Multimodal Circular Filtering Using Fourier Series

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

***Abstract*—Recursive filtering with multimodal likelihoods and transition densities on periodic manifolds is, despite the compact domain, still an open problem. We propose a novel filter for the circular case that performs well compared to other state-of-the-art filters adopted from linear domains. The filter uses a limited number of Fourier coefficients of the square root of the density. This representation is preserved throughout filter and prediction steps and allows obtaining a valid density at any point in time. Additionally, analytic formulae for calculating Fourier coefficients of the square root of some common circular densities are provided. In our evaluation, we show that this new filter performs well in both unimodal and multimodal scenarios while requiring only a reasonable number of coefficients.**

Keywords*—*Directional statistics, density estimation, nonlinear filtering, recursive Bayesian estimation.

I. Introduction

Although many users are unaware of the assumption, a lot of results in Bayesian statistics rely on a linear topology of the underlying domain. While a wide range of tools have emerged in the field of estimation for linear domains, many of them perform significantly worse on periodic domains. For example, using a Kalman filter and thus ignoring periodicities can lead to large errors, especially in regions close to the border of periodicity. While there have been attempts to reduce the severity of the problems arising, it is often impossible to fully eliminate the error introduced by the incorrect underlying assumption.

As a special case of periodic manifolds, we will regard circular manifolds throughout this paper, which are by no means rare. While the primary goal of tracking algorithms is usually to estimate the position of an object, the secondary goal is often to estimate its orientation, which is a 2π-periodic quantity. Just as filtering with arbitrary likelihoods is an open challenge on linear domains, the case is not settled in the circular case. However, while periodicity introduces additional complexity, there is the advantage of having a compact domain. Our proposed filter takes advantage of this property and constitutes an approach that is tailored to periodic domains.

A variety of work has been done on the topic of circular filtering. Most methods assume that the posterior density belongs to a specified class of circular densities. For example, Azmani et al. introduced a filter using the von Mises distribution [1] that only allows the use of identity system and measurement models. Unfortunately, unlike in the case of linear domains for which the Gaussian distribution can be used, there are no distributions that are popular, useful, and yield a distribution of the same class for both the multiplication and the convolution operation. Therefore, approximations are necessary in either the prediction or filter step to maintain such a representation.

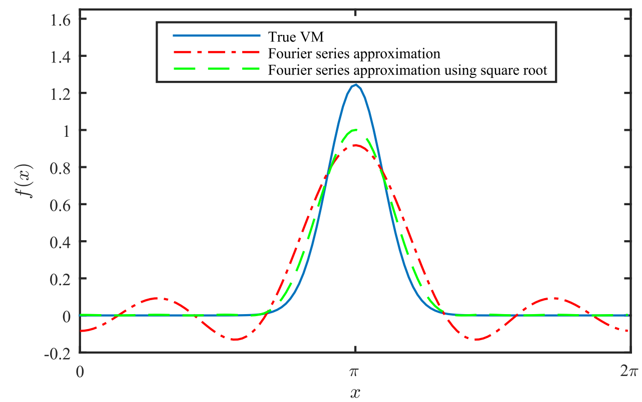


Figure 1. A von Mises distribution with $\kappa = 10$ and corresponding approximations using 7 Fourier coefficients.

Progress towards improved handling of more complicated system and measurement models was made in [2], [3], [4] but these approaches are still limited to unimodal densities. While there are filters for multimodal cases [5], [6], multimodality is only regarded if it is directly modeled by the distribution, e.g., for antipodally symmetric densities such as the density of the Bingham distribution. An easy approach would be to use mixtures, but this would entail the need for component reduction as in the Gaussian mixture case for linear domains.

Closely related work was done for phase estimation by Willsky who suggested using Fourier series for applications in communication scenarios [7]. He first theoretically derived solutions for certain stochastic differential equations under the assumption of using an infinite number of coefficients and then described practical implementations [8]. He reported obtaining poor results when using three Fourier coefficients and suggested approximating the results using wrapped normal densities. Fourier series were also considered by Fernández-Durán [9] who approximated densities using Fourier series and used an optimization constraint to ensure nonnegativity. Bayesian filtering was not dealt with and thus important challenges were not addressed.

The advantage of the approach presented in this paper is that coefficient vectors can be efficiently truncated throughout the filter steps while still maintaining the nonngeativity and thus validity of the density. In Figure 1, we demonstrate the effect of truncating a Fourier series approximating a density directly and compare it to the effect on our proposed solution. Our idea was inspired by an approach to nonlinear filtering on linear domains that was presented in [10] and for the multi-dimensional case in [11].

II. Prerequisites

Before we get to our contribution, we want to revisit some basics of circular statistics and Fourier series. We refer readers to the books [12] and [13] for more information about directional statistics and, regarding Fourier series, to the classic book by Zygmund about trigonometric series [14], books about harmonic analysis (such as [15]), and the extensive literature of the signal processing community.

A. Basics of Circular Statistics

Circular distributions are distributions defined on the interval $[0, 2\pi)$. The regarded manifold is understood to be periodic, with, visually speaking, the end of the interval being connected to its beginning.

When regarding periodic manifolds, many concepts from linear manifolds have to be adapted. Probability density functions (pdfs) assuming an unbounded domain (such as the normal distribution and many others) cannot be directly utilized and have to be modified for the use on circular domains. We will deal with some common circular distributions in Section IV-B. Another important concept for us are trigonometric moments that are also called circular moments. The k-th trigonometric moment [12, Ch. 2.1] m_k of a random variable $\boldsymbol{x}$ with density $f(x)$ is defined as

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

which can, in general, be complex valued. While the moments are often split up into a real and an imaginary part, the above representation is more convenient for us throughout this paper. Trigonometric moments differ from the standard (power) moments both in definition and intuition, e.g., the first trigonometric moment is not only a measure of the location of the density, but also of its spread. An important quantity used for estimation that describes only the density's location is the circular mean, which is defined as

$$\mu = \operatorname{atan2}(\mathcal{I}(m_1), \mathcal{R}(m_1)) \ .$$

B. Fourier Series

Fourier series allow representing a variety of functions by means of complex exponential functions. The speed of convergence of the Fourier series depends on certain properties that we will deal with in more detail later. For 2π-periodic functions, the complex Fourier series (see [16, Ch. 4.1] and [17, Ch. 4.1]) can be written in the form

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx} \quad \text{with } c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx \ .$$

In this representation, c_k are complex numbers, but when expanding real functions, $c_{-k} = \overline{c_k}$ always holds. For real functions, the concept can also be thought of as representing the function using sine and cosine functions. This leads to an alternative representation using only real coefficients. In the following, we will only consider the complex representation as this allows us to obtain easier formulae.

The calculation of Fourier coefficients is, as can be seen by comparing the integration formulae, closely related to the calculation of trigonometric moments via $c_k = \frac{m_{-k}}{2\pi}$. Since the circular mean can be calculated using the first trigonometric moment, we can also derive it directly from the Fourier coefficients.

III. Key Idea

As common circular densities are 2π-periodic and square integrable, Fourier series are well suited to represent them. In fact, some popular densities are often written in a Fourier series form or a form that closely resembles a Fourier series. However, as their exact representation usually involves an infinite amount of Fourier coefficients, truncation is necessary. As we will see in Section V, truncation is also necessary for the multiplication operation required for filter steps.

The truncation leads to an approximation error as the higher coefficients are implicitly assumed to be zero. Crucially, in regions where the density is close to zero, the approximation can become negative (as can be seen in Figure 1), which violates conditions for valid densities and can entail further problems. Therefore, as done similarly in [10], we approximate the square root of the density, rather than the density itself, by a Fourier series. This ensures that the reconstructed prior and posterior densities obtained by squaring the Fourier series yield nonnegative values for any angle.

Working with a Fourier series approximation of the square root of a density necessitates performing prediction and update steps while maintaining (or restoring) this representation. Details on this will be laid out in Section V. It is also useful to be able to calculate Fourier coefficients of the square root of common densities, which will be addressed in the following section.

IV. Approximating Common Circular Distributions

Many popular circular distributions are derived from linear distributions. One of the techniques to derive a circular distribution from a linear one is called wrapping [12, Ch. 2.1]. If a wrapped distribution is evaluated at a point x, the probability of $x + k\pi$ for all $k \in \mathbb{Z}$ is summed up. It can be easily verified that this procedure always returns a valid density on the circle. In the following, we will regard several densities derived by wrapping—the wrapped normal, the wrapped Cauchy, and the wrapped exponential distribution—and the von Mises distribution, another commonly used circular distribution that can be derived in a different manner.

These distributions and their moments are described in [13] and [18]. For distributions obtained by wrapping, Fourier coefficients can also be derived by evaluating the characteristic function of the density being wrapped at integer arguments. However, we require formulae for the Fourier coefficients of the square root of the density, which cannot be derived like this. A method to approximate the square root of an arbitrary density using a Fourier series will be introduced in Section V-C.

A. Handling the Parameter μ

Three of the considered circular distributions commonly use μ as a parameter to specify the circular mean of the distribution. To rotationally shift a function by μ towards the positive border of periodicity (e.g., for $[0, 2\pi)$, towards 2π), we can simply multiply all coefficients by $e^{-ki\mu}$. Thus, it is sufficient to derive formulae for $\mu = 0$ as we can obtain the formulae incorporating μ by using this additional factor.

B. Convergence of the Fourier Series

The convergence of the complex Fourier coefficients is easy to show. However, as we want to approximate the entire density, it is more important to regard how the squared integrated difference between the true function and the approximation behaves when increasing the number of coefficients. As all considered densities and their square roots are square integrable over $[0, 2\pi]$, we know from the Hausdorff–Young inequality [14, Vol. 2, Ch. XII] that for the Fourier coefficient vector $\underline{c}^{sqrt}$

$$\sum_{k=-\infty}^{\infty} |c_k^{sqrt}|^2 < \infty$$

holds. Let us denote the truncated Fourier series of the square root of the density that only considers coefficients from $-n$ to n with g_n. Using the theorem of Fischer–Riesz, we can deduce [14, Vol. 2, Ch. XII] that for $m < n$

$$\|g_n - g_m\|_2 \le \sqrt{\sum_{m+1}^{n} |c_k^{sqrt}|^2} \to 0$$

as $m \to \infty$ and $n \to \infty$, proving the convergence of g_n to $\sqrt{f}$ in $L^2(0, 2\pi)$. Using a similar argument, we can also show that g_n^2 converges to f in $L^2(0, 2\pi)$.

C. Von Mises Distribution

The von Mises distribution is a popular distribution that is sometimes referred to as the circular normal distribution. Its density is given by

$$f_{VM}(x; \mu = 0, \kappa) = \frac{e^{\kappa \cos(x)}}{2\pi I_0(\kappa)} .$$

Its trigonometric moments and thus, with little effort, the Fourier coefficients of the non-rooted density c_k^{id} are given by

$$\forall k \in \mathbb{Z} : c_k^{id} = \frac{1}{2\pi I_0(\kappa)} I_{|k|}(\kappa) ,$$

with $I_k(\cdot)$ being the modified Bessel function of the first kind. To obtain the Fourier series for the square root, we can rewrite $\sqrt{f(x; \mu = 0, \kappa)}$ in the following way

$$\sqrt{\frac{e^{\kappa \cos(x)}}{2\pi I_0(\kappa)}} = \frac{e^{\kappa/2 \cdot \cos(x)}}{\sqrt{2\pi I_0(\kappa)}} = \frac{2\pi I_0(\kappa/2)}{\sqrt{2\pi I_0(\kappa)}} \cdot \frac{e^{\kappa/2 \cdot \cos(x)}}{2\pi I_0(\kappa/2)} .$$

The factor in blue is the density of a von Mises distribution $f_{VM}(x; \mu = 0, \tilde{\kappa})$ with $\tilde{\kappa} = \kappa/2$, of which the Fourier coefficients can be calculated using the above described formula. Multiplying the resulting Fourier coefficients with the factor in green yields

$$\forall k \in \mathbb{Z} : c_k^{sqrt} = \frac{1}{\sqrt{2\pi I_0(\kappa)}} I_{|k|}(\kappa/2) .$$

The effect of increasing $|k|$ in $I_{|k|}(0.5\kappa)$ can be analyzed by using the definition of the modified Bessel function of the first kind

$$I_{|k|}(\kappa/2) = \sum_{n=0}^{\infty} \frac{(\frac{\kappa}{4})^{2n+|k|}}{n!\,\Gamma(|k| + n + 1)} ,$$

in which we can observe an exponential increase (or decrease, depending on κ) in the numerator, whereas there is an increase in the order of a factorial in the denominator due to the gamma function. Therefore, asymptotically, the Fourier coefficients are decreasing in the order of $(|k|!)^{-1}$.

D. Wrapped Normal Distribution

The density of the wrapped normal distribution is obtained by wrapping the density of a Gaussian distribution around the circle, yielding

$$f_{WN}(x; \mu = 0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left(\frac{-(x + 2\pi k)^2}{2\sigma^2}\right) .$$

A wrapped normal distribution with a certain first trigonometric moment is similar to the von Mises distribution with the same moment. The Fourier coefficients of the wrapped normal distribution are easily derived from its trigonometric moments. From the trigonometric moments, we get

$$\forall k \in \mathbb{Z} : c_k^{id} = \frac{1}{2\pi} e^{-\sigma^2 k^2/2}$$

as Fourier coefficients for the original density. So far, a closed-form expression for the Fourier coefficients of the square root of the density is not known.

E. Wrapped Cauchy Distribution

The wrapped Cauchy distribution is obtained by wrapping the density of the Cauchy distribution

$$f_{WC}(x; \mu = 0, a) = \sum_{k=-\infty}^{\infty} \frac{a}{\pi(a^2 + (x + 2\pi k)^2)} .$$

The Fourier coefficients for the density, derived from the trigonometric moments, are

$$\forall k \in \mathbb{Z} : c_k^{id} = \frac{1}{2\pi} e^{-|k|a} .$$

The Fourier coefficients of the square root of the density were derived with the help of a computer algebra system and some manual reformulations. For all $k \in \mathbb{Z}$, we get

$$c_k^{sqrt} = \sqrt{\frac{1}{2\pi} \tanh\left(\frac{a}{2}\right)} \frac{{}_3F_2(\frac{1}{2}, \frac{1}{2}, 1; 1 - k, 1 + k; \operatorname{sech}^2(\frac{a}{2}))}{\Gamma(1 - k)\Gamma(1 + k)}$$
$$= \sqrt{\frac{1}{2\pi^3} \tanh\left(\frac{a}{2}\right)} \sum_{n=0}^{\infty} \frac{\Gamma^2(\frac{1}{2} + n)\operatorname{sech}^{2n}(\frac{a}{2})}{\Gamma(1 - k + n)\Gamma(1 + k + n)} ,$$

with $\tanh(\cdot)$ being the hyperbolic tangent and $\operatorname{sech}(\cdot)$ the hyperbolic secant. The form in the second line is useful as it can be used as an easy and numerically more benign way to implement the expression if the regularized hypergeometric function ${}_3F_2(\underline{a}, \underline{b}, z) \cdot (\Gamma(b_1)\Gamma(b_2))^{-1}$ is not available as a library function. In this representation, it is also more evident that the dominating term, depending on the sign of k either $\Gamma(1+k+n)$ or $\Gamma(1-k+n)$, is in the denominator and causes an asymptotic convergence in the order of $(|k|!)^{-1}$.

F. Wrapped Exponential Distribution

The density of the wrapped exponential distribution, introduced in [18], is given by

$$f_{WE}(x;\lambda) = \sum_{k=0}^{\infty} \lambda e^{-\lambda(x+2\pi k)} \ .$$

As in general $\lim_{x\searrow 0} f_{WE}(x) \neq \lim_{x\nearrow 2\pi} f_{WE}(x)$, the wrapped exponential density is not continuous. We can easily derive its Fourier coefficients from its trigonometric moments and get the formula

$$\forall k \in \mathbb{Z} : c_k^{id} = \frac{1}{2\pi} \frac{\lambda^2 - \lambda k i}{\lambda^2 + k^2} \ .$$

As can be derived using computer algebra systems, the Fourier coefficients of the square root are given by

$$\forall k \in \mathbb{Z} : c_k^{sqrt} = \frac{\sqrt{\lambda}}{\pi(\lambda + 2ki)} \frac{(e^{\pi\lambda} - 1)}{\sqrt{e^{2\pi\lambda} - 1}} \ .$$

The convergence is significantly slower than for other densities as the denominator is merely growing approximately linearly faster than the numerator.

V. Prediction and Filter Steps

Our aim is to approximate the posterior density of our random variable $\boldsymbol{x}_t$ for the state at time step t, given all observed measurements $y_1, \ldots, y_t$. In other words, we aim to approximate the posterior density $f_t^e(x_t) = f(x_t|y_1, \ldots, y_t)$. Dealing with the whole density in Bayesian statistics allows us to build credible sets and quantify the probability that the state is in a specified region [19, Ch. 9]. On linear domains, the mean of the posterior density is of special interest as it constitutes the (Bayesian) minimum mean squared error estimator [20, Ch. 10]. In general, obtaining the correct mean over multiple time steps necessitates keeping track of the whole density. The prediction step for probabilistic models is described by

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

with transition density $f_t^T(x_t) = f(x_{t+1}|x_t)$, predicted density $f_{t+1}^p(x_{t+1}) = f(x_{t+1}|y_1, \ldots, y_t)$, and sample space Ω_x, while the Bayesian update step is given by the Bayes formula

$$f_t^e(x_t) = \frac{f_t^L(x_t) f_t^p(x_t)}{\int_{\Omega_x} f^L(x_t) f^p(x_t) dx_t} \propto f_t^L(x_t) f_t^p(x_t)$$

with the likelihood function $f_t^L(x_t) = f(y_t|x_t)$. Since determining and representing the whole density after prediction and filter steps is only feasible in special cases, a lot of effort in the field of nonlinear filtering for linear domains is geared towards estimating the mean of the true posterior density for every time step with minimal effort. However, the simplifications used can reduce estimation quality in less benign cases. Universal filters such as the particle filter [21] try to take the whole shape of the density into account.

Two efficient ways to estimate the posterior mean on periodic domains are the von Mises filter [1] and the wrapped normal (WN) filter [4] that aim to keep track of the posterior mean by approximating the predicted and posterior densities using von Mises or wrapped normal densities. While the WN filter correctly captures the mean after any single step, the resulting density does (even when the likelihood is a wrapped normal density), in general, not match the true posterior after filter steps, leading to imprecision in future time steps. Likewise, the same problem applies to prediction steps of the von Mises filter. Furthermore, these two filters do not take multimodality into account, significantly impeding the performance when the true posterior is multimodal.

The few available universal filters for periodic domains are approaches adopted from linear domains such as discrete (grid) filters and the particle filter. To derive a new universal filter for periodic domains, we need to be able to perform the necessary operations for the prediction and filter steps. Another basic requirement for any new filter is that the number of parameters must not increase indefinitely. For our proposed filter, in which we represent densities by approximating their square root as a Fourier series, this necessitates further approximations.

As we will see in the following subsections, it would be possible to perform the filter step in an exact fashion, but this would result in an increase in the number of parameters. Therefore, an approximation is necessary to limit the number of parameters. As we do not provide an exact formula for the prediction step, there is always an approximation involved. These are the two reasons why our proposed filter, while universal, is still an approximation even if all f_t^L, all f_t^T, and the initial prior f_0^p can be represented using a limited number of Fourier coefficients.

Until Section V-C, we will assume that we have a Fourier series approximation of the square root of f_t^T and f_t^L. If they belong to one of the densities described above, a Fourier series approximation can be derived using the respective formula. If they are arbitrary functions or non-rooted Fourier series, we need an additional step described in Section V-C that allows us to obtain an approximation with n Fourier coefficients in $O(n \log n)$. If the likelihood is time-invariant except for shifting operations, the coefficients calculated once can be reused in future steps using the shifting operation described in Section IV-A.

A. Filter Step

To perform a Bayes update, we need to be able to multiply two densities and normalize the result. Having both the square root of the likelihood and the prior density as Fourier series, obtaining the corresponding posterior density only requires simple operations. Multiplying two functions in Fourier series representation equals a discrete convolution of the Fourier coefficients ([16, Ch. 4.4]). As the square root of the posterior can be obtained by multiplying the square root of the prior and the square root of the likelihood, this property can also be used for the square root representation. In other words, the relationship

$$\sqrt{f_t^e} \propto \sqrt{f_t^p f_t^L} = \sqrt{f_t^p}\sqrt{f_t^L}$$

holds. Thus, the Fourier coefficient vector $\underline{c}^{e,sqrt}$ for the unnormalized, rooted posterior can be calculated from the coefficient vectors of the square root of the prior $\underline{c}^{p,sqrt}$ and of the likelihood $\underline{c}^{L,sqrt}$ via a discrete convolution

$$\underline{c}^{e,sqrt} \propto \underline{c}^{p,sqrt} * \underline{c}^{L,sqrt} \; .$$

As the discrete convolution leads to an increased length of the coefficient vector, the vector usually needs to be truncated to prevent an exponential increase in parameters over time. At this point, it should be noted that we implicitly assume that all coefficients with higher indicies are zero as they would otherwise affect those with lower indices as part of the convolution operation.

The second necessary operation for the Bayes update, the normalization, is also easy to perform. As integrating sines and cosines with a frequency that is an integer multiple of 2π from 0 to 2π yields zero, only c_0 is needed to integrate a Fourier series. For real functions, c_0 is real-valued and the integral becomes a multiplication by 2π. As a pdf is supposed to integrate to 1, dividing a non-rooted density by $2\pi c_0^{e,id}$, which can be achieved by dividing all Fourier coefficients by $2\pi c_0^{e,id}$, ensures normalization. For the square root form, we have to keep in mind that we want to normalize the actual density and not its square root. Therefore, we first have to calculate $c_0^{e,id}$, one element of the convolution result, from $\underline{c}^{e,sqrt}$, which can be achieved in $O(n)$. Then, we have to divide the Fourier series by $\sqrt{2\pi c_0^{e,id}}$ to ensure that the reconstructed density is normalized.

All in all, assuming we have $\sqrt{f_t^L}$ as a Fourier series, the asymptotic run time complexity when dealing with n Fourier coefficients is $O(n \log n)$ due to the discrete convolution required.

B. Prediction Step

Convolving Fourier series is usually not a problem as the convolution of two Fourier series can be obtained by calculating the Hadamard (element-wise) product of the coefficient vectors and then multiplying all coefficients by 2π [16, Ch. 4.4]. However, we are not interested in $\sqrt{f_t^e} * \sqrt{f_t^T}$, the convolution of the square roots, but rather in the square root of the actual predicted density for the next time step $t+1$. Written as formulae, due to the inequality

$$\sqrt{f_{t+1}^p} = \sqrt{f_t^e * f_t^T} \neq \sqrt{f_t^e} * \sqrt{f_t^T} \; ,$$

working with the square root form requires extra effort.

To obtain a valid density for f_{t+1}^p, we only require the posterior and transition densities in their square root form. We first square both functions individually by convolving each Fourier coefficient vector with itself to obtain coefficient vectors representing two valid densities. Afterwards, we can calculate the convolution using the Hadamard multiplication (denoted as $\circ$) of the Fourier coefficients to obtain

$$\underline{c}^{p,id} = 2\pi(\underline{c}^{e,sqrt} * \underline{c}^{e,sqrt}) \circ (\underline{c}^{T,sqrt} * \underline{c}^{T,sqrt}) \; .$$

Calculating the Fourier coefficients of the square root of a function is no trivial matter. One approach would be attempting to find $\underline{c}^{p,sqrt}$ for which $\underline{c}^{p,sqrt} * \underline{c}^{p,sqrt} = \underline{c}^{p,id}$ holds. However, especially as taking the square root can lead to higher frequencies, this is a considerably hard problem. As finding the optimal solution would still require an approximation in form of the truncation of the coefficient vector, we will instead introduce a cheap approximation in the next subsection that yields satisfactory results and is in $O(n \log n)$ for n coefficients. Taking this additional effort into account, we obtain a total asymptotic run time complexity of $O(n \log n)$ caused by the discrete convolution and the rooting step involved.

C. Obtaining Fourier Coefficients of the Square Root

Being able to obtain a Fourier series approximation of the square root of a density is a vital part for the prediction step and for dealing with arbitrary transition densities and likelihood functions. For an odd integer n, we can approximate the complex Fourier coefficients with indices $-\frac{n-1}{2}$ to $+\frac{n-1}{2}$ using n equally spaced function values at $[0, \frac{2\pi}{n}, 2 \cdot \frac{2\pi}{n}, \ldots, (n-1) \cdot \frac{2\pi}{n}]$. Using these function values, we can approximate the coefficients using a discrete Fourier transform, e.g., the very efficient method called fast Fourier transform (FFT) [17]. Although the FFT can only handle inputs with a size of a power of two (often necessitating zero padding), its run time complexity is always in $O(n \log n)$. It should be noted that the coefficients obtained are only approximations as higher frequencies are disregarded, entailing aliasing.

To optimize the run time needed for n function evaluations, we have to distinguish between two cases: the case, in which the cost of the function evaluation is independent from the number of coefficients and thus in $O(1)$ and the other case in which it is dependent on the number of Fourier coefficients. The latter applies when we aim to approximate the Fourier coefficients of the square root of a Fourier series as in Section V-B.

In the first case, obtaining n function values is in $O(n)$ and thus does not pose a problem. However, what we need to avoid in the latter case is a naïve calculation of n function values of a Fourier series with n coefficients as this would result in an unnecessarily high asymptotic run time complexity of $O(n^2)$. Just as we are able to approximate Fourier coefficients from function values, we can use the inverse FFT to calculate the function values at equidistant points from the Fourier coefficients in $O(n \log n)$. After obtaining the function values, we calculate the square root to obtain the function values of the square root of the density and afterwards use the FFT to approximate the Fourier coefficients. The procedure is shown as a diagram in Figure 2.

It is important to note that squaring the function values of the new Fourier series at arbitrary angles will, in general, not be equal to the values of the original Fourier series. This is because higher frequencies can result from taking the square root. This is apparent when regarding a sine function. While a sine function can be represented by three Fourier coefficients, the square root of a sine function cannot be represented using only these coefficients. All in all, the asymptotic run time

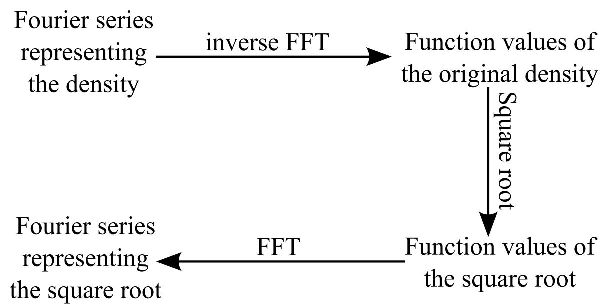


Figure 2. Diagram showing our method to obtain the Fourier coefficients for the square root of a density from the coefficients of the actual density.

complexity is in $O(n \log n)$ as one FFT and at most one inverse FFT is required.

An entirely different and less versatile approach to obtain Fourier coefficients for approximating the square root of a density would be to approximate the distribution by one of the distributions described in Section IV and use the given formulae to calculate the desired number of Fourier coefficients.

VI. Evaluation

To evaluate our new filter, we performed two simulative evaluations. They were designed to be fundamentally different and regard both the estimation quality of the mean, as well as the quality of the approximation of the density or distribution, which is a useful benchmark for bimodal densities. In both cases, an object moves along a circle for 40 time steps with an approximately known, time-varying velocity of $\boldsymbol{u}_k \sim \mathcal{WN}(\mu = 0.3, \sigma = 0.5)$ radians per time step. In other words, the state represented by the angle behaves according to $\boldsymbol{x}_{k+1} = (\boldsymbol{x}_k + \boldsymbol{u}_k) \mod 2\pi$. The main difference between the two scenarios lies in the choice of the (simulated) sensor as this choice determines the likelihood function.

The first scenario features only unimodal densities and is therefore well suited for the use of the WN filter [4] that can handle additive, unimodal noise well. This scenario demonstrates that our proposed filter can also handle easy scenarios well and with little effort. The second scenario, in which the (true) posterior can become bimodal, is more complicated. In this scenario, our proposed filter significantly outperforms the (unsuited) WN filter and two nonlinear filters adopted from linear domains: a discrete (grid) filter and a particle filter. All filters use the same transition densities and likelihood functions that were derived from the generative model. The particle filter was implemented as a simple SIR particle filter that restores equal weights using resampling after every filter step. As the evaluation criteria for the bimodal scenario were more expensive, we used fewer different numbers of coefficients, grid points, and samples and simulated fewer runs than in the first scenario.

A. Unimodal Scenario

In the first scenario, all measurements are angular measurements with measurement equation $\boldsymbol{y}_k = (\boldsymbol{x}_k + \boldsymbol{v}_k) \mod 2\pi$ and $\boldsymbol{v}_k \sim \mathcal{WN}(\mu = 0, \sigma = 0.3)$. We performed $R = 1000$ runs and evaluated the performance at the end of the 40th time step. To judge the performance, we calculated the angular RMSE between the circular mean and the ground truth over all runs using

$$\text{ARMSE}(\underline{x}^e, \underline{x}^{\text{true}}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R} d(x_r^e, x_r^{\text{true}})^2} \ ,$$

with $d(\alpha, \beta)$ being the smaller of the two arclengths between α and β [12, Ch. 1.3.2]

$$d(\alpha, \beta) = \min(\alpha - \beta, 2\pi - (\alpha - \beta)) \ .$$

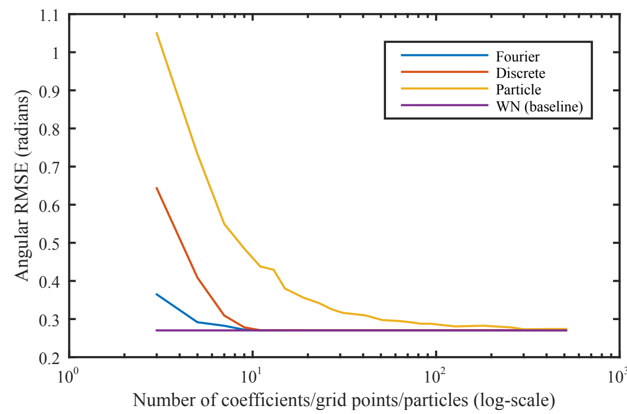


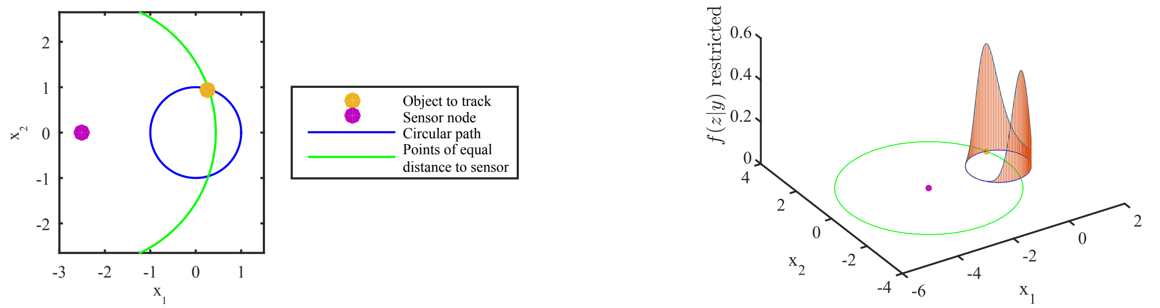
Figure 3. ARMSE of the filters for several numbers of coefficients, grid points, and particles.

The results are shown in Figure 3. In this scenario, prediction steps can be performed optimally by the WN filter, whereas an approximation is necessary in the filter step. However, since this filter handles such scenarios very well, we use it as a baseline to evaluate the performance of the other filters.

The universal filters differed in their convergence behavior. The performance of the discrete filter deviated less than 1% from baseline when 13 or more grid points were used. Our proposed Fourier-based approach attained this quality using 11 coefficients. This performance was not attained by the particle filter, even when using 500 particles. So all in all, the discrete filter and our Fourier-based approach attained good quality in this easy scenario, while the particle filter showed very slow convergence.

B. Bimodal Scenario

In the second and more complicated scenario, we obtain distance measurements instead of angular measurements. The scenario is shown in Figure 4a: the object moves along a known path shown as a blue circle. The sensor node at a fixed point 2.5 units away from the center of the blue circle can only measure the distance to the object. Based on a single distance measurement, all points on a whole circle in green are, if we disregard noise, possible positions of the measured object. All distance measurements are perturbed by Gaussian noise $\boldsymbol{v} \sim \mathcal{N}(\mu = 0, \sigma = 0.3)$. This additional uncertainty leads to a multimodal likelihood when restricted to the circle. The likelihood can be seen for one time step in Figure 4b. The distance between the modes depends on the position of the object, with the modes fusing when the object is close to the point with the smallest or the largest distance to the sensor.



(a) Sensor node, the object on its circular path, and a circle showing points with a certain distance to the sensor node.

(b) Bimodal measurement likelihood caused by the ambiguity of the distance measurement.

Figure 4. Overview over the bimodal scenario and the occurring likelihood.

The scenario was run 100 times and the performance was evaluated at the end of the 40th time step. To handle the bimodal likelihood, the WN Filter had to be used in its extended form for nonlinear problems [4]. For a comprehensive evaluation, we used two different criteria involving the pdf and the cumulative distribution function (cdf), allowing for better evaluation in this multimodal scenario. In both cases, we used a discrete filter with 1000 grid points as ground truth. We chose to use a definitely converging filter with a high number of components other than the Fourier filter to prevent any effects that may be favorable to our proposed filter. As the ground truth was created using 1000 grid points, we limited our evaluation to at most 500 grid points to prevent that the grid-based approach benefits too much from the fact that the filter and the approximation of the ground truth deviate from the actual true density in a similar fashion.

a) Kullback–Leibler Divergence: As the first criterion, we used the Kullback–Leibler divergence (KLD)

$$D_{KL}(f_{gt}||f_{approx}) = \int_{-\infty}^{\infty} f_{gt}(x) \ln\left(\frac{f_{gt}(x)}{f_{approx}(x)}\right) dx$$

between the (approximate) ground truth f_{gt} and the approximation f_{approx}.

In order to calculate the KLD using numerical integration, we require continuous densities. While a continuous density can always be derived for the Fourier-based approach, the discrete filter and the particle filter require some kind of conversion. To obtain a continuous density from the discrete filter, we derived a piecewise constant function from the discrete values. Matters are more difficult for the particle filter as there is no inherent continuous density associated with the filter result. Out of different approaches considered, we chose to use the conversion that we deemed to be the most favorable to the particle filter. First, we merged particles at numerically indistinguishable angles by replacing them with one particle that subsumes their weight. Then, we built Voronoi regions around the particles and distributed their weight evenly on the resulting intervals.

The results are shown in Figure 5 with logarithmic scales on both axes. The particle filter does not show definite convergence towards the ground truth while both the discrete filter and our proposed Fourier-based approach converged to the true density. Due to numerical issues and the lack of an exact ground truth, the Fourier filter could not improve beyond an

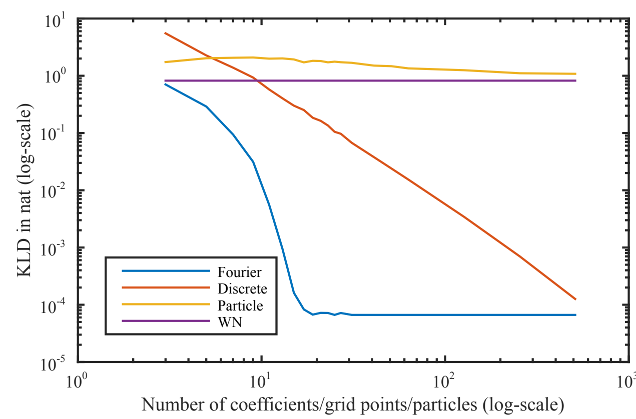


Figure 5. Mean of the KLD between the densities derived from the discrete filter with 1000 grid points and the densities derived from the respective filter results.

average KLD of approximately $5 \cdot 10^{-5}$. This performance is already achieved using only 17 Fourier coefficients, whereas the discrete filter is still not as close of an approximation with 500 grid points. The WN filter did not perform well, which was to be expected as the density is approximated by a unimodal WN in every time step.

b) Comparison of the Squared Integrated Difference of the Cumulative Distribution Function: Since deriving a cdf in one dimensional spaces is more straightforward for particle filters than generating a pdf, we performed a second evaluation comparing the squared integrated difference of the cdfs. However, to derive a cdf on periodic domains, it is necessary to specify a point at which the integration of the probability mass is started. As probability mass close to the border of periodicity plays a significant role when comparing cumulative distributions on periodic domains, we searched the probability mass function of the discrete filter used as the ground truth for a region of low probability mass. Out of the 1000 grid points, 101 subsequent points were searched that sum up to the lowest mass and the center was used as the starting point of the cumulation. Using this starting point, the cdfs for the discrete filter and for the particle filter were step functions that cumulate the weights from the starting point on. The cdf for the Fourier filter was obtained by using appropriate integration rules. The calculation of the squared integrated difference between the resulting cdfs of the ground truth and

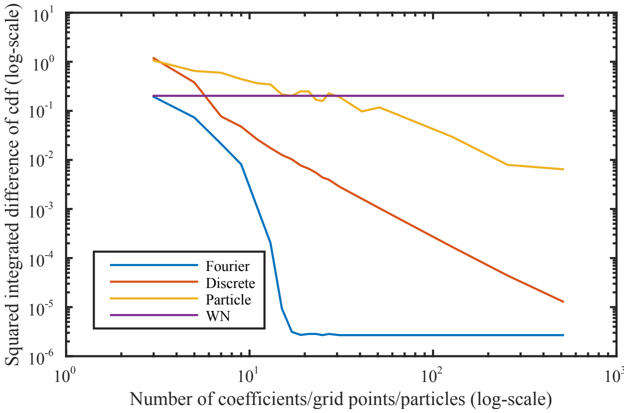


Figure 6. Mean of the squared integrated difference between the cdfs derived from the discrete filter with 1000 grid points and the cdfs derived from the respective filter results.

of the filter results was performed using numerical integration.

The results are shown, again using log scales, in Figure 6. Again, our proposed Fourier filter achieves what we believe to be optimal performance with only 17 Fourier coefficients. This is neither achieved by the discrete filter with 500 grid points nor by the particle filter with 500 particles.

All in all, our proposed filter needs far fewer coefficients than grid points are needed by the discrete filter and particles are needed by the particle filter. Since the discrete filter is not cheap when dealing with arbitrary densities, the Fourier filter with 17 components is far superior to the discrete filter with 500 grid points, while still achieving better results for both criteria. In our evaluation, the particle filter performed very badly per particle for the KLD and was orders of magnitude worse when the cdfs were compared. While a detailed run time analysis is out of scope for this paper, our unoptimized Matlab code was already fast enough for many real time applications. On a desktop PC with an Intel® Core™ i7-2700K, a test ran at 483 iterations of predictions and filter steps per second when using 101 Fourier coefficients.

VII. Conclusions and Discussion

In conclusion, filtering based on Fourier series approximations of the square root of densities allows obtaining valid, continuous approximations of the true posterior density. We showed that many common densities can be represented well in this form as their Fourier coefficients fall off rapidly asymptotically. The proposed filter is fully deterministic and can handle even multimodal densities well.

As a limitation, the quality of the filter depends on how well the density can be approximated using truncated Fourier series. This depends, for example, on how narrow the peaks of the distribution are and can result in a need for more coefficients for densities with lower variance. However, these densities are also hard to approximate using grid-based approaches while the particle filter also suffers from particle degeneracy in these cases. As the proposed approach has an asymptotic run time complexity of $O(n \log n)$ and requires significantly fewer coefficients than the discrete filter requires grid points, we deem it to be the more efficient solution for estimating the entire density.

While the Fourier filter outperformed the other filters in regard to run time in our implementation, a more thorough evaluation of the computational effort when using reasonable numbers of coefficients, grid points, and particles could be performed as future work. As the need for high performance filters is even greater for higher dimensions, we plan to publish an extension of the proposed filter for higher dimensions.

References

[1] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, “A Recursive Fusion Filter for Angular Data,” in *Proceedings of the 2009 International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2009.

[2] 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.

[3] ——, “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.

[4] ——, “Recursive Bayesian Filtering in Circular State Spaces,” *arXiv preprint: Systems and Control (cs.SY)*, Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.05151

[5] G. Kurz, I. Gilitschenski, S. J. Julier, and U. D. Hanebeck, “Recursive Estimation of Orientation Based on the Bingham Distribution,” in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.

[6] J. Glover, R. Rusu, and G. Bradski, “Monte Carlo Pose Estimation with Quaternion Kernels and the Bingham Distribution,” in *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

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

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

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

[10] D. Brunn, F. Sawo, and U. 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*, Sep. 2006.

[11] ——, “Nonlinear Multidimensional Bayesian Estimation with Fourier Densities,” in *Proceedings of the 2006 45th IEEE Conference on Decision and Control*, Dec. 2006.

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

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

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

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

[16] G. Strang, *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007.

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

[18] S. R. Jammalamadaka and T. J. Kozubowski, “New Families of Wrapped Distributions for Modeling Skew Circular Data,” *Communications in Statistics - Theory and Methods*, vol. 33, no. 9, 2004.

[19] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Cengage Learning, 2001.

[20] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[21] M. S. 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, 2002.