Nonlinear Toroidal Filtering
Based on Bivariate Wrapped Normal Distributions

Gerhard Kurz, **Florian Pfaff**, and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
gerhard.kurz@kit.edu, florian.pfaff@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—Estimation of periodic quantities such as angles or phase values is a common problem. However, standard approaches, for example the Kalman filter and extensions thereof, have difficulties when estimating periodic quantities. To address this problem, circular filtering algorithms have been proposed but they are limited to just a single angle. In order to deal with multiple, possibly correlated angles, toroidal filtering algorithms are necessary. We have previously proposed a bivariate filtering algorithm on the torus [1] that is limited to identity system and measurement models. In this paper, we show how the algorithm can be extended to handle nonlinear system and measurement models. The novel approach relies on the bivariate wrapped normal distribution for representing the uncertainty and it makes use of a deterministic sampling scheme for the torus. We provide a thorough evaluation of the proposed method using simulations.**

Keywords*—*circular correlation, deterministic sampling, recursive estimation, directional statistics, Bayesian filtering

I. Introduction

Periodic quantities are ubiquitous and of importance in a wide variety of fields ranging from geology [2] and biology [3] to signal processing [4] and robotics [5]. Such quantities often take the form of an angle or the phase of a signal. Ignoring their periodic nature in estimation tasks, e.g., by using the popular Kalman filter, can lead to unintended results. Estimating periodic quantities in a robust and semantically meaningful manner thus lays the foundation for high quality algorithms to extract information about a system and for influencing such a system's behavior in control tasks [6].

A variety of filters for angular quantities have been proposed, such as [7], [8] for simple models and [9]–[12] for more sophisticated models. Most filters cannot be trivially extended to a higher number of variates to estimate multiple possibly correlated angles. We have proposed a bivariate wrapped normal filter [1], but it is limited to simple identity system and measurement models. For this reason, we will extend the filter [1] to nonlinear models in this paper.

There is also a multivariate Fourier filter [13] that can handle nonlinear models and an arbitrary number of variates, but it is computationally fairly complex due to its very general density representation. Another alternative is the particle filter[14], which can be easily adapted to the torus. However, the particle filter is nondeterministic and its performance depends on the random values that are drawn when running the filter. Also, it typically requires a large number of particles and can suffer from the problem of particle degeneration.

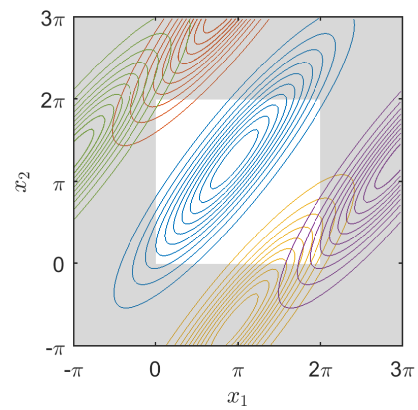


Fig. 1: Bivariate wrapped normal distribution on the torus $[0, 2\pi)^2$, which is indicated by the white square. Both the x_1 and the x_2 axes are 2π-periodic. It can be seen how the normal distribution shifted by multiples of 2π in either dimension wraps back onto the torus. All probability mass that is wrapped to the same point is summed up to obtain the value of the density.

Most circular filters assume that the resulting densities of filter and prediction steps are distributed according to a certain circular distribution such as the wrapped normal or von Mises distribution. The methods to approximate other occurring densities with the assumed density need to be modified to be applicable to the multivariate case. In this paper, we will focus on a filter that relies on the bivariate wrapped normal distribution (see Fig. 1).

Many nonlinear filters on linear domains (e.g., linear regression Kalman filters [15]) consist of three essential building blocks. The first is a sampling scheme that yields a finite set of points in the state space representative of the sampled density (e.g., [16]–[18]). For example, a naive approach to obtain such samples is to use random sampling. However, deterministic sampling schemes are often employed, which maintain certain properties of the distribution, for example the mean and the covariance.

The second building block is the processing of the samples in the prediction and filter step. This is usually performed either in a way that yields a new set of (potentially weighted) samples or in a way that yields parameters, e.g., mean and covariance, describing the filter result.

The last important building block, which can be seen as the counterpart of the sampling, is a matching step in which we aim to find a continuous density among a family of distributions that is supposed to be close to the actual posterior density. For this, the samples (or parameters) obtained as the result of the filter or prediction step are to be matched as closely as possible using the continuous density. In the linear case, the mean and covariance are often matched using Gaussian distributions as this yields a unique result.

For both the sampling and the matching part of the approaches, parameters describing relevant properties of the distribution and the set of samples are essential. On linear domains, power moments are often used in this context. For scalar angular quantities, trigonometric moments have found application as they serve as a useful analogue on circular domains. Trigonometric moments, however, do not feature an inherent way to represent correlations for multivariate angular quantities. Therefore, additional parameters such as circular correlation coefficients need to be considered in addition to trigonometric moments. As these are not present in the scalar case, modifications of circular algorithms are necessary to deal with filtering problems on the torus.

So far, progress has been made on sampling schemes [19] for bivariate circular quantities and estimation techniques to determine the parameters of bivariate wrapped normal distributions [20] have been proposed. These advances give us the tools for the first and the third building block of a nonlinear filter for bivariate circular estimation problems. In this paper, we focus on the second block and build upon our prior results to derive a complete sample-based filter that allows for recursive Bayesian estimation on the torus, in which all densities are approximated using bivariate wrapped normal distributions.

II. Statistics on the Torus

In this section, we give a brief introduction to the essential concepts for statistical methods on the torus. The torus is parameterized as $[0, 2\pi)^2$, i.e., the Cartesian product of two circles. One of the most common probability densities on the torus is the bivariate wrapped normal distribution. This distribution was mentioned in the 1970s by Johnson and Wehrly [21, Example 7.3]. It can be seen as a generalization of the univariate wrapped normal distribution proposed by Schmidt in 1917 [22].

Definition 1 (Bivariate Wrapped Normal Distribution) *The bivariate wrapped normal (BWN) distribution is defined by the probability density function*

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

where $\underline{x} = [x_1, x_2]^T, \underline{\mu} = [\mu_1, \mu_2]^T \in [0, 2\pi)^2$ *and*

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{1,2} & c_{2,2} \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

is symmetric positive definite.

An advantage of the BWN distribution compared with certain other toroidal distributions is that its normalization constant is trivial to compute. Also, it is closed under convolution, i.e., addition of random vectors modulo 2π. The convolution of two BWN densities $\mathcal{BWN}(\underline{\mu}_a, \mathbf{C}_a)$ and $\mathcal{BWN}(\underline{\mu}_b, \mathbf{C}_b)$ is given by

$$\mathcal{BWN}((\underline{\mu}_a + \underline{\mu}_b) \mod 2\pi, \mathbf{C}_a + \mathbf{C}_b)$$

according to [1, Lemma 4]. The BWN distribution is not closed under multiplication of densities, which complicates the measurement update in the identity case (see [1, Sec. III-B]). However, this limitation does not affect the nonlinear measurement update considered here because the proposed method uses a progressive reweighting technique that does not rely on the multiplication of two BWN densities. Also, the likelihood is not BWN distributed due to the effect of the nonlinear measurement model anyway.

Another toroidal probability distribution commonly found in literature is the bivariate von Mises distribution, which has been considered in different variations [23], [24]. Compared with the BWN distribution, it has certain advantages (e.g., it belongs to the exponential family) and disadvantages (such as its complicated normalization constant, the fact that it can become bimodal, and that it is not closed under convolution). In [25], we presented an algorithm for Bayesian fusion of toroidal data using the bivariate von Mises distribution. The advantage of that approach is a closed-form solution for the fusion of multiple densities, but it does not include a recursive filter. Even though many of the ideas presented in the following could also be applied in conjunction with the bivariate von Mises distribution, we will exclusively focus on the BWN distribution.

Aside from continuous distributions on the torus, we are also interested in discrete distributions, i.e., a set of weighted samples.

Definition 2 (Wrapped Dirac Mixture) *A wrapped Dirac mixture with weights* $w_1, \dots, w_L > 0, \sum_{j=1}^L w_j = 1$ *and positions* $\underline{\beta}_1, \dots, \underline{\beta}_L \in [0, 2\pi)^2$ *is given by*

$$\mathcal{BWD}(\underline{x}; \underline{\beta}_1, \dots, \underline{\beta}_L, w_1, \dots, w_L) = \sum_{j=1}^L w_j \delta(\underline{x} - \underline{\beta}_j) \ ,$$

where $\delta(\cdot)$ *refers to the Dirac delta distribution.*

A useful concept in scalar circular statistics is that of the n-th trigonometric moment given by $\mathbb{E}(\exp(inx))$, where i is the imaginary unit. We define a straightforward generalization to the toroidal case as follows.

Definition 3 (Trigonometric Moments) *The* n*-th trigonometric moment of a random variable* $\underline{x}$ *on the torus is given by*

$$\underline{m}_n = \begin{bmatrix} m_{n,1} \\ m_{n,2} \end{bmatrix} = \begin{bmatrix} \mathbb{E}(\exp(inx_1)) \\ \mathbb{E}(\exp(inx_2)) \end{bmatrix} \in \mathbb{C}^2 \ .$$

Note that each component is a complex number. For $n = 1$, its complex argument describes the direction and its complex norm describes the uncertainty, i.e., the first trigonometric moment on the circle can be compared with both mean and covariance in $\mathbb{R}^n$.

For the BWN distribution, the n-th trigonometric moment can be computed according to

$$\underline{m}_n = \begin{bmatrix} \exp(in\mu_1 - n^2 c_{1,1}/2) \\ \exp(in\mu_2 - n^2 c_{2,2}/2) \end{bmatrix}$$

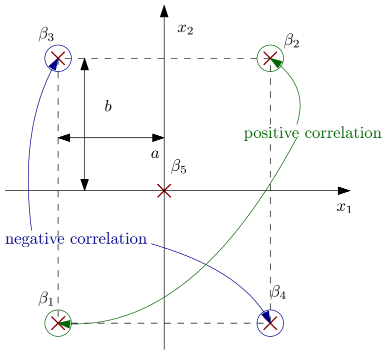


Fig. 2: Idea of the deterministic sampling scheme. The sample locations are noted by $\beta_1, \ldots, \beta_5$. The values a and b affect the spread of the samples. Large weights for β_1 and β_2 result in a positive correlation and large weights for β_3 and β_4 result in a negative correlation.

as shown in [1, Lemma 1]. It can be seen that the trigonometric moments are independent of the correlation parameter $c_{1,2}$. The circular mean in each dimension $j \in \{1, 2\}$ can be obtained according to

$$\mu_j = \operatorname{atan2}(\Im(m_{1,j}), \Re(m_{1,j})) \ .$$

Over the years, a number of different correlation coefficients for circular quantities have been proposed, e.g., [21], [26], [27]. In the following, we will use the correlation coefficient by Jammalamadaka and Sarma [28] due to its nice mathematical properties.

Definition 4 (Circular Correlation Coefficient) *The correlation coefficient for general periodic random variables x_1, x_2 with circular mean μ_1 and μ_2, respectively, is given by [28]*

$$r_c = \frac{\mathbb{E}(\sin(x_1 - \mu_1) \cdot \sin(x_2 - \mu_2))}{\mathbb{E}(\sin^2(x_1 - \mu_1) \cdot \mathbb{E}(\sin^2(x_2 - \mu_2)))} \ .$$

According to [1, Lemma 2], [28], the circular correlation coefficient for the specific case of a BWN distribution is given by

$$r_c = \frac{\sinh(c_{1,2})}{\sqrt{\sinh(c_{1,1}) \sinh(c_{2,2})}} \ .$$

III. BWN Operations

Before we can derive the novel filtering algorithm, we need to discuss how certain elementary operations can be performed on the BWN distribution, namely deterministic sampling and parameter estimation.

A. Deterministic Sampling

In this section, we present a deterministic sampling scheme for toroidal distributions such as the BWN originally published in [19]. As the sampling scheme only relies on trigonometric moments and the circular correlation coefficient, it is also applicable to many other toroidal distributions for which these quantities can be computed. The key idea is to find a set of weighted samples on the torus such that the first trigonometric moment and the circular correlation coefficient are matched.

In [19], we have derived the following solution with five weighted samples. For $\underline{\mu} = [0, 0]^T$, the samples are given by

$$\begin{aligned}
\underline{\beta}_1 &= [-a, -b]^T \mod 2\pi \ , \\
\underline{\beta}_2 &= [a, b]^T \mod 2\pi \ , \\
\underline{\beta}_3 &= [-a, b]^T \mod 2\pi \ , \\
\underline{\beta}_4 &= [a, -b]^T \mod 2\pi \ , \\
\underline{\beta}_5 &= [0, 0]^T \mod 2\pi
\end{aligned}$$

with weights $w_1 = w_2 > 0, w_3 = w_4 > 0$ and $w_5 = 1 - w_1 - w_2 - w_3 - w_4 > 0$ and parameters $a, b \in [0, \pi]$.

Intuitively, the parameters can be understood as illustrated in Fig. 2. The variables a and b affect the location of the samples in each dimension. They represent the variance $c_{1,1}$ and $c_{2,2}$, respectively, i.e., the samples are spread further apart for larger variances and closer together for small variances. The correlation is captured by the adjusting weights, i.e., for a positive correlation, w_1 and w_2 become larger and for a negative correlation, w_3 and w_4 become larger.

We define the abbreviation $\tilde{w} = w_1 + w_3$. By equating the first trigonometric moment of the sample set to a given moment $\underline{m}_1$, we obtain solutions for a and b according to

$$\begin{aligned}
a &= \arccos((m_{1,1} - 1)/(2\tilde{w}) + 1) \ , \\
b &= \arccos((m_{1,2} - 1)/(2\tilde{w}) + 1) \ .
\end{aligned}$$

Furthermore, by equating the circular correlation coefficient of the sample set to a given circular correlation coefficient r_c, we find the solutions for the weights

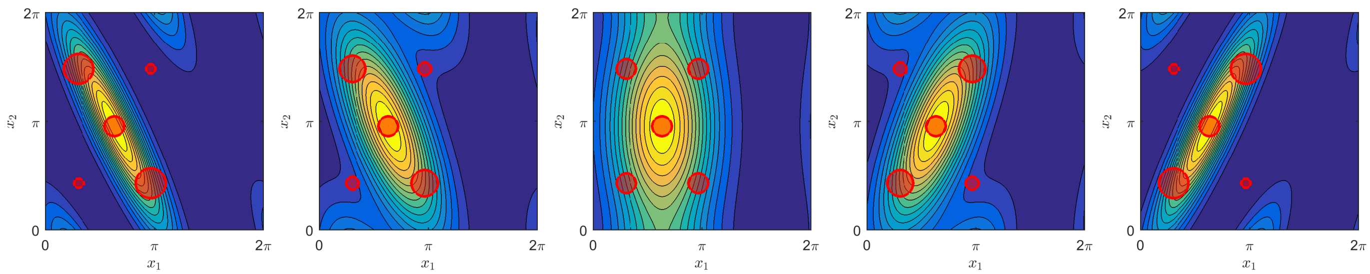
$$w_5 = 1 - 2\tilde{w}, \quad w_3 = (\tilde{w} - r_c|\tilde{w}|)/2, \quad w_1 = \tilde{w} - w_3 \ .$$

Finally, $\tilde{w}$ has to be chosen. We decided on $\tilde{w} = \frac{1}{5}$ because it guarantees positive weights in all cases and it results in uniform weights in the uncorrelated case $r_c = 0$. Some examples of the generated samples for different BWN distributions are shown in Fig. 3.

B. Parameter Estimation

Now that we have derived a method for approximating a continuous toroidal density with a set of weighted samples, we are interested in the reverse problem, i.e., fitting a BWN distribution to a set of weighted samples. We have discussed this problem in detail in [20]. That paper is limited to the case of uniform weights, but the generalization to the weighted case considered here is fairly straightforward.

A common technique for parameter estimation is the maximum likelihood estimation (MLE) method [20, Sec. III], where the parameters of the density are computed in such a way that the likelihood of obtaining the (weighted) sample set is maximized. While there is a closed form solution for certain distributions (e.g., the Gaussian), applying the maximum likelihood method to the BWN distribution requires numerical optimization, which is fairly costly. Alternatively, approximations based on Jensen's inequality can be performed [29], but the resulting parameters can be highly suboptimal.



(a) $\rho = -0.9$. (b) $\rho = -0.7$. (c) $\rho = 0$. (d) $\rho = 0.7$. (e) $\rho = 0.9$.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{BWN}\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & 2\rho \\ 2\rho & 4 \end{bmatrix} \right)$$

Fig. 3: Examples for deterministic samples of the BWN distribution given above for different correlation parameters $\rho \in (-1, 1)$. The size of each sample corresponds to its weight. Observe that both the x_1 and x_2-axes are 2π-periodic.

An alternative approach consists in moment-based methods [20, Sec. IV], where the parameters of the density are obtained using moment matching. Matching the first trigonometric moment $\underline{m}_1$ yields

$$\mu_1 = \operatorname{atan2}(\Im(m_{1,1}), \Re(m_{1,1})) \ , \quad (1)$$
$$\mu_2 = \operatorname{atan2}(\Im(m_{1,2}), \Re(m_{1,2})) \ , \quad (2)$$
$$c_{1,1} = -2 \log(|m_{1,1}|) \ , \quad (3)$$
$$c_{2,2} = -2 \log(|m_{1,2}|) \ . \quad (4)$$

Furthermore, one would typically try to match a quantity related to the correlation, for example one of the circular correlation coefficients found in literature. Moment-based methods are typically very fast to compute. Also, matching the same correlation coefficient that is used in deterministic sampling has the benefit that converting between continuous and discrete distributions is a lossless operation, i.e., repeated deterministic sampling and matching of a continuous density does not change the density. However, the main problem with the moment-based approach is that a solution does not always exist, especially in cases of high correlation.

For this reason, we have proposed the so-called mixed MLE method. The idea of this method is to compute all parameters except $c_{1,2}$ by matching the first trigonometric moment according to (1)–(4), but estimating $c_{1,2}$ using the maximum likelihood approach. Thus, we solve the univariate optimization problem

$$\arg\max_{c_{1,2}} \left(\sum_{j=1}^{5} w_j \log \mathcal{BWN}\left(\underline{\beta}_j; \underline{\mu}, \begin{bmatrix} c_{1,1} & c_{1,2} \\ c_{1,2} & c_{2,2} \end{bmatrix} \right) \right)$$

using numerical methods. The advantage of this method is that a solution will always be found and that the computational cost is much lower than that of the general maximum likelihood method due to the reduction to a one-dimensional optimization problem.

In order to reduce the computational cost, we use the moment-based method from [1, Lemma 3] by default and to fall back to the mixed MLE method only when the moment-based method fails.

IV. Nonlinear Filter

Based on the operations introduced before, we can now derive a nonlinear filtering algorithm that relies on BWN distributions. In the following, we will use k as the time index.

A. Prediction Step

First, we consider the prediction step. We distinguish predictions with additive and non-additive noise.

1) Additive Noise: In the additive noise case, the system model is given by

$$\underline{x}_{k+1} = \underline{a}_k(\underline{x}_k) + \underline{w}_k \ ,$$

where $\underline{a}_k : [0, 2\pi)^2 \to [0, 2\pi)^2$ is the system function and $\underline{w}_k$ is BWN-distributed noise.

In order to perform the prediction in this setting, we generalize the method from [30] to the bivariate case. First, we draw samples from the prior density using the deterministic sampling method introduced before. Then we propagate those samples through the system function $\underline{a}_k(\cdot)$ and fit a BWN distribution to the propagated samples. Finally, we compute the convolution of this density with the noise density (see Sec. II). Pseudocode of this method is shown in Algorithm 1.

Algorithm 1: Prediction with additive noise.

Input: $\mathcal{BWN}(\underline{\mu}_k^e, \mathbf{C}_k^e)$, $\mathcal{BWN}(\underline{\mu}_k^w, \mathbf{C}_k^w)$, $\underline{a}_k(\cdot)$
Output: $\mathcal{BWN}(\underline{\mu}_{k+1}^p, \mathbf{C}_{k+1}^p)$
$\underline{\beta}_1, \dots, \underline{\beta}_5, w_1, \dots w_5 \leftarrow$sampleDeterministic$(\underline{\mu}_k^e, \mathbf{C}_k^e)$;
for $j \leftarrow 1$ **to** 5 **do**
  $\underline{\tilde{\beta}}_j \leftarrow \underline{a}_k(\underline{\beta}_j)$;
end
$\underline{\mu}, \mathbf{C} \leftarrow$ parameterEstimation$(\underline{\tilde{\beta}}_1, \dots, \underline{\tilde{\beta}}_5, w_1, \dots w_5)$;
`/* perform convolution */`
$\mathcal{BWN}(\underline{\mu}_{k+1}^p, \mathbf{C}_{k+1}^p) \leftarrow \mathcal{BWN}(\underline{\mu}, \mathbf{C}) * \mathcal{BWN}(\underline{\mu}_k^w, \mathbf{C}_k^w)$;

2) Non-additive Noise: If the system noise is non-additive, we assume a system model

$$\underline{x}_{k+1} = \underline{a}_k(\underline{x}_k, \underline{w}_k) \ ,$$

where $\underline{a}_k : [0, 2\pi)^2 \times W \to [0, 2\pi)^2$ is the system function and $\underline{w}_k$ is noise in some noise space W. The noise space W can, for example, be the torus $[0, 2\pi)^2$, a real vector space $\mathbb{R}^n$ or even a discrete set. Furthermore, we assume that a (preferably deterministic) method for sampling from $\underline{w}_k$ is given.

To perform the prediction step in the presence of non-additive noise, we generalize the approach given in [9, Algorithm 3] to the bivariate case. First, we sample the prior state density using the toroidal deterministic sampling scheme. Furthermore, we obtain a set of (weighted) samples from the system noise $\underline{w}_k$. Then, we consider the Cartesian product[1] of the state and noise samples, i.e., all possible combinations. Each pair of a state and a noise sample is propagated through $\underline{a}_k(\cdot, \cdot)$ and the weight of the resulting sample is computed by multiplying the weights of the two samples involved. Finally, we reapproximate the propagated samples with a BWN distribution. Pseudocode can be found in Algorithm 2.

Algorithm 2: Prediction with non-additive noise.

Input: $\mathcal{BWN}(\underline{\mu}_k^e, \mathbf{C}_k^e)$, $\underline{a}_k(\cdot, \cdot)$, noise samples $\underline{\beta}_1^w, \dots, \underline{\beta}_{L^w}^w, w_1^w, \dots, w_{L^w}^w$

Output: $\mathcal{BWN}(\underline{\mu}_{k+1}^p, \mathbf{C}_{k+1}^p)$

$\underline{\beta}_1, \dots, \underline{\beta}_5, w_1, \dots w_5 \leftarrow$sampleDeterministic$(\underline{\mu}_k^e, \mathbf{C}_k^e)$;

for $j \leftarrow 1$ **to** 5 **do**

  for $l \leftarrow 1$ **to** L^w **do**

    $\tilde{\underline{\beta}}_{L^w \cdot (j-1)+l} \leftarrow \underline{a}_k(\underline{\beta}_j, \underline{\beta}_l^w)$;

    $\tilde{w}_{L^w \cdot (j-1)+l} \leftarrow w_j \cdot w_l^w$;

  end

end

$\mathcal{BWN}(\underline{\mu}_{k+1}^p, \mathbf{C}_{k+1}^p) \leftarrow$ parameterEstimation$(\tilde{\underline{\beta}}_1, \dots, \tilde{\underline{\beta}}_{5 \cdot L^w}, \tilde{w}_1, \dots \tilde{w}_{5 \cdot L^w})$;

B. Measurement Update Step

For the measurement update step, we assume that the likelihood $f(\underline{z}_k | \underline{x}_k)$ is given. For an additive noise measurement equation, the likelihood can easily be derived as discussed in [31, Sec. II].

In principle, the measurement update can be carried out according to the Bayes theorem by sampling the prior density, multiplying each weight with the likelihood and reapproximating the weighted samples with a continuous density. However, this approach often leads to sample degeneration, i.e., many samples get assigned a weight $w \approx 0$ and the effective sample size is significantly reduced. This is particularly problematic in our case because our toroidal deterministic sampling scheme uses just five samples.

To deal with this problem, we propose the use of a so-called progressive measurement update. The same approach has successfully been used on the unit circle [31], [9, Algorithm 4] and for Gaussians in $\mathbb{R}^n$ [32]. The basic idea of the progressive measurement update consists in rewriting the Bayes formula according to

$$\begin{aligned} f(\underline{x}_k | \underline{z}_k) &\propto f(\underline{z}_k | \underline{x}_k) f(\underline{x}_k) \\ &= f(\underline{z}_k | \underline{x}_k)^{\lambda_1} \cdot \ldots \cdot f(\underline{z}_k | \underline{x}_k)^{\lambda_s} f(\underline{x}_k) \ , \end{aligned}$$

where $\lambda_1, \dots, \lambda_s > 0$ and $\sum_{j=1}^s \lambda_j = 1$. Intuitively, the multiplication of the prior density with the likelihood is thus split into s multiplications, where each of them corresponds to a partial update. The key advantage is that a partial update is less likely to result in particle degeneration, provided the step sizes λ_j are chosen sufficiently small. After each progression step, we reapproximate the samples with a BWN density and use the deterministic sampling scheme to obtain new samples, which tend to have a more even weight distribution.

In order to derive the step size, we consider the condition

$$\frac{w_{\text{min}}}{w_{\text{max}}} \geq R \in (0, 1) \ ,$$

i.e., w_{min}, the smallest weight after reweighting divided by w_{max}, the largest weight after reweighting, has to be larger than a predefined constant R. If the prior Dirac mixture has equal weights $w_1 = \dots = w_L$, we have

$$\frac{w_{\text{min}}}{w_{\text{max}}} = \frac{\min_j f(\underline{z}_k | \underline{\beta}_j)^\lambda}{\max_j f(\underline{z}_k | \underline{\beta}_j)^\lambda} \ ,$$

which can easily be solved for λ. This idea has already been used in [31] and [32]. In [9, Sec. VI-B-3], we have generalized this idea to the case of non-uniform weights by providing the lower bound

$$\frac{w_{\text{min}}}{w_{\text{max}}} \geq \frac{\min_j w_j}{\max_j w_j} \cdot \frac{\min_j f(\underline{z}_k | \underline{\beta}_j)^\lambda}{\max_j f(\underline{z}_k | \underline{\beta}_j)^\lambda} \ .$$

However, this might lead to a very small step size λ if $\min_j w_j \approx 0$, which can occur in the sampling scheme given in Sec. III-A for strongly correlated densities. For this reason, we do not consider the weights after reweighting in this paper, but only the effect the likelihood has on the weights, i.e., the change of the weight

$$\frac{\min_j f(\underline{z}_k | \underline{\beta}_j)^\lambda}{\max_j f(\underline{z}_k | \underline{\beta}_j)^\lambda} \geq R \ .$$

Solving for λ yields

$$\lambda \leq \log(R) / \log \left(\frac{\min_j f(\underline{z}_k | \underline{\beta}_j)}{\max_j f(\underline{z}_k | \underline{\beta}_j)} \right) \ .$$

The resulting method for computing the measurement update is summarized in Algorithm 3.

V. Evaluation

In this section, we evaluate the proposed approach and compare it with a version of the UKF that has been modified for toroidal estimation and with an SIR particle filter. We consider the robot arm shown in Fig. 4 as our example system[2]. The

[1] An alternative to the use of the Cartesian product would be to perform deterministic sampling in the joint space of state and measurement directly. However, this depends on the specific noise space W and can be tricky in general.

[2] A similar scenario was considered in [13]. The main difference is the assumption that not only the end effector but also joint 2 can be observed, which ensures a unique solution for the joint angles and avoids multimodality.

Algorithm 3: Measurement update.

Input: $\mathcal{BWN}(\underline{\mu}_k^p, \mathbf{C}_k^p)$, $\hat{\underline{z}}$, likelihood $f(\underline{z}_k|\underline{x}_k)$
Output: $\mathcal{BWN}(\underline{\mu}_k^e, \mathbf{C}_k^e)$
$s \leftarrow 0$;
$\underline{\mu} \leftarrow \underline{\mu}_k^p$;
$\mathbf{C} \leftarrow \mathbf{C}_k^p$;
while $\sum_{j=1}^{s} \lambda_j < 1$ **do**
  $s \leftarrow s+1$;
  $\underline{\beta}_1, \dots, \underline{\beta}_5, w_1, \dots w_5 \leftarrow$sampleDeterministic$(\underline{\mu}, \mathbf{C})$;
  $\hat{\lambda} \leftarrow \log(R) / \log\left(\frac{\min_j f(\underline{z}_k|\underline{\beta}_j)}{\max_j f(\underline{z}_k|\underline{\beta}_j)}\right)$;
  $\lambda_s \leftarrow \min(1 - \sum_{j=1}^{s-1} \lambda_j, \hat{\lambda})$;
  for $j \leftarrow 1$ **to** 5 **do**
    $w_j \leftarrow w_j \cdot f(\hat{\underline{z}}_k)|\underline{\beta}_j)^{\lambda_s}$;
  end
  $\underline{\mu}, \mathbf{C} \leftarrow$ parameterEstimation$(\underline{\beta}_1, \dots, \underline{\beta}_5, w_1, \dots w_5)$;
end
$\underline{\mu}_k^e \leftarrow \underline{\mu}$;
$\mathbf{C}_k^e \leftarrow \mathbf{C}$;

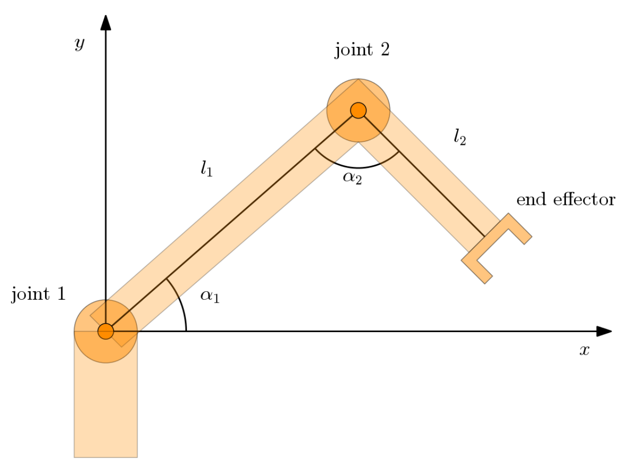


Fig. 4: Robot arm used in the evaluation scenario.

state of the system consists of the two joint angles $[\alpha_1, \alpha_2]^T$. For our purpose, we assume that both joint angles can cover the full range $[0, 2\pi)$. The system model is given by

$$\underline{x}_{k+1} = \underline{x}_k + [0.1, 0.2]^T + \underline{v}_k \mod 2\pi \ ,$$

i.e., we add a constant angular velocity to the state in every time step.

We assume that Cartesian measurements are obtained from the end effector and joint 2 (near α_2). Thus, the measurement model is given by

$$\underline{z}_k = \begin{bmatrix} l_1 \cos(x_{k,1}) + l_2 \cos(x_{k,1} + x_{k,2} + \pi) \\ l_1 \sin(x_{k,1}) + l_2 \sin(x_{k,1} + x_{k,2} + \pi) \\ l_1 \cos(x_{k,1}) \\ l_1 \sin(x_{k,1}) \end{bmatrix} + \underline{v}_k \ ,$$

where l_1 and l_2 are the lengths of the robot arm segments.

At each time step k, we consider the error measure

$$\sqrt{\sum_{j=1}^{2} d^2(x_{k,j}, x_{k,j}^{\text{true}})}$$

between the estimated state $\underline{x}_k$ and the true state $\underline{x}_k^{\text{true}}$, where $d(\cdot,\cdot)$ is the angular error

$$d(a,b) = \min(|a-b|, 2\pi - |a-b|) \ .$$

We simulate two different scenarios that differ in the noise parameters. In the low noise scenario, we have

$$\underline{w} \sim \mathcal{BWN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 0.0001 \cdot \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right) \ ,$$
$$\underline{v} \sim \mathcal{BWN}([0,0,0,0]^T, 0.0001 \cdot \mathbf{I}_{4\times4}) \ ,$$

and in the high noise scenario, we have

$$\underline{w} \sim \mathcal{BWN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 0.1 \cdot \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right) \ ,$$
$$\underline{v} \sim \mathcal{BWN}([0,0,0,0]^T, 1 \cdot \mathbf{I}_{4\times4}) \ .$$

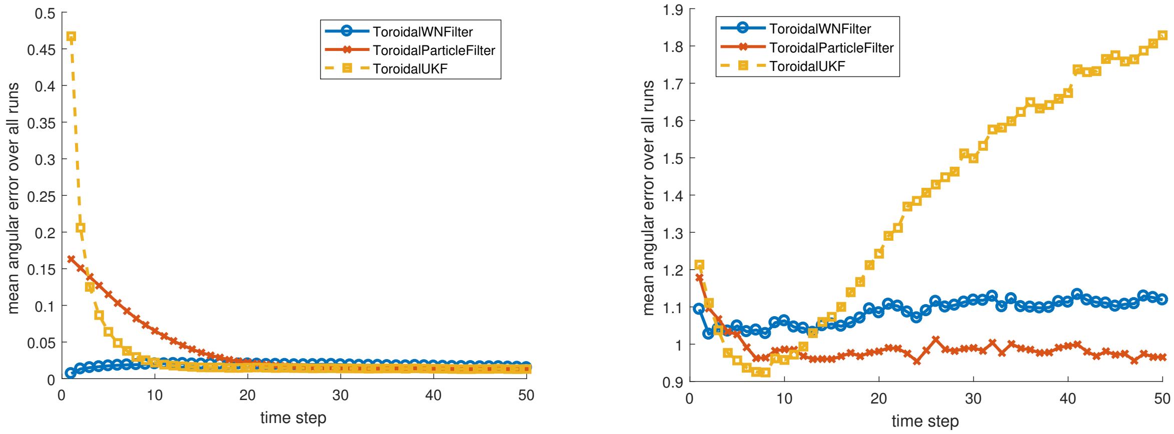
Observe that the system noise is correlated in both cases.

The number of particles for the particle filter was chosen as 500 and the progression threshold for the measurement update of the proposed filter was set to $R = 0.02$. The true initial state is always $x_0^{\text{true}} = [3,3]^T$ and the initial estimate is given by

$$x_0 \sim \mathcal{BWN}([2,4], \mathbf{I}_{2\times2}) \ .$$

In total, we performed 1000 Monte Carlo simulations and averaged the resulting error over all runs. The results are depicted in Fig. 5. In the low noise scenario, the proposed filter converges much more quickly than the UKF and the particle filter. After convergence, it yields comparable results. Especially the particle filter has trouble in the beginning because it suffers from particle degeneration due to the narrow likelihood. In the high noise scenario, the UKF is by far the worst filter because it does not properly account for the periodicity. The proposed filter and the particle filter are comparable with a slight advantage for the particle filter. As a result, it can be seen that the proposed filter works well in scenarios with both low and high noise, whereas the state-of-the-art filters have trouble with at least one of the scenarios.

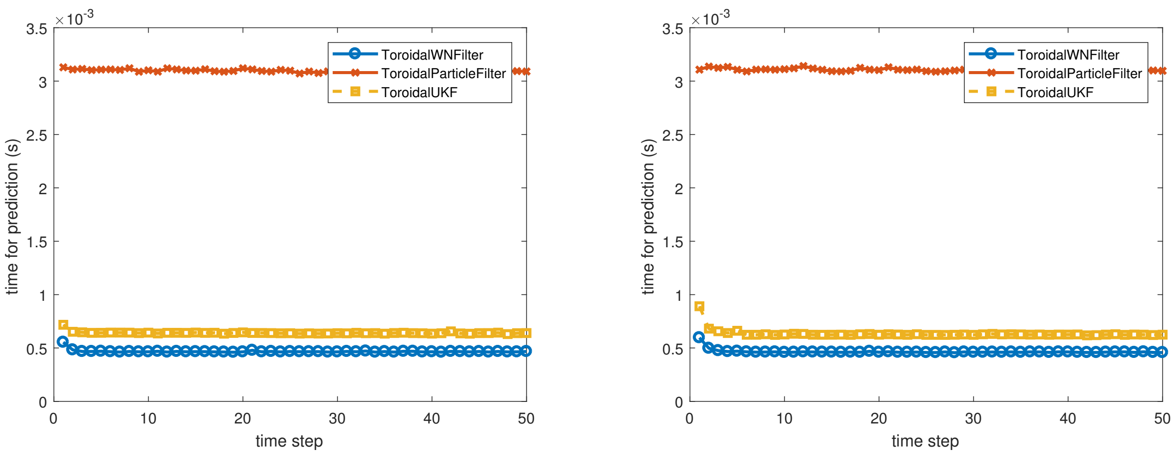
We also evaluated the runtime cost of the three considered methods. The evaluation was performed on a laptop with an Intel Core i7-2640M CPU, 8 GB RAM, and MATLAB 2017a. As can be seen in Fig. 6, the proposed filter and the UKF are very fast when performing the prediction step and the particle filter is significantly slower due to the large number of points that need to be propagated. When performing the update step, the proposed filter is somewhat slower than the UKF and the particle filter because the progressive update method needs several steps (see Fig. 7). This is especially obvious in the first time step, which is particularly slow because the initial estimate is poor, and thus, a large number of progression steps is required. However, the proposed approach is still easily fast enough for most real-time applications.



(a) Low noise scenario.

(b) High noise scenario.

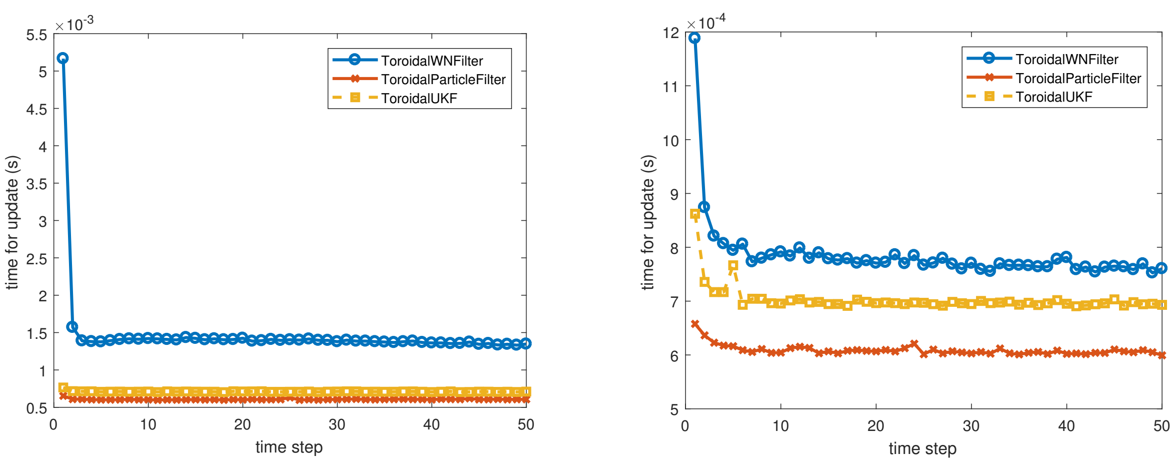
Fig. 5: Evaluation results for the proposed filter, an SIR particle filter, and a toroidal version of the UKF.



(a) Low noise scenario.

(b) High noise scenario.

Fig. 6: Runtime for prediction.



(a) Low noise scenario.

(b) High noise scenario.

Fig. 7: Runtime for measurement update.

VI. Conclusion

We have presented a novel nonlinear toroidal filtering algorithm. It makes use of a bivariate wrapped normal assumption and is based on the deterministic sampling scheme proposed in [19] as well as the parameter estimation scheme from [20]. Unlike our previous toroidal estimation algorithm [1], it can handle nonlinear system and measurement functions. The proposed approach was evaluated in simulations and its advantage compared with state-of-the-art approaches has been illustrated.

Future work may consist in generalization of the proposed method to the hypertorus in n dimensions. This generalization is nontrivial because the correlations cannot be quantified by a single value anymore, but rather, we have to take the pairwise correlation of any two dimensions into account, similar to the way the off-diagonal entries of a covariance matrix quantify the correlation in $\mathbb{R}^n$.

MATLAB implementations of the proposed algorithms are available in libDirectional, a library for directional statistics and estimation [33]. The algorithms used for comparison are also included in the library.

References

[1] 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.

[2] K. V. Mardia, "Directional Statistics in Geosciences," *Communications in Statistics - Theory and Methods*, vol. 10, no. 15, pp. 1523–1543, 1981.

[3] E. Batschelet, *Circular Statistics in Biology*, ser. Mathematics in Biology. London: Academic Press, 1981.

[4] J. Traa and P. Smaragdis, "A Wrapped Kalman Filter for Azimuthal Speaker Tracking," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1257–1260, 2013.

[5] J. Glover and L. P. Kaelbling, "Tracking the Spin on a Ping Pong Ball with the Quaternion Bingham Filter," in *Proceedings of the 2014 IEEE Conference on Robotics and Automation (ICRA 2014)*, Hong Kong, China, 2014.

[6] G. Kurz, M. Dolgov, and U. D. Hanebeck, "Nonlinear Stochastic Model Predictive Control in the Circular Domain," in *Proceedings of the 2015 American Control Conference (ACC 2015)*, Chicago, Illinois, USA, Jul. 2015.

[7] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A Recursive Fusion Filter for Angular Data," in *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)*, 2009, pp. 882–887.

[8] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun, "A Multi-temporal Multi-sensor Circular Fusion Filter," *Information Fusion*, vol. 18, pp. 86–100, Jul. 2013.

[9] 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, pp. 70–87, Mar. 2016.

[10] F. Pfaff, G. Kurz, and U. D. Hanebeck, "Multimodal Circular Filtering Using Fourier Series," in *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington D. C., USA, Jul. 2015.

[11] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Discrete Recursive Bayesian Filtering on Intervals and the Unit Circle," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sep. 2016.

[12] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "Non-Identity Measurement Models for Orientation Estimation Based on Directional Statistics," in *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington D. C., USA, Jul. 2015.

[13] F. Pfaff, G. Kurz, and U. D. Hanebeck, "Multivariate Angular Filtering Using Fourier Series," *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 206–226, Dec. 2016.

[14] 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.

[15] T. Lefebvre, H. Bruyninckx, and J. D. Schuller, "Comment on "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators" [with Authors' Reply]," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, Aug. 2002.

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

[17] B. Jia, M. Xin, and Y. Cheng, "High-degree Cubature Kalman Filter," *Automatica*, vol. 49, no. 2, pp. 510–518, 2013.

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

[19] G. Kurz and U. D. Hanebeck, "Deterministic Sampling on the Torus for Bivariate Circular Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 530–534, Feb. 2017.

[20] ——, "Parameter Estimation for the Bivariate Wrapped Normal Distribution," in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC 2015)*, Osaka, Japan, Dec. 2015.

[21] 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.

[22] W. Schmidt, "Statistische Methoden beim Gefügestudium krystalliner Schiefer," *Sitzungsberichte Akademie der Wissenschaften in Wien*, vol. 126, pp. 515–539, Jul. 1917.

[23] K. V. Mardia, C. C. Taylor, and G. K. Subramaniam, "Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data," *Biometrics*, vol. 63, no. 2, pp. 505–512, 2007.

[24] H. Singh, V. Hnizdo, and E. Demchuk, "Probabilistic Model for Two Dependent Circular Variables," *Biometrika*, vol. 89, no. 3, pp. 719–723, 2002.

[25] G. Kurz and U. D. Hanebeck, "Toroidal Information Fusion Based on the Bivariate von Mises Distribution," in *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015)*, San Diego, California, USA, Sep. 2015.

[26] 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.

[27] N. I. Fisher and A. J. Lee, "A Correlation Coefficient for Circular Data," *Biometrika*, vol. 70, no. 2, pp. 327–332, 1983.

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

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

[30] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[31] ——, "Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions," in *Proceedings of the 2014 American Control Conference (ACC 2014)*, Portland, Oregon, USA, Jun. 2014.

[32] J. Steinbring and U. D. Hanebeck, "Progressive Gaussian Filtering Using Explicit Likelihoods," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, Jul. 2014.

[33] G. Kurz, I. Gilitschenski, F. Pfaff, and L. Drude, "libDirectional," 2015. [Online]. Available: https://github.com/libDirectional