2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

# Bayesian Estimation for Linear Systems with Nonlinear Output and General Noise Distribution using Fourier Basis Functions

R.A.C. van Zuijlen, J.S. van Hulst, W.P.M.H. Heemels, D.J. Antunes

*Abstract*— **Bayesian estimation can be used to estimate the state of dynamical systems, but its applicability is hampered due to the curse of dimensionality. This paper aims to mitigate this bottleneck for a relevant class of systems consisting of a linear plant with bounded input, driven by stochastic disturbances with non-linear noisy output; the distributions of the disturbances and noise have a bounded support but are otherwise general. Using a frequency-domain interpretation of the operations of the Bayes' filter, we show that, under mild assumptions, exact Bayesian estimation can be pursued in a countable space of Fourier series coefficients, rather than in the usual functional space of probability densities. This fact leads to a natural approximate method, where the Fourier series coefficients corresponding to high frequencies are discarded. For this approximate method, the complexity of the conditioned state distribution, measured by the number of Fourier coefficients, remains constant at prediction steps and grows only linearly at each update step. The applicability of the results is illustrated in the context of electron microscopy, where a residual error analysis indicates that the approximation is accurate.**

## I. INTRODUCTION

Bayes' filtering is a general approach to the state estimation problem of dynamical systems [1]. It provides the state probability density function (pdf) given previous control inputs and measurements based on two recursive steps: update and prediction. The update step relies on the most recent measurement and the prediction step on the dynamical model.

When the dynamical and output equations are linear and the noise statistics are Gaussian, the Bayes' filter boils down to the Kalman filter. In such a case, the conditioned state distribution is Gaussian at every time step, provided that this is the case at time zero [2]. More generally, the conditioned state distribution is a sum of Gaussians, when the distributions of the initial state, the noise, and the disturbances are sums of Gaussians, although if the noise statistics cannot be represented by a single Gaussian, the number of Gaussians grows exponentially with time [3]. The Kalman filter is one of the rare cases in the context of multivariable systems with general state dimension $n_x$ for which Bayesian estimation is tractable. In fact, without the linear and Gaussian assumptions, one can typically only rely on discretization of the pdf; however, when $n_x \geq 3$ is large, keeping track of a pdf in $\mathbb{R}^{n_x}$ easily becomes computationally intractable.

There are many *approximate* methods, such as the Extended Kalman Filter [4], Unscented Kalman Filter [5], Particle Filter [6], Multiple Model Estimation Algorithm [7], Sequential Monte Carlo Algorithm [8], Variational Bayes' filtering [9], or methods that use Fourier densities to approximate pdfs and perform Bayes' filtering [10]. Although the results are certainly of interest, in general it is hard to quantify the quality of the approximation for these methods.

In this paper we are interested in a class of systems for which we show that *exact* Bayesian estimation can be pursued and accurately approximated in a tractable way. This class is characterized by:

(i) A linear system with a bounded set of admissible initial states.
(ii) General (non-linear) output equations with bounded independent and identically distributed measurement noise with a general distribution.
(iii) Bounded, independent, and identically distributed process noise with a general distribution.
(iv) Bounded control inputs.

Conceivably, ignoring the assumptions regarding the bounds on the sets constraining the inputs, noise and disturbances for a moment, this is a much broader class of systems when compared to the underlying assumptions in the context of the Kalman filter (or the sums of Gaussian approach).

For the class of systems satisfying (i)-(iv) above, we will interpret the operations of the Bayes' filter in the frequency domain through the Fourier transform of the conditioned pdf. In general, keeping track of such a Fourier transform would be intractable. Since the assumptions render bounded sets, it is possible to represent it in a lossless manner, with samples corresponding to the Fourier series. This means that exact Bayesian estimation can be pursued in the countable space of Fourier series coefficients uniquely representing the conditioned state pdf, rather than in the usual functional space of pdfs. This leads to a natural approximate method, where the Fourier series coefficients corresponding to high frequencies are discarded. For this approximate method, the complexity of the conditioned state distribution, measured by the number of Fourier coefficients, remains constant at prediction steps and grows only linearly at each update step, as we will see.

The research is carried out as part of the ITEA4 20216 ASIMOV project. The ASIMOV activities are supported by the Netherlands Organisation for Applied Scientific Research TNO and the Dutch Ministry of Economic Affairs and Climate (project number: AI211006). The research leading to these results is partially funded by the German Federal Ministry of Education and Research (BMBF) within the project ASIMOV-D under grant agreement No. 01IS21022G [DLR], based on a decision of the German Bundestag.

The authors are with the Control Systems Technology Section, Department of Mechanical Engineering, Eindhoven University of Technology, The Netherlands, E-mails: {r.a.c.zuijlen, j.s.v.hulst, m.heemels, d.antunes}@tue.nl

The use of Fourier series in a Bayesian estimation context can lead to large data efficiency gains (see the numerical examples in Section IV below) as well-known in other domains (compression into JPEG-images and MP3-files). Due to this, it can provide a viable solution for medium-scale problems $4 \leq n_x \leq 6$ where discretization would not be feasible. The use of Fourier series in this context has also been proposed in [10]. Contrary to [10], our proposed method uses Fourier series rather than their approximated Fourier densities, and does not require a transition density function but uses the linear system dynamics directly to compute the pdf of the state in the prediction step.

The proposed model and estimation method were inspired by a problem in the electron microscopy field. Electron microscopes require tuning by an expert operator to compensate for present aberrations [11]. The estimation of the aberration relies on the visual inspection of a so-called ronchigram [12], [13]. Based on one single ronchigram, it is impossible to uniquely identify the aberrations [14]. To address this problem in a systematic manner, it turns out to be useful to consider the electron microscope as an integrator (and thus linear) system, where the (unknown) states describe the aberrations and the control inputs are the tuning knobs to compensate for those aberrations. The output is a complex non-linear function based on the states. As this setting fits the assumptions of our framework, we apply the proposed method in this context and discuss its effectiveness through a residual error analysis. This analysis indicates that the approximation is accurate.

This paper is structured as follows. The proposed model is introduced in Section II, where the Bayes' filter in the time-domain is reviewed. The exact Bayes' filter written with Fourier transforms and Fourier series, and the approximate Bayes' filter using only a finite number of Fourier coefficients are given in Section III. Two numerical examples of the approximate method that demonstrates the performance of the Bayes' filter with Fourier basis functions are given in Section IV, including a comparison with the Unscented Kalman Filter. Finally, in Section V, conclusions are drawn and directions for future research are given.

## II. PROPOSED MODEL AND BAYES' FILTER

Consider the system

$$x_{t+1} = f(x_t, u_t, w_t) \quad (1)$$

$$y_t = g(x_t, v_t) \quad (2)$$

with linear dynamics, i.e.,

$$f(x, u, w) = Ax + Bu + w \quad (3)$$

and where $x_t \in \mathbb{R}^{n_x}$ is the state, $u_t \in \mathcal{U} \subseteq \mathbb{R}^{n_u}$ is the control input, $y_t \in \mathbb{R}^{n_y}$ the measured output at time $t \in \mathcal{T} := \{0, 1, \ldots, h\}$, where $h \in \mathbb{N} \cup \{\infty\}$ is a time horizon of interest. The initial state is assumed to belong to a state $x_0 \in \mathcal{X}_0 \subseteq \mathbb{R}^{n_x}$. Moreover, $w_t \in \mathcal{W} \subseteq \mathbb{R}^{n_x}$ and $v_t \in \mathcal{V} \subseteq \mathbb{R}^{n_y}$, $t \in \mathcal{T}$, form sequences of zero-mean, independent and identically distributed, and mutually independent, random variables, representing the process disturbances and the measurement noise, respectively. Disturbances and measurement noises $w_t$ and $v_t$, $t \in \mathcal{T}$, can be described by any probability density function (pdf), and are therefore not restricted to zero-mean Gaussian noise. We denote these pdfs by $p_w(w)$ and $p_v(v)$, respectively. Hence, for all $t \in \mathcal{T}$, $\mathcal{A} \subseteq \mathcal{W}$, $\mathcal{B} \subseteq \mathcal{V}$, it holds that

$$\int_{\mathcal{A}} p_w(w)dw = \text{Prob}[w_t \in \mathcal{A}], \quad \int_{\mathcal{B}} p_v(v)dv = \text{Prob}[v_t \in \mathcal{B}]. \quad (4)$$

The function $g(\cdot, \cdot)$ may be a general nonlinear function, and we do not put any particular restrictions on it. We can equivalently characterize the output equation (2) by a pdf $p_y(y; x)$ such that, for all $t \in \mathcal{T}$,

$$\int_{\mathcal{C}} p_y(y; x)dy = \text{Prob}[y_t \in \mathcal{C} | x_t = x] = \int_{\mathcal{B}_{\mathcal{C}}} p_v(v)dv, \quad (5)$$

where $\mathcal{B}_{\mathcal{C}} := \{v \in \mathcal{V} | g(x, v) \in \mathcal{C}\}$. Similarly, we can equivalently characterize the process equation (1) by a pdf $q(\bar{x}; x, u)$, such that, for all $t \in \mathbb{N}_0$,

$$\int_{\mathcal{D}} q(\bar{x}; x, u)d\bar{x} = \text{Prob}[x_{t+1} \in \mathcal{D} | x_t = x, u_t = u] = \int_{\mathcal{E}_{\mathcal{D}}} p_w(\bar{x})dw \quad (6)$$

with $\mathcal{E}_{\mathcal{D}} := \{w \in \mathcal{W} | f(x, u, w) \in \mathcal{D}\}$. The initial state $x_0$ is assumed to be distributed according to a pdf $\bar{p}_0(x_0)$, with support $\mathcal{X}_0$. A standing assumption is the following.

**Assumption 1.** The sets $\mathcal{W}$, $\mathcal{V}$, $\mathcal{U}$ and $\mathcal{X}_0$ are bounded.

A direct consequence of this assumption is that, for all $t \in \mathcal{T}$,

$$x_t \in \mathcal{X}_t, \text{ for some bounded sets } \mathcal{X}_t.$$

The sets $\mathcal{X}_t$ are assumed to be minimum volume sets such that $x_t \in \mathcal{X}_t$. We define the set of rectangular sets in $\mathbb{R}^{n_x}$ as

$$\mathsf{R} = \{[\underline{L}_1, \bar{L}_1] \times \cdots \times [\underline{L}_{n_x}, \bar{L}_{n_x}] | \bar{L}_i > \underline{L}_i, i \in \{1, ..., n_x\}\}.$$

Here $\underline{L}_i$ and $\bar{L}_i$ are the lower- and upper-bounds, respectively, for the $i$-th entry of $\mathcal{X}_t$. The volume of $[\underline{L}_1, \bar{L}_1] \times \cdots \times [\underline{L}_{n_x}, \bar{L}_{n_x}] \in \mathsf{R}$ is $\tilde{L} = \Pi_{i=1}^{n_x} L_i$, $L_i = \bar{L}_i - \underline{L}_i$. Let $\mathcal{R}_t$ be the set in $\mathsf{R}$ with minimal volume such that $\mathcal{X}_t \subseteq \mathcal{R}_t$. Let

$$\mathcal{R} := \cup_{t=1}^{h} \mathcal{R}_t.$$

**Assumption 2.** When $h = \infty$, $\mathcal{R}$ is bounded.

A sufficient condition for this assumption to hold is that the system matrix $A$ is Schur. If this holds, the system is input-to-state stable, which implies that when $\mathcal{X}_0$, $\mathcal{U}$ and $\mathcal{W}$ are bounded, $\mathcal{X}_t$ is also bounded.

**Assumption 3.** $A$ is invertible.

Assumption 3 is met, for instance, when the dynamical model results from an exact time-discretisation of a continuous-time dynamical model.

The Bayes' filter allows us to keep track of conditioned pdfs $p_{i|j}(\cdot)$, where $i \in \mathcal{T}$ is the timestep to be estimated and $j \leq i$ the input and output information available at timestep $i$, that is, $\{u_0, ..., u_{j-1}, y_0, ..., y_j\}$. The Bayes' filter

recursively performs the following two steps for every $t \in \mathcal{T}$ with $p_{0|-1}(x) = \bar{p}_0(x)$, and given $u_t$, for every $x \in \mathcal{X}_0$.

(i) Update step:

$$p_{t|t}(x_t) = \frac{1}{\alpha} p_y(y_t; x_t) p_{t|t-1}(x_t), \quad x_t \in \mathcal{X}_t, \tag{7}$$

with $\alpha = \int_{\mathcal{X}_t} p_y(y_t; s) p_{t|t-1}(s) ds$

(ii) Prediction step:

$$p_{t+1|t}(x_{t+1}) = \int_{\mathcal{X}_t} q(x_{t+1}; x_t, u_t) p_{t|t}(x_t) dx_t, \quad x_{t+1} \in \mathcal{X}_{t+1}. \tag{8}$$

For the linear model considered, we can write the prediction step as follows. Consider two scalar functions $a : \mathcal{A} \to \mathbb{R}$, $b : \mathcal{B} \to \mathbb{R}$ with domains in sets $\mathcal{A} \subseteq \mathbb{R}^n$, $\mathcal{B} \subseteq \mathbb{R}^n$. If $\mathcal{A}$ and $\mathcal{B}$ are bounded sets, extend these functions to $\mathbb{R}^n$ by setting $a(x) = 0$, $x \in \mathbb{R}^n \backslash \mathcal{A}$, $b(x) = 0$, $x \in \mathbb{R}^n \backslash \mathcal{B}$, respectively. Then, convolution $c = a \otimes b$ is denoted by $c(x) = (a \otimes b)(x) = \int_{\mathbb{R}^n} a(x - s) b(s) ds$ for $x \in \mathbb{R}^n$.

**Proposition 1.** The prediction step can equivalently be written as

$$p_{t+1|t}(x) = \frac{1}{\det(A)} p_{t|t}(A^{-1}(x - Bu_t)) \otimes p_w(w), x \in \mathcal{X}_{t+1} \tag{9}$$

for every $t \in \mathcal{T}$. □

While $p_{t+1|t}(x)$ and $p_{t|t}(x)$ have only been defined by the Bayes' filter in the intervals $\mathcal{X}_{t+1}$ and $\mathcal{X}_t$, it is convenient to define them for every $x \in \mathbb{R}^{n_x}$ by setting $p_{t+1|t}(x) = 0$ if $x \in \mathbb{R}^{n_x} \setminus \mathcal{X}_{t+1}$, $p_{t|t}(x) = 0$ if $x \in \mathbb{R}^{n_x} \setminus \mathcal{X}_t$. In turn $p_y(y; x)$ and $p_w(w)$ only needs to be defined for $x \in \mathcal{X}$ and $w \in \mathcal{W}$, but we also define them for every $x \in \mathbb{R}^{n_x}$ by setting $p_y(y; x) = 0$, if $x \in \mathbb{R}^{n_x} \setminus \mathcal{X}$, $p_w(w) = 0$ if $w \in \mathbb{R}^{n_w} \setminus \mathcal{W}$.

Performing the Bayes' filter operations is in general computationally unfeasible when $n_x$ is large (typically when $n_x \leq 3$ one can still rely on spatial discretization, but for larger $n_x$ this becomes unfeasible). This motivates the problem considered here which is to represent $p_{t+1|t}(x)$ and $p_{t|t}(x)$ in an efficient way and to provide a numerically efficient method to perform these operations.

## III. BAYES' FILTER WITH FOURIER TRANSFORM AND FOURIER BASIS FUNCTIONS

In this section, the Bayes' filter is pursued using the Fourier transform and Fourier basis functions. Section III-A provides an *exact* method, which requires an infinite number of coefficients. Section III-B provides an *approximate* method with a finite number of coefficients.

### A. Exact method

For every $t \in \mathcal{T}$, $\omega \in \mathbb{R}^{n_x}$, let

$$\begin{aligned} P_{t|t}(\omega) &:= \int_{\mathbb{R}^{n_x}} p_{t|t}(x) e^{-j\omega \cdot x} dx, \\ P_{t|t-1}(\omega) &:= \int_{\mathbb{R}^{n_x}} p_{t|t-1}(x) e^{-j\omega \cdot x} dx, \\ P_w(\omega) &:= \int_{\mathbb{R}^{n_w}} p_w(w) e^{-j\omega \cdot w} dw, \\ P_y(y; \omega) &:= \int_{\mathbb{R}^{n_y}} p_y(y; x) e^{-j\omega \cdot y} dy \end{aligned} \tag{10}$$

be the Fourier transforms of the pdfs $p_{t|t}(x)$, $p_{t|t-1}(x)$, $p_w(w)$ and of $p_y(y; x)$, where $a \cdot b$ denote the inner production of two vectors $a \in \mathbb{R}^n$, $b \in \mathbb{R}^n$. Then, the operations of the Bayes' filter can be written as follows.

**Lemma 1.** The Fourier transforms (10) are related for every $t \in \mathcal{T}$ by

- Update step:

$$P_{t|t}(\omega) = \frac{1}{\alpha} P_y(y_t; \omega) \otimes P_{t|t-1}(\omega) \tag{11}$$

for all $\omega \in \mathbb{R}^{n_x}$ with $\alpha = P_{y|x}(y_t; \omega) \otimes P_{t|t-1}(\omega)_{|_{\omega=0}}$

- Prediction step:

$$P_{t+1|t}(\omega) = P_{t|t}(A^T \omega) e^{-j\omega \cdot (Bu_t)} P_w(\omega), \omega \in \mathbb{R}^{n_x} \tag{12}$$

□

Interestingly, while the update and predictions steps in the time domain correspond to a product and a convolution (besides the affine transformation), respectively, in terms of the Fourier transform the update and predictions steps in the frequency domain correspond to a convolution and a product (besides the affine transformation). Hence, running the Bayes' filter with Fourier transforms is equally intractable. However, since the state belongs to a bounded set we only need enough samples to reconstruct the Fourier transform, and we can compute the Fourier series instead.

Let $k = [k_1 \dots k_n]^\intercal$ with $k_i \in \mathbb{Z}$, so that $k \in \mathcal{K} := \{[k_1 \dots k_n]^\intercal | k_i \in \mathbb{Z}\}$. Moreover, let $k./L := [k_1/L_1 \dots k_n/L_n]^\intercal$. Suppose that we sample and scale the Fourier transforms $P_{t|t}(\omega)$, $P_{t|t-1}(\omega)$, $P_y(y; \omega)$ obtaining the Fourier coefficients for every $k \in \mathcal{K}$, for every $t \in \mathcal{T}$

$$\begin{aligned} c_{t|t}[k] &= \frac{1}{\tilde{L}} P_{t|t}(\omega)_{|_{\omega=(\frac{2\pi k_1}{L_1}, \dots, \frac{2\pi k_n}{L_n})}} \\ c_{t|t-1}[k] &= \frac{1}{\tilde{L}} P_{t|t-1}(\omega)_{|_{\omega=(\frac{2\pi k_1}{L_1}, \dots, \frac{2\pi k_n}{L_n})}} \\ c_y[y; k] &= \frac{1}{\tilde{L}} P_y(y; \omega)_{|_{\omega=(\frac{2\pi k_1}{L_1}, \dots, \frac{2\pi k_n}{L_n})}}. \end{aligned} \tag{13}$$

These Fourier coefficients are enough to write the conditioned pdfs in $x_t \in \mathcal{R}$, $x_{t+1} \in \mathcal{R}$,

$$\begin{aligned} p_{t|t}(x_t) &= \sum_{k \in \mathcal{K}} c_{t|t}[k] e^{j2\pi(k./L) \cdot x_t}, \quad x_t \in \mathcal{R}, \\ p_{t|t-1}(x_{t+1}) &= \sum_{k \in \mathcal{K}} c_{t|t-1}[k] e^{j2\pi(k./L) \cdot x_{t+1}}, \quad x_{t+1} \in \mathcal{R}, \\ p_y(y; x) &= \sum_{k \in \mathcal{K}} c_y[y; k] e^{j2\pi(k./L) \cdot x}, \quad x \in \mathcal{R}. \end{aligned} \tag{14}$$

Outside $\mathcal{R}$, the formulas on the right-hand side define periodic replications of these pdfs.

Due to the crucial fact that $x_t \in \mathcal{R}$ for every $t \in \mathcal{T}$, where $\mathcal{R}$ is a bounded set, which follows from Assumption 1, we can use the sampling theorem, to conclude that these Fourier coefficients fully characterize the Fourier transforms $P_{t|t}(\omega)$, $P_{t|t-1}(\omega)$. In fact, let

$$B_1(\omega, L, k) := \prod_{i=1}^{n_x} \frac{2\sin((\omega_i - 2\pi/L_i k_i)\frac{Li}{2})}{(\omega_i - 2\pi/L_i k_i)} e^{-j(\omega_i - 2\pi k_i/L_i)(\bar{L}_i + \underline{L}_i)/2}$$

where $\frac{\sin(x)}{x}$ is assumed to be extended by continuity so that $\frac{\sin(x)}{x}_{|x=0} = 1$. Then, for every $\omega \in \mathbb{R}^{n_x}$, $t \in \mathcal{T}$,

$$\begin{aligned} P_{t|t}(\omega) &= \sum_{k\in\mathcal{K}} c_{t|t}[k] B_1(\omega, L, k), \\ P_{t|t-1}(\omega) &= \sum_{k\in\mathcal{K}} c_{t|t-1}[k] B_1(\omega, L, k) \\ P_y(y;\omega) &= \sum_{k\in\mathcal{K}} c_y[y;k] B_1(\omega, L, k) \end{aligned} \tag{15}$$

To perform the Bayes' filter operations with the Fourier series coefficients requires the discrete convolution operation. The convolution of $c : \mathcal{K} \to \mathbb{R}$ and $d : \mathcal{K} \to \mathbb{R}$, is denoted by $g[k] = c[k] \otimes d[k] = \sum_{r\in\mathcal{K}} c[k-r]d[r], \quad k \in \mathcal{K}$.

**Lemma 2.** The Fourier coefficients (13) are related for every $t \in \mathcal{T}$ by

- Update step: for $k \in \mathcal{K}$

$$c_{t|t}[k] = \frac{1}{\alpha} c_y[y_t;k] \otimes c_{t|t-1}[k] \tag{16}$$

  with $\alpha = c_y[y_t;k] \otimes c_{t|t-1}[k]_{|k=0}$
- Prediction step: for $k \in \mathcal{K}$

$$c_{t+1|t}[k] = \frac{1}{\tilde{L}} \left( \sum_{k\in\mathcal{K}} c_{t|t}[k] B_1(A^T\omega, L, k) \right) e^{-j\omega\cdot(Bu_t)} P_w(\omega)_{|\omega=(\frac{2\pi k_1}{L_1},\ldots,\frac{2\pi k_n}{L_n})} \tag{17}$$

□

Note that when $A = I$, the prediction step (17) boils down to a much simpler expression

$$c_{t+1|t}[k] = c_{t|t}[k] e^{-j\omega\cdot(Bu_t)} P_w(\omega)_{|\omega=(\frac{2\pi k_1}{L_1},\ldots,\frac{2\pi k_n}{L_n})}, \tag{18}$$

where each $c_{t+1|t}[k]$ only depends on $c_{t|t}[\ell]$ if $k = \ell$.

### *B. Approximate method*

Suppose we approximate the initial conditioned state distribution with a finite number of Fourier coefficients

$$\hat{c}_{0|-1}[k] := \begin{cases} c_{0|-1}[k] \text{ if } k \in \mathcal{K}_{0|-1}, \\ 0 \text{ otherwise }, \end{cases} \tag{19}$$

where

$$\mathcal{K}_{0|-1} = \{-m_{0|-1}, -m_{0|-1}+1 \ldots, m_{0|-1}-1, m_{0|-1}\}$$

with $m_{0|-1} \in \mathbb{N}$, i.e., we discard the coefficients corresponding to $k \in \mathcal{K}^c_{0|-1} = \mathcal{K} \setminus \mathcal{K}_{0|-1}$. Then

$$\hat{p}_{0|-1}(x) = \begin{cases} \sum\limits_{k\in\mathcal{K}_{0|-1}} \hat{c}_{0|-1}[k] e^{j2\pi(k./L)\cdot x}, & \text{if } x \in \mathcal{R}, \\ 0, \quad \text{if } x \in \mathbb{R}^{n_x} \setminus \mathcal{R}. \end{cases} \tag{20}$$

Note however that $\hat{p}_{0|-1}(x)$ is in general not a pdf, since it is not necessarily positive and does not necessarily integrate to one. However, given any $\hat{p}(x)$ in the space $\mathcal{P}$ of bounded functions in the bounded set $\mathcal{R}$ we can defined a normalization operator $\mathsf{N} : \mathcal{P} \to \mathcal{P}$ that to each $\hat{p}(x) \in \mathcal{P}$ provides $\tilde{p}(x) = \frac{\hat{p}(x)+\epsilon}{\alpha(\epsilon)}$ with $\epsilon = \min\{c \geq 0 | \hat{p}(x) + c \geq 0 \text{ for all } x \in \mathcal{R}\}$ and $\alpha(\epsilon) = \int_{\mathcal{R}} (\hat{p}(x) + \epsilon) dx$. If $\hat{p}(x) = \sum_{k\in\mathcal{K}} c[k] e^{2\pi(k./L)\cdot x}$ for a set $\mathcal{K}$ with finite cardinality $n_K = |\mathcal{K}|$ this induces a normalization procedure to the $\hat{c}[k]$, defined by map $\mathsf{N}_c : \mathbb{R}^{n_K} \to \mathbb{R}^{n_K}$,

$$\tilde{c}[k] = \begin{cases} \dfrac{\hat{c}[k]}{\alpha(\epsilon)}, \text{ if } k \in \mathcal{K} \setminus \{0\}, \\ \dfrac{\hat{c}[0] + \epsilon}{\alpha(\epsilon)} \text{ if } k = 0. \end{cases} \tag{21}$$

Thus, $\tilde{p}_{0|-1}(x) = \mathsf{N}(\hat{p}_{0|-1}(x))$ is described by

$$\tilde{p}_{0|-1}(x) = \begin{cases} \sum\limits_{k\in\mathcal{K}_{0|-1}} \tilde{c}_{0|-1}[k] e^{j2\pi(k./L)\cdot x}, & \text{if } x \in \mathcal{R}, \\ 0, \quad \text{if } x \in \mathbb{R}^{n_x} \setminus \mathcal{R}, \end{cases} \tag{22}$$

where

$$\tilde{c}_{0|-1}[k] = \mathsf{N}_c(\hat{c}_{0|-1}[k]), \tag{23}$$

which is a pdf.

Likewise suppose that we approximate $p_y(y;x)$ with a finite number of Fourier coefficients

$$\hat{c}_y[y;k] := \begin{cases} c_y[y;k] \text{ if } k \in \mathcal{K}_y, \\ 0 \text{ otherwise }, \end{cases} \tag{24}$$

where $\mathcal{K}_y = \{-m_y, -m_y+1 \ldots, m_y-1, m_y\}$ with $m_y \in \mathbb{N}$. Then, letting $\tilde{c}_y[y;k] = \mathsf{N}_c(\hat{c}_y[y;k])$, $k \in \mathcal{K}_y$, results in a pdf

$$\tilde{p}_y(y;x) = \begin{cases} \sum\limits_{k\in\mathcal{K}_y} \tilde{c}_y[y;k] e^{j2\pi(k./L)\cdot x}, & \text{if } x \in \mathcal{R}, \\ 0, \quad \text{if } x \in \mathbb{R}^{n_x} \setminus \mathcal{R}. \end{cases} \tag{25}$$

This leads naturally to the following approximate Bayes' algorithm, which provides estimates $\tilde{p}_{t|t}(x_t)$ and $\tilde{p}_{t+1|t}(x_{t+1})$ of $p_{t|t}(x_t)$ and $p_{t+1|t}(x_{t+1})$. Let $\mathcal{K}_{t|t} = \{-m_{t|t}, -m_{t|t}+1 \ldots, m_{t|t}\}$, $\mathcal{K}_{t+1|t} = \{m_{t+1|t}, -m_{t+1|t}+1 \ldots, m_{t+1|t}-1, m_{t+1|t}\}$.

**Algorithm 1.** *(Approximate Bayesian estimation)*
*Consider initially* (23) *and then, for every $t \in \mathcal{T}$:*

- *Update step: Set $m_{t|t} = m_{t|t-1} + m_y$ and for $k \in \mathcal{K}_{t|t}$ compute*

$$\tilde{c}_{t|t}[k] = \mathsf{N}_c(\tilde{c}_y[y_t;k] \otimes \tilde{c}_{t|t-1}[k]) \tag{26}$$

  *and set*

$$\tilde{p}_{t|t}(x_t) = \sum_{k\in\mathcal{K}_{t|t}} \tilde{c}_{t|t}[k] e^{j2\pi(k./L)\cdot x_t}, \quad x_t \in \mathcal{X}.$$

- *Prediction step: Set* $m_{t+1|t} = m_{t|t}$ *and for* $k \in \mathcal{K}_{t+1|t}$ *compute*

$$\tilde{c}_{t+1|t}[k] = \mathsf{N}_c\left(\frac{1}{\tilde{L}}\left(\sum_{k\in\mathcal{K}_{t|t}} \tilde{c}_{t|t}[k] B_1(A^T\omega, L, k)\right) e^{-j\omega\cdot(Bu_t)} P_w(\omega)\Big|_{\omega=(\frac{2\pi k_1}{L_1},\ldots,\frac{2\pi k_n}{L_n})}\right) \tag{27}$$

*and set*

$$\tilde{p}_{t+1|t}(x_{t+1}) = \sum_{k\in\mathcal{K}_{t+1|t}} \tilde{c}_{t+1|t}[k] e^{j2\pi(k./L)\cdot x_{t+1}}, \quad x_{t+1} \in \mathcal{X}.$$

□

Note that the factor $m_{t|t} = m_{t+1|t}$ determining the number of coefficients $(2m_{t|t-1} + 1)$ only grows linearly with time $t$, $m_{t|t-1} = m_{0|-1} + tm_y$. In fact, the complexity of the conditioned state distribution, measured by the number of coefficients, remains constant at prediction steps and grows only linearly at each update step. The number of coefficients $m_{0|-1}$ and $m_y$ to be selected depends on the balance between accuracy and computational effort; more coefficients increase the accuracy of the pdf at the expense of computational effort. A natural choice is to remove the coefficients, whose power is below a predefined threshold after the update step.

## IV. NUMERICAL EXAMPLES

In this section, two examples are discussed. The first example considers a scalar linear system with Gaussian noise and process disturances, which allows us to compare the accuracy of the proposed approximate method of Section III-B to the exact conditioned pdfs obtained using the Kalman filter. In the second example, we consider an example inspired by electron microscopy.

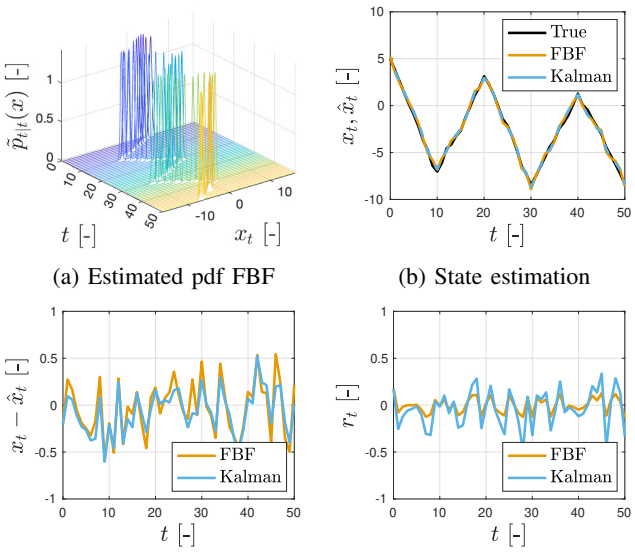
### A. Comparison with the Kalman filter for simple example

Consider the following first-order linear system

$$f(x_t, u_t, w_t) = x_t + u_t + w_t, \tag{28}$$
$$g(x_t, v_t) = x_t + v_t, \tag{29}$$

where $\mathcal{X} := \{x_t \in \mathbb{R} | -10 \leq x_t \leq 10\}, \mathcal{U} := \{u_t \in \mathbb{R} | -1 \leq u_t \leq 1\}$, and $t \in \mathbb{N}_0$. The disturbances and measurement noises are independent, zero mean, and distributed according to a Gaussian: $w_t \sim \mathcal{N}(0, 0.1)$ and $v_t \sim \mathcal{N}(0, 0.1)$. The Gaussian distributions are truncated and scaled such that $w_t \in [-5, 5]$ and $v_t \in [-5, 5]$. We set $\hat{x}_0 = 0$, and $\mathbb{E}[(x_0 - \hat{x}_0)^2] = 3$. The Fourier series coefficients that characterize the output equations $\tilde{c}_y[y_t; k]$ are obtained by sampling the pdf's of the output equations $p_y(y_t; x_t)$ for several values of $x_t \in \mathcal{X}$, and computing the Fourier series coefficients of these samples. The factor that determines the initial number of Fourier series coefficients is set to $m_{0|-1} = 1600$, and $m_y = m_{0|-1}$.

Fig. 1(a) shows how the approximated pdfs $\tilde{p}_{t|t}(x)$ vary over time $t$. Fig. 1(b) compares $\hat{x}_t = \mathbb{E}[x_t]$, computed from these pdfs, with the actual state $x_t$ in. The estimation closely follows the actual state. For comparison, the state estimates and residual error of both the Kalman filter as well as the Fourier basis method are visualized in Fig. 1(c). The performance is assessed based on the residual errors, similar to [3]. The residual errors are calculated according to $r_t = y_t - \mathbb{E}[g(\hat{x}_t, 0)]$, and shown in Fig. 1d. From Fig. 1, it can be concluded that the Fourier basis functions method has a similar performance to the (optimal) Kalman filter benchmark.




(a) Estimated pdf FBF (b) State estimation

(c) State estimation error (d) Residual error

Fig. 1: Comparison of the Fourier basis functions (FBF) method with the Kalman filter for the first example

### B. Application to electron microscopy

Part of the tuning of an electron microscope is determining the level of aberrations (modeled as states), which are coupled to features of a ronchigram (outputs) [14]. It is not possible to estimate the level of aberrations from a single ronchigram. In general multiple hypotheses on this level should be tracked. In this numerical example, two types of aberrations are considered: focus and astigmatism (modelled $x_1$ and $x_2$, respectively). Those two aberrations are linked to two outputs ($y_1$ and $y_2$) by non-linear output equations. The control inputs ($u_1$ and $u_2$) act on each aberration directly without couplings, while also each aberration is completely decoupled, thus the system has integrator dynamics. This leads to a simplified model of the electron microscope calibration problem, which considers the problem of inferring the state from features, and given by

$$f_1(\cdot) = x_{1,t} + u_{1,t} + w_{1,t}, \quad f_2(\cdot) = x_{2,t} + u_{2,t} + w_{2,t},$$
$$g_1(\cdot) = \alpha_1 x_{1,t}^2 + \alpha_2 x_{2,t}^2 + v_{1,t},$$
$$g_2(\cdot) = \frac{\alpha_3 x_{1,t}}{\alpha_4 + \alpha_5 (x_{2,t}^2 - x_{1,t}^2)^2} + v_{2,t}$$

where $\mathcal{X} := \{x_t \in \mathbb{R}^2 | [-5, 5] \times [-5, 0]\}$, $\mathcal{U} := \{u_t \in \mathbb{R}^2 | [-1, 1]^2\}$, $t \in \mathbb{N}_0$, and $\alpha_i$ with $i \in \{1, .., 5\}$ are constants. This model was obtained from system identification from available data; the details of the identification procedure are omitted here. The disturbances are independent and uniformly distributed in the sets: $w_{i,t} \in [-0.2, 0.2]$, $i \in \{1, 2\}$.

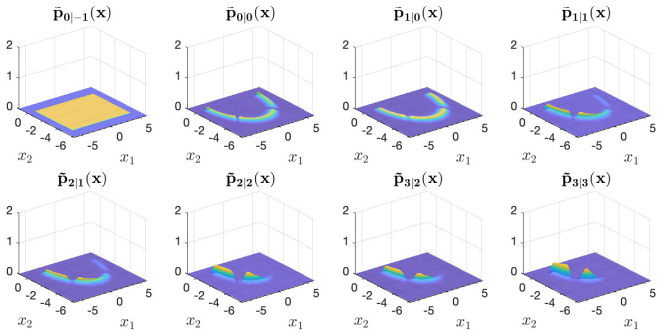



Fig. 2: Evolution of $\tilde{p}_{t|t}(x)$ over time

The measurement noises are independent, zero mean and distributed according to a Gaussian with a variance of 0.1: $v_{i,t} \sim \mathcal{N}(0, 0.1)$, $i \in \{1, 2\}$. Furthermore, they are truncated and scaled such that $v_{i,t} \in [-1, 1]$, $i \in \{1, 2\}$. The initial position estimate $\tilde{p}_{0,-1}(x)$ is considered as a uniform distribution with support $\mathcal{X}$. In the simulations, the control inputs are selected such that the states do not leave set $\mathcal{X}$. Both $\tilde{p}_{0|-1}(x)$ and $\tilde{p}(y; x)$ are sampled on a grid of $125 \times 75$ Fourier series coefficients.

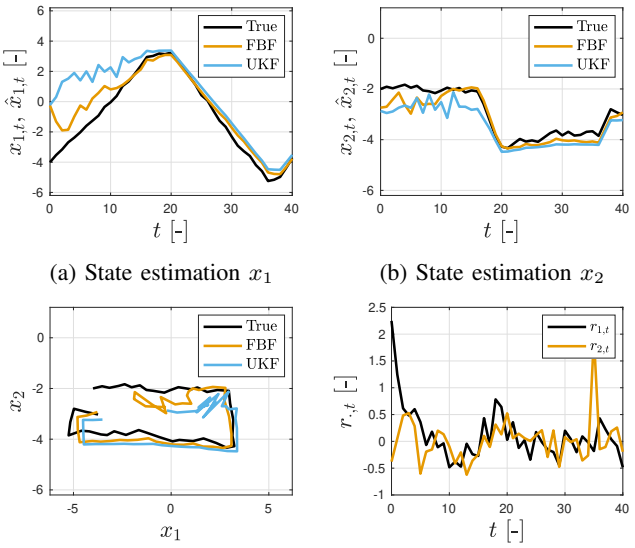
The evolution of the approximated pdf is visualized in Fig. 2 for the first 3 timesteps. The figure shows clearly the development from an initial uniform distribution towards a small region with high probability, indicating the certainty in the estimate improves over time. A more detailed figure on the quality of the estimation is given in Fig. 3, where a comparison is made with the Unscented Kalman filter (UKF), as proposed in [5]. Since the UKF cannot deal with uniform distributions, the disturbances are estimated by $\tilde{w}_{i,t} \sim \mathcal{N}(0, 0.15)$, $i \in \{1, 2\}$, the initial state estimate is set on $\hat{x}_0 = [0, -2.5]^T$, and

$$P_0 = \mathbb{E}\left[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T\right] = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

The Fourier basis function method's estimate converges faster to the true state than the UKF. In addition to the faster convergence, the proposed method has a better tracking performance, since it does not have the requirement of Gaussian noise statistics, as the UKF has. The computation time of the UKF on the other hand is less than the Fourier basis function method. Furthermore, Fig. 3(d) shows clearly that the residual error does not grow unboundedly.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, *exact* Bayesian estimation is pursued for the class of systems with linear dynamics and arbitrary (non-linear) output equations in the frequency domain through the Fourier transform, using the countable number of Fourier coefficients. This perspective naturally leads to an *approximate* method, where high frequencies are disregarded to obtain a finite number of Fourier coefficients. The approximate method is computationally tractable, since the number of coefficients grows only linearly with time. Future research directions include analyzing the estimation error, in order to give error bounds for the approximate method.




(a) State estimation $x_1$

(b) State estimation $x_2$

(c) State trajectory

(d) Residual error FBF

Fig. 3: Comparison of the Fourier basis functions (FBF) method with the Unscented Kalman filter (UKF) of the second example

## REFERENCES


[1] V. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation," *IEEE Pervasive Computing*, vol. 2, no. 3, pp. 24–33, 2003.

[2] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 03 1960.

[3] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[4] B. Anderson and J. Moore, *Optimal Filtering*. Information and system sciences series, Prentice-Hall, 1979.

[5] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Signal Processing, Sensor Fusion, and Target Recognition VI* (I. Kadar, ed.), vol. 3068, pp. 182 – 193, International Society for Optics and Photonics, SPIE, 1997.

[6] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113(6), April 1993.

[7] M. Athans and C. Chang, "Adaptive estimation and parameter identification using multiple model estimation algorithm," Tech. Rep. 1976-28, Lincoln Lab, Mass. Inst. Technol., Lexington, MA, 1976.

[8] A. Jasra, S. S. Singh, J. S. Martin, and E. McCoy, "Filtering via approximate Bayesian computation," *Statistics and Computing*, vol. 22, pp. 1223–1237, 2012.

[9] V. Smidl and A. Quinn, "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, 2008.

[10] D. Brunn, F. Sawo, and U. D. Hanebeck, "Nonlinear multidimensional Bayesian estimation with Fourier densities," in *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 1303–1308, 2006.

[11] P. W. Hawkes and J. C. Spence, *Science of microscopy*, vol. 1. Springer, 2007.

[12] V. Ronchi, "Forty years of history of a grating interferometer," *Appl. Opt.*, vol. 3, pp. 437–451, Apr 1964.

[13] N. Schnitzer, S. H. Sung, and R. Hovden, "Introduction to the ronchigram and its calculation with ronchigram. com," *Microscopy Today*, vol. 27, no. 3, pp. 12–15, 2019.

[14] H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Y. Tanishiro, and K. Takayanagi, "Measurement method of aberration from ronchigram by autocorrelation function," *Ultramicroscopy*, vol. 108, no. 11, pp. 1467–1475, 2008.

[15] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*. What's New in Statistics Series, Pearson, 2019.