DOI. No. 10.1109/MAES.2016.150083

Tutorial

Recursive Bayesian Filtering in Circular State Spaces

Gerhard Kurz, **Karlsruhe Institute of Technology, Karlsruhe, Germany**
Igor Gilitschenski, **Swiss Federal Institute of Technology, Zurich, Switzerland**
Uwe D. Hanebeck, **Karlsruhe Institute of Technology, Karlsruhe, Germany**

To facilitate recursive state estimation in the circular domain based on circular statistics, we introduce a general framework for estimation of a circular state based on different circular distributions. Specifically, we consider the wrapped normal (WN) distribution and the von Mises distribution. We propose an estimation method for circular systems with nonlinear system and measurement functions. This is achieved by relying on efficient deterministic sampling techniques. Furthermore, we show how the calculations can be simplified in a variety of important special cases, such as systems with additive noise, as well as identity system or measurement functions, which are illustrated using an example from aeronautics. We introduce several novel key components, particularly a distribution-free prediction algorithm, a new and superior formula for the multiplication of WN densities, and the ability to deal with nonadditive system noise. All proposed methods are thoroughly evaluated and compared with several state-of-the-art approaches.

I. INTRODUCTION

Estimation of circular quantities is an omnipresent issue, be it the wind direction, the heading of a ship, the angle of a robotic revolute joint, the orientation of a turntable, or the direction a car is facing. In particular, a variety of aerospace applications include circular estimation problems, such as heading estimation of aircrafts. Circular estimation is not limited to applications involving angles, however, and can be applied to a variety of periodic phenomena. For example, phase estimation is a common issue in signal processing, and tracking objects that periodically move along a certain trajectory is also of interest. Considering circular estimation problems is not only motivated by the immediate application on the circle. Beyond that, it forms an important foundation for research into higher-dimensional estimation problems, involving periodic quantities, for example, toroidal estimation of correlated angles, estimation of orientation, and estimation of rigid body motions. To solve these higher-dimensional problems, a good understanding of the circular case is essential.

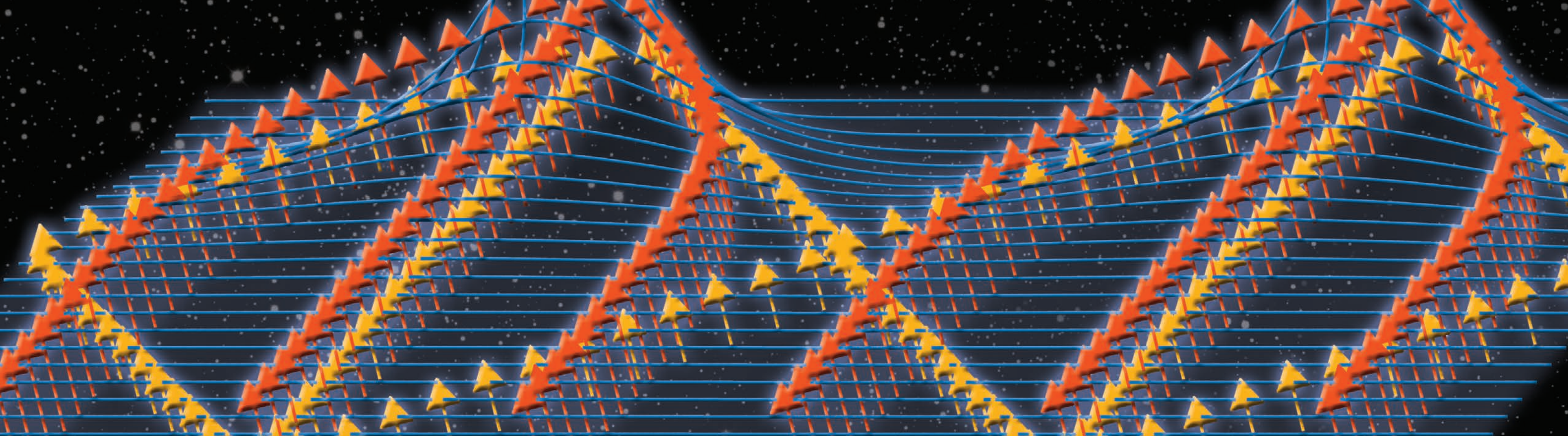
Manuscript received May 28, 2015; revised November 30, 2015; released for publication December 2, 2015.
Refereeing of this contribution was handled by W. Blair.
Authors' addresses: G. Kurz, U. Hanebeck, Intelligent Sensor-Actuator-Systems, Karlsruhe Institute of Technology, Institute for Anthropomatics and Robotics, Adenauer Ring 2, Karlsruhe, 76131 Germany, E-mail: (gerhard.kurz@kit.edu). I. Gilitschenski, Autonomous Systems Laboratory, Institute of Robotics and Intelligent Systems, Swiss Federal Institute of Technology, Zurich, Switzerland.

A. PRIOR WORK ON CIRCULAR ESTIMATION

Standard approaches to circular estimation are typically based on estimation techniques designed for linear scenarios that are tweaked to deal with some of the issues arising in the presence of circular quantities. Early work in this area dates back to the 1960s and was typically intended for aerospace applications [1]. Although much time has passed, this type of approach is still frequently used today [2, 3, 4]. However, modifying linear methods is not only tedious and error prone but also yields suboptimal results because certain assumptions of these methods are violated. For example, solutions based on Kalman filters [5], or nonlinear variants thereof [6], fundamentally neglect the true topology of the underlying manifold and assume a Gaussian distribution, which is only defined on $\mathbb{R}^n$. In the linear case, the use of a Gaussian distribution is frequently justified by the central limit theorem. This justification no longer holds in a circular setting, as the Gaussian is not a limit distribution on the circle.

To properly deal with circular estimation problems, we rely on circular statistics [7, 8], a subfield of statistics that deals with circular quantities. More broadly, the field of directional statistics [9] considers a variety of manifolds, such as the circle, the hypersphere, or the torus. Unlike standard approaches that assume linear state spaces, methods based on circular statistics correctly use the proper manifold and rely on probability distributions defined on this manifold. Circular statistics has been applied in a variety of sciences [9, Section 1.4], such as biology [8], bioinformatics [10], meteorology [11], medicine [12], space situational awareness [13], and geosciences [14].

Early results on circular filtering have been proposed by Willsky and Lo [15]. In their work, a Fourier series representation and wrapped normal (WN) mixture distributions are used, which yield nice theoretical results. However, questions of the issues arising in implementation, such as the truncation of the Fourier series to a finite number of terms and the reduction of the number of mixture components, are only briefly addressed. There has been some more recent work on filtering algorithms based on circular statistics by Azmani et al. [16], which was further investigated by Stienne et al. [17]. Their research is based on the von Mises (VM) distribution



and allows for recursive filtering of systems with a circular state space. However, it is limited to the identity with additive noise as the system equation and the measurement equation. The filter from [16] has been applied to phase estimation of global positioning system signals [18], as well as map matching [19]. A filter based on the WN distribution was developed by Traa and Smaragdis [20] for the purpose of azimuthal speaker tracking, but it is limited to the identity with additive noise for the system and measurement model as well. Chiuso and Picci [21] and Markovic et al. [22] have published similar filters based on the VM-Fisher distribution, a generalization of the VM distribution to the hypersphere.

We have previously published a recursive filter based on the WN distribution allowing for a nonlinear system equation [23]. The paper [24] extends this approach to make a nonlinear measurement update possible. Both papers rely on a deterministic sampling scheme that is based on the first trigonometric moment. This kind of sampling is reminiscent of the well-known unscented Kalman filter (UKF) [6]. We have extended this sampling scheme to the first two trigonometric moments in [25], so the proposed filters are, in a sense, circular versions of the UKF. The developed methods have been applied in the context of constrained tracking [26], bearings-only sensor scheduling [27], as well as circular model predictive control [28]. An overview of all of these filters and the considered distributions, as well as system and measurement models, is given in Table I.

B. RELATED TOPICS

There are a number of topics that are closely related to circular estimation. We will give a short overview of these topics in this section and briefly explain how they differ from circular estimation. However, we will focus on the specific problem of circular estimation for the remainder of the paper to maintain a reasonable length of the discussion.

Beyond estimation on the circle, it is possible to consider estimation on more complex manifolds, e.g., on the sphere [21], the torus [29], the group of rotations $SO(n)$ [30], the cylinder [13, 31], or the group of rigid body motions $SE(n)$ [32, 33]. Algorithms for estimation on these manifolds are often based on generalizations of circular methods. Thus, it is essential to understand estimation on the circle, as discussed in this paper, before moving on to higher-dimensional problems.

Furthermore, there is much interest in applications in which the estimated state is a linear quantity, but the measurements are of a periodic nature [34], e.g., bearings-only tracking [27]. Although there is a close relation to circular estimation and it is also possible to derive solutions based on circular statistics, this is a distinct problem and it has to be treated differently.

Another closely related problem is the propagation of random variables through trigonometric functions, i.e., given the distribution of a random variable x, we seek to find the probability den-

TABLE I

Circular Filters Based on Directional Statistics

		System		Measurement	
Publication	**Distribution**	**Model**	**Noise**	**Model**	**Noise**
Willsky and Lo [15]	WN mixture	“Linear”	Additive	“Linear”	Additive
Azmani et al. [16]	VM	Identity	Additive	Identity	Additive
Markovic et al. [22]	VM-Fisher	Identity	Additive	Identity	Additive
Traa and Smaragdis [20]	WN	Identity	Additive	Identity	Additive
Kurz et al. [23]	WN/VM	Nonlinear	Additive	Identity	Additive
Kurz et al. [24]	WN	Nonlinear	Additive	Nonlinear	Any
Kurz, Hanebeck, and Gilitschenski, 2016	WN/VM	Nonlinear	Any	Nonlinear	Any

sity function (pdf) or the moments of $\sin(x)$ or $\cos(x)$. This kind of problem is very widespread, e.g., in Rayleigh scattering [35, 36].

The problems of fitting a circle to data [37] or tracking an extended object [38] with a circular shape are also related to circular estimation, or, at least, they seem to be at first glance. However, neither state nor measurements are restricted to the circle in this case, i.e., this type of problem does not involve the circular topology and is usually not solved using directional statistics. Circular estimation may, however, be of interest when fitting or tracking more complicated shapes whose orientation has to be estimated.

A further well-known circular problem is Bertrand's problem. It considers "random" straight lines intersecting a circle and the resulting probability distribution of the length of the intersecting line segment [39]. The issue consists in the formal definition of what it means to draw "random" straight lines, as different methods for obtaining these lines affect the resulting distribution. Unlike circular estimation problems in which we are restricted to values on the circumference of the unit circle, this problem considers the area inside the circle.

C. CONTRIBUTION

This paper summarizes and combines our results, as well as extends the previous work by a number of additional contributions, and gives an introduction to the topic of circular filtering and provides an overview of our circular filters. Although many of the fundamentals can be found scattered through literature, e.g., [7, 8, 9, 40], we tried to make the paper as self-contained as possible.

The contributions of this paper can be summarized as follows: First of all, we propose a general filtering framework that can be used in conjunction with a variety of system and measurement equations, different types of noise, and both the WN and the VM distributions. Our previous publications [23, 24], as well as the work by Azmani et al. [16], can be seen as special cases of the proposed framework. Furthermore, we introduce a new multiplication formula for WN distributions that outperforms the solution proposed in [23]. We generalize the prediction step from [23] to a moment-based solution that does not need to assume any kind of distribution. Compared with [24], we add the ability to deal with nonadditive noise not only in the measurement update but also in the prediction step. Finally, we perform a thorough evaluation, where we compare the proposed techniques to several state-of-the-art approaches.

II. PROBLEM FORMULATION

In this section, we formulate the problems under consideration and summarize some standard approaches that have been used to address the issues associated with periodicity.

A. CIRCULAR FILTERING

Circular filtering considers estimation problems on the unit circle, which is commonly parameterized as the set of complex numbers with unit length, i.e., $\{x \in \mathbb{C} : |x| = 1\}$. To allow for a more convenient one-dimensional (1D) notation, we identify S^1 with the half-open interval $[0, 2\pi) \subset \mathbb{R}$, while keeping the topology of the circle. Together with the operation,

$$+ : S^1 \times S^1 \to S^1, \qquad x + y := x +_{\mathbb{R}} y \bmod 2\pi,$$

for all $x, y \in [0, 2\pi)$ with standard addition $+_{\mathbb{R}}$ on $\mathbb{R}$, the circle S^1 forms an Abelian group. Because S^1 with the topology given previously has the structure of a differentiable manifold and + is continuous with respect to that topology, $(S^1, +)$ is a Lie group. This implies that transition maps between different charts are differentiable and also serves as a justification for considering addition and subtraction on the unit circle.

We consider a system whose state x_k at time step k is a value on the unit circle S^1. System and measurement models are assumed to be given. We propose several methods to deal with different types of models. More complex models necessitate the use of more sophisticated algorithms, and conversely, simpler models allow the use of computationally less expensive algorithms.

1) *System Model*: In this work, we consider a system model whose state evolves according to the general system equation

$$x_{k+1} = a_k\left(x_k, w_k\right) \tag{1}$$

with (nonlinear) system function a_k: $S^1 \times W \to S^1$ and independent and identically distributed noise $w_k \in W$, stemming from some noise space W. Note that W is not necessarily S^1 but may be an arbitrary set, for example, the real-vector space $\mathbb{R}^n$, some manifold, or even a discrete set. Also, there is no assumption that the noise is zero mean. An interesting and practically relevant special case is a (nonlinear) system with additive noise

$$x_{k+1} = a_k\left(x_k\right) + w_k \bmod 2\pi, \tag{2}$$

with a_k: $S^1 \to S^1$ and $w_k \in S^1$. In particular, we also consider the special case, where a_k is the identity, i.e.,

$$x_{k+1} = x_k + w_k \bmod 2\pi. \tag{3}$$

2) *Measurement Model*: The system state cannot be observed directly but may only be estimated based on measurements that are disturbed by noise. A general measurement function is given by

$$\hat{z}_k = h_k\left(x_k, v_k\right), \tag{4}$$

where $\hat{z}_k \in Z$ is the measurement in the measurement space Z, h_k : $S^1 \times V \to Z$ is the measurement function, and $v_k \in V$ is arbitrary independent and identically distributed measurement noise in a certain noise space V. Note that the measurement and noise space can be arbitrary sets, in general. Interesting and special cases are measurement functions in which the measurement noise is additive, i.e.,

$$\hat{z}_k = h_k\left(x_k\right) + v_k \tag{5}$$

with measurement function $h_k : S^1 \to Z$ and $v_k \in Z$. In this case, we require Z to have a group structure with + as the operation. Additionally, we consider the more specific case in which h_k is the identity, i.e.,

$$\hat{z}_k = x_k + v_k \quad \mod 2\pi \tag{6}$$

with $\hat{z}_k, v_k \in S^1$.

Remark 1 *We do not consider linear system models because linearity is a concept of vector spaces, not manifolds [23]. For this reason, there are no linear functions on the circle.*

Example 1 Yaw angle estimation in aeronautics: *To illustrate the different models discussed in this paper, we consider the problem of estimating the yaw angle, or heading, of an airplane.*

If we had no other information, we could assume an identity system model with additive noise as in (3), i.e., a random walk model on the circle. By considering nonzero-mean noise, we could also introduce a known offset into the system model, e.g., the rotation since the last time step as obtained from a gyroscope or the expected rotation as a result of known control inputs by the pilot. A more complicated and nonlinear model with additive (2) or nonadditive (1) noise could, for example, consider the aerodynamic properties of the plane.

Models of different complexity can be used for consideration of measurements. The simplest case would be an identity measurement model (6), i.e., the yaw angle is directly observed, for example, with a magnetometer (if roll and pitch angles are assumed to be approximately zero). A more complex nonlinear model could, for example, consider position measurements of sensors, say time difference of arrival-based signal receivers, mounted at the nose and the tail of the plane. In this case, the noise could be modeled as either additive (5), i.e., added to the position measurements, or nonadditive (4), e.g., the time error in time-of-flight measurements.

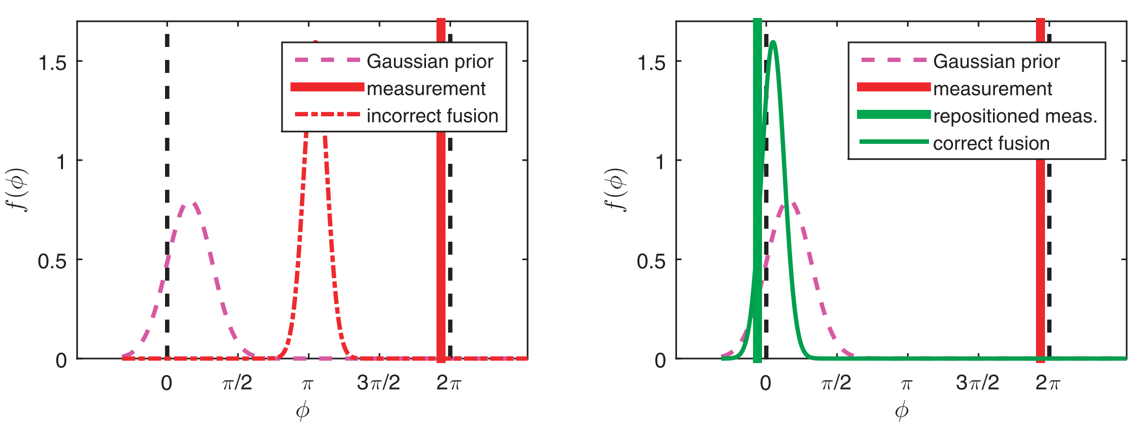
B. STANDARD APPROACHES

As circular estimation problems are widespread in a variety of applications, a number of standard approaches have been employed. We introduce some of the most common methods and explain their strengths and weaknesses.

1) *Gaussian-Based Approaches*: Gaussian-based methods (wrongly) assume a Gaussian distribution and use standard filtering techniques for Gaussians in conjunction with certain modifications to allow their application to circular problems.

a) *1D methods*: One common approach is the use of a standard Kalman filter [5] or in case of nonlinear system or measurement functions, UKF [6] with a scalar state x_k containing the angle θ_k, i.e., $x_k = \theta_k$ (that is, x_k is defined on a chart of the manifold). However, two modifications are necessary before this approach can be used in practice. First, the estimate has to be forced to stay within the interval $[0, 2\pi)$ by performing a modulo operation after every prediction, update step, or both. Second, if the measurement space is periodic, the measurement needs to be repositioned to be closer to the prediction step in certain cases. This problem occurs whenever the measurement and the current prediction are more than π apart. In this case, the measurement needs to be moved by $\pm 2\pi$ to an equivalent measurement that deviates at most π from the prediction. An illustration of this issue is given in Fig. 1. When the uncertainty is small, this kind of approach works fairly well, but it tends to produce unsatisfactory results if the uncertainty is high [23].

b) *2D methods*: Another common approach is based on the Kalman filter or the UKF with two dimensional (2D) state subject to a nonlinear constraint. More specifically, an angle θ_k is represented by the state vector $\underline{x}_k = \left[\cos(\theta_k), \sin(\theta)_k\right]^T$, and the constraint is $\| \underline{x}_k \| = 1$ to enforce that $\underline{x}_k$ is on the unit circle (i.e., $\underline{x}_k$ is defined on the space where the manifold is embedded). To enforce this constraint, $\underline{x}_k$ is projected to the unit circle after each



(a) Incorrect fusion.

(b) Correct fusion.

Fig. 1.
This figure illustrates that repositioning of the measurement is necessary to obtain satisfactory performance when using filters for linear spaces on circular problems.

prediction, update step, or both. More sophisticated approaches increase the covariance to reflect the fact that the projection operation constitutes an increase in uncertainty [41]. One of the issues of this approach is that the system and measurement models sometimes become more complicated when the angle θ_k is transformed to a 2D vector [26].

2) *Particle Filters*: Another method that can be applied is particle filtering [42]. Particle filters on nonlinear manifolds are fairly straightforward to implement because each particle can be treated separately. For the particle filter to work, the system function and the measurement likelihood both need to respect the underlying topology. The reweighting step, as well as the commonly used sequential importance resampling are independent of the underlying manifold and can be used in a circular setting as well. However, issues that are typically associated with particle filters arise. If the measurement likelihood function is very narrow, particle degeneration can occur, i.e., (almost) all particles have zero weight after the reweighting step. Furthermore, many particles are required to obtain stable results. Even though these problems are less critical in a 1D setting, there can still be issues if measurements with high certainty occur in areas with few particles. This can happen when, for example, information from sensors with very different degrees of accuracy is fused. Furthermore, note that sampling from certain circular distributions can be somewhat involved.

III. CIRCULAR STATISTICS

Because of the drawbacks of the approaches discussed previously, we propose a filtering scheme based on circular statistics [7]. In the following, we introduce the required fundamentals from the field of circular statistics.

A. CIRCULAR DISTRIBUTIONS

A variety of circular distributions has been proposed in literature [8]. We give definitions of all distributions that are required for the proposed filtering scheme.

Definition 1 WN distribution: *The WN distribution is given by the pdf*

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{k=-\infty}^{\infty}\exp\left(-\frac{(x-\mu+2\pi k)^2}{2\sigma^2}\right)$$

with $x \in S^1$*, location parameter* $\mu \in S^1$*, and concentration parameter* $\sigma > 0$.

The WN distribution is obtained by wrapping a 1D Gaussian distribution around the unit circle and adding up all probability mass that is wrapped to the same point. It appears as a limit distribution on the circle [23] in the following sense. A summation scheme of random variables that converges to the Gaussian distribution in the linear case will converge to the WN distribution if taken modulo 2π. Even though there is an infinite sum involved, evaluation of the pdf of a WN distribution can be performed efficiently, because only few summands need to be considered [43].

Definition 2 VM distribution: *The VM distribution is given by the pdf*

$$f(x;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)}\exp\big(\kappa\cos(x-\mu)\big)$$

with $x \in S^1$*, location parameter* $\mu \in S^1$*, and concentration parameter* $\kappa > 0$. $I_0(\cdot)$ *is the modified Bessel function of order* 0 *[44].*

The VM distribution has been used as a foundation for a circular filter by Azmani et al. [16].

Definition 3 Wrapped Dirac (WD) mixture distribution: *The WD distribution with L components is given by*

$$f\left(x;\gamma_1,\ldots,\gamma_L,\beta_1,\ldots,\beta_L\right) = \sum\nolimits_{j=1}^{L}\gamma_j\delta\left(x-\beta_j\right)$$

with Dirac delta distribution $\delta(\cdot)$*, Dirac positions* $\beta_1, \ldots, \beta_L \in S^1$*, and weights* $\gamma_1, \ldots, \gamma_L > 0$*, where* $\sum\nolimits_{j=1}^{L}\gamma_j = 1$.

Unlike the WN and VM distributions, the WD distribution is a discrete probability distribution on a continuous domain, obtained by wrapping a Dirac mixture in $\mathbb{R}$ around the unit circle. WD distributions can be used as discrete approximations of continuous distributions with a finite set of samples.

In this paper, we use the following notation. We denote the density function of a WN distribution with parameters μ and σ with $\mathcal{WN}(\mu,\sigma)$. If a random variable x is distributed according to this WN distribution, we write $x \sim \mathcal{WN}(\mu,\sigma)$. The terms $\mathcal{VM}(\mu,\kappa)$ and $\mathcal{WD}\left(\gamma_1,\ldots,\gamma_L,\beta_1,\ldots,\beta_L\right)$ are used analogously.

B. TRIGONOMETRIC MOMENTS

In circular statistics, there is a concept called trigonometric (or circular) moment.

Definition 4 Trigonometric moments: *For a random variable* $x \sim f(x)$ *defined on the circle, its* n*th trigonometric moment is given by*

$$m_n = \mathbb{E}\big(\exp(inx)\big) = \int_0^{2\pi}\exp(inx)\cdot f(x)\mathrm{d}x \in \mathbb{C}$$

with imaginary unit i.

The *n*th trigonometric moment is a complex number and, hence, has two degrees of freedom, the real and the imaginary part. For this reason, the first moment includes information about the circular mean $\arg m_1 = \operatorname{atan2}(\operatorname{Im} m_1, \operatorname{Re} m_1)$ as well as the concentration $|m_1| = \sqrt{\left(\operatorname{Re} m_1\right)^2 + \left(\operatorname{Im} m_1\right)^2}$ of the distribution, similar to the first two linear moments. The WN and VM distributions are uniquely defined by their first trigonometric moment [7]. For a Fourier series representation of the density function, the Fourier coefficients are closely related to the trigonometric moments according to $c_k = \frac{1}{2\pi}m_{-k}$; see [45, 46].

Lemma 1 Moments for WN, VM, and WD distributions: For WN, VM, and WD distributions with given parameters, the n*th* trigonometric moment can be calculated according to

$$m_n^{WN} = \exp\left(in\mu - n^2\sigma^2/2\right),$$

$$m_n^{VM} = \exp(in\mu)\frac{I_{|n|}(\kappa)}{I_0(\kappa)},$$

$$m_n^{WD} = \sum\nolimits_{j=1}^{L}\gamma_j\exp\left(in\beta_j\right).$$

A proof is given in [9].

C. TRIGONOMETRIC MOMENT MATCHING

As both WN and VM distributions are uniquely defined by their first trigonometric moment, it is possible to convert between them by matching the first trigonometric moment.

LEMMA 2 Trigonometric moment matching: *We define* $A(x) = \frac{I_1(x)}{I_0(x)}$ *as given in [9].*

1) *For a given first moment* m_1*, the WN distribution with this first moment has the density* $\mathcal{WN}\left(\operatorname{atan2}(\operatorname{Im} m_1, \operatorname{Re} m_1), \sqrt{-2\log(|m_1|)}\right)$.

2) *For a given first moment* m_1*, the VM distribution with this first moment has the density* $\mathcal{VM}\left(\operatorname{atan2}(\operatorname{Im} m_1, \operatorname{Re} m_1), A^{-1}(|m_1|)\right)$.

3) *For a given VM distribution with density* $\mathcal{VM}(\mu,\kappa)$*, the WN distribution with identical first moment has the density* $\mathcal{WN}\left(\mu, \sqrt{-2\log(I_1(\kappa)/I_0(\kappa))}\right)$.

4) *For a given WN distribution with density* $\mathcal{WN}(\mu,\sigma)$*, the VM distribution with identical first moment has the density* $\mathcal{VM}\left(\mu, A^{-1}\left(\exp\left(-\sigma^2/2\right)\right)\right)$.

The proof is given in [23]. Calculation of the function $A^{-1}(\cdot)$ is somewhat involved. In [23], we use the algorithm by Amos [47][1] to calculate $A(\cdot)$ and MATLAB's fsolve to invert this function. A more detailed discussion of approximations of $A^{-1}(\cdot)$ can be found in [12, Appendix], [48], [49, Section 2.3].

IV. DETERMINISTIC SAMPLING

To propagate continuous probability densities through nonlinear functions, it is a common technique to use discrete sample-based approximations of the continuous densities. A set of samples can be easily propagated by applying the nonlinear function to each sample individually. This approach can be used for both the prediction and the measurement update steps.

We distinguish between deterministic and nondeterministic sampling. Nondeterministic sampling relies on a randomized algorithm to stochastically obtain samples of a density. Typical examples include the samplers used by the particle filter [42] or the Gaussian particle filter [50]. Deterministic sampling selects samples in a deterministic way, for example, to fit certain moments (the sampler used by the UKF [6]) or to optimally approximate the shape of the density; see, e.g., [51]. Deterministic sampling schemes have the advantage of requiring a significantly smaller number of samples, which is why we will focus on this type of solution.

A naïve solution for approximating a WN density may be the application of a deterministic sampling scheme for the Gaussian distribution (such as the sampler used in [6]) and, subsequently, wrapping the samples. Even though this technique is valid for stochastic samples, it does not provide satisfactory results for deterministic samples. In extreme cases, wrapping can cause different samples to be wrapped to the same point, grossly misrepresenting the original density. This problem is illustrated in Fig. 2. For $\sigma \approx 2.5$, one sample is placed at μ, and two samples are placed on the opposite side of the circle, i.e., the mode of the approximation is opposite to the true mode. Furthermore, for $\sigma \approx 5$, all three UKF samples are wrapped to the same position, i.e., the sample-based approximation degenerates to a distribution with a single Dirac component even though the true distribution is nearly uniform.

[1] Pseudocode of this algorithm is given in [23].

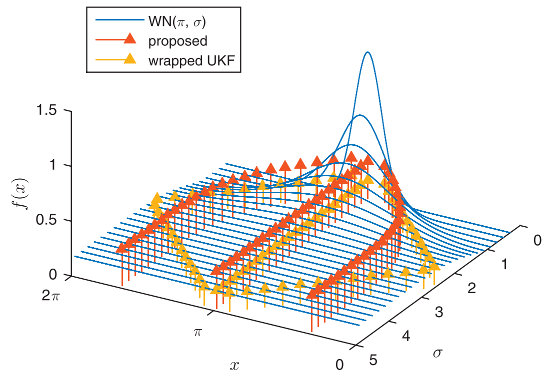


Fig. 2.
Proposed approaches for generating samples of WN distributions with different concentration parameters σ compared with the naïve approach of wrapping samples of a Gaussian with identical σ. As σ gets large, the UKF samples are eventually wrapped to the same location, which produces an extremely poor approximation.

A. ANALYTIC SOLUTIONS

First of all, we consider analytic solutions to obtain deterministic samples. These solutions are based on trigonometric moment matching and only provide a small, fixed number of Dirac components but are extremely fast to calculate, making them a good choice for real-time applications.

In [23], we presented a method to obtain a WD approximation with three equally weighted components, which is based on matching the first trigonometric moment (see Algorithm 1). We further extended this scheme to obtain a WD with five components by matching the first, as well as the second trigonometric moment (see Algorithm 2), which, as we proved in [25], necessitates the use of different weights.

B. OPTIMIZATION-BASED SOLUTIONS

If a larger number of samples is desired and there are more degrees of freedom in the samples than constraints (such as preservation of trigonometric moments), optimization-based solutions can be used. The number of samples can be adjusted by the user, and an optimal approximation is derived by minimizing a distance measure.

To simultaneously calculate optimal locations and weights for the samples, a systematic approach based on VM kernels has been proposed in [52]. For a WD mixture, an induced VM mixture is compared with the true distribution with a quadratic integral distance. A specific kernel width is considered for each component, which depends on the weight of the component and the value of the true distribution at the location of the component. Both the weights and the locations of a fixed even number of WD components are optimized to obtain an optimal symmetric approximation. Constraints in the optimization algorithms are used to maintain a predefined

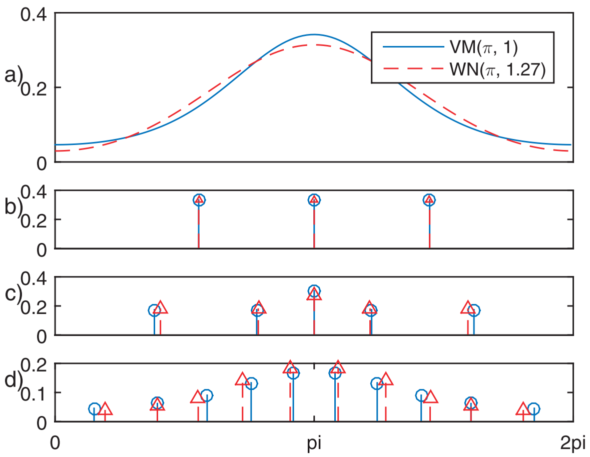


Fig. 3.
Example of the deterministic sampling of a VM and a WN distribution with equal first trigonometric moment. From top to bottom: (a) original densities, (b) result of Algorithm 1, (c) result of Algorithm 2, (d) approach based on VM kernels from [52] for 10 components. Note that the result of Algorithm 1 is identical for both densities because only the first trigonometric moment is matched.

number of trigonometric moments. This approach results in well-distributed Dirac mixtures that fulfill the moment constraints. Examples from all discussed methods for deterministic sampling are depicted in Fig. 3.

V. OPERATIONS ON DENSITIES

To derive a circular filtering algorithm, we need to be able to perform certain operations on the involved probability densities.

A. SHIFTING AND MIRRORING

For a given density $f(x)$, we want to obtain the density $f(c - x)$ for a constant $c \in S^1$. This operation is necessary in certain cases of the update step. We can split this operation into two steps: mirroring to obtain $f(-x)$, and subsequent shifting by c to obtain $f(c + (-x))$. Mirroring $\mathcal{WN}(\mu,\sigma)$ and $\mathcal{VM}(\mu,\kappa)$ yields $\mathcal{WN}(2\pi-\mu,\sigma)$ and $\mathcal{VM}(2\pi-\mu,\kappa)$ because the distributions are symmetric around their mean. Shifting $\mathcal{WN}(\mu,\sigma)$ and $\mathcal{VM}(\mu,\kappa)$ by c yields
$\mathcal{WN}(\mu - c \bmod 2\pi, \sigma)$
and
$\mathcal{VM}(\mu - c \bmod 2\pi, \kappa)$,
so the combined operation results in
$\mathcal{WN}\left((2\pi-\mu) - c \bmod 2\pi, \sigma\right)$
and
$\mathcal{VM}\left((2\pi-\mu) - c \bmod 2\pi, \kappa\right)$.

B. CIRCULAR CONVOLUTION

Given two independent circular random variables $x_1 \sim f_1(x_1)$, $x_2 \sim f_2(x_2)$, the sum $x_1 + x_2$ is distributed according to

$$\left(f_1 * f_2\right)(x) = \int_0^{2\pi} f_1(t) f_2(x-t)\,\mathrm{d}t,$$

where * denotes the convolution. This operation is necessary in the prediction step to incorporate additive noise.

WN distributions are closed under convolution, and the new pdf can be obtained just as in the Gaussian case [23], i.e., $\mu = \mu_1 + \mu_2 \mod 2\pi$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$. VM distributions are not closed under convolution. For this reason, Azmani et al. [16] used the approximation from [9], which is given by $\mu = \mu_1 + \mu_2$, $\kappa = A^{-1}(A(\kappa_1) \cdot A(\kappa_2))$. The function $A(\cdot)$ is the same as defined in Lemma 2. This approximation can be derived from an intermediate WN representation [17]. A similar approximation has been used by Markovic et al. for the VM-Fisher case [22, Eq. (7)].

In this paper, we present a more general result that calculates the convolution based on trigonometric moments.

LEMMA 3 Moments after addition of random variables: *Assume independent random variables $x_1 \sim f_1$, $x_2 \sim f_2$ defined on the circle. For the sum $x = x_1 + x_2$, it holds*
$\mathbb{E}\left(\exp(inx)\right) = \mathbb{E}\left(\exp\left(inx_1\right)\right)\,\mathbb{E}\left(\exp\left(inx_2\right)\right)$.

PROOF

$$\begin{aligned} m_n &= \mathbb{E}\left(\exp(inx)\right) = \int_0^{2\pi} \exp(inx) f(x)\,\mathrm{d}x \\ &= \int_0^{2\pi} \exp(in(x)) \int_0^{2\pi} f_1(y) f_2(x-y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^{2\pi}\int_0^{2\pi} \exp\left(in\left(x_1+x_2\right)\right) f_1\left(x_1\right) f_2\left(x_2\right)\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_0^{2\pi} \exp\left(inx_1\right) f_1\left(x_1\right)\mathrm{d}x_1 \int_0^{2\pi} \exp\left(inx_2\right) f_2\left(x_2\right)\mathrm{d}x_2 \\ &= \mathbb{E}\left(\exp\left(inx_1\right)\right)\,\mathbb{E}\left(\exp\left(inx_2\right)\right). \end{aligned}$$

If matching of the first trigonometric moment is used to fit a WN or VM to the density that results from the convolution, the solutions for WN and VM distributions from [16] and [23] arise as special cases of Lemma 3.

REMARK 2 *In fact, Lemma 3 allows us to calculate the convolution purely moment based. For this reason, we do not need to assume any particular distribution.*[2]

C. MULTIPLICATION

Multiplication of pdfs is an important operation for filtering algorithms, because it is required for Bayesian inference. In general, the product of two pdfs is not normalized and thus, not a pdf. For this reason, we consider the renormalized product, which is a valid pdf.

1) *VM*: VM densities are closed under multiplication [16]. It holds that $\mathcal{VM}\left(\mu_1,\kappa_1\right) \cdot \mathcal{VM}\left(\mu_2,\kappa_2\right) \propto \mathcal{VM}\left(\mu,\kappa\right)$, where $\mu = \operatorname{atan2}\left(\operatorname{Im} m_1, \operatorname{Re} m_1\right)$, $\kappa = |m_1|$ with $m_1 = \kappa_1 \exp(i\mu_1) + \kappa_2 \exp(i\mu_2)$.

2) *WN:* WN densities are not closed under multiplication. In the following, we consider two different methods to approximate the density of the product with a WN density.

a) *WN via VM*: In [23], we proposed a method to use the VM distribution to approximate the product of two WN densities. More specifically, we convert the WN densities to VM densities using Lemma 2, multiply according to the VM multiplication formula, and convert back to a WN distribution by applying Lemma 2 again. This method has the disadvantage that in general, the first trigonometric

[2] In general, the complexity of the density increases with each successive convolution, so considering a finite number of moments is an approximation.

Algorithm 1: Deterministic approximation with $L = 3$ components.

Input: first trigonometric moment m_1

Output: $\mathcal{WD}(\gamma_1, \gamma_2, \gamma_3, \beta_1, \beta_2, \beta_3)$

/* extract μ */

$\mu \leftarrow \text{atan2}(\operatorname{Im} m_1, \operatorname{Re} m_1)$;

/* obtain Dirac positions */

$\alpha \leftarrow \arccos\left(\frac{3}{2}|m_1| - \frac{1}{2}\right)$;

$\beta_1 \leftarrow \mu - \alpha \mod 2\pi$;

$\beta_2 \leftarrow \mu \mod 2\pi$;

$\beta_3 \leftarrow \mu + \alpha \mod 2\pi$;

/* obtain weights */

$\gamma_1, \gamma_2, \gamma_3 \leftarrow \frac{1}{3}$;

Algorithm 2: Deterministic approximation with $L = 5$ components.

Input: first and second trigonometric moment m_1 and m_2,

parameter $\lambda \in [0, 1]$ with default $\lambda = 0.5$

Output: $\mathcal{WD}(\gamma_1, \dots, \gamma_5, \beta_1, \dots, \beta_5)$

/* extract μ */

$\mu \leftarrow \text{atan2}(\operatorname{Im} m_1, \operatorname{Re} m_1)$;

$m_1 \leftarrow |m_1|, \;\; m_2 \leftarrow |m_2|$;

/* obtain weights */

$\gamma_5^{\text{min}} \leftarrow (4m_1^2 - 4m_1 - m_2 + 1)/(4m_1 - m_2 - 3)$;

$\gamma_5^{\text{max}} \leftarrow (2m_1^2 - m_2 - 1)/(4m_1 - m_2 - 3)$;

$\gamma_5 \leftarrow \gamma_5^{\text{min}} + \lambda(\gamma_5^{\text{max}} - \gamma_5^{\text{min}})$;

$\gamma_1, \gamma_2, \gamma_3, \gamma_4 \leftarrow (1 - \gamma_5)/4$;

/* obtain Dirac positions */

$c_1 \leftarrow \frac{2}{1-\gamma_5}(m_1 - \gamma_5)$;

$c_2 \leftarrow \frac{1}{1-\gamma_5}(m_2 - \gamma_5) + 1$;

$x_2 \leftarrow (2c_1 + \sqrt{4c_1^2 - 8(c_1^2 - c_2)})/4$;

$x_1 \leftarrow c_1 - x_2$;

$\phi_1 \leftarrow \arccos(x_1), \;\; \phi_2 \leftarrow \arccos(x_2)$;

$(\beta_1, \dots, \beta_5) \leftarrow \mu + (-\phi_1, +\phi_1, -\phi_2, +\phi_2, 0) \mod 2\pi$;

moment of the resulting WN does not match the first trigonometric moment of the true product. An example can be seen in Fig. 4.

b) *WN via moment matching*: In this paper, we present a new method for approximating the product of WN distributions. This method is based on directly approximating the true posterior moments.

THEOREM 1 *The first trigonometric moment of* $\mathcal{WN}(\mu_1,\sigma_1)\cdot\mathcal{WN}(\mu_2,\sigma_2)$ *after renormalization is given by*

$$m_1=\frac{\sum\limits_{j,k=-\infty}^{\infty} w(j,k)\int\limits_0^{2\pi}\exp(ix)\mathcal{N}(x;\mu(j,k),\sigma(j,k))dx}{\sum\limits_{j,k=-\infty}^{\infty} w(j,k)\int\limits_0^{2\pi}\mathcal{N}(x;\mu(j,k),\sigma(j,k))dx},$$

where $\mathcal{N}(x;\mu,\sigma)$ *is a 1D Gaussian density with mean* μ *and standard deviation* σ *and*

$$\mu(j,k)=\frac{(\mu_1+2\pi j)\sigma_2^2+(\mu_2+2\pi k)\sigma_1^2}{\sigma_1^2+\sigma_2^2}, \tag{7}$$

$$\sigma(j,k)=\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}, \tag{8}$$

$$w(j,k)=\frac{\exp\left(-\frac{1}{2}\frac{\left((\mu_1+2\pi j)-(\mu_2+2\pi k)\right)^2}{\sigma_1^2+\sigma_2^2}\right)}{\sqrt{2\pi\left(\sigma_1^2+\sigma_2^2\right)}}. \tag{9}$$

We give a proof of this theorem in the Appendix. The involved integrals can be reduced to evaluations of the complex error function erf [44, 7.1]. This yields

$$\begin{aligned}&\int_0^{2\pi}\exp(ix)\cdot\mathcal{N}(x;\mu(j,k),\sigma(j,k))\mathrm{d}x\\&=\frac{1}{2}\exp\left(i\mu(j,k)-\frac{\sigma(j,k)^2}{2}\right)\cdot\left(\operatorname{erf}\left(\frac{\mu(j,k)+i\sigma(j,k)^2}{\sqrt{2}\sigma(j,k)}\right)\right.\\&\qquad\left.-\operatorname{erf}\left(\frac{\mu(j,k)-2\pi+i\sigma(j,k)^2}{\sqrt{2}\sigma(j,k)}\right)\right)\end{aligned}$$

and

$$\begin{aligned}&\int_0^{2\pi}\mathcal{N}(x;\mu(j,k),\sigma(j,k))\mathrm{d}x\\&=\frac{1}{2}\left(\operatorname{erf}\left(\frac{\mu(j,k)}{\sigma(j,k)\sqrt{2}}\right)-\operatorname{erf}\left(\frac{\mu(j,k)-2\pi}{\sigma(j,k)\sqrt{2}}\right)\right).\end{aligned}$$

There are efficient implementations of the complex error function by means of the related Faddeeva function [53]. Furthermore, the infinite sums can be truncated to just a few summands without a significant loss in accuracy. For example, the multiplication in Fig. 4 requires $5\times5=25$ summands for an error smaller than the accuracy of the IEEE 754 64-bit double-precision data type [54]. Consequently, the proposed method allows for efficient calculation of the approximate multiplication of WN densities.

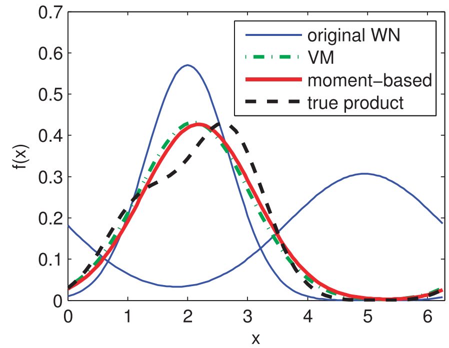


Fig. 4.
Multiplication of two WN densities with parameters $\mu_1=2$, $\sigma_1=0.7$ and $\mu_2=4.95$, $\sigma_2=1.3$. The true product and the results of both proposed approximation methods (VM and moment based) are depicted. Note that the true product is not a WN density.

VI. CIRCULAR FILTERING

Based on the results of the previous section, we derive recursive circular filtering algorithms for the scenarios described in Section II.A. We formulate the necessary steps without requiring a particular density whenever possible such that most of the methods can be directly applied to WN, as well as VM distributions, and might even be generalized to other continuous circular distributions. An overview of all considered prediction and measurement update algorithms is given in Table II.

A. PREDICTION

The prediction step is used to propagate the estimate through time.

1) *Identity System Model*: The transition density is given according to

$$f\left(x_{k+1}\mid x_k\right)=\int_0^{2\pi} f\left(x_{k+1},w_k\mid x_k\right)\,\mathrm{d}w_k=f^w\left(x_{k+1}-x_k\right),$$

where $f^w(\cdot)$ is the density of the system noise. For the predicted density, we consider the Chapman-Kolmogorov equation

$$f^p\left(x_{k+1}\right)=\int_0^{2\pi} f\left(x_{k+1}\mid x_k\right)f^e\left(x_k\right)\mathrm{d}x_k. \tag{10}$$

In the special case of an identity system model, this yields

$$f^p\left(x_{k+1}\right)=\int_0^{2\pi} f^w\left(x_{k+1}-x_k\right)f^e\left(x_k\right)\mathrm{d}x_k=\left(f^w * f^e\right)\left(x_{k+1}\right),$$

where * denotes the convolution as defined in Section V.B. For a VM distribution, this system model has been considered in [16]. If a WN distribution is assumed, (10) is a special case of [23] where we omit the propagation through the nonlinear function.

2) *Nonlinear System Model with Additive Noise*: Similar to the previous case, the transition density is given by

$$f\left(x_{k+1}\mid x_k\right)=\int_0^{2\pi}\delta\left(x_{k+1}-\left(a_k\left(x_k\right)+w_k\right)\right)f^w\left(w_k\right)\mathrm{d}w_k.$$

We approximate the prior density $f^e\left(x_k\right)\approx\sum_{j=1}^{L}\gamma_j\delta\left(x_k-\beta_j\right)$ using, for example, Algorithm 1 or Algorithm 2 ($L=3$ or $L=5$, respectively). Then, the predicted density can be approximated according to

$$
\begin{aligned}
f^{p}(x_{k+1}) &\approx \int_0^{2\pi}\left(\int_0^{2\pi}\delta\left(x_{k+1}-\left(a_k(x_k)+w_k\right)\right)f^{w}\left(w_k\right)\mathrm{d}w_k\right)\\
&\quad\cdot\left(\sum_{j=1}^{L}\gamma_j\delta\left(x_k-\beta_j\right)\right)\mathrm{d}x_k\\
&=\int_0^{2\pi}f^{w}\left(w_k\right)\sum_{j=1}^{L}\gamma_j\int_0^{2\pi}\delta\left(x_{k+1}-\left(a_k\left(x_k\right)+w_k\right)\right)\\
&\quad\cdot\delta\left(x_k-\beta_j\right)\mathrm{d}x_k\mathrm{d}w_k\\
&=\int_0^{2\pi}f^{w}\left(w_k\right)\underbrace{\sum_{j=1}^{L}\gamma_j\delta\left(x_{k+1}-\left(a_k\left(\beta_j\right)+w_k\right)\right)}_{\approx\tilde{f}\left(x_{k+1}-w_k\right)}\mathrm{d}w_k\\
&=\left(f^{w}*\tilde{f}\right)\left(x_{k+1}\right),
\end{aligned}
$$

where $\tilde{f}$ is obtained from the WD $\sum_{j=1}^{L}\gamma_j\delta\left(x-\left(a_k\left(\beta_j\right)\right)\right)$ using moment matching according to Lemma 2. The convolution can be calculated as described in Section V.B.

3) *Nonlinear System Model with Arbitrary Noise*: In this paper, we extend the previous results to deal with arbitrary noise in the prediction step. For arbitrary noise, the transition density is given by

$$
f\left(x_{k+1}\mid x_k\right)=\int_0^{2\pi}\delta\left(x_{k+1}-a_k\left(x_k,w_k\right)\right)f^{w}\left(w_k\right)\mathrm{d}w_k .
$$

We approximate the prior density $f^{e}\left(x_k\right)\approx\sum_{j=1}^{L}\gamma_j\delta\left(x_k-\beta_j\right)$, as well as the noise density $f^{w}\left(w_k\right)\approx\sum_{j=1}^{L^w}\gamma_l^w\delta\left(w_k-\beta_l^w\right)$. Note that the noise is not necessarily a circular quantity, and different approximation techniques may be required. If $W=\mathbb{R}^n$, the techniques presented in [51] may be applied. Then, the predicted density can be approximated according to

$$
\begin{aligned}
f^{p}\left(x_{k+1}\right) &\approx \int_0^{2\pi}\int_W\delta\left(x_{k+1}-a_k\left(x_k,w_k\right)\right)\\
&\quad\cdot\left(\sum_{l=1}^{L^w}\gamma_l^w\delta\left(w_k-\beta_l^w\right)\right)\mathrm{d}w_k\, f^{e}\left(x_k\right)\mathrm{d}x_k\\
&=\sum_{l=1}^{L^w}\gamma_l^w\int_0^{2\pi}\delta\left(x_{k+1}-a_k\left(x_k,\beta_l^w\right)\right)f^{e}\left(x_k\right)\mathrm{d}x_k\\
&\approx\sum_{l=1}^{L^w}\gamma_l^w\int_0^{2\pi}\delta\left(x_{k+1}-a_k\left(x_k,\beta_l^w\right)\right)\cdot\left(\sum_{j=1}^{L}\gamma_j\delta\left(x_k-\beta_j\right)\right)\mathrm{d}x_k\\
&=\sum_{j=1}^{L}\sum_{l=1}^{L^w}\gamma_j\gamma_l^w\delta\left(x_{k+1}-a_k\left(\beta_j,\beta_l^w\right)\right).
\end{aligned}
$$

A continuous density can then be fitted to this result by trigonometric moment matching (see Algorithm 3). This algorithm can be executed based purely on the trigonometric moments of $f(x_k)$ and $f^w(w_k)$, if the deterministic sampling scheme only depends on these moments. In that case, we do not necessarily need to fit a distribution $f(x_{k+1})$ to the resulting trigonometric moments but can just store the estimate by retaining those trigonometric moments.

B. MEASUREMENT UPDATE

The measurement update step fuses the prediction with a measurement that was obtained according to the measurement equation.

1) *Identity Measurement Model*: In the case of the identity measurement model and additive noise, the measurement likelihood is given by $f(z_k|x_k) = f^v(z_k - x_k)$. For the posterior density, application of Bayes' theorem yields

$$
f\left(x_k\mid z_k\right)\propto f\left(z_k\mid x_k\right)f\left(x_k\right),
$$

where $f(x_k)$ is the prior density. Thus, we obtain the posterior density $f\left(x_k\mid z_k\right)\propto f^{v}\left(z_k-x_k\right)f\left(x_k\right)$ as the product of the prior density and $f^v(z_k-x_k)$, which can be obtained as described in Section V.A. The multiplication depends on the assumed probability density and can be performed using the multiplication formulas given in Section V.C. For the VM case, this is equivalent to the measurement update from [16], and for the WN case, this is equivalent to the measurement update from [23].

2) *Nonlinear Model with Additive Noise*: For a nonlinear measurement function with additive noise, the measurement likelihood is calculated according to $f\left(\underline{z}_k\mid x_k\right)=f^{v}\left(\underline{z}_k-h_k\left(x_k\right)\right)$, as given in [24]. The remainder of the measurement update step is identical to the case of arbitrary noise as described in the following.

3) *Nonlinear Model with Arbitrary Noise*: For a nonlinear update with arbitrary noise, we assume that the likelihood is given. The key idea is to approximate the prior density $f(x_k)$ with a WD mixture and reweigh the components according to the likelihood $f(z_k|x_k)$. However, this can lead to degenerate solutions, i.e., most or all weights are close to zero, if the likelihood function is narrow.

We have shown in [24] that a progressive solution as introduced in [55] can be used to avoid this issue. For this purpose, we formulate the likelihood as a product of likelihoods

$$
f\left(\underline{z}_k\mid x_k\right)=f\left(\underline{z}_k\mid x_k\right)^{\lambda_1}\cdot\ldots\cdot f\left(\underline{z}_k\mid x_k\right)^{\lambda_s},
$$

TABLE II

Prediction and Measurement Update Algorithms

	System Model			Measurement Model		
Method	Identity	Additive Noise	Arbitrary Noise	Identity	Additive Noise	Arbitrary Noise
WN	[23] (special case)	[23]	Kurz, Hanebeck, and Gilitschenski, 2016	[23]	[24]	[24]
VM	[16]	Kurz, Hanebeck, and Gilitschenski, 2016	Kurz, Hanebeck, and Gilitschenski, 2016	[16]	Kurz, Hanebeck, and Gilitschenski, 2016	Kurz, Hanebeck, and Gilitschenski, 2016
Moment based	Kurz, Hanebeck, and Gilitschenski, 2016	Kurz, Hanebeck, and Gilitschenski, 2016	Kurz, Hanebeck, and Gilitschenski, 2016	—	—	—

Algorithm 3: Prediction with arbitrary noise.

Input: prior density $f(x_k)$, system noise density $f^w(w_k)$, system function $a_k(\cdot,\cdot)$

Output: predicted density $f(x_{k+1})$

`/* sample prior density and noise density */`

$(\gamma_1,\dots,\gamma_L,\beta_1,\dots,\beta_L) \leftarrow \text{sampleDeterm}(f(x_k))$;

$(\gamma_1^w,\dots,\gamma_{L^w}^w,\beta_1^w,\dots,\beta_{L^w}^w) \leftarrow \text{sampleDeterm}(f^w(w_k))$;

`/* obtain Cartesian product and propagate */`

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

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

    $\gamma_{j+L(l-1)}^p \leftarrow \gamma_j \cdot \gamma_l^w$;

    $\beta_{j+L(l-1)}^p \leftarrow a_k(\beta_j, \beta_l^w)$;

`/* obtain posterior density */`

$f(x_{k+1}) \leftarrow \text{momentMatching}(\gamma_1^p,\dots,\gamma_{L\cdot L^w}^p,\beta_1^p,\dots,\beta_{L\cdot L^w}^p)$;

Algorithm 4: Progressive measurement update for arbitrary noise.

Input: measurement $\hat{\underline{z}}_k$, likelihood $f(\underline{z}_k|\underline{x}_k)$, predicted density $f^p(x_k)$ as WN or VM, threshold parameter R

Output: estimated density as WN or VM

$s \leftarrow 0$;

$f(x_k) \leftarrow f^p(x_k)$;

while $\sum_{n=1}^{s} \lambda_n < 1$ **do**

  $s \leftarrow s+1$;

  `/* deterministic sampling from WN or VM density (Sec. IV) */`

  $(\gamma_1,\dots,\gamma_L,\beta_1,\dots,\beta_L) \leftarrow \text{sampleDeterm}(f(x_k))$;

  `/* calculate step size */`

$$\lambda_s \leftarrow \min\left(1-\sum_{n=1}^{s-1}\lambda_n, \frac{\log\left(R\cdot\frac{\max_{j=1,\dots,L}(\gamma_j)}{\min_{j=1,\dots,L}(\gamma_j)}\right)}{\log\left(\frac{\min_{j=1,\dots,L} f(\underline{z}_k|\beta_j)}{\max_{j=1,\dots,L} f(\underline{z}_k|\beta_j)}\right)}\right);$$

  `/* execute progression step */`

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

    $\gamma_j \leftarrow \gamma_j \cdot f(\hat{\underline{z}}_k|\beta_j)^{\lambda_s}$;

  `/* use moment matching to obtain WN or VM */`

  $f \leftarrow \text{momentMatching}(\gamma_1,\dots,\gamma_L,\beta_1,\dots,\beta_L)$;

where $\lambda_1, \ldots, \lambda_s > 0$ and $\sum_{j=1}^{s} \lambda_j = 1$. This decomposition of the likelihood allows us to perform the measurement update step gradually by performing s partial update steps. Each update step is small enough to prevent degeneration, and we obtain a new sample set after each step, which ensures that the differences between the sample weights remain small. To determine $\lambda_1, \ldots, \lambda_s$ and s, we require that after reweighing the quotient between the smallest weight $\gamma_{\min}$, and the largest weight $\gamma_{\max}$ is not below a certain threshold $R \in (0,1)$ i.e., $\frac{\gamma_{\min}}{\gamma_{\max}} \geq R$. Using the conservative bounds

$$\gamma_{\min} \geq \min_{j=1,\ldots,L}\left(\gamma_j^n\right) \cdot \min_{j=1,\ldots,L}\left(f\left(\underline{z}_k \mid \beta_j^n\right)\right),$$

$$\gamma_{\max} \leq \max_{j=1,\ldots,L}\left(\gamma_j^n\right) \cdot \max_{j=1,\ldots,L}\left(f\left(\underline{z}_k \mid \beta_j^n\right)\right),$$

this leads to the condition

$$\lambda_n \leq \frac{\log\left(R \cdot \frac{\max_{j=1,\ldots,L}\left(\gamma_j^n\right)}{\min_{j=1,\ldots,L}\left(\gamma_j^n\right)}\right)}{\log\left(\frac{\min_{j=1,\ldots,L} f\left(\underline{z}_k \mid \beta_j^n\right)}{\max_{j=1,\ldots,L} f\left(\underline{z}_k \mid \beta_j^n\right)}\right)},$$

where $\mathcal{WD}\left(\gamma_1^n, \ldots, \gamma_L^n, \beta_1^n, \ldots, \beta_L^n\right)$ is the deterministic approximation at the nth progression step.[3] The progression continues until $\Sigma_n \lambda_n = 1$. This method can be applied in conjunction with WN, as well as VM distributions (see Algorithm 4).

VII. EVALUATION

A. PROPAGATION ACCURACY

To evaluate the deterministic sampling as introduced in Section IV, we investigate the accuracy when performing propagation through the nonlinear function g: $S^1 \rightarrow S^1$,

$$g(x) = x + c \times_{\mathbb{R}} \sin(x) \mod 2\pi,$$

where $c \in [0, 1)$ is a parameter controlling the strength of the nonlinearity, and $\times_{\mathbb{R}}$ refers to multiplication in the field of real numbers $\mathbb{R}$. Furthermore, we consider the density $\mathcal{WN}(0, \sigma)$ that we want to propagate through $g(\cdot)$. For this purpose, we sample $\mathcal{WN}(0, \sigma)$ deterministically using the methods described in Section IV and obtain $\mathcal{WN}\left(\gamma_1, \ldots, \gamma_L, \beta_1, \ldots, \beta_L\right)$. Then, we apply $g(\cdot)$ componentwise, which yields $\mathcal{WD}\left(\gamma_1, \ldots, \gamma_L, g\left(\beta_1\right), \ldots, g\left(\beta_L\right)\right)$. The true posterior is given by

$$f^{true}(x) = \frac{f\left(g^{-1}(x); \mu, \sigma\right)}{g'(x)}$$

and can only be calculated numerically. We evaluate the first and the second trigonometric moment m_i^{WD}, $i = 1, 2$ of the resulting WD distribution and compare it to the first and the second trigonometric moment m_i^{true}, $i = 1, 2$ of the true posterior, which is obtained by numerical integration.[4] The considered error measure is given by $\left|m_i^{WD} - m_i^{true}\right|$, $i = 1, 2$, where $|\cdot|$ is the Euclidean norm in the complex plane. Additionally, we fit a WN density to the posterior WD by trigonometric moment matching and numerically compute the Kullback-Leibler divergence

$$D_{\mathrm{KL}}\left(f^{true} \,\|\, f^{fitted}\right) = \int_0^{2\pi} f^{true}(x) \log\left(\frac{f^{true}(x)}{f^{fitted}(x)}\right) \mathrm{d}x, \tag{11}$$

between the true posterior and the fitted WN. The Kullback-Leibler divergence is an information theoretic measure to quantify the information loss when approximating f^{true} with f^{fitted}. This concept is illustrated in Fig. 5.

The results for different values of σ are depicted in Fig. 6. We compare several samplers, the analytic methods with $L = 3$ components (Algorithm 1) and $L = 5$ components (Algorithm 2, parameter $\lambda = 0.5$) from Section IV.A, as well as a sampling with 50 equidistant samples that are weighted according to the pdf. It can be seen that the analytic solution with $L = 5$ components is significantly better than the solution with $L = 3$ components. The equidistant solution is computationally more demanding but gives almost optimal results. However, the analytic solution with $L = 5$ components has comparable performance in terms of the Kullback-Leibler divergence even though the posterior moments are not calculated as accurately.

B. MOMENT-BASED WN MULTIPLICATION

In this evaluation, we compare the two methods for WN multiplication given in Sections V.C.2a and Sections V.C.2b. For two WN densities $\mathcal{WN}\left(\mu_1, \sigma_1\right)$ and $\mathcal{WN}\left(\mu_2, \sigma_2\right)$, we calculate the true product $f^{true} = \mathcal{WN}\left(\mu_1, \sigma_1\right) \cdot \mathcal{WN}\left(\mu_2, \sigma_2\right)$ and compare it to the WN approximation f^{fitted}.

To determine the approximation quality, we compute the Kullback-Leibler divergence (11). The results for different values of σ_1, μ_2, and σ_2 are depicted in Fig. 7. We keep μ_1 fixed because only the difference between μ_2 and μ_1 affects the result. Multiplication is commutative, so we consider different sets of values for σ_1 and σ_2 to avoid redundant plots.

As can be seen, the moment-based approach derived in Section V.C.2b significantly outperforms the approach from Section V.C.2a in almost all cases. In particular, the new approach is superior for small uncertainties.

C. FILTERING

To evaluate the proposed filtering algorithms, we simulated several scenarios. First of all, we distinguish between models with additive and with a more complex noise structure. In the case of additive noise, we consider the system function

$$x_{k+1} = x_k + c_1 \times_{\mathbb{R}} \sin\left(x_k\right) + c_2 + w_k, \tag{12}$$

with two parameters $c_1 = 0.1$, $c_2 = 0.15$, noise $w_k \sim \mathcal{WN}(0, 0.2)$, and $\times_{\mathbb{R}}$ is multiplication in the field of real numbers $\mathbb{R}$. Intuitively, c_1 determines the degree of nonlinearity, and c_2 is a constant angu-

[3] Compared with [24], we extend the progressive scheme to handle discrete approximations with nonequally weighted components in this paper.

[4] Numerical integration produces very accurate results in this case but is too slow for use in practical filtering applications.

lar velocity that is added at each time step. For the case of arbitrary noise, the system function is given by

$$x_{k+1} = x_k + c_1 \times_{\mathbb{R}} \sin\left(x_k + w_k\right) + c_2 \tag{13}$$

with the same c_1, c_2, and w_k as shown previously. In both cases, the nonlinear measurement function is given by

$$\hat{\underline{z}}_k = \left[\cos\left(x_k\right), \sin\left(x_k\right)\right]^T + \underline{v}_k \in \mathbb{R}^2,$$

with additive noise $\underline{v}_k \sim \mathcal{N}\left(\underline{0}, \eta \cdot \mathrm{I}_{2\times 2}\right)$, $\eta \in \{3, 0.1, 0.01\}$. An overview of all considered scenarios is given in Table III.

In the scenarios with additive system noise, we compare the proposed filter to all standard approaches described in Section II.B, i.e., a UKF with 1D state vector, a UKF with 2D state vector, and particle filters with 10 and 100 particles. To handle nonadditive noise with the UKF, typically state augmentation is used, which is not applicable in scenarios with arbitrary noise. For this reason, we only compare the proposed approach to the particle

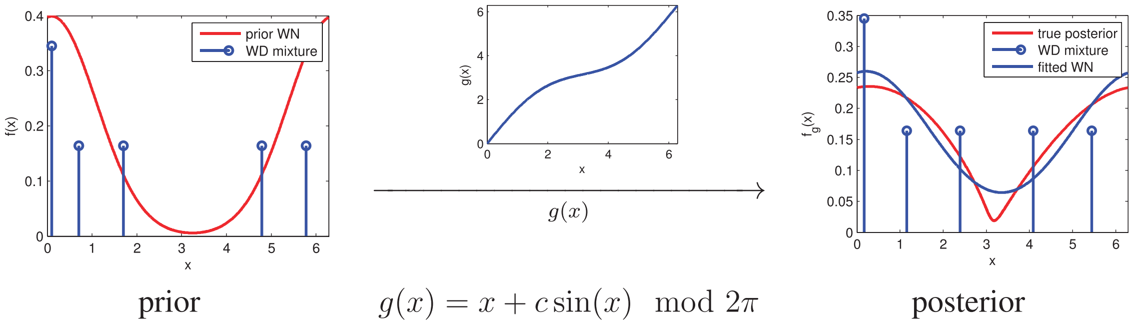


Fig. 5.
Propagation of a WN distribution with parameters $\mu = 0.1$, $\sigma = 1$ through the nonlinear function g by means of a deterministic WD approximation with five components. In this example, we use $c = 0.7$.

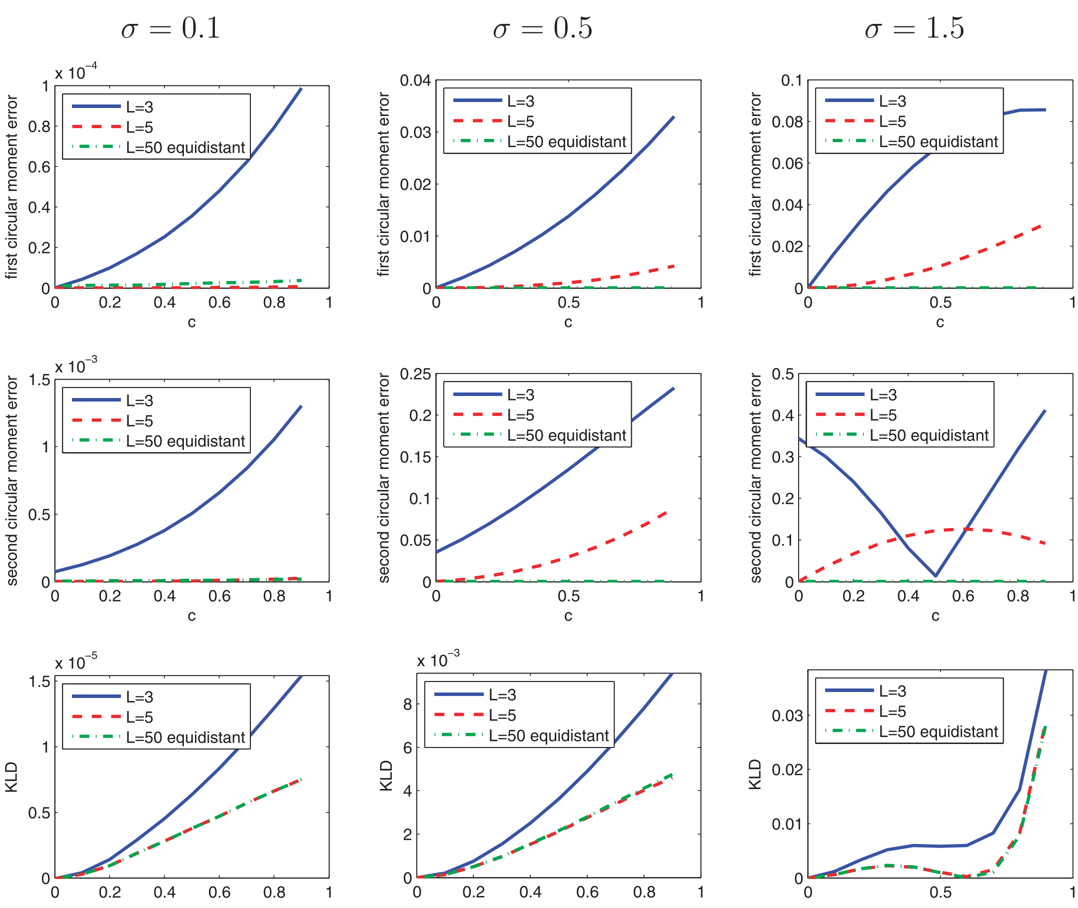


Fig. 6.
Propagation of $\mathcal{WN}(0, \sigma)$ through the function $g(\cdot)$ with nonlinearity parameter c.

filters in the nonadditive noise case. The initial estimate is given by $x_0 \sim \mathcal{WN}(0,1)$, whereas the true initial state is on the opposite side of the circle $x_0^{true} = \pi$, i.e., the initial estimate is poor, which is difficult to handle for noncircular filters. For the circular filtering algorithm, we use the deterministic sampling method given in Algorithm 2 with parameter $\lambda = 0.5$. The progression threshold is chosen as $R = 0.2$.

To evaluate the performance of different filters, we consider a specific error measure that takes periodicity into account. The angular error is defined as the shortest distance on the circle

$$d : \mathcal{S}^1 \times \mathcal{S}^1 \to [0,\pi], \qquad d(a,b) = \min(|a-b|, 2\pi - |a-b|).$$

This leads to an angular version of the commonly used root mean square error (RMSE)

$$\frac{1}{k_{\max}} \sqrt{\sum_{k=1}^{k_{\max}} d\left(x_k, x_k^{true}\right)^2}$$

between estimates x_k and true state variables x_k^{true}. We simulated the system for $k_{\max} = 100$ time steps and compared the angular RMSE of all estimators. The results from 100 Monte Carlo runs are depicted in Fig. 8.

In the scenarios with additive noise, it can be seen that the proposed filter performs very well, regardless of the amount of noise. Only the particle filter with 100 particles is able to produce similar results. However, it should be noted that the proposed filter uses just five samples. The particle filter with 10 particles performs much worse and fails completely for small noise as a result of particle degeneration issues. Both variants of the UKF perform worse than the proposed filter. Particularly the UKF with 2D state does not work very well, which can be explained by the inaccuracies in the conversion of the 1D into the 2D noise.

When nonadditive noise is considered, the proposed filter even significantly outperforms the particle filter with 100 particles. As a result of the low number of particles and the associated issues regarding particle degeneration, the particle filter with 10 particles has the worst performance.

Even though a particle filter with a 1D state and 100 particles is tractable to use in practice, the proposed approaches have several significant advantages. First of all, the methods are deterministic, i.e., the results are reproducible, and certain steps can be performed in an optimal way rather than using a stochastic approximation. Because of the smaller number of samples, they can be faster when the system or likelihood functions are expensive to evaluate. A further comparison with the particle filter has been performed in [46, Figs. 2 and 4], which shows fairly slow convergence with respect to the number of particles. Also, particle filtering approaches tend to scale very badly with the number of considered dimensions, whereas some of the proposed approaches can also be generalized to higher-dimensional manifolds.

Finally, we would like to mention that the proposed filter was also independently evaluated by Nitzan et al. [56] and shown to achieve the Bayesian Cramér-Rao lower bound. Also, it outperformed a particle filtering approach and the Fourier-based solution by Willsky and Lo [15].

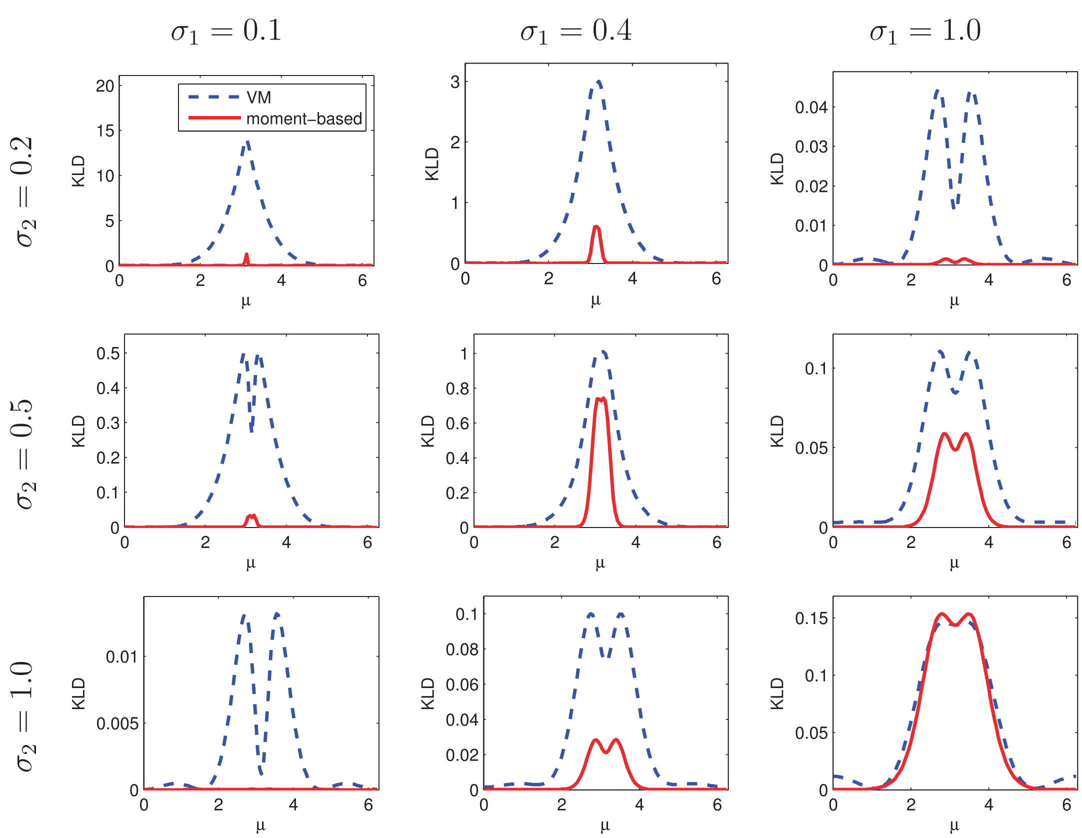


Fig. 7.
Kullback-Leibler divergence between the true product of WN densities and the proposed approximations.

TABLE III

Evaluation Scenarios

Scenario	System Function	Measurement Noise $\mathbf{C}^v$
Small	(12)	$0.01 \times \mathbf{I}_{2\times 2}$
Medium	(12)	$0.1 \times \mathbf{I}_{2\times 2}$
Large	(12)	$3 \times \mathbf{I}_{2\times 2}$
Small-nonadditive	(13)	$0.01 \times \mathbf{I}_{2\times 2}$
Medium-nonadditive	(13)	$0.1 \times \mathbf{I}_{2\times 2}$
Large-nonadditive	(13)	$3 \times \mathbf{I}_{2\times 2}$

VIII. CONCLUSION

In this paper, we presented a framework for recursive filtering on the circle. The proposed filtering algorithms can deal with arbitrary nonlinear system and measurement functions. Furthermore, they can be used in conjunction with different circular probability distributions. We have shown that the prediction step can be performed based on trigonometric moments only, without ever assuming a particular distribution. These algorithms are applicable to a wide range of problems, for example in aerospace applications, robotics, and signal processing.

For the purpose of evaluation, we have considered several aspects of the proposed methods. First of all, the accuracy of deterministic approximations was evaluated by considering the error when using them to propagate a continuous distribution through a nonlinear function. We have found that the proposed deterministic approximation with five samples yields good results for most practical scenarios. Second, we evaluated the novel moment-based WN multiplication method and showed that it is superior to the previously published method based on fitting a VM distribution. Finally, we evaluated the proposed filtering algorithms in several scenarios and compared it to state-of-the-art approaches. These simulations show the advantages of using a circular filtering scheme compared with traditional methods intended for the linear case.

Future work may include extensions of the proposed methods to other manifolds, such as the torus or the hypersphere. Additionally, consideration of multimodal circular distributions may be of interest, for example by means of WN or VM mixtures.

An implementation of the proposed algorithms is available as part of libDirectional, a MATLAB library for directional statistics and directional estimation [57]. ◆

APPENDIX. PROOF OF THEOREM I

The true renormalized product is given by $f(x) = c \cdot f(x; \mu_1, \sigma_1) \cdot f(x; \mu_2, \sigma_2)$, where c renormalizes the product, i.e.,

$$c = \left(\int_0^{2\pi} f\left(x; \mu_1, \sigma_1\right) \cdot f\left(x; \mu_2, \sigma_2\right) \, \mathrm{d}x \right)^{-1}$$

We calculate

$$\begin{aligned} m_1 &= c \cdot \int_0^{2\pi} \exp(ix) \cdot f\left(x; \mu_1, \sigma_1\right) \cdot f\left(x; \mu_2, \sigma_2\right) \mathrm{d}x \\ &= c \cdot \int_0^{2\pi} \exp(ix) \cdot \sum_{j=-\infty}^{\infty} \mathcal{N}\left(x; \mu_1 + 2\pi j, \sigma_1\right) \\ &\quad \cdot \sum_{k=-\infty}^{\infty} \mathcal{N}\left(x; \mu_2 + 2\pi k, \sigma_2\right) \mathrm{d}x \\ &= c \cdot \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_0^{2\pi} \exp(ix) \cdot \mathcal{N}\left(x; \mu_1 + 2\pi j, \sigma_1\right) \\ &\quad \cdot \mathcal{N}\left(x; \mu_2 + 2\pi k, \sigma_2\right) \mathrm{d}x \\ &= c \cdot \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_0^{2\pi} \exp(ix) \cdot w(j,k) \cdot \mathcal{N}(x; \mu(j,k), \sigma(j,k)) \mathrm{d}x \\ &= c \cdot \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w(j,k) \cdot \int_0^{2\pi} \exp(ix) \cdot \mathcal{N}(x; \mu(j,k), \sigma(j,k)) \mathrm{d}x, \end{aligned}$$

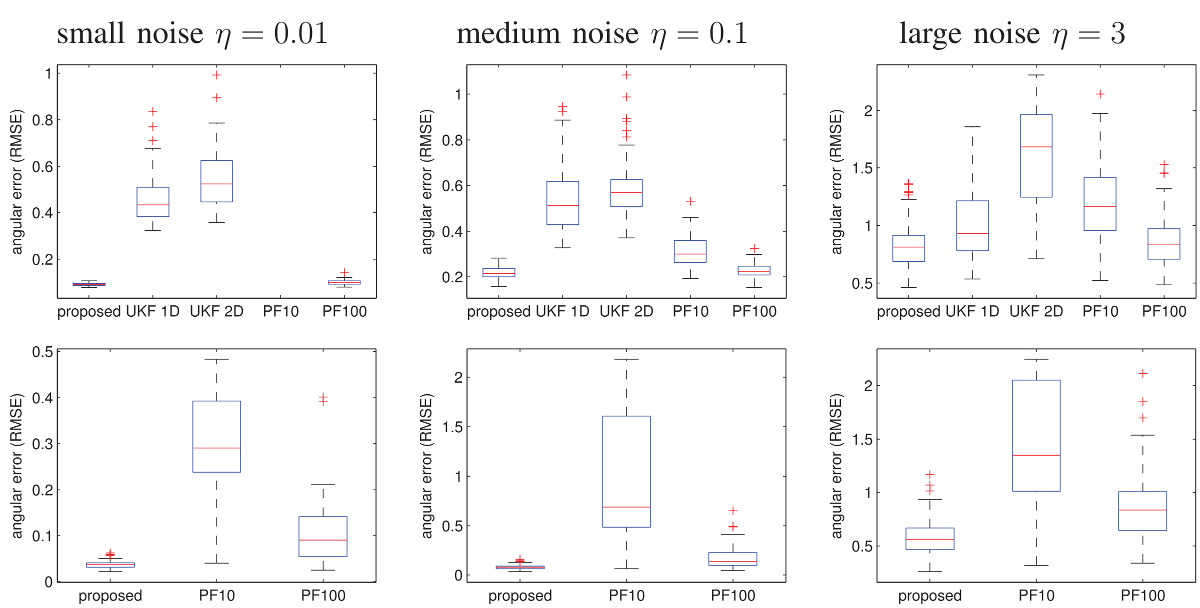


Fig. 8.
RMSE (in radians) for different filters obtained from 100 Monte Carlo runs for additive noise (top) and nonadditive noise (bottom).

where we use the dominated convergence theorem to interchange summation and integration. We use the abbreviations in (7)–(9) based on the multiplication formula for Gaussian densities given in [58,8.1.8].

To compute the renormalization factor c^{-1}, we use a similar derivation and obtain

$$c^{-1} = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w(j,k) \cdot \int_0^{2\pi} \mathcal{N}(x; \mu(j,k), \sigma(j,k)) \mathrm{d}x.$$

REFERENCES

[1] Smith, G. L., Schmidt, S. F., and McGee, L. A. *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle*. Washington, DC: National Aeronautics and Space Administration, 1962.

[2] Markley, F. L. Attitude error representations for Kalman filtering. *Journal of Guidance, Control, and Dynamics*, 26, 2 (2003), 311–317.

[3] Choukroun, D., Bar-Itzhack, I. Y., and Oshman, Y. Novel quaternion Kalman filter. *IEEE Transactions on Aerospace and Electronic Systems*, 42, 1 (2006), 174–190.

[4] Bry, A., Bachrach, A., and Roy, N. State estimation for aggressive flight in GPS-denied environments using onboard sensing. In *Proceedings of the 2012 IEEE International Conference on Robotics and Automation*, St. Paul, MN, May 2012.

[5] Kalman, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering*, 82 (1960), 35–45.

[6] Julier, S. J., and Uhlmann, J. K. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92, 3 (Mar. 2004), 401–422.

[7] Jammalamadaka, S. R., and Sengupta, A. *Topics in Circular Statistics*. Singapore: World Scientific, 2001.

[8] Batschelet, E. *Circular Statistics in Biology*. London, United Kingdom: Academic Press, 1981.

[9] Mardia, K. V., and Jupp, P. E. *Directional Statistics* (1st ed.). Chichester, United Kingdom: Wiley, 1999.

[10] Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36, 1 (2008), 99–109.

[11] Fisher, N. I. Problems with the current definitions of the standard deviation of wind direction. *Journal of Climate and Applied Meteorology*, 26, 11 (1987), 1522–1529.

[12] Kurz, G. Directional estimation for robotic beating heart surgery. Ph.D. dissertation, Karlsruhe Institute of Technology, Germany, 2015.

[13] Horwood, J. T., and Poore, A. B. Gauss von Mises distribution for improved uncertainty realism in space situational awareness. *SIAM/ASA Journal on Uncertainty Quantification*, 2, 1 (2014), 276–304.

[14] Mardia, K. V. Directional statistics in geosciences. *Communications in Statistics—Theory and Methods*, 10, 15 (1981), 1523–1543.

[15] Willsky, A. S., and Lo, J. T. Estimation for rotational processes with one degree of freedom—part II: discrete-time processes. *IEEE Transactions on Automatic Control*, 20, 1 (Feb. 1975), 22–30.

[16] Azmani, M., Reboul, S., Choquel, J.-B., and Benjelloun, M. A recursive fusion filter for angular data. In *Proceedings of the IEEE International Conference on Robotics and Biomimetics,* Guilin, China, 2009, 882–887.

[17] Stienne, G., Reboul, S., Azmani, M., Choquel, J., and Benjelloun, M. A multi-temporal multi-sensor circular fusion filter. *Information Fusion*, 18 (July 2013), 86–100.

[18] Stienne, G., Reboul, S., Choquel, J., and Benjelloun, M. Circular data processing tools applied to a phase open loop architecture for multi-channels signals tracking. In *Proceedings of the IEEE/ION Position Location and Navigation Symposium*, Myrtle Beach, SC, 2012, 633–642.

[19] Mokhtari, K. E., Reboul, S., Azmani, M., Choquel, J.-B. Salaheddine, H., Amami, B., and Benjelloun, M. A circular interacting multi-model filter applied to map matching. In *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, 2013.

[20] Traa, J., and Smaragdis, P. A wrapped Kalman filter for azimuthal speaker tracking. *IEEE Signal Processing Letters*, 20, 12 (2013), 1257–1260.

[21] Chiuso, A., and Picci, G. Visual tracking of points as estimation on the unit sphere. In *The Confluence of Vision and Control*. London, United Kingdom: Springer, 1998, pp. 90–105.

[22] Markovic, I., Chaumette, F., and Petrovic, I. Moving object detection, tracking and following using an omnidirectional camera on a mobile robot. In *Proceedings of the 2014 IEEE International Conference on Robotics and Automation*, Hong Kong, China, June 2014.

[23] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Recursive nonlinear filtering for angular data based on circular distributions. In *Proceedings of the 2013 American Control Conference*, Washington, DC, June 2013.

[24] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Nonlinear measurement update for estimation of angular systems based on circular distributions. In *Proceedings of the 2014 American Control Conference*, Portland, Oregon, June 2014.

[25] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Deterministic approximation of circular densities with symmetric Dirac mixtures based on two circular moments. In *Proceedings of the 17th International Conference on Information Fusion*, Salamanca, Spain, July 2014.

[26] Kurz, G., Faion, F., and Hanebeck, U. D. Constrained object tracking on compact one-dimensional manifolds based on directional statistics. In *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation*, Montbeliard, France, Oct. 2013.

[27] Gilitschenski, I., Kurz, G., and Hanebeck, U. D. Bearings-only sensor scheduling using circular statistics. In *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, July 2013.

[28] Kurz, G., Dolgov, M., and Hanebeck, U. D. Nonlinear stochastic model predictive control in the circular domain. In *Proceedings of the 2015 American Control Conference*, Chicago, IL, July 2015.

[29] Kurz, G., Gilitschenski, I., Dolgov, M., and Hanebeck, U. D. Bivariate angular estimation under consideration of dependencies using directional statistics. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, Dec. 2014.

[30] Glover, J., and Kaelbling, L. P. Tracking the spin on a ping pong ball with the quaternion Bingham filter. In *Proceedings of the 2014 IEEE Conference on Robotics and Automation*, Hong Kong, China, 2014.

[31] Mardia, K. V., and Sutton, T. W. A model for cylindrical variables with applications. *Journal of the Royal Statistical Society, Series B*, 40, 2 (1978), 229–233.

[32] Feiten, W., Lang, M., and Hirche, S. Rigid motion estimation using mixtures of projected Gaussians. In *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, 2013.

[33] Darling, J. E., and DeMars, K. J. Rigid body attitude uncertainty propagation using the Gauss-Bingham distribution. In *Proceedings of the 25th AAS/AIAA Space Flight Mechanics Meeting*, Williamsburg, VA, Jan. 2015.

[34] Crouse, D. Cubature/unscented/sigma point Kalman filtering with angular measurement models. In *Proceedings of the 18th International Conference on Information Fusion*, Washington, DC, 2015, 1550–1557.

[35] Gray, J. E., and Addison, S. R. The Rayleigh problem is everywhere!" In *Proceedings of the SPIE Defense, Security, and Sensing Conference*, Orlando, FL, Apr. 2010.

[36] Gray, J. E., and Alouani, A. T. Characterization of the PDFs of coordinate transformations in tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 50, 2 (2014), 1479–1492.

[37] Al-Sharadqah, and A., Chernov, N., Error analysis for circle fitting algorithms. *Electronic Journal of Statistics*, 3, (2009), 886–911.

[38] Baum, M., and Hanebeck, U. D. Extended object tracking with random hypersurface models. *IEEE Transactions on Aerospace and Electronic Systems*, 50 (Jan. 2014),149–159.

[39] Jaynes, E. T. The well-posed problem. *Foundations of Physics*, 3, 4 (1973), 477–492.

[40] Chirikjian, G. S. *Stochastic Models, Information Theory, and Lie Groups,* Vol. 1. New York: Birkhäuser, 2009.

[41] Julier, S. J., and LaViola, J. J. On Kalman filtering with nonlinear equality constraints. *IEEE Transactions on Signal Processing*, 55, 6 (2007), 2774–2784.

[42] Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50, 2 (2002), 174–188.

[43] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. 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*, Bonn, Germany, Oct. 2014.

[44] Abramowitz, M., and Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (10th ed.). New York: Dover, 1972.

[45] Willsky, A. S. Fourier series and estimation on the circle with applications to synchronous communication—part I: analysis. *IEEE Transactions on Information Theory*, 20, 5 (1974), 584–590.

[46] Pfaff, F., Kurz, G., and Hanebeck, U. D. Multimodal circular filtering using Fourier series. In *Proceedings of the 18th International Conference on Information Fusion*, Washington, DC, July 2015.

[47] Amos, D. E. Computation of modified Bessel functions and their ratios. *Mathematics of Computation*, 28, 125 (1974), 239–251.

[48] Sra, S. A short note on parameter approximation for von Mises–Fisher distributions: and a fast implementation of $I_s(x)$. *Computational Statistics*, 27, 1 (2012), 177–190.

[49] Stienne, G. Traitements des signaux circulaires appliqués à l'altimétrie par la phase des signaux GNSS. Ph.D. dissertation, Université du Littoral Côte d'Opale, Dunkirk, France, Dec. 2013.

[50] Kotecha, J. H., and Djuric, P. Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51, 10 (2003), 2592–2601.

[51] Hanebeck, U. D., and Klumpp, V. Localized cumulative distributions and a multivariate generalization of the Cramér-von Mises distance. In *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Seoul, Republic of Korea, Aug. 2008, 33–39.

[52] Hanebeck, U. D., and Lindquist, A. Moment-based Dirac mixture approximation of circular densities. In *Proceedings of the 19th IFAC World Congress*, Cape Town, South Africa, Aug. 2014.

[53] Johnson, S. G. Faddeeva package. Dec. 2012. Available: http://ab-initio.mit.edu/wiki/index.php/Faddeeva_Package.

[54] Goldberg, D. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys*, 23, 1 (Mar. 1991), 5–48.

[55] Hanebeck, U. D. PGF 42: progressive Gaussian filtering with a twist. In *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, July 2013.

[56] Nitzan, E., Routtenberg, T., and Tabrikian, J. Cyclic Bayesian Cramér-Rao Bound for filtering in circular state space. In *Proceedings of the 18th International Conference on Information Fusion*, Washington, DC, 2015, 734–741.

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

[58] Petersen. K. B., and Pedersen, M. S. The matrix cookbook, Nov. 2012. Available: http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf.

Gerhard Kurz

Gerhard Kurz received his diploma in computer science from the Karlsruhe Institute of Technology (KIT), Germany, in 2012. Afterwards, he obtained his Ph.D. in 2015 at the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT), Germany. His research interests include directional filtering, nonlinear estimation, and medical data fusion. He has authored multiple award-winning publications on these topics.

Igor Gilitschenski

Igor Gilitschenski is currently a postdoctoral researcher at the Autonomous Systems Laboratory, ETH Zurich. He obtained his Ph.D. degree at the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT) in 2015. Before joining the KIT, he obtained his diploma degree in mathematics from the University of Stuttgart, Germany. His research interests include stochastic filtering theory and dynamic state estimation with a focus on nonlinear systems and nonlinear domains.

Uwe D. Hanebeck

Uwe D. Hanebeck is a chaired professor of Computer Science at the Karlsruhe Institute of Technology (KIT) in Germany and director of the Intelligent Sensor-Actuator-Systems Laboratory (ISAS). Since 2005, he is the chairman of the Research Training Group RTG 1194 "Self-Organizing Sensor-Actuator-Networks" financed by the German Research Foundation.

Prof. Hanebeck obtained his Ph.D. degree in 1997 and his habilitation degree in 2003, both in electrical engineering from the Technical University in Munich, Germany. His research interests are in the areas of information fusion, nonlinear state estimation, stochastic modeling, system identification, and control with a strong emphasis on theory-driven approaches based on stochastic system theory and uncertainty models. Research results are applied to various application topics like localization, human-robot-interaction, assistive systems, sensor-actuator-networks, medical engineering, distributed measuring system, and extended range telepresence. Research is pursued in many academic projects and in a variety of cooperations with industrial partners.

Uwe D. Hanebeck was the General Chair of the "2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)", Program Co-Chair of the "11th International Conference on Information Fusion (Fusion 2008)", Program Co-Chair of the "2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)", Regional Program Co-Chair for Europe for the "2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)", and will be General Chair of the "19th International Conference on Information Fusion (Fusion 2016)". He is a Member of the Board of Directors of the International Society of Information Fusion (ISIF), Editor-in-chief of its Journal of Advances in Information Fusion (JAIF), and associate editor for the letter category of the IEEE Transactions on Aerospace and Electronic Systems (TAES). He is author and coauthor of more than 300 publications in various high-ranking journals and conferences.