

# Efficient Nonlinear Bayesian Estimation based on Fourier Densities

Dietrich Brunn, Felix Sawo, and Uwe D. Hanebeck

**Abstract**—Efficiently implementing nonlinear Bayesian estimators is still not a fully solved problem. For practical applications, a trade-off between estimation quality and demand on computational resources has to be found. In this paper, the use of nonnegative Fourier series, so-called Fourier densities, for Bayesian estimation is proposed. By using the absolute square of Fourier series for the density representation, it is ensured that the density stays nonnegative. Nonetheless, approximation of arbitrary probability density functions can be made by using the Fourier integral formula. An efficient Bayesian estimator algorithm with constant complexity for nonnegative Fourier series is derived and demonstrated by means of an example.

## I. INTRODUCTION

A common challenge in many technical systems is the problem of reconstructing unknown quantities from imprecise measurements. Typical examples are localization of vehicles or reconstruction of distributed phenomena by means of a sensor network.

The most common approach is to describe the uncertain quantities with random variables, which results in a Bayesian estimator. For linear systems with Gaussian random variables, the problem can be fully solved using the Kalman Filter [1], [2]. For the case of nonlinear systems, a generalized convolution integral has to be evaluated for the prediction step, which typically cannot be solved analytically. Furthermore, with an increasing number of measurements, the complexity of the density representation generally increases. This leads to unbounded computational demands for recursive Bayesian processing. For quite some time a vast variety of approaches exist [2], [3] to overcome this problem by employing appropriate approximations of the needed probability density functions. Very popular are the extended Kalman filter [2], particle filters [4], and set-based methods [5]. They all have in common that they are efficient, but determining the quality of their results is generally difficult and usually needs high computational costs.

To overcome this drawback, systems of functions can be employed for approximating probability densities. For that purpose, Gaussian Mixtures [6] and Gauss-Hermite series [7] are commonly used. Although methods exist for precisely determining Gaussian mixtures approximations for arbitrary density functions [8], approximating a density function, including reapproximating other Gaussian mixtures, has high computational demands as all parameters are interdependent.

Edgeworth-expansions [7] are much more efficient on this point, due to the fact that Gauss-Hermite series are,

in contrast to Gaussian mixtures, an orthogonal function system, which permits determining the required parameters independently of each other. Furthermore, the parameters can be ordered with respect to a distance measure. That reduces the computational demands drastically. Unfortunately, it cannot be generally ensured that a truncated Gauss-Hermite series is a valid density function as it can become negative.

We propose an alternative filtering approach, by employing Fourier series, which were proposed in [9] for estimating probability densities. When truncating Fourier series, again, nonnegativity cannot be ensured. Hence, we approximate the square root of all occurring probability densities by Fourier expansions. By taking the absolute square of those series, the so-called *Fourier probability density functions* (short: *Fourier densities*), the nonnegativity, i.e., a valid probability density, can be ensured. Exploiting the property that Fourier series are orthogonal expansions, it will be shown that required coefficients can be calculated independently and very effectively by evaluating a Fourier integral.

Furthermore, a recursive Bayesian estimator consisting of a filtering step and a prediction step will be derived. Employing Fourier densities Bayesian estimation can be performed very efficiently, since all required expressions can be calculated analytically and the type of probability density is preserved. Additionally, the complexity, i.e., the computational demands, stay constant, because the order of the Fourier density of the predicted density is independent of the order of the prior density. It can even be adjusted, because the length of the Fourier series can be reduced optimally with respect to the Hellinger metric, which is the employed distance measure.

The rest of the paper is structured as follows: We begin with discussing the Bayesian filter in the following section. In Sec. III Fourier probability densities are introduced. Sec. IV explains how to approximate arbitrary probability density functions with Fourier densities. Sec. V demonstrates the procedure of Bayesian filtering with Fourier densities by means of an example. Sec. VI summarizes results and gives an outlook to future work.

## II. PROBLEM FORMULATION

The main goal is to find an efficient realization of the Bayesian estimator, which is discussed in this section.

For brevity and clarity, only the case of one-dimensional random variables is discussed. Without loss of generality, finite intervals are limited to the interval  $\Omega = [-\pi, \pi]$  unless otherwise noted. The following notation is used:

Dietrich Brunn, Felix Sawo, and Uwe D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory, Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany. {brunn|sawo}@ira.uka.de, uwe.hanebeck@ieee.org

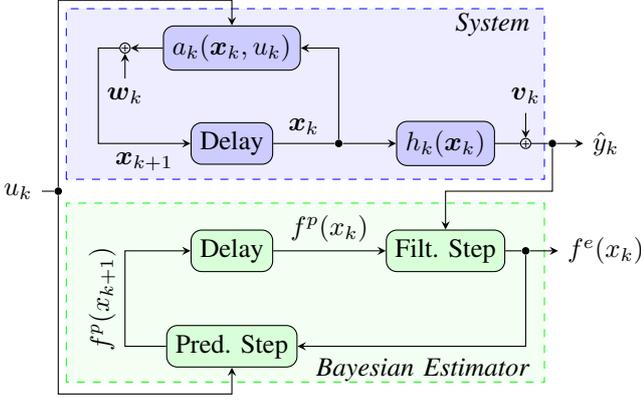


Fig. 1. Structure of a nonlinear discrete-time system with additive noise and a nonlinear discrete-time estimator. The output  $\hat{y}_k$  is a realization of the random variable  $\mathbf{y}_k$ .

$\mathbf{x}$  – random variable       $\mathbb{C}$  – set of complex numbers  
 $x_k$  –  $x$  at discr. time  $k$      $c^*$  – conjugate complex of  $c$   
 $j$  – imaginary unit       $\delta(x)$  – Dirac delta function

Consider a nonlinear discrete-time scalar system

$$\mathbf{x}_{k+1} = a_k(\mathbf{x}_k, u_k) + \mathbf{w}_k, \quad (1)$$

with state variable  $\mathbf{x}_k \in \Omega = [-\pi, \pi]$ , input  $u_k$ , the nonlinear system function  $a_k(\cdot, \cdot)$ , and additive noise  $\mathbf{w}_k$  with the probability density  $f_w(w)$ . The subscript  $k$  denotes the discrete time and bold variables, e.g.  $\mathbf{x}$ , denote random variables. Furthermore, we consider the measurement equation

$$\mathbf{y}_k = h_k(\mathbf{x}_k) + \mathbf{v}_k, \quad (2)$$

where  $\mathbf{y}_k$  is the measurement value at time  $k$ ,  $h(\cdot)$  the nonlinear measurement function and  $\mathbf{v}_k$  the additive noise with the density  $f_v(v_k)$ . Note that an actual measurement, denoted with  $\hat{y}_k$ , is a realization of (2). It is assumed that all the individual noise variables  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are statistically independent and have an expected value of zero.

The purpose of the estimator is to determine the probability density  $f(x_k)$  of  $\mathbf{x}_k$  as precise as possible for each time step  $k$ . To achieve this, two steps are performed alternately (Fig. 1), namely the prediction step and the filtering step.

**Prediction Step:** A prior density  $f(x_k)$  for  $\mathbf{x}_k$ , a known input  $u_k$ , and a noise density  $f_w(w_k)$  is assumed. With respecting (1) for the discrete time  $k+1$  the density  $f^p(x_{k+1})$  for  $\mathbf{x}_{k+1}$  can be determined by employing the well-known generalized convolution formula

$$\begin{aligned} f^p(x_{k+1}) &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \delta(x_{k+1} - a_k(\mathbf{x}_k, u_k) - w_k) \\ &\quad f(x_k) f_w(w_k) dw_k dx_k \\ &= \int_{-\pi}^{\pi} f_w(x_{k+1} - a_k(\mathbf{x}_k, u_k)) f(x_k) dx_k. \end{aligned} \quad (3)$$

The input for the prediction step is usually the estimated density  $f(x_k) = f^e(x_k)$  from the filtering step.

**Filtering Step:** The purpose is to incorporate the information of the measurement value  $\hat{y}_k$  including the measurement noise  $f_v(v_k)$  respecting (2). This can be achieved by using

the famous Bayes formula for additive noise

$$f^e(x_k) = \frac{f_v(\hat{y}_k - h(x_k)) f^p(x_k)}{\int_{\Omega} f_v(\hat{y}_k - h(x_k)) f^p(x_k) dx_k} = \frac{f^L(x_k) f^p(x_k)}{c_k}, \quad (4)$$

where

$$\begin{aligned} f^L(x_k) &= f(\hat{y}_k | x_k) = \int_{-\pi}^{\pi} \delta(\hat{y}_k - h(x_k) - v_k) f_v(v_k) dv_k \\ &= f_v(\hat{y}_k - h(x_k)) \end{aligned}$$

is called the likelihood function for the measurement value  $\hat{y}_k$ . It is the conditional density for the occurrence of the measurement  $\hat{y}_k$  under the condition  $x_k$ . The denominator  $c_k$  is a normalization constant for  $f^L(x_k)$ .

### III. CHARACTERISTICS OF FOURIER DENSITIES

As a probability density representation, we use Fourier expansions. As shown in Sec. IV, they are constructed in a way that they are nonnegative.

**DEFINITION 1 (FOURIER DENSITIES)** Consider a Fourier expansion  $f_{\Psi} : x \in \Omega = [-\pi, \pi] \mapsto \mathbb{C}$  of the form

$$f_{\Psi}(x) = \sum_{k=-2n}^{2n} \gamma_k e^{jkx} = \sum_{k=-2n}^{2n} (\alpha_k + j\beta_k) e^{jkx}, \quad (5)$$

with  $\gamma_k = \alpha_k + j\beta_k \in \mathbb{C}$  and  $j^2 = -1$ .

If the fundamental probability density properties

$$f_{\Psi} : x \mapsto \mathbb{R}_0^+ := \{f_{\Psi} | f_{\Psi} \in \mathbb{R} \cap f_{\Psi} \geq 0\} \quad (6)$$

$$\int_{-\pi}^{\pi} f_{\Psi}(x) dx = 1 \quad (7)$$

hold, then  $f_{\Psi}$  is called a Fourier probability density function or short a Fourier density on  $\Omega$  of the order  $2n$ . If (7) is not true, then  $f_{\Psi}$  is called an unnormalized Fourier density.

Note that for any real Fourier series  $f_{\Psi}(x) \in \mathbb{R}$   $\gamma_{-k} = \gamma_k^*$  holds [10], thus also for Fourier densities.

**LEMMA 1 (CUMULATIVE DENSITY)** The cumulative density function of a Fourier density is given by

$$F(x) = \int_{-\pi}^x f_{\Psi}(\xi) d\xi \quad (8)$$

$$= \gamma_0(x + \pi) + \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{jk} (e^{jkx} - (-1)^k).$$

The most widely used characterizations of probability densities are the expected value and the variance.

**THEOREM 1 (EXPECTATION VALUE AND VARIANCE)** Consider a Fourier density  $f_{\Psi}(x)$  as given in (5). The expected value and the variance are

$$\hat{x} = E\{\mathbf{x}\} = \int_{-\pi}^{\pi} \xi f_{\Psi}(\xi) d\xi = 2\pi \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{jk} (-1)^k, \quad (9)$$

$$\begin{aligned} \sigma_x^2 &= E\{(\mathbf{x} - \hat{x})^2\} = \\ &= \frac{2}{3} \gamma_0 \pi^3 - 4\pi \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{k^2} (-1)^k - \hat{x}^2. \end{aligned} \quad (10)$$

PROOF. Integration by parts gives

$$\begin{aligned}\hat{x} &= \int_{-\pi}^{\pi} \xi f_{\Psi}(\xi) d\xi = \xi F_{\Psi}(\xi) \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} 1 F_{\Psi}(\xi) d\xi \\ &= \left[ \gamma_0 \xi^2 + \xi \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{jk} (-1)^k \right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} \gamma_0 \xi + \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{jk} e^{jk\xi} d\xi \\ &= 2\pi \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{jk} (-1)^k - \left[ \gamma_0 \frac{\xi^2}{2} + \sum_{\substack{k=-2n \\ k \neq 0}}^{2n} \frac{\gamma_k}{-k^2} e^{jk\xi} \right]_{-\pi}^{\pi}\end{aligned}$$

equals (9), because  $e^{jk(\pi)} = e^{jk(-\pi)}$  holds for all integer values  $k \in \mathbb{Z}$ . Obviously, for the variance

$$\begin{aligned}\sigma_x^2 &= E\{(\mathbf{x} - \hat{x})^2\} = E\{\mathbf{x}^2 - 2\mathbf{x}\hat{x} + \hat{x}^2\} \\ &= E\{\mathbf{x}^2\} - E\{2\mathbf{x}\hat{x}\} + E\{\hat{x}^2\} = E\{\mathbf{x}^2\} - \hat{x}^2\end{aligned}$$

holds, so it is sufficient to calculate the central moment  $E\{\mathbf{x}^2\}$ . Again, integration by parts results in (10).  $\square$

As shown in (3) and (4), the Bayesian estimator uses two-dimensional probability densities. Consequently, two-dimensional Fourier densities are needed as well.

**DEFINITION 2 (2D FOURIER DENSITY)** A two-dimensional Fourier density  $f_{\Psi} : (x, y) \in \Omega^2 = [-\pi, \pi]^2 \mapsto \mathbb{R}_0^+$  of the order  $n \times m$  is defined by

$$f_{\Psi}(x, y) = \sum_{k,l} \gamma_{k,l} e^{j(kx+ly)}, \quad (11)$$

with  $\gamma_{k,l} \in \mathbb{C}$ ,  $k = -2n \dots 2n$  and  $l = -2m \dots 2m$ . Obviously, it needs to fulfill

$$f_{\Psi}(x, y) \geq 0 \quad \forall (x, y) \in \Omega^2 \quad \text{and} \quad \iint_{\Omega^2} f_{\Psi}(x, y) dx dy = 1.$$

Again, if the integral condition does not hold,  $f_{\Psi}(x, y)$  is called an unnormalized Fourier density. This definition can be easily generalized to higher dimensions. Analogous expressions to Theorem 1 can be derived as well.

#### IV. APPROXIMATING ARBITRARY DENSITIES WITH FOURIER DENSITIES

For representing arbitrary probability densities  $f(x)$  we need a method for determining an appropriate approximation by a  $n$ -th order Fourier density  $f_{\Psi}(x)$ . This, we interpret as an optimization problem

$$G(f, f_{\Psi}) \rightarrow \min,$$

where the parameters of  $f_{\Psi}(x)$  are adapted to minimize the quality measure  $G(\cdot, \cdot)$ , which compares the similarity of the two densities  $f(x)$  and  $f_{\Psi}(x)$ . Choosing an appropriate  $G(\cdot, \cdot)$  is a difficult task, because the quality of an approximation heavily depends on the application. E.g., optimizing for the expected value or for a certain fixed interval of the density require different measures. Deriving optimal quality measures in general is still, as far the authors know, an unsolved problem in estimation theory. Hence, we are satisfied with a measure, which is zero for identical

densities  $G(f, f) = 0$  and reflects similarity in the shape of the densities well.

In this work we utilize the Hellinger metric

$$G(f, f_{\Psi}) = \int_{-\pi}^{\pi} \left( \sqrt{f(x)} - \sqrt{f_{\Psi}(x)} \right)^2 dx,$$

which is employed for density estimation [11]. The Hellinger metric has the advantage over the usually used integral squared deviation that the coefficients can be determined independently of each other, which will be shown in the following. Furthermore, it can be shown that the Hellinger metric is an optimal choice for some classes of applications. The investigation of that is beyond the scope of this paper.

In our case, we do not determine the parameters of the Fourier density  $f_{\Psi}$  directly. In analogy to quantum mechanics, where the absolute square of  $\Psi$ -function are used to represent probability densities [12], we define  $\Psi$ -densities:

**DEFINITION 3 ( $\Psi$ -DENSITY)** A two-dimensional Fourier series  $\Psi : (x, y) \in \Omega^2 = [-\pi, \pi]^2 \mapsto \mathbb{C}$  which is given by

$$\Psi(x, y) = \sum_{k,l} c_{k,l} e^{j(kx+ly)},$$

with  $c_{k,l} \in \mathbb{C}$ ,  $k = -n \dots n$ ,  $l = -m \dots m$ , for which the absolute square

$$\Psi(x, y) \Psi^*(x, y) = |\Psi(x, y)|^2 = f_{\Psi}(x, y)$$

results in a Fourier density is called a  $\Psi$ -density. By setting  $y \equiv 0$ ,  $m \equiv 0$ , the one-dimensional case is obtained. Higher-dimensional cases can be defined analogously.

With  $\Psi$ -densities we are able to derive an approximation rule for the one and two-dimensional case.

**THEOREM 2 (APPROXIMATION)** Given a two-dimensional probability density function  $f : (x, y) \in \Omega^2 = [-\pi, \pi]^2 \mapsto \mathbb{R}_0^+$ , the optimal coefficients of the  $\Psi$ -density with respect to the Hellinger metric

$$G(f, f_{\Psi}) = \iint_{\Omega^2} \left( \sqrt{f(x, y)} - \Psi(x, y) \right)^2 dx dy \quad (12)$$

are given by the Fourier integral

$$c_{k,l} = a_{k,l} + jb_{k,l} = \frac{1}{4\pi^2} \iint_{\Omega^2} \sqrt{f(x, y)} e^{-j(kx+ly)} dx dy. \quad (13)$$

The Fourier density  $f_{\Psi}(x, y) = \Psi(x, y) \Psi^*(x, y)$  can be obtained, by taking the absolute square of  $\Psi(x, y)$ . Again, by setting  $y \equiv 0$  the one-dimensional case is obtained.

**PROOF.** For determining a minimum, the first derivative<sup>1</sup>  $G_{-k,-l} = \partial G / \partial a_{-k,-l}$  has to be zero, which results in

$$\begin{aligned}G_{-k,-l} &= \iint_{\Omega^2} \frac{\partial}{\partial a_{-k,-l}} \left( \sqrt{f(x, y)} - \Psi(x, y) \right)^2 dx dy \\ &= -2 \iint_{\Omega^2} \left( \sqrt{f(x, y)} - \Psi(x, y) \right) e^{-j(kx+ly)} dx dy.\end{aligned}$$

Since  $\int_{-\pi}^{\pi} e^{jkx} dx = 0$  for all integer  $k \neq 0$ , we obtain

$$G_{-k,-l} = -2 \iint_{\Omega^2} \sqrt{f(x, y)} e^{-j(kx+ly)} dx dy + 2(4\pi^2 c_{k,l}).$$

<sup>1</sup> $\partial / \partial c_{k,l}$  cannot be used, since  $\partial c_{k,l}^* / \partial c_{k,l}$  is not defined for complex  $c_{k,l} \in \mathbb{C}$ , but occurs when approximating real functions.

With setting  $G_{-k,-l} = 0$ , we obtain (13). Since it can easily be shown that  $\partial^2 G_{-k,-l} / \partial a_{-k,-l}^2$  is greater than zero, (13) is the condition for a minimum. The same result is obtained, when using  $b_{-k,-l}$  using instead of  $a_{-k,-l}$ .  $\square$

Note that the coefficients can be determined independently and there exist very efficient algorithms, like the fast Fourier transform [10], for calculating the Fourier integral of (13). That is the main reason for using the Hellinger metric. Other metrics, e.g. the integral squared deviation, could be employed as well. Comparing the quality of approximations of different metrics is a point of further research.

In the following, we discuss the special case of reapproximating a given Fourier density by a Fourier density with lower order. This is especially useful to reduce complexity, when performing filtering steps without prediction steps, which will be addressed in the next section. The reduction can be performed by first calculating the square root  $\sqrt{f_\Psi(x)}$  and then reducing the  $\Psi$ -density.

LEMMA 2 (CALCULATING  $\sqrt{f_\Psi(x)}$ ) *Given a valid Fourier probability density function  $f_\Psi(x)$ , if the coefficients  $c_k$  of  $\Psi(x)$  satisfy the following equations*

$$\begin{aligned} |c_n|^2 &= \gamma_{2n} \ , \\ 2c_n c_{n-1} + c_{n-1} c_n &= \gamma_{2n-1} \ , \end{aligned}$$

⋮

$$c_n c_{-n}^* + c_{n-1} c_{-n+1}^* + \dots + c_1 c_{-1}^* + |c_0|^2 = \gamma_0 \ ,$$

then  $f_\Psi(x) = \Psi(x)\Psi^*(x)$ . The first  $n$  equations can be solved sequentially from top to bottom, because the number of unknowns increase by two and each line represents two equations. The second  $n$  equations have to be used to ensure that the solution is correct.

The Hellinger metric implies an order on the coefficients of  $\Psi$ -density, as the next theorem shows.

THEOREM 3 (ORDERING FOURIER TERMS) *The coefficient  $c_{min}$  of a  $\Psi$ -density  $\Psi(x)$ , which influences the Hellinger metric minimally, has to satisfy*

$$c_{min} c_{min}^* \leq c_k c_k^* \text{ for } k \in [-n, n] \ , \quad (14)$$

where  $c_{min}$  is the  $c_k$  with the minimal influence. I.e., the coefficients  $c_k$  can be ordered with respect to the Hellinger metric by ordering them by their squared magnitude.

PROOF. By applying a  $\Psi$ -density reduced by one element

$$\Psi_{\cancel{k}}(x) := \Psi(x) - c_k e^{jkx}$$

to the Hellinger metric, we obtain

$$G(\Psi\Psi^*, \Psi_{\cancel{k}}\Psi_{\cancel{k}}^*) = c_k c_k^* \ ,$$

Choosing the  $c_k$ , which changes  $G$  minimally leads to (14).  $\square$

We end this section with an two-dimensional example.

EXAMPLE 1 (2D APPROXIMATION) *Fig. 2 (top) depicts*

$$f(x, y) = \mathcal{N}(y - 2x^3 - x, \sigma^2)$$

with  $\sigma^2 = \frac{1}{2}$  and the Gaussian

$$\mathcal{N}(z - \hat{z}, \sigma_z^2) = \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{1}{2} \frac{(z-\hat{z})^2}{\sigma_z^2}} \ . \quad (15)$$

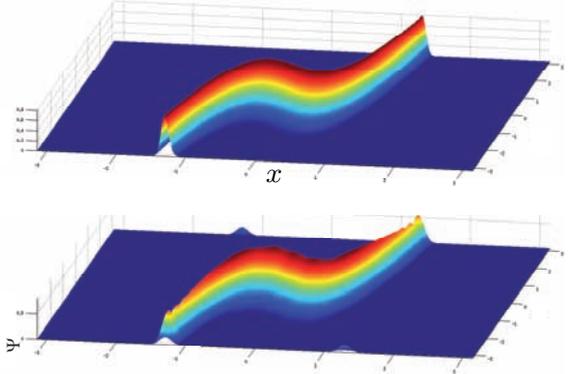


Fig. 2. The 2D-Density  $f(x, y) = \mathcal{N}(y - 2x^3 - x, \frac{1}{2})$  (top) and its Approximation with  $16 \times 16$ -order Fourier density (bottom).

The bottom shows the approximation with a  $16 \times 16$ -order Fourier density. The little bumps by  $(1.1, -\pi)$  and  $(-1.1, \pi)$  appear, since Fourier series are periodic functions and there is a sharp edge at  $y = \pm\pi$ . The bumpiness at the top and the bottom is caused by Gibbs phenomenon [10]. It can be reduced by using appropriate windowing functions or using Fourier densities of a higher order.

## V. NONLINEAR FILTERING WITH FOURIER DENSITIES

In this section we discuss how to perform Bayesian filtering with Fourier densities. We begin with the filtering step of (4). To perform it efficiently, we insert one constant function parameter in a two-dimensional Fourier density, which results in a one-dimensional slice.

THEOREM 4 (CALCULATING A SLICE) *Given a two-dimensional Fourier density  $f_\Psi(x, y)$  as in (11), the coefficients of the one-dimensional Fourier density*

$$f_\Psi(x, y = y_0) = f_\Psi^0(x) = \sum_{k=-2n}^{2n} \gamma_k^0 e^{jkx}$$

are given by

$$\gamma_k^0 = \sum_{l=-2m}^{2m} \gamma_{k,l} e^{jly_0} \ . \quad (16)$$

An analogous expression can be derived for  $f_\Psi(x = x_0, y = y)$  as well as for higher dimensional densities.

PROOF. Using  $e^{j(kx+ly)} = e^{jkx} e^{jly}$ , (16) can be easily derived from (11).  $\square$

Additionally, we need the product of two Fourier densities.

LEMMA 3 (PRODUCT) *Given are two Fourier densities  $f_\Psi^a(x)$ ,  $f_\Psi^b(x)$  of the order  $n_a$ ,  $n_b$ . For the product*

$$f_\Psi^a(x) f_\Psi^b(x) = f_\Psi^c(x) = \sum_{k=-2n_c}^{2n_c} \gamma_k^c e^{jkx} \quad (17)$$

the coefficients are given by

$$\gamma_k^c = \sum_{l=-n_a}^{n_a} \bar{\gamma}_l^a \bar{\gamma}_{l-k}^b \text{ with } \bar{\gamma}_p^{(\cdot)} = \begin{cases} \gamma_p^{(\cdot)} & \text{if } -n_{(\cdot)} \leq p \leq n_{(\cdot)} \ , \\ 0 & \text{otherwise,} \end{cases}$$

with  $n_c = n_a + n_b$ . The proof can be performed by reformulating (17). Note that an analogous expression can be derived for  $\Psi$ -densities.

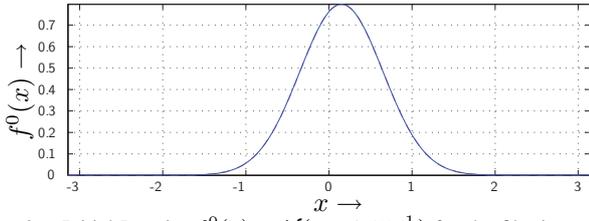


Fig. 3. Initial Density  $f^0(x) = \mathcal{N}(x - 0.15, \frac{1}{2})$  for the filtering step.

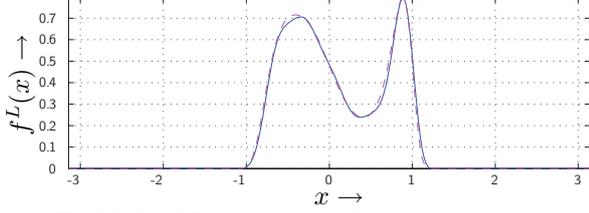


Fig. 4. The Likelihood Fourier series for the measurement equation (18) and the measurement value  $\hat{y}_1 = \frac{1}{2}$ . The dashed magenta line depicts the true function, which was determined numerically.

The filter step is demonstrated by means of an example.

**EXAMPLE 2 (FILTER STEP)** Given is the measurement equation consisting of a third order polynomial

$$\mathbf{y}_k = 2\mathbf{x}_k^3 + \mathbf{x}_k + \mathbf{v}_k, \quad (18)$$

with  $\mathbf{v}_k$  being zero mean Gaussian noise as in (15) with variance  $\sigma_v^2 = \frac{1}{2}$ . Furthermore, given is an initial Fourier density, which is an approximation of a Gaussian  $f^0(x) = \mathcal{N}(x - 0.15, \frac{1}{2})$  of the order 8 (Fig. 3). The maximum deviation to the true density is  $3 \cdot 10^{-6}$ . The filtering step is a three step procedure:

1) The first step, which can be performed off-line in time-invariant systems, is to determine the two-dimensional conditional Fourier density  $f_\Psi(y|x)$  from the measurement equation (18). Since  $\mathbf{v}_k$  is Gaussian, the true density is

$$f(y_k|x_k) = \mathcal{N}(y_k - 2x_k^3 - x_k, \sigma_v^2), \quad (19)$$

which has to be approximated by a 2D Fourier density. Fig. 2 depicts an approximation of the order  $16 \times 16$ .

2) The second step is to calculate the likelihood for a given measurement  $\hat{y}_1$ , which results in

$$f^L(x_1) = f(\hat{y}_1|x_1),$$

which can be performed by employing Theorem 4. Here, we assume  $\hat{y}_1 = \frac{1}{2}$ , where the corresponding likelihood Fourier series of the order 16 is depicted in Fig. 4.

3) The final step is performing the filtering step described in (4) using Lemma 3, which results in Fig. 5. The resulting density  $f_\Psi^e(x_1)$  has the order 24.

As expected, the density  $f_\Psi^e(x_1)$  is more complex than the previous density  $f_\Psi^0(x_0)$ . Thus, the order is higher as well.

Note that the order of the Fourier densities increases with each filtering step. In Theorem 5 we will show that the prediction step bounds the order of the Fourier density. When performing multiple sequential filter steps, the order can become very large. To limit the length, the corresponding  $\Psi$ -densities can be determined with Lemma 2 and then reapproximated using Theorem 3.

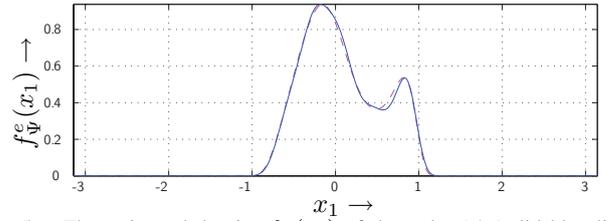


Fig. 5. The estimated density  $f_\Psi^e(x_1)$  of the order 24 (solid blue line: Fourier; dashed magenta line: true).

*Remark:* The dashed magenta lines in Fig. 4–7 show the corresponding numeric calculations with 256 grid points for each axis, which are assumed to represent the true functions.

We continue with deriving the prediction step.

**THEOREM 5 (PREDICTION STEP)** Given is a density

$$f(x_k) = \sum_l \gamma_l^{(k)} e^{jl x_k}$$

with  $l = -2n_l \dots 2n_l$  and a transition density

$$f^T(x_{k+1}, x_k) = f(x_{k+1}|x_k) = \sum_{p,q} \gamma_{p,q}^{(T)} e^{j(p x_{k+1} + q x_k)}$$

with  $p = -2n_p \dots 2n_p$  and  $q = -2n_q \dots 2n_q$ . With  $r = -2n_r \dots 2n_r$ , the predicted density results in

$$\begin{aligned} f^P(x_{k+1}) &= \int_{-\pi}^{\pi} f^T(x_{k+1}, x_k) f(x_k) dx_k = \\ &= \sum_p \gamma_p^{(k+1)} e^{j p x_{k+1}} \quad \text{with} \quad \gamma_p^{(k+1)} = 2\pi \sum_r \gamma_{-r}^{(k)} \gamma_{p,r}^{(T)}. \end{aligned} \quad (20)$$

**PROOF.** Obviously,

$$\begin{aligned} f^P(x_{k+1}) &= \int_{-\pi}^{\pi} f^T(x_{k+1}, x_k) f(x_k) dx_k \\ &= \int_{-\pi}^{\pi} \sum_{l,p,q} \gamma_l^{(k)} \gamma_{p,q}^{(T)} e^{j(l+q)x_k} e^{j p x_{k+1}} dx_k \\ &= \sum_{l,p,q} \left( \gamma_l^{(k)} \gamma_{p,q}^{(T)} e^{j p x_{k+1}} \int_{-\pi}^{\pi} e^{j(l+q)x_k} dx_k \right). \end{aligned} \quad (21)$$

Since  $\int_{-\pi}^{\pi} e^{j l x} dx = 0$  holds for all integers  $l \neq 0$ , (21) simplifies to

$$f^P(x_{k+1}) = 2\pi \sum_{p,r} \gamma_{-r}^{(k)} \gamma_{p,r}^{(T)} e^{j p x_{k+1}}$$

with  $r = -2n_r \dots 2n_r$ . This can be rewritten to (20).  $\square$

Note that the order of the predicted density is independent of the order of  $f(x_k)$ , i.e., for Bayesian prediction the complexity of the density stays constant. The following example demonstrates the utilization of the latter theorem.

**EXAMPLE 3 (PREDICTION STEP)** Given is a nonlinear system equation

$$\mathbf{x}_{k+1} = \frac{\mathbf{x}_k + 1}{2} + 25 \frac{\mathbf{x}_k + 1}{1 + (\mathbf{x}_k)^2} + \mathbf{w}_k, \quad (22)$$

without input. This is an adapted version of the nonstationary growth model, investigated by [13].  $\mathbf{w}_k$  is Gaussian  $f(\mathbf{w}_k) = \mathcal{N}(\mathbf{w}_k, \sigma_w^2)$  with variance  $\sigma_w^2 = \frac{1}{2}$ . The prior density is the result of the filter step of Example 2 depicted in Fig. 5.

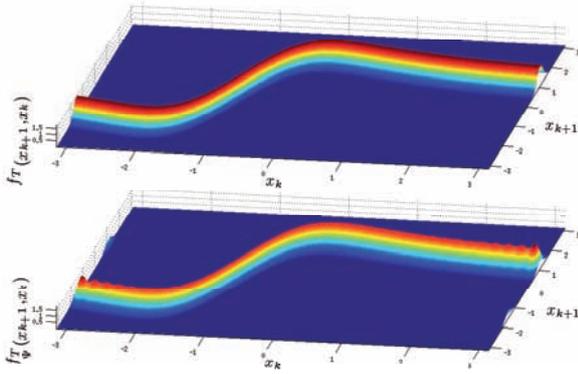


Fig. 6. The 2D-Density  $f^T(x_{k+1}, x_k)$  of Example 3. The top shows the real density and the bottom its approximation  $f_{\Psi}^T(x_{k+1}, x_k)$  with  $32 \times 32$ -order Fourier density (bottom).

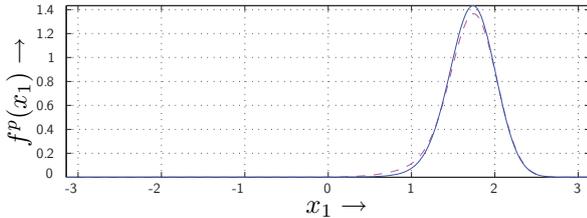


Fig. 7. The predicted density of Example 3 of the order 16 (solid blue line: Fourier; dashed magenta line: true).

The prediction step consists also of two steps. The first step, which can be performed off-line for time-invariant systems, is determining the approximation  $f_{\Psi}^T(x_{k+1}, x_k)$  of the transition density of (22)

$$f^T(x_{k+1}, x_k) = \mathcal{N}\left(x_{k+1} - \frac{x_k + 1}{2} - 25 \frac{x_k + 1}{1 + (x_k)^2}, \sigma_w^2\right)$$

with use of Theorem 2. The result of a  $32 \times 32$ -order Fourier density is depicted in Fig. 6.

The second step, which is always performed on-line, is the application of (20) to determine the resulting predicted density  $f^p(x_1)$ , which results in Fig. 7.

For treating systems with an input, a three dimensional  $f_{\Psi}^T(x_{k+1}, x_k, u_k)$  is needed, which is reduced to a two-dimensional density  $f_{\Psi}^T(x_{k+1}, x_k) = f_{\Psi}^T(x_{k+1}, x_k, \hat{u}_k)$  for a given  $\hat{u}_k$  (Theorem 4).

## VI. CONCLUSIONS AND FUTURE WORK

In this work, Bayesian estimation with Fourier densities was proposed. In Sec. II the Bayesian estimator was reviewed. In Sec. III Fourier densities as probability densities were introduced and some of their properties were discussed.

Sec. IV discussed the procedure of approximating arbitrary probability densities with Fourier densities. The Fourier coefficients can be determined independently evaluating a Fourier integral. With a fast Fourier transform, the coefficients can be calculated very efficiently. Additionally, it was shown that the complexity of a Fourier density can easily be reduced, since an ordering of coefficients exists. This allows to adjust the computational demands.

Sec. V derived a recursive Bayesian estimator consisting of a filtering step and a prediction step. It was addressed that

the prediction step bounds the complexity of the estimator. Furthermore, it was shown that the complexity, i.e., the order of the Fourier density, can be adjusted optimally with respect to the Hellinger metric. This is helpful for performing a large number of filter steps without intermediate prediction steps. Additionally, it was discussed that for time-invariant systems, the computationally expensive approximations can be done off-line. The on-line part can be performed very efficiently, since only multiplications and summations of coefficients are required.

A main aspect of future work includes evaluating the performance of Fourier densities on real life problems. Another aspect is tackling the problem of the fixed size of state space. One approach might be the use of an slideable and scalable state space, which could be realized by simple coordinate transformations. As already stated, a further point is the investigation of different quality measures for approximating densities. A final aspect is improving the approximation efficiency. Some classes of densities, especially those whose Fourier spectra have a lot of mass in high frequencies, Fourier densities of very high order are needed. Some other kinds of orthogonal function systems, e.g. orthogonal polynomials, might perform better in approximating those classes. A final result could be a family of orthogonal probability densities, which allow efficient Bayesian estimation for arbitrary nonlinear real life problems.

## REFERENCES

- [1] R. E. Kalman, "A new Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, no. 82, pp. 35–45, 1960.
- [2] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.
- [3] E. J. Wegman, "Nonparametric Probability Density Estimation: I. A Summary of Available Methods," *Technometrics*, vol. 14, no. 3, p. 5, August 1972.
- [4] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer-Verlag, 2001, ch. 1, pp. 6–12. [Online]. Available: <http://www-sigproc.eng.cam.ac.uk/~ad2/book.html>
- [5] J. Horn, U. D. Hanebeck, K. Riegel, K. Heesche, and W. Hauptmann, "Nonlinear set-theoretic position estimation of cellular phones," *Proceedings of the European Control Conference (ECC'03)*, 2003.
- [6] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972.
- [7] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy, "Nonlinear Filtering via Generalized Edgeworth Series and Gauss-Hermite Quadrature," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1816–1820, 2000.
- [8] U. D. Hanebeck, K. Briechele, and A. Rauh, "Progressive Bayes: A New Framework for Nonlinear State Estimation," in *Proceedings of SPIE*, vol. 5099. AeroSense Symposium, 2003.
- [9] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by fourier series methods," *Journal of the American Statistical Association*, vol. 63, no. 323, pp. 925–952, September 1968. [Online]. Available: <http://links.jstor.org/sici?sici=0162-1459%28196809%2963%3A323%3C925%3ATEOPDA%3E2.0.CO%3B2-2>
- [10] E. W. Weisstein, 2006. [Online]. Available: <http://mathworld.wolfram.com/>
- [11] R. Beran, "Minimum Hellinger Distance Estimates for Parametric Models," *The Annals of Statistics*, vol. 5, no. 3, pp. 445–463, 1977.
- [12] M. Alonso and E. J. Finn, *Physics*. Addison-Wesley, 1992.
- [13] H. Tanizki, *Handbook of Statistics 21: Stochastic Processes and Simulation*. North-Holland, 2003, ch. Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Techniques: A Survey and Comparative Study.