

Nonlinear Multidimensional Bayesian Estimation with Fourier Densities

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

Abstract— Efficiently implementing nonlinear Bayesian estimators is still an unsolved problem, especially for the multidimensional case. A trade-off between estimation quality and demand on computational resources has to be found. Using multidimensional Fourier series as representation for probability density functions, so called Fourier densities, is proposed. To ensure non-negativity, the approximation is performed indirectly via Ψ -densities, of which the absolute square represent the Fourier density. It is shown that Ψ -densities can be determined using the efficient fast Fourier transform algorithm and their coefficients have an ordering with respect to the Hellinger metric. Furthermore, the multidimensional Bayesian estimator based on Fourier densities is derived in closed form. That allows an efficient realization of the Bayesian estimator where the demands on computational resources are adjustable.

I. INTRODUCTION

Many technical applications involve dealing with unknown quantities. A common problem is determining an unknown state from imprecise measurements. Typical examples for the so called state estimation or filtering problem are localization of vehicles or reconstruction of temperature distributions by means of a sensor network.

The most common approach is to describe the uncertain quantities with stochastic variables, i.e., random vectors, and employ 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 vast variety of approaches exist [2], [3] for quite some time. The key idea is to find an appropriate approximation for probability density functions, which allows an efficient implementation of the Bayesian estimator. Very popular are the extended Kalman filter, particle filters [4], and set based methods [5]. They all have in common that they are efficient, scale well with an increasing number of dimensions. Unfortunately, determining the quality of their results is generally difficult and usually computationally costly.

To overcome this drawback, systems of functions can be employed for approximating probability densities. Gaussian mixtures [6] and Gauss-Hermite series [7] are commonly used. Although methods exist for precisely determining Gaussian mixtures for arbitrary density functions [8], approximating a density function, including another Gaussian mixture, demands much computational resources as all parameters are interdependent.

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

orthogonal function system, which permit determining the needed parameters independently of each other. Furthermore, the parameters can be ordered with respect to a distance measure. Unfortunately, it cannot be ensured in general that a truncated Gauss-Hermite series is a valid density function as they can become negative.

In this paper we use Fourier series as representations for probability density functions, as first proposed by Kronmal and Tarter [9]. To ensure non-negativity, we approximate the density indirectly via its square root utilizing the Hellinger metric. This concept was also used for density approximation with wavelets [10]. Unfortunately employing wavelet representations does result in a closed form Bayesian estimator.

The contribution of this paper is a full generalization of the approach we presented in [11] to the d -dimensional case. Particularly, the cumulative distribution, expected value and the covariance as well as the multidimensional Bayesian estimator are derived.

In this paper the following notation is used:

\mathbf{x} – random variable	\mathbb{C} – set of complex numbers
\underline{x} – vector	δ_{qp} – Kronecker symbol
$\underline{\mathbf{x}}$ – random vector	$\underline{a} \odot \underline{b}$ – elementwise multiplication
$\underline{1}$ – vector of ones	c^* – conjugate complex of c
j – imaginary unit	$\delta(x)$ – Dirac delta function

The rest of the paper is structured as follows: In the next Section the Bayesian estimator for the d -dimensional case is discussed. In Sec. III the multidimensional Fourier densities are defined and their key characteristics are derived. In Sec. V the multidimensional Bayesian estimator based on Fourier densities is derived. A localization example is shown in Sec. VI. The paper closes with a summary and an outlook to future work.

II. PROBLEM FORMULATION

Without loss of generality, finite intervals are limited to the interval $\Omega := [-\pi, \pi]$ unless otherwise noted.

Consider a nonlinear discrete-time multidimensional system

$$\underline{\mathbf{x}}_{k+1} = \underline{a}_k(\underline{\mathbf{x}}_k, \underline{u}_k) + \underline{\mathbf{w}}_k, \quad (1)$$

with the d -dimensional state vector $\underline{\mathbf{x}}_k \in \Omega^d = [-\pi, \pi]^d$, the input vector \underline{u}_k , the nonlinear system function $\underline{a}_k(\cdot, \cdot)$, and additive noise term $\underline{\mathbf{w}}_k$ with the probability density $f^w(\underline{w}_k)$. The subscript k denotes the discrete time, bold variables, e.g. $\underline{\mathbf{x}}$, denote random variables, and underlining, e.g. \underline{x} , denotes

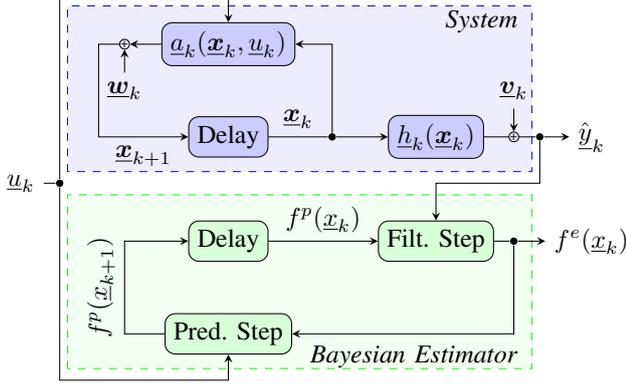


Fig. 1. Structure of a nonlinear discrete-time system with additive noise and a nonlinear discrete-time estimator. The output $\hat{\mathbf{y}}_k$ is a realization of the random variable \mathbf{y}_k .

vectors. Furthermore, we consider the measurement equation

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

where \mathbf{y}_k is the measurement vector at time k , $\mathbf{h}_k(\cdot)$ the nonlinear measurement function and \mathbf{v}_k the additive noise term with the density $f^v(\mathbf{v}_k)$. Note that an actual measurement, denoted with $\hat{\mathbf{y}}_k$, 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 $f(\mathbf{x}_k)$ 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 $f(\mathbf{x}_k)$ for \mathbf{x}_k , a known input \mathbf{u}_k , and a noise density $f^w(\mathbf{w}_k)$ are assumed to be given. With respecting (1) for the discrete time $k + 1$ the density $f^p(\mathbf{x}_{k+1})$ of \mathbf{x}_{k+1} can be determined by employing the well-known generalized convolution formula

$$\begin{aligned} f^p(\mathbf{x}_{k+1}) &= \int_{\Omega^d} \int_{\Omega^d} \delta(\mathbf{x}_{k+1} - \mathbf{a}_k(\mathbf{x}_k, \mathbf{u}_k) - \mathbf{w}_k) \\ &\quad f(\mathbf{x}_k) f^w(\mathbf{w}_k) d\mathbf{w}_k d\mathbf{x}_k \\ &= \int_{\Omega^d} f^w(\mathbf{x}_{k+1} - \mathbf{a}_k(\mathbf{x}_k, \mathbf{u}_k)) f(\mathbf{x}_k) d\mathbf{x}_k. \end{aligned} \quad (3)$$

The input for the prediction step is usually the estimated density $f(\mathbf{x}_k) = f^e(\mathbf{x}_k)$ from the filtering step.

Filtering Step: The purpose is to incorporate the information of the measurement value $\hat{\mathbf{y}}_k$ including the measurement noise $f^v(\mathbf{v}_k)$ respecting (2). This can be achieved by using the famous Bayes' formula for additive noise

$$\begin{aligned} f^e(\mathbf{x}_k) &= \frac{f^v(\hat{\mathbf{y}}_k - \mathbf{h}(\mathbf{x}_k)) f^p(\mathbf{x}_k)}{\int_{\Omega^d} f^v(\hat{\mathbf{y}}_k - \mathbf{h}(\mathbf{x}_k)) f^p(\mathbf{x}_k) d\mathbf{x}_k} \\ &= \frac{f^L(\mathbf{x}_k) f^p(\mathbf{x}_k)}{c_k}, \end{aligned}$$

where

$$\begin{aligned} f^L(\mathbf{x}_k) &= f(\hat{\mathbf{y}}_k | \mathbf{x}_k) = \int_{\Omega^d} \delta(\hat{\mathbf{y}}_k - \mathbf{h}(\mathbf{x}_k) - \mathbf{v}_k) f^v(\mathbf{v}_k) d\mathbf{v}_k \\ &= f^v(\hat{\mathbf{y}}_k - \mathbf{h}(\mathbf{x}_k)) \end{aligned}$$

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

III. MULTIDIMENSIONAL FOURIER DENSITIES

The use of Fourier series as density representation is very convenient. The function type is conserved over the operations $+$, $-$, \times , \int and $\sqrt{\cdot}$. It is shown in the previous section that these operations are sufficient for performing Bayesian estimation. Fourier series that can be interpreted as probability density functions are called Fourier densities.

DEFINITION 1 (FOURIER DENSITIES) Consider a d -dimensional Fourier expansion

$$f_\Psi : \mathbf{x} = [x_1, x_2 \dots x_d]^T \in \Omega^d := [-\pi, \pi]^d \mapsto \mathbb{C}$$

of the form

$$f_\Psi(\mathbf{x}) = \sum_{\mathbf{k} \in \mathcal{K}} \gamma_{\mathbf{k}} e^{j\mathbf{k}^T \mathbf{x}} = \sum_{\mathbf{k} \in \mathcal{K}} (\alpha_{\mathbf{k}} + j\beta_{\mathbf{k}}) e^{j\mathbf{k}^T \mathbf{x}}, \quad (4)$$

with $\gamma_{\mathbf{k}} = \alpha_{\mathbf{k}} + j\beta_{\mathbf{k}} \in \mathbb{C}$ and $j^2 = -1$. $\mathbf{k} = [\kappa_1, \kappa_2 \dots \kappa_d]^T \in \mathcal{K}$ is an index vector. When used as an index, it denotes multiple indices.

$$\begin{aligned} \mathcal{K} &= \{-\kappa_1^o, -\kappa_1^o + 1 \dots \kappa_1^o - 1, \kappa_1^o\} \\ &\quad \times \{-\kappa_2^o, -\kappa_2^o + 1 \dots \kappa_2^o - 1, \kappa_2^o\} \\ &\quad \times \dots \\ &\quad \times \{-\kappa_d^o, \kappa_d^o + 1 \dots \kappa_d^o - 1, \kappa_d^o\} \end{aligned}$$

denotes the set of all valid indices. Hence, (4) is a d -fold summation over the indices κ_1 from $-\kappa_1^o$ to κ_1^o , κ_2 from $-\kappa_2^o$ to κ_2^o , \dots , and κ_d from $-\kappa_d^o$ to κ_d^o . Note that the term $j\mathbf{k}^T \mathbf{x} = j \sum_{l=1}^d \kappa_l x_l$ denotes a scalar product.

If the fundamental probability density properties

$$\begin{aligned} f_\Psi : \mathbf{x} \mapsto \mathbb{R}_0^+ &:= \{f_\Psi | f_\Psi \in \mathbb{R} \cap f_\Psi \geq 0\} \\ \int_{\Omega^d} f_\Psi(\mathbf{x}) d\mathbf{x} &= 1 \end{aligned} \quad (5)$$

hold, then f_Ψ is called a Fourier probability density function or short a Fourier density on Ω^d of the order $\prod_{l=1}^d \kappa_l^o$. If (5) is not true, then f_Ψ is called an unnormalized Fourier density. Note that this definition can be generalized to arbitrary fixed intervals. This is omitted for the sake of brevity.

More important is the following theorem. It narrows the usable Fourier series to symmetric ones ($-\kappa_{(\cdot)}^o$ to $\kappa_{(\cdot)}^o$), i.e., a condition for real Fourier densities. Furthermore, it shows that only half of the coefficients need to be stored.

THEOREM 1 (REAL FOURIER SERIES) Consider a Fourier density $f_\Psi(\mathbf{x})$ of (4). For a coefficient $\gamma_{-\mathbf{k}}$ the condition

$$\gamma_{-\mathbf{k}} = \gamma_{\mathbf{k}}^*$$

holds, since $f_\Psi(\mathbf{x})$ is real, i.e., $f_\Psi(\mathbf{x}) \in \mathbb{R}$.

PROOF. A coefficient of a Fourier series can be determined with the Fourier integral [12]

$$\gamma_{\mathbf{k}} = \int_{\Omega^d} f_\Psi(\mathbf{x}) e^{j\mathbf{k}^T \mathbf{x}} d\mathbf{x}. \quad (6)$$

Conjugating (6) gives

$$\gamma_{\underline{\kappa}}^* = \int_{\Omega^d} f_{\Psi}^*(\underline{x}) e^{-j\underline{\kappa}^T \underline{x}} d\underline{x} = \int_{\Omega^d} f_{\Psi}(\underline{x}) e^{j(-\underline{\kappa})^T \underline{x}} d\underline{x} = \gamma_{-\underline{\kappa}}$$

since $f_{\Psi}^*(\underline{x}) = f_{\Psi}(\underline{x})$. \square

The following theorems derive the most important characteristics of probability density functions.

THEOREM 2 (CUMULATIVE DISTRIBUTION) *The cumulative distribution of a given d -dimensional Fourier density $f_{\Psi}(\underline{x})$ is*

$$F_{\Psi}(\underline{x}) = \int_{-\underline{\pi}}^{\underline{x}} f_{\Psi}(\underline{\xi}) d\underline{\xi} = \sum_{\underline{\kappa} \in \mathcal{K}} \gamma_{\underline{\kappa}} \prod_{l=1}^d \chi_{\kappa_l} \quad \text{with (7)}$$

$$\chi_{\kappa_l} = \begin{cases} \frac{e^{j\kappa_l x_l} - (-1)^{\kappa_l}}{j\kappa_l} & \kappa_l \neq 0 \\ x_l + \pi & \kappa_l = 0 \end{cases}$$

and $\underline{\pi} = [-\pi \dots -\pi]^T$ being a d -dimensional vector of the lower border.

PROOF. Since $e^{j\underline{\kappa}^T \underline{x}} = e^{j\kappa_1 x_1} e^{j\kappa_2 x_2} \dots e^{j\kappa_d x_d}$, the d -fold integral can be distributed to single integrals over each $e^{(\cdot)}$ -term. Integration over each term results in (7). \square

THEOREM 3 (EXPECTED VALUE) *The expected value of the l -th component of a d -dimensional random vector \underline{x} with the Fourier density $f_{\Psi}(\underline{x})$ is given as*

$$\hat{x}_l = E\{x_l\} = \int_{\Omega^d} x_l f(\underline{x}) d\underline{x} = (2\pi)^d \sum_{\substack{\kappa_l = -\kappa_{l,o} \\ \kappa_l \neq 0}}^{\kappa_{l,o}} \frac{\gamma_{(\underline{\delta}_l \odot \underline{\kappa})} (-1)^{\kappa_l}}{j\kappa_l} \quad (8)$$

with $\underline{\delta}_l = [\delta_{l1} \delta_{l2} \dots \delta_{ld}]^T$ being a vector of Kronecker symbols

$$\delta_{kl} = \begin{cases} 0 & k \neq l \\ 1 & l = q \end{cases}.$$

Since \odot denotes elementwise multiplication, $\gamma_{(\underline{\delta}_l \odot \underline{\kappa})}$ is the coefficient with all indices being zero except κ_l . The expected value of the random vector \underline{x} is given by

$$E\{\underline{x}\} = [E\{x_1\} \ E\{x_2\} \ \dots \ E\{x_d\}]^T.$$

PROOF. Eq. (8) can be written as a product of one-dimensional integrals. Since

$$\int_{-\pi}^{\pi} e^{jq\xi} d\xi = 2\pi\delta_{q0}$$

is zero for integer $q \neq 0$, only the coefficients $\gamma_{(\underline{\delta}_l \odot \underline{\kappa})}$ influence the expected value. Using integration by parts on the term $\int \xi_l e^{j\kappa_l \xi_l} d\xi_l$ results in (8). \square

THEOREM 4 (VARIANCE) *The variance of the l -th component of a d -dimensional random vector \underline{x} with the Fourier density $f_{\Psi}(\underline{x})$ is given as*

$$C_{ll} = E\{(x_l - \hat{x}_l)^2\} = \frac{2}{3}\pi^2 \gamma_{\underline{0}} + 2(2\pi)^d \sum_{\substack{\kappa_l = -\kappa_{l,o} \\ \kappa_l \neq 0}}^{\kappa_{l,o}} \frac{\gamma_{(\underline{\delta}_l \odot \underline{\kappa})}}{\kappa_l^2} (-1)^{\kappa_l} - \hat{x}_l^2.$$

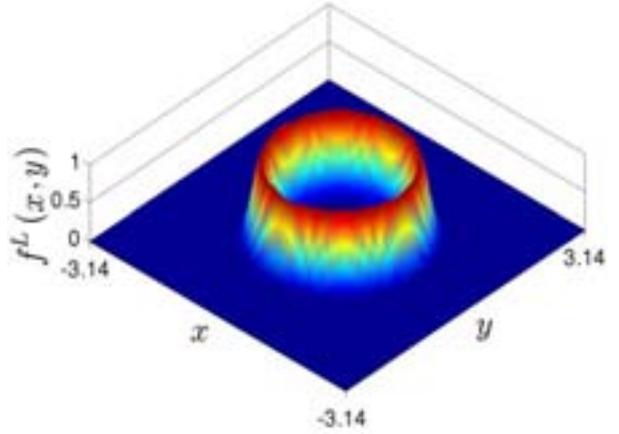


Fig. 2. The approximation of the 2D-Density $f(x, y) = \mathcal{N}(\sqrt{x^2 + y^2} - 2, 0.09)$ with a 63×63 -order Fourier density.

PROOF. Obviously,

$$C_{ll} = E\{(x_l - \hat{x}_l)^2\} = E\{x_l^2 - 2x_l \hat{x}_l + \hat{x}_l^2\} = E\{x_l^2\} - E\{2x_l \hat{x}_l\} + E\{\hat{x}_l^2\} = E\{x_l^2\} - \hat{x}_l^2.$$

Determining $E\{x_l^2\}$ can be performed analogously to $E\{x_l\}$. \square

THEOREM 5 (COVARIANCE) *The covariance of the l -th and the q -th component of a d -dimensional random vector \underline{x} with the Fourier density $f_{\Psi}(\underline{x})$ is given as*

$$C_{lq} = \text{Cov}\{x_l, x_q\} = E\{(x_l - \hat{x}_l)(x_q - \hat{x}_q)\} = (2\pi)^d \sum_{\substack{\kappa_l = -\kappa_{l,o}, \kappa_l \neq 0 \\ \kappa_q = -\kappa_{q,o}, \kappa_q \neq 0}}^{\kappa_{l,o}, \kappa_{q,o}} \frac{\gamma_{((\underline{\delta}_l + \underline{\delta}_q) \odot \underline{\kappa})} (-1)^{(\kappa_l + \kappa_q)}}{\kappa_l \kappa_q} - \hat{x}_l \hat{x}_q.$$

$\gamma_{((\underline{\delta}_l + \underline{\delta}_q) \odot \underline{\kappa})}$ is the coefficient with all indices being zero, except κ_q and κ_l . The covariance matrix

$$\text{Cov}\{\underline{x}\} = \begin{bmatrix} C_{11} & \dots & C_{1d} \\ \vdots & \ddots & \vdots \\ C_{d1} & \dots & C_{dd} \end{bmatrix}$$

can be constructed from the covariances.

PROOF. Obviously,

$$E\{(x_l - \hat{x}_l)(x_q - \hat{x}_q)\} = E\{x_l x_q\} - E\{x_l\} \hat{x}_q - \hat{x}_l E\{x_q\} + \hat{x}_l \hat{x}_q = E\{x_l x_l\} - \hat{x}_l \hat{x}_q.$$

Determining $E\{x_l x_q\}$ can be performed in analogy to $E\{x_l\}$. \square

IV. APPROXIMATING ARBITRARY DENSITIES

Determining the Fourier series from an arbitrary function can be accomplished via the Fourier integral [12]. Generally, exact Fourier representations of a given probability density function are not possible, because it would yield a Fourier series with an infinite number of coefficients. Thus, a truncated Fourier series has to be used. Unfortunately, truncating a Fourier series does not preserve its non-negativity. To

ensure valid Fourier density functions, the approximation is performed indirectly. A function $\Psi(\underline{x})$, a so-called Ψ -density, is determined, of which the absolute square $|\Psi(\underline{x})|^2 = f_\Psi(\underline{x})$ corresponds to the Fourier density.

DEFINITION 2 (Ψ -DENSITY) *If a d -dimensional Fourier expansion*

$$\Psi : \underline{x} = [x_1, x_2 \dots x_d]^T \in \Omega^d := [-\pi, \pi]^d \mapsto \mathbb{C}$$

of the form

$$\Psi(\underline{x}) = \sum_{\underline{\kappa} \in \mathcal{K}} \gamma_{\underline{\kappa}} e^{j\underline{\kappa}^T \underline{x}} = \sum_{\underline{\kappa} \in \mathcal{K}} (\alpha_{\underline{\kappa}} + j\beta_{\underline{\kappa}}) e^{j\underline{\kappa}^T \underline{x}},$$

analogously to Eq. (4) fulfills the condition

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

then $\Psi(\underline{x})$ is called a Ψ -density of the Fourier density $f_\Psi(\underline{x})$.

With the Ψ -densities, we are able to approximate arbitrary density functions.

THEOREM 6 (APPROXIMATION) *Consider a d -dimensional probability density function $f : \underline{x} \in \Omega^d = [-\pi, \pi]^d \mapsto \mathbb{R}_0^+$, the optimal coefficients of the Ψ -density with respect to the Hellinger metric*

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

are given by the Fourier integral

$$c_{\underline{\kappa}} = a_{\underline{\kappa}} + jb_{\underline{\kappa}} = \frac{1}{(2\pi)^d} \int_{\Omega^d} \sqrt{f(\underline{x})} e^{-j\underline{\kappa}^T \underline{x}} d\underline{x}. \quad (9)$$

The Fourier density $f_\Psi(\underline{x}) = \Psi(\underline{x})\Psi^*(\underline{x})$ can be obtained by taking the absolute square of $\Psi(\underline{x})$.

PROOF. For determining a minimum, the first derivative¹ $G_{-\underline{\kappa}} = \partial G / \partial a_{-\underline{\kappa}}$ has to be zero, which results in

$$\begin{aligned} G_{-\underline{\kappa}} &= \int_{\Omega^d} \frac{\partial}{\partial a_{-\underline{\kappa}}} \left(\sqrt{f(\underline{x})} - \Psi(\underline{x}) \right)^2 d\underline{x} \\ &= -2 \int_{\Omega^d} \left(\sqrt{f(\underline{x})} - \Psi(\underline{x}) \right) e^{-j\underline{\kappa}^T \underline{x}} d\underline{x}. \end{aligned}$$

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

$$G_{-\underline{\kappa}} = -2 \int_{\Omega^d} \sqrt{f(\underline{x})} e^{-j\underline{\kappa}^T \underline{x}} d\underline{x} + 2 \left((2\pi)^d c_{\underline{\kappa}} \right).$$

With setting $G_{-\underline{\kappa}} = 0$, we obtain (9). Since it can easily be shown that $\partial^2 G_{-\underline{\kappa}} / \partial a_{-\underline{\kappa}}^2$ is greater than zero, (9) is the condition for a minimum. The same result is obtained, when using $b_{-\underline{\kappa}}$ instead of $a_{-\underline{\kappa}}$. \square

Note that the fast Fourier transform [13] can be employed for efficiently calculating all coefficients at once. It has an algorithmic complexity of $O(n \log n)$, with n being the number of sampled data points.

Determining the Ψ -density from a given Fourier density is also possible. This is important for reducing the order of a given density when performing multiple filtering steps.

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

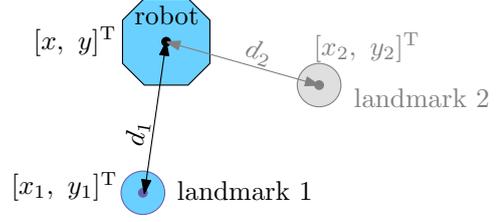


Fig. 3. Robot localization example. The distance d_i is measured to different landmarks with known position.

LEMMA 1 (CALCULATING $\sqrt{f_\Psi(\underline{x})}$) *A Ψ -density $\Psi(\underline{x}) = \sqrt{f_\Psi(\underline{x})}$ can be determined by solving the equation system*

$$\Psi(\underline{x})\Psi^*(\underline{x}) = f_\Psi(\underline{x}).$$

It can be solved recursively, since the equation

$$c_{\underline{\kappa}_0} c_{-\underline{\kappa}_0}^* = \gamma_{2\underline{\kappa}_0}$$

has one unknown variable. Decreasing any component of the index vector $\underline{\kappa}$ by one adds one unknown variable. Note that when solving the latter equation system recursively, it is very sensitive to numeric deviations in $c_{\underline{\kappa}_0}$.

We conclude this section with an approximation example.

EXAMPLE 1 (2D APPROXIMATION) *Consider the probability density*

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

with the variance $\sigma^2 = 0.09$ and the Gaussian

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

$f(x, y)$ resembles a Gaussian pulled along a circle with center $(0, 0)$ and a radius $\sqrt{2}$. Fig. 2 shows an approximation of the order 63×63 . The Gaussian ring can clearly be seen.

V. NONLINEAR FILTERING WITH FOURIER DENSITIES

As discussed in Sec. II the Bayesian estimator consists of a filtering step and a prediction step.

A. The Bayesian Prediction Step

The prediction step (3) is a generalized convolution integral, which simplifies to a summation over a product of two coefficients when Fourier densities.

LEMMA 2 (PREDICTION STEP) *Consider a Fourier density*

$$f_\Psi^e(\underline{x}_k) = \sum_{\underline{\mu} \in \mathcal{K}_k} \gamma_{\underline{\mu}}^e e^{j\underline{\mu}^T \underline{x}_k}$$

and a transition density

$$f_\Psi^T(\underline{x}_{k+1}, \underline{x}_k) = \sum_{\substack{\underline{\kappa} \in \mathcal{K}_{k+1} \\ \underline{\mu} \in \mathcal{K}_k}} \gamma_{\underline{\kappa}, \underline{\mu}}^e e^{j\underline{\kappa}^T \underline{x}_{k+1} + j\underline{\mu}^T \underline{x}_k}.$$

The generalized convolution integral is given by

$$f_\Psi^p(\underline{x}_{k+1}) = \sum_{\underline{\kappa} \in \mathcal{K}_{k+1}} \gamma_{\underline{\kappa}}^p e^{j\underline{\kappa}^T \underline{x}_{k+1}} = \int_{\Omega} f_\Psi^T(\underline{x}_{k+1}, \underline{x}_k) f_\Psi^e(\underline{x}_k) d\underline{x}_k$$

$$\text{with } \gamma_{\underline{\kappa}}^p = \sum_{\underline{\mu} \in \mathcal{K}_{k+1}} \gamma_{-\underline{\mu}}^e \gamma_{\underline{\kappa}, \underline{\mu}}^T.$$

An interesting property of the prediction step with Fourier densities is that the order of the resulting density $f_\Psi(\underline{x}_{k+1})$ depends only on the order of the transition density $f_\Psi^T(\underline{x}_{k+1}, \underline{x}_k)$ but not on the order of prior density $f_\Psi^e(\underline{x}_k)$. Hence, the prediction step automatically limits the algorithmic complexity of the Bayesian estimator.

B. The Bayesian Filtering Step

For the case of having low-dimensional measurements \underline{y}_k , determining the conditional density $f(\underline{y}_k | \underline{x}_k)$ off-line can be more efficient. For a given measurement $\hat{\underline{y}}_k$, the likelihood can then be calculated on-line by evaluating $f^L(\underline{x}_k) = f(\underline{y}_k = \hat{\underline{y}}_k | \underline{x}_k)$. That can be interpreted as slicing the conditional density along $\hat{\underline{y}}_k$.

LEMMA 3 (CALCULATING A SLICE) *Consider a d -dimensional Fourier density $f_\Psi(\underline{x})$. Setting one component of \underline{x} to $x_l = x_l^0$ results in a Fourier density, which is reduced by one dimension,*

$$f_\Psi(\underline{z}) = \sum_{\underline{\mu} \in \mathcal{K}_z} \gamma_\mu^0 e^{j\underline{\mu}^T \underline{z}} \text{ with } \gamma_\mu^0 = \sum_{\kappa_l = -\kappa_{l,0}}^{\kappa_{l,0}} \gamma_\kappa e^{j\kappa x_l^0}, \quad (10)$$

where the vector \underline{z} consists of the components

$$z_j = \begin{cases} x_j & j < l \\ x_{j-1} & j > l \end{cases}, \quad j = 1 \dots (d-1)$$

and the set $\mathcal{K}_z = \mathcal{K} \setminus \{-\kappa_l^0 \dots \kappa_l^0\}$ is a subset of \mathcal{K} excluding the indices of the l -th dimension. For slicing in higher dimensions, (10) can be applied repeatedly.

The main part of the filtering step is a multiplication of two densities. For Fourier densities this results in a convolution. LEMMA 4 (PRODUCT) *Consider two Fourier densities $f^a : \underline{x} \in \mathbb{R}^d \mapsto \mathbb{R}_0^+$ and $f^b : \underline{x} \in \mathbb{R}^d \mapsto \mathbb{R}_0^+$. The product is given as*

$$f^c(\underline{x}) = f^a(\underline{x})f^b(\underline{x}) = \sum_{\underline{\kappa} \in \mathcal{K}_c} \gamma_\kappa^c e^{j\kappa^T \underline{x}} \quad \text{with}$$

$$\gamma_\kappa^c = \sum_{\underline{\mu} \in \mathcal{K}_c} \bar{\gamma}_\mu^a \bar{\gamma}_{(\underline{\mu}-\underline{\kappa})}^b,$$

where the bar denotes a valid index

$$\bar{\gamma}_\mu^{(\cdot)} = \begin{cases} \gamma_\mu^{(\cdot)} & \underline{\mu} \in \mathcal{K}_{(\cdot)} \\ 0 & \text{otherwise} \end{cases}$$

of γ^a or γ^b . The order of f_Ψ^c is $\prod_{l=1}^d (\kappa_l^{o,a} + \kappa_l^{o,b})$, i.e., the number of coefficients is significantly higher than of f_Ψ^a and of f_Ψ^b . When performing recursive filtering steps, the number of coefficients increase exponentially. To limit the complexity, Lemma 1 can be used to calculate a corresponding Ψ -density and then eliminate coefficients. A simple heuristic would be the removal of the highest order coefficients, which does not guarantee satisfactory results. A better approach is given by the following Theorem. It allows the removal of the coefficients with minimal influence on f_Ψ .

THEOREM 7 (ORDERING FOURIER TERMS) *The coefficient c_{\min} of a Ψ -density $\Psi(\underline{x})$, which minimally influences the Hellinger metric, has to satisfy*

$$c_{\min} c_{\min}^* \leq c_\kappa c_\kappa^* \text{ for } \underline{\kappa} \in \mathcal{K}, \quad (11)$$

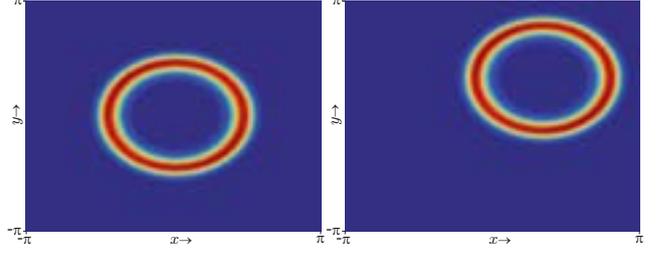


Fig. 4. Likelihoods for landmark 1 at position (0, 0) (left) and landmark 2 at position (1, 1) (right). The distance is both times $d_1 = d_2 = \sqrt{2}$.

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

PROOF. By applying a Ψ -density reduced by one element

$$\Psi_{\underline{\kappa}}(\underline{x}) := \Psi(\underline{x}) - c_\kappa e^{j\kappa^T \underline{x}}$$

to the Hellinger metric, we obtain

$$G(\Psi\Psi^*, \Psi_{\underline{\kappa}}\Psi_{\underline{\kappa}}^*) = c_\kappa c_\kappa^*,$$

Choosing the c_κ , which changes G minimally leads to (11). \square

NOTE: When exclusively performing filtering steps, the product can also be calculated directly with a convolution of the Ψ -densities instead of the convolution with the Fourier densities.

Normalizing a Fourier density is quite simple, as the following Theorem shows.

THEOREM 8 *The normalization constant of a Fourier density is*

$$\int_{\Omega^d} f_\Psi(\underline{\xi}) d\underline{\xi} = \frac{1}{(2\pi)^d \gamma_0}.$$

PROOF. This result can be easily derived using Eq. (7). \square

Dividing the Fourier density by the normalization constant concludes the filtering step.

VI. EXAMPLE: FILTERING STEP

In this section a simple localization problem is discussed. A position of a robot is determined using distance measurements as depicted in Fig. 3. The measurement equation

$$\underline{y} = h\left(\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}, \underline{v}\right) = \sqrt{(\underline{x} - x_l)^2 + (\underline{y} - y_l)^2} + \underline{v}$$

with (x_l, y_l) being the known landmark position and \underline{v} is additive zero-mean Gaussian noise with a variance of 0.09. In this example there is no prior knowledge, i.e., $f^0(x, y) = \frac{1}{4\pi^2}$ is uniform distribution. Two measurements to landmark 1 with the position (0, 0) and landmark 2 with the position (1, 1) are performed. Both measurements are given by $\hat{y}_1 = \hat{y}_2 = \sqrt{2}$. Fig. 2 depicts the Fourier densities of the likelihood of \hat{y}_1 and Fig. 4 show the likelihoods of \hat{y}_1 and \hat{y}_2 . These functions have been sampled with 125×125 values and then been transformed with a fast Fourier transform. To avoid aliasing, the likelihoods have $63^2 = 4225$ components. It can be seen that approximation is quite precise. Only some small ripples can be seen.

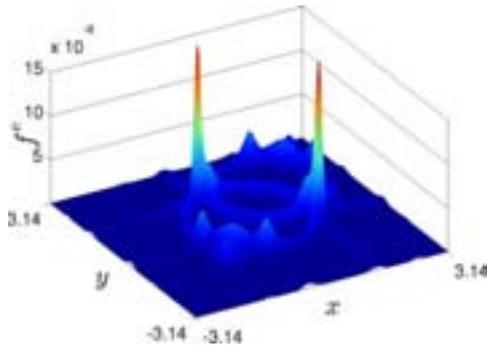


Fig. 5. Unnormalized position estimate after filtering. The peaks denote the intersection of the two rings.

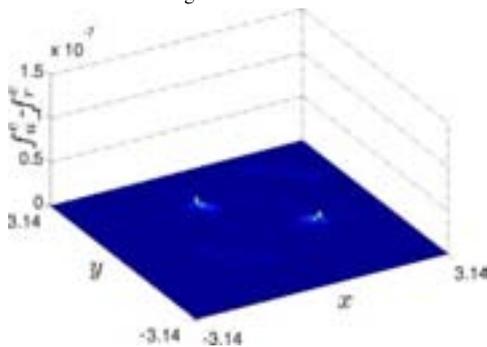


Fig. 6. Error of the reduction of the position estimate to 1000 components.

Performing the filtering step with a uniform distribution results in the likelihood. The result of the second filtering step is depicted in Fig. 5. It has $125^2 = 15625$ components. The peaks are at the intersection points of the two circles. The bumps show that the approximation of the likelihoods is not exact. Using a higher sampling rate for determining the coefficients would reduce those bumps considerably.

That a reduction can be performed without significantly changing the shape of the density is demonstrated in Fig. 6. There the error is depicted of the density of Fig. 5 and its reduction from 15625 to 1000 components. It can be seen that only a small amount of “density mass” was taken away.

VII. SUMMARY AND FUTURE WORK

In this work, Bayesian estimation with Fourier densities is generalized to the multidimensional case. In Sec. II the d -dimensional Bayesian estimator is reviewed. In Sec. III d -dim. Fourier densities as probability densities are introduced. The most important density properties, particularly cumulative distribution, expected value, and covariance are derived. Interestingly, the moments of the l -th components depend solely on the coefficients of that dimensions. That opens perspectives for developing a decentralized estimator.

Sec. IV discusses the procedure of approximating arbitrary probability densities with Fourier densities. With a fast Fourier transform, the coefficients can be calculated very efficiently. Additionally, it is 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 derives a recursive Bayesian estimator consisting of a filtering step and a prediction step. It is addressed that the prediction step bounds the complexity of the estimator. Furthermore, it is shown that the complexity, i.e., the number of coefficients, 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.

Sec. VI discusses a simple localization example. It is shown that reducing the number of coefficients is reasonable, because it changes the shape of the density insignificantly.

A main aspect of future work is the implementation of an efficient d -dimensional C++ library. A first version was used in example of Sec. VI. This also includes the investigation of the required number of coefficients of a Fourier density to perform efficient Bayesian estimation. Furthermore, the use of the fast Fourier transform to reduce the computational complexity of the filtering step is planned.

Some further aspects are the introduction of slideable and scalable domains as well as investigating orthogonal function series for coping with densities that need a large number of coefficients in the Fourier space.

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,” *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 48, no. 6, pp. 1816–1820, June 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] A. Pinheiro and B. Vidakovic, “Estimating the Square Root of a Density via Compactly supported Wavelets,” *Computational Statistics & Data Analysis*, vol. 25, no. 4, pp. 399–415, 1997.
- [11] D. Brunn, F. Sawo, and U. D. Hanebeck, “Efficient Nonlinear Bayesian Estimation based on Fourier Densities,” *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 312–322, September 2006.
- [12] E. W. Weisstein, “MathWorld – A Wolfram Web Resource,” 2006. [Online]. Available: <http://mathworld.wolfram.com/>
- [13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992. [Online]. Available: <http://www.numerical-recipes.com/>