Fourier Filters, Grid Filters, and the Fourier-Interpreted Grid Filter

Florian Pfaff, **Kailai Li**, and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
florian.pfaff@kit.edu, kailai.li@kit.edu, uwe.hanebeck@kit.edu

***Abstract*—While estimation problems on periodic manifolds are inherently nonlinear, filters can benefit from the boundedness of the domain. In this paper, we describe how densities on the unit circle are represented in four recently proposed filters that take advantage of the finite size of the domain. Specifically, we address representations based on trigonometric polynomials, piece-wise constant functions, and probability mass functions. Further, we propose a novel filter based on function values on a grid. To provide a continuous density function, we propose two interpolations. The first involves a trigonometric polynomial with the function values at the grid points. The second ensures a nonnegative approximation of the density by interpolating the square roots and then squaring the interpolation. Prediction and update steps are provided for the filter, both of which are independent of the interpolation employed. In an evaluation of the accuracy of the resulting probability density, the newly proposed filter outperforms the filter based on piece-wise constant functions and achieves results comparable to those of the previously proposed filters based on trigonometric polynomials.**

***Index Terms*—Directional estimation, Fourier series, grid filter**

I. Introduction

Recursive Bayesian estimation is a wide field covering estimation problems of various degrees of difficulty. While the general nonlinear estimation problem remains a great challenge, an increasing number of estimation problems in real-world applications can be handled with success. One class of problems involves the estimation of periodic quantities, in particular angles. Such estimation problems have the underlying topology of the unit circle and arise, e.g., in speaker tracking [1], phase estimation [2], and when tracking the orientation of objects on a conveyor belt [3, Ch. 7].

Due to the periodic nature of the estimation problem, a simple Kalman filter should not be used. For example, consider an update step in which a prior density is fused with a likelihood function that is equal to the prior density shifted by π. In this case, the information of the prior density and the likelihood is contradictory, and thus, a uniform distribution on the circle is the correct solution. In contrast, a Kalman filter would yield a unimodal density whose probability mass is more concentrated around its mode than that of the prior density and the likelihood. During the last decade, filters based on densities on the unit circle that properly account for the specifics of the underlying domain have been introduced [1], [4]. Even nonlinear variants that involve deterministic sampling schemes in their prediction and update steps have been proposed [5]. However, such filters are always limited by the range of possible densities that can be accurately represented. For example, two commonly used densities—the von Mises density and the wrapped normal density—are unimodal and symmetric around the mode.

In order to allow for more accurate estimation results, some filters have recently been proposed that are not limited to specific parametric densities and instead use the boundedness of the domain to their advantage. Two such filters are the Fourier filters [6], [7], in which the densities (or their square roots) are represented by trigonometric polynomials, i.e., Fourier series with finite numbers of nonzero coefficients. Approximating the square roots allows for obtaining approximations of the densities with only nonnegative function values. This desirable feature comes at the expense of a more complicated prediction step. Besides the Fourier filters, grid-based approaches were proposed [8]. More details on these approaches are provided in Sec. II.

In this paper, we provide an amalgamation of the grid-based approaches and the Fourier filters. Our aim is not to provide a filter that outperforms the Fourier filters but rather to provide a novel grid filter that integrates the ideas of the Fourier filters to yield results comparable to those of the Fourier filters with comparable run times. The newly proposed filter is based on function values on a grid. To evaluate the density, the function values are interpolated using trigonometric functions up to a certain degree. As we show in Sec. III, the interpolation can be easily changed to ensure the nonnegativity of the approximation. Unlike for the Fourier filters, no changes to the prediction and update steps are required to obtain nonnegative approximations. The quality of the density approximations is evaluated in Sec. IV and a conclusion and an outlook are provided in Sec. V.

II. Fourier and Grid Filters

In this section, we briefly describe how densities are represented in specific filters for the unit circle. In our papers, each point on the unit circle is described by a value in the half-open interval $[0, 2\pi)$. For all filters considered, we illustrate approximations of the wrapped normal density

$$f_{\mathrm{WN}}(x; \mu, \sigma) = \sum_{j \in \mathbb{Z}} \mathcal{N}\big(x + 2\pi j; \mu, \sigma\big) \tag{1}$$

with $\mu = \frac{\pi}{2}$ and $\sigma = 1$ as an example in Fig. 1.

In [6], [7], we introduced two filters using trigonometric polynomials. In the Fourier identity filter (IFF), the density is directly represented using a trigonometric polynomial. Based on a complex trigonometric polynomial with coefficients from $-k_{\text{max}}$ to k_{max}, we obtain the formula

$$f_{\text{IFF}}(x; \underline{c}^{\text{id}}) = \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} c_k^{\text{id}} \mathrm{e}^{ikx} \tag{2}$$

for the probability density function (pdf), in which i denotes the imaginary unit. The approximation is parameterized by a Fourier coefficient vector $\underline{c}^{\text{id}}$ that can either be obtained via explicit formulae as presented in [6] or by applying a transformation to the function values on a grid with $L = 2k_{\text{max}} + 1$ equidistant grid points. A fast transformation from function values to Fourier coefficients can be realized in $O(L \log L)$ using algorithms such as the one proposed in [9]. This algorithm and related algorithms have since become popular for calculating the closely related discrete Fourier transform [10, Ch. 7] as the fast Fourier transform (FFT) [10, Ch. 8]. In our papers, we say the FFT directly yields the desired Fourier coefficient vector.

Irrespective of how the coefficients are determined, the approximation can have negative function values, which is undesirable for approximations of densities. An example of an approximation with negative function values is depicted in Fig. 1a. As shown in Fig. 1b, a nonnegative approximation of the considered wrapped normal density is attained using five coefficients. However, large numbers of coefficients may be required for other densities with regions closer to zero.

The Fourier square root filter (SqFF) is based on approximating the square root of the density using a trigonometric polynomial with coefficient vector $\underline{c}^{\text{sqrt}}$. The values of the approximation of the pdf are obtained via

$$f_{\text{SqFF}}(x; \underline{c}^{\text{sqrt}}) = \left(\sum_{k=-k_{\text{max}}}^{k_{\text{max}}} c_k^{\text{sqrt}} \mathrm{e}^{ikx} \right)^2 . \tag{3}$$

It is always ensured that $\underline{c}^{\text{sqrt}}$ describes a real-valued function, and thus, the square of the trigonometric polynomial yields only nonnegative values even for small numbers of coefficients. Thus, the negative function values observed in Fig. 1a can be prevented in the approximation of the SqFF shown in Fig. 1c. As can be seen when comparing these two figures and the Figs. 1b and 1d, approximating the square root of the density does not necessarily lead to worse approximations. Based on our experience with the Fourier filters, the convergence speeds of the approximations of the IFF and SqFF to the true density are often comparable. However, in the prediction step described in [6], [11], transformations in the form of an inverse FFT and an FFT are required in each step. These additional operations increase the theoretical complexity and run time of the algorithm.

Two representations based on discretization were proposed in [8]. The first is based on piece-wise constant (PWC) densities, which are illustrated in Figs. 1e and 1f. In this representation, the $[0, 2\pi)$ domain is subdivided into L equal-sized grid cells. The probability density of the approximation stays constant within each grid cell. Filters using such approximations of the densities are also referred to as histogram filters [12, Sec. 4.1] because the plot of a one-dimensional density resembles a histogram if the edges of the grid cells are shown using vertical lines. The L intervals describing the grid cells are defined as

$$I_j = \left[(j-1)\frac{2\pi}{L},\ j\frac{2\pi}{L} \right) \text{ for } j \in \{1, \ldots, L\}$$

in the circular case. The density is then parameterized by the vector $\underline{\gamma}$ that describes the function values of the approximation in all intervals. Using $\mathbb{1}_{x \in I_j}$ as the indicator function that returns 1 if $x \in I_j$ and 0 otherwise, the corresponding density function can be written as

$$f_{\text{PWC}}(x; \underline{\gamma}) = \sum_{j=1}^{L} \gamma_j \mathbb{1}_{x \in I_j} .$$

Because modulo arithmetic can be employed to determine the interval in which x is contained, the evaluation of the pdf is always in $O(1)$. As the integral over each interval I_j of size $2\pi/L$ is $\gamma_j 2\pi/L$, the density can be normalized to integrate to one by ensuring that $\sum_{j=1}^{L} \gamma_j = \frac{L}{2\pi}$ holds.

The second grid-based representation explained in [8] is the one used by the so-called discrete filter. For the discrete filter, the continuous probability density function is converted into a discrete probability mass function (pmf) that is only defined on the grid points. The underlying domain is thus no longer continuous but only contains L points in the interval $[0, 2\pi)$. In [8], the grid points are placed at the centers of the intervals used by the PWC distribution, i.e.,

$$\beta_j = \left(j - \frac{1}{2} \right) \frac{2\pi}{L} \text{ for } j \in \{1, \ldots, L\} . \tag{4}$$

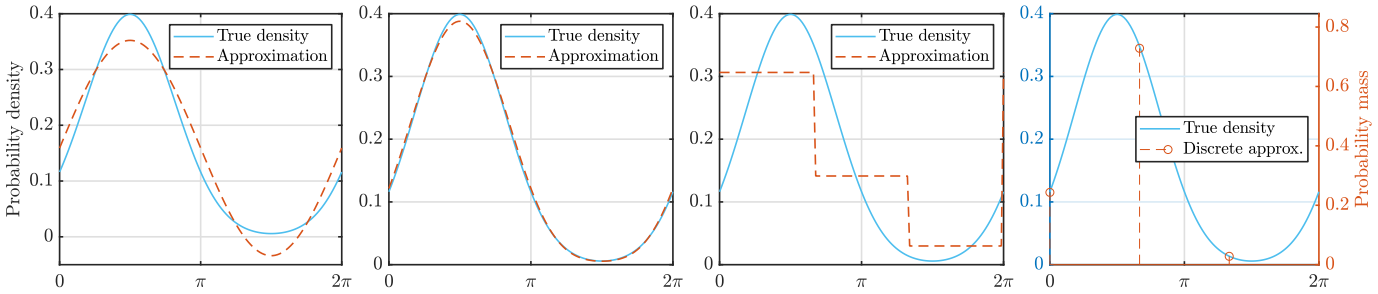
The corresponding pmf is

$$p(x; \underline{\gamma}) = \sum_{j=1}^{L} \gamma_j \mathbb{1}_{x = \beta_j} .$$

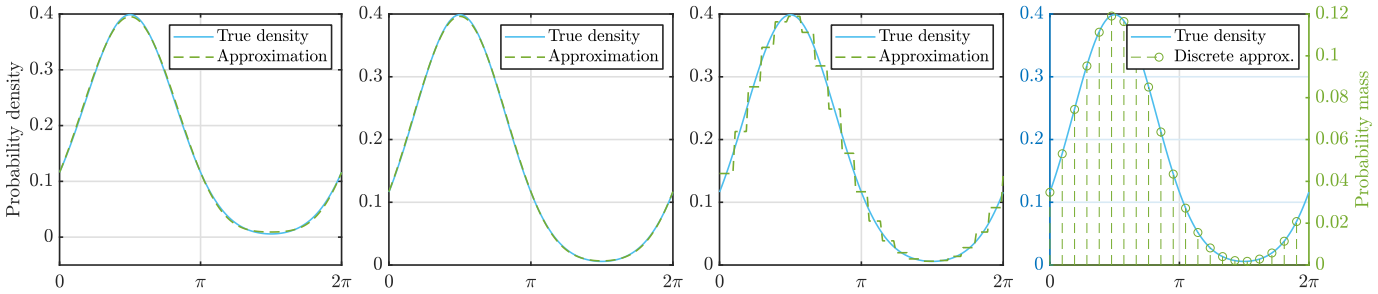
Since the pmf directly provides the probability at the grid points, the probability mass can be easily normalized to one by ensuring that the sum of the vector $\underline{\gamma}$ is 1.

Although the pmf does not describe a density, there is a relationship between the pmf and the pdf that is approximated using the pmf. To illustrate this in Figs. 1g and 1h, we have chosen the right y-axis such that the circles showing the pmf values coincide with the function values of the original density at the respective points. Further, a continuous density in the form of a PWC density could be obtained from the pmf by evenly distributing the probability mass at the grid points in the intervals used for PWC densities.

In the next section, we describe a filter that is based on a grid similar to that of the discrete filter. Instead of using a pmf on the grid points, we keep track of approximate function values on the grid and provide ways to interpolate them to provide a continuous pdf.



(a) Approximation used by the IFF with 3 coefficients.
(c) Approximation used by the SqFF with 3 coefficients.
(e) Approximation used by the PWC filter with 3 intervals.
(g) Approximation used by the discrete filter with 3 grid points.



(b) Approximation used by the IFF with 5 coefficients.
(d) Approximation used by the SqFF with 5 coefficients.
(f) Approximation used by the PWC filter with 21 intervals.
(h) Approximation used by the discrete filter with 21 grid points.

Fig. 1. Approximations used in the Fourier filters, the piece-wise constant filter, and the discrete filter for two different numbers of parameters.

III. Fourier-Interpreted Grid Filter

In our novel filter, the continuous pdf is represented using the function values $\underline{\gamma}$ on the grid points $\underline{\alpha}$. The grid points are spaced equidistantly with the first point stored at 0, i.e.,

$$\alpha_j = j\frac{2\pi}{L} \text{ for } j \in \mathcal{J},\ \mathcal{J} = \{0, \dots, L-1\}\ . \tag{5}$$

The grid is slightly different from the one used in (4), which differs by a shift of half of the length of the grid cells. For the FIGF, we start indexing the grid points with zero to simplify the formulae in the course of this paper.

While evaluating the approximation of the pdf is not required for the predict–update cycles of our estimator, being able to provide a continuous pdf is valuable to describe the current knowledge about the state. In the first subsection of this section, we address how to obtain a continuous pdf by interpolating the function values at the grid points. Two different approaches are considered—one that does not ensure that all function values are nonnegative and one that ensures their nonnegativity. The interpolations employed are reminiscent of the Fourier filters and involve determining Fourier coefficients. Therefore, we call our novel filter Fourier-interpreted grid filter (FIGF). The way the density is interpolated is only relevant to evaluate the pdf. While the IFF and SqFF differ in how the prediction and update steps are performed, the operations of the FIGF are independent of the interpolation used.

To approximate a density for the FIGF, the first step is to evaluate the density at all grid points in $\underline{\alpha}$ and store the function values in $\underline{\gamma}$. At least one of the values has to be strictly positive to obtain a valid approximation. Interpolating the function values may not yield a normalized density as the interpolation may not be equal to the original density. Thus, an additional normalization, as explained in the first subsection of this section, may be required. As we will see, the normalization constant is strictly positive for valid approximations. Since all function values are nonnegative, the values in $\underline{\gamma}$ are all nonnegative.

In the second and third subsection, we describe a simple prediction step and an update step. Both steps preserve the nonnegativity of the values in $\underline{\gamma}$. In the FIGF, we perform the prediction and update steps using operations that are the analogues of (or at least related to) those performed on the Fourier coefficient vectors in the IFF. Therefore, we explain the filter side-by-side with the IFF. The prediction and update steps of the FIGF are closely related to the alternative prediction and update steps of the Fourier filters described in [3, App. C].

A. Obtaining Normalized Density Values

In this subsection, we explain two ways to interpolate the function values on the grid to obtain a continuous pdf and ensure its normalization. For the interpolation, we convert the grid representation into one of the Fourier-based representations. In this paper, we limit ourselves to odd numbers of grid points. While it is also possible to interpolate the function values for an even number of grid points, we can only obtain the real part of the highest and lowest order complex coefficient (corresponding to the cosine part in a real Fourier series [13, Volume I, Section I.4]) from the grid values. Thus, special treatment of this coefficient would be required.

a) Interpolation Based on Trigonometric Polynomials: For the first interpolation, which permits negative function values, the Fourier coefficients are directly calculated from the function values. For L grid points, the Fourier coefficients c_k^{id} with $k \in \mathcal{K}$, $\mathcal{K} = \{-(L-1)/2, -(L-1)/2+1, \ldots, (L-1)/2\}$ can be calculated according to [10, Ch. 7]

$$c_k^{\text{id}} = \frac{1}{L} \sum_{j=0}^{L-1} \gamma_j \mathrm{e}^{2\pi i k j / L} \ . \tag{6}$$

All higher order coefficients, i.e., those with indices $|k| > (L-1)/2$ are assumed to be zero. If the full Fourier series expansion of the function includes nonzero values for terms of higher order, these evidently influence the sampled values. However, in the L function values we have, the effects of terms with order higher than $(L-1)/2$ are indistinguishable from those of lower order terms. In the context of analog–digital conversion of signals, this phenomenon is known as aliasing [10, Ch. 1]. Due to such effects, the approximation is generally not equivalent to the original density.

As discussed in the context of the IFF in Sec. II, the FFT can be used to efficiently determine Fourier coefficients from function values on a grid. The interpolation of the FIGF allowing negative function values (FIGFAN) can thus be given as

$$f_{\text{FIGFAN}}(x; \underline{\gamma}) = \sum_{k \in \mathcal{K}} c_k^{\text{id}} \mathrm{e}^{ikx} \text{ with } \underline{c}^{\text{id}} = \mathrm{FFT}(\underline{\gamma}) \ . \tag{7}$$

As all L coefficients of the resulting Fourier series have to be considered for the evaluation of the function, evaluating the approximation of the density at n points is in $O(L \log L + nL)$. If the function should be evaluated at all points on an equidistant grid starting at 0, lower run times may be achieved by padding the coefficient vector $\underline{c}^{\text{id}}$ with zeros to the next multiple of the number of desired function values and then using an inverse FFT to obtain the actual values.

As we prove in App. A, the integral of (7) over $[0, 2\pi)$ is 2π multiplied by the average of all function values in $\underline{\gamma}$, which we denote by $\mathrm{avg}(\underline{\gamma})$. Thus, the interpolation is normalized if $\mathrm{avg}(\underline{\gamma}) = \frac{1}{2\pi}$. In this paper, we use $\breve{\underline{\gamma}}$ to denote a vector that describes a potentially unnormalized density. To obtain a vector $\underline{\gamma}$ describing a normalized density from a vector $\breve{\underline{\gamma}}$ describing an unnormalized one, we can use

$$\underline{\gamma} = \frac{1}{2\pi \mathrm{avg}(\breve{\underline{\gamma}})} \breve{\underline{\gamma}} \ .$$

The second interpolation that only allows nonnegative function values uses the nonnegativity of the values in $\underline{\gamma}$, which is ensured by the initial approximation and the formulae for the prediction and update steps. Due to the nonnegativity, we can first calculate the square roots of all the values in $\underline{\gamma}$ and then apply an FFT to $\sqrt{\underline{\gamma}}$ to obtain a coefficient vector $\underline{c}^{\text{sqrt}}$. The trigonometric polynomial with Fourier coefficient vector $\underline{c}^{\text{sqrt}}$ attains precisely $\sqrt{\gamma_j}$ at α_j for all $j \in \mathcal{J}$. To obtain an approximation that passes through the actual values in $\underline{\gamma}$, the value of the trigonometric polynomial is squared. Thus, we obtain

$$f_{\text{FIGFDN}}(x; \underline{\gamma}) = \left(\sum_{k \in \mathcal{K}} c_k^{\text{sqrt}} \mathrm{e}^{ikx} \right)^2 \text{with } \underline{c}^{\text{sqrt}} = \mathrm{FFT}(\sqrt{\underline{\gamma}}) \tag{8}$$

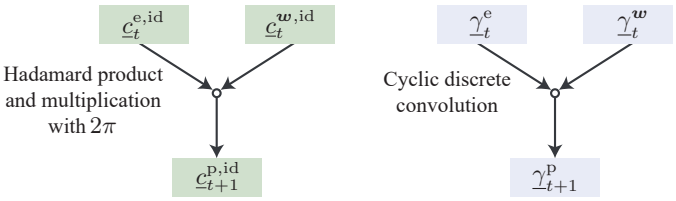
as the formula for the density of the FIGF disallowing negative function values (FIGFDN). Because all values in $\underline{\gamma}$ are nonnegative, the square roots are real numbers. Thus, the Fourier series with coefficient vector $\underline{c}^{\text{sqrt}}$ describes a real function. Further, the square thereof yields only nonnegative values. The integral over the interpolation is likewise $2\pi \mathrm{avg}(\underline{\gamma})$, which is proven in App. B. Hence, the same vector $\underline{\gamma}$ can be used for both interpolations.

Remark 1. In the Fourier filters, the Fourier coefficients are approximated based on the function values on a grid (or the square roots thereof) if no closed-form or at least analytic formula for the coefficients is available. Hence, if we initially transform a density for which no such formula is available, the density described by the Fourier coefficients of the IFF is identical to the FIGFAN interpolation of the function values. Likewise, the FIGFDN interpolation is equal to the approximation used by the SqFF. However, if a closed-form formula is used for the Fourier filters, the approximations generally differ. In this case, the lower order coefficients are calculated precisely (although they may be scaled afterward due to the normalization) and the higher order coefficients are disregarded. Thus, the approximation used by the IFF is not affected by aliasing and corresponds to a (normalized) low-pass filtered version of the original density.

b) Interpolation Using Cardinal Series: An interpolation can also be provided based on a cardinal series [14], which involves writing the function as a sum of cardinal sines $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$. If we define $\gamma[j] = \gamma_j$ for improved readability and assume the original density f_{Orig} is 2π-periodic on $\mathbb{R}$, we can write the pdf using a cardinal series (CS) according to

$$\begin{aligned} f_{\text{CS}}(x; \underline{\gamma}) &= \sum_{j=-\infty}^{\infty} f_{\text{Orig}}\left(\frac{2\pi j}{L}\right) \mathrm{sinc}\left(\frac{xL}{2\pi} - j\right) \\ &= \sum_{j=-\infty}^{\infty} \gamma[j \mod L] \, \mathrm{sinc}\left(\frac{xL}{2\pi} - j\right) \ . \end{aligned}$$

It is also possible to ensure that the interpolation is nonnegative by evaluating the cardinal series based on $\sqrt{\underline{\gamma}}$ and squaring the obtained function values. However, for both interpolations, the infinite sum cannot be truncated without changing the values of the interpolation because the cardinal sine never vanishes completely. Nonetheless, since the cardinal sine is asymptotically decreasing, reasonably close values can be obtained for $x \in [0, 2\pi)$ if a finite but sufficiently large number of terms is used. However, as the range to consider would be an additional parameter of the filter, we do not use this formula in our implementation.



(a) Operations performed by the IFF [3, Fig. 6.5].

(b) Operations performed by the FIGF.

Fig. 2. Prediction step of the IFF and FIGF for the topology-aware identity model with additive noise.

B. Prediction Step for the Identity Model with Additive Noise

For the prediction step, we limit ourselves to the topology-aware identity model with additive noise. Based on the random variable $\boldsymbol{x}_t$ describing the state at time step t and the additive system noise term $\boldsymbol{w}_t$, the state of the system at time step $t+1$ can be written as

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \boldsymbol{w}_t \mod 2\pi \ .$$

The posterior density of $\boldsymbol{x}_t$ that includes the information of all measurements $\hat{z}$ obtained up to time step t shall be written as $f_t^{\mathrm{e}}(x_t|\hat{z}_1,\dots,\hat{z}_t)$. The prior density $f_{t+1}^{\mathrm{p}}(x_{t+1}|\hat{z}_1,\dots,\hat{z}_t)$ of $\boldsymbol{x}_{t+1}$ based on the same measurements can be derived using the Chapman–Kolmogorov equation. For the identity model with additive noise, the Chapman–Kolmogorov equation involving an integral over the entire sample space $[0,2\pi)$ can be simplified to [5]

$$f_{t+1}^{\mathrm{p}}(x_{t+1}|\hat{z}_1,\dots,\hat{z}_t) = \int_0^{2\pi} f_t^{\boldsymbol{w}}(x_{t+1}-x_t) f_t^{\mathrm{e}}(x_t|\hat{z}_1,\dots,\hat{z}_t)\,dx_t,$$

which is a continuous cyclic convolution of the posterior density f_t^{e} and the density of the additive noise term $f_t^{\boldsymbol{w}}$.

Before considering the FIGF, we regard the prediction step of the IFF. In the IFF, the cyclic convolution can be realized efficiently via a Hadamard (entry-wise) product of the coefficient vectors $\underline{c}_t^{\mathrm{e,id}}$ and $\underline{c}_t^{\boldsymbol{w},\mathrm{id}}$ and a multiplication with 2π (see Fig. 2a). Assuming that the coefficient vector $\underline{c}_t^{\mathrm{e,id}}$ was obtained as the result of the previous update step, only $\underline{c}_t^{\boldsymbol{w},\mathrm{id}}$ has to be determined via analytic formulae (if available) or an FFT applied to function values on a grid. For time-invariant systems, $\underline{c}_t^{\boldsymbol{w},\mathrm{id}}$ can be determined once and reused in future time steps. If $\underline{c}_t^{\mathrm{e,id}}$ is unknown but the function f_t^{e} is given, a transformation can likewise be performed via an FFT.

Aside from errors stemming from previous time steps and the error resulting from the use of a finite number of coefficients to represent $f_t^{\boldsymbol{w}}$, the prediction step realized via the Hadamard product yields the precise predicted density without introducing an additional error. As the convolution of the square roots of two functions is not equal to the square root of the convolution, transformations that introduce additional errors are required for the SqFF [6].

For the FIGF, we perform operations that are the analogues of the operations performed on the Fourier coefficients for the IFF. We assume the vector $\underline{\gamma}_t^{\mathrm{e}}$ of function values of the prior density on the grid is given. Then, $f_t^{\boldsymbol{w}}$ is evaluated at the same grid points. Similar as in the IFF, the vector $\underline{\gamma}_t^{\boldsymbol{w}}$ can be stored and reused for time-invariant noise terms. As illustrated in Fig. 2b, the prediction step of the FIGF is realized via a discrete cyclic convolution of the vectors $\underline{\gamma}_t^{\mathrm{e}}$ and $\underline{\gamma}_t^{\boldsymbol{w}}$, i.e.,

$$\gamma_{t+1}^{\mathrm{p}}[k] = \sum_{j=0}^{L-1} \gamma_t^{\mathrm{e}}[j]\gamma_t^{\boldsymbol{w}}[k-j \mod L] \ . \tag{9}$$

If $\underline{c}_t^{\mathrm{e,id}} = \mathrm{FFT}(\underline{\gamma}_t^{\mathrm{e}})$ and $\underline{c}_t^{\boldsymbol{w},\mathrm{id}} = \mathrm{FFT}(\underline{\gamma}_t^{\boldsymbol{w}})$ then $\underline{c}_{t+1}^{\mathrm{p,id}} = \mathrm{FFT}(\underline{\gamma}_{t+1}^{\mathrm{p}})$, and thus, there is a direct correspondence between this prediction step and that of the IFF[1]. However, there is no direct correspondence between the predictions steps of the FIGF and the SqFF.

For the FIGF, the nonnegativity of the resulting values in $\underline{\gamma}_{t+1}^{\mathrm{p}}$ is ensured as (9) only involves multiplications and summations, which yield nonnegative values if the original values are nonnegative. The prediction step is in $O(L\log L)$ for the FIGF due to the discrete cyclic convolution involved and in $O(L)$ for the IFF (assuming $k_{\max} = (L-1)/2$).

C. Update Step

For the update step, we require a likelihood function $f_t^{\mathrm{L}}(z_t|x_t)$ that describes the probability that a measurement z_t is obtained for a state x_t. For a fixed measurement $\hat{z}_t$, the likelihood $f_t^{\mathrm{L}}(\hat{z}_t|x_t)$ is only a function of $x_t \in [0,2\pi)$, and thus, approximations based on grid points and Fourier coefficients are possible even when the measurement space is, e.g., a multidimensional Euclidean space.

Based on the prior density $f_t^{\mathrm{p}}(x_t|\hat{z}_1,\dots,\hat{z}_{t-1})$ and the likelihood $f_t^{\mathrm{L}}(\hat{z}_t|x_t)$, the posterior density can be written according to Bayes' rule as

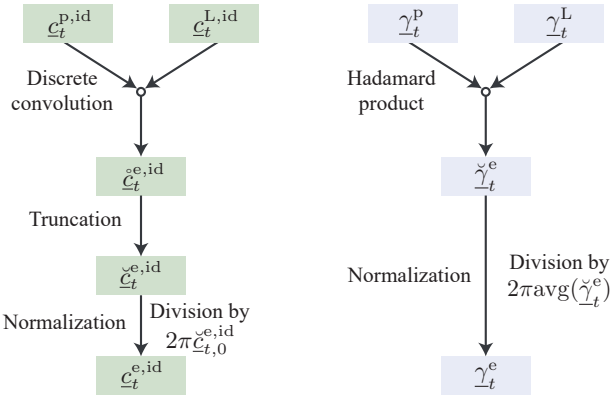
$$\begin{aligned} f_t^{\mathrm{e}}(x_t|\hat{z}_1,\dots,\hat{z}_t) &= \frac{f_t^{\mathrm{L}}(\hat{z}_t|x_t) f_t^{\mathrm{p}}(x_t|\hat{z}_1,\dots,\hat{z}_{t-1})}{\int_0^{2\pi} f_t^{\mathrm{L}}(\hat{z}_t|x_t) f_t^{\mathrm{p}}(x_t|\hat{z}_1,\dots,\hat{z}_{t-1})\,dx_t} \\ &\propto f_t^{\mathrm{L}}(\hat{z}_t|x_t) f_t^{\mathrm{p}}(x_t|\hat{z}_1,\dots,\hat{z}_{t-1}) \ . \end{aligned}$$

Thus, the posterior density can be obtained via a multiplication of the prior density and the likelihood with a subsequent normalization.

The update step of the IFF is illustrated in Fig. 3a. The multiplication of the two functions can be realized as a (non-cyclic) discrete convolution of the Fourier coefficient vectors. When convolving two coefficient vectors comprising nonzero coefficients from $-k_{\max}$ to $k_{\max}$, the full convolution result $\mathring{\underline{c}}_t^{\mathrm{e,id}}$ may have nonzero coefficients for indices from $-2k_{\max}$ to $2k_{\max}$. If $\underline{c}_t^{\mathrm{p,id}}$ and $\underline{c}_t^{\mathrm{L,id}}$ perfectly describe the prior density and likelihood, $\mathring{\underline{c}}_t^{\mathrm{e,id}}$ describes the error-free multiplication result.

To avoid an ever-increasing size of the coefficient vector, all coefficients with indices $k > |k_{\max}|$ are discarded in the IFF. Discarding the higher order coefficients corresponds to low-pass filtering the result, which introduces an approximation error. As the integral of a Fourier series over $[0,2\pi)$ is $2\pi c_0^{\mathrm{id}}$ (see App. A), the normalization can be realized by dividing

[1] In fact, the cyclic convolution is often realized by transforming the values to Fourier coefficients via an FFT, calculating a Hadamard product, and finally transforming the result back via an inverse FFT. This is done, for example, in the `cconv` function of Matlab 2019a.



(a) Operations performed by the IFF [3, Fig. 6.3].

(b) Operations performed by the FIGF.

Fig. 3. Update step of the IFF and FIGF.

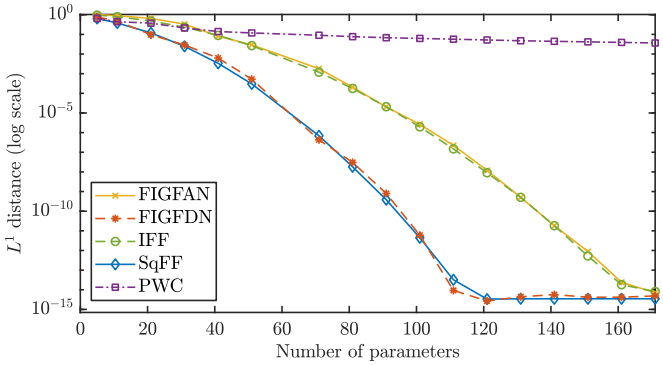


Fig. 4. Evaluation results for the different approximations of the density.

all coefficients by $2\pi c_0^{\mathrm{id}}$. Because the multiplication of the square roots of two functions is equal to the square root of the multiplication of the original functions, the update step can be realized similarly (albeit with a different normalization constant) for the SqFF.

For the FIGF, we determine the function values of the multiplication result of f_t^{p} and f_t^{L} on the grid. As depicted in Fig. 3b, the vector containing the function values of the unnormalized multiplication result can be obtained via a Hadamard product of the vectors $\underline{\gamma}_t^{\mathrm{p}}$ and $\underline{\gamma}_t^{\mathrm{L}}$. The normalization is performed as described in Sec. III-A. The values in $\underline{\gamma}_t^{\mathrm{e}}$ are proportional to the actual function values of f_t^{e} on the grid. However, the interpolation of $\underline{\gamma}_t^{\mathrm{e}}$ does generally not correspond to the normalized multiplication of the interpolations of $\underline{\gamma}_t^{\mathrm{p}}$ and $\underline{\gamma}_t^{\mathrm{L}}$. As discussed for the IFF, the multiplication of two functions can introduce higher order terms. Thus, even if f_t^{p} and f_t^{L} can be perfectly interpolated based on the L function values, L values may not suffice to describe f_t^{e} precisely. We thus only have L function values of a function that may require $2L-1$ function values on a grid to interpolate precisely. The error is reminiscent of the error encountered due to aliasing when approximating the density using too few grid points. Due to the different characteristics of the approximation errors in the IFF and the FIGF, $\underline{c}_t^{\mathrm{e,id}} \neq \mathrm{FFT}(\underline{\gamma}_t^{\mathrm{e}})$ generally holds even if $\underline{c}_t^{\mathrm{p,id}} = \mathrm{FFT}(\underline{\gamma}_t^{\mathrm{p}})$ and $\underline{c}_t^{\mathrm{L,id}} = \mathrm{FFT}(\underline{\gamma}_t^{\mathrm{L}})$.

The run time complexity is in $O(L)$ as the Hadamard product and the division are both in $O(L)$. For the IFF, a discrete convolution is required that is in $O(L \log L)$ when $L = 2k_{\max}+1$. The update step of the FIGF preserves the nonnegativity of the function values on the grid. If two nonnegative numbers are multiplied and the result is then multiplied by a strictly positive normalization constant, the result is always nonnegative. However, a problem that may arise for likelihoods with very narrow support is that all values in $\underline{\check{\gamma}}_t^{\mathrm{e}}$ may be zero. In this case, the result is invalid and it may be viable to retry the update step using a denser grid. The likelihood can usually simply be evaluated for more inputs, and the function values of the prior density on a denser grid can be obtained by performing an FFT on $\underline{\gamma}_t^{\mathrm{p}}$, padding the resulting coefficient vector with zeros, and then transforming the result back via an inverse FFT.

IV. Evaluation

We implemented the FIGF in Matlab, and the implementation is publicly available as `FIGFilter` as part of libDirectional [15]. In our evaluation, we compare the densities obtained via the FIGF, the PWC filter, and the Fourier filters. The discrete filter is not considered because it does not provide a continuous pdf. For the FIGF, we regard both the FIGFAN and the FIGFDN interpolation. Three evaluations are performed. The first is concerned with the quality of the approximation of a given density. Then, we show the approximation quality after prediction and update steps in the second and third evaluation. The quality is always measured by the L^1 distance (i.e., the integral over the absolute difference) between the true density and the approximation.

In our first evaluation, we approximate a von Mises density

$$f_{\mathrm{VM}}(x;\mu,\kappa) = \frac{\mathrm{e}^{\kappa\cos(x-\mu)}}{2\pi I_0(\kappa)}$$

with $\mu = 3$ and $\kappa = 100$, which is a quite concentrated density that requires many coefficients or grid points to approximate accurately. For von Mises densities, closed-form formulae for the Fourier coefficients are available (otherwise, as explained in Rem. 1, the approximation of the IFF would match the FIGFAN interpolation and the approximation of the SqFF would match the FIGFDN interpolation). In the results in Fig. 4, the FIGF and the Fourier filters show an at least exponential convergence speed until they reach their numerically optimal result. The convergence speed is in line with the convergence of the coefficients as described in [6, Sec. IV.C]. The nonnegative approximations outperform those also involving negative values. Compared with the other approaches, the approximation of the PWC filter converges slowly.

In the second evaluation, we evaluate the prediction step for the identity model with additive noise for the wrapped normal densities $f_t^{\mathrm{e}}(x) = f_{\mathrm{WN}}(x;1,0.5)$ and $f_t^{\boldsymbol{w}}(x) = f_{\mathrm{WN}}(x;0,0.5)$. As wrapped normal densities are closed under the topology-aware convolution [16, Section 2.2.6], the filter results can be efficiently compared with the ground truth. The evaluation

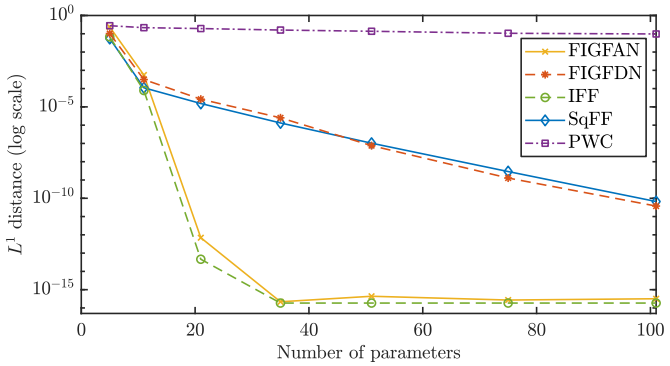


Fig. 5. Evaluation results for the prediction step with additive noise.

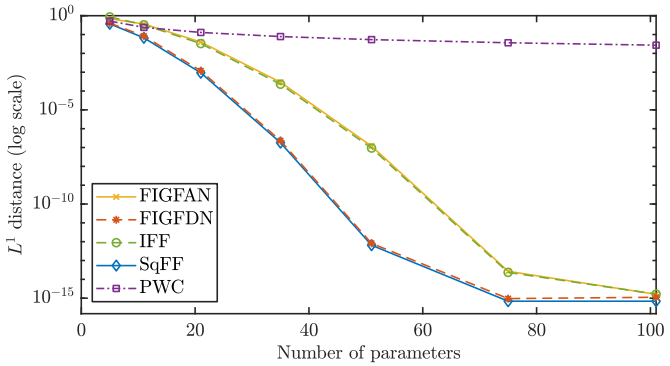


Fig. 6. Evaluation results for the update step.

results are shown in Fig. 5. In this evaluation, the FIGFAN and IFF are superior to the FIGFDN and the SqFF. The convergence speed of the PWC filter is the worst again.

In the third evaluation, we compare the filters' results after an update step. For the update step, we use a von Mises-distributed prior density $f_t^{\mathrm{p}}(x) = f_{\mathrm{VM}}(x; 1, 0.5)$ and likelihood function $f_t^{\mathrm{L}}(z|x) = f_{\mathrm{VM}}(z; x, 0.5)$. Von Mises densities were chosen because they are closed under multiplication with a subsequent normalization [4]. Because the update step is dependent on the actual measurement, we do 10000 runs. In each run, the true state is drawn from the prior density, and then, a measurement is drawn according to the likelihood function. As depicted in Fig. 6, the FIGFDN and SqFF achieve better results than the FIFGAN and IFF. All approaches outperform the PWC filter regarding the approximation quality.

V. Conclusion

In this paper, we have discussed very general estimation approaches for filtering on the unit circle based on grids and trigonometric polynomials. We considered previously proposed grid-based approaches to estimation on the unit circle that were not tailored to accurately approximating densities. The discrete filter, which employs a pmf, does not provide a continuous pdf, and the piece-wise constant filter achieves far worse approximation quality than the Fourier filters. We introduced the FIGF as a new grid-based approach that is closely related to the Fourier filters. For the Fourier filters, the variant ensuring the nonnegativity of the approximation requires special prediction and update steps that lead to a higher theoretical and computational complexity. For the FIGF, we obtain a single grid-based filter that achieves results comparable to those of the Fourier filters. No changes to the prediction and update steps are required and only the interpolation of the grid values has to be modified to ensure the nonnegativity.

While the IFF and FIGF are clearly related, the run time complexities of the operations differ. For L parameters, the prediction step of the IFF for time-invariant system noise terms is in $O(L)$ and the update step is in $O(L \log L)$, whereas the prediction step of the FIGF is in $O(L \log L)$ and the update step in $O(L)$. If multiple consecutive prediction or update steps are to be performed, an advantage in the run time may be achieved by switching from one filter to the other, which is possible in $O(L \log L)$ using the FFT or inverse FFT. Temporarily switching from the FIGF to the IFF (without using closed-form formulae for $\underline{c}_t^{\boldsymbol{w},\mathrm{id}}$) for the prediction step does not introduce the risk of obtaining negative values from the inverse FFT because the prediction step based on the Fourier coefficients precisely corresponds to the prediction step based on the function values on the grid.

In future work, the FIGF could be generalized to higher dimensions, analogous to the Fourier filters for multivariate estimation problems [7]. Further, efficient procedures to randomly sample densities in the representations of the FIGF or the Fourier filters may allow for additional interesting insights.

Appendix
Proof for the Normalization Formula

For both the FIGFAN and the FIGFDN interpolation, the integral of the interpolation over $[0, 2\pi)$ is $2\pi \mathrm{avg}(\underline{\gamma})$. First, we prove this for the interpolation allowing negative values, and then we provide a proof for the interpolation disallowing negative function values.

A. Interpolation Allowing Negative Function Values

Theorem 1. *The integral of the FIGFAN interpolation described in (7) over* $[0, 2\pi)$ *is* $2\pi \mathrm{avg}(\underline{\gamma})$.

Proof: Because the integral of e^{ikx} over $[0, 2\pi)$ is zero unless $k = 0$, we obtain

$$\int_0^{2\pi} \sum_{k \in \mathcal{K}} c_k^{\mathrm{id}} \mathrm{e}^{ikx} = 2\pi c_0^{\mathrm{id}} \tag{10}$$

for the integral over the trigonometric polynomial. By using the formula for the Fourier coefficients (6) for c_0^{id}, we obtain

$$c_0^{\mathrm{id}} = \frac{1}{L} \sum_{j=0}^{L-1} \gamma_j \mathrm{e}^0 = \mathrm{avg}(\underline{\gamma}) \ ,$$

which proves that the integral is $2\pi \mathrm{avg}(\underline{\gamma})$. ∎

B. Interpolation Disallowing Negative Function Values

Theorem 2. *The integral of the FIGFDN interpolation, as given by (8), over* $[0, 2\pi)$ *is* $2\pi \mathrm{avg}(\underline{\gamma})$.

Proof: The interpolation involves squaring a trigonometric polynomial with Fourier coefficient vector $\underline{c}^{\mathrm{sqrt}}$ comprising

coefficients with indices from $-(L-1)/2$ to $(L-1)/2$. By calculating a (non-cyclic) discrete convolution of $\underline{c}^{\text{sqrt}}$ with itself, we can obtain a Fourier coefficient vector $\underline{c}^{\text{conv}}$ for the trigonometric polynomial that directly describes the interpolation. As higher-order coefficients may become nonzero, we obtain a trigonometric polynomial with potentially nonzero coefficients from $-(L-1)$ to $L-1$. The interpolation can be described based on the Fourier coefficient vector $\underline{c}^{\text{conv}}$ according to

$$f_{\text{FIGFDN}}(x;\underline{\gamma}) = \sum_{k=-(L-1)}^{L-1} c_k^{\text{conv}} \mathrm{e}^{ikx} \text{ with } \underline{c}^{\text{conv}} = \underline{c}^{\text{sqrt}} * \underline{c}^{\text{sqrt}} \ .$$

As we know from App. A, only c_0^{conv} is required to determine the integral of a trigonometric polynomial over $[0, 2\pi)$. Using the formula for the convolution, we obtain

$$\begin{aligned} c_0^{\text{conv}} &= \sum_{j=-(L-1)/2}^{(L-1)/2} c_j^{\text{sqrt}} c_{-j}^{\text{sqrt}} \\ &= \sum_{j=-(L-1)/2}^{(L-1)/2} \left(\left(\frac{1}{L} \sum_{n=0}^{L-1} \sqrt{\gamma_n} \mathrm{e}^{2\pi i j n/L} \right) \right. \\ &\qquad \left. \cdot \left(\frac{1}{L} \sum_{m=0}^{L-1} \sqrt{\gamma_m} \mathrm{e}^{-2\pi i j m/L} \right) \right) . \end{aligned} \tag{11}$$

The complex exponential function is 2π-periodic. Due to the additional factors in the exponents, the exponential functions in (11) are L-periodic in j. As this holds for all terms in both inner sums, we can change the range of the outer summation. The value obtained for any negative j is equivalent to that obtained for $j+L$. Thus, we can replace the terms for j from $-(L-1)/2$ to -1 with the terms from $(L-1)/2+1$ to $L-1$. Hence, (11) is equivalent to

$$\begin{aligned} c_0^{\text{conv}} &= \sum_{j=0}^{L-1} \left(\frac{1}{L} \sum_{n=0}^{L-1} \sqrt{\gamma_n} \mathrm{e}^{2\pi i j n/L} \right) \left(\frac{1}{L} \sum_{m=0}^{L-1} \sqrt{\gamma_m} \mathrm{e}^{-2\pi i j m/L} \right) \\ &= \frac{1}{L} \sum_{n=0}^{L-1} \sum_{m=0}^{L-1} \sqrt{\gamma_n} \sqrt{\gamma_m} \underbrace{\frac{1}{L} \sum_{j=0}^{L-1} \mathrm{e}^{2\pi i j (n-m)/L}}_{\Lambda(n-m)} . \end{aligned}$$

The expression defined as $\Lambda(n-m)$ is equal to the formula for the Fourier coefficient (6) with index $n-m$ when $\gamma_j = 1$ for all j. Thus, we obtain the Fourier coefficients of a function that is constantly 1. It is important to note that the formula (6) only yields the actual Fourier coefficient for indices in $\mathcal{K}$. For values of $n-m$ that are not in $\mathcal{K}$, a different Fourier coefficient is returned. Due to the L-periodicity of $\Lambda(n-m)$, L can be added to or subtracted from the input argument without changing the result. Thus, for example, we know that for $n=0$ and $m=L-1$, the expression $\Lambda(-L+1)$ corresponds to the formula for the Fourier coefficient with index 1. For all other $(n-m) \notin \mathcal{K}$, the terms are likewise equal to the formula for the Fourier coefficient with index $n-m+L$ if $(n-m) < -(L-1)/2$ and $n-m-L$ if $(n-m) > (L-1)/2$.

For a function that is constantly 1, the zeroth Fourier coefficient is 1 and all other coefficients are 0. The zeroth coefficient is only returned when $n-m$ modulo L is zero. As n and m are nonnegative and $n-m$ is always an integer in the range from $-L+1$ to $L-1$, the zeroth coefficient is only obtained for $n=m$. For all other values of n and m, a Fourier coefficient that is 0 is obtained. Hence, $\Lambda(n-m) = \mathbb{1}_{n=m}$. Thus, we obtain

$$\begin{aligned} c_0^{\text{conv}} &= \frac{1}{L} \sum_{n=0}^{L-1} \sum_{m=0}^{L-1} \sqrt{\gamma_n} \sqrt{\gamma_m} \mathbb{1}_{n=m} \\ &= \frac{1}{L} \sum_{n=0}^{L-1} \sqrt{\gamma_n} \sqrt{\gamma_n} = \operatorname{avg}(\underline{\gamma}) \ , \end{aligned}$$

which proves that the interpolated density integrates to $2\pi \operatorname{avg}(\underline{\gamma})$. ■

Acknowledgment

This work is supported by the German Research Foundation (DFG) under grant HA 3789/16-1.

References

[1] 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.

[2] R. S. Bucy and A. J. Mallinckrodt, "An Optimal Phase Demodulator," *Stochastics*, vol. 1, no. 1-4, pp. 3–23, 1975.

[3] F. Pfaff, "Multitarget Tracking Using Orientation Estimation for Optical Belt Sorting," Ph.D. dissertation, Karlsruhe Institute of Technology, defended on Nov. 13, 2018.

[4] 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)*, Dec. 2009.

[5] 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.

[6] 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.

[7] ——, "Multivariate Angular Filtering Using Fourier Series," *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 206–226, Dec. 2016.

[8] 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.

[9] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[10] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 4th ed. Prentice Hall, 2006.

[11] 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.

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

[13] A. Zygmund, *Trigonometric Series*, 3rd ed. Cambridge University Press, 2003, vol. 1 and 2.

[14] E. T. Whittaker, "On the Functions Which are Represented by the Expansions of the Interpolation-Theory," *Proceedings of the Royal Society of Edinburgh*, vol. 35, pp. 181–194, 1915.

[15] G. Kurz, I. Gilitschenski, F. Pfaff, L. Drude, U. D. Hanebeck, R. Haeb-Umbach, and R. Y. Siegwart, "Directional Statistics and Filtering Using libDirectional," *Journal of Statistical Software*, May 2019.

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