Kernel density estimates in particle filter

Coufal, David
2014

Kernel density estimates
in particle filter

David Coufal

Institute of Computer Science
Academy of Sciences of the Czech Republic

Kernel density estimates
in particle filter

David Coufal[1]

Technical report No. 1210

Abstract:

The paper deals with kernel density estimates of filtering densities in the particle filter. The convergence of the estimates is investigated by means of Fourier analysis. It is shown that the estimates converge to the theoretical filtering densities in the mean integrated squared error under a certain assumption on the Sobolev character of the filtering densities. A sufficient condition is presented for the persistence of this Sobolev character over time. Both results are extended to partial derivatives of the estimates and filtering densities.

Keywords:
particle filters, kernel methods, Fourier analysis

---

[1]Department of Nonlinear Dynamics and Complex Systems of the Institute of Computer Science AS CR, Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic. E-mail: david.coufal@cs.cas.cz.

1 Introduction

The particle filter enables its user to efficiently compute integral characteristics (moments) of distributions of interest. In the filtering problem, these distributions are traditionally referred to as the *filtering distributions*. In the particle filter, the filtering distribution is approximated by an empirical measure. This measure is implemented in the form of a weighted sum of Dirac measures located at randomly (empirically) generated points called *particles*. Particles are generated sequentially by the algorithm which is an instance of the *sequential Monte Carlo methods* [1, 2].

The theoretical result that justifies the application of the particle filter is that the generated empirical measures converge to the theoretical filtering distribution as the number of particles goes to infinity [1, 3]. Approximating the filtering distribution by an empirical measure is extremely useful for estimating moments of the distribution because they correspond to weighted sums of values of moment functions over generated particles.

The filtering distribution has typically a density with respect to the corresponding Lebesgue measure. This density is called the *filtering density*. The knowledge of a suitable analytical approximation of the filtering density has several advantages. Let us mention, for example, the possibility of computing densities of related conditional distributions and conditional expected values in an analytical form. The other benefit is that one can get a deeper insight into the character of the filtering distribution through the analysis of its density approximation.

From these practical, and of course also theoretical, reasons the issue of the analytical approximation of the filtering densities is the subject of ongoing research. The problem has been addressed in [1], Chapter 12, [4, 5] and recently in [6].

In this paper, we deal with the estimation/approximation of filtering densities using the nonparametric kernel density estimation methodology. We use an approach based on Fourier analysis inspired by the book of Tsybakov [7]. We will show that the convergence of kernel density estimates is assured even if the particles generated by the particle filter are not i.i.d., which is the common assumption in the application of kernel methods.

The paper presents two main results. The first result is the convergence of the kernel density estimates to the theoretical filtering density at a fixed time of operation of the filter, provided that the number of generated particles goes to infinity. The result is based on the notion of the Sobolev character of the filtering density. The second result gives a condition under which this Sobolev character is retained over time. Thus, the first result applies at any time of operation of the filter. Both results are extended to partial derivatives of the estimates and filtering densities.

The rest of the paper is organized as follows. In the next section we review the basics of the particle filter's theory together with the related convergence results. Section 3 deals with a review of nonparametric kernel density estimation methods with the focus on the Fourier analysis approach. Sections 4 and 5 present the announced main results of the paper. Section 6 shows an application of the developed theory in an example related to the Kalman filter. The paper is concluded by Section 7.

2 Particle filter

The basics of the particle filter and general filtering theory can be found in [1, 2, 3, 8] and [9]. However, there is a plenty of other literature specialized in these subjects. Nevertheless, we present here the essential framework of the related methodology in order that the paper be self-contained.

2.1 Filtering problem

The filtering problem is the task of determining the optimal estimate of an inaccessible value of the actual state of a stochastic process on the basis of knowledge of accessible observations. The observations establish a stochastic process called the *observation process*. The observation process is interconnected with a principal stochastic process which is called the *signal process*. Let us be more specific.

Let $(\Omega, \mathcal{A}, P)$ be a probabilistic space with two stochastic processes $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$, $\{\boldsymbol{Y}_t\}_{t=1}^{\infty}$ specified on it. The first process $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$, $\boldsymbol{X}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_x}, \mathcal{B}(\mathbb{R}^{d_x}))$, $t \in \mathbb{N}_0$, $d_x \in \mathbb{N}$ is the signal process. The

signal process is considered to represent generally an inhomogeneous Markov chain with a continuous state space. The probabilistic behavior of the chain is determined by the initial distribution $\pi_0(d\boldsymbol{x}_0)$ of $\boldsymbol{X}_0$ and by the set of transition kernels $K_{t-1} : \mathcal{B}(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x} \to [0, 1]$, $t \in \mathbb{N}$. We denote by $K_{t-1}(d\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ the measure represented by the transition kernel K_{t-1} for $\boldsymbol{x}_{t-1} \in \mathbb{R}^{d_x}$ being fixed.

Let $\{\boldsymbol{Y}_t\}_{t=1}^\infty$, $\boldsymbol{Y}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_y}, \mathcal{B}(\mathbb{R}^{d_y}))$, $t \in \mathbb{N}$, $d_y \in \mathbb{N}$ be the observation process specified on the basis of the signal process by formula

$$\boldsymbol{Y}_t = h_t(\boldsymbol{X}_t) + \boldsymbol{V}_t, \quad t \in \mathbb{N}, \tag{2.1}$$

where $h_t : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$, $t \in \mathbb{N}$ are Borel functions and $\boldsymbol{V}_t$ are (all)-other-variables independent random variables specified on $(\Omega, \mathcal{A}, P)$. That is, $\boldsymbol{V}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_y}, \mathcal{B}(\mathbb{R}^{d_y}))$, $t \in \mathbb{N}$, $d_y \in \mathbb{N}$ and $P(\boldsymbol{V}_t \in d\boldsymbol{v}_t | \boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t-1}, \boldsymbol{V}_{1:t-1}) = P(\boldsymbol{V}_t \in d\boldsymbol{v}_t)$ for all $t \in \mathbb{N}$. The (all)-other-variables independence of $\boldsymbol{V}_t$ transfers on observations in the following way:

$$P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t | \boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t-1}) = P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t | \boldsymbol{X}_t). \tag{2.2}$$

Indeed, we have $\sigma(\boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t-1}) = \sigma(\boldsymbol{X}_{0:t}, \boldsymbol{V}_{1:t-1})$ due to (2.1). $\boldsymbol{V}_{1:t-1}$ is independent of $(\boldsymbol{Y}_t, \boldsymbol{X}_{0:t})$, therefore $P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t | \boldsymbol{X}_{0:t}, \boldsymbol{V}_{1:t-1}) = P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t | \boldsymbol{X}_{0:t})$. The assertion is finally obtained by the Markov property of the signal process. Remark that for $t = 1$, the left-hand side of (2.2) reads as $P(\boldsymbol{Y}_1 \in d\boldsymbol{y}_1 | \boldsymbol{X}_{0:1})$.

2.2 Filtering distribution and filtering density

As stated, the purpose of filtering is to present the optimal estimate of the actual state $\boldsymbol{x}_t \in \mathbb{R}^{d_x}$ of the signal process using the actual and past observations $\boldsymbol{y}_{1:t} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$. This is done at each time instant $t \in \mathbb{N}$. It is the classical result that under the assumption of L_2 integrability of $\boldsymbol{X}_t$, the L_2-optimal estimate corresponds to the conditional expectation $\mathbb{E}[\boldsymbol{X}_t | \boldsymbol{Y}_{1:t}]$. In what follows we will assume that $\boldsymbol{X}_t \in L_2(\Omega, \mathcal{A}, P)$ for each $t \in \mathbb{N}_0$.

For fixed observations $\boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t}$, the conditional expectation $\mathbb{E}[\boldsymbol{X}_t | \boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t}]$ can be determined on the basis of the related conditional distribution $P(\boldsymbol{X}_t \in d\boldsymbol{x}_t | \boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t})$. This distribution then represents the filtering distribution at time $t \in \mathbb{N}$ and will be approximated by an empirical measure generated by the particle filter.

In the standard setting of the filtering problem, all the involved finite-dimensional distributions have bounded and continuous densities with respect to the corresponding Lebesgue measures. Especially, we assume that $\pi_0(d\boldsymbol{x}_0) = p_0(\boldsymbol{x}_0) \, d\boldsymbol{x}_0$, $K_{t-1}(d\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \, d\boldsymbol{x}_t$ and $P(\boldsymbol{V}_t \in d\boldsymbol{v}_t) = g_t^v(\boldsymbol{v}_t) \, d\boldsymbol{v}_t$. This enables us to identify the respective filtering density, which is the density of $P(\boldsymbol{X}_t \in d\boldsymbol{x}_t | \boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t})$.

The conditional density of $P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t | \boldsymbol{X}_t = \boldsymbol{x}_t)$ is determined by formula (2.1). The density is denoted $g_t(\boldsymbol{y}_t | \boldsymbol{x}_t)$ and writes as

$$g_t(\boldsymbol{y}_t | \boldsymbol{x}_t) = g_t^v(\boldsymbol{y}_t - h_t(\boldsymbol{x}_t)). \tag{2.3}$$

The joint density of $(\boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t})$ has then form

$$p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t}) = p_0(\boldsymbol{x}_0) \prod_{k=1}^t g_k(\boldsymbol{y}_k | \boldsymbol{x}_k) \, K_{k-1}(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}). \tag{2.4}$$

These specifications are induced by the conditional independence of observations (2.2) and by the standard theory of Markov chains with a continuous state space.

The filtering density is $p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t})$ for $t \in \mathbb{N}$. Employing the joint distribution (2.4), we have

$$p(\boldsymbol{x}_t | \boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{x}_t, \boldsymbol{y}_{1:t})}{p(\boldsymbol{y}_{1:t})} = \frac{\int p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t}) \, d\boldsymbol{x}_{0:t-1}}{\int p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t}) \, d\boldsymbol{x}_{0:t}}. \tag{2.5}$$

The above integrals are generally inexpressible in a closed form. However, certain recursive analytical relations can be stated. These relations are called the *filtering equations* and are addressed in the next section.

2.3 Filtering equations

The filtering equations describe recursively the evolution of the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ over time. They consists of the *prediction formula* (2.6) and the *update formula* (2.7).

Lemma 1. *Let the joint density be given by formula (2.4), then*

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})\,d\boldsymbol{x}_{t-1} \tag{2.6}$$

for $t \geq 2$, and $p(\boldsymbol{x}_1) = \int K_0(\boldsymbol{x}_1|\boldsymbol{x}_0)p_0(\boldsymbol{x}_0)\,d\boldsymbol{x}_0$ for $t = 1$.

Proof. We get the result from (2.4) by series of integrations. Let us start with $t = 1$. In this case, formula (2.4) reads as $p(\boldsymbol{x}_{0:1}, \boldsymbol{y}_1) = g_1(\boldsymbol{y}_1|\boldsymbol{x}_1)K_0(\boldsymbol{x}_1|\boldsymbol{x}_0)\,p_0(\boldsymbol{x}_0)$. By integrating out $\boldsymbol{y}_1$ we get $p(\boldsymbol{x}_{0:1}) = K_0(\boldsymbol{x}_1|\boldsymbol{x}_0)\,p_0(\boldsymbol{x}_0)$ and the result is obtained by integration with respect to $\boldsymbol{x}_0$.

In the general case of $t \geq 2$, we get the following expressions by the transcription of (2.4) and integrating out $\boldsymbol{y}_t$,

$$\begin{aligned} p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t}) &= g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{0:t-1}, \boldsymbol{y}_{1:t-1}), \\ p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t-1}) &= K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{0:t-1}, \boldsymbol{y}_{1:t-1}). \end{aligned}$$

Subsequently, the integration w.r.t. $\boldsymbol{x}_{0:t-2}$ and $\boldsymbol{x}_{t-1}$ gives

$$\begin{aligned} p(\boldsymbol{x}_{t-1:t}, \boldsymbol{y}_{1:t-1}) &= K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:t-1}), \\ p(\boldsymbol{x}_t, \boldsymbol{y}_{1:t-1}) &= \int K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:t-1})\,d\boldsymbol{x}_{t-1}. \end{aligned}$$

Finally, dividing both sides of the last formula by the marginal density $p(\boldsymbol{y}_{1:t-1})$ gives the result. $\square$

Lemma 2. *Let the joint density be given by formula (2.4), then*

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})}{\int g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\,d\boldsymbol{x}_t}, \quad t \in \mathbb{N}, \tag{2.7}$$

with $p(\boldsymbol{x}_1|\boldsymbol{y}_{1:0})$ understood as $p(\boldsymbol{x}_1)$ for $t = 1$.

Proof. We start with the Bayes' rule and rearrange

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_t)p(\boldsymbol{x}_t)}{p(\boldsymbol{y}_{1:t})}, \\ p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t)}{p(\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1})}, \\ p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{y}_{1:t-1})p(\boldsymbol{y}_{1:t-1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t)}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})p(\boldsymbol{y}_{1:t-1})}. \end{aligned}$$

We again use the Bayes' rule on $p(\boldsymbol{y}_{1:t-1}|\boldsymbol{x}_t)$, which gives

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{y}_{1:t-1})p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})p(\boldsymbol{y}_{1:t-1})p(\boldsymbol{x}_t)}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})p(\boldsymbol{y}_{1:t-1})p(\boldsymbol{x}_t)}.$$

Considering the conditional independence of $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{y}_{1:t-1})$, which is expressed by $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{y}_{1:t-1}) = p(\boldsymbol{y}_t|\boldsymbol{x}_t)$, and cancelling out the $p(\boldsymbol{y}_{1:t-1})p(\boldsymbol{x}_t)$ terms we get the final formula

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})}.$$

In the denominator, the normalizing constant is obtained by integration

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\,d\boldsymbol{x}_t.$$

3

As we have $p(\boldsymbol{y}_t|\boldsymbol{x}_t) = g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)$, this finishes the proof. □

The development of the filtering density over time is split into two sub-steps by the filtering equations. The prediction density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$ is obtained in the first sub-step and, in the second one, it is updated to the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ on the basis of the actual observation $\boldsymbol{y}_t$.

Speaking in the language of distributions, the filtering distribution is usually denoted by π_t, i.e., $\pi_t(d\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})\,d\boldsymbol{x}_t$. π_t is also alternatively referred to as the *update distribution (measure)*. The prediction density then corresponds to the density of the so-called *prediction distribution (measure)* denoted by $\overline{\pi}_t$, i.e., $\overline{\pi}_t(d\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\,d\boldsymbol{x}_t$.

2.4 Particle filter

The time evolution of the filtering distribution can be seen as a recursive alternation between the prediction and update distributions $\overline{\pi}_t$ and π_t. This characterization fits to the particle filter operation because the filter alternately generates empirical prediction and update measures.

In the particle filter, empirical measures are constructed as weighted sums of Dirac measures localized at particles generated by the filter. The justification of this representation stems from the Strong Law of Large Numbers (SLLN). Assuming that $\{\boldsymbol{X}_i = \boldsymbol{x}_i\}_{i=1}^n$, $n \in \mathbb{N}$ is an i.i.d. sample from a given distribution μ and constructing the empirical measure $\delta_n(d\boldsymbol{x})$ as

$$\delta_n(d\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{x}_i}(d\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{X}_i}(d\boldsymbol{x}), \tag{2.8}$$

the SLLN states that for any integrable function f, the integral over this empirical measure converges a.s. to the integral over the distribution μ. Note that in (2.8), the second expression points out the random character of $\delta_n(d\boldsymbol{x})$, in fact, $\delta_n(d\boldsymbol{x})$ is a random measure.

Dealing with the filtering problem practically, we are not able to directly generate i.i.d. samples from π_t because we do not have any closed-form representation of the filtering density at our disposal. However, due to the product character of the joint density $p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t})$, one can state an algorithm which recursively generates samples (particles) that are used for constructing empirical counterparts of $\overline{\pi}_t$ and π_t distributions.

The construction of empirical measures proceeds sequentially. The particles generated in the previous cycle of operation are employed in the actual cycle. A stochastic update of particles and their weights is taken in each cycle. The weights are updated on the basis of the actual observation. The procedure is in fact an instance of the sequential Monte Carlo methods applied in the context of the filtering problem [1] and the algorithm follows the recursion described by the filtering equations. However, there is one extension.

In the raw mode of operation, the update measure is constructed as a non-uniformly weighted sum of Dirac measures. As explained in [1], as $t \in \mathbb{N}$ increases the distribution of weights becomes more and more skewed and practically, after a few time steps, only one particle has a non-zero weight. To avoid this degeneracy, the *resampling step* is introduced.

During the resampling step, a non-uniformly weighted empirical measure is resampled into its uniformly weighted counterpart. The basic type of resampling is based on the idea of discarding particles with low weights (with respect to $1/n$) and promote those with high weights. Practically, it is done by sampling from the multinomial distribution $\mathcal{M}$ over original particles with the probabilities of selection given by particles' weights. This type of resampling corresponds to the sampling with replacement from the set of original particles with the probabilities of individual selections corresponding to the individual weights. Let us stress here that the resampled particles *does not constitute an i.i.d. sample.*

We are now ready to present the operation of the particle filter in the algorithmic way:

- **0. declarations**
  $n \in \mathbb{N}$ - the number of particles,
  $T \in \mathbb{N}$ - the computational horizon,
  $p_0(\boldsymbol{x}_0)$ - the initial density of $\boldsymbol{X}_0$,
  $K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, $t = 1, \ldots T$ - the transition densities.

- **1. initialization**
  $t = 0$,
  sample $\{\overline{\boldsymbol{x}}_0^i \sim p_0(\boldsymbol{x}_0)\}_{i=1}^n$,
  constitute $\widehat{\pi}_0^n(d\boldsymbol{x}_0) = \frac{1}{n}\sum_{i=1}^n \delta_{\overline{\boldsymbol{x}}_0^i}(d\boldsymbol{x}_0)$,
  set $\pi_0^n(d\boldsymbol{x}_0) = \widehat{\pi}_0^n(d\boldsymbol{x}_0)$, i.e., $\{\boldsymbol{x}_0^i = \overline{\boldsymbol{x}}_0^i\}_{i=1}^n$.

- **2. sampling**
  $t = t + 1$,
  sample $\{\overline{\boldsymbol{x}}_t^i \sim K_{t-1}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^i)\}_{i=1}^n$,
  for $i = 1:n$ compute
  $$\widetilde{w}(\overline{\boldsymbol{x}}_t^i) = \frac{g_t(\boldsymbol{y}_t - h_t(\overline{\boldsymbol{x}}_t^i))}{\sum_{j=1}^n g_t(\boldsymbol{y}_t - h_t(\overline{\boldsymbol{x}}_t^j))},$$
  constitute $\widehat{\pi}_t^n(d\boldsymbol{x}_t) = \sum_{i=1}^n \widetilde{w}(\overline{\boldsymbol{x}}_t^i)\,\delta_{\overline{\boldsymbol{x}}_t^i}(d\boldsymbol{x}_t)$.

- **3. resampling**
  using $\mathcal{M}(n, \widetilde{w}(\overline{\boldsymbol{x}}_t^1), \ldots, \widetilde{w}(\overline{\boldsymbol{x}}_t^n))$, resample $\{\boldsymbol{x}_t^i\}_{i=1}^n$ from $\{\overline{\boldsymbol{x}}_t^i\}_{i=1}^n$ and constitute
  $\pi_t^n(d\boldsymbol{x}_t) = \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{x}_t^i}(d\boldsymbol{x}_t)$.

- **4.** if $t = T$ end, else go to step 2.

Algorithm 1. Operation of the particle filter.

The particle filter sequentially generates three empirical measures in each single cycle of its operation. These are the empirical prediction measure $\overline{\pi}_t^n$, the empirical update measure before resampling $\widehat{\pi}_t^n$ and the empirical update measure after resampling π_t^n. The third measure then forms the empirical counterpart of the filtering distribution π_t.

A comparison of the evolution of the empirical measures with the evolution of the theoretical distributions can be done by means of the following schema:

$$
\begin{array}{ccccccccc}
\pi_0 \to & \overline{\pi}_1^n & \to & \widehat{\pi}_1^n \to \pi_1^n & \to \ldots \to & \overline{\pi}_t^n & \to & \widehat{\pi}_t^n \to \pi_t^n \\
\pi_0 \to & \overline{\pi}_1 & \to & \pi_1 & \to \ldots \to & \overline{\pi}_t & \to & \pi_t
\end{array}
$$

Figure 2.1: The evolution of the empirical and theoretical distributions in the particle filter.

2.5 Convergence results

The particle filter algorithm is known that the empirical measures $\overline{\pi}_t^n$ and π_t^n converge weakly a.s. (they are random measures) to their theoretical counterparts as the number of generated particles goes to infinity. We will not go into details of the proof of the assertion, we only mention the result and its L_2 variant related to our research.

To present the convergence theorems, we denote the class of all real bounded and continuous functions over $\mathbb{R}^{d_x}$ by $\mathcal{C}_b(\mathbb{R}^{d_x})$, the supremum norm of a function $f : \mathbb{R}^{d_x} \to \mathbb{R}$ by $||f||_\infty$, i.e., $||f||_\infty = \sup_{\boldsymbol{x}}\{|f(\boldsymbol{x})|\}$, and the integral of f over the measure μ by μf. Further, it is assumed that the transition kernels of the signal process possess the Feller property. That is, $K_{t-1}f \in \mathcal{C}_b(\mathbb{R}^{d_x})$ for any $f \in \mathcal{C}_b(\mathbb{R}^{d_x})$ and $t \in \mathbb{N}$, where $(K_{t-1}f)(\boldsymbol{x}_{t-1}) = \int f(\boldsymbol{x}_t)K_{t-1}(d\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$. The other assumption is that the densities $g_t(\boldsymbol{y}_t | \cdot)$ of (2.3), $t \in \mathbb{N}$ are bounded, continuous and strictly positive functions.

Theorem 1. *Let $\{\overline{\pi}_t^n\}_{t=1}^T$ and $\{\pi_t^n\}_{t=1}^T$ be the sequences of empirical measures generated by the particle filter for some fixed observation history $\{\boldsymbol{Y}_t = \boldsymbol{y}_t\}_{t=1}^T$, $T \in \mathbb{N}$. Then for all $t \in \{1, \ldots, T\}$ and $f \in \mathcal{C}_b(\mathbb{R}^{d_x})$,*
$$\lim_{n \to \infty} |\overline{\pi}_t^n f - \overline{\pi}_t f| = 0 \ \ a.s., \quad \lim_{n \to \infty} |\pi_t^n f - \pi_t f| = 0 \ \ a.s.$$

Proof. See [1], Chapter 2 for a discussion of the convergence theorems. Other source is [3], Section IV. Paper [6] has a proof even for unbounded functions in Proposition 1(b). $\qquad\square$

In our research we employ the L_2 version of the theorem for π_t^n. It reads as follows:

Theorem 2. *Let $\{\pi_t^n\}_{t=1}^T$ be the sequence of empirical measures generated by the particle filter for some fixed observation history $\{\boldsymbol{Y}_t = \boldsymbol{y}_t\}_{t=1}^T$, $T \in \mathbb{N}$. Then for all $t \in \{1, \dots, T\}$ and $f \in \mathcal{C}_b(\mathbb{R}^{d_x})$,*

$$\mathbb{E}[|\pi_t^n f - \pi_t f|^2] \leq \frac{c_t^2 ||f||_\infty^2}{n} \tag{2.9}$$

with $c_t > 0$ being a constant for fixed $t \in \{1, \dots, T\}$.

Proof. In this formulation, the theorem is presented in [3], Section V (authors use c_t instead ours c_t^2). $\qquad\square$

Remark that the L_1 version, i.e., $\mathbb{E}[|\pi_t^n f - \pi_t f|]$, is treated in [1], Theorem 2.4.1. The theorem is mentioned for general L_p norm, $p \geq 1$ in [6], Proposition 1(a).

The theorem holds also for the class $\mathcal{C}_b^{\mathbb{C}}(\mathbb{R}^{d_x})$ of bounded and continuous complex functions of real variables over $\mathbb{R}^{d_x}$. That is, it holds also for functions $h : \mathbb{R}^{d_x} \to \mathbb{C}$, $h(\boldsymbol{x}) = f(\boldsymbol{x}) + \mathrm{i}g(\boldsymbol{x})$, $f, g \in \mathcal{C}_b(\mathbb{R}^{d_x})$, where i denotes the imaginary unit. Clearly, the extension on complex functions is due to the triangle inequality for the absolute value (the modulus) of a complex number.

3   Kernel methods

Kernel methods are widely used for nonparametric estimation of densities of probability distributions with the vast literature available on the topic. Here we review the very basics of the related methodology. We focus in more details on the application of Fourier analysis in this field. Our review is mainly based on the standard works of [11] and [12], and the recent book by Tsybakov [7].

3.1   Basics of kernel methods

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, $n \in \mathbb{N}$ be a set of independent random variables identically distributed as the real random variable $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Let the distribution of $\boldsymbol{X}$ have the density $f : \mathbb{R}^d \to [0, \infty)$ with respect to the d-dimensional Lebesgue measure. A nonparametric kernel density estimate of f is constructed on the basis of an i.i.d. sample $\{\boldsymbol{X}_i = \boldsymbol{x}_i\}_{i=1}^n$ from the distribution of $\boldsymbol{X}$. The estimate is constructed as a generalization of the classical histogram by replacing the indicator function, which specifies individual bins of the histogram, by a more general function $K : \mathbb{R}^d \to \mathbb{R}$ which is commonly referred to as the *kernel function* or simply as the *kernel*.

The definition formula of the standard d-variate nonparametric kernel density estimate writes as

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\right) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h}\right). \tag{3.1}$$

In the formula, the second expression points out the random character of the estimate. That is, for each $\boldsymbol{x} \in \mathbb{R}^d$, the estimate $\hat{f}_n(\boldsymbol{x})$ constitutes a random variable whose distribution is determined by the distribution of $\boldsymbol{X}$ and by the value of the parameter $h > 0$ which is called the *bandwidth*.

Due to the random character of $\hat{f}_n(\boldsymbol{x})$, there is the relevant question of the consistency and unbiasedness of the estimate. In the univariate case, the classical result of Parzen [14] (see also [11], p. 71) states the conditions under which the estimate is consistent. The result extends on the multivariate case, see e.g. [15]. The conditions are imposed on the properties of the kernel function and on the evolution of the bandwidth h in dependence on the sample size $n \in \mathbb{N}$. We mention only that h is required to evolve in such a way that 1) $\lim_{n \to \infty} h(n) = 0$ and 2) $\lim_{n \to \infty} nh^d(n) = \infty$.

The investigation on the bias of $\hat{f}_n(\boldsymbol{x})$ is closely related to the investigation on the quality of the estimate in terms of the *mean squared error* - $\mathrm{MSE}_{\boldsymbol{x}}(\hat{f}_n)$. For a fixed point $\boldsymbol{x} \in \mathbb{R}^d$, the error is specified as $\mathrm{MSE}_{\boldsymbol{x}}(\hat{f}_n) = \mathbb{E}[(\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}))^2]$. Employing properties of mean and variance, it writes as

$$\mathrm{MSE}_{\boldsymbol{x}}(\hat{f}_n) = (\mathbb{E}[\hat{f}_n(\boldsymbol{x})] - f(\boldsymbol{x}))^2 + var[\hat{f}_n(\boldsymbol{x})] = (b[\hat{f}_n](\boldsymbol{x}))^2 + \sigma^2[\hat{f}_n](\boldsymbol{x}), \tag{3.2}$$

where the term $b[\hat{f}_n](\boldsymbol{x}) = \mathbb{E}[\hat{f}_n(\boldsymbol{x})] - f(\boldsymbol{x})$ is the *bias* and $\sigma^2[\hat{f}_n](\boldsymbol{x}) = var[\hat{f}_n(\boldsymbol{x})]$ the *variance* of the kernel density estimate $\hat{f}_n(\boldsymbol{x})$ at the point $\boldsymbol{x} \in \mathbb{R}^d$.

The $\text{MSE}_{\boldsymbol{x}}(\hat{f}_n)$ is the local measure of the quality of the estimate. It is desirable to have also a corresponding global measure. Expectedly, such the measure deals with local errors accumulated over the whole domain of the estimated density. Mathematically, the accumulation is performed by integration. This leads to the notion of the *mean integrated squared error* (MISE) of a kernel density estimate.

The MISE of the kernel density estimate $\hat{f}_n$ is defined and expressed on the basis of (3.2) using the Fubini's theorem as

$$
\begin{aligned}
\text{MISE}(\hat{f}_n) &= \mathbb{E} \int [(\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}))^2]\, d\boldsymbol{x} = \int \text{MSE}_{\boldsymbol{x}}(\hat{f}_n)\, \boldsymbol{x} \\
&= \int (\mathbb{E}[\hat{f}_n(\boldsymbol{x})] - f(\boldsymbol{x}))^2\, d\boldsymbol{x} + \int var[\hat{f}_n(\boldsymbol{x})]\, d\boldsymbol{x} \\
&= \int (b[\hat{f}_n](\boldsymbol{x}))^2\, d\boldsymbol{x} + \int \sigma^2[\hat{f}_n](\boldsymbol{x})\, d\boldsymbol{x}.
\end{aligned}
$$

The formula consists of two summands which are the integrated versions of the squared bias and variance terms of the $\text{MSE}_{\boldsymbol{x}}(\hat{f}_n)$. The value of the $\text{MISE}(\hat{f}_n)$ depends on the value of the bandwidth h.

It is a standard observation that the bias and variance terms behave in the opposite way with respect to the magnitude of the bandwidth. That is, for $n \in \mathbb{N}$ fixed, if h decreases, i.e., if $h \to 0$, then the bias goes to zero, and we have the asymptotic unbiasedness of the $\hat{f}_n(\boldsymbol{x})$ estimate. However, the variance increases. If h increases, i.e., if $h \to \infty$, the bias increases too, but the variance term diminishes. Thus, we encounter here the situation of the *bias-variance trade-off* when minimizing the $\text{MISE}(\hat{f}_n)$ by adjusting the bandwidth h.

The specification of the optimal value h^*_{MISE} minimizing (3.3) can be made analytically only if (3.3) has a closed-form expression. This is known only in some specific cases, for example, when the estimated density f is a convex sum of normal densities, see [11], p. 37 or [12], p. 102 for the related explicit formulas for $\text{MISE}(\hat{f}_n)$. To deal with the minimization problem generally, the widely used approach is to investigate the asymptotic behavior of the MISE with respect to the sample size $n \in \mathbb{N}$ going to infinity (AMISE analysis). The result based on the Taylor's expansion of the estimated density f states ([11], p. 85, [12], p. 99) that

$$
\text{MISE}(\hat{f}_n) \approx n^{-1} h^{-d} R(K) + \frac{1}{4} h^4 (\mu_2(K))^2 \int (\nabla^2 f(\boldsymbol{x}))^2\, d\boldsymbol{x} \tag{3.3}
$$

for $R(K) = \int K^2(\boldsymbol{u})\, d\boldsymbol{u}$, $\mu_2(K) = \int u_1^2 K(\boldsymbol{u})\, d\boldsymbol{u}$, $\nabla^2 f(\boldsymbol{x}) = \sum_{i=1}^d (\partial^2/\partial x_i^2) f(\boldsymbol{x})$. Using standard calculus, the minimizer of the above formula reads as

$$
h^*_{\text{AMISE}} = \left[\frac{d \cdot R(K)}{\mu_2(K)^2 \int (\nabla^2 f(\boldsymbol{x}))^2\, d\boldsymbol{x}} \frac{1}{n} \right]^{1/(d+4)}. \tag{3.4}
$$

In (3.3), the terms $R(K)$ and $\mu_2(K)$ can be further minimized over a set of appropriate kernels. The minimizer is known as the *Epanechnikov kernel* which is specified as $K_e(\boldsymbol{u}) = \frac{1}{2} \vartheta_d^{-1} (d+2)(1 - ||\boldsymbol{u}||^2)_+$ where ϑ_d is the volume of the d-dimensional unit sphere, $||\cdot||$ is the Euclidean norm and $(\cdot)_+ = \max\{0, \cdot\}$ is the positive part.

AMISE analysis represents the standard approach to the analytic specification of a suitable value of the bandwidth when constructing a kernel density estimate, even though the specification of h^*_{AMISE} requires the knowledge of partial derivatives of the density f under estimation. Typically, to overcome the deadlock, the respective entities are somehow estimated from data [12].

However, in Section 1.2.4 of his book [7], Tsybakov provides a deeper criticism of the asymptotic approach. It stems from the fact that the optimality of h^*_{AMISE} is related to a *fixed density* f and not to a well defined class of densities. In Proposition 1.7, Tsybakov shows that for a given fixed density f it is possible to construct such a non-negative kernel estimate that the $\text{MISE}(\hat{f}_n)$ diminishes, but this

cannot be done uniformly over a sufficiently broad class of densities. Examples of such classes, e.g. Hölder, Sobolev or Nikol'ski classes, are presented in [7]. The Sobolev class is treated in Definition 2 below.

Based on this criticism, Tsybakov presents a different approach to the MISE analysis in Section 1.3 of [7]. The approach relies on Fourier analysis.

3.2 Fourier analysis

In this section, we deal with the application of Fourier analysis in the area of nonparametric kernel density estimation. We mainly follow the presentation of Tsybakov given in Chapter 1 of [7]. In [7], results are provided for the univariate case. In order to the results could be applied in our research presented in Section 4, we have extended them into multiple dimensions.

In the probability theory, Fourier analysis is intimately interconnected with the notion of the characteristic function. Let $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a d-variate real random vector with the joint distribution $\mu(d\boldsymbol{x})$. The characteristic function $\phi(\boldsymbol{\omega}) : \mathbb{R}^d \to \mathbb{C}$ of $\boldsymbol{X}$ is defined as the integral transform

$$\phi(\boldsymbol{\omega}) = \mathbb{E}[e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}\rangle}] = \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}\rangle} \mu(d\boldsymbol{x}), \ \ \boldsymbol{\omega} \in \mathbb{R}^d, \tag{3.5}$$

where $\langle\cdot,\cdot\rangle$ denotes the dot product. It is well known that the transform provides the complete characterization of the distribution of $\boldsymbol{X}$; and we often speak about the Fourier transform of the random vector $\boldsymbol{X}$.

The other quite common view of the Fourier transform comes from the area of applied mathematics. Let $f : \mathbb{R}^d \to \mathbb{R}$ be an integrable function (a signal in electrical engineering), i.e., let $f \in L_1(\mathbb{R}^d)$, then its Fourier transform is specified as

$$\mathcal{F}[f](\boldsymbol{\omega}) = \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}\rangle} f(\boldsymbol{x}) \, d\boldsymbol{x}, \ \ \boldsymbol{\omega} \in \mathbb{R}^d. \tag{3.6}$$

Formula (3.6) can be treated as the special case of formula (3.5) when the distribution of $\boldsymbol{X}$ is absolutely continuous with respect to the d-dimensional Lebesgue measure and has the density f, i.e., $\mu(d\boldsymbol{x}) = f(\boldsymbol{x}) \, d\boldsymbol{x}$. On the other hand, in (3.6), f need not be necessarily a density, only the integrability is assumed.

Let $f, g \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$, i.e., we consider functions both L_1 and L_2 integrable over $\mathbb{R}^d$, then the following properties of the multivariate Fourier transform are relevant to our research:

- continuity: $\mathcal{F}[f]$ is uniformly continuous on $\mathbb{R}^d$,

- linearity: $\mathcal{F}[af + bg](\boldsymbol{\omega}) = a\mathcal{F}[f](\boldsymbol{\omega}) + b\mathcal{F}[g](\boldsymbol{\omega}), \quad a, b \in \mathbb{R}$,

- shifting: $\mathcal{F}[f(\boldsymbol{x} - \boldsymbol{s})](\boldsymbol{\omega}) = e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{s}\rangle}\mathcal{F}[f](\boldsymbol{\omega}), \ \ \boldsymbol{s} \in \mathbb{R}^d$,

- scaling: $\mathcal{F}[f(\boldsymbol{x}/h)/h^d](\boldsymbol{\omega}) = \mathcal{F}[f](h\boldsymbol{\omega}), \ \ h > 0$,

- shifting & scaling: $\mathcal{F}[f((\boldsymbol{x} - \boldsymbol{s})/h)/h^d] = e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{s}\rangle}\mathcal{F}[f](h\boldsymbol{\omega}), \ \ \boldsymbol{s} \in \mathbb{R}^d$,

- complex conjugate: $\overline{\mathcal{F}[f](\boldsymbol{\omega})} = \mathcal{F}[f](-\boldsymbol{\omega})$,

- convolution: $\mathcal{F}[f * g](\boldsymbol{\omega}) = \mathcal{F}[f](\boldsymbol{\omega})\mathcal{F}[g](\boldsymbol{\omega})$,

- derivative: $\mathcal{F}[f_{i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega}) = (-\mathrm{i})^m(\omega_1^{i_1} \cdot \ldots \cdot \omega_d^{i_d})\mathcal{F}[f](\boldsymbol{\omega})$,

- symmetry: if $f(-\boldsymbol{x}) = f(\boldsymbol{x})$, then $\mathcal{F}[f](-\boldsymbol{\omega}) = \mathcal{F}[f](\boldsymbol{\omega})$,

- isometry, due to the Plancheler's formula for $f \in L_2(\mathbb{R}^d)$:

$$\int f^2(\boldsymbol{x}) \, d\boldsymbol{x} = \frac{1}{(2\pi)^d} \int |\mathcal{F}[f](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}.$$

Now, the uniformly weighted sum of Dirac measures $\delta_n(d\boldsymbol{x})$ introduced in formula (2.8) represents the probability distribution which does not have any density with respect to the corresponding Lebesgue measure. Its characteristic function $\phi_n(\boldsymbol{\omega})$ is specified as

$$\phi_n(\boldsymbol{\omega}) = \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}\rangle}\delta_n(d\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n}e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_j\rangle}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{3.7}$$

Note that $\phi_n(\boldsymbol{\omega})$ constitutes a random variable for $\boldsymbol{\omega} \in \mathbb{R}^d$ being fixed.

Under the assumption of $L_1(\mathbb{R}^d)$ integrability of the employed kernel K, we can consider the Fourier transform of the multivariate density kernel estimate (3.1). Using the linearity and the shifting & scaling property of the Fourier transform, $\mathcal{F}[\hat{f}_n](\boldsymbol{\omega})$ is specified by formula

$$\mathcal{F}[\hat{f}_n](\boldsymbol{\omega}) = \frac{1}{n}\sum_{j=1}^{n}\mathcal{F}\left[\frac{1}{h^d}K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h}\right)\right] = \frac{1}{n}\sum_{j=1}^{n}e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_j\rangle}\mathcal{F}[K](h\boldsymbol{\omega}). \tag{3.8}$$

Writing $K_{\mathcal{F}}(\boldsymbol{\omega})$ for $\mathcal{F}[K](\boldsymbol{\omega})$ we obtain the compact expression of $\hat{f}_n$ in the form

$$\mathcal{F}[\hat{f}_n](\boldsymbol{\omega}) = \phi_n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}). \tag{3.9}$$

This shows that the standard kernel estimator which is based on an i.i.d. sample is obtained by the convolution of the employed kernel with the uniformly weighted sum of Dirac measures corresponding to the sample.

To proceed with the investigation of the MISE of density kernel estimates in the frequency domain, we present a multivariate version of Lemma 1.2 from [7].

Lemma 3. *Let $\{\boldsymbol{X}_j\}_{j=1}^{n}$ be an i.i.d. sample from a distribution with the density f. Let the characteristic function of $\boldsymbol{X}_j$ be $\phi(\boldsymbol{\omega})$. Then for ϕ_n of (3.7) we have*

(i) $\mathbb{E}[\phi_n(\boldsymbol{\omega})] = \phi(\boldsymbol{\omega})$,
(ii) $\mathbb{E}[|\phi_n(\boldsymbol{\omega})|^2] = \left(1 - \frac{1}{n}\right)|\phi(\boldsymbol{\omega})|^2 + \frac{1}{n}$,
(iii) $\mathbb{E}[|\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2] = \frac{1}{n}(1 - |\phi(\boldsymbol{\omega})|^2)$.

Proof. To show (i), consider the i.i.d. character of $\{\boldsymbol{X}_j\}_{j=1}^{n}$,

$$\mathbb{E}[\phi_n(\boldsymbol{\omega})] = \frac{1}{n}\sum_{j=1}^{n}\int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}\rangle}f(\boldsymbol{x})\,d\boldsymbol{x} = \frac{1}{n}\sum_{j=1}^{n}\phi(\boldsymbol{\omega}) = \phi(\boldsymbol{\omega}).$$

To show (ii), note that

$$
\begin{aligned}
\mathbb{E}[|\phi_n(\boldsymbol{\omega})|^2] &= \mathbb{E}[\phi_n(\boldsymbol{\omega})\overline{\phi_n(\boldsymbol{\omega})}] = \mathbb{E}[\phi_n(\boldsymbol{\omega})\phi_n(-\boldsymbol{\omega})] \\
&= \mathbb{E}\left[\frac{1}{n^2}\sum_{j,k:j\neq k}e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_j\rangle}e^{-\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_k\rangle}\right] + \frac{n}{n^2} \\
&= \frac{1}{n^2}\sum_{j,k:j\neq k}\mathbb{E}\left[e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_j\rangle}\right]\mathbb{E}\left[e^{-\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{X}_k\rangle}\right] + \frac{1}{n} \\
&= \frac{n^2-n}{n^2}\phi(\boldsymbol{\omega})\phi(-\boldsymbol{\omega}) + \frac{1}{n} \\
&= \left(1 - \frac{1}{n}\right)|\phi(\boldsymbol{\omega})|^2 + \frac{1}{n}. \tag{3.10}
\end{aligned}
$$

Case (iii) folows from (ii) a (i). Indeed, $\mathbb{E}[|\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2] =$

$$
\begin{aligned}
&= \mathbb{E}[(\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))\overline{(\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))}] \\
&= \mathbb{E}[(\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))(\phi_n(-\boldsymbol{\omega}) - \phi(-\boldsymbol{\omega}))] \\
&= \mathbb{E}[|\phi_n(\boldsymbol{\omega})|^2 - \phi_n(\boldsymbol{\omega})\phi(-\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})\phi_n(-\boldsymbol{\omega}) + |\phi(\boldsymbol{\omega})|^2] \\
&= \mathbb{E}[|\phi_n(\boldsymbol{\omega})|^2] - \phi(\boldsymbol{\omega})\phi(-\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})\phi(-\boldsymbol{\omega}) + |\phi(\boldsymbol{\omega})|^2 \\
&= \mathbb{E}[|\phi_n(\boldsymbol{\omega})|^2] - 2|\phi(\boldsymbol{\omega})|^2 + |\phi(\boldsymbol{\omega})|^2 \\
&= \left(1 - \frac{1}{n}\right)|\phi(\boldsymbol{\omega})|^2 + \frac{1}{n} - |\phi(\boldsymbol{\omega})|^2 \\
&= \frac{1}{n}(1 - |\phi(\boldsymbol{\omega})|^2).
\end{aligned}
$$

This concludes the proof. $\qquad\square$

Let us assume that both density f and kernel K belong also to $L_2(\mathbb{R}^d)$. Then employing the Plancherel's theorem and (3.9), we get for the MISE of (3.3) the expression

$$
\mathrm{MISE}(\hat{f}_n) = \frac{1}{(2\pi)^d}\,\mathbb{E}\int |\phi_n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega}. \tag{3.11}
$$

The next theorem provides the exact computation of the $\mathrm{MISE}(\hat{f}_n)$ for any fixed $n \in \mathbb{N}$.

Theorem 3. *Let $f \in L_2(\mathbb{R}^d)$ be a density and $K \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ a kernel. Then for all $n \geq 1$ and $h > 0$ the MISE of the i.i.d. based kernel estimator $\hat{f}_n$ of (3.1) has the form*

$$
\begin{aligned}
\mathrm{MISE}(\hat{f}_n) &= \frac{1}{(2\pi)^d}\left[\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2|\phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} + \frac{1}{n}\int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\,d\boldsymbol{\omega}\right] \\
&\quad - \frac{1}{(2\pi)^d}\frac{1}{n}\int |\phi(\boldsymbol{\omega})|^2|K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\,d\boldsymbol{\omega}. \tag{3.12}
\end{aligned}
$$

Proof. As $\phi, K \in L_2(\mathbb{R}^d)$ and $|\phi(\boldsymbol{\omega})| \leq 1$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$, all the integrals are finite. To obtain the Fourier MISE formula it suffices to develop (3.11),

$$
\begin{aligned}
&\mathbb{E}\int |\phi_n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} \\
&= \mathbb{E}\int |(\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))K_{\mathcal{F}}(h\boldsymbol{\omega}) - (1 - K_{\mathcal{F}}(h\boldsymbol{\omega}))\phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} \\
&= \mathbb{E}\int ((\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))K_{\mathcal{F}}(h\boldsymbol{\omega}) - (1 - K_{\mathcal{F}}(h\boldsymbol{\omega}))\phi(\boldsymbol{\omega})) \\
&\qquad\quad \cdot \overline{((\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega}))K_{\mathcal{F}}(h\boldsymbol{\omega})} - \overline{(1 - K_{\mathcal{F}}(h\boldsymbol{\omega}))\phi(\boldsymbol{\omega}))}\,d\boldsymbol{\omega} \\
&= \int \mathbb{E}[|\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2]|K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 + |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2|\phi(\boldsymbol{\omega})|^2 \\
&\qquad + (\mathbb{E}[(\phi_n(\boldsymbol{\omega})] - \phi(\boldsymbol{\omega}))K_{\mathcal{F}}(h\boldsymbol{\omega})\,\overline{(1 - K_{\mathcal{F}}(h\boldsymbol{\omega}))\phi(\boldsymbol{\omega})}) \\
&\qquad + (1 - K_{\mathcal{F}}(h\boldsymbol{\omega}))\phi(\boldsymbol{\omega}))\,(\mathbb{E}[\overline{\phi_n(\boldsymbol{\omega})}] - \overline{\phi(\boldsymbol{\omega})})\overline{K_{\mathcal{F}}(h\boldsymbol{\omega})}\,d\boldsymbol{\omega} \\
&= \int \mathbb{E}[|\phi_n(\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2]|K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 + |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2|\phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} \\
&= \frac{1}{n}\int (1 - |\phi(\boldsymbol{\omega})|^2)|K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} + \int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2|\phi(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega}
\end{aligned}
$$

After rearranging we obtain the assertion of the theorem. $\qquad\square$

We are now going to discuss the individual terms in the Fourier MISE formula (3.12). We start with the notion of the *order of a kernel*.

10

Definition 1. *Let $\ell \geq 1$ be an integer. We say that the kernel $K \colon \mathbb{R}^d \to \mathbb{R}$ is of order ℓ, if K is $L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ integrable, its Fourier transform $K_{\mathcal{F}}(\boldsymbol{\omega})$ is real, satisfies $K_{\mathcal{F}}(\mathbf{0}) = 1$ and has all partial derivatives $K_{\mathcal{F},i_1,\ldots i_d}^{(m)} = \partial^m K_{\mathcal{F}}/\partial_{i_1} \ldots \partial_{i_d}$, $m = i_1 + \cdots + i_d$, $m \in \mathbb{N}$ up to the ℓ-th order and it holds that $K_{\mathcal{F},i_1,\ldots i_d}^{(m)}(\mathbf{0}) = 0$ for all $m = 1, \ldots, \ell$.*

Remark that the above definition imposes the following conditions on a multivariate kernel to be of order $\ell \geq 1$, $\ell \in \mathbb{N}$:

- $\int K(\boldsymbol{u}) \, d\boldsymbol{u} = 1$,

- $\int u_1^{i_1} \cdots u_d^{i_d} K(\boldsymbol{u}) \, d\boldsymbol{u} = 0$ for $m = 1, \ldots, \ell$.

Indeed, at the origin we have $K_{\mathcal{F}}(\mathbf{0}) = \int e^{\mathrm{i}\langle \mathbf{0}, \boldsymbol{u} \rangle} K(\boldsymbol{u}) \, d\boldsymbol{u} = \int K(\boldsymbol{u}) \, d\boldsymbol{u} = 1$. For the m-th partial derivative, we get

$$K_{\mathcal{F},i_1,\ldots i_d}^{(m)}(\boldsymbol{\omega}) = \int (\mathrm{i}u_1)^{i_1} \cdots (\mathrm{i}u_d)^{i_d} \, e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{u} \rangle} K(\boldsymbol{u}) \, d\boldsymbol{u},$$

hence $0 = K_{\mathcal{F},i_1,\ldots i_d}^{(m)}(\mathbf{0}) = \mathrm{i}^m \int u_1^{i_1} \cdots u_d^{i_d} K(\boldsymbol{u}) \, d\boldsymbol{u}$.

From the remark, it follows that kernels of order $\ell \geq 2$ must take negative values. If such kernels are allowed in kernel estimates, then $\hat{f}_n$ of (3.1) may also take negative values. However, this is not a serious drawback because we can always take as the final estimate the positive part of $\hat{f}_n$, i.e., $\hat{f}_n^+ = \max\{0, \hat{f}_n\}$. At each point $\boldsymbol{x} \in \mathbb{R}^d$, the $\mathrm{MSE}_{\boldsymbol{x}}$ of $\hat{f}_n^+(\boldsymbol{x})$ is always smaller than that of negative $\hat{f}_n(\boldsymbol{x})$. Therefore we have also $\mathrm{MISE}(\hat{f}_n^+) \leq \mathrm{MISE}(\hat{f}_n)$.

The first term

For the first term in the Fourier MISE formula (3.12), we are able to say something more specific if we consider the order of the kernel involved in the estimate.

Theorem 4. *Let $K \colon \mathbb{R}^d \to \mathbb{R}$ be a kernel of order $\ell \geq 1$, $\ell \in \mathbb{N}$. Then there exists a constant $A > 0$ such that*

$$\sup_{\boldsymbol{\omega} \in \mathbb{R}^d \setminus \{\mathbf{0}\}} \frac{|1 - K_{\mathcal{F}}(\boldsymbol{\omega})|}{||\boldsymbol{\omega}||^\ell} \leq A, \tag{3.13}$$

and

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \leq A^2 h^{2\ell} \int ||\boldsymbol{\omega}||^{2\ell} |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \tag{3.14}$$

for any function f with the Fourier transform $\phi(\boldsymbol{\omega})$ and $h > 0$.

Proof. We employ the multidimensional Taylor's theorem. Because the kernel K is of order $\ell \geq 1$, its Fourier transform $K_{\mathcal{F}}(\boldsymbol{\omega})$ is real and by the Taylor's theorem

$$K_{\mathcal{F}}(\boldsymbol{\omega}) = K_{\mathcal{F}}(\mathbf{0}) + \frac{1}{1!} \sum_{i=1}^d K_{\mathcal{F},i}^{(1)}(\mathbf{0}) \, \omega_i + \cdots + \frac{1}{\ell!} \sum_{i_1,\ldots,i_d=1}^d K_{\mathcal{F},i_1,\ldots,i_d}^{(\ell)}(\mathbf{0}) \, \omega_{i_1} \cdots \omega_{i_d} + R_\ell(\boldsymbol{\omega})$$

with $\lim_{\boldsymbol{\omega} \to \mathbf{0}} R_\ell(\boldsymbol{\omega})/||\boldsymbol{\omega}||^\ell = 0$ for the reminder, where $|| \cdot ||$ is the Euclidean norm.

Because the involved partial derivatives equal to zero, the remainder writes $R_\ell(\boldsymbol{\omega}) = K_{\mathcal{F}}(\boldsymbol{\omega}) - K_{\mathcal{F}}(\mathbf{0}) = K_{\mathcal{F}}(\boldsymbol{\omega}) - 1$ and $\lim_{\boldsymbol{\omega} \to \mathbf{0}} |1 - K_{\mathcal{F}}(\boldsymbol{\omega})|/||\boldsymbol{\omega}||^\ell = 0$ by the Taylor's theorem.

Let us define $A_\ell(\boldsymbol{\omega}) = |1 - K_{\mathcal{F}}(\boldsymbol{\omega})|/||\boldsymbol{\omega}||^\ell$ for $\boldsymbol{\omega} \neq \mathbf{0}$, and $A_\ell(\mathbf{0}) = 0$. The function $A_\ell \colon \mathbb{R}^d \to [0, \infty)$ is continuous on $\mathbb{R}^d$ and attains its maximum on the unit ball $||\boldsymbol{\omega}|| \leq 1$. We denote this maximum by M_1, $M_1 = \max_{\{||\boldsymbol{\omega}|| \leq 1\}}\{A_\ell(\boldsymbol{\omega})\}$. Because $K \in L_1(\mathbb{R}^d)$, we have $0 \leq |K_{\mathcal{F}}(\boldsymbol{\omega})| \leq M_2 < \infty$. Indeed, $|K_{\mathcal{F}}(\boldsymbol{\omega})| \leq \int |e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{u} \rangle}| |K(\boldsymbol{u})| \, d\boldsymbol{u} \leq \int |K(\boldsymbol{u})| \, d\boldsymbol{u} = M_2 < \infty$. Therefore, $|1 - K_{\mathcal{F}}|/||\boldsymbol{\omega}||^\ell \leq 1 + M_2$ for $||\boldsymbol{\omega}|| > 1$. Composing both cases one gets $A_\ell(\boldsymbol{\omega}) \leq \max\{M_1, 1 + M_2\} = A < \infty$ for $\boldsymbol{\omega} \in \mathbb{R}^d$.

The inequality (3.14) is implied by (3.13) as follows:

$$\sup_{\boldsymbol{\omega} \in \mathbb{R}^d \setminus \{\mathbf{0}\}} \frac{|1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|}{||h\boldsymbol{\omega}||^{\ell}} \leq A,$$

$$|1 - K_{\mathcal{F}}(h\boldsymbol{\omega})| \leq A||h\boldsymbol{\omega}||^{\ell},$$

$$|1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \leq A^2||h\boldsymbol{\omega}||^{2\ell},$$

$$|1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 \leq A^2 h^{2\ell}||\boldsymbol{\omega}||^{2\ell}|\phi(\boldsymbol{\omega})|^2,$$

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \leq A^2 h^{2\ell} \int ||\boldsymbol{\omega}||^{2\ell}|\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}.$$

This concludes the proof. □

The other terms in formula (3.12) refer to individual properties of the kernel and density under considerations. We mention only two straightforward observations.

The second term

The second term can be directly translated from the frequency to the "time" domain by the Plancherel's theorem and the scaling property of the Fourier transform:

$$\frac{1}{n} \int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} = \frac{(2\pi)^d}{nh^{2d}} \int K^2(\boldsymbol{x}/h) \, d\boldsymbol{x} = \frac{(2\pi)^d}{nh^d} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u}. \tag{3.15}$$

The third term

The third term is actually the correction term. For this term we have the following inequality:

$$\frac{1}{(2\pi)^d} \frac{1}{n} \int |\phi(\boldsymbol{\omega})|^2 |K(h\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq \frac{||K_{\mathcal{F}}||_{\infty}^2}{n} \int f^2(\boldsymbol{x}) \, d\boldsymbol{x},$$

where $||K_{\mathcal{F}}||_{\infty} = \sup_{\boldsymbol{\omega}}\{|K_{\mathcal{F}}(\boldsymbol{\omega})|\}$.

3.3 The upper bound on the Fourier MISE formula

Concerning an upper bound on the Fourier MISE formula (3.12), we actually sum up the results obtained in the preceding sections. First of all, to obtain the upper bound we can omit the correction (the third) term in (3.12). The second term is solely determined by the properties of the kernel, which is expressed by formula (3.15). Finally, to obtain a bound on the first term, the properties of the density the data are sampled from and the properties of the kernel have to be matched somehow. To do this we introduce the so-called Sobolev class of densities.

Definition 2. *Let $\beta \geq 1$ be an integer and $L > 0$. The Sobolev class of densities $\mathcal{P}_{S(\beta, L)}$ consists of all probability density functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying*

$$\int ||\boldsymbol{\omega}||^{2\beta}|\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq (2\pi)^d L^2, \tag{3.16}$$

where $\phi(\boldsymbol{\omega}) = \mathcal{F}[f](\boldsymbol{\omega})$ and $|| \cdot ||$ is the Euclidean norm.

The condition (3.16) is related to the boundedness of partial derivatives of densities in the Sobolev class; e.g., it can be shown that if $\int (\partial f/\partial x_j)^2 \, d\boldsymbol{x} \leq L_j < \infty$ for all $j = 1, \ldots, d$, then (3.16) holds for $\beta = 1$ and $L = ||(L_1, \ldots, L_d)||$. Furthermore, if $f \in \mathcal{P}_S(\beta, L)$, for some $\beta \in \mathbb{N}$ and $L > 0$, then $f \in L_2(\mathbb{R}^d)$.

Now, the announced matching is provided by the fitting the order of the kernel to the Sobolev character of the estimated density. The next theorem, which is the variant of Theorem 1.5 in [7], provides the final result.

Theorem 5. *Let $n \in \mathbb{N}$ be the number of i.i.d. samples from a distribution with the density $f : \mathbb{R}^d \to [0, \infty)$ which is β-Sobolev for some $\beta \in \mathbb{N}$ and $L > 0$, i.e., $f \in \mathcal{P}_S(\beta, L)$. Let K be a kernel of order β. Assume that inequality (3.13) holds for some constant $A > 0$. Fix $\alpha > 0$ and set $h = \alpha n^{-\frac{1}{2\beta+d}}$. Then for any $n \geq 1$ the kernel density estimate $\hat{f}_n$ satisfies*

$$\sup_{f \in \mathcal{P}_S(\beta, L)} \mathbb{E} \int (\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}))^2 \, d\boldsymbol{x} \leq C \cdot n^{-\frac{2\beta}{2\beta+d}}, \tag{3.17}$$

where $C > 0$ is a constant depending only on α, β, d, A, L and on the kernel K.

Proof. By Theorem 4 and from the definition of the Sobolev class of densities, we have

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq (2\pi)^d A^2 h^{2\beta} L^2.$$

Plugging this into the Fourier MISE formula (3.12) and employing

$$\frac{1}{(2\pi)^d n} \int |K(h\boldsymbol{\omega})|^2 d\boldsymbol{\omega} = \frac{1}{nh^d} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u},$$

we get for $h = \alpha n^{-\frac{1}{2\beta+d}}$ the following:

$$h^{2\beta} = \alpha^{2\beta} n^{-\frac{2\beta}{2\beta+d}}, \ (nh^d)^{-1} = n^{-1}\alpha^{-d} n^{\frac{d}{2\beta+d}} = \alpha^{-d} n^{-\frac{2\beta}{2\beta+d}}$$

and

$$\begin{aligned}
\mathrm{MISE}(\hat{f}_n) &\leq \frac{1}{(2\pi)^d} \left[\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \ + \ \frac{1}{n} \int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \right] \\
&\leq A^2 h^{2\beta} L^2 + \frac{1}{nh^d} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u}, \\
&\leq (AL)^2 \alpha^{2\beta} n^{-\frac{2\beta}{2\beta+d}} + \alpha^{-d} n^{-\frac{2\beta}{2\beta+d}} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u}, \\
&\leq \left[(AL)^2 \alpha^{2\beta} + \alpha^{-d} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u} \right] \cdot n^{-\frac{2\beta}{2\beta+d}}, \\
&\leq C(\alpha, \beta, d, A, L, K) \cdot n^{-\frac{2\beta}{2\beta+d}}.
\end{aligned}$$

This concludes the proof. $\qquad\square$

The theorem provides the upper bound on the MISE of the multivariate kernel density estimate (3.1), if the order of the employed kernel fits to the Sobolev character of the density the employed data are sampled from.

4 Particle filter and kernel methods

This section presents our own research in the area of the combination of the particle filter and kernel methods. The main question here is if the kernel density estimates constructed on the basis of empirical measures approximate the related filtering densities reasonably well. The main obstacle to a direct application of the presented kernel estimation methodology is the fact that the generated empirical measures are not based on i.i.d. samples due to the resampling step of the filter.

Our results are twofold. First, we show that, despite the mentioned obstacle, the standard kernel density estimates still converge to the related filtering densities. The proof of the assertion is based on Fourier analysis of the convergence result for the particle filter.

The second result concerns a deeper analysis of the obtained convergence formula. The convergence result is based on the assumption on the Sobolev character of the filtering densities. We present a sufficient condition for the persistency of this Sobolev character over time.

We extend both results to the partial derivatives of the kernel density estimates and to the partial derivatives of the filtering densities, respectively.

4.1 Convergence of kernel density estimates

To start, let us remind that the particle filter generates at each time step $t = 1, \ldots, T$, $T \in \mathbb{N}$ the empirical measure $\pi_t^n(d\boldsymbol{x}_t) = \frac{1}{n} \sum_{i=1}^n \delta_{\boldsymbol{x}_t^i}(d\boldsymbol{x}_t)$. This measure approximates the related filtering distribution π_t that is assumed to have the density $p_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ with respect to the d-dimensional Lebesgue measure, i.e., $\pi_t(d\boldsymbol{x}_t) = p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t$.

A carrier of the empirical measure π_t^n is the set of particles $\{\boldsymbol{x}_t^i\}_{i=1}^n$, $n \in \mathbb{N}$. This set does not constitute an i.i.d. sample from π_t. If one constructs the standard kernel density estimate on the basis of $\{\boldsymbol{x}_t^i\}_{i=1}^n$ and the selected kernel K, i.e., the estimate

$$\hat{p}_t^n(\boldsymbol{x}_t) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\boldsymbol{x}_t - \boldsymbol{x}_t^i}{h}\right), \tag{4.1}$$

then we ask if $\hat{p}_t^n$ converges in the MISE to the filtering density p_t, provided that the number of particles goes to infinity.

Theorem 6. *In the filtering problem, let $\{\pi_t\}_{t=0}^T$, $\{p_t\}_{t=0}^T$, $T \in \mathbb{N}$ be the sequences of filtering distributions and corresponding filtering densities. Let p_t, $t \in \{0, 1, \ldots, T\}$ be β-Sobolev for some $\beta \in \mathbb{N}$ and $L_t > 0$, i.e., $p_t \in \mathcal{P}_{S(\beta, L_t)}$. Let $\{\pi_t^n\}_{t=1}^T$, $\{\hat{p}_t^n\}_{t=1}^T$, $n \in \mathbb{N}$ be the sequences of the empirical measures generated by the particle filter and related kernel density estimates (4.1) with the bandwidth varying as $h(n) = \alpha n^{-\frac{1}{2\beta+d}}$ for some $\alpha > 0$. Let the kernel K employed in the estimates be of order β. Then we have the following evolution of the MISE of $\hat{p}_t^n$ over time $t \in \{1, \ldots, T\}$:*

$$\mathbb{E}\left[\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2\, d\boldsymbol{x}_t\right] \le C_t^2 \cdot n^{-\frac{2\beta}{2\beta+d}}, \tag{4.2}$$

where

$$C_t = AL_t\alpha^\beta + c_t\alpha^{-d/2}||K||. \tag{4.3}$$

In (4.3), A is the constant of Theorem 4, c_t, $t \in \{1, \ldots, T\}$ are the constants of Theorem 2 and $||K||$ is the L_2 norm of the kernel K.

Proof. The proof is based on the employment of the Fourier transform. We start by the assertion of Theorem 2:

$$\mathbb{E}[|\pi_t^n f - \pi_t f|^2] \le \frac{c_t^2||f||_\infty^2}{n}, \tag{4.4}$$

where we replace a general function $f \in \mathcal{C}_b^{\mathbb{C}}(\mathbb{R}^{d_x})$ by the complex exponential specified on $\mathbb{R}^d$. Note that $d_x = d$.

Let $f(\boldsymbol{x}_t) = e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle}$, then $||f||_\infty = 1$. Denoting $\psi_t^n = \mathcal{F}[\pi_t^n]$ and $\psi_t = \mathcal{F}[\pi_t]$ we have from the above

$$\mathbb{E}[|\psi_t^n(\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2] \quad \le \quad \frac{c_t^2}{n},$$

$$|K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \mathbb{E}[|\psi_t^n(\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2] \quad \le \quad |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n},$$

$$\mathbb{E}\left[|\psi_t^n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\right] \quad \le \quad |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n},$$

$$\mathbb{E}\left[\int |\psi_t^n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\, d\boldsymbol{\omega}\right] \quad \le \quad \frac{c_t^2}{n}\int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\, d\boldsymbol{\omega},$$

$$\mathbb{E}\left[\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t^*(\boldsymbol{x}_t))^2\, d\boldsymbol{x}_t\right] \quad \le \quad \frac{c_t^2}{nh^d}\int K^2(\boldsymbol{u})\, d\boldsymbol{u}. \tag{4.5}$$

For any density p_t and its convolution $p_t^* = p_t * (h^{-d}K(\cdot/h))$,

$$\int (p_t^*(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2\, d\boldsymbol{x}_t \quad = \quad \frac{1}{(2\pi)^d}\int |\psi_t(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2\, d\boldsymbol{\omega}$$

$$= \quad \frac{1}{(2\pi)^d}\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2|\psi_t(\boldsymbol{\omega})|^2\, d\boldsymbol{\omega}. \tag{4.6}$$

14

We assume that the employed kernel has order β and $p_t \in \mathcal{P}_{S(\beta, L_t)}$. Therefore the right-hand side of (4.6) is bounded according to Theorem 4. Further, there is nothing random here and we can apply the expectation with no effect to obtain

$$\mathbb{E}\left[\int (p_t^*(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right] \leq A^2 h^{2\beta} L_t^2. \tag{4.7}$$

To proceed, let us consider the product measure $\lambda^d \otimes P$ with the corresponding norm $|| \cdot ||_{\lambda^d \otimes P} = [\int\int | \cdot |^2 d(\lambda^d \otimes P)]^{1/2}$. We have

$$||\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t)||_{\lambda^d \otimes P} \leq A h^\beta L_t + \frac{c_t}{(nh^d)^{1/2}}||K|| \tag{4.8}$$

by (4.5), (4.7) and the triangle inequality for $|| \cdot ||_{\lambda^d \otimes P}$.

Let the bandwidth h develop with n as $h(n) = \alpha n^{-\frac{1}{2\beta+d}}$ for some $\alpha > 0$. We have $h^\beta = \alpha^\beta n^{-\frac{\beta}{2\beta+d}}$. Further, $(nh^d)^{-1} = n^{-1}\alpha^{-d}n^{\frac{d}{2\beta+d}} = \alpha^{-d}n^{-\frac{2\beta}{2\beta+d}}$ and therefore $(nh^d)^{-1/2} = \alpha^{-d/2}n^{-\frac{\beta}{2\beta+d}}$. Inequality (4.8) then reads as

$$\begin{aligned}
||\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t)||_{\lambda^d \otimes P} &\leq AL_t\alpha^\beta n^{-\frac{\beta}{2\beta+d}} + c_t\alpha^{-d/2}n^{-\frac{\beta}{2\beta+d}}||K|| \\
&\leq (AL_t\alpha^\beta + c_t\alpha^{-d/2}||K||) \cdot n^{-\frac{\beta}{2\beta+d}}.
\end{aligned}$$

Squaring to obtain the MISE we get

$$\mathbb{E}\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t \leq (AL_t\alpha^\beta + c_t\alpha^{-d/2}||K||)^2 \cdot n^{-\frac{2\beta}{2\beta+d}}$$

or in the more compact form

$$\mathbb{E}\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t \leq C_t^2 \cdot n^{-\frac{2\beta}{2\beta+d}}$$

for $C_t = AL_t\alpha^\beta + c_t\alpha^{-d/2}||K||$. □

Let us discuss the theorem.

1) First of all, the theorem is proved without any assumption on the i.i.d. character of samples (particles) constituting the empirical measures π_t^n. This is the crucial observation, as we know that due to the resampling step the generated particles are not i.i.d.

2) Convergence. For $t \in \mathbb{N}$ fixed, we immediately see from (4.2) that the MISE of kernel estimates goes to zero as the number of particles increases and the bandwidth decreases accordingly, i.e., $\lim_{n\to\infty} \mathbb{E}\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t = 0$.

3) Consistency. The theorem proposes that the bandwidth develops with the number of particles n as $h(n) = \alpha n^{-\frac{1}{2\beta+d}}$ for some $\alpha > 0, \beta, d \in \mathbb{N}$. Obviously, $\lim_{n\to\infty} h(n) = 0$, and $\lim_{n\to\infty} nh(n) = \lim_{n\to\infty} \alpha n^{\frac{2\beta+d-1}{2\beta+d}} = \infty$.

4) The dimension matters. We have $n^{-\frac{2\beta}{2\beta+d_1}} < n^{-\frac{2\beta}{2\beta+d_2}}$ for $d_1 < d_2$, and therefore we must increase the number of particles in order to assure a given accuracy as the dimension increases.

5) The order helps. Contrary to the previous result, we have $n^{-\frac{2\beta_1}{2\beta_1+d}} > n^{-\frac{2\beta_2}{2\beta_2+d}}$ for $\beta_1 < \beta_2$. Hence the greater is the order of the employed kernel, the tighter is the bound on the related MISE, in fact, it tends towards n^{-1}. There are techniques available for constructing kernels of arbitrary orders [7], however, the order of the employed kernel is primarily driven by the Sobolev character of the filtering densities.

6) The theorem assumes that the filtering densities p_t are β-Sobolev for some $L_t > 0, t \in \{0, \ldots, T\}$, $T \in \mathbb{N}$ and $\beta \in \mathbb{N}$ being constant over time. It is the question when this assumption holds. In Section 5, we show that the Sobolev character of the filtering densities is retained over time, if a certain condition holds on the transition kernels of the signal process.

7) For $\alpha = 1$, the specification of C_t simplifies to $C_t = AL_t + c_t||K||$ and C_t consists of four terms. Two of them, A and $||K|| = [\int K^2(\boldsymbol{u})\,d\boldsymbol{u}]^{1/2}$ are the constants determined by the employed kernel. The other two, L_t and c_t, develop with time. The L_t term is discussed in Section 5.

8) The c_t constant (with respect to the number of particles) comes from Theorem 2. It can be shown that its values can be computed recursively as $c_t = c_{t-1}\left(1 + \frac{4||g_t^v||_\infty}{\overline{\pi}_t g_t}\right), c_0 = 1$. The integral $\overline{\pi}_t g_t$ depends on the values of the observation process and c_t generally develops exponentially with time, see the remark in concluding Section 7.

4.2 Extension to partial derivatives

The result of Theorem 6 can be straightforwardly extended to the convergence of partial derivatives of kernel density estimates to partial derivatives of the filtering densities. The proof of the assertion substantially overlaps with the proof of Theorem 6, however, we present it here in full detail for the convenience of the reader.

In what follows we denote by $p_{t,i_1,\ldots,i_d}^{(m)} = \partial^m p_t/\partial x_1^{i_1}\ldots\partial x_d^{i_d}$ the m-th partial derivative of the filtering density p_t, $t \in \{0,\ldots,T\}$, $T \in \mathbb{N}$ for $i_1,\ldots,i_d \in \mathbb{N}_0$, such that $m = i_1 + \cdots + i_d$, and $m \in \mathbb{N}_0$. Similarly, we will use $\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)} = \partial^m \hat{p}_t^n/\partial x_1^{i_1}\ldots\partial x_d^{i_d}$ for the partial derivative of kernel estimate (4.1), $p_{t,i_1,\ldots,i_d}^{*,(m)}$ for the partial derivative of the convolution $p_t^* = p_t * (h^{-d}K(\cdot/h))$ and $K_{i_1,\ldots,i_d}^{(m)} = \partial^m K/\partial u_1^{i_1}\ldots\partial u_d^{i_d}$ for the partial derivative of the kernel employed in the estimates. Clearly, the zero value of i_j, $j = 1,\ldots,d$, corresponds to the situation when no differentiation is applied in the respective dimension.

Theorem 7. *In the filtering problem, let $\{\pi_t\}_{t=0}^T$, $\{p_{t,i_1,\ldots,i_d}^{(m)}\}_{t=0}^T$, $T \in \mathbb{N}$, $m \in \mathbb{N}_0$ be the sequences of filtering distributions and m-th partial derivatives of corresponding filtering densities for some $i_1,\ldots,i_d \in \mathbb{N}_0$, $m = i_1 + \cdots + i_d$. Let $p_{t,i_1,\ldots,i_d}^{(m)}$, $t \in \{0,\ldots,T\}$ satisfy (3.16) for some $\beta \in \mathbb{N}$ and $L_{t,(m)} > 0$. Let $\{\pi_t^n\}_{t=1}^T$, $\{\hat{p}_{t,i_1,\ldots,i_d}^{(m)}\}_{t=1}^T$, $n \in \mathbb{N}$ be the sequences of the empirical measures generated by the particle filter and m-th partial derivatives of the related kernel density estimates (4.1) with the bandwidth varying as $h(n) = \alpha n^{-\frac{1}{2\beta+d+2m}}$ for some $\alpha > 0$. Let the kernel K employed in the estimates be of order β. Then we have the following evolution of the MISE of the m-th partial derivatives of kernel estimates $\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}$ over time $t \in \{1,\ldots,T\}$:*

$$\mathbb{E}\left[\int (\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{(m)}(\boldsymbol{x}_t))^2\,d\boldsymbol{x}_t\right] \leq C_{t,(m)}^2 \cdot n^{-\frac{2\beta}{2\beta+d+2m}}, \tag{4.9}$$

where

$$C_{t,(m)} = AL_{t,(m)}\alpha^\beta + c_t\alpha^{-(d/2+m)}||K_{i_1,\ldots,i_d}^{(m)}||. \tag{4.10}$$

In (4.10), A is the constant of Theorem 4, c_t, $t \in \{1,\ldots,T\}$ are the constants of Theorem 2 and $||K_{i_1,\ldots,i_d}^{(m)}||$ is the L_2 norm of the corresponding m-th partial derivative of kernel K.

Proof. To start remind that for any function $f : \mathbb{R}^d \to \mathbb{R}$ and its m-th partial derivative $f_{i_1,\ldots,i_d}^{(m)} : \mathbb{R}^d \to \mathbb{R}$, both assumed in $L_1(\mathbb{R}^d)$, one has for their Fourier transforms $\mathcal{F}[p](\boldsymbol{\omega})$ and $\mathcal{F}[p_{i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega})$, respectively, the equality

$$\mathcal{F}[f_{i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega}) = (-\mathrm{i})^m (\omega_{i_1}^{i_1} \cdot \cdots \cdot \omega_{i_d}^{i_d})\mathcal{F}[f](\boldsymbol{\omega}). \tag{4.11}$$

Now, in order to prove the theorem, we just mimic the proof of Theorem 6. Employing the complex

exponential in (4.4) and the equality (4.11) we have

$$\mathbb{E}[|\psi_t^n(\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2] \leq \frac{c_t^2}{n},$$

$$|(-\mathrm{i})^m(\omega_1^{i_1} \cdots \cdot \omega_d^{i_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \mathbb{E}[|\psi_t^n(\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2]$$
$$\leq |(-\mathrm{i})^m(\omega_1^{i_1} \cdots \cdot \omega_d^{i_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n},$$

$$\mathbb{E}\left[|(-\mathrm{i})^m(\omega_1^{i_1} \cdots \cdot \omega_d^{i_d})(\psi_t^n(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega})) - (-\mathrm{i})^m(\omega_1^{i_1} \cdots \cdot \omega_d^{i_d})(\psi_t(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega}))|^2\right]$$
$$\leq |(-\mathrm{i})^m(\omega_1^{i_1} \cdots \cdot \omega_d^{i_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n},$$

$$\mathbb{E}\left[\int |\mathcal{F}[\partial^m \hat{p}_t^n / \partial x_1^{i_1} \ldots \partial x_d^{i_d}](\boldsymbol{\omega}) - \mathcal{F}[\partial^m p_t^* / \partial x_1^{i_1} \ldots \partial x_d^{i_d}](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}\right]$$
$$\leq \frac{c_t^2}{n}\int |\mathcal{F}[\partial^m h^{-d} K(\boldsymbol{x}/h) / \partial x_1^{i_1} \ldots \partial x_d^{i_d}]|^2 \, d\boldsymbol{\omega},$$

$$\mathbb{E}\left[\int |\mathcal{F}[\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t)](\boldsymbol{\omega}) - \mathcal{F}[p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t)](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}\right]$$
$$\leq \frac{c_t^2}{n}\int |\mathcal{F}[\partial^m h^{-d} K(\boldsymbol{x}/h) / \partial x_1^{i_1} \ldots \partial x_d^{i_d}]|^2 \, d\boldsymbol{\omega},$$

$$\mathbb{E}\left[\int (\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right]$$
$$\leq \frac{c_t^2}{nh^{2d}}\int (\partial^m K(\boldsymbol{x}/h) / \partial x_1^{i_1} \ldots \partial x_d^{i_d})^2 \, d\boldsymbol{x},$$

$$\mathbb{E}\left[\int (\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right]$$
$$\leq \frac{c_t^2}{nh^{d+2m}}\int (\partial^m K(\boldsymbol{u}) / \partial u_1^{i_1} \ldots \partial u_d^{i_d})^2 \, d\boldsymbol{u},$$

$$\mathbb{E}\left[\int (\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right]$$
$$\leq \frac{c_t^2}{nh^{d+2m}}\int (K_{i_1,\ldots,i_d}^{(m)}(\boldsymbol{u}))^2 \, d\boldsymbol{u}.$$

Using the L_2 norm of $K_{i_1,\ldots,i_d}^{(m)}(\boldsymbol{u})$, i.e., $||K_{i_1,\ldots,i_d}^{(m)}||^2 = \int (K_{i_1,\ldots,i_d}^{(m)}(\boldsymbol{u}))^2 \, d\boldsymbol{u}$, the above reads as

$$\mathbb{E}\int (\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t \leq \frac{c_t^2}{nh^{d+2m}} ||K_{i_1,\ldots,i_d}^{(m)}||^2. \tag{4.12}$$

For given $i_1,\ldots i_d \in \mathbb{N}_0$, we assume that $p_{t,i_1,\ldots,i_d}^{(m)}$, $t \in \{0,\ldots,T\}$ exist and are $\beta \in \mathbb{N}$ Sobolev in the sense of validity of (3.16). That is, for the Fourier transforms $\mathcal{F}[p_{t,i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega})$ there exist positive constants $L_{t,(m)}$ such that

$$\int ||\boldsymbol{\omega}||^{2\beta} |\mathcal{F}[p_{t,i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq (2\pi)^d L_{t,(m)}^2. \tag{4.13}$$

Using (4.13) we have under the assumptions of Theorem 4 the formula

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\mathcal{F}[p_{t,i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq (2\pi)^d A^2 h^{2\beta} L_{t,(m)}^2. \tag{4.14}$$

By (4.14) we get the counterpart of (4.6) that writes as

$$\int (p_{t,i_1,\ldots,i_d}^{*,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t$$

$$= \frac{1}{(2\pi)^d} \int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\mathcal{F}[p_{t,i_1,\ldots,i_d}^{(m)}](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}$$

$$\leq A^2 h^{2\beta} L_{t,(m)}^2.$$

We proceed in the same way as in the proof of Theorem 6. We consider the $||\cdot||_{\lambda^d \otimes P}$ norm and employ the triangle inequality to get

$$||\hat{p}_{t,i_1,\ldots,i_d}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_d}^{(m)}(\boldsymbol{x}_t)||_{\lambda^d \otimes P} \leq A h^\beta L_{t,(m)} + \frac{c_t}{(nh^{d+2m})^{-1/2}} ||K_{i_1,\ldots,i_d}^{(m)}||. \qquad (4.15)$$

The bandwidth h develop with n as $h(n) = \alpha n^{-\frac{1}{2\beta+d+2m}}$ for some $\alpha > 0$. So we have $h^\beta = \alpha^\beta n^{-\frac{\beta}{2\beta+d+2m}}$. Further, $(nh^{d+2m})^{-1} = n^{-1}\alpha^{-(d+2m)}n^{\frac{d+2m}{2\beta+d+2m}} = \alpha^{-(d+2m)}n^{-\frac{2\beta}{2\beta+d+2m}}$ and therefore $(nh^{d+2m})^{-1/2} = \alpha^{-(d+2m)/2}n^{-\frac{\beta}{2\beta+d+2m}}$. This gives us after squaring (4.15) the statement of the theorem:

$$\mathbb{E} \int (\hat{p}_{t,i_1,\ldots,i_m}^{n,(m)}(\boldsymbol{x}_t) - p_{t,i_1,\ldots,i_m}^{(m)}(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t \leq C_{t,(m)}^2 \cdot n^{-\frac{2\beta}{2\beta+d+2m}} \qquad (4.16)$$

for $C_{t,(m)} = AL_{t,(m)}\alpha^\beta + c_t\alpha^{-(d+2m)/2}||K_{i_1,\ldots,i_d}^{(m)}||$. $\qquad\qquad\square$

The structure of formula (4.16) is the same as that of formula (4.2) of Theorem 6. Only two constants are replaced. Therefore, the discussion of its corollaries remains valid, especially, it implies the convergence of partial derivatives of the kernel density estimates to the respective derivatives of the related filtering densities.

On the other hand, we see that the order of the partial derivative m slows down the convergence. In fact, it has the same effect on the convergence as the dimension d, see the discussion concerning the influence of the dimension below Theorem 6.

5 Sobolev character of filtering densities

In Theorem 6, we have assumed that the filtering densities p_t, $t \in \{0,\ldots,T\}$, $T \in \mathbb{N}$ are β-Sobolev over time. This assumption can be verified for p_0, but for other time instants $t > 0$ a direct verification is typically impossible. That is why we are interested in a practical tool for performing the verification indirectly so that the assumptions for the convergence result of Theorem 6 were fulfilled. As a result, we present a sufficient condition on the densities of transition kernels of the signal process such that the Sobolev character of the filtering densities is retained over time.

In the statement below, we work with the prediction and update formulas, (2.6) and (2.7), respectively, of Section 2.3. We rewrite these formulas in the more compact form using the following shortcuts: $\overline{p}_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$, $p_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ (in fact, this shortcut was already used in Theorem 6) and $g_t(\boldsymbol{x}_t) = g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)$ for the respective densities; and $\overline{\pi}_t g_t = \int g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) \, d\boldsymbol{x}_t = \int g_t(\boldsymbol{x}_t)\overline{p}_t(\boldsymbol{x}_t) \, d\boldsymbol{x}_t$ for the normalizing integral. Using the introduced shortcuts we have (2.6) and (2.7) written as

$$\overline{p}_t(\boldsymbol{x}_t) = \int K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p_{t-1}(\boldsymbol{x}_{t-1}) \, d\boldsymbol{x}_{t-1}, \qquad (5.1)$$

$$p_t(\boldsymbol{x}_t) = \frac{g_t(\boldsymbol{x}_t)\overline{p}_t(\boldsymbol{x}_t)}{\overline{\pi}_t g_t}. \qquad (5.2)$$

Definition 3. *Let K_{t-1} be the transition kernel in the filtering problem for time $t-1$, $t-1 \in \mathbb{N}_0$. As the conditional characteristic function $\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})$ of the transition kernel K_{t-1} we denote the characteristic function of the conditional distribution determined by this kernel, i.e.,*

$$\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1}) = \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle} K_{t-1}(d\boldsymbol{x}_t|\boldsymbol{x}_{t-1}).$$

18

Theorem 8. *In the filtering problem, let $p_0 \in \mathcal{P}_{S(\beta, L_0)}$. Let K_{t-1}, $t \in \mathbb{N}$ be the set of the transition kernels and $\mathcal{F}[K_{t-1}]$, $t \in \mathbb{N}$ be the set of the corresponding conditional characteristic functions. For all $t \in \mathbb{N}$, let $\mathcal{F}[K_{t-1}]$ be bounded by a function $K_b \colon \mathbb{R}^d \to \mathbb{C}$ in such a way that for any $\boldsymbol{x}_{t-1} \in \mathbb{R}^d$ and $\boldsymbol{\omega} \in \mathbb{R}^d$*

$$|\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})| \leq |K_b(\boldsymbol{\omega})|. \tag{5.3}$$

Let the function K_b satisfy (3.16) for some $\beta \in \mathbb{N}$ and $L_{K_b} > 0$. Then the filtering densities p_t are β-Sobolev for all $t \in \mathbb{N}$, i.e., $p_t \in \mathcal{P}_{S(\beta, L_t)}$, with the recurrence for L_t written as

$$L_t = ||g_t^v||_\infty L_{K_b}/\overline{\pi}_t g_t, \tag{5.4}$$

where $||g_t^v||_\infty = \sup_{\boldsymbol{u}}\{|g_t^v(\boldsymbol{u})|\}$.

Proof. The theorem holds for p_0 by the assumption. Let $t \in \mathbb{N}$, then by multiplying both sides of (5.1) by the complex exponential we get from the prediction formula

$$e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}\,\overline{p}_t(\boldsymbol{x}_t) = e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}\int K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}.$$

By integration, the left-hand side gives the characteristic function $\overline{\psi}_t(\boldsymbol{\omega})$ of $\overline{p}_t(\boldsymbol{x}_t)$, i.e.,

$$\overline{\psi}_t(\boldsymbol{\omega}) = \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}\overline{p}_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t.$$

The right-hand side has then form

$$\int\int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\,d\boldsymbol{x}_t$$

$$= \int p_{t-1}(\boldsymbol{x}_{t-1})\left(\int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_t\right)d\boldsymbol{x}_{t-1},$$

$$= \int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}.$$

The equality of two complex numbers is equivalent to the equality of their complex conjugates. Hence we can multiply both sides by their complex conjugates with the equality retained. This gives us the expression

$$|\overline{\psi}_t(\boldsymbol{\omega})|^2 = \left|\int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right|^2.$$

By the Jensen's inequality and assumed boundedness of $\mathcal{F}[K_{t-1}]$, we have

$$|\overline{\psi}_t(\boldsymbol{\omega})|^2 \leq \left(\int |\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})|\,p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2$$

$$\leq \left(|K_b(\boldsymbol{\omega})|\int p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2 = |K_b(\boldsymbol{\omega})|^2.$$

Thus,

$$\int ||\boldsymbol{\omega}||^{2\beta}|\overline{\psi}_t(\boldsymbol{\omega})|^2\,d\boldsymbol{\omega} \leq \int ||\boldsymbol{\omega}||^{2\beta}|K_b(\boldsymbol{\omega})|^2 \leq (2\pi)^d L_{K_b}^2. \tag{5.5}$$

The above formula shows that $\overline{p}_t \in \mathcal{P}_{(\beta, L_{K_b})}$ for any $t \in \mathbb{N}$. We proceed with the specification of the Sobolev constant L_t of the update (filtering) density p_t.

In Section 2.2, in formula (2.3), there was shown that the function $g_t(\boldsymbol{x}_t)$ of the update formula (5.2) has form $g_t(\boldsymbol{x}_t) = g_t^v(\boldsymbol{y}_t - h(\boldsymbol{x}_t))$. Function g_t^v is the density of the noise term in the observation process and is assumed to be bounded. Thus, we have $\sup_{\boldsymbol{x}_t,\boldsymbol{y}_t}\{|g_t^v(\boldsymbol{y}_t - h(\boldsymbol{x}_t))|\} = \sup_{\boldsymbol{u}}\{|g_t^v(\boldsymbol{u})|\} = ||g_t^v||_\infty < \infty.$

Again, multiplying the update formula (5.2) by the complex exponential, integrating and multiplying by the respective conjugates gives us

$$(\overline{\pi}_t g_t)\, p_t(\boldsymbol{x}_t) = g_t(\boldsymbol{x}_t)\, \overline{p}_t(\boldsymbol{x}_t),$$

$$(\overline{\pi}_t g_t)\!\int\! e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle} p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t = \int e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle} g_t(\boldsymbol{x}_t)\, \overline{p}_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t,$$

$$(\overline{\pi}_t g_t)^2 |\psi_t(\boldsymbol{\omega})|^2 \leq ||g_t^v||_\infty^2 |\overline{\psi}_t(\boldsymbol{\omega})|^2,$$

$$||\boldsymbol{\omega}||^{2\beta} |\psi_t(\boldsymbol{\omega})|^2 \leq \frac{||g_t^v||_\infty^2}{(\overline{\pi}_t g_t)^2}\, ||\boldsymbol{\omega}||^{2\beta} |\overline{\psi}_t(\boldsymbol{\omega})|^2,$$

$$(2\pi)^{-d}\!\int\! ||\boldsymbol{\omega}||^{2\beta} |\psi_t(\boldsymbol{\omega})|^2\, d\boldsymbol{\omega} \leq \frac{||g_t^v||_\infty^2 L_{K_b}^2}{(\overline{\pi}_t g_t)^2} = L_t^2.$$

This concludes the proof. □

The theorem tells us that, in the particle filter, the β-Sobolev character of the filtering densities is retained over time if the set of the conditional characteristic functions of transition kernels $\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})$, $t \in \mathbb{N}$ is uniformly bounded.

5.1 Extension to partial derivatives

Considering preservation of the Sobolev character of partial derivatives (in the sense of validity of (3.16)) of the filtering densities $p_{t,i_1,\ldots,i_m}^{(m)}$, the theorem holds as well. The difference is that we assume that $p_{0,i_1,\ldots,i_m}^{(m)}$ is β-Sobolev[2] and, in (5.3), instead of considering boundedness of $\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})$, we consider the boundedness of $\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})$ for any $\boldsymbol{x}_{t-1} \in \mathbb{R}^d$, $t \in \mathbb{N}$.

Theorem 9. *In the filtering problem, let $p_{0,i_1,\ldots,i_m}^{(m)}$ be β-Sobolev for some $\beta \in \mathbb{N}$, $L_{0,(m)>0}$ and $i_1,\ldots,i_d \in \mathbb{N}_0$ such that $m = i_1 + \cdots + i_d$, $m \in \mathbb{N}_0$. Let K_{t-1}, $t \in \mathbb{N}$ be the set of the transition kernels, and $K_{t-1,i_1,\ldots,i_m}^{(m)} = \partial^m K_{t-1}/\partial x_1^{i_1}\ldots\partial x_d^{i_d}$, $t \in \mathbb{N}$ the set of corresponding partial derivatives. Let $\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})$, $t \in \mathbb{N}$ be the set of the corresponding conditional Fourier transforms, i.e.,*

$$\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1}) = \int e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle} K_{t-1,i_1,\ldots,i_m}^{(m)}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, d\boldsymbol{x}_t.$$

For all $t \in \mathbb{N}$, let $\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}]$ be bounded by some function $K_{b,(m)}\colon \mathbb{R}^d \to \mathbb{C}$ in such a way that for any $\boldsymbol{x}_{t-1} \in \mathbb{R}^d$ and $\boldsymbol{\omega} \in \mathbb{R}^d$,

$$|\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})| \leq |K_{b,(m)}(\boldsymbol{\omega})|. \tag{5.6}$$

Let the function $K_{b,(m)}$ satisfy (3.16) for the above $\beta \in \mathbb{N}$ and some $L_{K_{b,(m)}} > 0$. Then the partial derivatives of filtering densities $p_{t,i_1,\ldots,i_m}^{(m)}$, $t \in \mathbb{N}$ are β-Sobolev with the recurrence for L_t written as

$$L_t = ||g_t||_\infty L_{K_{b,(m)}}/\overline{\pi}_t g_t, \tag{5.7}$$

where $||g_t||_\infty = \sup_{\boldsymbol{u}}\{|g_t(\boldsymbol{u})|\}$.

Proof. The theorem holds for $p_{0,i_1,\ldots,i_m}^{(m)}$ by the assumption. From the prediction formula, multiplying both sides of (5.1) by the complex exponential, we get

$$e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle} \overline{p}_t(\boldsymbol{x}_t) = e^{i\langle \boldsymbol{\omega}, \boldsymbol{x}_t\rangle}\!\int\! K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) p_{t-1}(\boldsymbol{x}_{t-1})\, d\boldsymbol{x}_{t-1}.$$

---

[2] Strictly speaking, we cannot say that $p_{0,i_1,\ldots,i_m}^{(m)}$ is β-Sobolev or write $p_{0,i_1,\ldots,i_m}^{(m)} \in \mathcal{P}_{(\beta, L_{0,(m)})}$ as the partial derivative is not a density anymore. But, if we still do it for a general function, then we mean that the Fourier transform of this function exists and satisfies the inequality (3.16) for some $\beta \in \mathbb{N}$ and $L > 0$.

By integration, the left-hand side just gives the characteristic function $\overline{\psi}_t(\boldsymbol{\omega})$ of $\overline{p}_t(\boldsymbol{x}_t)$, i.e.,

$$\overline{\psi}_t(\boldsymbol{\omega}) = \int e^{i\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}\overline{p}_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t.$$

The right-hand side has then form

$$\int\int e^{i\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\,d\boldsymbol{x}_t$$

$$= \int p_{t-1}(\boldsymbol{x}_{t-1})\left(\int e^{i\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_t\right)d\boldsymbol{x}_{t-1},$$

$$= \int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}.$$

Multiplying both sides by $(-i)^m(\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d})$, we move both sides to the Fourier transforms of the corresponding partial derivatives. That is,

$$(-i)^m(\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d})\overline{\psi}_t(\boldsymbol{\omega}) = \mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]$$

and

$$\int p_{t-1}(\boldsymbol{x}_{t-1})(-i)^m(\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d})\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1} =$$

$$= \int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}.$$

Further multiplying both sides by the complex conjugates gives the expression

$$|\mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]|^2 = \left|\int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right|^2.$$

Now, by the assumed boundedness of $\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}]$ and the Jensen's inequality, we have

$$|\mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]|^2 \leq \left(\int|\mathcal{F}[K_{t-1,i_1,\ldots,i_m}^{(m)}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})|\,p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2$$

$$\leq \left(|K_{b,(m)}(\boldsymbol{\omega})|\int p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2 = |K_{b,(m)}(\boldsymbol{\omega})|^2.$$

Thus,

$$\int||\boldsymbol{\omega}||^{2\beta}|\mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]|^2\,d\boldsymbol{\omega} \leq \int||\boldsymbol{\omega}||^{2\beta}|K_b^{(m)}(\boldsymbol{\omega})|^2 \leq (2\pi)^d L_{K_{b,(m)}}^2. \tag{5.8}$$

The above formula shows that $\overline{p}_{t,i_1,\ldots,i_m}^{(m)} \in \mathcal{P}_{(\beta,L_{K_{b,(m)}})}$ for any $t \in \mathbb{N}$. We proceed with the specification of the Sobolev constant $L_{t,(m)}$ of the partial derivative $p_{t,i_1,\ldots,i_m}^{(m)}$.

Similarly as in the proof of Theorem 8, we have $\sup_{\boldsymbol{x}_t,\boldsymbol{y}_t}\{|g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)|\} = ||g_t||_\infty < \infty$. Further, multiplying the update formula (5.2) by the complex exponential, integrating, multiplying by $(-i)^m(\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d})$ and the respective conjugates we shift to the Fourier transforms of partial derivatives and get

$$\begin{aligned}
(\overline{\pi}_t g_t)\,p_t(\boldsymbol{x}_t) &= g_t(\boldsymbol{x}_t)\,\overline{p}_t(\boldsymbol{x}_t),\\
(\overline{\pi}_t g_t)\int e^{i\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}p_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t &= \int e^{i\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle}g_t(\boldsymbol{x}_t)\,\overline{p}_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t,\\
(\overline{\pi}_t g_t)\,\psi_t(\boldsymbol{\omega}) &\leq ||g_t||_\infty^2\,\overline{\psi}_t(\boldsymbol{\omega}),\\
(\overline{\pi}_t g_t)\,(-i)^m\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d}\,\psi_t(\boldsymbol{\omega}) &\leq ||g_t||_\infty^2\,(-i)^m\omega_{i_1}^{i_1}\cdot\ldots\cdot\omega_{i_d}^{i_d}\,\overline{\psi}_t(\boldsymbol{\omega}),\\
(\overline{\pi}_t g_t)^2\,|\mathcal{F}[p_{t,i_1,\ldots,i_m}^{(m)}]|^2 &\leq ||g_t||_\infty^2\,|\mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]|^2,\\
||\boldsymbol{\omega}||^{2\beta}\,|\mathcal{F}[p_{t,i_1,\ldots,i_m}^{(m)}]|^2 &\leq \frac{||g_t||_\infty^2}{(\overline{\pi}_t g_t)^2}\,||\boldsymbol{\omega}||^{2\beta}|\mathcal{F}[\overline{p}_{t,i_1,\ldots,i_m}^{(m)}]|^2,\\
(2\pi)^{-d}\int||\boldsymbol{\omega}||^{2\beta}|\mathcal{F}[p_{t,i_1,\ldots,i_m}^{(m)}]|^2\,d\boldsymbol{\omega} &\leq \frac{||g_t||_\infty^2 L_{K_{b,(m)}}^2}{(\overline{\pi}_t g_t)^2} = L_{t,(m)}^2.
\end{aligned}$$

21

This concludes the proof. □

6 Example

In this section, we demonstrate an application of the presented theory. Because our research has not been driven by any concrete application, we apply the particle filtering and kernel density estimation methodologies on the filtering problem for a multivariate Gaussian process. This problem has the analytical solution - the well-known Kalman filter [16, 8, 17, 9].

The purpose of this choice is to check if empirical results from computer simulations follow the analytic counterpart. By replacing the Gaussian transition kernel and Gaussian observation density by general entities we can build up the appropriate particle filter for a general Markov process, but without the possibility of checking against the analytical solution.

6.1 Multivariate Gaussian process

Let the signal and observation processes introduced in Section 2.1 be specified as multivariate Gaussian. That is, we assume that the formulas driving evolution of states and observations are specified, for a general dimension $d \geq 1$, as

$$\boldsymbol{X}_t = \mathbf{F}\boldsymbol{X}_{t-1} + \boldsymbol{W}_t, \quad \boldsymbol{Y}_t = \mathbf{H}\boldsymbol{X}_t + \boldsymbol{V}_t, \quad t \in \mathbb{N}, \tag{6.1}$$

where $\mathbf{F}$, $\mathbf{H}$ are $d \times d$ regular matrices and $W_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $V_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ are multivariate normal noise terms with $d \times d$ covariance matrices $\mathbf{Q}$ and $\mathbf{R}$. The signal process $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ forms a multivariate Markov chain with Gaussian transition kernels. The initial distribution is considered also multivariate normal, i.e., $\boldsymbol{X}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\mu}_0 \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_0$ is a $d \times d$ covariance matrix.

Mathematically, the filtering task is to find the conditional expected values $\mathbb{E}[\boldsymbol{X}_t|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t]$ for $t \geq 1$. At the given time instant $t \in \mathbb{N}$, the conditional expected value is the integral characteristic of the related conditional distribution which represents the filtering distribution we are interested in.

The vector $(\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{Y}_1, \ldots, \boldsymbol{X}_t, \boldsymbol{Y}_t)$ is multivariate normal because it is determined by a linear transformation of the vector $(\boldsymbol{X}_0, \boldsymbol{W}_1, \boldsymbol{V}_1, ..., \boldsymbol{W}_t, \boldsymbol{V}_t)$ which is multivariate normal. Therefore, the filtering distribution is also multivariate normal, and is determined by its mean vector $\boldsymbol{\mu}_t$ and its covariance matrix $\boldsymbol{\Sigma}_t$ at time $t \in \mathbb{N}$. The preservation of the normal character of the filtering distribution over time allows us to obtain an analytic expression for its parameters. The result is known as the *multivariate Kalman filter*.

6.2 Multivariate Kalman filter

The theoretical analysis presented in [17] gives the following recursive Kalman's equations for $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. The parameters are computed in several steps using some auxiliary variables for $t \geq 1$:

$$\begin{aligned}
\boldsymbol{\mu}_{t|t-1} &= \mathbf{F}\boldsymbol{\mu}_{t-1}, \\
\boldsymbol{\Sigma}_{t|t-1} &= \mathbf{F}\boldsymbol{\Sigma}_{t-1}\mathbf{F}^T + \mathbf{Q}, \\
\mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{H}^T[\mathbf{H}\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}^T + \mathbf{R}]^{-1}, \\
\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t[\boldsymbol{Y}_t - \mathbf{H}\boldsymbol{\mu}_{t|t-1}], \\
\boldsymbol{\Sigma}_t &= [\mathbf{I}_d - \mathbf{K}_t\mathbf{H}]\boldsymbol{\Sigma}_{t|t-1}.
\end{aligned}$$

Using the above formulas, one can recursively compute the determining parameters of the filtering distribution over time. Due to the normal character of the distribution, we have apparently $\mathbb{E}[\boldsymbol{X}_t|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t] = \boldsymbol{\mu}_t$. Further, the formula for the evolution of the covariance matrix $\boldsymbol{\Sigma}_t$ is deterministic. That is, it is not affected by observations.

6.3 Multivariate Gaussian particle filter

The incorporation of schema (6.1) into the particle filter's computation, presented in Section 2.4, stems from the specification of the initial density $p_0(\boldsymbol{x}_0)$ and the set of transition kernels K_{t-1}, $t \in \mathbb{N}$.

As already mentioned, the initial density is multivariate normal with some mean $\boldsymbol{\mu}_0 \in \mathbb{R}^d$ and a $d \times d$ covariance matrix $\boldsymbol{\Sigma}_0$, i.e.,

$$p_0(\boldsymbol{x}_0) = (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_0 - \boldsymbol{\mu})\right].$$

The densities of Gaussian transition kernels $K_{t-1}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$, $t \in \mathbb{N}$ are specified as

$$K_{t-1}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) = (2\pi)^{-\frac{d}{2}} |\mathbf{Q}|^{-\frac{1}{2}} \exp\left[-\boldsymbol{u}_t^T \mathbf{Q}^{-1} \boldsymbol{u}_t\right] \tag{6.2}$$

with $\boldsymbol{u}_t = \boldsymbol{x}_t - \mathbf{F}\boldsymbol{x}_{t-1}$.

The above formula reflects the multivariate normal character of the noise term $\boldsymbol{W}_t$ in (6.1) and, in fact, corresponds to the specification of the density of the multivariate normal distribution $\mathcal{N}(\mathbf{F}\boldsymbol{x}_{t-1}, \mathbf{Q})$.

The Sobolev character of the filtering densities is given by the Sobolev character of the Gaussian transition kernels. We show that the conditional characteristic functions of the Gaussian kernels (6.2) are uniformly bounded, which implies the Sobolev character according to Theorem 8.

We have $\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1}) = \mathcal{F}[\mathcal{N}(\mathbf{F}\boldsymbol{x}_{t-1}, \mathbf{Q})]$, and therefore

$$\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1}) = e^{\mathrm{i}\langle \boldsymbol{\omega}, \mathbf{F}\boldsymbol{x}_{t-1}\rangle} \exp\left[-\frac{1}{2}\boldsymbol{\omega}^T \mathbf{Q}\,\boldsymbol{\omega}\right].$$

Further,

$$
\begin{aligned}
|\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})| &\leq \exp\left[-\frac{1}{2}\boldsymbol{\omega}^T \mathbf{Q}\,\boldsymbol{\omega}\right] \\
&\leq \exp\left[-\frac{1}{2}\lambda_{\min}||\boldsymbol{\omega}||^2\right] = K_b(\boldsymbol{\omega}),
\end{aligned}
$$

where $\lambda_{\min}$ is the minimal eigenvalue of the covariance matrix $\mathbf{Q}$.

For the Sobolev constant L_{K_b} of $K_b(\boldsymbol{\omega})$ and $\beta = 1$, we have the integral

$$(2\pi)^{-d} \int ||\boldsymbol{\omega}||^2 \exp\left[-\lambda_{\min}||\boldsymbol{\omega}||^2\right] = \frac{\pi^{-\frac{d}{2}}}{4^d(\sqrt{\lambda_{\min}})^{d+2}}.$$

From this result we also see that any multivariate normal initial distribution with the covariance matrix $\boldsymbol{\Sigma}_0$ is 1-Sobolev with the constant $L_0 = \pi^{-\frac{d}{2}}/[4^d(\sqrt{\lambda_{\min}^0})^{d+2}]$, where $\lambda_{\min}^0$ is the minimal eigenvalue of $\boldsymbol{\Sigma}_0$.

The obtained result on the Sobolev character of the filtering densities is consistent with the fact that all densities in the multivariate Gaussian process (6.1) are normal, i.e., the character of the involved densities does not change during operation of the filter.

6.4 Multivariate Gaussian convolution kernel

In the multivariate Gaussian particle filter, kernel density estimates are made using the multivariate standard normal (convolution) kernel

$$K(\boldsymbol{u}) = (2\pi)^{-\frac{d}{2}} \exp\left[-\frac{1}{2}||\boldsymbol{u}||^2\right].$$

The specification of the L_2 norm of the kernel is straightforward. We have

$$||K||^2 = (2\pi)^{-d} \int \exp(-||\boldsymbol{u}||^2)\, d\boldsymbol{u} = (4\pi)^{-\frac{d}{2}},$$

hence $||K|| = (4\pi)^{-\frac{d}{4}}$.

Concerning the A constant of Theorem 4, we start with the Fourier transform of the multivariate standard normal kernel which corresponds to the characteristic function of the $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ distribution. That is, $K_{\mathcal{F}}(\boldsymbol{\omega}) = e^{-\boldsymbol{\omega}^T \mathbf{I}_d \boldsymbol{\omega}} = e^{-\frac{1}{2}||\boldsymbol{\omega}||^2}$. In order to specify some constant A, we need to determine a bound on the spectral matrix norm of the Hessian of $K_{\mathcal{F}}$. The entries of the Hessian matrix $\mathcal{H}(K_{\mathcal{F}})$ reads as

$$\frac{\partial^2 K_{\mathcal{F}}}{\partial \omega_j^2} = (\omega_j^2 - 1)K_{\mathcal{F}}(\boldsymbol{\omega}), \quad \frac{\partial K_{\mathcal{F}}}{\partial \omega_j \partial \omega_k} = K_{\mathcal{F}}(\boldsymbol{\omega})\omega_j\omega_k, \quad j \neq k.$$

In the matrix notation, the Hessian writes as $\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\omega}) = K_{\mathcal{F}}(\boldsymbol{\omega})(\boldsymbol{\omega}\boldsymbol{\omega}^T - \mathbf{I}_d)$. Using the spectral matrix norm we get

$$
\begin{aligned}
||\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\omega})||_{spc} &\leq K_{\mathcal{F}}(\boldsymbol{\omega})||\boldsymbol{\omega}\boldsymbol{\omega}^T - \mathbf{I}_d||_{spc} \\
&\leq K_{\mathcal{F}}(\boldsymbol{\omega})(||\boldsymbol{\omega}\boldsymbol{\omega}^T||_{spc} + ||\mathbf{I}_d||_{spc}) \\
&\leq K_{\mathcal{F}}(\boldsymbol{\omega})(||\boldsymbol{\omega}^T||||\boldsymbol{\omega}|| + 1) \\
&\leq K_{\mathcal{F}}(\boldsymbol{\omega})(||\boldsymbol{\omega}||^2 + 1).
\end{aligned}
$$

Note that for a vector $\boldsymbol{\omega} \in \mathbb{R}^d$, $||\boldsymbol{\omega}||_{spc} = ||\boldsymbol{\omega}||$ (the standard Euclidean norm). Let $\boldsymbol{\omega} = \boldsymbol{\xi}$ such that $||\boldsymbol{\xi}|| \leq 1$. Then we clearly have $||\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\xi})||_{spc} \leq 2$ as $K_{\mathcal{F}}(\boldsymbol{\xi}) \leq 1$.

The multidimensional Taylor's theorem for $K_{\mathcal{F}}$ writes as

$$K_{\mathcal{F}}(\boldsymbol{\omega}) = K_{\mathcal{F}}(\mathbf{0}) + (\nabla K_{\mathcal{F}}(\mathbf{0}))\boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\omega}^T[\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\xi})]\boldsymbol{\omega}$$

for a suitable $\boldsymbol{\xi} \in \mathbb{R}^d$, $||\boldsymbol{\xi}|| \leq ||\boldsymbol{\omega}||$. For the gradient, we have $\nabla K_{\mathcal{F}}(\mathbf{0}) = \mathbf{0}$ and $K_{\mathcal{F}}(\mathbf{0}) = 1$, therefore the above Taylor's theorem gives for any $||\boldsymbol{\omega}|| \leq 1$,

$$
\begin{aligned}
K_{\mathcal{F}}(\boldsymbol{\omega}) - 1 &= \frac{1}{2}\boldsymbol{\omega}^T[\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\xi})]\boldsymbol{\omega}, \\
|K_{\mathcal{F}}(\boldsymbol{\omega}) - 1| &\leq \frac{1}{2}||\boldsymbol{\omega}^T|| \cdot ||\mathcal{H}(K_{\mathcal{F}})(\boldsymbol{\xi})||_{spc} \cdot ||\boldsymbol{\omega}||, \\
\frac{|K_{\mathcal{F}}(\boldsymbol{\omega}) - 1|}{||\boldsymbol{\omega}||} &\leq ||\boldsymbol{\omega}^T|| = ||\boldsymbol{\omega}||.
\end{aligned}
$$

Further $|K_{\mathcal{F}}(\boldsymbol{\omega}) - 1| \leq 1$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$ and therefore $|K_{\mathcal{F}}(\boldsymbol{\omega}) - 1|/||\boldsymbol{\omega}|| \leq 1$ for $||\boldsymbol{\omega}|| > 1$. Thus, joining the two inequalities we finally get

$$\frac{|K_{\mathcal{F}}(\boldsymbol{\omega}) - 1|}{||\boldsymbol{\omega}||} \leq \max\{1, 1\} = 1, \quad \boldsymbol{\omega} \in \mathbb{R}^d \backslash \{\mathbf{0}\},$$

and the A constant equals to 1, i.e., $A = 1$.

The above considerations immediately lead to the specification of the order of the multivariate standard normal kernel. As mentioned, the Fourier transform of the kernel is $K_{\mathcal{F}}(\boldsymbol{\omega}) = e^{-\frac{1}{2}||\boldsymbol{\omega}||^2}$ and $K_{\mathcal{F}}(\mathbf{0}) = \mathbf{1}$. The related gradient writes as $\nabla K_{\mathcal{F}}(\boldsymbol{\omega}) = -e^{-\frac{1}{2}||\boldsymbol{\omega}||^2}\boldsymbol{\omega}$, thus $\nabla K_{\mathcal{F}}(\mathbf{0}) = \mathbf{0}$. For the Hessian of $K_{\mathcal{F}}(\boldsymbol{\omega})$, we have $\text{diag}(\mathcal{H}(K_{\mathcal{F}})(\mathbf{0})) = -\mathbf{1}$. Hence the order of the kernel is $\ell = \beta = 1$.

6.5 MATLAB implementation and experiments

In this section we introduce our implementation of the multivariate Kalman filter and its particle filter counterpart to show results of several experiments.

We have implemented both filters in the form of a MATLAB function. The inputs into the function are $\mathbf{F}, \mathbf{Q}, \mathbf{H}, \mathbf{R}$ matrices of formula (6.1), the computational horizon $T \in \mathbb{N}$ and the selected number of particles $n \in \mathbb{N}$. The outputs are the means and covariance matrices from the particle and Kalman filters, respectively. If the dimension of the signal process is $d = 1$ or $d = 2$, then the script provides a graphical output illustrating the estimated density and its theoretical counterpart from the Kalman filter. The source code of the function is presented in Appendix 1.

We have performed several experiments in order to check if the computational behavior of the multivariate Gaussian particle filter coincides with the analytical results. The experiments were performed for the following setting of parameters: $\mathbf{F} = \mathbf{I}_d$, $\mathbf{Q} = 2\mathbf{I}_d$, $\mathbf{H} = 2\mathbf{I}_d$, $\mathbf{R} = \mathbf{I}_d$. In the script, the density of the multivariate standard normal distribution is used as the initial density. Computational horizon was set to $T = 100$.

The results of three $d = 2$ experiments for different numbers of particles $n = 10, 100$ and $n = 1000$ are presented in Table 6.1. Graphically, the obtained kernel density estimate and theoretical filtering density are presented in Fig. 6.1 for $n = 100$.

$T{=}100$	$\hat{\boldsymbol{\mu}}_T$	$\boldsymbol{\mu}_T$	$\widehat{\boldsymbol{\Sigma}}_T$ - PF		$\boldsymbol{\Sigma}_T$ - KF	
$n{=}10$	32.25	31.92	0.1472	0.0992	0.2247	0
	-18.43	-18.65	0.0992	0.4290	0	0.2247
$n{=}100$	0.46	0.48	0.1557	-0.0212	0.2247	0
	-2.16	-2.04	-0.0212	0.2144	0	0.2247
$n{=}1000$	-2.76	-2.75	0.2207	-0.0036	0.2247	0
	-29.18	-29.18	-0.0036	0.2206	0	0.2247

Table 6.1: Comparison of bivariate particle and Kalman filters.

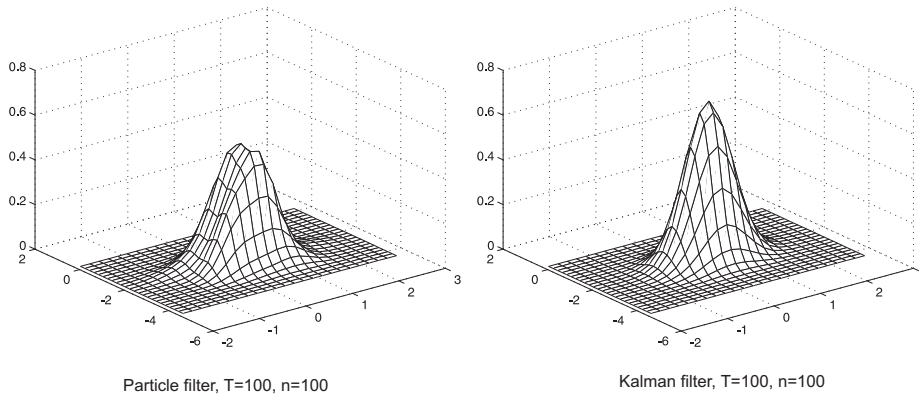Particle filter, T=100, n=100   Kalman filter, T=100, n=100

Figure 6.1: The kernel density estimate generated by the bivariate Gaussian particle filter and the corresponding filtering density from the Kalman filter.

On the basis of the inspection of the numerical results presented in Table 6.1, we can state a good agreement of numerical characteristics delivered by the Gaussian particle filter with the theoretical characteristics of the filtering distributions.

7 Conclusion

In the paper, we have demonstrated that the standard methodology of kernel density estimates can be applied in the area of particle filtering. We have proved that the kernel density estimates constructed on the basis of particles generated by the particle filter converge in the MISE to the theoretical filtering density at each time instant of operation of the filter. The result holds even though the generated particles do not constitute an i.i.d. sample from the filtering distribution. Moreover, we have stated the sufficient condition for the preservation of the Sobolev character of the filtering densities over time. The extension of both results to the partial derivatives of the kernel estimates and filtering densities has been provided as well.

In Theorem 2, the constant c_t is known that it typically grows exponentially with time, see e.g., [1] p. 87, therefore C_t of (4.3) does so; and, if one wants to assure the given precision of the density approximation, then one must increase the number of generated particles exponentially, too. This

is an unpleasant property of the particle filter. On the other hand, there are results available, e.g., [18] or [19], that under additional conditions, uniformly convergent particle filters can be constructed. That is, that c_t of (2.9) is constant over time.

The constant C_t depends on L_t. Under the conditions of Theorem 8, we know the evolution of L_t over time. In fact, the evolution is somehow similar to the evolution of c_t constant and there is again the risk of an exponential growth of L_t. The study of the conditions when L_t evolves uniformly over time is the issue of the future research in this field.

1   MATLAB implementation

```
function [PFm,PFcov,KFm,KFcov] = mvpf(F,Q,H,R,T,n);

%---HMM---
d=size(Q,1);
m0=zeros(d,1);S0=eye(d);
X0=mvnrnd(m0',S0)';
X=zeros(d,T);Y=X;
X(:,1)=F*X0+mvnrnd(zeros(1,d),Q)';
Y(:,1)=H*X(:,1)+mvnrnd(zeros(1,d),R)';
for t=2:T,
 W=mvnrnd(zeros(1,d),Q)';
 V=mvnrnd(zeros(1,d),R)';
 X(:,t)=F*X(:,t-1)+W;
 Y(:,t)=H*X(:,t)+V;
end;

%---Kalman filter---
M=zeros(d,T);
m=m0;S=S0;
for t=1:T,
 m1=F*m;
 S1=F*S*F'+Q;
 K=S1*H'*inv(H*S1*H'+R);
 m=m1+K*(Y(:,t)-H*m1);
 S=(eye(d)-K*H)*S1;
 M(:,t)=m;
end;
KFm=M(:,T)
KFcov=S;

%---PF filter---
P=mvnrnd(m0',S0,n)';
for t=1:T,
 disp(t);
 pp=zeros(d,n);w=zeros(1,n);
 for j=1:n;
  pp(:,j)=F*P(:,j)+mvnrnd(zeros(1,d),Q)';
  w(j)=mvnpdf((Y(:,t)-H*pp(:,j))',zeros(1,d),R);
 end;
 if n>1, wn=w/sum(w); else wn=1; end;
 mn=randsample(n,n,true,wn);
 P=pp(:,mn);
end;
PT=P;
PFm=mean(PT')';
PFcov=cov(PT');

%---kernel estimate for d=1 with graphical output---
```

```
if d==1,
 alpha=1;beta=1;
 hn=alpha*n^(-1/(2*beta+1));
 xx=[KFm-5*sqrt(KFcov):0.1:KFm+5*sqrt(KFcov)];
 fx=zeros(1,length(xx));
 for j=1:n,
  fx=fx+1/(n*hn)*1/sqrt(2*pi)*exp(-(xx-PT(j)).^2/(2*hn^2));
 end;
 plot(xx,fx,'b',xx,normpdf(xx,KFm,sqrt(KFcov)),'r');
 figure(1);
end;

%---kernel estimate for d=2 with graphical output---
if d==2,
 alpha=1;beta=1;
 hn=alpha*n^(-1/(2*beta+d));
 x1=[KFm(1)-5*sqrt(KFcov(1,1)):0.2:KFm(1)+5*sqrt(KFcov(1,1))];
 x2=[KFm(2)-5*sqrt(KFcov(2,2)):0.2:KFm(2)+5*sqrt(KFcov(2,2))];
 [X1,X2]=meshgrid(x1,x2);
 Xr=[X1(:) X2(:)];
 nXr=size(Xr,1);fxr=zeros(nXr,1);
 for j=1:n,
  mvn=mvnpdf((Xr-ones(nXr,1)*PT(:,j)')/hn,zeros(1,d),eye(d));
  fxr=fxr+1/(n*hn^d)*mvn;
 end;
 colormap([0 0 0]);
 mesh(X1,X2,reshape(fxr,length(x2),length(x1)));
 figure(1);
 pause;
 pr=mvnpdf(Xr,KFm',KFcov);
 mesh(X1,X2,reshape(pr,length(x2),length(x1)));
 figure(1);
end;
```

Acknowledgment

Bibliography

[1] A. Doucet, N. de Freitas, and N. G. (Eds.), *Sequential Monte Carlo Methods in Practice*. New York, LLC: Springer-Verlag, 2001.

[2] A. Doucet and A. M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in *The Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovskii, Eds. Oxford University Press, 2011.

[3] D. Crisan and A. Doucet, "A Survey of Convergence Results on Particle Filtering Methods for Practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.

[4] F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 144–187, 2004.

[5] M. Hürzeler and H. R. Künsch, "Monte Carlo Approximations for General State-Space Models," *Journal of Computational and Graphical Statistics*, vol. 7, no. 2, pp. 175–193, 1998.

[6] D. Crisan and J. Míguez, "Particle Approximation of the Filtering Density for State-Space Markov Models in Discrete Time," *arXiv:1111.5866*, pp. 1–42, 2013.

[7] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer, 2009.

[8] B. Fristedt, N. Jain, and N. Krylov, *Filtering and Prediction: A Primer*. American Mathematical Society, 2007.

[9] S. Sarkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[10] R. L. Schilling, *Measures, Integrals and Martingales*. Cambridge University Press, 2005.

[11] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, New York: Chapman and Hall/CRC, 1986.

[12] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, New York: Chapman and Hall/CRC, 1995.

[13] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, Inc., 1992.

[14] E. Parzen, "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[15] G. H. Givens, "Consistency of the local kernel density estimator," *Statistics & Probability Letters*, vol. 25, pp. 55–61, 1995.

[16] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[17] D. S. G. Pollock, *Handbook of Time Series Analysis, Signal Processing, and Dynamics*. Academic Press, 1999.

[18] P. D. Morral and A. Guionnet, "On the Stability of Interacting Processes with Applications to Filtering and Genetic Algorithms," *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, vol. 37, no. 2, pp. 155–194, 2001.

[19] K. Heine and D. Crisan, "Uniform Approximations of Discrete-Time Filters," *Advances in Applied Probability*, vol. 40, no. 4, pp. 979–1001, 2008.