Discrete Recursive Bayesian Filtering on Intervals and the Unit Circle

Gerhard Kurz[1], Florian Pfaff[1], and Uwe D. Hanebeck[1]

***Abstract*— Various applications necessitate the estimation of quantities defined on intervals or the unit circle, which can also be parameterized as an interval. These applications include estimation of joint angles that are either limited to a certain range or that are 360-degree-periodic. For this purpose, we consider two approaches based on discretizing the state space that use fundamentally different density representations. We show how prediction and measurement update for systems with nonlinear dynamics and nonlinear measurement models can be performed in each representation. In particular, we discuss the choices that go into designing discrete filters, which are sometimes taken for granted. A thorough comparison and a numerical evaluation of both approaches show the advantages and disadvantages of each method.**

I. Introduction

We consider state estimation where the state is defined on certain compact spaces, in particular intervals in $\mathbb{R}$ and the unit circle. The unit circle in $\mathbb{R}^2$ can be parameterized as the interval $[0, 2\pi)$ with a special topology using the mapping $\phi \mapsto [\cos(\phi), \sin(\phi)]^T$. State estimation in these spaces differs significantly from estimation on the real vector space $\mathbb{R}^n$ because the domain is bounded.

Applications where state values inside an interval are of interest include the estimation of joint angles with limited range [1, Sec. III-A], [2] of the position of an object constrained to a certain area, and of the speed of a vehicle between zero and its maximum speed. Estimation on the unit circle plays an important role in signal processing [3], [4], meteorology [5], aerospace (e.g., the heading of an airplane) [6], and many other fields.

For bounded spaces, it is possible to perform a discretization and consider estimation on a space with a finite but sufficiently large set of possible values. While the idea of discretization may seem obvious, there are a number of important choices when designing a discrete filter for bounded spaces.

Let us imagine that we want to discretize the state using a grid with L equidistant values. The key question that has to be considered deals with the way the discrete probability is interpreted. The first interpretation relies on a piecewise constant distribution (i.e., a histogram), where the original space is subdivided into L intervals and we implicitly assume a uniform distribution within each interval. Thus, the resulting density is actually a continuous density on a continous domain with a finite number of different values, which is determined by a discrete set of parameters. However, as we will see later, the filtering problem can be reduced to filtering on a discrete domain by integrating over each interval. As a result, we only need to consider a finite number of possible states, rather than a continuous-valued state.

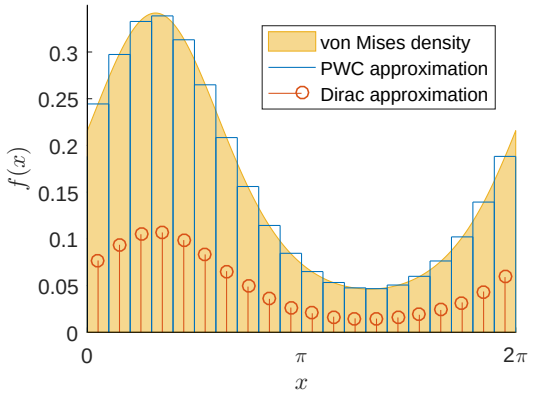

Fig. 1: Discrete approximations of a continous von Mises density on the unit circle with $L = 20$ discretization steps.

On the contrary, the second approach consists in a particle approach, where we define a mixture of L weighted Dirac delta functions located on a grid. This corresponds to taking each of the L intervals from the first interpretation and concentrating all of its probability mass into a point mass located at the center of the interval. Thus, it can be seen as a discrete distribution on a continous domain. This is quite similar to a particle filter [7], except that the particle positions are fixed. Both interpretations require the same amount of memory (L values), but the induced density has different properties and the computational effort of estimation algorithms based thereon can differ significantly. These concepts are illustrated in Fig. 1.

Grid-based filtering has a long history dating back to the Wonham filter [8]. It has been applied in robotics, e.g., by Burgard et al. [9], and is sometimes called histogram filtering [10, Sec. 4.1]. Approaches relying on grid-based discretization have also been commonly used for solving partially observable Markov decision processes (POMDPs) [11], [12], [13].

In contrast to discrete methods, there are also continuous approaches to recursive filtering on intervals and the unit circle. Many researchers rely on classical approaches that are designed for real numbers $\mathbb{R}$, e.g., based on the Kalman filter [14], and do not consider the bounded nature of the interval or the periodicity of the unit circle. Sometimes, neglecting

[1] The authors are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany (e-mail: gerhard.kurz@kit.edu, florian.pfaff@kit.edu, uwe.hanebeck@ieee.org.

the true topology of the considered state space still yields acceptable results, but under certain circumstances (e.g., large uncertainties) these approaches tend to fail. For this reason, filters designed for the interval, e.g., truncated Gaussian filters [15], [16, Sec. 7.5.4], or for the unit circle, e.g, based on Matrix Lie groups [17] or on von Mises or wrapped normal distributions [18], [6], have been proposed. The downside of those filters is that a particular density is assumed and they are thus limited to problems where this density assumption holds (at least approximately). Finally, methods based on Fourier series have been proposed [19], [20], [21] which can represent a wide range of densities, but are limited to the periodic case and require special tweaking to ensure positivity and normalization of the density. Moreover, they need many coefficients to handle high frequencies and their complexity is typically not linear in the number of coefficients.

The contribution of this paper can be summarized as follows. We present two different interpretations of discrete filtering and show how they can be applied to estimation on the unit circle and on intervals. The two approaches are first compared regarding their theoretical properties and then put to the test in a simulated example.

In this paper, we use the following notation.

$P(\cdot)$	discrete probability mass function
$f(\cdot)$	continous probability density function
$\mathbf{A}_k(i,j)$	matrix $\mathbf{A}_k$, element at row i, column j
$\mathbb{1}_M(\cdot)$	indicator function of the set M
$\delta(\cdot)$	Dirac delta function

II. Problem Formulation

We consider an arbitrary bounded interval contained in $\mathbb{R}$ as the domain of the system state. For our purposes, it does not matter if the interval is closed, open, or half-open. The topology of the interval can either be the subset topology of $\mathbb{R}$ or the topology of a circle, i.e., the lower and upper borders of the interval are connected. In the following, we will assume that the domain is without loss of generality defined as $[0, 2\pi)$ as this interval is commonly used as a parameterization for the circle. At this point, we do not choose one of the two topologies, however. Instead, we distinguish the topologies only where it actually makes a difference.

A discrete-time system with state $x_k \in [0, 2\pi)$ at time step k is given by the general system equation

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

where $w_k \in W$ is noise in some noise space W. Sometimes, it is helpful if the transition density $f(x_{k+1}|x_k)$ is also known. In the case of additive system noise, the transition density can easily be derived from (1) (see [6]). The general measurement model is given by

$$z_k = h_k(x_k, v_k) \ , \tag{2}$$

where $z_k \in Z$ is the measurement in some measurement space Z and $v_k \in V$ is noise in some noise space V. For some of the approaches discussed in this paper, we also require the likelihood $f(z_k|x_k)$ to be known. In the case of additive measurement noise, the likelihood can easily be derived from (2) as discussed, e.g., in [6].

In order to simplify the calculations, it is sometimes helpful to approximate the noise densities $f_k^w(w_k)$ and $f_k^v(v_k)$ using (weighted) samples. These samples can be obtained using stochastic or deterministic sampling. In stochastic sampling, the samples are drawn randomly from the given distribution, whereas deterministic sampling places the samples optimally with respect to an optimality criterion. The benefit of deterministic sampling is that typically a much smaller number of samples is required for the same approximation quality. Methods for deterministic sampling of circular densities are given in [22], [23], [24]. Algorithms for Gaussian densities can be found in [25], [26].

III. Piecewise Constant Approach

In this section, we discuss approaches for discretized filtering based on piecewise constant distributions. The derived filters can be seen as discrete-time Wonham filters [8]. It is noteworthy that all methods discussed here are independent of the underlying topology.

Definition 1 (Piecewise Constant Distribution)**.** A piecewise constant (PWC) distribution with L discretization steps on the interval $[0, 2\pi)$ is given by the probability density function

$$\mathcal{PWC}(x; \gamma_1, \ldots, \gamma_L) = \begin{cases} \gamma_1 \ , & x \in I_1 \\ \vdots & \vdots \\ \gamma_L \ , & x \in I_L \end{cases} ,$$

where $I_j = [I_j^{\min}, I_j^{\max})$ with equidistant interval borders

$$I_j^{\min} = (j-1)\frac{2\pi}{L} \quad \text{and} \quad I_j^{\max} = j\frac{2\pi}{L}$$

for $j = 1, \ldots, L$, and $\frac{2\pi}{L}\sum_{j=1}^{L} \gamma_j = 1$.

Obviously, it holds that $I_i \cap I_j = \emptyset$ for $i \neq j$ and $\bigcup_{j=1}^{L} I_j = [0, 2\pi)$, i.e., $I_1, \ldots, I_L$ constitute a partition of the set $[0, 2\pi)$. In the following, we assume $x_k \sim \mathcal{PWC}(x; \gamma_1, \ldots, \gamma_n)$, i.e., x_k is PWC-distributed. Thus, we have

$$P(x_k \in I_i) = \int_{I_i} \mathcal{PWC}(x; \gamma_1, \ldots, \gamma_L) \, \mathrm{d}x_k = \frac{2\pi}{L}\gamma_i \tag{3}$$

for $i = 1, \ldots, L$. As a result, the probability mass within each interval is proportional to the corresponding parameter γ_i.

A. Prediction

In order to perform the prediction step, we first obtain the system matrix $\mathbf{A}_k(i,j) = P(x_{k+1} \in I_i | x_k \in I_j)$. The entry $\mathbf{A}_k(i,j)$ corresponds to the probability of the state at time step $k+1$ being in the interval I_i given that the state at time step k was in the interval I_j. This probability can be computed according to

$$\begin{aligned} &P(x_{k+1} \in I_i | x_k \in I_j) \\ &= \int_W \int_{I_j} f_k^w(w_k) \mathbb{1}_{I_i}(a_k(x_k, w_k)) \, \mathrm{d}x_k \, \mathrm{d}w_k \ . \end{aligned}$$

In general, these integrals have to be evaluated numerically. If the noise w_k is approximated using L^w samples with positions $\beta_1^w, \ldots, \beta_{L^w}^w$ and weights $\gamma_1^w, \ldots, \gamma_{L^w}^w$, we can simplify the integral and obtain

$$P(x_{k+1} \in I_i | x_k \in I_j) \approx \sum_{l=1}^{L^w} \gamma_l^w \int_{I_j} \mathbb{1}_{I_i}(a_k(x_k, \beta_l^w)) \, \mathrm{d}x_k \ .$$

Alternatively, if the transition density $f(x_{k+1}|x_k)$ is known, we can obtain the entries of $\mathbf{A}$ according to

$$P(x_{k+1} \in I_i | x_k \in I_j) = \int_{I_i} \int_{I_j} f(x_{k+1}|x_k) \, \mathrm{d}x_k \, \mathrm{d}x_{k+1} \ .$$

While this representation also necessitates the evaluation of a two-dimensional integral, numerical evaluation is typically more stable because the discontinuous indicator function is not present. Either way, the evaluation of the integral(s) can be performed offline if the system function $a_k(\cdot,\cdot)$ and the distribution $f_k^w(w_k)$ of the system noise are time-invariant. For certain system models or certain noise distributions, it may be possible, to find analytical solutions to the aforementioned integrals.

Once the system matrix is known, the predicted PWC density can be obtained using

$$P(x_{k+1} \in I_i) = \sum_{j=1}^{L} \mathbf{A}_k(i,j) P(x_k \in I_j) \ .$$

B. Measurement Update

When performing the measurement update, we consider two different approaches. The first possibility is to discretize the measurement space just as we discretize the state space. The second possibility is to retain a continuous measurement space and compute the likelihood of one specific measurement.

1) Discretized Measurement Space: If the measurement space Z is an interval in $\mathbb{R}$, we can discretize it into intervals $\hat{I}_1, \ldots, \hat{I}_m$. The number of discretization steps does not need to be identical to the number of discretization steps used for the state space.

Now, we construct the measurement matrix $\mathbf{H}_k(i,j) = P(z_k \in \hat{I}_i | x_k \in I_j)$, where

$$\begin{aligned} &P(z_k \in \hat{I}_i | x_k \in I_j) \\ &\quad = \int_V \int_{I_j} f_k^v(v_k) \mathbb{1}_{\hat{I}_i}(h_k(x_k, v_k)) \, \mathrm{d}x_k \, \mathrm{d}v_k \ . \end{aligned}$$

These integrals can in general only be evaluated numerically. If the noise v_k is approximated by L^v samples with positions $\beta_1^v, \ldots, \beta_{L^v}^v$ and weights $\gamma_1^v, \ldots, \gamma_{L^v}^v$, we can simplify the integral and obtain

$$P(z_k \in \hat{I}_i | x_k \in I_j) \approx \sum_{l=1}^{L^w} \gamma_l^v \int_{I_j} \mathbb{1}_{\hat{I}_i}(h_k(x_k, \beta_l^v)) \, \mathrm{d}x_k \ .$$

Alternatively, if the likelihood $f(z_k|x_k)$ is known, we can obtain the entries of $\mathbf{H}_k$ using

$$P(z_k \in \hat{I}_i | x_k \in I_j) = \int_{\hat{I}_i} \int_{I_j} f(z_k|x_k) \, \mathrm{d}x_k \, \mathrm{d}z_k \ .$$

Once again, this allows avoiding integration over the discontinuous indicator function. In all cases, the evaluation of the integrals can be performed offline if the measurement function $h_k(\cdot,\cdot)$ and measurement noise $f_k^v(v_k)$ are time-invariant.

When we receive a measurement, we can perform the measurement update by finding the interval $\hat{I}_i$ that contains z_k and selecting the corresponding row of the matrix $\mathbf{H}_k$. The update is then performed using Bayes' rule according to

$$P(x_k \in I_j | z_k \in \hat{I}_i) \propto P(z_k \in \hat{I}_i | x_k \in I_j) P(x_k \in I_j) \ .$$

Note that we omit the normalization term here as it is easy to renormalized the density afterwards.

2) Continuous Measurement Space: If we cannot or do not want to discretize the measurement space (e.g., it is very large or even unbounded), we can use the following approach instead. For a fixed z_k, we use Bayes' theorem to obtain

$$\begin{aligned} P(x_k \in I_j | z_k) &= \int_{I_j} f(x_k|z_k) \, \mathrm{d}x_k \\ &\propto \int_{I_j} f(z_k|x_k) f(x_k) \, \mathrm{d}x_k \\ &= \frac{2\pi}{L} \gamma_j \int_{I_j} f(z_k|x_k) \, \mathrm{d}x_k \ , \end{aligned}$$

where we use integration over the continous likelihood function. The integral has to be evaluated online in every measurement update step for each I_j even if the likelihood function $f(z_k|x_k)$ is time-invariant because the integration depends on the actual measurement z_k. In the end, we once again renormalize the result to ensure a valid density.

IV. Dirac-based Approach

This section is devoted to the second type of discretization approach—the use of a Dirac mixture.

Definition 2 (Dirac Mixture)**.** A Dirac mixture density on $[0, 2\pi)$ with L components is given by

$$\mathcal{D}(x; \beta_1, \ldots, \beta_L, \gamma_1, \ldots, \gamma_L) = \sum_{j=1}^{L} \gamma_j \delta(\beta_j - x) \ ,$$

where $\beta_1 < \cdots < \beta_L \in [0, 2\pi)$ with $\gamma_1, \ldots, \gamma_L \geq 0$ and $\sum_{j=1}^{L} \gamma_j = 1$.

Note that the Dirac mixture is sometimes referred to as wrapped Dirac mixture if the underlying space is periodic (see [6], [27]).

In the following, we use Dirac mixtures where the Dirac components have fixed equidistant positions

$$\beta_j = (j - 1/2)\frac{2\pi}{L}, \quad j = 1, \ldots, L \tag{4}$$

on a grid and variable weights to represent the state estimate. Contrary to a sequential importance resampling (SIR) particle filter [7], the particles can only change their weights but they cannot move. This design choice has the advantage that the state space is always evenly covered by particles and particle agglutination as well as inadequate state space coverage as present in SIR particle filters is avoided. Also, contrary

to a SIR particle filter, the proposed approach is entirely deterministic. On the other hand, the ability to represent very concentrated densities is obviously limited by the resolution of the underlying grid.

The probabilistic interpretation of the Dirac-based approach is given by

$$P(x_k = \beta_j) = \gamma_j \ . \tag{5}$$

In contrast to (3), we do not consider L intervals, but just L discrete points in (5).

A. Prediction

In order to compute the prediction, we first propagate the samples through the system function. Let us ignore for now that the resulting density is supposed to be defined on the grid (4). Then, the exact propagated density is given by

$$f(x_{k+1}) = \sum_{i=1}^{L} \gamma_i \int_W \delta(x_{k+1} - a_k(\beta_i, w_k)) \, \mathrm{d}w_k \ .$$

This integral can be precomputed for time-invariant a_k and w_k. If the noise w_k is approximated by L^w samples according to $\sum_{j=1}^{L^w} \gamma_j^w \delta(x - \beta_j^w)$, we can simplify the integral and obtain

$$f(x_{k+1}) \approx \sum_{i=1}^{L} \sum_{j=1}^{L^w} \gamma_i \gamma_j^w \delta(x_{k+1} - a_k(\beta_i, \beta_j^w)) \ .$$

Now, we present strategies to reapproximate $f(x_{k+1})$ with samples that are located on the grid defined in (4).

1) Nearest Neighbor: The nearest neighbor approach starts by initializing all weights of the predicted density with zero. For every component, we find the grid point closest to $a_k(\beta_i, \beta_j^w)$, i.e.,

$$\arg\min_l d(\beta_l, a_k(\beta_i, \beta_j^w))$$

where $d(\cdot, \cdot)$ is topology-aware distance function. If we are performing estimation on an interval, we would typically use $d(x, y) = |x - y|$. In the case of estimation on the unit circle, we would use the geodetic distance

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

instead. Either way, it is possible to find the closest grid point in $\mathcal{O}(1)$ using modulo arithmetic due to the equidistant grid. Then, we add the weight $\gamma_i \gamma_j^w$ to the weight of the considered Dirac delta component.

While this approach is very easy to implement, it suffers from significant problems. In particular, small changes to the density are completely disregarded in certain cases. For example, consider $a_k(x_k, w_k) = x_k + \epsilon$ with $|\epsilon| < \pi/L$. In this case, all probability mass would be assigned to the same Dirac component as before, i.e., the density would not change at all.

2) Proportional: An alternative to the nearest neighbor approach that alleviates the aforementioned problem is the proportional method. The basic idea consists in finding the two nearest neighbors and dividing the probability mass between them depending on the distance.

In the following, we distinguish several cases. If we have $\beta_l < a_k(\beta_i, \beta_j^w) < \beta_{l+1}$, we can assign the weight depending on the distance in a straightforward manner. The component at β_l is assigned the weight

$$\frac{d(\beta_{l+1}, a_k(\beta_i, \beta_j^w))}{d(\beta_l, a_k(\beta_i, \beta_j^w)) + d(\beta_{l+1}, a_k(\beta_i, \beta_j^w))} \gamma_i \gamma_j^w$$

and the component at β_{l+1} is assigned the weight

$$\frac{d(\beta_l, a_k(\beta_i, \beta_j^w))}{d(\beta_l, a_k(\beta_i, \beta_j^w)) + d(\beta_{l+1}, a_k(\beta_i, \beta_j^w))} \gamma_i \gamma_j^w \ .$$

Obviously, both weights sum up to $\gamma_i \gamma_j^w$, i.e., the total probability mass is retained.

Certain special cases arise at the borders depending on topology. If the underlying topology is that of an interval, we consider two cases. For $a_k(\beta_i, \beta_j^w) < \beta_1$, the component at β_1 is assigned the entire weight $\gamma_i \gamma_j^w$. Correspondingly, for $a_k(\beta_i, \beta_j^w) > \beta_L$, the component at β_L is assigned the entire weight $\gamma_i \gamma_j^w$.

In a periodic topology, we proceed as follows. For $a_k(\beta_i, \beta_j^w) < \beta_1$ or $a_k(\beta_i, \beta_j^w) > \beta_L$, we divide the probability mass between β_1 and β_L using topology-aware distance, i.e, the component at β_1 is assigned the weight

$$\frac{d(\beta_L, a_k(\beta_i, \beta_j^w))}{d(\beta_1, a_k(\beta_i, \beta_j^w)) + d(\beta_L, a_k(\beta_i, \beta_j^w))} \gamma_i \gamma_j^w$$

and the component at β_L is assigned the weight

$$\frac{d(\beta_1, a_k(\beta_i, \beta_j^w))}{d(\beta_l, a_k(\beta_1, \beta_j^w)) + d(\beta_L, a_k(\beta_i, \beta_j^w))} \gamma_i \gamma_j^w \ .$$

All cases discussed above can be handled in $\mathcal{O}(1)$ again.

B. Measurement Update

Computation of the measurement update is straightforward in the Dirac-based approach. We can directly apply Bayes' rule

$$f(x_k|z_k) \propto f(z_k|x_k) f(x_k) = \sum_{j=1}^{L} \gamma_j f(z_k|\beta_j) \delta(\beta_j - x_k) \ ,$$

i.e., we multiply all weights with the likelihood. Afterwards, the weights are renormalized to sum up to one. This method for performing the measurement update does not require discretization of the measurement space Z.

V. Comparison

An overview of the presented algorithms is given in Table I and Table II. In each table, we show different cases where we distinguish the density representation, whether the system model and noise distribution are time-variant, whether the noise is given as a sample set, and whether the measurement space is discretized. It can be seen that

in most cases, the PWC-based algorithms tend to have a higher computational cost than the Dirac-based algorithms. Typically, the computation time is dominated by the numerical integration algorithm, i.e., it is highly desirable to avoid numerical integration at runtime. Also, we would like to avoid 2D numerical integration altogether as it is substantially more expensive than 1D numerical integration.

There are also some theoretical differences between the two discretization approaches. It can be shown that for a continous pdf $f(\cdot)$, the PWC approximation converges pointwise to the true pdf as L approaches infinity. Formally, we have

$$\lim_{L\to\infty} |\mathcal{PWC}(x;\gamma_1,\ldots,\gamma_L) - f(x)| = 0 \quad \forall x \in [0,2\pi) \ ,$$

where $\gamma_i = \int_{I_i} f(x)\,\mathrm{d}x$ for $1 \leq i \leq L$. Such a statement is not possible for the Dirac distribution as it never approaches a continous density. Still, it can be shown for piecewise continuous $f(\cdot)$ that the limit

$$\lim_{L\to\infty} \left| \int_a^b \mathcal{D}(x;\beta_1,\ldots,\beta_L,\gamma_1,\ldots,\gamma_L)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x \right|$$

is zero for all fixed $a < b \in [0,2\pi)$, where $\beta_1,\ldots,\beta_L$ are given by (4) and $\gamma_i = f(\beta_i)$ for $1 \leq i \leq L$. Intuitively, this means that the probability mass of the approximation that is contained in $[a,b]$ converges to the true probability mass in $[a,b]$.

Another interesting aspect to consider is the entropy, i.e., the information content of the densities. The continuous entropy for a PWC distribution is given by

$$\begin{aligned} &-\int_0^{2\pi} \mathcal{PWC}(x;\gamma_1,\ldots,\gamma_L) \log \mathcal{PWC}(x;\gamma_1,\ldots,\gamma_L)\,\mathrm{d}x \\ &\quad = -\frac{2\pi}{L}\sum_{i=1}^{L} \gamma_i \log \gamma_i \ , \end{aligned}$$

i.e., it can be simplified to a sum. On the contrary, the continuous entropy for a Dirac distribution is not well defined. However, it is possible to compute the discrete entropy for a Dirac distribution according to

$$-\sum_{i=1}^{L} \gamma_i \log \gamma_i \ ,$$

which is identical to the entropy of the PWC distribution with the same weights up to a factor of $2\pi/L$. Note, however, that the discrete entropy does not depend on the location parameters $\beta_1,\ldots,\beta_L$ at all, i.e., it does not depend on the grid. As a result, the discrete entropy is not a very useful measure if the Dirac locations encode important information as is the case in SIR particle filters [7].

It should be noted that there are some further alternative approaches. First of all, it is possible to consider a piecewise linear rather than a piecewise constant density. However, some of the simplifications that are possible for piecewise constant densities are not possible in this case. Second, the grid does not necessarily need to be uniform. Non-equidistant grids may allow a more accurate representation of the density in areas that are of particular interest and adaptive discretization approaches may be achieve better results by dynamically adjusting the grid resolution. Finally, it is also possible to discretize the frequency domain as is done in approaches based on Fourier series, e.g., [19], [20].

Problem			Solution	
density	time-variant	noise sampled	numerical integration	online complexity
PWC	no	no	offline 2D	$\mathcal{O}(L^2)$
PWC	yes	no	online 2D	$\mathcal{O}(L^2 + L^2 \cdot I_2)$
PWC	no	yes	offline 1D	$\mathcal{O}(L^2)$
PWC	yes	yes	online 1D	$\mathcal{O}(L^2 + L \cdot I_1)$
Dirac	no	no	offline 1D	$\mathcal{O}(L)$
Dirac	yes	no	online 1D	$\mathcal{O}(L + L \cdot I_1)$
Dirac	no	yes	no	$\mathcal{O}(L \cdot L^w)$
Dirac	yes	yes	no	$\mathcal{O}(L \cdot L^w)$

TABLE I: Prediction algorithms. Note that we assume the cost for a single numerical integration in d dimensions to be I_d. The complexity refers to the online complexity only.

Problem				Solution	
density	time-variant	noise sampled	meas. space	numerical integration	online complexity
PWC	no	no	discr.	offline 2D	$\mathcal{O}(L)$
PWC	yes	no	discr	online 2D	$\mathcal{O}(L + L \cdot I_2)$
PWC	no	yes	discr.	offline 1D	$\mathcal{O}(LL^v)$
PWC	yes	yes	discr.	online 1D	$\mathcal{O}(LL^v + L \cdot I_1)$
PWC	no	N/A	cont.	online 1D	$\mathcal{O}(L + L \cdot I_1)$
PWC	yes	N/A	cont.	online 1D	$\mathcal{O}(L + L \cdot I_1)$
Dirac	yes/no	N/A	cont.	no	$\mathcal{O}(L)$

TABLE II: Measurement update. Note that we assume the cost for a single numerical integration in d dimensions to be I_d. The complexity refers to the online complexity only.

VI. Evaluation

In the following, we will evaluate the proposed approaches in a strongly nonlinear scenario on the unit circle. For this purpose, we consider the function proposed by Gilitschenski et al. in [28] as a system and a measurement model. It is given by

$$h_c(\theta) = \pi \cdot \left(\sin\left(\frac{\operatorname{sign}(\theta-\pi)}{2} \cdot \frac{|\theta-\pi|^c}{\pi^{c-1}} \right) + 1 \right)$$

for $\theta \in [0,2\pi)$, $c \in \mathbb{R}_+$. This function is a continous bijection on $[0,2\pi)$ (see Fig. 2). The nonlinearity of the function can be adjusted by changing the parameter c.

A. Prediction

For the evaluation of the prediction, we assume that the state at time step k is distributed according to a wrapped normal [29, Sec. 2.2.6] distribution

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

with parameters $\mu = \pi$ and $\sigma = 1$, which we denote by $\mathcal{WN}(\pi,1)$. The system model is given by

$$x_{k+1} = h_c(x_k) + w_k \ ,$$

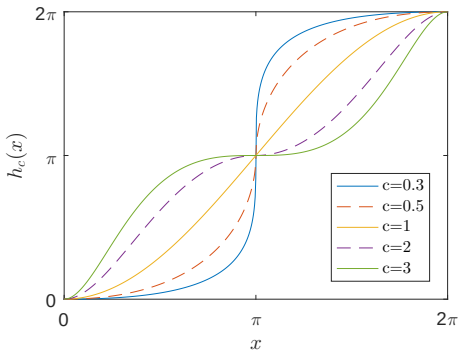

Fig. 2: Nonlinear function used for evaluation.

where the system noise w_k is distributed according to $\mathcal{WN}(0, 0.2)$. Now, we perform a single prediction step using several methods and compare the first trigonometric moment of the result to the first trigonometric moment of the ground truth[1]. The first trigonometric moment of a random variable x is defined as the complex number $\mathbb{E}(\exp(ix))$ and contains both the location and the uncertainty of the distribution (see [30, Sec. 2.4] for a more thorough introduction). Thus, the error measure is given by

$$\left| \int_0^{2\pi} f^{\text{result}}(x) \exp(ix) \, \mathrm{d}x - \int_0^{2\pi} f^{\text{true}}(x) \exp(ix) \, \mathrm{d}x \right| ,$$

where i is the imaginary unit and $|\cdot|$ is the Euclidean norm in the complex plane.

We compare a wrapped normal based filter, a circular version of the one-dimensional UKF, the discrete Dirac-based filter proposed in Sec. IV, the piecewise constant filter proposed in Sec. III, and a circular SIR particle filter (see [6]). For the PWC filter, we obtained the system matrix using 1D numerical integration by using sampled noise consisting of five samples obtained using the sampling scheme given in [22]. For the discrete Dirac-based filter, we applied the proportional weighting scheme.

The evaluation results are depicted in Fig. 3. It can be seen that the discrete Dirac-based filter with $L = 50$ performs best in most cases. The PWC filter with $L = 50$ is better for small c but somewhat worse in most cases even though it has much higher computational cost. Even though it is not as good, the Dirac-based filter with $L = 10$ still produces acceptable results. The WN filter is worse because it reapproximates with a WN density and uses only $L = 5$ particles for propagation. The UKF has an even lower accuracy because it only uses $L = 3$ samples and it does not properly consider the circular nature of the problem. We find that the circular particle filters yield quite poor results. Even when $L = 1000$ particles are used, the performance of the PWC and discrete approaches with $L = 50$ cannot be matched. This illustrates the advantages of grid-based approaches where particles have fixed positions.

[1]The ground truth is obtained using a Dirac-based filter with $L = 5000$.

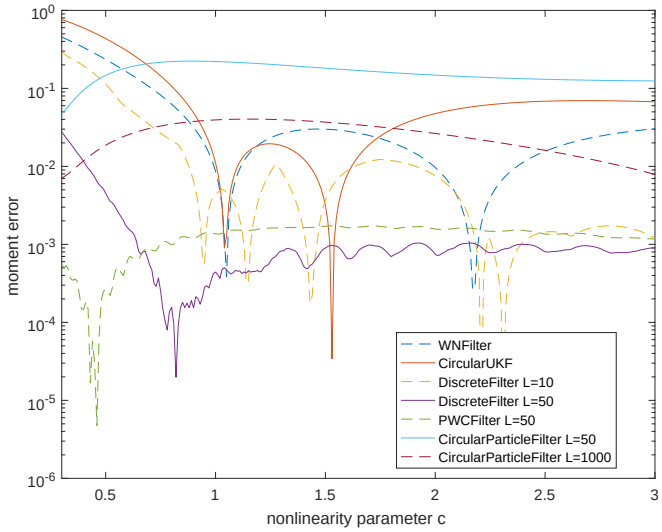

Fig. 3: Moment error after one prediction step.

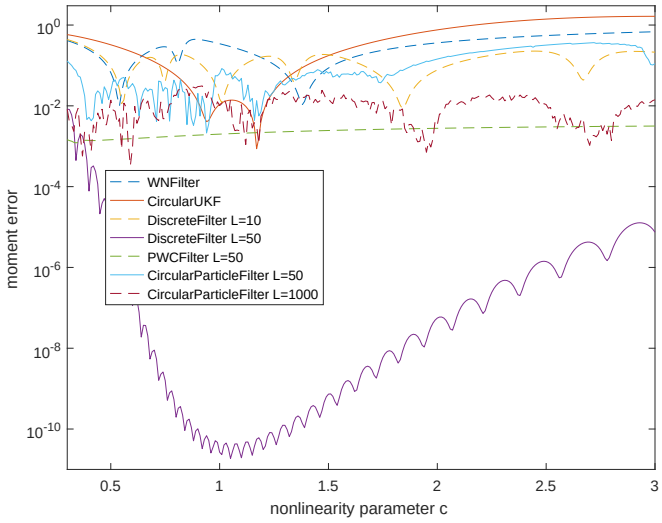

Fig. 4: Moment error after one update step.

B. Measurement Update

When evaluating the measurement update, we assume that the prior density is given by $\mathcal{WN}(\pi, 1)$ and the measurement model is based on the same function as before, i.e.,

$$z_k = h_c(x_k) + v_k \ ,$$

where the measurement noise v_k is distributed according to $\mathcal{WN}(0, 0.2)$. Now, we perform a single measurement update for the measurement $z_k = 5$ and compare the same filters as before with respect to the error in the first trigonometric moment. For the PWC filter, we chose to use the version with the continuous measurement space (see Sec. III-B.2), which requires online numerical integration.

In Fig. 4, we show the results of the evaluation of the measurement update. In this case, the discrete Dirac-based filter with $L = 50$ clearly performs best and even beats the PWC filter by orders of magnitude while being faster to compute. All other approaches are significantly worse.

In particular, we once again notice that the grid filter with $L = 50$ is significantly better than a circular particle filter with $L = 50$.

VII. Conclusion

We have presented two fundamentally different approaches for filtering on intervals and the unit circle. While both approaches have advantages and disadvantages, our evaluation indicates that the Dirac-based approach is usually superior and also much faster to compute.

The proposed approaches can be generalized to a higher number of dimensions. However, it is expected that the number of required grid points typically increases exponentially with the number of dimensions. It is also possible to consider a grid on the unit hypersphere for hyperspherical estimation problems. In that case, it is not obvious how to choose and evenly spread the grid points [31], [32].

Implementations of the proposed filters as well as the other filters used in the evaluation are available as part of libDirectional [33], a MATLAB library for directional filtering.

References

[1] J. Steinbring, C. Mandery, F. Pfaff, F. Faion, T. Asfour, and U. D. Hanebeck, "Real-Time Whole-Body Human Motion Tracking Based on Unlabeled Markers (to appear)," in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sep. 2016.

[2] M. El-Gohary and J. McNames, "Human Joint Angle Estimation with Inertial Sensors and Validation with a Robot Arm," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1759–1767, 2015.

[3] 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.

[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] N. I. Fisher, "Problems with the Current Definitions of the Standard Deviation of Wind Direction," *Journal of Climate and Applied Meteorology*, vol. 26, no. 11, pp. 1522–1529, 1987.

[6] 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.

[7] 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.

[8] W. M. Wonham, "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 2, no. 3, pp. 347–369, 1964.

[9] W. Burgard, D. Fox, D. Hennig, and T. Schmidt, "Estimating the Absolute Position of a Mobile Robot Using Position Probability Grids," in *Proceedings of the National Conference on Artificial Intelligence*, 1996, pp. 896–901.

[10] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents Series. The MIT Press, 2005.

[11] W. S. Lovejoy, "Computationally Feasible Bounds for Partially Observed Markov Decision Processes," *Operations research*, vol. 39, no. 1, pp. 162–175, 1991.

[12] R. Zhou and E. A. Hansen, "An Improved Grid-based Approximation Algorithm for POMDPs," in *International Joint Conference on Artificial Intelligence*, vol. 17, no. 1, 2001, pp. 707–716.

[13] H. Yu and D. P. Bertsekas, "Discretized Approximations for POMDP with Average Cost," in *Proceedings of the 20th conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2004, pp. 619–627.

[14] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[15] C. Lauvernet, J.-M. Brankart, F. Castruccio, G. Broquet, P. Brasseur, and J. Verron, "A Truncated Gaussian Filter for Data Assimilation with Inequality Constraints: Application to the Hydrostatic Stability Condition in Ocean Models," *Ocean Modelling*, vol. 27, no. 1–2, pp. 1–17, 2009.

[16] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[17] I. Markovic, J. Cesic, and I. Petrovic, "On Wrapping the Kalman Filter and Estimating with the SO(2) Group," in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, Jul. 2016.

[18] 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.

[19] A. S. Willsky, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part I: Analysis," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 584–590, 1974.

[20] 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.

[21] ——, "Nonlinear Prediction for Circular Filtering Using Fourier Series," in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, Jul. 2016.

[22] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Deterministic Approximation of Circular Densities with Symmetric Dirac Mixtures Based on Two Circular Moments," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, Jul. 2014.

[23] G. Kurz, I. Gilitschenski, R. Y. Siegwart, and U. D. Hanebeck, "Methods for Deterministic Approximation of Circular Densities (to appear)," *Journal of Advances in Information Fusion*, 2016.

[24] I. Gilitschenski, G. Kurz, U. D. Hanebeck, and R. Siegwart, "Optimal Quantization of Circular Distributions," in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, Jul. 2016.

[25] I. Gilitschenski and U. D. Hanebeck, "Efficient Deterministic Dirac Mixture Approximation," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[26] 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.

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

[28] 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.

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

[30] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Baffins Lane, Chichester, West Sussex, England: Wiley, 1999.

[31] S. Schaefer, J. Hakenberg, and J. Warren, "Smooth Subdivision of Tetrahedral Meshes," in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, ser. SGP '04. New York, USA: ACM, 2004, pp. 147–154.

[32] A. Yershova and S. M. LaValle, "Deterministic Sampling Methods for Spheres and SO (3)," in *2004 IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04.*, vol. 4, 2004, pp. 3974–3980.

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