

Multivariate Angular Filtering Using Fourier Series

FLORIAN PFAFF
GERHARD KURZ
UWE D. HANEBECK

Filtering for multiple, possibly dependent angular variates on higher-dimensional manifolds such as the hypertorus is challenging as solutions from the circular case cannot easily be extended. In this paper, we present an approach to recursive multivariate angular estimation based on Fourier series. Since only truncated Fourier series can be used in practice, implications of the approximation errors need to be addressed. While approximating the density directly can lead to negative function values in the approximation, this problem can be solved by approximating the square root of the density. As this comes at the cost of additional complexity in the algorithm, we present both a filter based on approximating the density and a filter based on approximating its square root and closely regard the trade-offs. While the computational effort required for the filters grows exponentially with increasing number of dimensions, our approach is more accurate than a sampling importance resampling particle filter when comparing configurations of equal run time.

Refereeing of this contribution was handled by William Blair.

Authors' address: Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany (E-mail: florian.pfaff@kit.edu gerhard.kurz@kit.edu uwe.hanebeck@ieee.org).

I. INTRODUCTION

Periodic quantities are ubiquitous both in nature [1], [2], [3] and technology [4], [5], [6]. The most common periodic quantities are in the form of angles, such as the orientation in a two-dimensional space, but a variety of other periodic quantities exist, such as the phase of a signal [7], [8]. When dealing with orientations, we are usually only interested in the current orientation in our coordinate system and do not aim to count how often the object has revolved around the axis of rotation. While neglecting the latter simplifies the task, estimators need to be carefully crafted to properly account for the effects of periodicity.

For recursive Bayesian estimation, uncertainties in the system and measurement models have to be represented, e.g., via transition densities and likelihoods. Filters on linear domains that assume the support of the prior and posterior densities to be unbounded, such as the Kalman filter, have underlying assumptions that are incompatible with periodic manifolds. However, most periodic domains are locally similar to linear ones when regarding a very narrow region of the domain. Since the importance of a region to the estimation problem strongly depends on the probability mass in the region, densities that are concentrated on a very narrow region can still be handled with sufficient accuracy using approaches that rely on the linearity of the domain. Therefore, problems featuring very little uncertainty can still be handled properly using a modified Kalman filter or an unscented Kalman filter [9].

However, the wider the probability mass is spread on the domain, the less the estimation problem behaves like on a linear domain. For higher uncertainties, filters assuming linearity of the domain degrade and can become entirely misleading. In these cases, approaches based on directional statistics specifically crafted for periodic domains become a necessity. Directional statistics [10], [11] puts the focus on properly handling periodic manifolds by providing many analogues to concepts that are commonly used on linear manifolds. Fields in which directional statistics is applied include, e.g., geosciences [1], biology [2], [12], analysis of crystal structures [13], scattering theory [14], MIMO radar systems [15], robotics [4], and signal processing [5], [6], [16].

The main focus of directional statistics is on two classes of topologies. One of the two classes is the

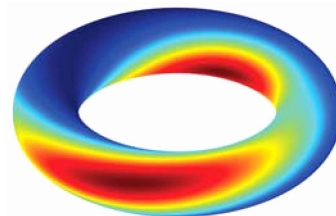


Fig. 1. A mixture of two bivariate wrapped normal distributions shown as a heatmap. The two modes are shown in dark red.

hypersphere S^d , which is the surface of the unit ball in \mathbb{R}^{d+1} . The other important topology is the Cartesian product of multiple circles $S^1 \times \dots \times S^1$ called hypertorus, which will be the focus of this paper. The hypertorus is the proper topology to use when dealing with multiple, possibly stochastically dependent angles in the range of $[0, 2\pi)$. Examples featuring a hypertoroidal topology include angles at different points in time or connected rotatory robotic joints. If the variates of the random vector are stochastically independent, they can be treated as independent random variables and filters for circular topologies can be used. For the circle, several filters have been proposed [17], [18], [19], [20], [21], [22]. However, filters for multivariate angular problems are necessary if the random vector cannot be separated into multiple independent random variables. For example, if the random variables \mathbf{x} and \mathbf{y} are independent, the two variates of the vector $[\mathbf{x} \ \mathbf{x} + \mathbf{y}]^T$ are (usually, but not necessarily) dependent. They can, however, be easily transformed into a problem featuring independent variates that can be estimated independently. The obtained estimates can then be transformed back for use in the original problem featuring dependent variates. In practice, transforming the problem in a way that causes the components to be independent is usually nontrivial and frequently impossible.

On linear domains, an important filter is the Kalman filter that scales well with increasing number of variates and yields optimal estimation results if certain conditions are met. The Kalman filter can handle correlations and scales no more than cubically in the number of variates. For nonlinear system or measurement models, such an efficient and optimal solution does not exist in general. Therefore, even on linear domains, nonlinear estimation is a challenging task, especially for multivariate problems. On periodic domains, estimation problems are inherently nonlinear and thus challenging to deal with for higher numbers of variates.

For multivariate angular estimation, a recursive filter has been suggested in [23] for the two-dimensional torus. However, due to the approximation with a single bivariate wrapped normal distribution, multimodal bivariate posteriors, such as the one shown in Fig. 1, cannot be modeled adequately. As an alternative for toroidal and hypertoroidal estimation problems, the very general concept of a sampling importance resampling (SIR) particle filter [24] that is popular on linear domains can be adapted in a straightforward manner to periodic manifolds.

In this paper, we generalize a recursive Bayesian estimator based on Fourier series, which we proposed for univariate densities on the circle in [21]. Our new approach to multivariate angular estimation problems on the hypertorus is based on representing the density or its square root using a multidimensional Fourier series. Based on the representation used, we either call the filter the (angular) *Fourier identity filter* or the (angular)

Fourier square root filter and we abbreviate their names as IFF and SqFF. We show that the prediction and filter operations can be performed in computationally efficient ways, never exceeding an asymptotic complexity of $O(n \log n)$ for n Fourier coefficients and a fixed number of variates. However, for estimation problems with a higher number of variates, it is advisable to use more Fourier coefficients.

The rest of this paper is structured as follows. We give a brief explanation on how to perform recursive Bayesian estimation in general and lay out the required operations in the next section. We introduce the basics of Fourier series and directional statistics that are a prerequisite for our proposed approach in the third and fourth section. In the fifth section, we address related work and explicate the key idea of our proposed filters. In Sec. VI, we introduce an actual implementation of the Fourier filters. A comparison of our proposed filters and an evaluation comparing the Fourier filters with other approaches is given in Sec. VII. A conclusion and an outlook is provided in Sec. VIII. Finally, we present some useful properties of the Fourier series representations in the appendix.

II. RECURSIVE BAYESIAN ESTIMATION

While a lot of effort in estimation theory is geared towards estimating the mean of the posterior density, keeping track of the whole density is usually required for accurate results over multiple time steps. The focus on approximating the mean can be explained by the fact that on linear domains, the mean of the posterior density is the minimum mean squared error estimator [25, Sec. 10.3]. However, if the density is to be reused in future time steps, the mean of the posterior density itself does in general not suffice for an accurate calculation of the mean of the posterior density in future time steps. In the case of linear systems with Gaussian noise on linear domains, keeping track of the whole density is easy—the resulting Gaussian posterior density is precisely described by the mean vector and the covariance matrix.

For noise terms that are not exactly Gaussian distributed and especially for nonlinear models, describing the precise posterior density using a limited number of parameters is challenging or impossible even on linear domains. Many estimators focus on approximating the posterior density using a parameterized density of a prespecified family of densities. For example, Gaussian assumed density filters [26], [27], [28], [29] try to find a suitable Gaussian approximation for the true posterior density. To increase the accuracy of the approximation, Gaussian mixtures [26] can be used but mixtures entail further problems concerning component reduction [30]. Another popular approach is to use an SIR particle filter [24]. While simple particle filters can be used to asymptotically approximate moments such as the mean over multiple time steps, they are non-deterministic and do not directly provide a continuous approximation of the density.

In the subsections of this section, we introduce the general formulae that can be used for recursive Bayesian estimation with given likelihoods and transition densities and lay out the required operations for implementing a recursive Bayesian estimator. The formulae presented are suitable for manifolds with a topological group structure, such as \mathbb{R}^n and the hypertorus.

A. Prediction Step

The prediction step can be described by the Chapman–Kolmogorov equation

$$\begin{aligned} f_{t+1}^p(\mathbf{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t) \\ = \int_{\Omega_x} f_t^T(\mathbf{x}_{t+1} | \mathbf{x}_t) f_t^e(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_t) d\mathbf{x}_t, \end{aligned}$$

in which Ω_x denotes the sample space—for hypertori, $[0, 2\pi)^d$ — f_t^T the transition density, f_t^e the posterior density, and f_{t+1}^p the prior density based on measurements up to the time step t . For identity models involving additive noise on linear domains, we can write

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{w}_t, \quad (1)$$

with \mathbf{w}_t distributed according to f_t^w . On periodic domains, this becomes

$$\mathbf{x}_{t+1} = (\mathbf{x}_t + \mathbf{w}_t) \bmod 2\pi,$$

which includes a nonlinear transformation. In the case of additive noise, we can simplify the formula for f_{t+1}^p to

$$\begin{aligned} f_{t+1}^p(\mathbf{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t) \\ = \int_{\Omega_x} f_t^w(\mathbf{x}_{t+1} - \mathbf{x}_t) f_t^e(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_t) d\mathbf{x}_t \\ = (f_t^w * f_t^e)(\mathbf{x}_{t+1}) \end{aligned}$$

for both linear and circular domains and are able to perform the prediction step by calculating a continuous convolution.

B. Filter Step

If we obtain a measurement and know the corresponding measurement likelihood, we can use Bayes' formula for the filter step. This essential concept can be formulated as

$$\begin{aligned} f_t^e(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_t) &= \frac{f_t^L(\underline{z}_t | \mathbf{x}_t) f_t^p(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})}{\int_{\Omega_x} f_t^L(\underline{z}_t | \mathbf{x}_t) f_t^p(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1}) d\mathbf{x}_t} \\ &\propto f_t^L(\underline{z}_t | \mathbf{x}_t) f_t^p(\mathbf{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1}), \end{aligned}$$

with the likelihood function f_t^L . It is important to note that the denominator is independent of \mathbf{x}_t and can thus be treated as a constant. Since we know that a proper pdf integrates to one, we can ignore the denominator if we have other means to normalize the density.

III. BASICS OF DIRECTIONAL STATISTICS

In this section, we introduce important concepts of directional statistics. In directional statistics, there are

counterparts to many important concepts used in the context of linear domains, some of which are addressed in this section. We always assume that our periodic region has a size of 2π along each dimension. While most formulae given in this paper do not explicitly depend on the precise region used, we say that our periodic quantities are always in $[0, 2\pi)$ in the scalar case and in $[0, 2\pi)^d$ in the d -variate case.

In this paper, we use the von Mises distribution as well as the wrapped normal distribution and generalize the latter to an arbitrary number of variates. We chose to use the multivariate wrapped normal distribution as it can be trivially generalized from its bivariate definition. Another important concept is that of trigonometric moments. While these moments are seldom used on linear domains, they are a useful concept to employ instead of power moments when dealing with periodic densities. We also introduce concepts for describing correlations between the variates of a random vector on hypertoroidal manifolds. For further reading, we recommend the two classic books about directional statistics [10], [11].

A. Von Mises Distribution

The von Mises distribution [10, Sec. 3.5], also called the circular normal distribution [11, Sec. 2.2.4], is a popular circular distribution. A useful property of this distribution is that the product of two von Mises distributions yields an (unnormalized) von Mises distribution again. The density of the von Mises distribution is given by

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

with $I_0(\cdot)$ being the modified Bessel function of the first kind, $\mu \in [0, 2\pi)$ being the location parameter, and $\kappa \geq 0$ describing its concentration.

B. Wrapped Normal Distribution

The wrapped normal distribution [10, Sec. 3.5] can be, visually speaking, obtained by wrapping a Gaussian distribution around the circle and summing up all probability mass at each point. The density is defined as

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

in which we parameterize the density based on the mean $\mu \in [0, 2\pi)$ and the variance σ^2 of the underlying normal distribution. We use the variance instead of the standard deviation to provide a definition that is consistent with the multivariate case introduced in the next subsection.

C. Multivariate Wrapped Normal Distribution

The concept of a wrapped normal distribution can be generalized to higher dimensions, e.g., to the torus as a bivariate wrapped normal (also called wrapped bivariate normal) distribution [11, Sec. 2.3.2]. For the d -variate wrapped normal distribution, we wrap a

d -variate normal distribution onto the d -dimensional hypertorus. This leads to the formula

$$f_{\text{WN}}(\underline{x}; \underline{\mu}, \mathbf{C}_{\text{WN}}) = \sum_{\underline{j} \in \mathbb{Z}^d} \mathcal{N}(\underline{x} + 2\pi \underline{j}; \underline{\mu}, \mathbf{C}_{\text{WN}})$$

for the density of the d -variate distribution with the vector-valued mean $\underline{\mu} \in [0, 2\pi)^d$ and covariance matrix \mathbf{C}_{WN} .

D. Trigonometric Moments

One important concept on linear domains are power moments, usually simply referred to as moments. Moments describe useful properties of the distribution and are important for estimators. As previously mentioned, the first moment of the posterior density is the MMSE estimator on linear domains. Distributions of certain types can be parametrized by some of their moments, e.g., for the Gaussian distribution, the combination of the first moment and the second central moment (or its root) is the most commonly used parameterization.

On periodic manifolds, trigonometric (also called raw) moments feature some of the properties that power moments have on linear domains. The k th trigonometric moment ($k \in \mathbb{N}$) for scalar random variables is given by [11, Sec. 2.1]

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

It is also common practice to write trigonometric moments as vectors instead of as a complex number. For this representation, used for example in [10, Sec. 3.4.1], the parts represented by the real and imaginary part are calculated using separate integrals. However, using Euler's formula, it can be shown that the conversion from one representation to the other is straightforward.

Unlike power moments, trigonometric moments consist of two values—either the real and complex parts or the two components of the vector—and thus, a single moment can describe multiple properties. For example, the first trigonometric moment is not only a measure of the density's position but also of its dispersion. The parameters for some distributions, such as the wrapped normal and the von Mises distribution, can be derived from the first trigonometric moment [18]. As trigonometric moments express a lot about the density, several filters in the circular case approximate trigonometric moments and make use of them. Due to the close relationship of some of the Fourier coefficients to trigonometric moments, the Fourier filters are no exception.

For the k th moment of multivariate densities, we simply stack all k th moments of all variates. Thus, in

the d -variate case, the k th moment is given by

$$\underline{m}_k = \begin{bmatrix} m_{k,1} \\ m_{k,2} \\ \vdots \\ m_{k,d} \end{bmatrix} = \begin{bmatrix} \mathbb{E}(e^{ikx_1}) \\ \mathbb{E}(e^{ikx_2}) \\ \vdots \\ \mathbb{E}(e^{ikx_d}) \end{bmatrix} = \begin{bmatrix} \int_{[0,2\pi)^d} f(\underline{x}) e^{ikx_1} d\underline{x} \\ \int_{[0,2\pi)^d} f(\underline{x}) e^{ikx_2} d\underline{x} \\ \vdots \\ \int_{[0,2\pi)^d} f(\underline{x}) e^{ikx_d} d\underline{x} \end{bmatrix}. \quad (2)$$

E. Circular Mean Direction

The circular mean direction, which can be thought of as an analogue to the linear mean, only describes the density's location and can be calculated from the first trigonometric moment via [10, Sec. 2.2]

$$\mu = \text{atan2}(\mathcal{I}(m_1), \mathcal{R}(m_1)). \quad (3)$$

A useful property on linear domains is the linearity of the expected value, which does not hold for the circular mean direction [11, Sec. 2.2.1]. For the circular mean direction in the d -variate case, we simply calculate the circular mean direction for every component of the moment vector according to (3), yielding

$$\underline{\mu} = \begin{bmatrix} \text{atan2}(\mathcal{I}(m_{1,1}), \mathcal{R}(m_{1,1})) \\ \vdots \\ \text{atan2}(\mathcal{I}(m_{1,d}), \mathcal{R}(m_{1,d})) \end{bmatrix}.$$

F. Angular Correlation Coefficients

A variety of measures of correlation of two angular random variables have been introduced in the literature. Examples include the correlation coefficients by Jammalamadaka and Sarma [31], Johnson and Wehrley [32], and Jupp and Mardia [33]. One of the correlation coefficients is used by the only assumed density filter for bivariate toroidal problems [23] that we know of. As shown in Appendix D, a limited number of Fourier coefficients contain all information necessary to calculate all correlation coefficients mentioned.

Key to this is the close relationship of the correlation coefficients to a certain covariance matrix. In the bivariate case, all of these correlation coefficients can be calculated using entries of the covariance matrix

$$\Sigma = \mathbb{E} \left(\left(\begin{bmatrix} \cos(x_1) \\ \sin(x_1) \\ \cos(x_2) \\ \sin(x_2) \end{bmatrix} - \underline{\mu}_c \right) \left(\begin{bmatrix} \cos(x_1) \\ \sin(x_1) \\ \cos(x_2) \\ \sin(x_2) \end{bmatrix} - \underline{\mu}_c \right)^T \right)$$

with

$$\underline{\mu}_c = \mathbb{E}([\cos(x_1) \quad \sin(x_1) \quad \cos(x_2) \quad \sin(x_2)]^T),$$

which can be calculated via

$$\underline{\mu}_c = [\mathcal{R}(m_{1,1}) \quad \mathcal{I}(m_{1,1}) \quad \mathcal{R}(m_{1,2}) \quad \mathcal{I}(m_{1,2})]^T.$$

Thus, efficient calculation of this matrix allows efficient calculation of all correlation coefficients. It is easy to

extend this covariance matrix to arbitrary multivariate distributions by stacking the terms for all individual variates, yielding

$$\Sigma = \mathbb{E} \left(\left(\begin{pmatrix} \cos(x_1) \\ \sin(x_1) \\ \vdots \\ \cos(x_d) \\ \sin(x_d) \end{pmatrix} - \underline{\mu}_c \right) \left(\begin{pmatrix} \cos(x_1) \\ \sin(x_1) \\ \vdots \\ \cos(x_d) \\ \sin(x_d) \end{pmatrix} - \underline{\mu}_c \right)^T \right)$$

with

$$\underline{\mu}_c = \mathbb{E}([\cos(x_1) \quad \sin(x_1) \quad \cdots \quad \cos(x_d) \quad \sin(x_d)]^T).$$

IV. BASICS OF FOURIER SERIES

In this section, we give a brief introduction to multidimensional Fourier series, the second concept essential to this paper. For details regarding Fourier series, we refer the reader to the two-volume book series about trigonometric series by Zygmund [34] and books about harmonic analysis [35].

A. One-Dimensional Fourier Series

Using Fourier series, it is possible to approximate functions on $[0, 2\pi)$ using complex exponential functions. The set of functions

$$\{e^{ikx} \mid k \in \mathbb{Z}\}$$

is an orthogonal basis that can be used to represent any square-integrable (also called square-summable) complex function defined on $[0, 2\pi)$ using a square-summable sequence of Fourier coefficients [35, Sec. I-5]. Since densities encountered in practice are usually square-integrable, we can write their density $f(x)$ as a Fourier series

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}, \quad (4)$$

where the Fourier coefficients $c_k \in \mathbb{C}$ fulfill

$$\sum_{k \in \mathbb{Z}} |c_k|^2 < \infty.$$

The Fourier coefficients can be calculated from the density according to

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

For real functions, $c_{-k} = \bar{c}_k$ holds [34, Ch. I], causing imaginary parts to cancel out in (4). Furthermore, it is also possible to use real basis functions and coefficients to represent real functions. In this alternative representation, the series in (4) becomes a weighted sum of sine and cosine functions of different frequencies.

B. Higher-Dimensional Fourier Series

The straightforward generalization of one-dimensional Fourier series to the d -dimensional case is to use

functions of the orthogonal system

$$\{e^{i(k_1 x_1 + k_2 x_2 + \cdots + k_d x_d)} \mid \underline{k} \in \mathbb{Z}^d\}$$

to represent functions on the d -dimensional hypercube [34, Ch. XVII] (or in our case hypertorus) $[0, 2\pi)^d$. In the following, we write the basis functions described above using a dot product as $e^{i\underline{k} \cdot \underline{x}}$ and use a vector-valued index $\underline{k} \in \mathbb{Z}^d$ to specify individual entries $c_{\underline{k}}$ of the d -dimensional Fourier coefficient tensor. Using this notation, a multidimensional Fourier series can be written as

$$f(\underline{x}) = \sum_{\underline{k} \in \mathbb{Z}^d} c_{\underline{k}} e^{i\underline{k} \cdot \underline{x}}.$$

The individual Fourier coefficients can then be calculated via

$$c_{\underline{k}} = \frac{1}{(2\pi)^d} \int_{[0, 2\pi)^d} f(\underline{x}) e^{-i\underline{k} \cdot \underline{x}} d\underline{x}. \quad (5)$$

In this paper, our focus is on Fourier series for which only n specific coefficients are nonzero and we only consider sets of indices \mathcal{J} that are subsets of the integer lattice \mathbb{Z}^d with an equal subset of \mathbb{Z} in each dimension. If every subset of \mathbb{Z} ranges from $-k_{\max}$ to k_{\max} in each dimension, the total number of Fourier coefficients is $n = (2k_{\max} + 1)^d$. Similar to the one-dimensional case, $c_{-\underline{k}} = \bar{c}_{\underline{k}}$ holds for real functions.

REMARK 1. The formulae for the Fourier coefficients bear a close resemblance to the trigonometric moments. The k th trigonometric moment \underline{m}_k can be calculated from the Fourier coefficients via

$$\underline{m}_k = (2\pi)^d [c_{-k, 0, \dots, 0} \quad c_{0, -k, 0, \dots, 0} \cdots c_{0, \dots, 0, -k}]^T.$$

Thus, all trigonometric moments and especially the first trigonometric moment required for the calculation of the circular mean direction can be calculated efficiently from the Fourier coefficients. On the other hand, calculating arbitrary Fourier coefficients from the trigonometric moments is, in general, only possible in the one-dimensional case. For higher dimensions, many entries of the Fourier coefficient tensor do not have corresponding entries in the moment vectors as defined in (2). We visualize this for the two-dimensional case in Fig. 2. In Appendix D, we show that there is a relationship between the covariance matrix described in Sec. III-F and other entries of the Fourier coefficient tensor.

V. RELATED WORK AND KEY IDEA

For our approach, it is important to note that many important multivariate densities—such as the multivariate wrapped normal density—are square-integrable and thus lend themselves well to approximations using Fourier series. Mardia also states this observation for the univariate case in his book [10, Ch. 3–4], noting that if the Fourier coefficients are square-summable, the corresponding density is equal to the Fourier series almost everywhere.

Willsky discusses optimal filtering using non-truncated Fourier series with an infinite number of co-

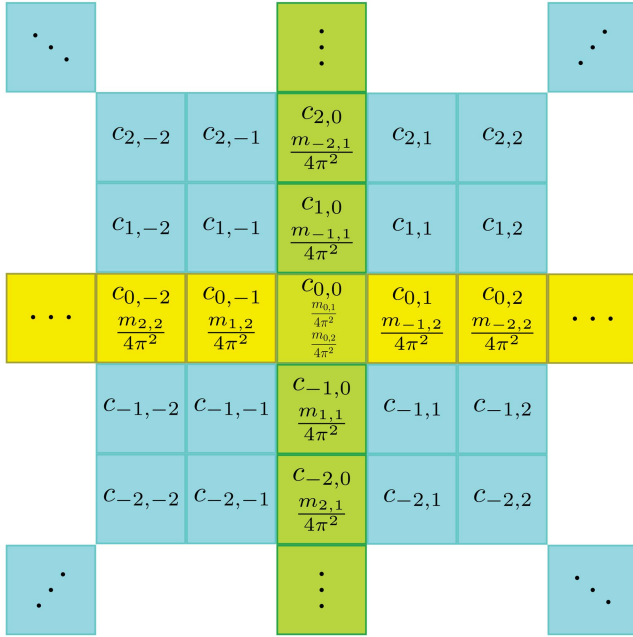


Fig. 2. Visualization of the relationship between the Fourier coefficients and the trigonometric moments of a bivariate distribution. Fourier coefficients that are unrelated to all trigonometric moments are shown in blue, coefficients that are related to the first entry of a trigonometric moment vector are shown in green, and coefficients that are related to the second entry are shown in yellow. The zeroth coefficient in the middle determines both entries of the zeroth trigonometric moment vector and is identical for all (normalized) densities.

efficients in [6]. Since infinite series cannot be handled computationally, he discusses practical implementations in [16]. Willsky deems the performance of a filter working with truncated Fourier series to be insufficient for few coefficients and suggests making the assumption that the density is distributed according to a wrapped normal distribution, which we believe is too restrictive.

Fernández-Durán [36] observes that when approximating densities using Fourier series, the truncation of the coefficient vector can cause negative function values and suggests a computationally expensive way to ensure nonnegativity. In the context of nonlinear filtering for linear domains, Brunn et al. [37], [38] argue that approximating a transformed version of the pdf allows the reconstruction of a valid density with only nonnegative values in every time step.

Based on this idea, we have presented a filter for univariate periodic densities in [21] that ensures the validity of a density function

$$f : [0, 2\pi) \rightarrow \mathbb{R}_0^+$$

by approximating the square root of the density. Approximating the square root $g(x) = \sqrt{f(x)}$ of the density is reasonable as the square root of every density is square-integrable since

$$\int_{[0,2\pi)} g(x)^2 dx = \int_{[0,2\pi)} f(x) dx = 1$$

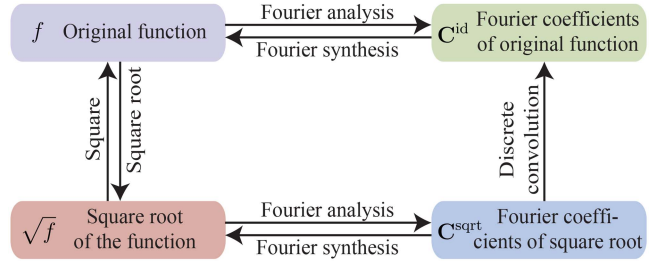


Fig. 3. Different representations that are used by the filters.

holds. Therefore, the Fourier coefficients are square-summable and thus converge to zero, facilitating the approximation via a Fourier series. Furthermore, if $g_{k_{\max}}$ denotes the truncated Fourier series respecting the coefficients from $-k_{\max}$ to k_{\max} , the convergence of

$$g_{k_{\max}} \xrightarrow{k_{\max} \rightarrow \infty} \sqrt{f} \text{ also implies } g_{k_{\max}}^2 \xrightarrow{k_{\max} \rightarrow \infty} f$$

almost everywhere. For this representation, we showed in [21] how prediction and filter steps can be calculated efficiently and accurately using the Fourier coefficients for univariate estimation problems. In a scenario featuring bimodality, the filter presented outperformed both a grid filter and a particle filter.

In this paper, we generalize the approach presented in [21] to higher dimensions and approximate multivariate periodic densities or their square root using multi-dimensional Fourier series. Based on these approximations, we describe how the prediction and filter steps of a recursive Bayesian estimator can be calculated efficiently in the multivariate case. Since run time is crucial for high numbers of variates, we put emphasis on comparing configurations of equal run time. Furthermore, we closely regard the IFF, which has disadvantages in theory but features a lower run time when using an identical number of coefficients.

VI. RECURSIVE BAYESIAN ESTIMATION BASED ON FOURIER SERIES

In this section, we present how a recursive Bayesian estimator can be implemented based on approximating the density or its square root using a Fourier series. For the SqFF, we are dealing with a total of four representations that can be used to obtain values of the probability density (see Fig. 3). These include the original function, the Fourier coefficients of the original function, the square root of the function, and the Fourier coefficients thereof. Only the first two representations are used in the IFF. Deriving a Fourier series representation for a function is commonly referred to as Fourier analysis whereas reconstructing the function is frequently called Fourier synthesis. As the different representations have implications for the algorithmic implementation of the operations, we will make a clear notational distinction between Fourier coefficient tensors representing the actual density \mathbf{C}^{id} and coefficient tensors representing the

square root of the density \mathbf{C}^{sqrt} and apply the same notation to the individual entries of the tensors denoted by c_k^{id} and c_k^{sqrt} with the vector-valued index \underline{k} .

For the upper two representations in Fig. 3, there are important dualities that we make use of. The convolution of two functions corresponds to a Hadamard (entrywise) product of their Fourier coefficient tensors. Furthermore, the multiplication of two functions represented as a d -dimensional Fourier series can be performed using a d -dimensional discrete convolution of the coefficient tensors.

In order to perform (or at least approximate) the prediction and filter steps described in Sec. II, we first need to be able to transform arbitrary densities and likelihoods into the two Fourier series representations. Second, we need to be able to perform multiplications and normalizations for the filter step in the respective representation. Third, convolutions are necessary to perform prediction steps for an identity model with additive noise (1). While exact results for both operations can be obtained in the IFF, an increase in the number of coefficients would be inevitable. While we may allow the number of coefficients to vary over time, parameter reduction becomes inevitable in the long run. Parameter reduction generally induces an approximation and is also necessary for the SqFF. For our implementation of the filters, we do not allow the number of coefficients to vary over time and truncate to an identical number of coefficients after each prediction and filter step.

In the following, we describe how the filter step, the prediction step, and the parameter reduction can be performed in $O(n \log n)$ for n Fourier coefficients. Further properties that are useful to applying the filter in practice but are not essential to the filter and prediction step are given in the appendix.

A. Transforming Multivariate Densities

An efficient approach to Fourier analysis was proposed by Cooley and Tukey [39]. This approach has become widespread for calculating the closely related discrete Fourier transform [40, Ch. 2] and is nowadays known as the fast Fourier transform (FFT) [41], which is also how we will refer to it and its higher-dimensional generalizations for the remainder of this paper.

The complexity of the FFT is $O(n \log n)$ for a total number of n coefficients. However, for a fixed k_{max} , the number of coefficients still grows exponentially with the dimensionality of the space. This is not a serious problem for low numbers of variates and we have verified good filter results with fast run times for densities with up to five variates. To obtain a Fourier series approximation of the square root of a density using the FFT, there is (aside from calculating the square root of each function value) no additional overhead involved. In the one-dimensional case, we were able to derive closed-form formulae for the coefficients for many important univariate densities [21]. In Appendix A of this

paper, we provide the formula for the Fourier coefficients of the multivariate wrapped normal distribution. While closed-form formulae can lead to lower run times, the cost of calculating n coefficients is always at best in $O(n)$ when at least n coefficients are required for an exact representation of the density.

For identity system and measurement models with independent, time-invariant additive noise terms, we can reduce the computational effort involved in obtaining the required Fourier coefficients. In these cases, it is not necessary to transform the density of the system noise and the likelihood in each time step. The density of the system noise simply stays identical while the likelihood is only influenced by the measurement via a shift.

For example, for an additive system noise that is distributed according to a multivariate wrapped normal distribution, the density of the system noise is

$$f_t^w(\underline{w}_t) = f_{\text{WN}}(\underline{w}_t; \underline{\mu}, \mathbf{C}_{\text{WN}})$$

in every time step. Similarly, we can avoid the need for transforming the likelihood multiple times for additive noise terms. If the likelihood is

$$f_t^L(\underline{z}_t | \underline{x}_t) = f_{\text{WN}}(\underline{z}_t; \underline{x}_t, \mathbf{C}_{\text{WN}}) = f_{\text{WN}}(\underline{x}_t; \underline{z}_t, \mathbf{C}_{\text{WN}})$$

we can initially transform the likelihood when setting $\underline{z} = \underline{0}$ (meaning, $f_{\text{WN}}(\underline{x}; \underline{0}, \mathbf{C}_{\text{WN}})$ in our case) and then calculate the Fourier coefficients of the actual likelihood respecting the current measurement \underline{z}_t from these coefficients. The individual Fourier coefficients c_k^{shifted} for a function shifted by \underline{z} can be calculated according to

$$c_k^{\text{shifted}} = c_k e^{-ik \cdot \underline{z}}, \quad (6)$$

which is a straightforward generalization of the shifting operation for the scalar case (Theorem 1.1 (iv) in [34, Sec. II-1]).

B. Filter Step

To implement the filter step, two operations have to be performed for the two filters. The first operation is a multiplication of the prior density $f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})$ and the likelihood $f_t^L(\underline{z}_t | \underline{x}_t)$. The coefficient tensors for the intermediate, unnormalized results will be called $\mathbf{C}_t^{e,\text{id}}$ and $\mathbf{C}_t^{e,\text{sqrt}}$. In the second step, the densities are normalized to yield the coefficient tensors $\mathbf{C}_t^{e,\text{id}}$ and $\mathbf{C}_t^{e,\text{sqrt}}$ to be used in the next prediction (or filter) step. The filter steps for the two Fourier filters are illustrated in Fig. 4 and the necessary operations are explained in detail in the following.

1) **Multiplication of Two Densities:** The first operation necessary to perform the filter step of our Bayesian filter is the multiplication operation. Let us now denote the Fourier coefficient tensor of $f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})$ as $\mathbf{C}_t^{p,\text{id}}$ and refer to the Fourier coefficient tensor of $f_t^L(\underline{z}_t | \underline{x}_t)$ (for a fixed measurement \underline{z}_t , depending only on the state \underline{x}_t) as $\mathbf{C}_t^{L,\text{id}}$. For the IFF, we can then directly obtain

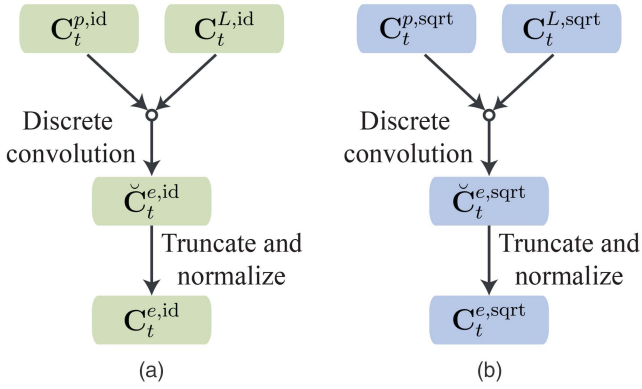


Fig. 4. Filter step of the IFF and the SqFF. (a) Illustration of the filter step of the IFF. (b) Illustration of the filter step of the SqFF.

the coefficient tensor $\check{C}_t^{e,id}$ for the new Fourier series representing the unnormalized multiplication of the two functions via

$$\check{C}_t^{e,id} = C_t^{p,id} * C_t^{L,id},$$

in which $*$ denotes the discrete convolution.

The discrete convolution can be performed in $O(n \log n)$ using FFT-based convolution approaches or by using alternative convolutions methods tailored to the multidimensional tensor convolution [42]. Since the discrete convolution of two tensors results in a larger tensor, parameter reduction as explained in Sec. VI-C becomes necessary.

The multiplication can be performed similarly for the SqFF. Owing to the fact that for all functions $f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})$ and $f_t^L(\underline{z}_t | \underline{x}_t)$ and for all vectors \underline{x}_t and \underline{z}_t in the respective domains

$$\begin{aligned} & \sqrt{f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})} \cdot \sqrt{f_t^L(\underline{z}_t | \underline{x}_t)} \\ &= \sqrt{f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1}) f_t^L(\underline{z}_t | \underline{x}_t)} \end{aligned}$$

holds, we can simply multiply the functions in the square root representation. Thus, the multiplication can be performed analogously to the IFF. For the Fourier coefficient tensors $C_t^{p,sqrt}$ of $\sqrt{f_t^p(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_{t-1})}$ and $C_t^{L,sqrt}$ of $\sqrt{f_t^L(\underline{z}_t | \underline{x}_t)}$ (again, for a fixed measurement), we can calculate the unnormalized coefficient tensor $\check{C}_t^{e,sqrt}$ in the square root representation using

$$\check{C}_t^{e,sqrt} = C_t^{p,sqrt} * C_t^{L,sqrt}.$$

2) Normalization: The second operation necessary for the filter step is the normalization. We use $\check{c}_{t,\underline{k}}$ to refer to an entry of the coefficient tensor to be normalized, such as $\check{C}_t^{e,id}$ or $\check{C}_t^{e,sqrt}$ as obtained from the multiplication above. For the normalization, we need to integrate over the whole domain of the function, which is easy when a Fourier series is used to represent a real function. In the integral, all terms except the one stemming from the first coefficient integrate to zero. This is obvious since the exponential functions for all other

coefficients can be converted (pairwise) into sine and cosine functions of differing (but nonzero) frequencies using Euler's formula. Integrating these trigonometric functions over $[0, 2\pi)$ always yields zero. Therefore, the integral over the whole function can be calculated according to

$$\int_{[0, 2\pi)^d} \sum_{\underline{k} \in \mathbb{Z}^d} \check{c}_{t,\underline{k}}^{id} e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{x} = \int_{[0, 2\pi)^d} \check{c}_{t,\underline{0}}^{id} d\mathbf{x} = (2\pi)^d \check{c}_{t,\underline{0}}^{id}$$

in the non-rooted representation. Thus, we can normalize the density by dividing all coefficients by $(2\pi)^d \check{c}_{t,\underline{0}}^{id}$, which ensures that the coefficient with index $\underline{0}$ of the new, normalized density is $c_{t,\underline{0}}^{id} = 1/(2\pi)^d$.

For the square root version, we can calculate the integral of the function by determining the Fourier coefficient with index $\underline{0}$ of the non-rooted representation. The coefficient $\check{c}_{t,\underline{0}}^{id}$ can be calculated from the coefficients $\check{c}_{t,\underline{k}}^{sqrt}$ of the square root representation via

$$\check{c}_{t,\underline{0}}^{id} = \sum_{\underline{k} \in \mathcal{J}} \check{c}_{t,\underline{k}}^{sqrt} \check{c}_{t,-\underline{k}}^{sqrt} = \sum_{\underline{k} \in \mathcal{J}} |\check{c}_{t,\underline{k}}^{sqrt}|^2.$$

Based on this insight, we can ensure that the density represented by the Fourier series integrates to one by dividing all $\check{c}_{t,\underline{k}}^{sqrt}$ by $\sqrt{(2\pi)^d \sum_{\underline{k} \in \mathcal{J}} |\check{c}_{t,\underline{k}}^{sqrt}|^2}$. Since calculating the sum of the square of the absolute values is in $O(n)$ and the division of all coefficients is also in $O(n)$, normalization is possible in $O(n)$ for both representations.

C. Parameter Reduction

The need for parameter reduction in the Fourier filters is reminiscent of nonlinear filters on linear domains. Except for very simple problems, representing the exact result of both convolution and multiplication operations usually results in an increase in the number of parameters. In the long run, parameter reduction becomes necessary and usually requires approximations. A popular example of nonlinear estimators are estimators based on Gaussian mixtures. For mixtures, component reduction is a non-trivial and expensive operation, in which the density has to be approximated using a lower number of components while preserving the shape of the density as well as possible [30].

Parameter reduction is also essential for our proposed approach using Fourier series. As described in Sec. VI-B.1, the multiplication operation requires calculating a discrete convolution of the coefficient tensors, resulting in an increase in the size of the tensor. As common densities and especially their square roots are square-integrable, the coefficients tend to zero in every dimension. Therefore, the influence of the coefficients on the shape of the function converges to zero for increasing $\|\mathbf{k}\|$ and it is thus reasonable to use truncation as an easy way for parameter reduction.

As can be seen in Sec. VI-B.2, renormalization is necessary after truncation for the SqFF as the integral

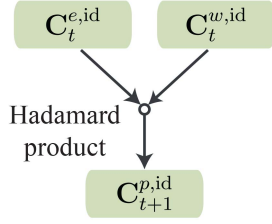


Fig. 5. Prediction step of the IFF.

depends on the sum of the squared absolute values of the coefficients. For the IFF, the integral only depends on $c_{t,0}^{\text{id}}$ and renormalization is thus not required.

D. Prediction Step

To perform the prediction step for an identity system model with additive noise (1), we only need to be able to convolve two densities. Convoluting two functions using their Fourier coefficient tensors is possible very efficiently as this only involves a Hadamard product. The prediction step for nonlinear system models usually requires more than a convolution operation. An approach for arbitrary transition densities that can easily be extended to the multivariate case is laid out in [43].

1) Prediction Step of the IFF: As illustrated in Fig. 5, we can directly use the rule for the convolution of two functions to obtain the Fourier coefficient tensor of the result. If $\mathbf{C}_t^{e,\text{id}}$ and $\mathbf{C}_t^{w,\text{id}}$ are the truncated Fourier coefficient tensors of $f_t^e(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_t)$ and $f_t^w(\cdot)$, we can obtain the coefficient tensor $\mathbf{C}_{t+1}^{p,\text{id}}$ for

$$f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t) = (f_t^e * f_t^w)(\underline{x}_{t+1})$$

via the use of the Hadamard product \odot according to

$$\mathbf{C}_{t+1}^{p,\text{id}} = \mathbf{C}_t^{e,\text{id}} \odot \mathbf{C}_t^{w,\text{id}}.$$

2) Prediction Step of the SqFF: Matters are more complicated when aiming to obtain the coefficient tensor $\mathbf{C}_t^{p,\text{sqrt}}$ representing the square root of the convolution of the densities. Since we intend to obtain the Fourier coefficients for

$$\sqrt{f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t)} = \sqrt{(f_t^e * f_t^w)(\underline{x}_{t+1})},$$

we cannot simply use the Hadamard product of the coefficient tensors $\mathbf{C}_t^{e,\text{sqrt}}$ and $\mathbf{C}_t^{w,\text{sqrt}}$ as this would yield the coefficient tensor for the Fourier series representing

$$\sqrt{f_t^e(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_t)} * \sqrt{f_t^w(\cdot)},$$

which is, in general, unequal to

$$\sqrt{(f_t^e * f_t^w)(\underline{x}_{t+1})}.$$

Instead, we first calculate the Fourier coefficient tensor for $(f_t^e * f_t^w)(\underline{x}_{t+1})$ and then determine the coefficients of the square root, which we lay out in the following and illustrate in Fig. 6. First, we derive the Fourier coefficients for the non-rooted densities $f_t^e(\underline{x}_t | \underline{z}_1, \dots, \underline{z}_t)$

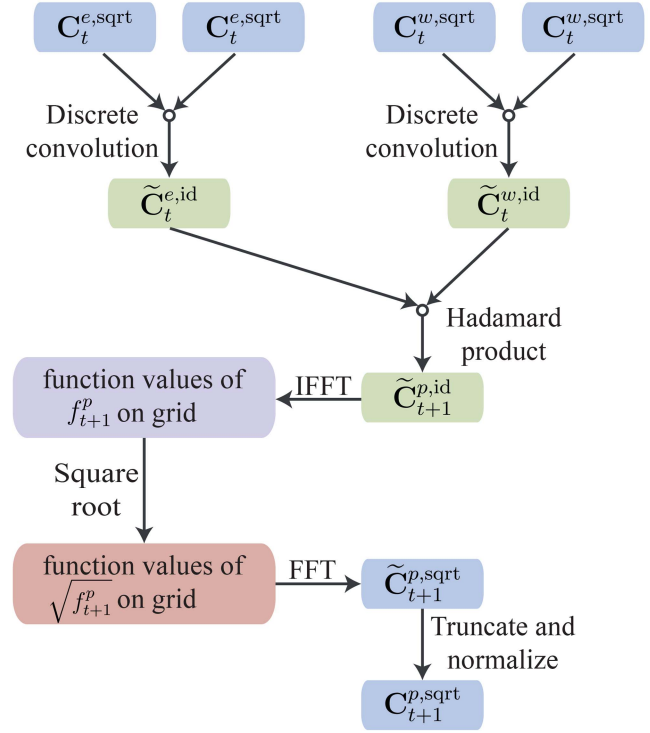


Fig. 6. Prediction step of the SqFF. The colors indicate the respective representation as laid out in Fig. 3.

and $f_t^w(\cdot)$ from the square root representations. These coefficient tensors are denoted by $\tilde{\mathbf{C}}_t^{e,\text{id}}$ and $\tilde{\mathbf{C}}_t^{w,\text{id}}$ to emphasize that they contain more coefficients than the original coefficient tensors $\mathbf{C}_t^{e,\text{sqrt}}$ and $\mathbf{C}_t^{w,\text{sqrt}}$. We can use the multiplication explained in Sec. VI-B.1 to derive the formulae

$$\tilde{\mathbf{C}}_t^{e,\text{id}} = \mathbf{C}_t^{e,\text{sqrt}} * \mathbf{C}_t^{e,\text{sqrt}} \quad \text{and} \quad \tilde{\mathbf{C}}_t^{w,\text{id}} = \mathbf{C}_t^{w,\text{sqrt}} * \mathbf{C}_t^{w,\text{sqrt}}.$$

Based on this, we can use the discrete convolution to obtain the Fourier coefficient tensor $\tilde{\mathbf{C}}_{t+1}^{p,\text{id}}$ for $f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t)$ via

$$\tilde{\mathbf{C}}_{t+1}^{p,\text{id}} = \tilde{\mathbf{C}}_t^{e,\text{id}} \odot \tilde{\mathbf{C}}_t^{w,\text{id}}.$$

As the next step, we wish to obtain the Fourier coefficient tensor for the square root representation with a specified number of coefficients. For this, we first calculate n function values of $f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t)$ on an equidistant grid, take the square root these function values, and then use the FFT to obtain the Fourier coefficient tensor. To obtain the function values on an equidistant grid, we could now naïvely evaluate the Fourier series with coefficient tensor $\tilde{\mathbf{C}}_{t+1}^{p,\text{id}}$ on all n grid points. However, to achieve a good run time behavior, we have to keep one more issue in mind. Unless there is a reason to do otherwise, it is reasonable that the number of coefficients to calculate for the result is linearly dependent on the number of coefficients used in the coefficient tensor $\tilde{\mathbf{C}}_{t+1}^{p,\text{id}}$. Since each function evaluation is in $O(n)$, n function evaluations would be in $O(n^2)$. Therefore, it is significantly cheaper to use the inverse

FFT with a complexity of $O(n \log n)$ to calculate the required n function values.

With the function values of $f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t)$ at our disposal, we can calculate the square root of each value to obtain n function values of $\sqrt{f_{t+1}^p(\underline{x}_{t+1} | \underline{z}_1, \dots, \underline{z}_t)}$. These function values can then be used to approximate n Fourier coefficients via the FFT. Since the most expensive operations are the inverse FFT, the FFT, and the discrete convolution, all of which are in $O(n \log n)$, the total effort is in $O(n \log n)$.

In the step of calculating the square root, approximation errors are caused because taking the square root induces higher frequencies that are not accounted for. Furthermore, as the discrete convolutions used in the calculation of $\tilde{\mathbf{C}}_t^{e,\text{id}}$ and $\tilde{\mathbf{C}}_t^{w,\text{id}}$ result in an increase in the number of coefficients, we need to perform parameter reduction if the coefficients should not be allowed to increase with every prediction step. The parameter reduction can be performed as described in Sec. VI-C. Afterward, a normalization step is required due to the approximations performed.

VII. EXPERIMENTS AND EVALUATION

In this section, we compare our implementations of the two Fourier filters to each other and to other applicable filters. Both filters are available as part of libDirectional [44], a Matlab toolbox for directional statistics with a focus on recursive Bayesian estimation. In Sec. VII-A, we lay out theoretical benefits of the SqFF and describe an experiment to compare the filter steps of the Fourier filters in the univariate case. In the experiment, measurements are simulated regardless of how likely they are. This means we simulate both very likely measurements and unlikely measurements that we would rarely obtain if all underlying assumptions of the filter are correct. For all of these measurements, we evaluate the circular mean direction (3) and the approximation quality of the posterior density provided by the IFF and SqFF.

In the second subsection, we evaluate the error in the form of an angular distance for the two Fourier filters in two bivariate scenarios and one trivariate scenario with additive, wrapped normally distributed noise terms and compare the results with those of other applicable filters in these scenarios. While there is a bivariate wrapped normal filter for toroidal manifolds [23], the only approach to the knowledge of the authors that is applicable to arbitrary multivariate angular estimation problems is the particle filter. Since the number of coefficients used by the Fourier filters and the number of particles used by the particle filter has a major impact on the filter performance, we evaluated several possible configurations. In the scenarios in the second subsection, the likelihoods to be used in each time step can be efficiently calculated using an initial transformation of the likelihood that is shifted according to the measurement obtained.

In the third subsection, we simulate an application of the filters to estimating the angles of a robotic arm based on measurements of the position of the end effector. This application shows that we can use our filter as long as we have a likelihood function and do not require a periodic measurement space. However, this scenario requires more computational effort for the Fourier filters as a Fourier series approximation of the likelihood has to be performed in each time step.

Unlike our evaluation of the circular SqFF in [21], we do not regard the quality of the pdfs and cdfs for the multivariate cases. First, it is difficult to derive a continuous pdf from an SIR particle filter and comparing the cdf becomes increasingly difficult for higher dimensions, especially as a starting point of the integration has to be chosen on periodic domains. Second, we do not have any other filter to numerically approximate the ground truth with at our disposal. In our current evaluation, we also compare the run time of the filters and take the run times into account when assessing the individual filters.

A. Comparing the IFF and the SqFF for Varying Measurements

Allowing the function approximating the density to become negative has many inherent theoretic disadvantages as several useful concepts depend on valid pdfs.¹ Based on an approximation of the pdf with negative function values, no valid cdf can be derived. Furthermore, sampling the density using equally weighted samples may not be possible and sampling schemes such as Metropolis–Hastings sampling [46] do not work. Moreover, some measures of similarity between two densities, such as the Kullback–Leibler divergence [47, Sec. 8.5] and the Hellinger distance [48], cannot be calculated due to the logarithm or square root involved.

In most cases, we observed the approximation quality of the prior density or likelihood function to be already superior for the square root representation when compared with the non-rooted representation with an identical number of coefficients. An example of this is shown in Fig. 7. In the following, we regard the filter step of the Fourier filters and show how both filters perform when comparing the posterior densities and circular mean directions obtained.

Even if no truncation is performed in the filter operations described in Sec. II-B, the approximation errors in the prior density and the likelihood function cause errors in the filter result. The severity of this effect strongly depends on the actual prior density and likelihood function. One important factor influencing the quality of the approximation of the posterior density

¹Negative probabilities [45] can be a viable tool as long as they only appear as intermediate results or if implications are made for other properties that are unobservable simultaneously in the context of physics. In our case, they have no special semantic but stem from approximation errors and have a negative impact on our estimator.

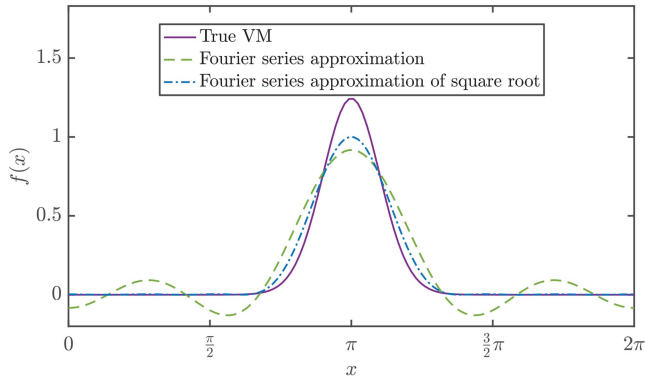


Fig. 7. Von Mises distribution with $\mu = \pi$ and $\kappa = 10$ and corresponding Fourier series approximations using 7 coefficients.

is the overlap of regions of high and low function values of the prior density and the likelihood function.

To understand this effect, we have to take a closer look at the convergence of the Fourier series. A Fourier series converges in the L^2 distance, which is a measure of the squared absolute deviation. Due to the close relationship between the L^2 distance and the Hellinger distance, optimizing the L^2 distance is reasonable when regarding one density individually. However, when regarding the product of two functions, one also has to regard the relative deviation to make statements about the quality of the approximation of the (normalized) multiplication result. Since the Fourier series converges regarding the absolute deviation and not regarding the relative deviation, we expect to see a higher relative deviation in regions of low density than in regions of high density.

This has profound implications for the result of the multiplication of two functions. Let us assume both the prior density and the likelihood are unimodal and are close to zero except for values $\pm\pi/2$ around their modes, such as the von Mises distribution with $\kappa = 10$ that we show in Fig. 7. The prior distribution shall now have a mean of $\mu = \frac{3}{4}\pi$ and the likelihood of $\mu = \frac{3}{2}\pi$. Then, as can be seen in Fig. 8, the relative error of the approximation of the likelihood using 7 coefficients is very high in regions far from the mode, especially when approximating the function directly. When multiplying the prior with the likelihood, the small deviations visible in Fig. 7 are massively amplified in regions of high relative deviation of the likelihood (and also the other way around). This leads to a high deviation from the actual posterior in total. In this example, we can see that the approximation used in the SqFF is advantageous as the relative error is significantly less.

However, there is another effect occurring in the filter step that strongly differs for the two filters. While it is possible to efficiently ensure that a density represented by its Fourier coefficients integrates to one both for the non-rooted and the square root representation (see Sec. VI-B.2), the underlying pdfs that are integrated differ substantially. For the square root representation,

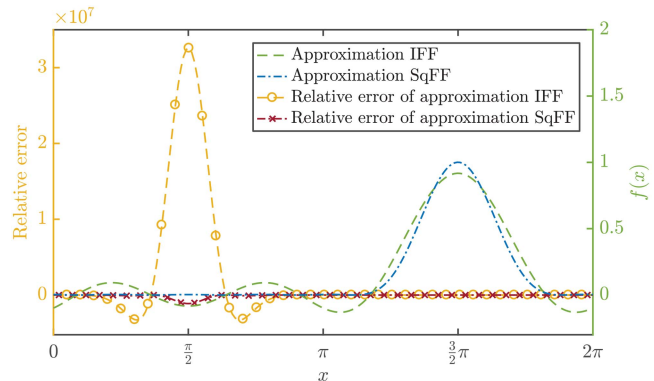


Fig. 8. Approximations of the likelihood and the relative error of the approximations when using Fourier series with 7 coefficients. The likelihood follows a von Mises distribution with $\mu = \frac{3}{2}\pi$ and $\kappa = 10$.

the integral is over nonnegative values, whereas negative function values are possible for the non-rooted version. Always normalizing to one induces a higher total deviation from the abscissa for functions with negative parts as positive and negative parts cancel out. Furthermore, if more than half of the supposed density is negative, normalization to one causes negative parts to become positive and vice versa.

In Fig. 9, we show the filter results for different distances between the modes when the prior density and the likelihood are von Mises distributions with $\kappa = 10$. Both the prior density and the likelihood function were approximated using five Fourier coefficients for both filters and the Fourier series for the posterior densities were also truncated to five coefficients. Fig. 9a and Fig. 9b show the posterior density for the two filters when the modes of the prior density and likelihood function are $\pi/2$ apart. While the normalized result of the Fourier identity filter shows highly negative parts and has the lowest probability density around the peak of the true posterior, the main peak of the SqFF matches that of the true posterior. Furthermore, the result of the SqFF captures the circular mean direction of the posterior density correctly, while the IFF is off by π . Fig. 9c and Fig. 9d show the results when the modes of the two functions are almost π apart. In this case, the true posterior is much flatter than the filter results that bear more of a resemblance to a mixture of the original densities. However, the mean is still correctly captured by the SqFF, while the IFF is off by π .

In Fig. 10, we provide an evaluation depending on the distance α between the two modes of the von Mises distributions. As the first criterion, we evaluate the quality of the posterior density. Since we are unable to use the Kullback–Leibler divergence and the Hellinger distance to compare the result of the IFF to the ground truth, we calculate the total variation [48] between the true densities and the filter results according to

$$d(h_1, h_2) = \int_{[0, 2\pi)} |h_1(x) - h_2(x)| dx,$$

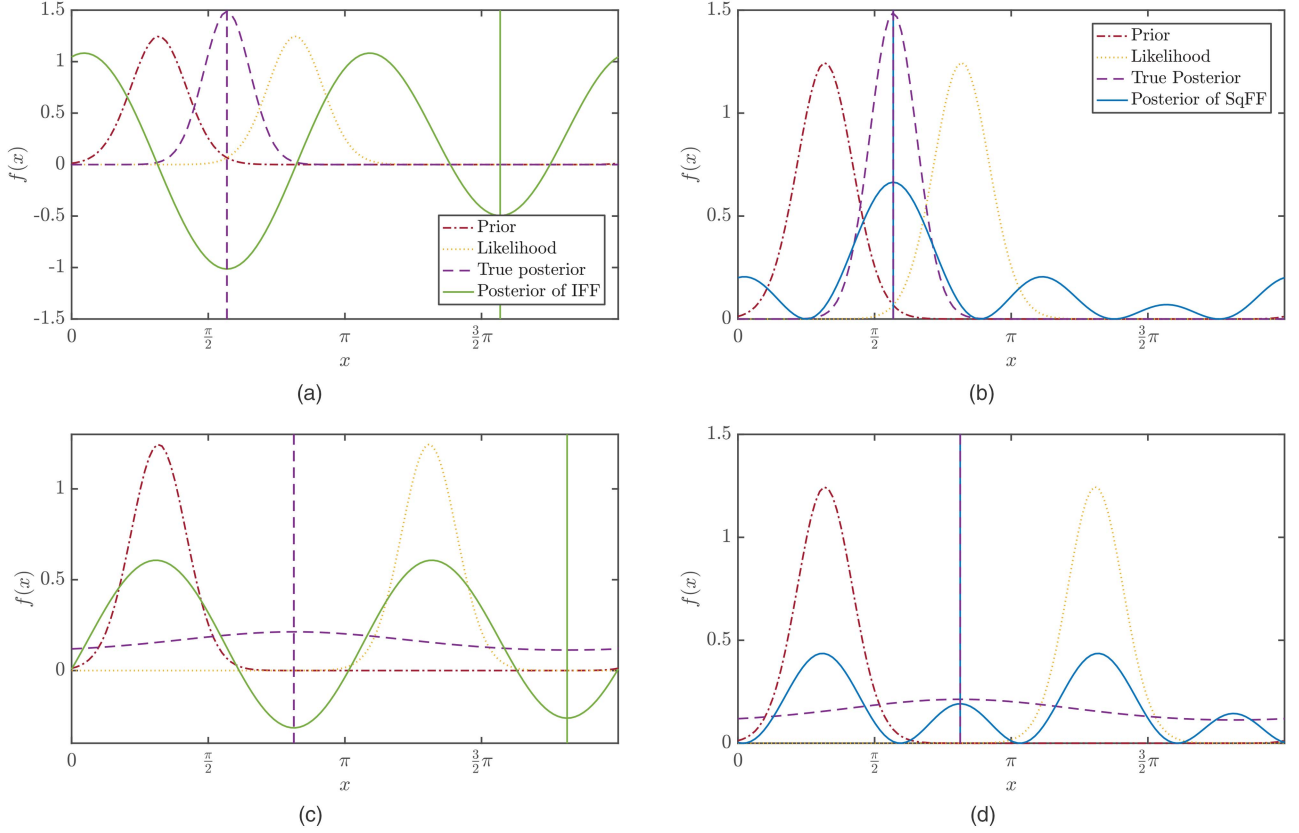


Fig. 9. Examples for posterior densities provided by the two Fourier filters after a single filter step when using five coefficients. The circular mean directions of the posterior densities are shown as a vertical line in the respective color. (a) Result obtained using the IFF when the densities are at a distance of $\pi/2$. (b) Result obtained using the SqFF when the densities are at a distance of $\pi/2$. (c) Result obtained using the IFF when the densities are at a distance of almost π . (d) Result obtained using the SqFF when the densities are at a distance of almost π .

in which $h_1 = f_1 f_2 / \|f_1 f_2\|_{L^1}$ denotes the true density after the filter step and h_2 is the result obtained by the respective variant of the filter. As the second criterion, we regard the quality of the point estimates. For this, we calculate the circular mean direction of the Fourier series approximation from the first trigonometric moment that can be easily obtained as described in Remark 1. Then, we calculate the shorter of the two arc lengths between the true circular mean direction β and the circular mean direction of the filter result γ via [11, Ch. 1.3.2]

$$d_{UV}(\beta, \gamma) = \min(\beta - \gamma, 2\pi - (\beta - \gamma)). \quad (7)$$

A more in-depth discussion of metrics on the circle is given in [49, Sec. 2.2.2].

In the experiment, we limit ourselves to the range $\alpha \in [0, \pi - 0.01]$ as the true posterior for $\alpha = \pi$ is the wrapped uniform distribution that does not have a circular mean direction. The plots clearly show the sensitivity of the filters to larger distances between the modes. The two filters tend to perform worse for $\alpha \in [\pi/2, \pi - 0.01]$ than for $\alpha \in [0, \pi/2]$ both regarding the circular mean direction and the total variation. Comparing the two filters with each other, the total variation between the filter result and the true posterior, as shown in Fig. 10a, can be seen to be higher in most cases when using the IFF. This was to be expected as negative function values are

possible for the result of the IFF. Both the negative parts and the highly positive parts required to cancel out the negative parts result in a higher total deviation. Therefore, deviations that are higher than the usual maximum of the total variation² can be observed, e.g., in the case shown in Fig. 9a. The circular mean direction shown in Fig. 10b is correct for most distances between the means for both filters. This is because the circular mean direction of the result is calculated only from the first trigonometric moment, which is approximated well. Errors in the trigonometric moment usually occur over multiple time steps as an incorrectly shaped density is used for the next filter step, resulting in a wrong circular mean direction. However, as apparent in Fig. 10b, the circular mean direction can also become totally off in a single filter step. Using this criterion, the IFF is also more susceptible to errors than the SqFF. All in all, our results suggest that the SqFF is more robust when unlikely measurements are obtained. Furthermore, we can see that results obtained by the IFF can become totally unlike the true posterior density and can strongly violate the properties of valid densities.

²The total variation usually cannot exceed the value of two for valid densities.

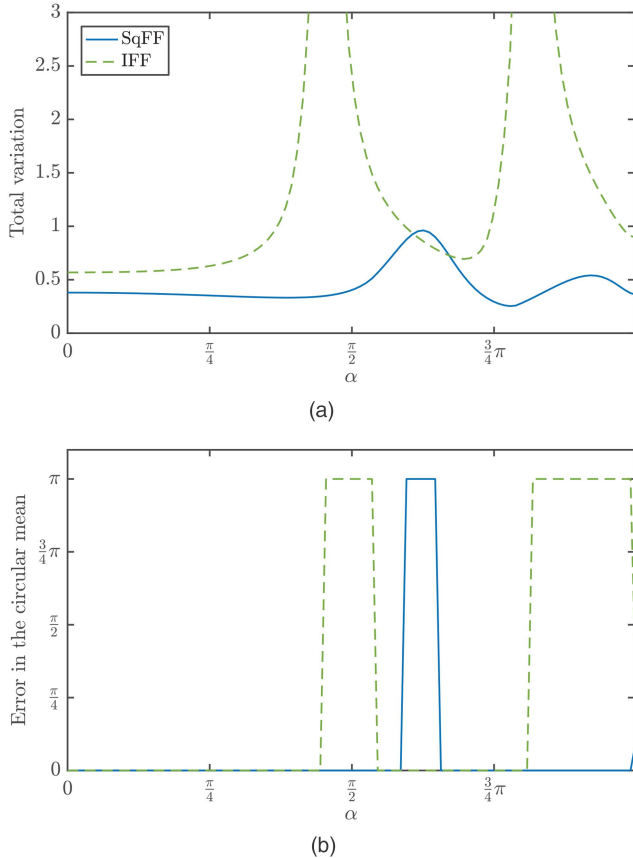


Fig. 10. Comparison of the performance of the IFF and the SqFF depending on the distance between the modes of the functions multiplied. (a) Total variation between the true posterior density and the densities obtained by the two filters. (b) Distance d_{UV} between the true circular mean direction and the circular mean directions provided by the filters.

REMARK 2. Approximating the square root of the density is not the only option. However, using the square root ensures nonnegativity of the reconstructed density while necessitating only few approximations to maintain this representation. An alternative would be to use the logarithm, but the logarithm is highly nonlinear and especially regions with density values close to zero (in which the logarithm approaches $-\infty$) need to be treated with caution. Nonetheless, we plan to consider other transformations than the square root in future work.

B. Evaluation of the Circular Mean Direction

Having compared the two proposed Fourier filters for one filter step in the previous subsection, we now evaluate the Fourier filters against other state-of-the-art approaches to multivariate angular filtering by comparing the estimation performance and run times when multiple filter and prediction steps are performed. In all scenarios, we initialized the filters using an approximation of the actual prior density used for the simulation. We have simulated the entire system behavior and the measurements for 50 time steps and performed alternating prediction and filter steps. A total of 1500 runs

were performed and the true states and the estimation results at each time step were saved for the calculation of our evaluation criterion. As the regarded scenarios feature multivariate estimation problems, we generalize the distance d_{UV} given in (7) to a vector of angles $\underline{\beta}$ and $\underline{\gamma}$ as

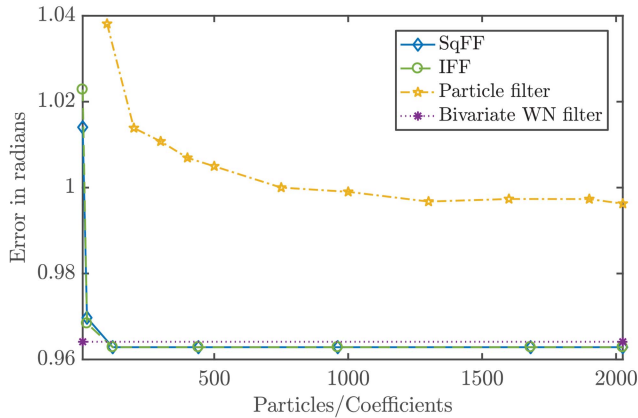
$$d_{MV}(\underline{\beta}, \underline{\gamma}) = \sqrt{\sum_{i=1}^d (\min(\beta_i - \gamma_i, 2\pi - (\beta_i - \gamma_i)))^2}.$$

This distance measure is similar to the Euclidean distance but takes periodicity into account. We calculate the distance between the ground truth vector and the estimate provided by the respective filter as a measure of the error of the estimator. To assess the estimation quality over all time steps and runs, we calculate the average of the errors over all time steps and all runs.

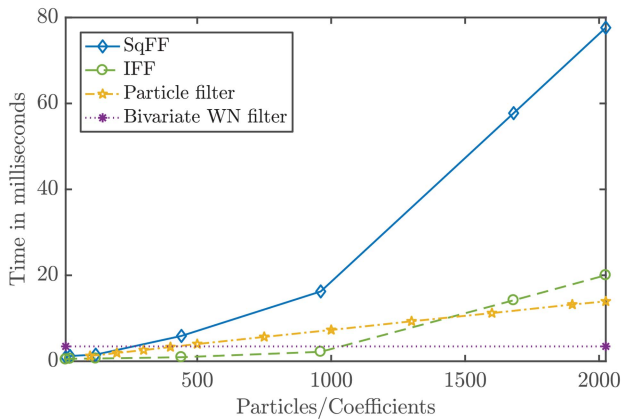
Comparing the two Fourier filters regarding the average error is also an important part of our evaluation. While the results regarding the approximation of the density in the previous subsection were much more promising for the SqFF, the error in the circular mean direction was identical when the modes of the prior density and the likelihood function were close, which is a very common case. Furthermore, as the SqFF requires more convolution and FFT operations in the prediction step, the IFF can be used with more coefficients at an identical run time. Therefore, it is not obvious a priori which Fourier filter performs better when evaluating the average error for configurations of equal run time.

All filters were implemented in Matlab without sophisticated optimizations and were compared on a laptop with an Intel Core i7-5500U processor, 12 GB of RAM, and Matlab 2016b running on Windows 10. The run times given are the average run times for each of the 50 time steps. For all filters, multiple different numbers of parameters were used. As k_{\max} was chosen to be equal in every dimension, the numbers of parameters for the Fourier filter were always odd integers taken to the second (bivariate scenarios) or third (trivariate scenario) power. In all evaluations in this subsection, the likelihood at each time step can be obtained by shifting an initially transformed likelihood and this was used to reduce the effort necessary for the Fourier filters.

In the following, we provide evaluation results for three scenarios. An identity model with additive noise (1) is used as the system and the measurement model in all scenarios. The first scenario features bivariate, unimodal transition densities and likelihood functions on the torus. The measurement and system noises are additive noise terms distributed according to multivariate wrapped normal distributions. This scenario is well suited for the use of the bivariate wrapped normal filter. The second scenario evaluated is also a bivariate scenario but features bimodal likelihoods instead of unimodal ones. The bimodality is introduced by using



(a)

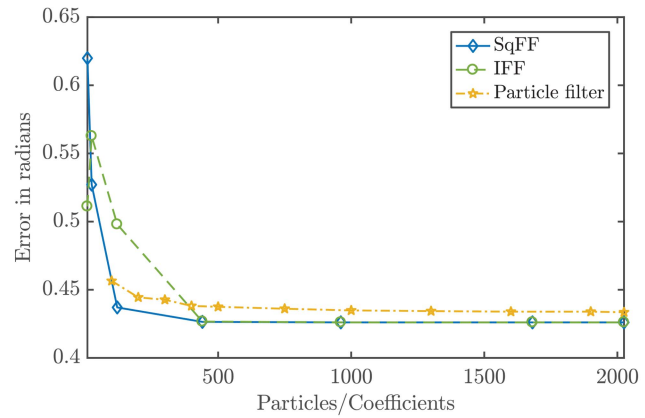


(b)

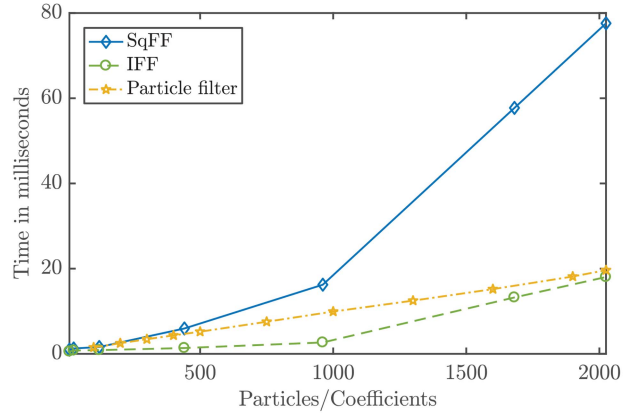
Fig. 11. Average errors and run times for the different filters in the bivariate scenario with unimodal likelihoods. (a) Error of the different filters depending on the number of particles or coefficients used. (b) Run times of the different filters for one time step depending on the number of particles or coefficients used.

a mixture of two multivariate wrapped normal distributions at a distance of one radian in each dimension as additive measurement noise. Both mixture components are centered 0.5 radians away from the actual mean of the density along each axis. Approximating an arbitrary likelihood function using a (potentially unnormalized) wrapped normal distribution is no trivial matter and therefore we have not used the bivariate wrapped normal filter for this scenario and only compared our filters with the particle filter. The third scenario features trivariate densities on the three-dimensional hypertorus. The likelihoods are bimodal again and consist of two multivariate wrapped normal distributions at a distance of one radian in each dimension.

1) Bivariate, Unimodal Scenario: The results shown in Fig. 11 confirm our intuition that the bivariate wrapped normal filter is well suited to this scenario. However, as the results of the Fourier filters show, the performance of the wrapped normal filter can be slightly exceeded by the Fourier filters that approximate the whole posterior density more accurately, facilitating better performance in future time steps. While the Fourier



(a)



(b)

Fig. 12. Average errors and run times for the different filters in the bivariate scenario with bimodal likelihoods. (a) Error of the different filters depending on the number of particles or coefficients used. (b) Run times of the different filters for one time step depending on the number of particles or coefficients used.

filters already appear to reach their optimal accuracy using approximately 100 coefficients, the particle filter does not yet achieve as good of an accuracy for thousands of particles.

When taking the run times into account, the Fourier filters outperform the bivariate wrapped normal filter. The quality of the result of the bivariate wrapped normal filter can be surpassed with numbers of coefficients that still compare favorably in terms of the run time. While fast, the particle filter never comes close to the accuracy of the Fourier filters and the Fourier filters achieve better results using fast configurations with few coefficients. The IFF and the SqFF have almost identical estimation quality, but the IFF is faster. However, the filters achieve a quality that is close to its optimal estimation quality using only approximately 100 coefficients—a configuration at which the two filters differ only little in their run time.

2) Bivariate, Bimodal Scenario: In the results of this scenario, as shown in Fig. 12, the SqFF can be seen to provide better convergence when compared with

the IFF with identical numbers of coefficients. However, one can argue that the IFF is significantly faster and that, when comparing configurations with approximately equal run time, the IFF performs better than the SqFF.

The results of the particle filter do not reach the accuracy of the Fourier filters even for high numbers of coefficients. The particle filter using over 2000 particles is outperformed by the SqFF with less than 500 coefficients in both accuracy and run time. Similarly, this holds for the IFF, whose estimation quality quickly surpasses that of the particle filter and which is significantly faster, resulting in a better performance for configurations with comparable run time.

This scenario shows that the run time performance of the particle filter depends heavily on how fast the likelihood can be evaluated and even a mixture of two components instead of a single wrapped normal distribution can make a clear difference. For the Fourier filters, this does not hold for the identity model with additive noise as the likelihood is only transformed once and shifted afterward.

3) Trivariate, Bimodal Scenario: In the trivariate scenario, the possible configurations of the Fourier filters become more limited as we only use numbers of coefficients that can be written as an odd integer taken to the third power. However, as can be seen in Fig. 13, the superiority of the SqFF over the particle filter is evident. Its performance with corresponding numbers of parameters is better in terms of both the error and the run time, yielding a significant advantage when comparing configurations of comparable run time. Compared on a run time basis, the IFF also outperforms the particle filter since it can handle over 2000 coefficients with a run time that is lower than that of the particle filter with 200 particles.

The bad run time performance of the particle filter is caused by the high computational effort required for evaluating the trivariate wrapped normal density used as the likelihood. As the number of variates increases, the effort involved in evaluating the wrapped normal distribution grows exponentially. To the knowledge of the authors, there is, in general, no other way to calculate the density of a wrapped normal distribution with an arbitrary number of variates other than to sum up some of the addends of the infinite sum. The number of addends required to approximate the density with sufficient accuracy increases exponentially with the number of variates. Thus, the particle filter not only requires more particles for an increasing number of variates, the evaluation of the likelihood also becomes exponentially more expensive. If the filter is used for a large number of time steps, the problem is far less severe for the Fourier filters. For an arbitrary number of time steps, the likelihood and the noise density only have to be approximated once if the likelihood is shifted in the computationally efficient way given in (6) in each time

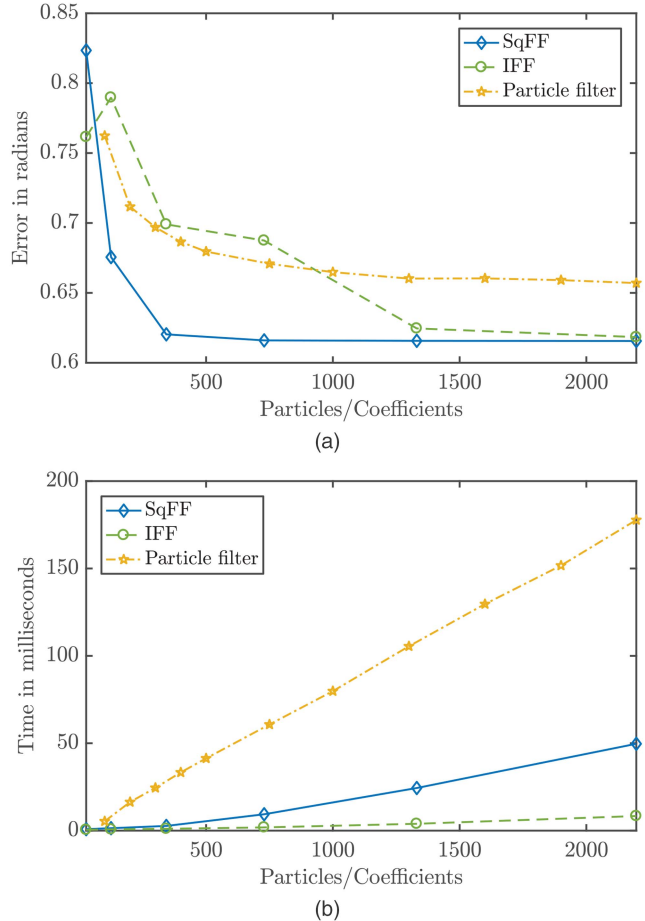


Fig. 13. Average errors and run times for the different filters in the trivariate scenario with bimodal likelihoods. (a) Error of the different filters depending on the number of particles or coefficients used. (b) Run times of the different filters for one time step depending on the number of particles or coefficients used.

step. Therefore, the expensive likelihood only has to be evaluated on a grid once at the beginning.

C. Estimating the Joint Angles of a Robotic Arm

In this subsection, we estimate the joint angles of a robotic arm based on measurements of the position of the end effector that are perturbed by multivariate Gaussian noise. Such a task could, e.g., arise when trying to validate the proper functionality of the robotic arm using external observations. The robotic arm, its joints, and the angles to be estimated are illustrated in Fig. 14. For simplicity, we assume that both joints can move freely and attain any angle. Only point measurements of the point in red on the end effector are obtained. As the end effector can only attain positions on a two-dimensional plane, the measurement is a vector comprising two components. To simplify the measurement equation, we set the zero coordinate of our coordinate system to the center of the first joint indicated in green in Fig. 14.

Given the kinematics of the system, we obtain the measurement equation

$$h(\underline{\alpha}) = \begin{bmatrix} \cos(\alpha_1) \\ \sin(\alpha_1) \end{bmatrix} l_1 + \begin{bmatrix} \cos(\alpha_1 + \alpha_2) \\ \sin(\alpha_1 + \alpha_2) \end{bmatrix} l_2$$

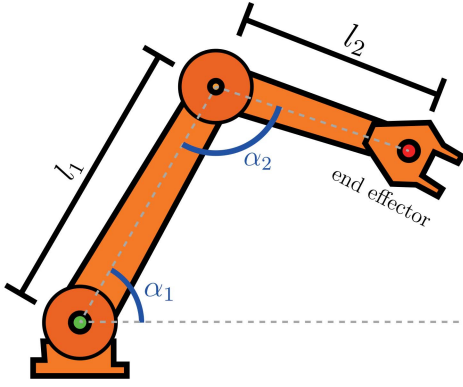


Fig. 14. Illustration of the robotic arm modeled in the simulation.

for $\underline{\alpha} = [\alpha_1 \ \alpha_2]^T$. In our evaluation, we set the joint lengths $l_1 = 2$ m and $l_2 = 1$ m. To simulate sensor noise, the measurements are generated according to

$$\underline{z}_t = h(\underline{\alpha}_t) + \underline{v}_t$$

with a multivariate Gaussian distributed noise term $\underline{v}_t \sim \mathcal{N}(\underline{v}; \underline{\mu}, \mathbf{C})$ with the parameters

$$\underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ m} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix} \text{ m}^2.$$

While the measurement noise is uncorrelated in the measurement space, the likelihood function

$$f^L(\underline{z}_t | \underline{\alpha}_t) = \mathcal{N}(\underline{z}_t; h(\underline{\alpha}_t), \mathbf{C})$$

is, as shown in Fig. 15 for $\underline{z}_t = [0 \ 2.3]^T$ m, asymmetric and the estimation problem thus cannot be trivially split up into univariate problems. Unlike in the previous subsection, the measurement equation is nonlinear and the likelihoods for differing measurements cannot be obtained by shifting an initial approximation of the likelihood. Therefore, a Fourier series approximation has to be performed in each time step, negatively affecting the run times of the Fourier filters.

As the system model, we use a periodic analogue to a random walk model. This means $\underline{\alpha}_t$ evolves according to

$$\underline{\alpha}_{t+1} = \underline{\alpha}_t + \underline{w}_t \quad \text{mod } 2\pi,$$

with a time-invariant, multivariate wrapped normally distributed additive system noise term \underline{w}_t and a modulo operator that ensures that the angles are always between 0 and 2π . The parameters $\underline{\mu}$ and \mathbf{C} of the system noise \underline{w}_t are identical to those of the measurement noise but the units are rad and rad^2 instead of m and m^2 .

The scenario was simulated for 50 time steps with alternating filter and prediction steps. As in Sec. VII-B, we determine the error d_{MV} in each time step and calculate the average over all 50 time steps and over 1500 runs. The results are depicted in Fig. 16 and are in line with the results obtained in the other bivariate scenarios. Both Fourier filters outperform the particle filter as configurations using only few coefficients provide better results than the particle filter using 2000 particles.

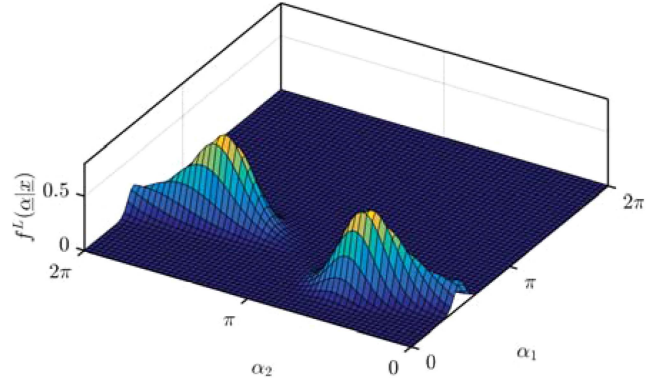


Fig. 15. Likelihood when the measurement $\underline{z}_t = [0 \ 2.3]^T$ m is obtained.

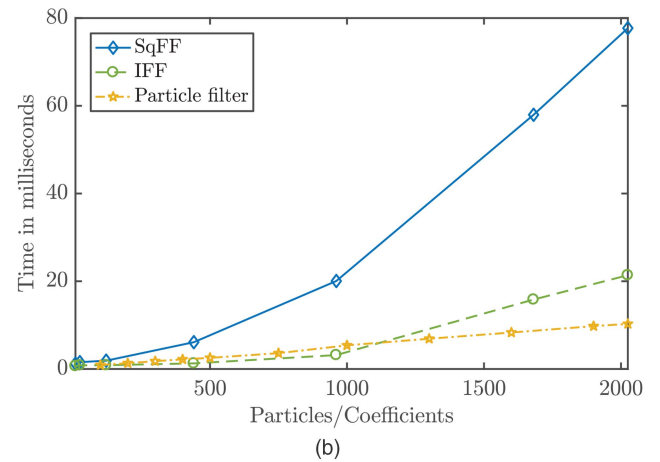
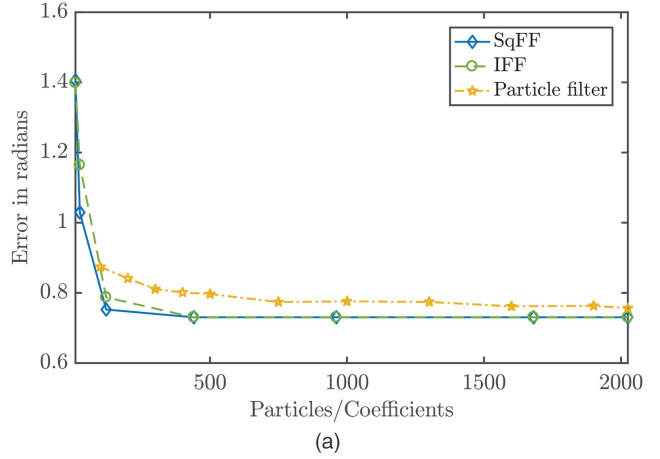


Fig. 16. Average errors and run times for the different filters in the scenario featuring a simulated robotic arm. (a) Error of the different filters depending on the number of particles or coefficients used. (b) Run times of the different filters for one time step depending on the number of particles or coefficients used.

However, as expected, the run time of the Fourier filters is slightly worse than in the previous scenarios as the approximation of the likelihood in every time step causes additional overhead. The IFF is faster than the SqFF with equal results, but the difference in the run time is not that pronounced when using the lowest number of coefficients necessary to obtain the highest accuracy achievable for the Fourier filters in this scenario.

VIII. CONCLUSION

In this paper, we have proposed filters based on Fourier series for multivariate angular estimation problems that fill a gap in recursive Bayesian estimation on hypertoroidal manifolds and allow for good estimation results even when likelihood functions or densities occurring are multimodal. In our evaluation of the error in the circular mean direction, the proposed IFF and SqFF achieve better results than the particle filter when comparing configurations of equal run time. The Fourier filters also outperform the bivariate wrapped normal filter in a scenario for which the latter is well suited.

Out of the proposed Fourier filters, the SqFF tends to perform better when comparing the error on a per coefficient basis. However, compared on a run time basis, the IFF is superior when only the estimation quality of the circular mean direction is evaluated. Since the SqFF prevents negative function values in the resulting approximation of the posterior density, it has significant theoretical advantages. As shown in Sec. VII-A, the SqFF is clearly the better choice if an accurate approximation of the posterior pdf is to be provided. Based on these experiments, we also recommend using the SqFF for additional robustness when the likelihood and the prior density share only little common regions of high function values (such as when an unlikely measurement is observed). All in all, we advise the users to employ the SqFF whenever feasible given the run time constraints to benefit from its higher robustness and the expressiveness of the pdf and to use the IFF when run time constraints are tight.

In future research, we intend to inspect the differences between the IFF and the SqFF more closely to be able to give recommendations for which filter to use given the likelihoods, transition densities, run time requirements, and measures of deviation to be minimized. Furthermore, automatically finding the lowest number of coefficients that results in close to optimal results may help users that want to utilize the filter with high estimation quality while saving computational effort. The number of parameters could be further reduced by using sparse representations of the Fourier coefficient tensors. Additional insights could also result from considering other transformations than the square root. Finally, using other basis functions or targeting other periodic manifolds will also be a subject of future research.

ACKNOWLEDGMENT

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

APPENDIX

In the following, we describe some useful formulae and properties that, despite not being essential to our filter, are important when working with filters and densities in general. We only present the formulae for the non-rooted representation. To use the formula for the Fourier coefficients of the multivariate wrapped normal distribution presented in Appendix A for the SqFF, the approach to derive the Fourier coefficients of the square root from the Fourier coefficients of the original density, as presented in the prediction step in Sec. VI-D.2, can be used. This, however, will not yield a higher accuracy than directly evaluating the square root of the density on a grid and then using the FFT. To use the formulae and properties in Appendix B–D for coefficient tensors representing the square root, we can simply derive the coefficients for the non-rooted representation in the computationally inexpensive way described in Sec. VI-B.1.

A. Fourier Coefficients for the Multivariate Wrapped Normal Distribution

Calculating Fourier coefficients for multivariate wrapped normal distributions in the non-rooted representation is possible using closed-form formulae. For this, we show that we can use the characteristic function of a regular multivariate normal distribution to derive the Fourier coefficients for the multivariate wrapped normal distribution. The characteristic function $\varphi_{\underline{x}}(\underline{k})$ of a random vector \underline{x} on \mathbb{R}^d with density $f_{\underline{x}}$ is defined as

$$\varphi_{\underline{x}}(\underline{k}) = \int_{\mathbb{R}^d} e^{i\underline{k} \cdot \underline{x}} f_{\underline{x}}(\underline{x}) d\underline{x}.$$

Now, let us assume that \underline{x} is normally distributed. Since $e^{i\underline{k} \cdot \underline{x}}$ is 2π -periodic in every dimension,

$$\begin{aligned} \varphi_{\underline{x}}(\underline{k}) &= \int_{\mathbb{R}^d} e^{i\underline{k} \cdot \underline{x}} \mathcal{N}(\underline{x}, \underline{\mu}, \mathbf{C}) d\underline{x} \\ &= \int_{[0, 2\pi)^d} e^{i\underline{k} \cdot \underline{x}} \sum_{\underline{j} \in \mathbb{Z}^d} \mathcal{N}(\underline{x} + 2\pi \underline{j}, \underline{\mu}, \mathbf{C}) d\underline{x} \quad (8) \end{aligned}$$

holds. Aside from a constant factor and a difference in a sign, the last line of the equation (8) is identical to the formula for the Fourier coefficients (5) of a multivariate wrapped normal distribution parametrized by $\underline{\mu}$ and \mathbf{C} . Now, we write the formula for the Fourier coefficients of the multivariate wrapped normal distribution depending on the characteristic function of the multivariate normal distribution [50, Table C] to obtain a closed-form solution. This leads to the formula

$$c_{\underline{k}} = \frac{1}{(2\pi)^d} \varphi_{\underline{x}}(-\underline{k}) = \frac{1}{(2\pi)^d} e^{-i\underline{k} \cdot \underline{\mu} - \underline{k}^T \mathbf{C} \underline{k} / 2}.$$

B. Integrating Fourier Series over Hyperrectangles

For a density given as a Fourier series, it is possible to efficiently calculate the integral over any axis-aligned

hyperrectangle directly from the Fourier coefficients. Let us first regard the one-dimensional case. In this case, we can rewrite the integral over a (truncated) Fourier series from l to r as

$$\int_l^r \sum_{k=-k_{\max}}^{k_{\max}} c_k e^{ikx} dx = \sum_{k=-k_{\max}}^{k_{\max}} c_k \underbrace{\int_l^r e^{ikx} dx}_{h_k}$$

and regard each addend h_k separately to obtain

$$h_0 = c_k [x]_l^r = c_k (r - l)$$

and

$$\begin{aligned} \forall k \neq 0, |k| \leq k_{\max} : h_k &= c_k \frac{1}{ik} [e^{ikx}]_l^r \\ &= -c_k \frac{i}{k} (e^{ikr} - e^{ikl}) \end{aligned}$$

via common integration rules. We now use $c_k = \bar{c}_{-k}$ and obtain

$$\begin{aligned} h_k + h_{-k} &= -c_k \frac{i}{k} (e^{ikr} - e^{ikl}) - \bar{c}_k \frac{i}{-k} (e^{-ikr} - e^{-ikl}) \\ &= \frac{i}{k} (-c_k (e^{ikr} - e^{ikl}) + \overline{c_k (e^{ikr} - e^{ikl})}) \\ &= \frac{2}{k} \mathcal{I}(c_k (e^{ikr} - e^{ikl})), \end{aligned}$$

with $\mathcal{I}(\cdot)$ denoting the imaginary part of the term. As expected, the sum of the pairs are real values. Based on this, we can calculate the integral via

$$\int_l^r \sum_{k=-k_{\max}}^{k_{\max}} c_k e^{ikx} dx = c_k (r - l) + \sum_{k=1}^{k_{\max}} \frac{2}{k} \mathcal{I}(c_k (e^{ikr} - e^{ikl})).$$

The integration formula provided can easily be extended to higher dimensions. If we use $\mathcal{J} \subset \mathbb{Z}^d$ to denote the index set comprising all indices of the nonzero Fourier coefficients,

$$\begin{aligned} \int_l^r \sum_{\underline{k} \in \mathcal{J}} c_{\underline{k}} e^{i\mathbf{k} \cdot \mathbf{x}} &= \sum_{\underline{k} \in \mathcal{J}} c_{\underline{k}} \int_l^r e^{ik_1 x_1} \dots e^{ik_d x_d} d\mathbf{x} \\ &= \sum_{\underline{k} \in \mathcal{J}} c_{\underline{k}} \left(\int_{l_1}^{r_1} e^{ik_1 x_1} dx_1 \dots \int_{l_d}^{r_d} e^{ik_d x_d} dx_d \right) \end{aligned}$$

holds and we can thus split the integration up and use the integration rule for the one-dimensional case to obtain the result in $O(nd)$.

C. Marginalizing Out Specific Dimensions

Calculating the Fourier coefficients for a marginalized density is computationally inexpensive. Without loss of generality, we marginalize the first dimension out. We denote the index set of all nonzero Fourier coefficients by \mathcal{J} and rewrite the integral

$$\begin{aligned} \int_0^{2\pi} \sum_{\underline{k} \in \mathcal{J}} c_{\underline{k}} e^{ik_1 x_1} \dots e^{ik_d x_d} dx_1 \\ = \sum_{\underline{k} \in \mathcal{J}} c_{\underline{k}} e^{ik_2 x_2} \dots e^{ik_d x_d} \int_0^{2\pi} e^{ik_1 x_1} dx_1 \end{aligned}$$

and then use that the integral is always zero for $k_1 \neq 0$

$$\begin{aligned} &= \sum_{k_2=-k_{\max}}^{k_{\max}} \dots \sum_{k_d=-k_{\max}}^{k_{\max}} e^{ik_2 x_2} \dots e^{ik_d x_d} c_{0, k_2, \dots, k_d} \int_0^{2\pi} 1 dx_1 \\ &= \sum_{k_2=-k_{\max}}^{k_{\max}} \dots \sum_{k_d=-k_{\max}}^{k_{\max}} 2\pi e^{ik_2 x_2} \dots e^{ik_d x_d} c_{0, k_2, \dots, k_d}. \end{aligned}$$

Thus, we can calculate the new coefficient tensor by discarding all entries for which the index of the respective dimension is unequal to zero and then multiplying all remaining entries by 2π .

D. Calculating the Covariance Matrix

The covariance matrix mentioned in Sec. III-F is useful for calculating angular correlations and can be calculated efficiently for densities in a Fourier series representation. For higher dimensions, we first introduce a notational trick for Fourier coefficients representing a density after certain dimensions have been marginalized out. We write $c_{k_r, k_t} = (2\pi)^{d-2} c_{0, \dots, 0, k_r, 0, \dots, 0, k_t, 0, \dots, 0}$ for the Fourier coefficients of the density with all dimensions unequal to $r \in \{1, 2, \dots, d\}$ and $t \in \{1, 2, \dots, d\}$ marginalized out and specify the exact coefficient via $c_{k_r, a, k_t, b}$ for $a \in \mathbb{Z}, b \in \mathbb{Z}$. We use a similar notation to index Fourier coefficients of densities with all dimensions except one marginalized out and write them as $c_{k_r} = (2\pi)^{d-1} c_{0, \dots, 0, k_r, 0, \dots, 0}$ and denote specific entries via $c_{k_r, a}$. For additional brevity and clarity, we assign the real and the imaginary part of the first trigonometric moment the names $\underline{p} = \mathcal{R}(\underline{m}_1)$ and $\underline{q} = \mathcal{I}(\underline{m}_1)$.

The general term for entries of the covariance matrix Σ is

$$\mathbb{E}((\text{trig}_1(x_r) - u_r)(\text{trig}_2(x_t) - v_t)),$$

in which trig_1 and trig_2 can be sine or cosine functions. If trig_1 is cos, then $u_r = p_r$, if trig_1 is sin, then $u_r = q_r$ and the same applies to trig_2 and v_t . We present the formulae for the individual entries of the covariance matrix Σ without a derivation as the derivations are long and not essential to this paper.

As higher coefficients are lost in the integral involved in calculating the expectation value, only few coefficients are necessary for each entry. Furthermore, we can use the redundancy in the complex Fourier coefficients to arrive at even easier formulae depending on the real and complex parts of the Fourier coefficients. We further use

$\forall i \in \{1, \dots, d\} : p_i = 2\pi \mathcal{R}(c_{k_i=1})$ and $q_i = -2\pi \mathcal{I}(c_{k_i=1})$ to write the formulae in an even more compact manner.

For the formulae for the individual entries of the covariance matrix, keep in mind that Σ is a $2d \times 2d$ matrix. In the covariance matrix, an odd column index (i.e., $2r-1$) indicates that trig_1 is the sine function, while an even column index (i.e., $2r$) indicates that trig_1 is the cosine function. The relationship is the same between trig_2 and the row index.

For all $r \neq t$, we obtain the formulae

$$\begin{aligned}\sigma_{2r-1,2t-1} &= \mathbb{E}((\cos(x_r) - p_r)(\cos(x_t) - p_t)) \\ &= 2\pi^2 \mathcal{R}(c_{k_r=1, k_t=1}) + 2\pi^2 \mathcal{R}(c_{k_r=1, k_t=-1}) \\ &\quad - p_r p_t\end{aligned}$$

and

$$\begin{aligned}\sigma_{2r-1,2t} &= \mathbb{E}((\cos(x_r) - p_r)(\sin(x_t) - q_t)) \\ &= -2\pi^2 \mathcal{I}(c_{k_r=1, k_t=1}) + 2\pi^2 \mathcal{I}(c_{k_r=1, k_t=-1}) - p_r q_t\end{aligned}\tag{9}$$

for the respective entries of the covariance matrix. Due to the symmetry of the covariance matrix Σ , we can use $\sigma_{2r,2t-1} = \sigma_{2t-1,2r}$ to calculate these entries using (9). For the terms with two sines, we obtain

$$\begin{aligned}\sigma_{2r,2t} &= \mathbb{E}((\sin(x_r) - q_r)(\sin(x_t) - q_t)) \\ &= -2\pi^2 \mathcal{R}(c_{k_r=1, k_t=1}) + 2\pi^2 \mathcal{R}(c_{k_r=1, k_t=-1}) - q_r q_t\end{aligned}$$

If $r = t$, other coefficients are involved in the calculation. In this case, we obtain

$$\begin{aligned}\sigma_{2r-1,2r-1} &= \mathbb{E}((\cos(x_r) - p_r)^2) \\ &= \pi \mathcal{R}(c_{k_r=2}) - p_r^2 + \frac{1}{2}, \\ \sigma_{2r-1,2r} &= \mathbb{E}((\cos(x_r) - p_r)(\sin(x_r) - q_r)) \\ &= \pi \mathcal{I}(c_{k_r=2}) - p_r q_r.\end{aligned}\tag{10}$$

Again, we can use the symmetry of Σ to obtain $\sigma_{2r,2r-1} = \sigma_{2r-1,2r}$, allowing us to use (10). Lastly, we get

$$\begin{aligned}\sigma_{2r,2r} &= \mathbb{E}((\sin(x_r) - q_r)^2) \\ &= -\pi \mathcal{R}(c_{k_r=2}) - q_r^2 + \frac{1}{2}.\end{aligned}$$

REFERENCES

- [1] K. V. Mardia
“Directional Statistics in Geosciences,”
Communications in Statistics—Theory and Methods, vol. 10, no. 15, pp. 1523–1543, 1981.
- [2] E. Batschelet
Circular Statistics in Biology, ser. Mathematics in Biology. London: Academic Press, 1981.
- [3] J. L. Crassidis, F. L. Markley, and Y. Cheng
“Survey of Nonlinear Attitude Estimation Methods,”
Journal of Guidance, Control, and Dynamics, vol. 30, no. 1, pp. 12–28, 2007.
- [4] J. Glover and L. P. Kaelbling
“Tracking the Spin on a Ping Pong Ball with the Quaternion Bingham Filter,”
in *Proceedings of the 2014 IEEE Conference on Robotics and Automation (ICRA 2014)*, Hong Kong, China, 2014.
- [5] J. Lo and A. Willsky
“Estimation for Rotational Processes with One Degree of Freedom—Part I: Introduction and Continuous-Time Processes,”
IEEE Transactions on Automatic Control, vol. 20, no. 1, pp. 10–21, Feb. 1975.
- [6] A. S. Willsky
“Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part I: Analysis,”
IEEE Transactions on Information Theory, 1974.
- [7] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun
“A Multi-Sensor Circular Particle Filter Applied to the Fusion of the GPS-L2C Channels,”
in *Proceedings of the 14th International Conference on Information Fusion (Fusion 2011)*, Chicago, Illinois, USA, 2011.
- [8] G. Kurz and U. D. Hanebeck
“Heart Phase Estimation Using Directional Statistics for Robotic Beating Heart Surgery,”
in *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington D. C., USA, Jul. 2015.
- [9] 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.
- [10] K. V. Mardia and P. E. Jupp
Directional Statistics,
1st ed. Baffins Lane, Chichester, West Sussex, PO19 1UD England: Wiley, 1999.
- [11] S. R. Jammalamadaka and A. Sengupta
Topics in Circular Statistics.
World Scientific, 2001.
- [12] W. Boomsma, K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck
“A Generative, Probabilistic Model of Local Protein Structure,”
Proceedings of the National Academy of Sciences, vol. 105, no. 26, pp. 8932–8937, 2008.
- [13] Y.-H. Chen, D. Wei, G. Newstadt, M. DeGraef, J. Simmons, and A. Hero
“Parameter Estimation in Spherical Symmetry Groups,”
IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1152–1155, 2015.
- [14] F. Chatelain and N. Le Bihan
“Von Mises–Fisher Approximation of Multiple Scattering Process on the Hypersphere,”
in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6461–6465.
- [15] M. Akcakaya and A. Nehorai
“MIMO Radar Detection and Adaptive Design Under a Phase Synchronization Mismatch,”
IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 4994–5005, Oct 2010.
- [16] A. S. Willsky
“Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part II: Implementation,”
IEEE Transactions on Information Theory, 1974.
- [17] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun
“A Recursive Fusion Filter for Angular Data,”
in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2009.
- [18] G. Kurz, I. Gilitschenski, and U. D. Hanebeck
“Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions,”
in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

- [19] ———
“Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions,”
in *Proceedings of the 2014 American Control Conference (ACC 2014)*, Portland, Oregon, USA, Jun. 2014.
- [20] ———
“Recursive Bayesian Filtering in Circular State Spaces,”
IEEE Aerospace and Electronic Systems Magazine, vol. 31, no. 3, pp. 70–87, Mar. 2016.
- [21] 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.
- [22] 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.
- [23] G. Kurz, I. Gilitschenski, M. Dolgov, and U. D. Hanebeck
“Bivariate Angular Estimation Under Consideration of Dependencies Using Directional Statistics,”
in *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2014)*, Los Angeles, California, USA, Dec. 2014.
- [24] 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.
- [25] S. Kay
Fundamentals of Statistical Signal Processing: Estimation Theory.
Prentice Hall, 1993.
- [26] M. F. Huber, F. Beutler, and U. D. Hanebeck
“Semi-Analytic Gaussian Assumed Density Filter,”
in *Proceedings of the 2011 American Control Conference (ACC 2011)*, San Francisco, California, USA, Jun. 2011.
- [27] U. D. Hanebeck
“PGF 42: Progressive Gaussian Filtering with a Twist,”
in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.
- [28] J. Steinbring and U. D. Hanebeck
“LRKF Revisited: The Smart Sampling Kalman Filter (S2KF),”
Journal of Advances in Information Fusion, vol. 9, no. 2, pp. 106–123, Dec. 2014.
- [29] J. Steinbring, M. Pander, and U. D. Hanebeck
“The Smart Sampling Kalman Filter with Symmetric Samples,”
Journal of Advances in Information Fusion, vol. 11, no. 1, pp. 71–90, Jun. 2016.
- [30] M. Huber, P. Krauthausen, and U. D. Hanebeck
“Superficial Gaussian Mixture Reduction,”
in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2011)*, Berlin, Germany, Oct. 2011.
- [31] S. R. Jammalamadaka and Y. Sarma
“A Correlation Coefficient for Angular Variables,”
Statistical Theory and Data Analysis II, pp. 349–364, 1988.
- [32] R. A. Johnson and T. Wehrly
“Measures and Models for Angular Correlation and Angular-Linear Correlation,”
Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 2, pp. 222–229, 1977.
- [33] P. E. Jupp and K. V. Mardia
“A General Correlation Coefficient for Directional Data and Related Regression Problems,”
Biometrika, vol. 67, no. 1, pp. 163–173, 1980.
- [34] A. Zygmund
Trigonometric Series,
3rd ed. Cambridge University Press, 2003, vol. 1 and 2.
- [35] Y. Katznelson
An Introduction to Harmonic Analysis,
3rd ed. Cambridge University Press, 2004.
- [36] J. J. Fernández-Durán
“Circular Distributions Based on Nonnegative Trigonometric Sums,”
Biometrics, vol. 60, no. 2, Jun. 2004.
- [37] D. Brunn, F. Sawo, and U. D. Hanebeck
“Efficient Nonlinear Bayesian Estimation Based on Fourier Densities,”
in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sep. 2006, pp. 312–322.
- [38] ———
“Nonlinear Multidimensional Bayesian Estimation with Fourier Densities,”
in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, USA, Dec. 2006, pp. 1303–1308.
- [39] 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.
- [40] D. E. Dudgeon and R. M. Mersereau
Multidimensional Digital Signal Processing.
Prentice Hall, 1984.
- [41] 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.
- [42] S. P. Jan-P. Calliess, Michael Mai
“On the Computational Benefit of Tensor Separation for High-Dimensional Discrete Convolutions,”
Multidimensional Systems and Signal Processing, 2012.
- [43] 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.
- [44] G. Kurz, I. Gilitschenski, F. Pfaff, and L. Drude
“libDirectional,”
2015. [Online]. Available: <https://github.com/libDirectional>.
- [45] R. P. Feynman
“Negative Probability,”
Quantum Implications: Essays in Honour of David Bohm, pp. 235–248, 1987.
- [46] W. K. Hastings
“Monte Carlo Sampling Methods Using Markov Chains and Their Applications,”
Biometrika, vol. 57, no. 1, pp. 97–109, 1970.
- [47] T. M. Cover and J. A. Thomas
Elements of Information Theory.
John Wiley & Sons, 2012.
- [48] F. Liese and I. Vajda
“On Divergences and Informations in Statistics and Information Theory,”
IEEE Transactions on Information Theory, 2006.
- [49] G. Kurz
“Directional Estimation for Robotic Beating Heart Surgery,”
Ph.D. dissertation, Karlsruhe Institute of Technology, Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe, Germany, 2015.
- [50] F. Oberhettinger
Fourier Transforms of Distributions and Their Inverses.
Academic Press, 1973.

Florian Pfaff started as a Ph.D. student at the Karlsruhe Institute of Technology (KIT) in 2014 after completing his diploma for which he received an award due to his outstanding performance. His research interests include directional estimation, distributed estimation, and multitarget tracking and its application to bulk material sorting.

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

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

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

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