

# Application of Discrete Recursive Bayesian Estimation on Intervals and the Unit Circle to Filtering on SE(2)

Gerhard Kurz<sup>1</sup>, Florian Pfaff<sup>1</sup>, and Uwe D. Hanebeck<sup>1</sup>

**Abstract**—Many applications require state estimation where possible values of the state are constrained to an interval (say, the valve position in percent) or the unit circle (say, the direction a robot is facing). We present two approaches that rely on a discretization of the state space, which differ in their interpretation of the discretized density. The first option is a piecewise constant density and the second option is a Dirac-mixture density. We show how circular filters can be derived and discuss the advantages and disadvantages of both approaches. In addition, we show how to extend the Dirac-based approach to estimation on the special Euclidean group in 2D, the group of rigid body motions in the plane, using Rao–Blackwellization. All presented methods are thoroughly evaluated in simulations.

This paper is an extended version of [1].

## I. INTRODUCTION

Many practical problems involve estimation of a state contained in certain compact spaces, such as an interval  $I \subset \mathbb{R}$  or the unit circle. We parameterize the unit circle as an angle in the interval  $[0, 2\pi)$  while keeping in mind that it has a different topology.

There are many applications in which the state of a system is restricted to an interval, for example, the estimation of joint angles within predefined admissible limits [2], the location of a robot constrained to a room, and a car’s speed that has to be larger than zero and smaller than its maximum speed. The problem of performing estimation on the unit circle is crucial in signal processing [3], [4], wind energy research [5], aerospace [6], and various other fields.

Bounded spaces can be discretized and nonlinear estimation can be performed more easily. Even though discretization may be an obvious approach, we have to make some important decisions when deriving a filter based on discretizing a bounded space.

Suppose we seek to discretize the state space using  $L$  equidistant points. The key question consists in the semantics of the discrete values. The first interpretation is based on a piecewise constant distribution [7], which can be visualized as a histogram. Here, the original space is partitioned into  $L$  intervals of equal size. Within each interval, we assume a uniform distribution. Thus, the probability density is a continuous density defined on a continuous domain that is parameterized with a finite set of values. As we will show,

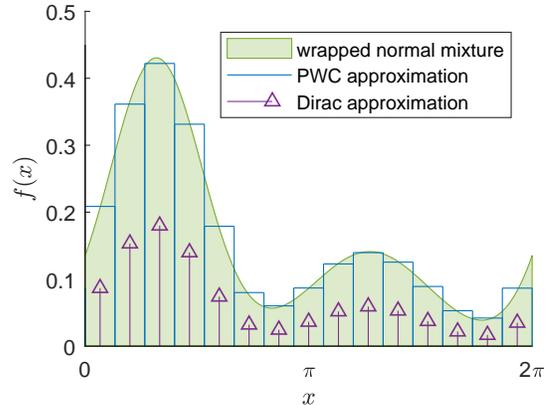


Fig. 1: A continuous wrapped normal mixture density together with its discrete approximation with  $L = 15$  steps.

we can reduce the continuous filtering problem to a discrete problem by performing integration over each of the intervals. Thus, we do not need to consider a continuous state space but can limit ourselves to a finite number of possible states.

In contrast, the second approach uses a sample-based representation. We consider a weighted mixture consisting of  $L$  Dirac delta pulses on an equidistant grid. This can be understood as concentrating all probability mass of each interval from the first interpretation into a point mass [8], [9] in the center of the interval. Hence, we have a discrete distribution defined on a continuous domain. This bears some resemblance to particle filters [10], except that the locations of the particles are fixed. An identical amount of storage ( $L$  values) is required in either interpretation, but the parameterized densities have different properties and the computational complexity may differ significantly. We illustrate both concepts in Fig. 1.

The concept of grid-based filtering dates back to the Wonham filter [11] and has been used, e.g., in robotics [12]. It is sometimes also referred to as histogram filtering [13, Sec. 4.1]. Furthermore, grid-based methods have been commonly applied to partially observable Markov decision processes (POMDPs) [14].

There are also continuous approaches to recursive estimation on intervals as well as the unit circle. Often, classical approaches designed for real numbers, e.g., extensions of the Kalman filter [15, Sec. 5], are employed even though they ignore that the interval is actually bounded or that the unit circle is periodic. Sometimes, these methods yield decent

<sup>1</sup> 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).

results nonetheless, but under certain circumstances, such as large uncertainties, they tend to fail. These issues motivate designing filters specifically for interval-valued problems, e.g., truncated Gaussian filters [15, Sec. 7.5.4], or on the unit circle, e.g., using Matrix Lie groups [16] or wrapped normal or von Mises distributions [3], [6]. The disadvantage of these filters is that they assume a particular probability distribution. To address this limitation, algorithms using a Fourier series as the density representation have been proposed [17], [18]. They can approximate almost any density, provided a sufficient number of Fourier coefficients is used, but they are more suitable to periodic scenarios. Also, it is nontrivial to ensure that the density is positive everywhere. Moreover, many coefficients are required to deal with high spatial frequencies and the computational complexity typically grows faster than linearly with respect to the number of coefficients.

Many practical applications necessitate simultaneous estimation of an interval-valued or circular quantity and a real vector. A very common case is estimation on SE(2), the group of rigid body motions in 2D, for example for tracking the movement of a robot or a vehicle in a plane. A rigid body motion consists of a translation vector in  $\mathbb{R}^2$  and a rotation angle in  $[0, 2\pi)$ . Some approaches to estimation on SE(2) have been presented in [19], [20], and some authors have also considered the related three-dimensional problem [21], [22]. However, these methods are limited by assuming a particular distribution, e.g., variants of the Bingham distribution. As a result, generalizing the grid-based methods to SE(2) is a promising way to remove this limitation.

The contributions of this paper are the following.

- We present two different approaches for discrete filtering on intervals and the unit circle based on the different interpretations as illustrated in Fig. 1.
- We compare these two approaches with respect to their theoretical properties and their filtering performance in simulations.
- We provide a novel extension of the Dirac-based method to SE(2) using Rao–Blackwellization and evaluate it in simulations.

Compared with [1], we have completely rewritten the paper for greater clarity and extended the method to estimation on SE(2).

We use the following notation.

$\delta(\cdot)$	Dirac delta distribution
$\mathbb{1}_M(\cdot)$	indicator function of the set $M$
	$\mathbb{1}_M(x) = 1 \Leftrightarrow x \in M, \mathbb{1}_M(x) = 0 \Leftrightarrow x \notin M$
$P(\cdot)$	probability mass function (discrete)
$f(\cdot)$	probability density function (continuous)
$\mathbf{A}_k(i, j)$	element at column $j$ and row $i$ of matrix $\mathbf{A}_k$

## II. SYSTEM AND MEASUREMENT MODEL

In the following, the system state is assumed to be inside a bounded interval contained in  $\mathbb{R}$ . It is not necessary to distinguish closed, open, or half-open intervals. In the case of a state inside an interval, we assume the subset topology of  $\mathbb{R}$  and for a circular state, we assume the topology of the circle. From now on, we will assume without loss of generality that

the domain is defined as  $[0, 2\pi)$ . Also, we only distinguish these two topologies when it makes a practical difference.

We consider a discrete-time system whose state at time step  $k$  is  $x_k \in [0, 2\pi)$ . The system dynamics is

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

with noise  $w_k \in W$  in the noise space  $W$  and system function  $a_k : [0, 2\pi) \times W \rightarrow [0, 2\pi)$ . In some cases, it is advantageous to use the transition density  $f(x_{k+1}|x_k)$  rather than (1). If the system noise is additive, we can easily derive the transition density from (1) as discussed in [6]. Furthermore, we assume a measurement model

$$z_k = h_k(x_k, v_k), \quad (2)$$

with a measurement  $z_k \in Z$  in the measurement space  $Z$  and noise  $v_k \in V$  in the noise space  $V$ . For certain approaches discussed in the following, we also need the likelihood  $f(z_k|x_k)$ . If the measurement noise is additive, we can derive the likelihood from (2) as described, e.g., in [6].

To simplify some calculations, it can be beneficial to use (weighted) samples as an approximation of the noise densities  $f_k^w(w_k)$  and  $f_k^v(v_k)$ . We can obtain these samples with stochastic or deterministic sampling. When using stochastic sampling, we draw the samples at random from the given distribution. In deterministic sampling, we determine the sample locations based on an optimality criterion. Deterministic sampling is advantageous because usually significantly fewer samples are necessary to achieve the same approximation quality [23, Sec. 4.2]. Deterministic sampling of circular densities can be performed using the methods in [6].

## III. PIECEWISE CONSTANT METHOD

In the following, we present filtering algorithms based on piecewise constant distributions. These filters are essentially Wonham filters [11] in discrete time. All methods discussed in this section are applicable to the topology of both the circle and an interval.

**Definition 1** (Piecewise Constant Distribution). The piecewise constant (PWC) distribution on the interval  $[0, 2\pi)$  has the probability density function

$$\mathcal{PWC}(x; \omega_1, \dots, \omega_L) = \sum_{i=1}^L \omega_i \mathbb{1}_{I_i}(x),$$

where  $L$  is the number of discretization steps, and  $I_j = [I_j^{\min}, I_j^{\max})$ ,  $j = 1, \dots, L$  are intervals and  $\frac{2\pi}{L} \sum_{j=1}^L \omega_j = 1$ .

In the remainder of the paper, we assume equidistant borders

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

It can easily be seen that  $I_1, \dots, I_L$  are a partition of  $[0, 2\pi)$ . For the remainder of the section, we assume that  $x_k$  is PWC-distributed, which we denote by  $x_k \sim \mathcal{PWC}(x; \omega_1, \dots, \omega_n)$ . We can derive the probability associated with an interval  $I_i$  according to

$$P(x_k \in I_i) = \int_{I_i} \mathcal{PWC}(x; \omega_1, \dots, \omega_L) dx_k = \frac{2\pi}{L} \omega_i \quad (3)$$

for  $i = 1, \dots, L$ . Hence,  $\omega_i$  is directly proportional to the probability mass contained in the  $i$ -th interval. Observe that

$$P(x_k \in (I_1 \cup \dots \cup I_L)) = \sum_{i=1}^L P(x_k \in I_i) = \sum_{i=1}^L \frac{2\pi}{L} \omega_i = 1$$

due to the normalization condition.

#### A. Prediction Step

For the prediction step, we start by determining the system matrix  $\mathbf{A}_k(i, j) = P(x_{k+1} \in I_i | x_k \in I_j)$ . We can compute these conditional probabilities as

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

In general, these integrals cannot be evaluated analytically. By approximating  $w_k$  using  $L^w$  samples with locations  $\theta_1^w, \dots, \theta_{L^w}^w$  and weights  $\omega_1^w, \dots, \omega_{L^w}^w$ , the integral can be reformulated as follows

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

In case the state transition density  $f(x_{k+1} | x_k)$  is available, we can use the equation

$$P(x_{k+1} \in I_i | x_k \in I_j) = \int_{I_i} \int_{I_j} f(x_{k+1} | x_k) dx_k dx_{k+1}$$

instead. The resulting integral is still two-dimensional, but numerical integration algorithms tend to perform much better because the integrand does not contain the discontinuous indicator function. Either way, the integral(s) can be evaluated offline if neither the function  $a_k(\cdot, \cdot)$  nor the distribution of the system noise  $f_k^w(w_k)$  are time-variant. Finally, we obtain the predicted PWC density 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

For the measurement update, we can either discretize the measurement space (just as the state space) or compute the likelihood of a given measurement without performing a discretization. Note that discretizing general unbounded measurement spaces using a finite grid is not possible.

1) *Discretized Measurement Space*: In the following, we assume that the measurement space  $Z$  has been subdivided into intervals  $\hat{I}_1, \dots, \hat{I}_m$ . Note that the number of discretization steps for the measurement space can be different from the number used for the state space.

The measurement matrix  $\mathbf{H}_k(i, j) = P(z_k \in \hat{I}_i | x_k \in I_j)$  can be computed according to

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

These integrals can, in general, only be evaluated using numerical methods. If we approximate the noise  $v_k$  using  $L^v$  samples with positions  $\theta_1^v, \dots, \theta_{L^v}^v$  and weights  $\omega_1^v, \dots, \omega_{L^v}^v$ , the integral simplifies to

$$P(z_k \in \hat{I}_i | x_k \in I_j) \approx \sum_{l=1}^{L^v} \omega_l^v \int_{I_j} \mathbb{1}_{\hat{I}_i}(h_k(x_k, \theta_l^v)) dx_k .$$

If we know the likelihood  $f(z_k | x_k)$ , we can alternatively compute the entries of  $\mathbf{H}_k$  according to

$$P(z_k \in \hat{I}_i | x_k \in I_j) = \int_{\hat{I}_i} \int_{I_j} f(z_k | x_k) dx_k dz_k .$$

Just as before, this yields better results when using numerical integration. If the measurement function  $h_k(\cdot, \cdot)$  as well as the measurement noise  $f_k^v(v_k)$  are time-invariant, we can compute these integrals offline.

For a given measurement, we perform the measurement update by determining the interval  $\hat{I}_i$  containing  $z_k$  and choosing the corresponding row of  $\mathbf{H}_k$ . Then, the update is performed with the Bayes' theorem according to

$$P(x_k \in I_j | z_k \in \hat{I}_i) \propto \mathbf{H}_k(i, j) \cdot P(x_k \in I_j) .$$

To complete the measurement update step, we renormalize the density so it integrates to one.

2) *Continuous Measurement Space*: Sometimes it may be undesirable or impossible to discretize the measurement space. Then, we can use Bayes' theorem for a fixed  $z_k$  to obtain

$$\begin{aligned} P(x_k \in I_j | z_k) &= \int_{I_j} f(x_k | z_k) dx_k \\ &\propto \frac{2\pi}{E} \omega_j \int_{I_j} f(z_k | x_k) dx_k \end{aligned}$$

by integrating over the continuous likelihood function. We need to evaluate this integral online in each measurement update step and for every interval  $I_j$  because the result depends on the actual value of  $z_k$ .

## IV. DIRAC-BASED METHOD

Now, we consider the second discretization approach, which is based on a Dirac mixture<sup>1</sup>.

**Definition 2** (Dirac Mixture). A Dirac mixture on the interval  $[0, 2\pi)$  with  $L$  components is defined as

$$\mathcal{D}(x; \theta_1, \dots, \theta_L, \omega_1, \dots, \omega_L) = \sum_{j=1}^L \omega_j \delta(\theta_j - x) ,$$

where  $0 \leq \theta_1 < \dots < \theta_L < 2\pi$  with  $\omega_1, \dots, \omega_L \geq 0$  and  $\sum_{j=1}^L \omega_j = 1$ .

The Dirac-based approach induces the probabilistic interpretation  $P(x_k = \theta_j) = \omega_j$ , i.e., we only consider  $L$  discrete points rather than  $L$  intervals as in (3).

In this paper, we use Dirac mixtures for which the locations of the Dirac components are fixed on an equidistant grid

$$\theta_j = (j - 1/2) \frac{2\pi}{L}, \quad j = 1, \dots, L . \quad (4)$$

Unlike in sequential importance resampling (SIR) particle filters [10], the particles have fixed locations and only their weights are modified. The benefit of immutable sample locations is that we can ensure that the state space is uniformly covered by particles, and thus, inadequate state space coverage cannot occur. Furthermore, the proposed approach is completely deterministic. However, the drawback of an evenly-spaced grid is that highly concentrated densities can only be represented accurately if  $L$  is chosen very large because the grid does not adaptively put more discretization points in areas with large amounts of probability mass.

<sup>1</sup>In the context of periodic state spaces, Dirac mixtures are sometimes called Wrapped Dirac Mixtures (see [6]).

### A. Prediction Step

To perform the prediction, we first propagate each sample through the system function. The propagated density can be obtained as

$$f(x_{k+1}) = \sum_{i=1}^L \omega_i \int_W \delta(x_{k+1} - a_k(\theta_i, w_k)) f^w(w_k) dw_k,$$

which is, in general, a continuous density. If we approximate the noise  $w_k$  by  $L^w$  samples as  $\sum_{j=1}^{L^w} \omega_j^w \delta(x - \theta_j^w)$ , the integral can be simplified to

$$f(x_{k+1}) \approx \sum_{i=1}^L \sum_{j=1}^{L^w} \omega_i \omega_j^w \delta(x_{k+1} - a_k(\theta_i, \theta_j^w)). \quad (5)$$

Observe that the resulting Dirac delta components are not necessarily located on the grid (4) anymore. Now, we discuss how to find samples located on the grid (4) that approximate  $f(x_{k+1})$  as given in (5). We only consider the case with sampled noise in this paper.

1) *Nearest Neighbor Method*: The nearest neighbor approach initializes the weights of all samples with zero. Then, we consider each component in  $f(x_{k+1})$  and determine the grid point closest to  $\phi := a_k(\theta_i, \theta_j^w)$ , i.e.,

$$\arg \min_l \Delta(\theta_l, \phi), \quad (6)$$

where  $\Delta(\cdot, \cdot)$  is a topology-aware distance function. On an interval, a typical distance function is  $\Delta(a, b) = |a - b|$  and in the circular case, we employ

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

i.e., the geodesic distance. Either way, we can use modulo arithmetic to solve (6) in constant time due to the equidistant grid. Then, we increase the weight of the closest sample by  $\omega_i \cdot \omega_j^w$ .

Implementing this approach is straightforward, but it suffers from significant issues. Consider the example  $a_k(x_k, w_k) = x_k + \epsilon$  for any constant  $|\epsilon| < \pi/L$ . Here, the estimate does not change during the prediction step, because each sample gets assigned the same weight as before. A more detailed analysis of this behavior can be found in the supplementary material.

2) *Proportional Method*: In order to solve the aforementioned problem, we propose the proportional method. Instead of just considering the nearest neighbor, we find the two closest neighbors and split the probability mass between them while taking the distance into account.

To achieve this, we need to consider multiple cases. If we have  $\theta_l < \phi \leq \theta_{l+1}$ , we can distribute the weight in relation to the distance. The weight of the component at location  $\theta_l$  is increased by

$$\frac{\Delta(\theta_{l+1}, \phi)}{\Delta(\theta_l, \phi) + \Delta(\theta_{l+1}, \phi)} \omega_i \omega_j^w$$

and the weight of the component at location  $\theta_{l+1}$  is increased by

$$\frac{\Delta(\theta_l, \phi)}{\Delta(\theta_l, \phi) + \Delta(\theta_{l+1}, \phi)} \omega_i \omega_j^w.$$

The sum of both weights is  $\omega_i \cdot \omega_j^w$ , and hence, the total probability mass does not change.

Depending on the topology, we have to deal with some special cases at the borders. On an interval, we consider two scenarios. If  $\phi \leq \theta_1$ , the component at  $\theta_1$  receives the complete

weight  $\omega_i \cdot \omega_j^w$ . Correspondingly, if  $\theta_L < \phi$ , the component at  $\theta_L$  is assigned the entire weight  $\omega_i \cdot \omega_j^w$ .

On the unit circle, we have to deal with these cases differently. For  $\phi \leq \theta_1$  or  $\theta_L < \phi$ , the probability mass is distributed between  $\theta_1$  and  $\theta_L$  using the topology-aware distance. Hence, the component at  $\theta_1$  is assigned the weight

$$\frac{\Delta(\theta_L, \phi)}{\Delta(\theta_1, \phi) + \Delta(\theta_L, \phi)} \omega_i \omega_j^w$$

and the component at  $\theta_L$  is assigned the weight

$$\frac{\Delta(\theta_1, \phi)}{\Delta(\theta_1, \phi) + \Delta(\theta_L, \phi)} \omega_i \omega_j^w.$$

Once again, the propagation of a single Dirac component is possible in constant time in all cases.

### B. Measurement Update

The measurement update is a fairly straightforward operation. We simply apply Bayes' theorem

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

that is, each weight is multiplied with the likelihood. Afterward, the weights are renormalized. Discretizing the measurement space  $Z$  is not necessary here.

## V. COMPARISON

Table I and Table II provide an overview of the proposed prediction and measurement update algorithms, respectively. We consider several cases where we distinguish the interpretation of the density (PWC or Dirac-based), the type of system model as well as the noise distribution (time-variant or time-invariant), whether the noise is continuous or sampled, and whether or not the measurement space has been discretized. In most scenarios, the PWC-based methods are computationally more expensive than their Dirac-based counterparts. Usually, the most expensive step is the numerical integration, i.e., approaches that do not require numerical integration, at least not at runtime, are much more efficient. In particular, numerical integration in 2D is significantly more expensive than numerical integration in 1D. It is also worth mentioning that the space requirements for the PWC method are higher under some circumstances. If the system matrix  $\mathbf{A}_k \in \mathbb{R}^{L \times L}$  is precomputed offline,  $\mathcal{O}(L^2)$  space is required and if the measurement matrix  $\mathbf{H}_k \in \mathbb{R}^{m \times L}$  is precomputed,  $\mathcal{O}(L \cdot m)$  is needed. On the contrary, the Dirac-based method only requires  $\mathcal{O}(L)$  space.

As far as theoretical differences are concerned, we consider the limiting behavior. Proofs of the following claims are given as supplementary material. When  $L$  approaches infinity, the PWC density converges pointwise to the actual density provided  $f(\cdot)$  is piecewise continuous. More formally, it holds for all  $x \in [0, 2\pi)$  where  $f(\cdot)$  is continuous

$$\lim_{L \rightarrow \infty} |f(x) - \mathcal{PWC}(x; \omega_1, \dots, \omega_L)| = 0$$

with  $\omega_i = \int_{I_i} f(x) dx$  for  $i \in \{1, \dots, L\}$ . As the Dirac-based approach does not provide a continuous probability density, a similar statement is not possible. However, for Riemann integrable  $f(\cdot)$ , it is possible to show that

$$\lim_{L \rightarrow \infty} \left| \int_a^b f(x) dx - \int_a^b \mathcal{D}(x; \theta_1, \dots, \theta_L, \omega_1, \dots, \omega_L) dx \right|$$

Scenario			Solution	
density	time-variant	sampled noise	numerical integration	online complexity
PWC	-	-	2D offline	$\mathcal{O}(L^2)$
PWC	✓	-	2D online	$\mathcal{O}(L^2 + L^2 \cdot C_2)$
PWC	-	✓	1D offline	$\mathcal{O}(L^2)$
PWC	✓	✓	1D online	$\mathcal{O}(L^2 + L \cdot C_1)$
Dirac	-	✓	-	$\mathcal{O}(L \cdot L^w)$
Dirac	✓	✓	-	$\mathcal{O}(L \cdot L^w)$

TABLE I: Prediction methods. The cost for one  $d$ -dimensional numerical integration is given by  $C_d$ .

Scenario				Solution	
density	time-variant	sampled noise	meas. space	numerical integration	online complexity
PWC	-	-	discr.	2D offline	$\mathcal{O}(L)$
PWC	✓	-	discr.	2D online	$\mathcal{O}(L + L \cdot C_2)$
PWC	-	✓	discr.	1D offline	$\mathcal{O}(LL^v)$
PWC	✓	✓	discr.	1D online	$\mathcal{O}(LL^v + L \cdot C_1)$
PWC	-	N/A	cont.	1D online	$\mathcal{O}(L + L \cdot C_1)$
PWC	✓	N/A	cont.	1D online	$\mathcal{O}(L + L \cdot C_1)$
Dirac	✓/-	N/A	cont.	-	$\mathcal{O}(L)$

TABLE II: Measurement update methods. The cost for one  $d$ -dimensional numerical integration is given by  $C_d$ .

is equal to zero for any fixed  $a, b \in [0, 2\pi)$  with  $a < b$ , where  $\omega_i = \frac{f(\theta_i)}{\sum_{k=1}^L f(\theta_k)}$  and  $\theta_i$  is given by (4) for all  $i \in \{1, \dots, L\}$ . The intuition behind this statement is that the probability mass contained in the interval  $[a, b]$  approaches the true probability mass in that interval for an increasing number of discretization points.

## VI. EVALUATION

For evaluation purposes, we consider a nonlinear circular scenario. Note that we use slightly different parameters compared with [1]. As the system and the measurement model, we use the function

$$h_c(x) = \pi \cdot \left( 1 + \sin \left( \frac{|x-\pi|^c}{\pi^{c-1}} \cdot \frac{\text{sign}(x-\pi)}{2} \right) \right) \quad (7)$$

for  $x \in [0, 2\pi)$  and  $c \in \mathbb{R}_+$ , which was also considered in [24]. The function (7) is a continuous bijection on  $[0, 2\pi)$ , whose nonlinearity can be controlled using the parameter  $c$  (see Fig. 2).

### A. Prediction Step

To evaluate the prediction step, we assume that  $x_k$  follows a wrapped normal [25, Sec. 2.2.6] distribution

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

with parameters  $\mu = 2$  and  $\sigma = 1$ . We denote this density by  $\mathcal{WN}(2, 1)$ . The system dynamics is

$$x_{k+1} = h_c(x_k) + w_k \quad \text{mod } 2\pi,$$

where the system noise is  $w_k \sim \mathcal{WN}(0, 0.5)$ . We perform one prediction step using each method and compare the result with the ground truth using two distance measures.

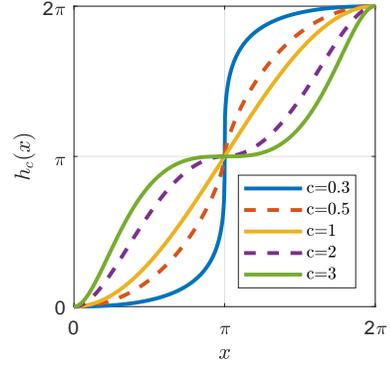


Fig. 2: Our evaluation makes use of the nonlinear function  $h_c(\cdot)$  whose behavior can be modified by varying  $c$ .

The first error measure consists in the difference with respect to the first trigonometric moment. For a random variable  $x$ , the first trigonometric moment is given by the complex number  $\mathbb{E}(e^{ix}) \in \mathbb{C}$  and describes the location as well as the dispersion of the distribution [26, Sec. 2.4]. Consequently, the error measure is defined as

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

where  $i \in \mathbb{C}$  is the imaginary unit and the norm  $|\cdot|$  is defined as  $|a + bi| = \sqrt{a^2 + b^2}$ . Numerical integration [27] of the exact Bayesian filtering equations was used to compute the true trigonometric moment.

As a second error measure, we consider the Kullback–Leibler divergence (KLD) [28, Sec. 1.3]

$$\text{KLD}(f^{\text{true}} || f^{\text{result}}) = \int_0^{2\pi} f^{\text{true}}(x) \cdot \log \left( \frac{f^{\text{true}}(x)}{f^{\text{result}}(x)} \right) dx$$

between the exact predicted density  $f^{\text{true}}$  and the density predicted by the filter  $f^{\text{result}}$ . In this case, the ground truth is computed using the Fourier square root filter [29] with a very large number of  $L = 301$  coefficients. As the KLD is only defined for continuous densities, the result obtained from the Dirac-based approach is reinterpreted as a PWC density to allow computation of the KLD. We do not consider the KLD for the particle filter because it is nontrivial to obtain a continuous density based on the estimate given as particles.

Our evaluation compares the wrapped normal filter [6, Sec. VI], the Dirac-based filter (see Sec. IV), the piecewise constant filter (see Sec. III), and an SIR particle filter (as discussed in [6, Sec. II-B-2]). The PWC filter requires the system matrix  $\mathbf{A}$ , which we computed with one-dimensional numerical integration using sampled noise. The noise was approximated using five weighted samples obtained using [6, Algorithm 2]. For the Dirac-based filter, the proportional method was chosen.

Fig. 3 and Fig. 4 show the results of our evaluation. It is obvious that the Dirac-based filter with  $L = 60$  has the best performance for most values of  $c$  according to both error measures. For small values of  $c$ , the PWC filter with  $L = 60$  performs slightly better when considering the trigonometric moment. However, it is inferior otherwise even though it is much more expensive in terms of computation. The Dirac-based filter with  $L = 15$  cannot match the filters with  $L = 60$ ,

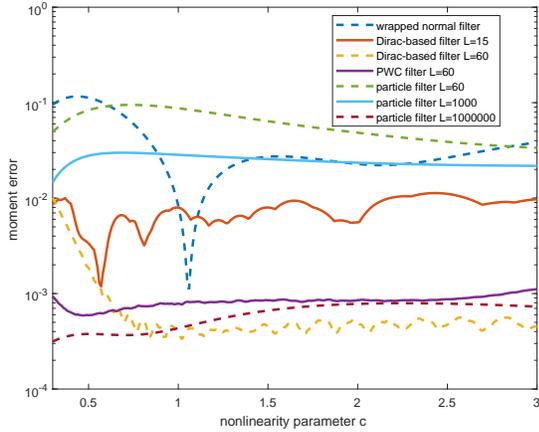


Fig. 3: Norm of the error in the first trigonometric moment after a single prediction step.

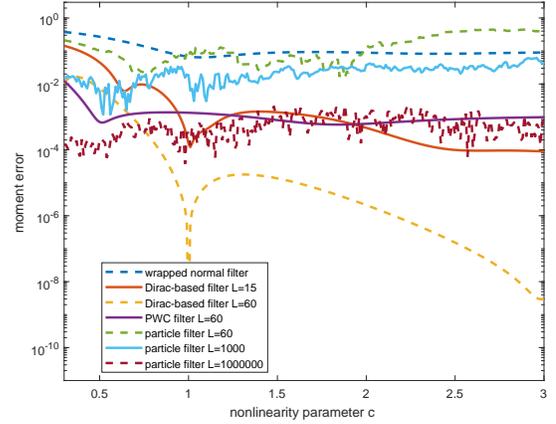


Fig. 5: Norm of the error in the first trigonometric moment after a single measurement update step.

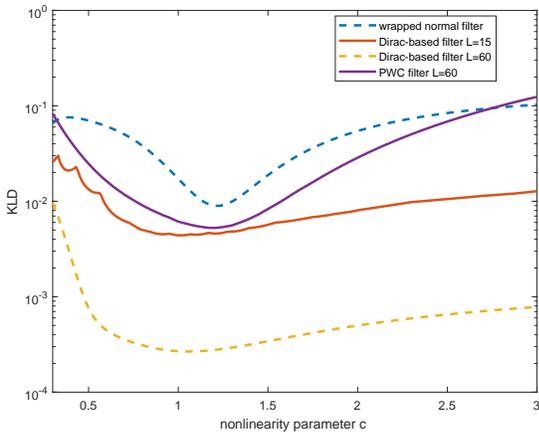


Fig. 4: Kullback–Leibler divergence from the ground truth to the density after a single prediction step.

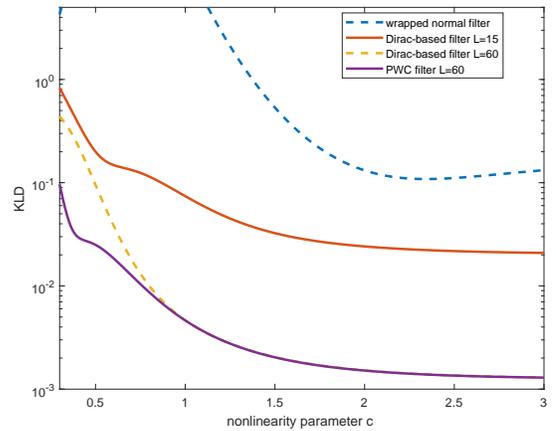


Fig. 6: Kullback–Leibler divergence from the ground truth to the density after a single measurement update step.

but still yields decent results. The WN filter does not perform very well as it only uses  $L = 5$  samples (see [6]). Furthermore, the results of the circular particle filters are quite unsatisfactory. Even for  $L = 1000$  particles, the performance is significantly worse than that of the PWC filter and the Dirac-based approach with  $L = 15$  and  $L = 60$ . For  $L = 1000000$ , the error in the first trigonometric moment is similar to the PWC and Dirac-based approaches. This example shows the benefits of grid-based approaches with particles at fixed locations.

### B. Measurement Update

In the evaluation of the measurement update, we assume the prior density  $\mathcal{WN}(2, 1)$  and we use the measurement model

$$z_k = h_c(x_k) + v_k \pmod{2\pi},$$

with the measurement noise  $v_k \sim \mathcal{WN}(0, 0.5)$ . We consider the same filters and error measures as before and perform a single measurement update with the measurement  $z_k = 4$ . Then, we compute the first trigonometric moment of the posterior distribution and compare it with the true trigonometric

moment. For the KLD-based evaluation, we use Bayes' theorem to obtain the true posterior

$$f(x_k|z_k) = \frac{f(z_k|x_k)f(x_k)}{\int_0^{2\pi} f(z_k|x_k)f(x_k) dx_k}.$$

The numerator is calculated directly and the denominator is computed using numerical integration [27].

To keep the comparison as fair as possible, we avoid introducing an additional discretization error by using a continuous measurement space for the PWC filter (see Sec. III-B2), even though we have to perform numerical integration online.

We depict the results of the evaluation in Fig. 5 and Fig. 6. Obviously, the Dirac-based filter with  $L = 60$  yields the lowest error in most cases and even significantly outperforms the PWC filter with regards to the first trigonometric moment while having a lower computational cost. No other approach can match this accuracy except for small  $c$ . Once again, the particle filter is a lot worse than the Dirac-based approach, even with a much larger number of samples, especially for large values of the nonlinearity parameter  $c$ . In the KLD-based evaluation, the PWC approach performs best, but the Dirac-based method is almost identical for large values of the nonlinearity parameter  $c$ .

## VII. RAO-BLACKWELLIZED SE2 FILTER

The Dirac-based filter presented before can be used to derive a filter on SE(2), the group of rigid body motions in  $\mathbb{R}^2$ . A rigid body motion can be parameterized by a translation  $\underline{x}^l \in \mathbb{R}^2$  and a rotation  $x^c \in [0, 2\pi)$ . Their joint density is given by

$$f(\underline{x}^l, x^c) = f(\underline{x}^l | x^c) f(x^c)$$

according to the definition of conditional probability densities. The key idea consists in factoring the joint density in this way, and thus, separating the translation (for a fixed rotation) from the rotation. This technique is referred to as Rao-Blackwellization and has also been successfully applied in various other applications such as simultaneous localization and mapping (SLAM) [30]. Most existing approaches based on Rao-Blackwellization rely on a particle filter for the discrete part of the state space [31], but we use the Dirac-based filter from Sec. IV instead, somewhat similar to [32], [33].

In the following, we assume  $f(\underline{x}^l | x^c)$  to be a Gaussian density<sup>2</sup> for fixed  $x^c$  and  $f(x^c)$  to be a Dirac mixture density. As a result, we can parameterize the joint density as a set of  $L$  tuples  $(\theta_i, \omega_i, \underline{\mu}_k^i, \mathbf{C}_k^i)$  for  $i \in \{1, \dots, L\}$ , where  $\theta_i$  and  $\omega_i$  correspond to the location and weight of the  $i$ -th Dirac component and  $\underline{\mu}_k^i, \mathbf{C}_k^i$  are the mean and the covariance of the corresponding Gaussian.

### A. Prediction Step

In the translatory part, we assume a linear system model<sup>3</sup> given  $x_k^c$ . The model is given by

$$\begin{bmatrix} \underline{x}_{k+1}^l \\ x_{k+1}^c \end{bmatrix} = \begin{bmatrix} \mathbf{A}(x_k^c) \underline{x}_k^l + \mathbf{B}(x_k^c) \underline{u}_k + \underline{w}_k^l \\ a_k(x_k^c, \underline{u}_k, \underline{w}_k^c) \end{bmatrix}, \quad (8)$$

where  $\mathbf{A}(x_k^c) \in \mathbb{R}^{2 \times 2}$  is a matrix depending on  $x_k^c$ ,  $\underline{w}_k^l$  is zero-mean Gaussian noise with covariance  $\mathbf{C}_k^w$ , and  $\underline{w}_k^c \in W$  is noise affecting the circular part given as a set of weighted samples. The matrix  $\mathbf{B}(x^c) \in \mathbb{R}^{2 \times r}$  is the input matrix depending on  $x^c$  and  $\underline{u}_k \in \mathbb{R}^r$  is the deterministic control input. For the sake of simplicity, we assume here that  $a_k(\cdot, \cdot, \cdot) : [0, 2\pi) \times \mathbb{R}^r \times W \rightarrow [0, 2\pi)$  does not depend on  $\underline{x}_k^l$ .

Pseudo code for the prediction step is given in Algorithm 1. Essentially, we perform a Kalman filter prediction step for the Gaussian distribution associated with each particle according to

$$\begin{aligned} \underline{\mu}_{k+1}^i &= \mathbf{A}(\theta_k^i) \underline{\mu}_k^i + \mathbf{B}(\theta_k^i) \underline{u}_k \\ \mathbf{C}_{k+1}^i &= \mathbf{A}(\theta_k^i) \mathbf{C}_k^i \mathbf{A}(\theta_k^i)^T + \mathbf{C}_k^w. \end{aligned}$$

Then, the circular part is predicted according to (5) using the noise samples  $\sum_{j=1}^L \omega_k^{j,w} \delta(x - \theta_k^{j,w})$ . Because of the advantages discussed in Sec. IV-A, we use the proportional assignment method.

Assigning weight from one Dirac component to another is, however, nontrivial in this case because we also have to

<sup>2</sup>In principle, this method can be generalized to allow other parametric distributions such as Gaussian mixtures, but we limit ourselves to Gaussians in this paper.

<sup>3</sup>If the system model is nonlinear, we can employ an EKF, UKF [34], or similar Gaussian-assumed filters instead of the Kalman filter.

### Algorithm 1: SE(2) Prediction

---

**Input:**  $(\theta_k^i, \omega_k^i, \underline{\mu}_k^i, \mathbf{C}_k^i)$  for  $1 \leq i \leq L$ , control input  $\underline{u}_k$   
**Output:**  $(\theta_{k+1}^i, \omega_{k+1}^i, \underline{\mu}_{k+1}^i, \mathbf{C}_{k+1}^i)$  for  $1 \leq i \leq L$

```

for  $i \leftarrow 1$  to  $L$  do
  /* Initialize new particles */
   $\omega_{k+1}^i = 0$ ;
   $\theta_{k+1}^i = \theta_k^i$ ;
  /* Kalman filter prediction */
   $\underline{\mu}_{k+1}^i = \mathbf{A}(\theta_k^i) \underline{\mu}_k^i + \mathbf{B}(\theta_k^i) \underline{u}_k$ ;
   $\mathbf{C}_{k+1}^i = \mathbf{A}(\theta_k^i) \mathbf{C}_k^i \mathbf{A}(\theta_k^i)^T + \mathbf{C}_k^w$ ;
end
 $\mathbf{W} \leftarrow \mathbf{0}_{L \times L}$ ;
for  $i \leftarrow 1$  to  $L$  do
  for  $j \leftarrow 1$  to  $L^w$  do
    /* Discrete prediction */
     $\phi \leftarrow a_k(\theta_k^i, \underline{u}_k, \theta_k^{j,w})$ ;
     $l_1 \leftarrow \arg \min_{l_1} \Delta(\phi, \theta_{k+1}^{l_1})$ ;
     $l_2 \leftarrow \arg \min_{l_2, l_2 \neq l_1} \Delta(\phi, \theta_{k+1}^{l_2})$ ;
     $w_1 \leftarrow \frac{\Delta(\theta_{k+1}^{l_2}, \phi)}{\Delta(\theta_{k+1}^{l_1}, \phi) + \Delta(\theta_{k+1}^{l_2}, \phi)} \omega_k^i \omega_k^{j,w}$ ;
     $w_2 \leftarrow \frac{\Delta(\theta_{k+1}^{l_1}, \phi)}{\Delta(\theta_{k+1}^{l_1}, \phi) + \Delta(\theta_{k+1}^{l_2}, \phi)} \omega_k^i \omega_k^{j,w}$ ;
     $\omega_{k+1}^{l_1} \leftarrow \omega_{k+1}^{l_1} + w_1$ ;
     $\omega_{k+1}^{l_2} \leftarrow \omega_{k+1}^{l_2} + w_2$ ;
     $\mathbf{W}(i, l_1) \leftarrow \mathbf{W}(i, l_1) + w_1$ ;
     $\mathbf{W}(i, l_2) \leftarrow \mathbf{W}(i, l_2) + w_2$ ;
  end
end
for  $i \leftarrow 1$  to  $L$  do
  /* Reduce Gaussian mixtures */
   $\underline{\mu}_{k+1}^i \leftarrow \mathbb{E} \left( \sum_{j=1}^L \frac{\mathbf{W}(j,i)}{\sum_{r=1}^L \mathbf{W}(r,i)} \mathcal{N}(\underline{\mu}_{k+1}^j, \mathbf{C}_{k+1}^j) \right)$ ;
   $\mathbf{C}_{k+1}^i \leftarrow \text{Cov} \left( \sum_{j=1}^L \frac{\mathbf{W}(j,i)}{\sum_{r=1}^L \mathbf{W}(r,i)} \mathcal{N}(\underline{\mu}_{k+1}^j, \mathbf{C}_{k+1}^j) \right)$ ;
end

```

---

consider the associated Gaussian density<sup>4</sup>. For this reason, we introduce the matrix  $\mathbf{W}$  whose entries  $\mathbf{W}(i, j)$  quantify how much of the Gaussian component associated with  $\theta_k^i$  has been assigned to  $\theta_{k+1}^j$ . As a result, the continuous density associated with a particle  $\theta_{k+1}^i$  is now the Gaussian mixture  $\sum_{j=1}^L \frac{\mathbf{W}(j,i)}{\sum_{r=1}^L \mathbf{W}(r,i)} \mathcal{N}(\underline{\mu}_{k+1}^j, \mathbf{C}_{k+1}^j)$ . In order to avoid an increasing complexity of the density in each time step, we reapproximate each Gaussian mixture with a single Gaussian by computing the mixture's mean and covariance.

### B. Update Step

For the measurement equation, we assume a linear measurement model<sup>5</sup> for given  $x_k^c$

$$\underline{z}_k = \mathbf{H}(x_k^c) \underline{x}_k^l + \underline{v}_k, \quad (9)$$

<sup>4</sup>Note that this step is not required in Rao-Blackwellized particle filters because they never assign probability mass from one particle to another.

<sup>5</sup>If the measurement model is nonlinear, we can employ an EKF, UKF [34], or similar Gaussian-assumed filters instead of the Kalman filter. Note that we require the likelihood function as well, which can easily be derived in the case of additive noise.

where  $\underline{v}_k$  is additive zero-mean Gaussian noise with covariance  $\mathbf{C}_k^v$ . To perform the update, we apply the Kalman filter to update the Gaussian part for each particle

$$\begin{aligned} \mathbf{K}_k^i &= \mathbf{C}_k^i \mathbf{H}(\theta_k^i)^T (\mathbf{H}(\theta_k^i) \mathbf{C}_k^i \mathbf{H}(\theta_k^i)^T + \mathbf{C}_k^v)^{-1}, \\ \underline{\mu}_k^i &\leftarrow \underline{\mu}_k^i + \mathbf{K}_k^i \left( \underline{z}_k - \mathbf{H}(\theta_k^i) \underline{\mu}_k^i \right), \\ \mathbf{C}_k^i &\leftarrow (\mathbf{I} - \mathbf{K}_k^i \mathbf{H}(\theta_k^i)) \mathbf{C}_k^i. \end{aligned}$$

Afterward, we re-weight the Dirac delta components corresponding to the circular part using the likelihood

$$f(z_k | \theta_k^i) = \mathcal{N}(z_k; \mathbf{H}(\theta_k^i) \underline{\mu}_k^i, \mathbf{C}_k^v)$$

as discussed in Sec. IV-B.

### C. Evaluation

We evaluate the Rao–Blackwellized discrete filter on SE(2) in the following scenario. A vehicle moves in the plane and obtains position-only measurements, e.g., via GPS. We assume a two-wheel vehicle with bicycle dynamics [35, Sec. 3.1.5].

The system model has two inputs, the velocity of the vehicle  $v_{H,k}$  and the steering angle  $\alpha_k$ , i.e.,

$$\underline{u}_k = [v_{H,k}, \alpha_k]^T.$$

The dynamics are given by (8), where

$$\begin{aligned} \mathbf{A}(\theta_i) &= \mathbf{I}, \\ \mathbf{B}(\theta_i) &= \begin{bmatrix} T \cos(\theta_i) & 0 \\ T \sin(\theta_i) & 0 \end{bmatrix}, \end{aligned}$$

$$a_k(x_k^c, \underline{u}_k, w_k^c) = x_k^c + T \cdot v_{H,k} / \lambda \cdot \tan \alpha_k + w_k^c \pmod{2\pi}.$$

Here, the constants  $T$  and  $\lambda$  describe the duration of a time step and the length of the vehicle, respectively. In our evaluation, we use the values  $\lambda = 1$  m,  $T = 1$  s.

Furthermore, the measurement model is given by (9), where  $\mathbf{H}(\theta_i) = \mathbf{I}$  is the identity matrix, i.e., the position is measured directly, but the orientation is not contained in the measurement. The parameters were chosen as follows. For the system noise, we have  $\mathbf{C}_k^w = 0.01 \cdot \mathbf{I}_{2 \times 2} \text{m}^2$ ,  $w_k^c \sim \mathcal{WN}(0, 0.1)$  and for the measurement noise, we have  $\mathbf{C}_k^v = 10 \cdot \mathbf{I}_{2 \times 2} \text{m}^2$ . The input is  $\underline{u}_k = [1 \text{ m s}^{-1}, 0.1 \text{ rad}]^T$  at all 30 time steps.

We compare our approach with the SIR particle filter [10]. Our approach uses  $L = 20$  discrete points and the particle filter uses 200 particles, which results in a comparable computational effort. We evaluated the accuracy in terms of both position and orientation. For the position, we consider the mean Euclidean error, and for the orientation, we consider the mean angular error as given by  $\Delta(\cdot)$ . Both filters are initialized with a circular uniform distribution for  $x^c$  and a very high uncertainty  $\underline{x}^l \sim \mathcal{N}(\underline{0}, 100 \cdot \mathbf{I}_{2 \times 2})$ .

The results from 100 Monte Carlo runs are depicted in Fig. 7. New instances of the system noise and the measurement noise are sampled in each run. We observe that the proposed approach significantly outperforms the particle filter even though it uses fewer discrete points. The figure also depicts the expected error according to the square root of the trace of the covariance matrix and the expected angular error. It can be seen that the proposed filter is conservative, whereas the particle filter

	proposed	SIRPF
Average time for prediction step	3.7 ms	5.6 ms
Average time for update step	1.8 ms	0.4 ms
Total	5.5 ms	6.0 ms

TABLE III: Runtime for the proposed filter and the particle filter.

sometimes tends to underestimate its uncertainty. The runtime on a laptop with an Intel Core i7-2640M and 8 GB of RAM using MATLAB 2017a is shown in Table III. It can be seen that both approaches require a very similar amount of time.

## VIII. CONCLUSION

Two approaches for estimation on intervals and on the unit circle have been proposed, namely the piecewise constant and the Dirac-based method. Our evaluation suggests that the Dirac-based approach usually yields better results and is also faster to compute. Furthermore, we have shown how to extend the Dirac-based approach to estimation of rigid body motions in SE(2) using Rao–Blackwellization.

Future work may consider a generalization of the proposed estimation algorithms to higher-dimensional manifolds such as the torus or the sphere. Due to the exponential growth of the number of points in an evenly spaced grid, we expect these approaches to be limited to a fairly small number of dimensions unless an adaptive non-uniform discretization scheme is employed.

MATLAB implementations of the proposed filters are freely available to other researchers as part of libDirectional [36], a library for directional estimation. This library also contains implementations of all reference methods used in the evaluation section.

## ACKNOWLEDGMENT

The IGF project 18798 N of the research association Forschungs-Gesellschaft Verfahrens-Technik e.V. (GVT) was supported via the AiF in a program to promote the Industrial Community Research and Development (IGF) by the Federal Ministry for Economic Affairs and Energy on the basis of a resolution of the German Bundestag. This work was also supported by the German Research Foundation (DFG) under grant HA 3789/13-1.

## REFERENCES

- [1] G. Kurz, F. Pfaff, and U. D. Hanebeck, “Discrete Recursive Bayesian Filtering on Intervals and the Unit Circle,” in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sep. 2016.
- [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, J. Choquel, and M. Benjelloun, “Circular Data Processing Tools Applied to a Phase Open Loop Architecture for Multi-Channels Signals Tracking,” in *IEEE/ION Position Location and Navigation Symposium (PLANS 2012)*, 2012, pp. 633–642.
- [4] J. Traa and P. Smaragdīs, “A Wrapped Kalman Filter for Azimuthal Speaker Tracking,” *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1257–1260, 2013.

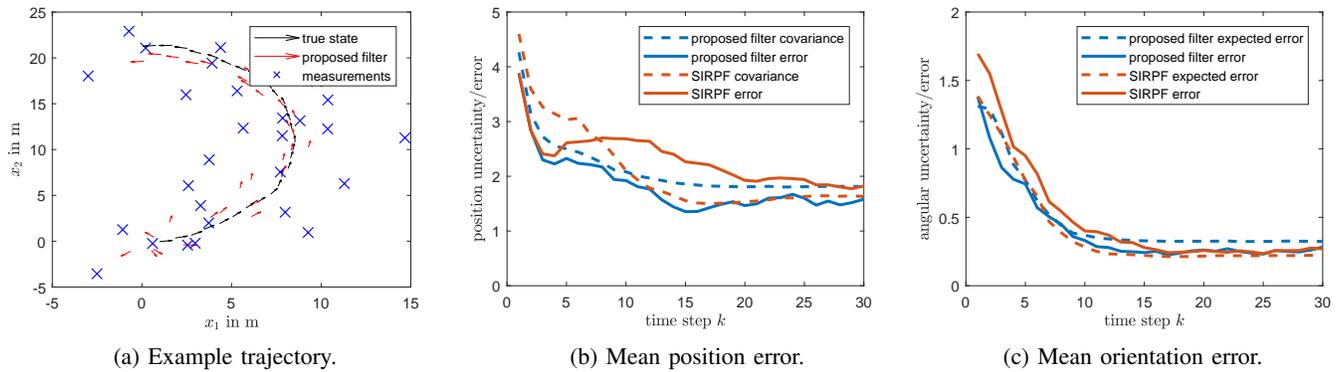


Fig. 7: Evaluation results for the SE(2) scenario. We compare the proposed approach to an SIR particle filter. In addition to the mean error, we also show the error expected according to the uncertainty of the estimate.

- [5] A. Martins, A. Carvalho, and J. Sousa, “Comparing Wind Generation Profiles: A Circular Data Approach,” in *12th International Conference on the European Energy Market (EEM)*, 2015.
- [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] S. C. Kramer and H. W. Sorenson, “Recursive Bayesian Estimation Using Piece-wise Constant Approximations,” *Automatica*, vol. 24, no. 6, pp. 789–801, 1988.
- [8] R. S. Bucy and K. D. Senne, “Digital Synthesis of Non-linear Filters,” *Automatica*, vol. 7, no. 3, pp. 287–298, 1971.
- [9] N. Bergman, “Recursive Bayesian Estimation: Navigation and Tracking Applications,” Ph.D. dissertation, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 1999.
- [10] A. Doucet and A. M. Johansen, “A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later,” *Handbook of Nonlinear Filtering*, vol. 12, pp. 656–704, 2009.
- [11] 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.
- [12] 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.
- [13] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents Series. The MIT Press, 2005.
- [14] 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.
- [15] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.
- [16] I. Markovic, J. Cestic, 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, pp. 2245–2250.
- [17] 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.
- [18] F. Pfaff, G. Kurz, and U. D. Hanebeck, “Nonlinear Prediction for Circular Filtering Using Fourier Series,” in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, Jul. 2016.
- [19] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, “A Stochastic Filter for Planar Rigid-Body Motions,” in *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015)*, San Diego, California, USA, Sep. 2015.
- [20] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, “The Partially Wrapped Normal Distribution for SE(2) Estimation,” in *Proceedings of the 2014 IEEE International Conference on Multisensor Fusion and Information Integration (MFI 2014)*, Beijing, China, Sep. 2014.
- [21] J. E. Darling and K. J. DeMars, “Rigid Body Attitude Uncertainty Propagation Using the Gauss–Bingham Distribution,” in *25th AAS/AIAA Space Flight Mechanics Meeting*, Williamsburg, Virginia, USA, Jan. 2015.
- [22] W. Feiten, M. Lang, and S. Hirche, “Rigid Motion Estimation using Mixtures of Projected Gaussians,” in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, 2013.
- [23] I. Gilitschenski, “Deterministic Sampling for Nonlinear Dynamic State Estimation,” Ph.D. dissertation, Karlsruhe Institute of Technology, Intelligent Sensor-Actuator-Systems Laboratory, 2015.
- [24] 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.
- [25] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*. World Scientific, 2001.
- [26] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Baffins Lane, Chichester, West Sussex, England: John Wiley & Sons, 1999.
- [27] L. F. Shampine, “Vectorized Adaptive Quadrature in Matlab,” *Journal of Computational and Applied Mathematics*, vol. 211, no. 2, pp. 131–140, 2008.
- [28] S. Kullback, *Information Theory and Statistics*. Dover, 1978.
- [29] 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.
- [30] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, “FastSLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges,” in *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, 2003, pp. 1151–1156.
- [31] T. Schön, F. Gustafsson, and P.-J. Nordlund, “Marginalized Particle Filters for Mixed Linear/Nonlinear State-Space Models,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.
- [32] V. Šmídl and M. Gašperin, “Rao–Blackwellized Point Mass Filter for Reliable State Estimation,” in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, 2013, pp. 312–318.
- [33] M. Lindfors, G. Hendeby, F. Gustafsson, and R. Karlsson, “Vehicle Speed Tracking Using Chassis Vibrations,” in *Intelligent Vehicles Symposium (IV), 2016 IEEE*. IEEE, 2016, pp. 214–219.
- [34] 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.
- [35] G. Dudek and M. Jenkin, *Computational Principles of Mobile Robotics*, 2nd ed. Cambridge University Press, 2010.
- [36] G. Kurz, I. Gilitschenski, F. Pfaff, and L. Drude, “libDirectional,” 2015. [Online]. Available: <https://github.com/libDirectional>