & \tilde{c}_{3,3} & 0 \\ 0 & \tilde{c}_{2,4} & 0 & \tilde{c}_{4,4} \end{bmatrix} \ ,$$

where

$$\begin{aligned} \tilde{c}_{1,1} &= \frac{1}{2}(1 - \exp(-c_{1,1}))^2 \ , \\ \tilde{c}_{2,2} &= \frac{1}{2}(1 - \exp(-2c_{1,1})) \ , \\ \tilde{c}_{3,3} &= \frac{1}{2}(1 - \exp(-c_{2,2}))^2 \ , \\ \tilde{c}_{4,4} &= \frac{1}{2}(1 - \exp(-2c_{2,2})) \ , \\ \tilde{c}_{1,3} &= \exp(-c_{1,1}/2 - c_{2,2}/2)(\cosh(c_{1,2}) - 1) \ , \\ \tilde{c}_{2,4} &= \exp(-c_{1,1}/2 - c_{2,2}/2)(\sinh(c_{1,2})) \ . \end{aligned}$$

Proof. The moments can be derived using the characteristic function of the bivariate normal distribution (see [11, Example 7.3]). A detailed derivation including the case of $\underline{\mu} \neq \underline{0}$ can be found in [12, Sec. 2.3.3, Theorem 1]. □

The empirical sample moments for $\underline{x}^{(1)}, \dots, \underline{x}^{(n)} \in [0, 2\pi)^2$ are given by

$$\bar{\underline{\mu}} = \frac{1}{n}\sum_{j=1}^{n} \xi(\underline{x}^{(j)}), \ \bar{\mathbf{C}} = \frac{1}{n}\sum_{j=1}^{n} \xi(\underline{x}^{(j)} - \bar{\underline{\mu}})\xi(\underline{x}^{(j)} - \bar{\underline{\mu}})^T \ , \tag{2}$$

where $\xi(\underline{x}) = [\cos(x_1), \sin(x_1), \cos(x_2), \sin(x_2)]^T$.

[1] Three samples are not always sufficient as there are certain special cases, e.g., multiple identical samples, collinear samples, etc. where not all degrees of freedom are covered.

$$\tilde{\mathbf{C}} = \left[\begin{array}{cc|cc} \mathrm{Cov}[\cos(x_1),\cos(x_1)] & \mathrm{Cov}[\cos(x_1),\sin(x_1)] & \mathrm{Cov}[\cos(x_1),\cos(x_2)] & \mathrm{Cov}[\cos(x_1),\sin(x_2)] \\ \mathrm{Cov}[\sin(x_1),\cos(x_1)] & \mathrm{Cov}[\sin(x_1),\sin(x_1)] & \mathrm{Cov}[\sin(x_1),\cos(x_2)] & \mathrm{Cov}[\sin(x_1),\sin(x_2)] \\ \hline \mathrm{Cov}[\cos(x_2),\cos(x_1)] & \mathrm{Cov}[\cos(x_2),\sin(x_1)] & \mathrm{Cov}[\cos(x_2),\cos(x_2)] & \mathrm{Cov}[\cos(x_2),\sin(x_2)] \\ \mathrm{Cov}[\sin(x_2),\cos(x_1)] & \mathrm{Cov}[\sin(x_2),\sin(x_1)] & \mathrm{Cov}[\sin(x_2),\cos(x_2)] & \mathrm{Cov}[\sin(x_2),\sin(x_2)] \end{array}\right]$$

$$= \underbrace{\left[\begin{array}{cc|cc} \mathbb{E}(\cos(x_1)\cos(x_1)) & \mathbb{E}(\cos(x_1)\sin(x_1)) & \mathbb{E}(\cos(x_1)\cos(x_2)) & \mathbb{E}(\cos(x_1)\sin(x_2)) \\ \mathbb{E}(\sin(x_1)\cos(x_1)) & \mathbb{E}(\sin(x_1)\sin(x_1)) & \mathbb{E}(\sin(x_1)\cos(x_2)) & \mathbb{E}(\sin(x_1)\sin(x_2)) \\ \hline \mathbb{E}(\cos(x_2)\cos(x_1)) & \mathbb{E}(\cos(x_2)\sin(x_1)) & \mathbb{E}(\cos(x_2)\cos(x_2)) & \mathbb{E}(\cos(x_2)\sin(x_2)) \\ \mathbb{E}(\sin(x_2)\cos(x_1)) & \mathbb{E}(\sin(x_2)\sin(x_1)) & \mathbb{E}(\sin(x_2)\cos(x_2)) & \mathbb{E}(\sin(x_2)\sin(x_2)) \end{array}\right]}_{=:\tilde{\mathbf{A}}}$$

$$- \underbrace{\left[\begin{array}{cc|cc} \mathbb{E}(\cos(x_1))^2 & \mathbb{E}(\cos(x_1))\,\mathbb{E}(\sin(x_1)) & \mathbb{E}(\cos(x_1))\,\mathbb{E}(\cos(x_2)) & \mathbb{E}(\cos(x_1))\,\mathbb{E}(\sin(x_2)) \\ \mathbb{E}(\sin(x_1))\,\mathbb{E}(\cos(x_1)) & \mathbb{E}(\sin(x_1))^2 & \mathbb{E}(\sin(x_1))\,\mathbb{E}(\cos(x_2)) & \mathbb{E}(\sin(x_1))\,\mathbb{E}(\sin(x_2)) \\ \hline \mathbb{E}(\cos(x_2))\,\mathbb{E}(\cos(x_1)) & \mathbb{E}(\cos(x_2))\,\mathbb{E}(\sin(x_1)) & \mathbb{E}(\cos(x_2))^2 & \mathbb{E}(\cos(x_2))\,\mathbb{E}(\sin(x_2)) \\ \mathbb{E}(\sin(x_2))\,\mathbb{E}(\cos(x_1)) & \mathbb{E}(\sin(x_2))\,\mathbb{E}(\sin(x_1)) & \mathbb{E}(\sin(x_2))\,\mathbb{E}(\cos(x_2)) & \mathbb{E}(\sin(x_2))^2 \end{array}\right]}_{=:\tilde{\mathbf{B}}}$$

Fig. 2: Covariance matrix of the vector $\underline{\tilde{x}}$ as defined in (1).

B. Measures of Correlation

Several measures of correlation between two circular random variables x_1 and x_2 have been proposed. There does not seem to be any clear consensus which of those measures is to be preferred. Several authors [10], [11], [13], [14] have proposed correlation coefficients based on the matrix $\mathbf{D} := \tilde{\mathbf{C}}_{1:2,1:2}^{-1} \cdot \tilde{\mathbf{C}}_{1:2,3:4} \cdot \tilde{\mathbf{C}}_{3:4,3:4}^{-1} \cdot \tilde{\mathbf{C}}_{3:4,1:2}$. The notation $\tilde{\mathbf{C}}_{a:b,c:d}$ refers to the submatrix of $\tilde{\mathbf{C}}$ with rows a to b and columns c to d. In the case of a zero-mean BWN distribution, we obtain

$$\mathbf{D} = \begin{bmatrix} \tilde{c}_{1,3}^2/(\tilde{c}_{1,1}\tilde{c}_{3,3}) & 0 \\ 0 & \tilde{c}_{2,4}^2/(\tilde{c}_{2,2}\tilde{c}_{4,4}) \end{bmatrix} .$$

For example, Jupp [10] defines $\rho^2 = \operatorname{trace}\mathbf{D}$, and Johnson [11] defines ρ^2 as the largest eigenvalue of $\mathbf{D}$. In both cases, the sign of ρ is lost, but can be reconstructed by using the sign of $\det \tilde{\mathbf{C}}_{1:2,3:4}$. Other closely related correlation coefficients can be found in the papers by Mardia et al. [13] and Rivest [14]. In 1988, Jammalamadaka and Sarma [2] proposed the correlation coefficient

$$\rho_c = \frac{\mathbb{E}\left(\sin(x_1-\mu_1)\sin(x_2-\mu_2)\right)}{\sqrt{\mathbb{E}\left(\sin^2(x_1-\mu_1)\right)\cdot\mathbb{E}\left(\sin^2(x_2-\mu_2)\right)}} , \quad (3)$$

where μ_1 and μ_2 are the circular means of x_1 and x_2, respectively. This coefficient was, for example, used in [3]. In the case of $\underline{\mu} = [\mu_1, \mu_2]^T = [0, 0]^T$, ρ_c simplifies to

$$\rho_c = \frac{\mathbb{E}\left(\sin(x_1)\sin(x_2)\right)}{\sqrt{\mathbb{E}\left(\sin^2(x_1)\right)\cdot\mathbb{E}\left(\sin^2(x_2)\right)}} = \frac{\tilde{a}_{2,4}}{\sqrt{\tilde{b}_{2,2}\cdot\tilde{b}_{4,4}}} .$$

For a BWN distribution, ρ_c is given by

$$\rho_c = \sinh(c_{1,2})/\sqrt{\sinh(c_{1,1})\sinh(c_{2,2})} .$$

III. Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a common parameter estimation technique based on finding the parameters that maximize the likelihood of obtaining the samples $\underline{x}^{(1)}, \dots, \underline{x}^{(n)}$ given those parameters. For a BWN distribution with parameters $\underline{\mu}$ and $\mathbf{C}$, and i.i.d. samples, the likelihood is given by $L = \prod_{l=1}^n \mathcal{BWN}(\underline{x}^{(l)}; \underline{\mu}, \mathbf{C})$. Because the logarithm is a strictly increasing function and $L > 0$, it is equivalent to maximizing the log-likelihood instead. The resulting optimization problem

$$\arg\max_{\underline{\mu},\mathbf{C}}(\log(L))$$

can be solved using numerical methods such as the Nelder-Mead simplex algorithm [15]. This is quite computationally intensive, however. Also, as this problem is not convex, the results may depend on the chosen initial value and there is no guarantee that the global optimum will be found.

For some probability distributions, it is possible to analytically calculate the parameters that maximize the likelihood, e.g., for the Gaussian distribution. This does not seem to be the case for the BWN distribution, as not even the wrapped normal distribution on the circle allows maximum likelihood estimation in closed form. Thus, some authors have proposed approximations based on Jensen's inequality. This method has been applied to both the circular [16] and the toroidal cases [4]. Jensen's inequality states that for a concave function $\phi(\cdot)$, values x_j, and coefficients $\alpha_j > 0$ with $\sum_j \alpha_j = 1$,

$$\phi\left(\sum\nolimits_j \alpha_j x_j\right) \geq \sum\nolimits_j \alpha_j \phi(x_j)$$

holds. Particularly, for $\phi(\cdot) = \log(\cdot)$, we can bound the log-likelihood function from below according to

$$\begin{aligned} \log(L) &= \sum_{l=1}^n \log\left(\mathcal{BWN}(\underline{x}^{(l)}; \underline{\mu}, \mathbf{C})\right) \\ &= \sum_{l=1}^n \log\left(\sum_{j,k=-\infty}^{\infty} \mathcal{N}(\underline{x}^{(l)} + 2\pi[j,k]^T; \underline{\mu}, \mathbf{C})\right) \\ &= \sum_{l=1}^n \log\left(\sum_{j,k=-\infty}^{\infty} \alpha_{j,k}^l \frac{\mathcal{N}(\underline{x}^{(l)} + 2\pi[j,k]^T; \underline{\mu}, \mathbf{C})}{\alpha_{j,k}^l}\right) \\ &\geq \sum_{l=1}^n \sum_{j,k=-\infty}^{\infty} \alpha_{j,k}^l \log\left(\frac{\mathcal{N}(\underline{x}^{(l)} + 2\pi[j,k]^T; \underline{\mu}, \mathbf{C})}{\alpha_{j,k}^l}\right) \end{aligned}$$

$$=: Q$$

for suitable $\alpha^l_{j,k} > 0$ with $\sum_{j,k=-\infty}^{\infty} \alpha^l_{j,k} = 1$ for all $l = 1, \dots, n$. Now, we can find the parameters $\underline{\mu}$ and $\mathbf{C}$ by analytically determining

$$\arg\max_{\underline{\mu}, \mathbf{C}}(Q)$$

as follows. We compute the partial derivatives of Q with respect to $\underline{\mu}$ and $\mathbf{C}$, and obtain the solution

$$\underline{\mu} = \frac{1}{n} \sum_{l=1}^{n} \sum_{j,k=-\infty}^{\infty} (\underline{x}^{(l)} - 2\pi[j,k]^T)\alpha^l_{j,k} \ ,$$

$$\mathbf{C} = \frac{1}{n} \sum_{l=1}^{n} \sum_{j,k=-\infty}^{\infty} (\underline{x}^{(l)} - \underline{\mu} - 2\pi[j,k]^T) \cdot (\underline{x}^{(l)} - \underline{\mu} - 2\pi[j,k]^T)^T \alpha^l_{j,k} \ .$$

For a practical implementation, the infinite sums can be truncated to a small number of terms, which is reasonable if $\alpha^l_{j,k}$ falls off quickly[2]. It is, however, quite obvious that the parameters $\underline{\mu}$ and $\mathbf{C}$ maximizing the function Q are, in general, not identical to the parameters maximizing the likelihood L. For this reason, the solution based on Jensen's inequality constitutes an approximation and this estimator is usually not even asymptotically unbiased. It can be seen that the resulting parameters depend on the choice of the constants $\alpha^l_{j,k}$. Even though Jensen's inequality holds for any $\alpha^l_{j,k}$ satisfying the aforementioned conditions, the bound is not equally tight for any choice of $\alpha^l_{j,k}$ and the location of the maximum of Q and the maximum of L may differ significantly. This difference depends on the choice of $\alpha^l_{j,k}$. A possible choice of $\alpha^l_{j,k}$ is

$$\alpha^l_{j,k} = \frac{\mathcal{N}(\underline{x}^{(l)} + 2\pi[j,k]^T; \underline{\mu}^\alpha, \mathbf{C}^\alpha)}{\sum_{p,q=-\infty}^{\infty} \mathcal{N}(\underline{x}^{(l)} + 2\pi[p,q]^T; \underline{\mu}^\alpha, \mathbf{C}^\alpha)} \ , \quad (4)$$

where $\underline{\mu}^\alpha$ and $\mathbf{C}^\alpha$ are parameters of another BWN distribution, possibly a (rough) estimate of the true parameters that are to be obtained.

IV. Moment-based Estimation

An alternative to the MLE method for parameter estimation are moment-based solutions. There, the moments of the given sample set $\underline{x}^{(1)}, \dots, \underline{x}^{(n)}$ are calculated, and then, the parameters of the distribution are chosen such that it has the same moments. For certain distributions, the MLE method and the moment-based estimation yield the same results[3], e.g., for the Gaussian distribution (when matching mean and covariance) or the von Mises distribution (when matching the first trigonometric moment).

In the remainder of this section, we assume that the sample moments $\bar{\underline{\mu}}$ and $\bar{\mathbf{C}}$ are given as in (2). We assume that $\bar{\mu}_1^2 + \bar{\mu}_2^2 > 0$ (i.e., not both are zero) and $\bar{\mu}_3^2 + \bar{\mu}_4^2 > 0$ because the circular mean is undefined otherwise. Furthermore, we consider a BWN distribution with parameters $\underline{\mu}$ and $\mathbf{C}$, whose moments are given by $\tilde{\underline{\mu}}$ and $\tilde{\mathbf{C}}$. The moment matrix $\tilde{\mathbf{C}}$ is decomposed into $\tilde{\mathbf{A}}$ with entries $\tilde{a}_{i,j}$ and $\tilde{\mathbf{B}}$ with entries $\tilde{b}_{i,j}$ as illustrated in Fig. 2.

By solving the equation $\bar{\underline{\mu}} = \tilde{\underline{\mu}}$, we obtain the solution for $\underline{\mu}$ and $c_{1,1}$ and $c_{2,2}$ (analogous to [18, Lemma 2], [19, Sec. III-A2])

$$\mu_1 = \operatorname{atan2}(\bar{\underline{\mu}}_2, \bar{\underline{\mu}}_1) \ , \qquad \mu_2 = \operatorname{atan2}(\bar{\underline{\mu}}_4, \bar{\underline{\mu}}_3) \ ,$$
$$c_{1,1} = -\log(\bar{\underline{\mu}}_1^2 + \bar{\underline{\mu}}_2^2) \ , \qquad c_{2,2} = -\log(\bar{\underline{\mu}}_3^2 + \bar{\underline{\mu}}_4^2) \ .$$

This leaves the question of how to obtain the remaining parameter $c_{1,2}$. Because it is not obvious how to calculate this parameter, we present several alternative solutions. It should be noted that the parameter matrix $\mathbf{C}$ has to be positive definite to obtain a well-defined BWN distribution. This is the case if and only if $c_{1,1}c_{2,2} > c_{1,2}^2$.

All purely moment-based estimators discussed in this paper can be shown to be consistent, i.e., the estimate converges in probability to the true parameters as the number of samples n approaches infinity. The proof is based on the fact that the sample covariance $\bar{\mathbf{C}}$ converges in probability towards the true covariance $\tilde{\mathbf{C}}$ and the proposed estimators are continuous functions that yield the correct parameters for $\bar{\mathbf{C}} = \tilde{\mathbf{C}}$.

A. *Jammalamadaka's Method*

In 1988, Jammalamadaka and Sarma [2, Sec. 3.1] proposed a method based on matching the expectation value $\mathbb{E}(\exp(ix_1)\exp(ix_2))$. This can be rewritten as

$$\begin{aligned}
&\mathbb{E}(\exp(ix_1)\exp(ix_2)) \\
&= \mathbb{E}(\cos(x_1 + x_2)) + i\mathbb{E}(\sin(x_1 + x_2)) \\
&= \mathbb{E}(\cos(x_1)\cos(x_2)) - \mathbb{E}(\sin(x_1)\sin(x_2)) \\
&\quad + i(\mathbb{E}(\sin(x_1)\cos(x_2)) + \mathbb{E}(\cos(x_1)\sin(x_2))) \\
&= \tilde{a}_{1,3} - \tilde{a}_{2,4} + i(\tilde{a}_{1,4} + \tilde{a}_{2,3}) \ .
\end{aligned}$$

For a BWN distribution with zero mean, we have $\tilde{a}_{1,4} = \tilde{a}_{2,3} = 0$. For this reason, we only match the real part. Solving the equation $\tilde{a}_{1,3} - \tilde{a}_{2,4} = \bar{a}_{1,3} - \bar{a}_{2,4}$ yields

$$\begin{aligned}
c_{1,2} &= -\log(\exp(c_{1,1}/2 + c_{2,2}/2) \cdot (\bar{a}_{1,3} - \bar{a}_{2,4})) \\
&= -c_{1,1}/2 - c_{2,2}/2 - \log(\bar{a}_{1,3} - \bar{a}_{2,4}) \ .
\end{aligned}$$

However, this solution does not guarantee that $\mathbf{C}$ is positive definite. Even worse, it even causes $\mathbf{C}$ to lack positive definiteness in many common and practically relevant cases.

B. *Jammalamadaka's Correlation Coefficient*

In the same paper [2], the authors proposed a circular-circular correlation coefficient (3). We proposed a parameter estimation technique based on this correlation coefficient in [3]. Matching the correlation coefficient according to

$$\tilde{a}_{2,4} \cdot (\tilde{b}_{2,2} \cdot \tilde{b}_{4,4})^{-1/2} = \bar{a}_{2,4} \cdot (\bar{b}_{2,2} \cdot \bar{b}_{4,4})^{-1/2}$$

[2] For the choice of $\alpha^l_{j,k}$ proposed in (4), the convergence is similar to that discussed in [17].

[3] This only holds if the moment-based method uses the sample variance without Bessel's correction for unbiasedness.

yields

$$c_{1,2} = \operatorname{arcsinh}\left(\frac{\sqrt{\sinh(c_{1,1})\sinh(c_{2,2})} \cdot \bar{a}_{2,4}}{\sqrt{\bar{b}_{2,2} \cdot \bar{b}_{4,4}}} \right) .$$

Experimental results have shown that this method produces a positive definite matrix in most practically relevant cases, but it does not guarantee positive definiteness, particularly in cases with very strong correlation. This can happen because no BWN with the given correlation coefficient may exist (see also [3, Fig. 2b]).

C. Jupp's Correlation Coefficient

A similar technique can be derived based on Jupp's correlation coefficient [10] as follows. Matching this correlation coefficient according to

$$\begin{aligned} &\operatorname{trace}(\tilde{\mathbf{C}}_{1:2,1:2}^{-1} \cdot \tilde{\mathbf{C}}_{1:2,3:4} \cdot \tilde{\mathbf{C}}_{3:4,3:4}^{-1} \cdot \tilde{\mathbf{C}}_{3:4,1:2}) \\ &= \operatorname{trace}(\bar{\mathbf{C}}_{1:2,1:2}^{-1} \cdot \bar{\mathbf{C}}_{1:2,3:4} \cdot \bar{\mathbf{C}}_{3:4,3:4}^{-1} \cdot \bar{\mathbf{C}}_{3:4,1:2}) , \\ &\operatorname{sign}(\det \tilde{\mathbf{C}}_{1:2,3:4}) = \operatorname{sign}(\det \bar{\mathbf{C}}_{1:2,3:4}) \end{aligned}$$

yields after a lengthy calculation

$$c_{1,2} = \operatorname{sign}(\det \bar{\mathbf{C}}_{1:2,3:4}) \operatorname{arccosh}(\tau) ,$$

where

$$\begin{aligned} \tau &= \frac{\omega_2 + \sqrt{\omega_2^2 - (\omega_2+\omega_1)(\omega_2 - \omega_1 - \rho^2\omega_1\omega_2/\eta^2)}}{\omega_1+\omega_2} , \\ \eta &= \exp(-c_{1,1}/2 - c_{2,2}/2) , \\ \rho^2 &= \operatorname{trace}(\bar{\mathbf{C}}_{1:2,1:2}^{-1} \cdot \bar{\mathbf{C}}_{1:2,3:4} \cdot \bar{\mathbf{C}}_{3:4,3:4}^{-1} \cdot \bar{\mathbf{C}}_{3:4,1:2}) , \\ \omega_1 &= \tilde{c}_{1,1}\tilde{c}_{3,3} , \\ \omega_2 &= \tilde{c}_{2,2}\tilde{c}_{4,4} . \end{aligned}$$

The terms for ω_1 and ω_2 can be calculated according to Theorem 1. Once again $\mathbf{C}$ may not be positive definite in all cases, but experiments have shown that this yields a positive definite $\mathbf{C}$ in most practically relevant cases.

D. Covariance Matrix

The previous approaches (except Jupp's correlation coefficient) suffer from the fact that they only consider somewhat arbitrarily chosen entries of the covariance matrix $\tilde{\mathbf{C}}$ or its decomposition (see Fig. 2). For this reason, we consider a new approach, which is based on directly matching the entire upper right submatrix

$$\tilde{\mathbf{C}}_{1:2,3:4} = \begin{bmatrix} \tilde{c}_{1,3} & \tilde{c}_{1,4} \\ \tilde{c}_{2,3} & \tilde{c}_{2,4} \end{bmatrix}$$

related to correlation. Now we seek to minimize the squared Frobenius norm of $\tilde{\mathbf{C}}_{1:2,3:4} - \bar{\mathbf{C}}_{1:2,3:4}$. In other words, we seek to solve the overdetermined system of four equations

$$\tilde{c}_{1,3} = \bar{c}_{1,3} \ , \ \tilde{c}_{1,4} = \bar{c}_{1,4} \ , \ \tilde{c}_{2,3} = \bar{c}_{2,3} \ , \ \tilde{c}_{2,4} = \bar{c}_{2,4}$$

in a least-squares sense. Because $\tilde{c}_{1,4} = \tilde{c}_{2,3} = 0$ for a zero-mean BWN distribution, we can neglect these equations and obtain $c_{1,2}$ by solving the optimization problem

$$\arg\min_{c_{1,2}} \left((\tilde{c}_{1,3} - \bar{c}_{1,3})^2 + (\tilde{c}_{2,4} - \bar{c}_{2,4})^2 \right) . \quad (5)$$

Name	**Uses pdf**	**Guarantees pos. def.**	**Numerical**	**New in this paper**
MLE	yes	yes	yes	no
MLE (Jensen) [4]	yes	yes	no	no
Jammalamadaka [2]	no	no	no	no
Jamm.'s coeff. [3]	no	no	no	no
Jupp's coefficient	no	no	no	yes
Covariance	no	no	no	yes
Mixed MLE	yes	yes	yes	yes
Unwrapping EM [6]	no	yes	yes	no

TABLE I: Parameter estimation approaches.

Note that $\tilde{c}_{1,3}$ and $\tilde{c}_{2,4}$ depend on $c_{1,2}$ as shown in Theorem 1. A lengthy calculation yields the result $c_{1,2} = \log(z)$, where z is one of the roots of the polynomial

$$\eta X^4 + (-\bar{c}_{1,3} - \bar{c}_{2,4} - \eta)X^3 + (\bar{c}_{1,3} - \bar{c}_{2,4} + \eta)X - \eta$$

with $\eta = \exp(-c_{1,1}/2 - c_{2,2}/2)$. As this is a 4th degree polynomial, a closed-form solution exists and can be implemented [20]. The optimal solution can be obtained by plugging all possible roots into (5). Unfortunately, this method does not guarantee that $\mathbf{C}$ is positive definite either, even though it seems to work for most practically relevant scenarios.

E. Mixed MLE

As none of the approaches above can guarantee that $\mathbf{C}$ is positive definite, we propose a novel method that is based on a combination of moment-based parameter estimation and MLE. We use moment-based estimation to obtain $\underline{\mu}$ and $c_{1,1}, c_{1,2}$ as above and subsequently consider the log-likelihood as a function of $c_{1,2}$. To be precise, we seek a solution to the optimization problem

$$\arg\max_{c_{1,2}} \left(\sum_{l=1}^{n} \log \mathcal{BWN}\left(\underline{x}^{(l)}; \underline{\mu}, \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{1,2} & c_{2,2} \end{bmatrix} \right) \right) .$$

This optimization problem can be solved using numerical methods similar to the optimization problem in the general MLE method. However, as this is a one-dimensional optimization problem rather than a five-dimensional optimization problem, it can be solved much more efficiently.

V. Unwrapping-based Estimation

Another approach for parameter estimation of wrapped distributions is based on the concept of data augmentation [5]. The idea is to introduce latent variables that describe how many times each sample was wrapped. For toroidal variables $\underline{x}^{(1)}, \dots, \underline{x}^{(n)} \in [0, 2\pi)^2$, we introduce $2n$ wrapping numbers $[k_1^{(1)}, k_2^{(1)}]^T \dots, [k_1^{(n)}, k_2^{(n)}]^T \in \mathbb{Z}^2$ such that

$$\underline{x}^{(l)} + 2\pi[k_1^{(l)}, k_2^{(l)}]^T = \underline{y}^{(l)}, \quad l = 1, \dots, n$$

where $\underline{y}^{(1)}, \dots, \underline{y}^{(n)} \in \mathbb{R}^2$ represent the random variables before wrapping occurred. If the wrapping numbers were known, the parameters of a bivariate normal distribution could be easily estimated from the unwrapped random variables $\underline{y}^{(1)}, \dots, \underline{y}^{(n)} \in \mathbb{R}^2$. Fisher et al. [6] proposed an expectation maximization (EM) algorithm that tries to estimate the parameters of the PWN distribution by alternating between calculating the probability

$$P(k_1^{(l)}=K_1, k_2^{(l)}=K_2|\underline{\mu}, \mathbf{C}) \propto \mathcal{N}(\underline{x}^{(l)}+2\pi[K_1, K_2]^T; \underline{\mu}, \mathbf{C}) \quad l = 1, \dots, n$$

of the wrapping numbers given the parameters and estimation of the parameters according to

$$\begin{aligned} \underline{\mu} &= \frac{1}{n} \sum_{K_1, K_2} \sum_{l=1}^{n} P(k_1^{(l)}=K_1, k_2^{(l)}=K_2|\underline{\mu}, \mathbf{C}) \\ &\quad \cdot (\underline{x}^{(l)}+2\pi[K_1, K_2]^T) \ , \\ \mathbf{C} &= \frac{1}{n} \sum_{K_1, K_2} \sum_{l=1}^{n} P(k_1^{(l)}=K_1, k_2^{(l)}=K_2|\underline{\mu}, \mathbf{C}) \\ &\quad \cdot (\underline{x}^{(l)}+2\pi[K_1, K_2]^T)(\underline{x}^{(l)} + 2\pi[K_1, K_2]^T)^T \ . \end{aligned}$$

For a practical implementation, suitable maximum values of K_1 and K_2 have to be chosen. The procedure is repeated until convergence.

Because of the high computational cost, some authors have investigated the use of Markov Chain Monte-Carlo (MCMC) methods for unwrapping. Coles [7] proposed a scheme to obtain samples for the wrapping coefficients as well as the parameters $\underline{\mu}$ and $\mathbf{C}$ using the Metropolis–Hastings [21] algorithm. However, implementation of this method involves many choices (e.g., prior densities, a proposal density, a conditional density for cycling though different variables, the burn-in duration, a stopping criterion). As Coles does not give clear recommendations regarding these choices, we decided not to implement his algorithm. A similar method based on Gibbs sampling [22] was proposed by Ravindran [9], but his derivations are limited to scalar distributions. Further work in this area was done by Jona-Lasinio et al. [8] for the univariate case and for wrapped Gaussian processes.

VI. Evaluation

To compare all presented parameter estimation methods (see Table I), we consider the BWN distributions

$$\mathcal{BWN}\left(\underline{x}; \underline{\mu}_a = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{C}_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}\right) \ ,$$
$$\mathcal{BWN}\left(\underline{x}; \underline{\mu}_b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{C}_b = \begin{bmatrix} 1 & 1.2 \\ 1.2 & 2 \end{bmatrix}\right) \ ,$$

which represent cases with weak ($\rho_a \approx 0.35$) and with strong ($\rho_b \approx 0.84$) correlation. From each distribution, we randomly draw a predefined number of samples and then apply all parameter estimation methods to fit a BWN distribution to the samples. Finally, we compare the reconstructed BWN distribution to the original BWN distribution. Because the main difficulty consists in the estimation of the correlation parameter $\rho = \frac{c_{1,2}}{\sqrt{c_{11} c_{2,2}}}$, we evaluate its bias and variance.

For each BWN distribution and every number of samples, we performed 100 Monte-Carlo runs. The results are depicted in Fig. 3 and the results for 300 samples and $\mathcal{BWN}(\underline{x}; \underline{\mu}_b, \mathbf{C}_b)$ are also given in Table II. Beyond the bias and variance of ρ, we also give the computation time and the number of times the parameter estimation failed as a result of a non-positive-definite $\mathbf{C}$ matrix. All algorithms were implemented in MATLAB and the computation times were obtained on a Core i7-2640M with 8 GB RAM. Note, however, that the implementations were not optimized for performance and it might be possible to significantly improve the computation time of some approaches.

Name	bias	variance	time	failures
MLE	-0.001201	0.000196	0.459018s	0
MLE (Jensen)	0.336868	0.002501	0.583134s	0
Jammalamadaka [2]	0.159217	0.028777	0.000446s	29
Jamm.'s coeff. [3]	0.003026	0.000670	0.000658s	0
Jupp's coefficient	-0.002165	0.000324	0.001553s	0
Covariance	0.001294	0.000399	0.001273s	0
Mixed MLE	0.000126	0.000221	0.039276s	0
Unwrapping EM [6]	-0.001246	0.000195	0.677304s	0

TABLE II: Results for 300 samples of $\mathcal{BWN}(\underline{x}; \underline{\mu}_b, \mathbf{C}_b)$.

The results indicate that the method by Jammalamadaka and the MLE approach based on Jensen's inequality yield fairly inaccurate results. The latter appears to be strongly biased even as the number of samples approaches infinity, which can be explained by the approximation of the likelihood function. All other methods are fairly similar for $\mathcal{BWN}(\underline{x}; \underline{\mu}_b, \mathbf{C}_b)$, but for $\mathcal{BWN}(\underline{x}; \underline{\mu}_a, \mathbf{C}_a)$, the method based on Jupp's correlation coefficient is also quite inaccurate. Furthermore, it can be seen that Jammalamadaka's method fails very frequently even for many samples, particularly in the case of $\mathcal{BWN}(\underline{x}; \underline{\mu}_b, \mathbf{C}_b)$. The other moment-based methods also fail in a few cases, but only when there are too few samples. The remaining methods are always successful. As far as the runtime is concerned, all moment-based methods are very fast. The mixed MLE approach is somewhat slower, but still fast enough for many applications. The numerical MLE, the MLE based on Jensen's inequality and the unwrapping method are very slow and seem impractical for most real-time applications.

All in all, the moment-based methods based on Jammalamadaka's correlation coefficient and the covariance matrix performed very well and resulted in high accuracy at a small computational cost. However, in certain cases these methods can fail. Thus, it might be a good option to fall back to the mixed MLE method in the case that $\mathbf{C}$ is not positive definite (or very close to singular). This method is somewhat slower, but it is the fastest method that works in all cases, and it is also highly accurate.

VII. Conclusion

This paper summarized a number of methods for parameter estimation of the bivariate wrapped normal distribution found in literature and proposed several new methods as well. All methods were compared in a thorough evaluation and the new method based on matching a submatrix of the covariance as well as the new method based on a combination of maximum likelihood and moment matching were shown to outperform state-of-the-art approaches.

Even though we did not consider weighted samples in this paper for reasons of clarity, all discussed methods can easily

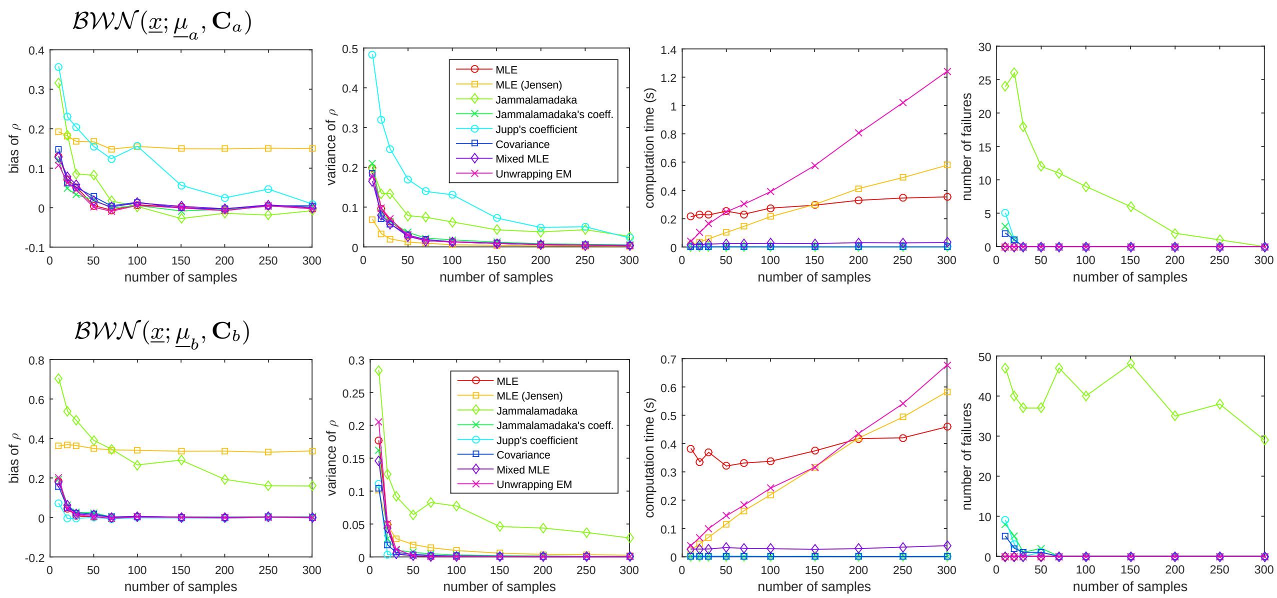

Fig. 3: Evaluation results based on 100 runs.

be applied to weighted samples as well. Most approaches can, in principle, be generalized to the n-torus, but parameter estimation on the n-torus is still an open problem because some methods have exponential computational complexity with respect to the dimension, whereas others may have an increased difficulty of obtaining a positive definite parameter matrix $\mathbf{C}$.

Furthermore, the approaches discussed in this paper may be applied to the problem of fitting mixtures, for example using the EM algorithm (see e.g., [4]). Also, an application to the problem of recursive toroidal filtering is highly interesting [3].

References

[1] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Wiley, 1999.

[2] S. R. Jammalamadaka and Y. Sarma, "A Correlation Coefficient for Angular Variables," *Statistical Theory and Data Analysis II*, pp. 349–364, 1988.

[3] G. Kurz, I. Gilitschenski, M. Dolgov, and U. D. Hanebeck, "Bivariate Angular Estimation Under Consideration of Dependencies Using Directional Statistics," in *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2014)*, Los Angeles, California, USA, Dec. 2014.

[4] J. Traa, "Expectation Maximization — Math and Pictures," http://cal.cs.illinois.edu/ johannes/research/EM%20derivations.pdf, Feb. 2014.

[5] M. A. Tanner and W. H. Wong, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.

[6] N. I. Fisher and A. J. Lee, "Time Series Analysis of Circular Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 56, no. 2, pp. 327–339, 1994.

[7] S. Coles, "Inference for Circular Distributions and Processes," *Statistics and Computing*, vol. 8, no. 2, pp. 105–113, 1998.

[8] G. Jona-Lasinio, A. Gelfand, and M. Jona-Lasinio, "Spatial Analysis of Wave Direction Data Using Wrapped Gaussian Processes," *The Annals of Applied Statistics*, vol. 6, no. 4, pp. 1478–1498, Dec. 2012.

[9] P. Ravindran and S. K. Ghosh, "Bayesian Analysis of Circular Data Using Wrapped Distributions," *Journal of Statistical Theory and Practice*, vol. 5, no. 4, pp. 547–561, 2011.

[10] P. E. Jupp and K. V. Mardia, "A General Correlation Coefficient for Directional Data and Related Regression Problems," *Biometrika*, vol. 67, no. 1, pp. 163–173, 1980.

[11] R. A. Johnson and T. Wehrly, "Measures and Models for Angular Correlation and Angular-Linear Correlation," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 2, pp. 222–229, 1977.

[12] G. Kurz, "Directional Estimation for Robotic Beating Heart Surgery," Ph.D. dissertation, Karlsruhe Institute of Technology, Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe, Germany, 2015.

[13] K. V. Mardia and M. L. Puri, "A Spherical Correlation Coefficient Robust Against Scale," *Biometrika*, vol. 65, no. 2, pp. 391–395, 1978.

[14] L.-P. Rivest, "Some Statistical Methods for Bivariate Circular Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 1, pp. 81–90, 1982.

[15] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[16] A. Roy, S. K. Parui, and U. Roy, "SWGMM: A Semi-wrapped Gaussian Mixture Model for Clustering of Circularlinear Data," *Pattern Analysis and Applications*, 2014.

[17] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Efficient Evaluation of the Probability Density Function of a Wrapped Normal Distribution," in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2014)*, Bonn, Germany, Oct. 2014.

[18] ——, "Recursive Bayesian Filtering in Circular State Spaces," *arXiv preprint: Systems and Control (cs.SY)*, Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.05151

[19] ——, "Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[20] S. L. Shmakov, "A Universal Method of Solving Quartic Equations," *International Journal of Pure and Applied Mathematics*, vol. 71, no. 2, pp. 251–259, 2011.

[21] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[22] G. Casella and E. I. George, "Explaining the Gibbs Sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.