Nonlinear Prediction for Circular Filtering Using Fourier Series

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

***Abstract*—While nonlinear filtering for circular quantities is closely related to nonlinear filtering on linear domains, the underlying manifold enables the development of novel filters that take advantage of the boundedness of the domain. Previously, we introduced Fourier filters that approximate the density or its square root using Fourier series. For these filters, we proposed filter steps for arbitrary likelihoods and prediction steps for the identity system model with additive noise. This paper adds the capability of handling arbitrary transition densities in the prediction step, which facilitates, e.g., the use of the filters for nonlinear systems with additive noise. In the evaluation, the new prediction steps for the Fourier filters outperform an SIR particle filter, a grid filter, and a nonlinear variant of the von Mises filter.**

***Index Terms*—Directional statistics, Chapman–Kolmogorov equation, nonlinear filtering, Fourier series, recursive Bayesian estimation**

I. Introduction

In recursive Bayesian estimation, a lot of research revolves around derivatives of the Kalman filter, a technique that has many inherent limitations and underlying assumptions. A strong assumption of most current filters is the unboundedness of the domain. The need to properly account for the underlying manifold in stochastic modeling has been known for a long time and was mentioned, e.g., in 1953 by Fisher [1]. Recursive Bayesian estimation is evolving quickly recently and developing filters that are tailored to special systems and topologies is a promising means to increase estimation quality while maintaining real time performance. In this paper, we limit ourselves to circular manifolds but the concepts can also be used for multivariate estimation problems on the hypertorus.

Circular topologies occur in a variety of applications. While stochastic modeling of circular quantities has been popular for a long time in geology [2], [3] and biology [4], [5], new applications are emerging in technological applications. For example, modern signal processing tasks such as phase estimation [6], [7] and speaker tracking [8] feature an underlying circular domain and can benefit from specialized filters.

Currently, there are few filters that are suited to arbitrary system functions on periodic domains. One suitable filter is presented in [9] but it involves approximations using a unimodal density, which imposes a limit on the performance of the filter. The limitations of this sample-based filter are similar to that of nonlinear filters on linear domains such as the UKF [10] that have inspired the approach. A very universal approach is to use particle filters [11] adapted to nonlinear domains. However, particle filters may require many particles to achieve good results and are thus potentially costly. Furthermore, deriving a meaningful continuous density from particles without making any assumptions is no trivial task.

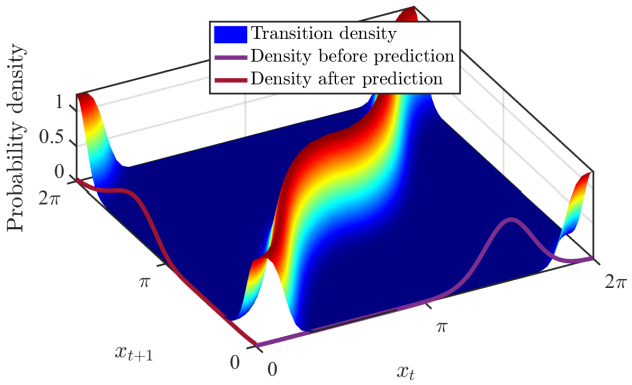


Figure 1. Visualization of a prediction step for a nonlinear system function. The density before the prediction step is shown in 2D on the right, the transition density is shown in 3D, and the result of the prediction step is shown in 2D on the left. Note that the nonlinear prediction even voids the symmetry of the distribution.

Previously, we have proposed Fourier filters [12] that are based on the idea of approximating the density or its square root using a truncated Fourier series. We have further extended this concept to multivariate estimation problems in [13]. Since using only a finite number of Fourier coefficients usually induces an approximation error, approximating the density directly can lead to an approximation that has negative function values. Approximating the square root of the density as a means to provide a valid density in each time step is a concept that has been previously used by Brunn et al. [14], [15] on linear domains. Depending on whether we approximate the density or its square root as a Fourier series, we refer to the filter as the Fourier identity filter or as the Fourier square root filter. The filters also build upon the idea of approximating densities on circular manifolds via Fourier series, which has been considered, e.g., by Willsky [16], [17] and Fernández-Durán [18].

The Fourier filters introduced by us allow for the use of arbitrary likelihood functions in the filter step. However, in

the version described in [12] and [13], the filters only support an identity system model with additive noise in the prediction step. In this paper, we introduce a more sophisticated version of the prediction steps that can handle transition densities of highly nonlinear systems, such as the one shown in Fig. 1. As the transition density is easy to derive for nonlinear system models with additive noise, these problems are covered as a special case.

II. General Prediction Step for Recursive Bayesian Estimation

For the prediction step, the most common way to express one's knowledge about the behavior of a system over time is in terms of a system equation based on random variables. Most generally, we can write such a system equation as

$$\boldsymbol{x}_{t+1}^p = a_t(\boldsymbol{x}_t^e, u_t, \boldsymbol{w}_t) \ ,$$

in which the random variable $\boldsymbol{x}_t^e \sim f_t^e(x_t|z_1,\dots,z_t)$ describes the state at time step t based on all measurements up to time step t and $\boldsymbol{x}_{t+1}^p \sim f_{t+1}^p(x_{t+1}|z_1,\dots,z_t)$ describes the state at time $t+1$ without incorporating knowledge from the measurement in time step $t+1$. Finally, u_t denotes a known system input and $\boldsymbol{w}_t \sim f_t^w(w_t)$ the system noise term. As u_t is known, we can include it in the nonlinear function to write

$$\boldsymbol{x}_{t+1}^p = \overline{a}_t(\boldsymbol{x}_t^e, \boldsymbol{w}_t) \ .$$

A special case regarded very often in recursive Bayesian estimation is that of additive system noise, i.e.,

$$\boldsymbol{x}_{t+1}^p = \tilde{a}_t(\boldsymbol{x}_t^e) + \boldsymbol{w}_t \ . \tag{1}$$

An important alternative to describing the system behavior using random variables is to describe it using probability densities. In this case, we can write the prediction step using the Chapman–Kolmogorov equation

$$f_{t+1}^p(x_{t+1}|z_1,\dots,z_t) = \int_{\Omega_x} f_t^T(x_{t+1}|x_t) f_t^e(x_t|z_1,\dots,z_t) dx_t \ ,$$

in which Ω_x denotes the sample space (e.g., $[0, 2\pi)$ in the circular case) and $f_t^T(x_{t+1}|x_t)$ the transition density for the state from time step t to $t+1$. In this paper, we describe how f_{t+1}^p can be approximated via the Chapman–Kolmogorov equation when f_t^T and f_t^e can be adequately represented by a truncated Fourier series.

It is important to note that solving the Chapman–Kolmogorov equation also covers nonlinear system models with additive noise (as shown in (1)) as a special case since we can write the corresponding transition density as

$$f_t^T(x_{t+1}|x_t) = f_t^w(x_{t+1} - a(x_t)) \ .$$

Solutions to generate transition densities from a variety of other system equations exist, such as for an identity system model with multiplicative noise.

III. Basics of Directional Statistics and Fourier Series

As a brief reminder, we revisit some concepts of directional statistics and Fourier series. Directional statistics is a wide field of statistics and we recommend the two popular books by Mardia and Jupp [19] and Jammalamadaka and Sengupta [20] for a more detailed overview of the field. While a variety of circular distributions have been proposed, we only introduce the von Mises distribution as it is the only distribution used in this paper. However, the concepts presented can also be used for other circular distributions as explained in [12].

Regarding Fourier series, we explain one-dimensional Fourier series and their relationship to trigonometric moments. Furthermore, we introduce the most common definition of two-dimensional Fourier series as we will later use them to approximate the transition density. For more information on Fourier series, the reader is referred to the works of Zygmund [21].

A. The von Mises Distribution

The von Mises distribution [2], [19, Sec. 3.5] is a popular circular distribution of which the density f_{VM} can be written according to

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

in which $I_0(\cdot)$ is the modified Bessel function of the first kind. While the von Mises distribution is closed under (normalized) multiplication, it is not closed under convolution. Thus, a prediction step with an identity system model does not yield a von Mises distribution even if all densities involved are von Mises distributed.

B. Trigonometric Moments

In directional statistics, trigonometric moments serve as a counterpart to power moments on linear domains. Each trigonometric moment consists of two components [19, Sec. 3.4.1] but can also be written using a single complex value [20, Sec. 2.1] according to

$$m_k = \mathbb{E}(\mathrm{e}^{ik\boldsymbol{x}}) = \int_0^{2\pi} f(x)\mathrm{e}^{ikx} dx \ . \tag{2}$$

As some distributions such as the von Mises distribution can be parametrized by the first trigonometric moment, trigonometric moments are well suited for moment matching. Nonlinear filters based on samples, such as [9], rely on matching the resulting trigonometric moment of the samples using a suitable distribution. For some distributions, moment matching results in the minimization of the information loss in form of the Kullback–Leibler divergence [22] as shown in [23] for the von Mises distribution.

C. One-dimensional Fourier Series

Densities of distributions on the circle are real functions on $[0, 2\pi)$. When representing a function using a Fourier series,

the function is represented by a Fourier coefficient vector $\underline{c}$ such that

$$f(x) = \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{ikx}$$

holds. As densities are usually square integrable, they lend themselves well to approximations by Fourier series. A square integrable function can be represented by a Fourier series with a square summable coefficient vector [24, Sec. I-5]. This implies that (at least asymptotically) the Fourier coefficients c_k and thus the effect on the density tends to zero for $k \to -\infty$ and $k \to \infty$. As a valid density integrates to one, the square root of the density is always guaranteed to be square integrable.

The corresponding Fourier coefficients can be calculated via the integral

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) \mathrm{e}^{-ikx} dx \ .$$

Due to the close relationship to (2), we can see that trigonometric moments are also matched when approximating the (non-rooted) density for the Fourier identity filter. While it is possible to solve this integral analytically in special cases (as we have shown for some distributions on the circle in [12]), a more general solution to approximate arbitrary functions is necessary for a truly versatile filter. To approximate the Fourier coefficients of arbitrary functions, we use the efficient way proposed in [25], which has since become popular as the fast Fourier transform (FFT) [26] for the related problem of calculating the discrete Fourier transform [27, Ch. 2]. If we wish to approximate a density using Fourier coefficients from $-k_{\max}$ to $k_{\max}$, we need to evaluate the function at $n = 2k_{\max}+1$ equidistantly spaced points to obtain the desired number of n Fourier coefficients via the FFT.

D. Two-dimensional Fourier Series

As an intuitive two-dimensional extension of a Fourier series, we represent the function of the vector $\underline{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$ by a series

$$f(\underline{x}) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} c_{k_1,k_2} \mathrm{e}^{ik_1x_1} \mathrm{e}^{ik_2x_2} \ ,$$

which is a special case of the Fourier series for the arbitrary dimensional case presented in [21, Ch. XVII]. The Fourier coefficients are then calculated according to

$$c_{k_1,k_2} = \frac{1}{(2\pi)^2} \int_{[0,2\pi)^2} f(\underline{x}) \mathrm{e}^{-ik_1x_1} \mathrm{e}^{-ik_2x_2} d\underline{x} \ .$$

To approximate the Fourier coefficients of a function of a two-dimensional vector argument, we can use a two-dimensional variant of the FFT used in the one-dimensional case.

IV. Prediction Steps for Fourier Filters

In this chapter, we first explain the prediction steps for identity system models with additive noise, which we have introduced in [12]. Afterwards, we present the new prediction steps that allow for the use of arbitrary transition densities. While we only present the novel prediction steps for the univariate case, they can also be extended to the multivariate case. To model the transition density for d-variate problems, we need to use a $2d$-dimensional Fourier series. As we have already proposed the filter steps for arbitrary likelihoods in [12], [13], we do not explain or evaluate the filter steps in this paper.

A. Prediction Steps for Identity System Models

In the case of an identity system model with additive noise, the Chapman–Kolmogorov equation reduces to

$$f_{t+1}^p(x_{t+1}|z_1,\ldots,z_t) = \int_{\Omega_x} f_t^w(x_{t+1}-x_t) f_t^e(x_t|z_1,\ldots,z_t) dx_t$$

and it is thus sufficient to calculate the convolution of f_t^w and f_t^e. Convolving two functions in a Fourier series representation is computationally easy. The Fourier coefficients of the result of the convolution can be obtained by calculating the Hadamard (entrywise) product of the Fourier coefficient vectors. Thus, the prediction step is trivial for the Fourier identity filter.

For the Fourier square root filter, a more sophisticated prediction step is necessary since naïvely calculating the Hadamard product of the coefficient vectors would yield the coefficient vector of $\sqrt{f_t^e} * \sqrt{f_t^w}$ instead of $\sqrt{f_t^e * f_t^w}$. In our proposed approach, we first square $\sqrt{f_t^e}$ and $\sqrt{f_t^w}$ individually. Functions represented as a Fourier series can be multiplied (and thus also be squared) by calculating the discrete convolution of the Fourier coefficient vectors. Thus, using two discrete convolutions, we can obtain the Fourier coefficient vectors of f_t^e and f_t^w and both functions represented by the coefficient vectors are guaranteed to be nonnegative. From these coefficient vectors, the coefficient vector of $f_t^e * f_t^w$ can be calculated using the Hadamard product and, as no error is made in the convolution operation, the resulting density is also nonnegative. As long as only prediction steps are performed, the next prediction step can be performed by squaring $\sqrt{f_{t+1}^w}$ via a discrete convolution again and then using the Hadamard product to perform the next convolution.

To restore the square root representation for the next filter step, we use an efficient way to approximate the Fourier coefficients of the square root of a function when the function is given by its Fourier coefficients. As sketched in Fig. 2, we use the inverse FFT to calculate the function values of the density on an equidistant grid. Then, we calculate the square root of all function values. Based on these values, we can use the FFT to approximate the Fourier coefficient vector representing the square root of the density.

B. New Prediction Steps for Arbitrary Transition Densities

Finding an exact solution to the Chapman–Kolmogorov equation is only possible in special cases. In this section, we show how to provide an approximate result for densities given in a Fourier series representation. Instead of approximating $f_t^w(x_{t+1} - x_t)$ or its square root using a one-dimensional Fourier series, we now represent $f_t^T(x_{t+1}|x_t)$ or its square root using a two-dimensional Fourier series. By calculating the function values of the transition density (or the square root of

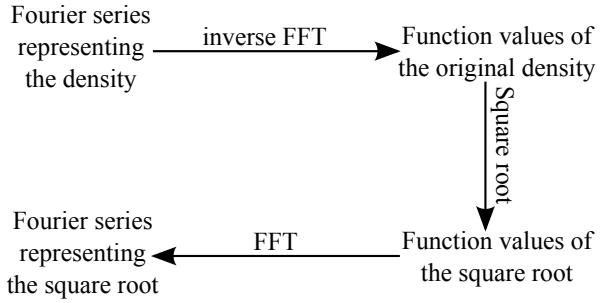


Figure 2. Illustration of the algorithm to approximate the Fourier coefficients of the square root of a density, given the Fourier coefficients of the density.

the function values) on a grid of values of x_{t+1} and x_t, we can use the two-dimensional FFT to approximate the Fourier coefficient matrix. If only a system equation is available, we need to convert it into a transition density, which is, e.g., easily possible for the frequently used case of a nonlinear system model with additive noise (see Sec. II). We will first lay out the procedure for the Fourier identity filter and then explain the changes necessary for the Fourier square root filter.

1) Fourier Identity Filter: The basic idea is to calculate the joint density

$$f_{t+1}^{j}(x_{t+1}, x_t | z_1, \ldots, z_t) = f_t^T(x_{t+1}|x_t) f_t^e(x_t|z_1, \ldots, z_t)$$

and then to obtain the desired predicted density f_{t+1}^p via a marginalization

$$f_{t+1}^{p}(x_{t+1}|z_1, \ldots, z_t) = \int_0^{2\pi} f_{t+1}^{j}(x_{t+1}, x_t|z_1, \ldots, z_t) dx_t \,.$$

We visualize this for an example in Fig. 3. First, we multiply f_t^e shown in Fig. 3a with the transition density f_t^T shown in Fig. 3b. Multiplying these two functions in every point, we obtain the joint density f_{t+1}^j shown in Fig. 3c. In the second step, we marginalize this function to obtain the predicted density f_{t+1}^p that is shown in Fig. 3d.

To numerically calculate the Fourier coefficient matrix $\mathbf{C}^{\text{id}}$ of f_{t+1}^j, we calculate the discrete convolution of the coefficient vector $\underline{a}^{\text{id}}$ of f_t^e with the coefficient matrix $\mathbf{B}^{\text{id}}$ of f_t^T. Afterwards, we marginalize x_t out of $f_{t+1}^j(x_{t+1}, x_t|z_1, \ldots, z_t)$, which is possible very efficiently. We can calculate the Fourier coefficients of f_{t+1}^p with

$$f_{t+1}^{p}(x_{t+t}|z_1, \ldots, z_t) = \int_0^{2\pi} f_{t+1}^{j}(x_{t+1}, x_t|z_1, \ldots, z_t) dx_t$$

from the individual Fourier coefficients c_{k_1,k_2} of $\mathbf{C}^{\text{id}}$ by using

$$\begin{aligned}
&\int_0^{2\pi} \sum_{k_2=-k_{\max}}^{k_{\max}} \sum_{k_1=-k_{\max}}^{k_{\max}} c_{k_1,k_2} \mathrm{e}^{ik_2 x_{t+1}} \mathrm{e}^{ik_1 x_t} dx_t \\
&= \sum_{k_2=-k_{\max}}^{k_{\max}} \mathrm{e}^{ik_2 x_{t+1}} \sum_{k_1=-k_{\max}}^{k_{\max}} c_{k_1,k_2} \int_0^{2\pi} \mathrm{e}^{ik_1 x_t} dx_t
\end{aligned}$$

and since the integral is always zero for $k_1 \neq 0$

$$\begin{aligned}
&= \sum_{k_2=-k_{\max}}^{k_{\max}} \mathrm{e}^{ik_2 x_{t+1}} c_{0,k_2} \int_0^{2\pi} 1 \, dx_t \\
&= \sum_{k_2=-k_{\max}}^{k_{\max}} 2\pi c_{0,k_2} \mathrm{e}^{ik_2 x_{t+1}} \,.
\end{aligned}$$

Thus, we get n Fourier coefficients of f_{t+1}^p by building a vector out of the entries c_{k_1,k_2} of $\mathbf{C}^{\text{id}}$ for which $k_1 = 0$ and $k_2 \in \{-k_{\max}, \ldots, k_{\max}\}$ hold and then multiplying all entries by 2π.

If $\underline{a}^{\text{id}}$ is a vector containing n Fourier coefficients, we use an $n \times n$ Fourier coefficient matrix for $\mathbf{B}^{\text{id}}$ by default in our implementation and the required convolution of an $n \times 1$ vector with an $n \times n$ matrix is in $O(n^2 \log n)$. The marginalization is possible in $O(n)$. Thus, in total, the prediction step has an asymptotic run time complexity of $O(n^2 \log n)$. Step by step instructions of the algorithm are shown in Algorithm 1.

Remark 1. For the implementation, it is advantageous to combine the convolution with the marginalization. This allows us to calculate fewer entries of the convolution result, yielding a significantly improved run time.

Algorithm 1: Prediction step for the Fourier identity filter. For an efficient implementation, all steps can be combined into one. One simply only calculates the required elements of the convolution result and multiplies them by 2π.

Input: $\underline{a}^{\text{id}} \in \mathbb{R}^{n\times 1}$: Fourier coefficient vector of f_t^e,
$\mathbf{B}^{\text{id}} \in \mathbb{R}^{n\times n}$: Fourier coefficient matrix of f_t^T
Output: $\underline{d}^{\text{id}} \in \mathbb{R}^{n\times 1}$: Approximation of the Fourier coefficient vector of f_{t+1}^p

$\mathbf{C}^{\text{id}} \leftarrow \text{convolve}(\underline{a}^{\text{id}}, \mathbf{B}^{\text{id}})$;
$\tilde{\mathbf{C}}^{\text{id}} \leftarrow \text{truncate}(\mathbf{C}^{\text{id}})$;
$\underline{d}^{\text{id}} \leftarrow 2\pi \cdot \text{getCentralRow}(\tilde{\mathbf{C}}^{\text{id}})^\top$;

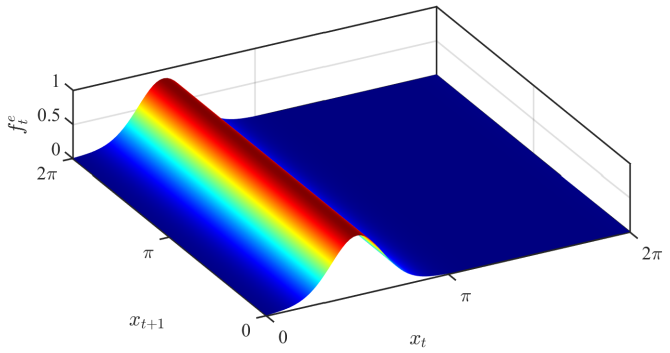
2) Fourier Square Root Filter: For the Fourier square root filter, we ensure the nonnegativity of the pdf in every processing step. For this, we approximate $\sqrt{f_t^T}$ using a two-dimensional Fourier series with coefficient matrix $\mathbf{B}^{\text{sqrt}}$. Afterwards, we use the discrete convolution of the coefficient matrix with itself to obtain a Fourier series representation of f_t^T that is nonnegative for any x_t and x_{t+1}. As in the prediction step for the identity system model, we also square $\sqrt{f_t^e}$ using its coefficient vector. In total, we get the Fourier coefficient matrix for f_{t+1}^j via

$$\mathbf{C}^{\text{id}} = \underline{a}^{\text{sqrt}} * \underline{a}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}} \,.$$

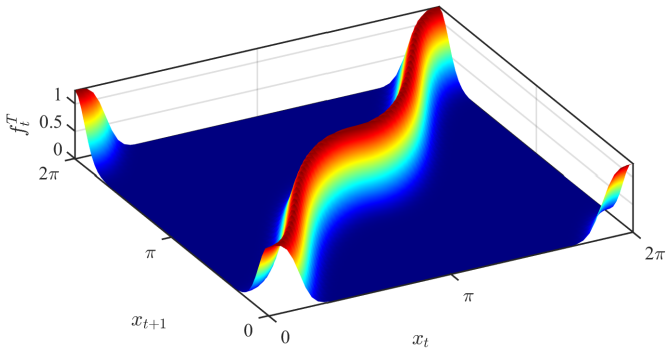
We can now marginalize x_t out of f_{t+1}^j as done in the case of the Fourier identity filter. Afterwards, we can obtain the Fourier coefficients of the square root of the density using the Fourier coefficients of the density via the approach sketched in Fig. 2. The entire algorithm is laid out in Algorithm 2.

Remark 2. For the implementation, we use

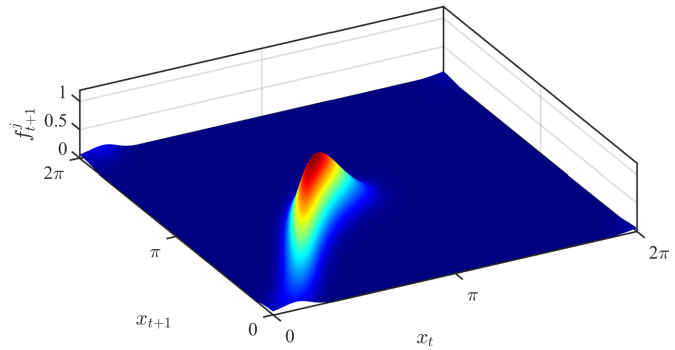
$$\underline{a}^{\text{sqrt}} * \underline{a}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}} = (\underline{a}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}}) * (\underline{a}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}}) \,.$$



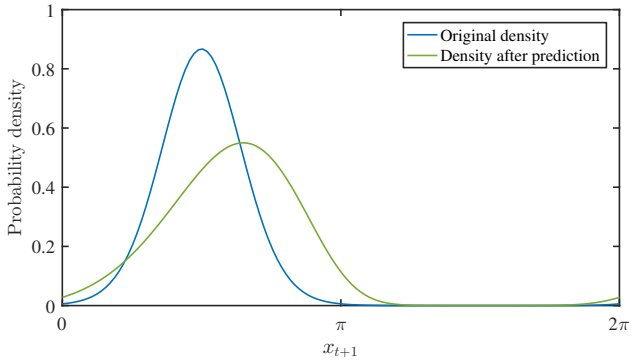
(a) Visualization of $f_t^e(x_t|z_1,\ldots,z_t)$ when plotted as a function of x_t and x_{t+1}. While f_t^e only depends on x_t, it it shown depending on x_t and x_{t+1} to allow for a more comprehensive visualization of how the joint density is generated.



(b) Transition density $f_t^T(x_t|x_{t+1})$.



(c) Joint density $f_{t+1}^j(x_{t+1},x_t|z_1,\ldots,z_t)$.



(d) $f_{t+1}^p(x_{t+1}|z_1,\ldots,z_t)$ as obtained from marginalization in comparison with $f_t^e(x_t|z_1,\ldots,z_t)$ shown in 2D.

Figure 3. Visualization of the densities multiplied for the generation of the joint density, the resulting joint density, and the predicted density for an example with a nonlinear system equation.

Calculating $\mathbf{C}^{\text{id}}$ this way has multiple advantages. First, we only need to calculate two discrete convolutions as the intermediate result $\mathbf{C}^{\text{sqrt}} = \underline{a}^{\text{sqrt}} * \mathbf{B}^{\text{sqrt}}$ can simply be convolved with itself. Second, we can truncate $\mathbf{C}^{\text{sqrt}}$ and still ensure that the Fourier series described by $\mathbf{C}^{\text{id}} = \mathbf{C}^{\text{sqrt}} * \mathbf{C}^{\text{sqrt}}$ is nonnegative. Since the end result has to be truncated anyways, truncating $\mathbf{C}^{\text{sqrt}}$ helps to reduce computational costs. First, because the truncated entries do not have to be calculated and second, because the resulting coefficient matrix $\mathbf{C}^{\text{id}}$ used in future operations is reduced in size.

V. Evaluation

In our evaluation, we compared multiple filters regarding their performance after a single prediction step for a nonlinear system model. We chose the nonlinear system function

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

which has been used for evaluation purposes in [28] as a nonlinear measurement function. Using c, the degree of nonlinearity can be influenced. To evaluate a system function with significant nonlinearity, we chose $c = 2$ for the main part of our evaluation. However, we also did some tests with higher c that lead to similar results. As our system noise, we used an additive noise $\boldsymbol{w} \sim \mathcal{VM}(w; \mu = 0, \kappa = 10)$. The resulting transition density, shown in Fig. 3b, can be seen to have the highest degree of nonlinearity around 0 and π.

We initialized our filter using a von Mises distributed prior

Algorithm 2: Prediction step for the Fourier square root filter. If convolve is followed by truncate or getCentralRow, the discarded entries can be omitted to save computation time.

Input: $\underline{a}^{\text{sqrt}} \in \mathbb{R}^{n\times 1}$: Fourier coefficient vector of $\sqrt{f_t^e}$,
$\mathbf{B}^{\text{sqrt}} \in \mathbb{R}^{n\times n}$: Fourier coefficient matrix of $\sqrt{f_t^T}$
Output: $\tilde{\underline{d}}^{\text{sqrt}} \in \mathbb{R}^{n\times 1}$: Approximation of the Fourier coefficient vector of $\sqrt{f_{t+1}^p}$

$\mathbf{C}^{\text{sqrt}} \leftarrow \text{convolve}(\underline{a}^{\text{sqrt}}, \mathbf{B}^{\text{sqrt}})$;
$\tilde{\mathbf{C}}^{\text{sqrt}} \leftarrow \text{truncate}(\mathbf{C}^{\text{sqrt}})$;
$\mathbf{C}^{\text{id}} \leftarrow \text{convolve}(\tilde{\mathbf{C}}^{\text{sqrt}}, \tilde{\mathbf{C}}^{\text{sqrt}})$;
$\underline{d}^{\text{id}} \leftarrow 2\pi \cdot \text{getCentralRow}(\mathbf{C}^{\text{id}})^\top$;
$\underline{g}^{\text{id}} \leftarrow \text{IFFT}(\underline{d}^{\text{id}})$;
$\underline{g}^{\text{sqrt}} \leftarrow \sqrt{\underline{g}^{\text{id}}}$;
$\underline{d}^{\text{sqrt}} \leftarrow \text{FFT}(\underline{g}^{\text{sqrt}})$;
$\tilde{\underline{d}}^{\text{sqrt}} \leftarrow \text{truncate}(\underline{d}^{\text{sqrt}})$;

$\boldsymbol{x}_0 \sim \mathcal{VM}(x_0; \mu_0, \kappa = 5)$, which we deem to be a reasonably concentrated density before a prediction step. It is important to note that μ_0 has a profound influence on how the system function influences the prediction result. For a more thorough investigation, we performed our evaluation for both $\mu_0 = \frac{\pi}{2}$ and $\mu_0 = \pi$. This can also be seen as evaluating the approach for varying degrees of nonlinearity.

A. Evaluation Methodology

In the evaluation, we compared a total of five approaches. The first two are the Fourier identity filter and the Fourier square root filter, both used with up to 1001 Fourier coefficients. Third, we used a von Mises filter with its nonlinear extension using deterministic sampling [9]. The von Mises filter only aims to match the first trigonometric moment and approximate the result using a von Mises distribution—no additional parameters can be used to match the pdf or cdf more closely. The fourth is the discrete filter already used in the evaluation of [12]. For this filter, we approximate the function using grid points with weightings assigned to them relative to the function value at the respective grid point. The fifth filter evaluated is an SIR particle filter [11] adjusted to the periodic manifold. The last two filters were evaluated using up to 5000 grid points or particles, respectively.

We generated an approximate ground truth using a discrete filter with a significantly higher number (20000) of grid points. While tests showed that the Fourier filters can achieve far better performance in only a fraction of the time, we have decided against using a Fourier filter for approximating the ground truth to prevent any unintended advantages of the Fourier filters in the comparison. While the utilized choice of the ground truth may give the discrete filter a bit of an advantage (the result of a discrete filter using 20000 grid points would be exactly identical to the ground truth used), we believe that this effect does not significantly influence our verdict about the Fourier filters.

To evaluate how closely our result matches the (approximate) ground truth, we calculated the L^2-distance

$$\|F_{\text{filter}} - F_{\text{gt}}\|_2 = \sqrt{\int_0^{2\pi} (F_{\text{filter}}(x) - F_{\text{gt}}(x))^2 \, dx}$$

between the cdf provided by the filter F_{filter} and the cdf F_{gt} used as ground truth. We chose to perform the evaluation using the cdf instead of the pdf due to the lack of an intuitive or established way to convert the result of the particle filter to a pdf. For the cdf, the intuitive way is to cumulate all probability mass represented by the particles. On circular domains, it is necessary to specify a starting point for the integration. Since the comparison of two cdfs is sensitive to shifts of probability mass beyond the starting point of integration, we have chosen $\mu + \pi$—a point of low probability density—as the starting point.

To be able to evaluate the approaches fairly, we also measured the run time. This is not only important for comparing the Fourier approaches with other approaches, but also for a more comprehensive comparison of the Fourier filters with each other. In comparisons performed in [13], the Fourier square root filter showed better estimation results than the Fourier identity filter but was slower when compared on a per coefficient basis. Thus, neglecting the risk of negative values and the theoretical disadvantages, the Fourier identity filter can perform better when comparing configurations of equal run time. All filters were implemented in Matlab and the source code is available as part of libDirectional [29], a library for directional estimation. We encourage readers to evaluate the filters on their own nonlinear filtering problems on the circle. All run times were measured on a laptop with an Intel Core i7-5500U CPU, $12\,\text{GB}$ of RAM, running Windows 10 and Matlab 2016a.

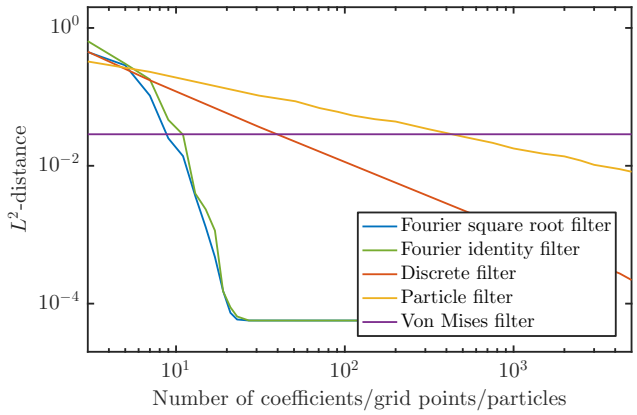
We performed 200 evaluation runs of all filters. As the Fourier filters, the von Mises filter, and the discrete filter are deterministic and return the same result in each run, multiple runs were performed merely to allow for reliable statements about the run times by calculating the average. For the particle filter, the L^2-distance was calculated in each run and the average over all runs of each configuration was used for comparison with the other filters.

B. Evaluation Results

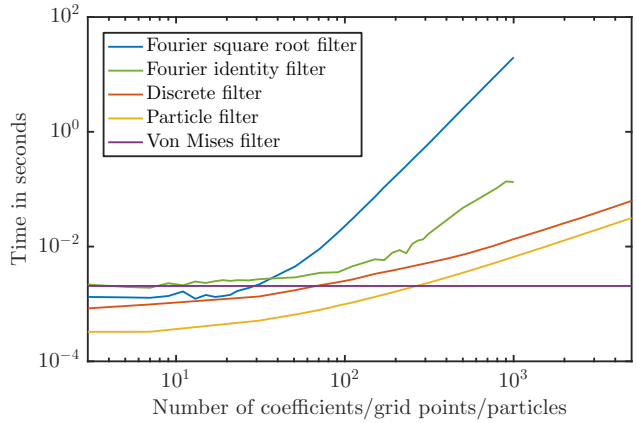
The resulting L^2-distances and run times are shown using logarithmic scales for $\mu_0 = \frac{\pi}{2}$ in Fig. 4 and for $\mu_0 = \pi$ in Fig. 5. As can easily be seen, the results for the chosen two μ_0 (and thus for varying degrees of nonlinearity) differ only little. For the run times, the results are so remarkably similar that even some unexpected non-monotonic parts are present in both evaluated scenarios. The most prominent difference between the two scenarios is the quality of the von Mises filter which can handle the second scenario better as the resulting predicted density is symmetric.

In both scenarios, we can see that all filters (except the von Mises filter) show convergence to the approximation of the ground truth. The reason why the Fourier filters stop improving is that the evaluation is limited by the accuracy of the approximation of the ground truth used. Once the filter results become significantly closer to the actual ground truth than the approximation of the ground truth used, the improvement will not show when calculating the distance to the approximation.

To validate this supposition, we did an additional test for the Fourier filters. We defined a function that calculates the function value of f_{t+1}^p from f_t^T and f_t^e (for $\mu_0 = \frac{\pi}{2}$) via numerical integration for every x_{t+1} the function is evaluated at. We then used this function as a ground truth and calculated the L^2-distance to the result of the Fourier square root filter via numerical integration. Despite obvious limitations of the numerical integration performed by Matlab, we believe the calculated L^2-distances in the order of 10^{-8} for 101 coefficients and in the order of 10^{-12} for 1001 coefficients indicate that further improvement is achieved by the Fourier filters. While we deem the ground truth using numerical integration to be more precise, we were unable to use it as the ground truth for the entire evaluation as this would have resulted in very long run times, rendering us unable to evaluate all filters for

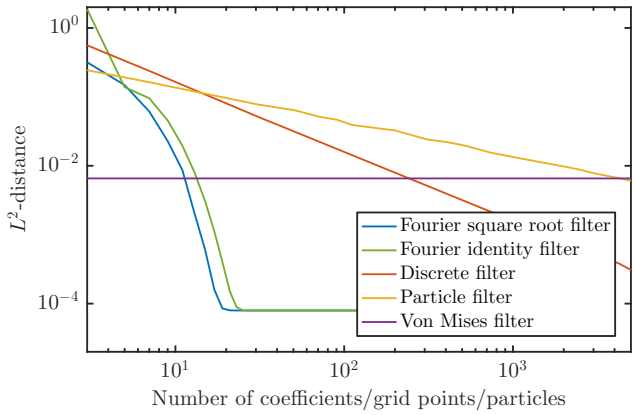


(a) L^2-distance between the cdfs.

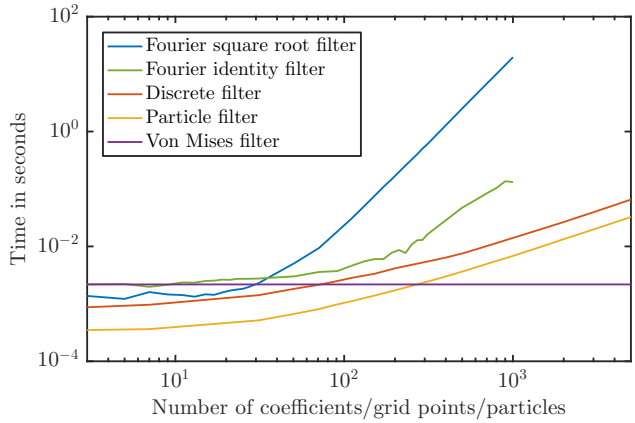


(b) Time taken by the individual filters.

Figure 4. Results for $\mu_0 = \frac{\pi}{2}$ for the individual filters using different configurations shown on logarithmic scales. Only one configuration is possible for the von Mises filter and it is therefore shown as a straight line.



(a) L^2-distance between the cdfs.



(b) Time taken by the individual filters.

Figure 5. Results for $\mu_0 = \pi$ for the individual filters using different configurations shown on logarithmic scales. The von Mises filter is shown as a straight line again.

that many configurations and to average over that many runs, which is particularly important for the particle filter.

In both scenarios, the discrete filter shows faster convergence than the particle filter. While the particle filter is faster with a comparable quality for low numbers of particles, it performs significantly worse for higher numbers of particles. Compared on a run time basis, the discrete filter clearly outperforms the particle filter for higher numbers of grid points. It is also important to note that the quality depicted for the particle filter is an average. Due to the nondeterministic behavior of the particle filter, the resulting estimation quality varies, especially when using only few particles. When comparing the von Mises filter with the configurations of the discrete filter and the particle filter of comparable run time, the von Mises filter outperforms both in the second scenario and loses only to the discrete filter in the first scenario.

When comparing the Fourier filters with the other approaches for very low number of coefficients, the particle filter and discrete filter achieve a comparable estimation quality with lower run times. However, the Fourier filters converge rapidly and reach a quality that we deem to be optimal for the utilized ground truth for about 23 to 27 Fourier coefficients. Using such low numbers of coefficients, both Fourier filters provide a better approximation of the cdfs than the discrete filter does using 5000 grid points and the particles filter does using 5000 particles while featuring a run time that is lower by over one order of magnitude. Based on this, we assess both Fourier filters to be clearly superior to the other filters evaluated. However, it is hard to declare a clear winner among the two Fourier filters. For the very low numbers of coefficients that are necessary in this scenario, the run time and quality of the Fourier square root filter is very similar to that of the Fourier identity filter. Based on this, we recommend using the Fourier square root filter due to its advantageous property of providing a valid density in every time step. For multivariate estimation problems, a trade-off may arise between the run time advantage of the Fourier identity filter and the advantageous properties of the Fourier square root filter, as was observed for the multivariate Fourier filters using the prediction steps for identity models with additive noise [13].

VI. Conclusion

The novel prediction steps for nonlinear system equations proposed in this paper make our Fourier filters applicable to a variety of new scenarios. While the nonlinear prediction step of the Fourier square root filter ensures a valid probability density, the prediction step of the Fourier identity filter features lower run times for high numbers of coefficients. Thus, our proposed approaches add versatility to the Fourier filters while preserving the strengths of the individual filters. We observed promising performance that is reminiscent of the results we have previously obtained when evaluating the Fourier filters over multiple time steps [12], using the filter steps for arbitrary likelihoods and the prediction steps for identity models with additive noise. We see this as an indicator that the new prediction scheme can handle nonlinearity well and allows for similar estimation results as the alternative for identity system models, only at the cost of a moderate increase in run time.

In the evaluated scenarios, the Fourier filters show rapid convergence. Even in configurations in which the Fourier filters outperform the expensive approximation of the ground truth used, they are still fast enough for real time applications. All in all, the Fourier filters surpass the other filters when comparing the quality of the cdf for configurations of comparable run time. While the other filters also showed performance suitable for most real-time applications, the differences in run times for results of comparable estimation quality are expected to scale with the number of variates, allowing only for the use of the most efficient filters for multivariate estimation problems in real time applications.

One possible area of future work is to inspect the performance of the filters more closely, e.g., with regard to the cyclic version of the Bayesian Cramér–Rao lower bound [30]. Furthermore, inspecting and evaluating other possible transformations than the square root may lead to further insights and to new variants of the filter that are particularly suited to certain classes of scenarios.

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.

References

[1] R. Fisher, "Dispersion on a Sphere," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 217, no. 1130, pp. 295–305, 1953.

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

[3] K. Kunze and H. Schaeben, "The Bingham Distribution of Quaternions and its Spherical Radon Transform in Texture Analysis," *Mathematical Geology*, vol. 36, pp. 917–943, 2004.

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

[5] J. A. Carta, C. Bueno, and P. Ramírez, "Statistical Modelling of Directional Wind Speeds Using Mixtures of von Mises Distributions: Case Study," *Energy Conversion and Management*, vol. 49, no. 5, pp. 897–907, May 2008.

[6] A. S. Willsky and J. T. Lo, "Estimation for Rotational Processes with One Degree of Freedom—Part II: Discrete-time Processes," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 22–30, Feb. 1975.

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

[8] 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.

[9] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Bayesian Filtering in Circular State Spaces," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, pp. 70–87, Mar. 2016.

[10] 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.

[11] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[12] 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.

[13] ——, "Multivariate Angular Filtering Using Fourier Series (submitted)," *Journal of Advances in Information Fusion*.

[14] D. Brunn, F. Sawo, and U. D. Hanebeck, "Efficient Nonlinear Bayesian Estimation Based on Fourier Densities," in *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sep. 2006.

[15] ——, "Nonlinear Multidimensional Bayesian Estimation with Fourier Densities," in *2006 45th IEEE Conference on Decision and Control*, Dec. 2006.

[16] A. S. Willsky, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part I: Analysis," *IEEE Transactions on Information Theory*, 1974.

[17] ——, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part II: Implementation," *IEEE Transactions on Information Theory*, 1974.

[18] J. J. Fernández-Durán, "Circular Distributions Based on Nonnegative Trigonometric Sums," *Biometrics*, vol. 60, no. 2, Jun. 2004.

[19] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 1999.

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

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

[22] S. Kullback, *Information Theory and Statistics*. Dover, 1978.

[23] G. Kurz and U. D. Hanebeck, "Trigonometric Moment Matching and Minimization of the Kullback–Leibler Divergence," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 3480–3484, Oct. 2015.

[24] Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd ed. Cambridge University Press, 2004.

[25] 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.

[26] P. Duhamel and M. Vetterli, "Fast Fourier Transforms: A Tutorial Review and a State of the Art," *Signal Processing*, vol. 19, no. 4, pp. 259–299, 1990.

[27] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Prentice Hall, 1984.

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

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

[30] E. Nitzan, T. Routtenberg, and J. Tabrikian, "Cyclic Bayesian Cramér–Rao Bound for Filtering in Circular State Space," in *18th International Conference on Information Fusion (Fusion)*, 2015, pp. 734–741.