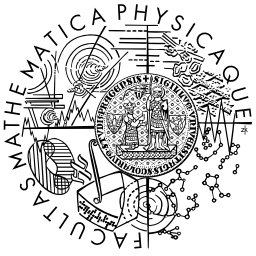

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

David Coufal

Kernel Methods in Particle Filtering

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis:	prof. RNDr. Viktor Beneš, DrSc.
Study programme:	Mathematics
Study branch:	Probability and Mathematical Statistics

Prague 2018

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague, 28th June 2018

I am grateful to my supervisor, Dr. Viktor Beneš, for providing me with an interesting topic of study and his helpful comments. I am also grateful to the Institute of Computer Science of Academy of Sciences of the Czech Republic for technical support during writing the thesis. Finally, special thanks go to Angélique, marquise du Plessis-Bellieré for her inspiring comment on lifelong education.

Title: Kernel Methods in Particle Filtering

Author: **David Coufal**

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Viktor Beneš, DrSc., Department of Probability and Mathematical Statistics

Abstract: The thesis deals with the use of kernel density estimates in particle filtering. In particular, it examines the convergence of the kernel density estimates to the filtering densities. The estimates are constructed on the basis of an output from particle filtering. It is proved theoretically that using the standard kernel density estimation methodology is effective in the context of particle filtering, although particle filtering does not produce random samples from the filtering densities. The main theoretical results are: 1) specification of the upper bounds on the MISE error of the estimates of the filtering densities and their partial derivatives; 2) specification of the related lower bounds and 3) providing a suitable tool for checking persistence of the Sobolev character of the filtering densities over time. In addition, the thesis also focuses on designing kernels suitable for practical use.

Keywords: particle filtering, kernel density estimates of filtering densities, upper and lower bounds

Contents

List of Symbols

- $\mathbb{N}, \mathbb{N}_0, \mathbb{Z}, \mathbb{R}, \mathbb{C}$ - sets of natural, $\mathbb{N} \cup \{0\}$, integer, real and complex numbers
- $\mathbb{R}^d$ - d-dimensional Euclidean space, $\mathbb{R}^d = \mathbb{R} \times \cdots \times \mathbb{R}$
- $\mathcal{B}(\mathbb{R}^d)$ - σ-algebra of Borel subsets of $\mathbb{R}^d$
- $B(\mathbb{R}^d), B(\mathbb{C}^d)$ - set of bounded real functions over $\mathbb{R}^d$, $\mathbb{C}^d$
- $C_b(\mathbb{R}^d), C_b(\mathbb{C}^d)$ - set of continuous and bounded functions over $\mathbb{R}^d$, $\mathbb{C}^d$
- $L^1(\mathbb{R}^d, \mu), L^2(\mathbb{R}^d, \mu)$ - $L^{p=1,2}$ integrable functions w.r.t. measure μ on $\mathbb{R}^d$
- $||\boldsymbol{x}||_2 = (\sum_{i=1}^d x_i^2)^{1/2}$ - Euclidean norm of the vector $\boldsymbol{x} \in \mathbb{R}^d$
- $\boldsymbol{x} - s$, $\boldsymbol{x}/h$ - $(x_1 - s, \ldots, x_d - s)$, $(x_1/h, \ldots, x_d/h)$, $\boldsymbol{x} \in \mathbb{R}^d$, $h, s \in \mathbb{R}, h \neq 0$
- $||f||_2 = (\int |f(\boldsymbol{x})|^2 \, d\boldsymbol{x}))^{1/2}$ - L^2 norm of the function $f : \mathbb{R}^d \to \mathbb{R}$
- $||f||_\infty = \sup_{\boldsymbol{x}} |f(\boldsymbol{x})|$ - supremum norm of the function $f : \mathbb{R}^d \to \mathbb{R}$
- $\delta_{\boldsymbol{x}_i}(d\boldsymbol{x})$ - Dirac measure located at $\boldsymbol{x}_i \in \mathbb{R}^d$
- μf - image of the function f over the measure μ, $\mu f = \int f \mu(d\boldsymbol{x})$
- $K_{t-1}(A \in \mathcal{B}(\mathbb{R}^d), \boldsymbol{x}_{t-1})$ - transition kernels of the signal process for $t \in \mathbb{N}$
- $K_{t-1}(\boldsymbol{x}_t, \boldsymbol{x}_{t-1})$ - densities of K_{t-1}
- $g_t(\boldsymbol{y}_t | \boldsymbol{x}_t)$ - conditional densities of the observation process
- $\hat{f}, \hat{p}$ - kernel density estimates of densities f and p
- $|\cdot|$ - absolute value of a real or complex number
- $|\alpha|$, $\alpha!$ - for the multi-index $\alpha \in \mathbb{N}_0^d$, $|\alpha| = \sum_{i=1}^d |\alpha_i|$, $\alpha! = \alpha_1! \cdots \alpha_d!$
- D^α - differential operator for $\alpha \in \mathbb{N}_0^d$, i.e., $D^\alpha = \partial^{|\alpha|} / (\partial^{\alpha_1} \ldots \partial^{\alpha_d})$
- $\mathcal{F}[\cdot], \mathscr{H}_\nu[\cdot]$ - Fourier and Hankel transform of order ν, resp.
- $\mathcal{P}^\infty_{S(\beta,L)}$, $\mathcal{P}_{S(\beta,L)}$ - the classes of β-Sobolev functions and densities
- Γ, B - Gamma and Beta functions
- $(\cdot)_n = \Gamma(\cdot + n)/\Gamma(\cdot)$ - the Pochhammer symbol
- $(\cdot)_+ = \max\{0, \cdot\}$ - positive part

1. Introduction

Particle filtering enables its users to efficiently compute integral characteristics (moments) of distributions of interest. In the area of filtering and its applications, these distributions are traditionally referred to as the *filtering distributions*. In particle filtering, a filtering distribution is approximated by an empirical measure. This measure is constructed 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* [Doucet et al., 2001, Doucet and Johansen, 2011, Crisan and Rozovskii, 2011].

The theoretical result that justifies the application of particle filtering is convergence of the generated empirical measures to the theoretical filtering distribution as the number of used particles goes to infinity [Crisan and Doucet, 2002, Doucet et al., 2001]. Approximating the filtering distribution by an empirical measure is beneficial for estimating moments of the distribution because they correspond to weighted sums of the values of moment functions over generated particles.

The filtering distribution has typically a density with respect to the corresponding (in the sense of the dimension) Lebesgue measure. This density is called the *filtering density*. Knowing an analytical approximation of the filtering density has advantages. For example, the possibility of computing analytical approximations of densities of the related conditional distributions. 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. Various authors have addressed this topic, such as [Musso et al., 2001, Le Gland and Oudjane, 2004, Hürzeler and Künsch, 1998, Künsch, 2005] and recently [Crisan and Míguez, 2014].

1.1 State of the art

In [Musso et al., 2001], the authors refer to their previous works in which they introduced the particle filters that employ kernel estimates at different places in their computational schemes. The filters are called the pre-regularized and post-regularized particle filter, respectively, and differ in where exactly the

kernel density estimate is applied in the classical particle filtering algorithm. They further introduce the local rejection regularized particle filter (L2RPF filter) and show that it generalizes the post-regularized particle filter and the KF filter introduced in [Hürzeler and Künsch, 1998]. The convergence analysis of the post-regularized filter is presented in details in [Le Gland and Oudjane, 2004].

In [Künsch, 2005], the author investigates configurations if the acceptance-rejection method and importance sampling with an additional resampling step are used in particle filtering. The author shows that rejection sampling has a smaller asymptotic variance than the standard importance sampling resampling method. However, the computational effort for rejection sampling is generally greater than for the importance sampling. Whilst the author provides several convergence results for a kernel estimate to converge to the corresponding filtering density in terms of convergence in probability and a version of the central limit theorem, the assumptions of Theorem 2 in [Künsch, 2005, p. 2006], which applies to the importance sampling resampling method, exclude the common filtering settings that consider an additive Gaussian noise.

A summary discussion of the above papers is also presented in Section 3.1 of [Crisan and Míguez, 2014]. In fact, the paper [Crisan and Míguez, 2014] is the closest to our work as it addresses the application of kernel density estimation in particle filtering in a very similar way to what we do. However, our work, which is mainly inspired by Tsybakov's book [Tsybakov, 2009], builds on Fourier analysis of kernel density estimates. This fact enables us to obtain a stronger version of certain results presented in Crisan and Míguez [2014], which is discussed in details in the last chapter, Chapter 5, of the thesis.

1.2 Structure of thesis and main results

The thesis comprises of five chapters. The first introductory chapter is followed by Chapter 2 that recalls notation, concepts and theorems we need in order to present our work. Namely, we review essentials of the orthogonal polynomials, the Fourier and Hankel integral transforms, basics of particle filtering with special emphasis put on the related convergence results and the analysis of kernel density estimates in the frequency domain.

Chapter 3 delivers the main theoretical results. We present the upper bounds on the mean integrated squared error (MISE) of the kernel density estimates of the filtering densities and its partial derivatives. The bounds first show that the standard kernel estimation methodology is applicable in particle filtering as the MISE error converge to zero with an increasing number of generated particles. Second, they provide estimates on the convergence rate in dependance on the number of particles, the dimensionality of the signal process and smoothness of the estimated filtering density. In its basic version, the result has been published in [Coufal, 2016].

We are interested in lower bounds too. We extend the approach presented in Chapter 2 of [Tsybakov, 2009] to the multivariate design. The approach, rooted in information theory, investigates the minimax lower bounds on estimation error if a set of deliberately constructed probability measures is considered. It turns out that the standard kernel density estimates are efficient estimators because the lower bounds meet the upper ones. The result together with the extension of the upper bounds on partial derivatives has been submitted for publication in [Coufal, 2018a].

Finally, we address the filtering density's smoothness assumption in the introduced convergence theorem. There is assumed that the filtering density has certain character, namely the Sobolev one. This assumption can be checked directly for the initial density of the signal process, however, this does not hold for other operation times as we do not have an explicit representation of the filtering density at our disposal. The purpose of the third part is to provide a handy tool for checking persistence of the Sobolev character of the filtering density over time, and, as the consequence, to ensure that the kernel estimates converge through whole operation time of the filter. This result has been published in [Coufal, 2018b].

Chapter 4 deals with designing kernels to be practically used for estimating the filtering densities as theoretically introduced in Chapter 3. The convergence theorem assumes that the kernels of the specific order are used to construct the filtering density estimates. The chapter focuses on designing both univariate and multivariate kernels of the given order. The presented approach draws on using the orthogonal polynomials to construct the kernels with required properties.

Chapter 5 delivers a discussion on originality of the presented work in the context of the results reported in [Crisan and Míguez, 2014].

1.3 Notation and typography

The mathematical symbols used throughout the thesis are summarized in the related list. Special symbols and notation are defined and mentioned before they are used for the first time.

Within the thesis, *italic* is used to emphasize new or important concepts.

Definitions, lemmas and theorems are typeset in italic as well. The ends of proofs are denoted by the $\square$ symbol.

The thesis was typeset in LaTeX2$_\varepsilon$, MiKTeX 2.9 distribution, using standard Computern Modern Fonts with several enhancements, especially $\mathcal{AMS}$-LaTeX mathematical symbols.

2. Preliminaries

This chapter reviews notions, definitions and theorems we need in the sequel. Providing references to original sources would make the thesis lighter, however, it could lead to inconvenience due to the inconsistency in notation. So we put all the material here to have the thesis self-referencing and make a reader as comfortable as possible.

2.1 Orthogonal polynomials

This section reviews the orthogonal polynomials. Whilst the basic facts are known for the classical families such as Hermite or Jacobi family (with the special cases of Gegenbauer, Chebyshev and Legendre polynomials), this is generally not the case for the polynomials in multiple dimensions. Below we recall some essential facts about the selected families, certain interrelations and the construction of the orthogonal polynomials on the unit disc and unit ball in $\mathbb{R}^d$. The main source of information provided here are the comprehensive books [Dunkl and Xu, 2014], [Abramowitz and Stegun, 1964] and [Weisstein, 2002].

2.1.1 Univariate orthogonal polynomials

Let μ be a non-negative Borel measure on $\mathbb{R}$ with an infinite support. The sequence of univariate polynomials $\{P_n(x) \in L^2(\mathbb{R}, \mu)\}_{n=0}^{\infty}$ is called orthogonal with respect to the measure μ if for each $n, m \in \mathbb{N}_0$, $P_n(x)$ is a polynomial of degree n and

$$\int P_n(x) P_m(x)\, d\mu = h_n \delta_{nm}$$

with h_n being a normalization constant and δ_{nm} the Kronecker symbol. The sequence is called orthonormal, if $h_n = 1$, $n \in \mathbb{N}_0$. It is known that if $\int |x|^n \, d\mu < \infty$ for all $n \in \mathbb{N}_0$, then the sequence $\{P_n(x) \in L^2(\mathbb{R}, \mu)\}_{n=0}^{\infty}$ can be constructed explicitly by the standard Gram-Schmidt orthogonalization process applied on the monomial sequence $\{1, x, x^2, \dots\}$. Since the Gram-Schmidt process has the property that the first n orthogonal vectors spans the same space as the original first n vectors, i.e., that $P_0, \dots, P_n$ span the same space as $1, x, \dots, x^n$, $P_0, \dots, P_n$ form a basis for polynomials of degree n or less.

Depending on specification of the orthogonal measure μ, various families of the orthogonal polynomials are introduced. Below we cover some of the classical ones. Namely, Hermite, Legendre, Gegenbauer and Jacobi families. For each of these families the orthogonality measure is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}$ and has the form $d\mu(x) = w(x)\,dx$, with a positive weight function $w(x)$ on some interval and normalization constant $c = [\int w(x)\,dx]^{-1}$.

2.1.2 Hermite polynomials

The (physicists') Hermite polynomials $\{H_n(x)\}_{n=0}^{\infty}$ can be generated by progressive differentiation of the unscaled Gaussian function e^{-x^2},

$$H_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}, \qquad n \in \mathbb{N}_0.$$

The first five polynomials read as

$$\begin{aligned}
&H_0(x) = 1,\\
&H_1(x) = 2x,\\
&H_2(x) = 4x^2 - 2,\\
&H_3(x) = 8x^3 - 12x,\\
&H_4(x) = 16x^4 + 8x^2 + 12.
\end{aligned}$$

The Hermite polynomials are orthogonal with respect to the weight function $w(x) = e^{-x^2}$ for $x \in \mathbb{R}$, $c = \pi^{-1/2}$ and therefore

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2}\,dx = h_n \delta_{nm}, \qquad n, m \in \mathbb{N}_0$$

with $h_n = \sqrt{\pi} 2^n n!$.

2.1.3 Legendre polynomials

The Legendre polynomials $\{P_n(x)\}_{n=0}^{\infty}$ can be generated according to the so-called Rodrigues' formula

$$P_n(x) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} (x^2 - 1)^n.$$

The first five polynomials reads as

$$
\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= \frac{1}{2}(3x^2-1),\\
P_3(x) &= \frac{1}{2}(5x^3-3x),\\
P_4(x) &= \frac{1}{8}(35x^4-30x^2+3).
\end{aligned}
$$

The weight function corresponds to the indicator of the $[-1,1]$ interval, i.e., $w(x) = 1_{[-1,1]}(x)$, $c = 1/2$ and

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx = h_n\delta_{nm}, \qquad n,m \in \mathbb{N}_0$$

with $h_n = 2/(2n+1)$.

2.1.4 Gegenbauer polynomials

The Gegenbauer polynomials $\{C_n^{(\lambda)}(x)\}_{n=0}^{\infty}$ are also called the ultraspherical polynomial. For a parameter $\lambda > -1/2$, the related weight function writes $w(x) = (1-x^2)^{\lambda-1/2}$ on the interval $[-1,1]$ with the constant $c = B(\frac{1}{2}, \lambda+\frac{1}{2})^{-1}$ where B is the Beta function. The case $\lambda = 1/2$ corresponds to the Legendre polynomials, i.e.,

$$C_n^{(1/2)}(x) = P_n(x).$$

The generating formula for the Gegenbauer polynomials writes

$$C_n^{(\lambda)}(x) = \frac{(-1)^n}{2^n\,n!}\frac{\Gamma(2\lambda+n)\Gamma(\lambda+\frac{1}{2})}{\Gamma(2\lambda)\Gamma(\lambda+\frac{1}{2}+n)}(1-x^2)^{\frac{1}{2}-\lambda}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(1-x^2)^{\lambda+n-\frac{1}{2}}.$$

The first four polynomials reads as

$$
\begin{aligned}
C_0^{(\lambda)}(x) &= 1,\\
C_1^{(\lambda)}(x) &= 2\lambda x,\\
C_2^{(\lambda)}(x) &= 2\lambda(\lambda+1)x^2-\lambda,\\
C_3^{(\lambda)}(x) &= \frac{4}{3}\lambda(\lambda+1)(\lambda+2)x^3-2\lambda(\lambda+1)x.
\end{aligned}
$$

Finally, for $\lambda > -1/2$ orthogonality writes

$$\int_{-1}^{1} C_n^{(\lambda)}(x) C_m^{(\lambda)}(x)(1-x^2)^{\lambda-1/2}\,dx = h_n \delta_{nm} \qquad n, m \in \mathbb{N}_0$$

with $h_n = (\pi 2^{1-2\lambda}\Gamma(n+2\lambda))/(n!(n+\lambda)[\Gamma(\lambda)]^2)$.

2.1.5 Jacobi polynomials

The Jacobi polynomials $\{P_n^{(\alpha,\beta)}(x)\}_{n=0}^{\infty}$ are the most general family of the classical orthogonal polynomials. They are orthogonal with respect to the weight function $w(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha, \beta > -1$ on the interval $[-1, 1]$ with constant $c = 2^{-\alpha-\beta-1}B(\alpha+1, \beta+1)^{-1}$. They generalize (not only) these families of the orthogonal polynomials:

- Gegenbauer polynomials

$$C_n^{(\lambda)}(x) = \frac{\Gamma(\lambda+\frac{1}{2})\Gamma(n+2\lambda)}{\Gamma(n+\lambda+\frac{1}{2})\Gamma(2\lambda)}\, P_n^{(\lambda-\frac{1}{2},\lambda-\frac{1}{2})}(x) = \frac{(2\lambda)_n}{(\lambda+\frac{1}{2})_n}\, P_n^{(\lambda-\frac{1}{2},\lambda-\frac{1}{2})}(x),$$

- Legendre polynomials $P_n(x) = C_n^{(1/2)}(x) = P_n^{(0,0)}(x)$.
- Zernike polynomials (see below) for $0 \le m \le n$, $n-m$ even

$$R_n^m(\rho) = (-1)^{(n-m)/2}\rho^m P_{(n-m)/2}^{(m,0)}(1-2\rho^2).$$

An explicit representation of the Jacobi polynomials involves the hypergeometric function. As we are only interested in the above special cases, of the many interrelations valid for the the Jacobi polynomials [Abramowitz and Stegun, 1964, Sec. 22.5], we only mention the symmetry, evaluation at $x = 1$ and a specific relation to the Gegenbuer polynomials. It holds

$$P_n^{(\alpha,\beta)}(-x) = (-1)^n P_n^{(\beta,\alpha)}(x), \qquad P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n} \tag{2.1}$$

and

$$P_n^{(\alpha,-\frac{1}{2})}(x) = \frac{\left(\frac{1}{2}\right)_n}{(\alpha+\frac{1}{2})_n} C_{2n}^{(\alpha+\frac{1}{2})}\left(\sqrt{\frac{x+1}{2}}\right). \tag{2.2}$$

The normalizing constant for the Jacobi polynomials reads as [Abramowitz and Stegun, 1964, p. 774],

$$h_n^{(\alpha,\beta)} \;=\; \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)}.$$

2.1.6 Multivariate orthogonal polynomials

Moving to multiple dimensions makes the situation more complicated. The point is that sets of multivariate monomials (multivariate counterparts of the univariate monomial sequence $\{1, x, x^2, \dots\}$) can be ordered in various ways, which leads to different orthogonal sets when the Gram-Schmidt process is used. Moreover, the orthogonality domains, i.e., the supports of the related weight functions, can be much more complicated. Here we survey results which are important from the point of view of Chapter 4. That is, we focus on the multivariate orthogonal polynomials over the unit disc and the unit ball in $\mathbb{R}^d$, $d \geq 3$. Namely, the Zernike polynomials covers the $d = 2$ case and the $d \geq 3$ case is covered by the polynomials rooted in spherical harmonics [Dunkl and Xu, 2014]. As in the previous section, we denote $c = [\int w(\boldsymbol{x})\, d\boldsymbol{x}]^{-1}$.

2.1.7 Zernike polynomials

The Zernike polynomials $\{Z_n^m(\rho, \varphi) : m \in \mathbb{Z}, |m| \leq n\}_{n=0}^{\infty}$ are orthogonal polynomials over the unit disc. They are used extensively in optics in connection with modelling human vision and image analysis. Though a compact representation of Z_n^m exists in complex coordinates, we follow the standard approach of splitting Z_n^m into the even and odd polynomials [Weisstein, 2002, p. 3234]. These are specified as

$$\begin{aligned} Z_n^m(\rho, \varphi) &= R_n^m(\rho)\cos(m\varphi), \quad 0 \leq m \leq n, \\ Z_n^{-m}(\rho, \varphi) &= R_n^m(\rho)\sin(m\varphi), \quad 0 < m \leq n, \end{aligned}$$

respectively, for $n, m \in \mathbb{N}_0, m \leq n,\ \rho \in [0, 1]$, and $\varphi \in [0, 2\pi)$.

In this representation, $R_n^m : [0, 1] \to \mathbb{R}$, $0 \leq m \leq n$ are the radial polynomials which are specified as

$$R_n^m(\rho) = \begin{cases} \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k!(\frac{n+m}{2}-k)!(\frac{n-m}{2}-k)!}\rho^{n-2k} & \text{for } n-m \text{ even}, \\ 0 & \text{for } n-m \text{ odd}. \end{cases}$$

Using the Jacobi polynomials, the even part is compactly written

$$R_n^m(\rho) = (-1)^{(n-m)/2}\rho^m P_{(n-m)/2}^{(m,0)}(1 - 2\rho^2). \tag{2.3}$$

The first non-zero radial polynomial up to the fourth order read as

$$
\begin{aligned}
&R_0^0(\rho) = 1,\\
&R_1^1(\rho) = \rho,\\
&R_0^2(\rho) = 2\rho^2 - 1,\\
&R_2^2(\rho) = \rho^2,\\
&R_1^3(\rho) = 3\rho^3 - 2\rho,\\
&R_3^3(\rho) = \rho^3,\\
&R_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1,\\
&R_4^2(\rho) = 4\rho^4 - 3\rho^2,\\
&R_4^4(\rho) = \rho^4.
\end{aligned}
$$

The weight function for the Zernike polynomials corresponds to the indicator of the unit disc, i.e., $w(\boldsymbol{x}) = 1_{||\boldsymbol{x}||_2 \leq 1}$, $\boldsymbol{x} \in \mathbb{R}^2$ with $c = \pi^{-1}$. In polar coordinates, orthogonality writes

$$\int_0^1 \int_0^{2\pi} Z_n^m(\rho,\varphi) Z_{n'}^{m'}(\rho,\varphi) \rho \, d\varphi \, d\rho = \frac{\epsilon_m \pi}{2(n+1)} \delta_{mm'} \delta_{nn'} \tag{2.4}$$

where $\epsilon_m = 2$ if $m = 0$ and $\epsilon_m = 1$ if $m > 0$.

2.1.8 Orthogonal polynomials on the unit ball

We start with some notation from Section 3.1 of [Dunkl and Xu, 2014]. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$, $d \in \mathbb{N}$ and $\boldsymbol{x} = (x_1, \ldots, x_d)$, a monomial in the variables $x_1, \ldots, x_d$ is the product $\boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. The number $|\alpha| = \alpha_1 + \cdots + \alpha_d$ is called the total degree of the monomial $\boldsymbol{x}^\alpha$. Note that by definition $x^0 \equiv 1$, $x \in \mathbb{R}$, i.e., if $|\alpha| = 0$, then $\boldsymbol{x}^\alpha = 1$, $\boldsymbol{x} \in \mathbb{R}^d$ including the $\boldsymbol{x} = \mathbf{0} = (0, \ldots, 0)$ case; and $\mathbf{0}^\alpha = 0$, if $|\alpha| > 0$.

A polynomial P in d variables is a finite linear combination of monomials. The degree of a polynomial P, $\deg P$, is defined as the highest total degree of its monomials. By Π^d we denote the set of polynomials in d real variables. For $n \in \mathbb{N}_0$, Π_n^d denote the linear space of polynomials in several variables of degree *at most* n. A polynomial is called homogeneous of degree n if all the monomials appearing in it have the same total degree n. We denote the linear space of homogeneous polynomials of degree n in d variables by $\mathcal{P}_n^d$. Every polynomial in Π^d can be written as a linear combination of homogeneous polynomials. Formally written, one has

$$\mathcal{P}_n^d = \operatorname{span}\{\boldsymbol{x}^\alpha : |\alpha| = n, \alpha \in \mathbb{N}_0^d\}, \quad \Pi_n^d = \operatorname{span}\{\boldsymbol{x}^\alpha : |\alpha| \leq n, \alpha \in \mathbb{N}_0^d\}.$$

For $P, Q \in \Pi^d$, denote by $\langle \cdot, \cdot \rangle$ the inner product on Π^d with respect to the weight function W_μ, i.e.,

$$\langle P, Q \rangle = \int_B P(\boldsymbol{x}) Q(\boldsymbol{x}) W_\mu(\boldsymbol{x})\, d\boldsymbol{x}.$$

The polynomials P, Q are said to be orthogonal with respect to the weight function W_μ if $\langle P, Q \rangle = 0$. A polynomial P is called an orthogonal polynomial w.r.t. W_μ if P is orthogonal to all polynomials of lower degree, i.e.,

$$\langle P, Q \rangle = 0, \quad \forall Q \in \Pi^d \quad \text{with} \quad \deg Q \leq \deg P.$$

Let $\mathcal{V}_n^d(W_\mu^B)$ denote the space of orthogonal polynomials w.r.t. W_μ^B of order *exactly* n, that is,

$$\mathcal{V}_n^d(W_\mu^B) = \{P \in \Pi_n^d : \langle P, Q \rangle = 0, \forall Q \in \Pi_{n-1}^d\}.$$

The dimension of $\mathcal{V}_n^d(W_\mu^B)$ is the same as that of $\mathcal{P}_n^d$ so it is natural to use a multi-index to index the elements of an orthogonal basis of $\mathcal{V}_n^d(W_\mu^B)$. The elements of such a basis are denoted $\{P_\alpha : |\alpha| = n\}$ or $\{P_{\alpha(j)}^{d,n} : |\alpha| = n\}_{j=0}^{N^{d,n}}$ where $N^{d,n}$ is cardinality of $\{P_\alpha : |\alpha| = n\}$. Clearly, in the second case some total ordering of multi-indices in $\{|\alpha| = n\}$ is considered. Lexicographic or graded lexicographic orders are discussed in [Dunkl and Xu, 2014, p. 59].

Denote $S_n(W_\mu^B; f)$ the n-th partial sum of the Fourier orthogonal expansion of function $f \in L^2(\mathbb{R}^d, W_\mu^B)$ w.r.t. $\mathcal{V}_n^d(W_\mu^B)$. That is

$$S_n(W; f) = \sum_{k=0}^{n} \sum_{|\alpha|=k} \left[\int f(\boldsymbol{x}) P_\alpha(\boldsymbol{x}) W_\mu^B(\boldsymbol{x})\, d\boldsymbol{x} \right] P_\alpha = \sum_{k=0}^{n} \sum_{j=0}^{N^{d,k}} b_j^k P_{\alpha(j)}^{d,k}.$$

One important property of this operator is that it is a projection operator onto Π_n^d, i.e., $S_n(W; P) = P$ if $P \in \Pi_n^d$ [Dunkl and Xu, 2014, p. 293]. Hence considering a homogeneous monomial $\boldsymbol{x}^\alpha$ of degree n, i.e., if $|\alpha| = n$, it holds that $\boldsymbol{x}^\alpha \in \text{span}\{\mathcal{V}_n^d(W_\mu^B)\}$. We will use this fact in the proof of Lemma 4.1.

To specify the orthogonal polynomials on the unit ball, we mainly follow [Dunkl and Xu, 2014], Sections 4.1 and 5.2. For $d \geq 3$ and every $n \in \mathbb{N}_0$, let $\mathcal{H}_n^d$ be the linear space of harmonic polynomials (homogeneous solutions of the Laplace equation $\Delta P = 0$) of degree n on $\mathbb{R}^d$ [Dunkl and Xu, 2014, Def. 4.1.1, p. 115]. It is known that the dimension of this space is

$$\dim \mathcal{H}_n^d = \binom{n+d-1}{d-1} - \binom{n+d-3}{d-1}.$$

We are interested in constructing a basis of $\mathcal{H}_n^d$.

Let $T_n(u)$ and $U_n(u)$ be the Chebychev polynomials of the first and the second kind, respectively. (We have not covered these polynomials explicitly in the preceeding list, as this is the only place we need them, we ask the reader to go elsewhere for checking their properties.) For $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$, define

$$\begin{aligned} g_{0,n}(x_1, x_2) &= ||\boldsymbol{x}||^n \cdot T_n(x_2/||\boldsymbol{x}||), \\ g_{1,n-1}(x_1, x_2) &= ||\boldsymbol{x}||^{n-1} \cdot U_{n-1}(x_2/||\boldsymbol{x}||). \end{aligned}$$

These are homogeneous polynomials of degree n in two dimensions and the set $\{g_{0,n}, g_{1,n-1}\}$ constitutes a mutually orthogonal basis for $\mathcal{H}_n^2$. For $d \geq 3$ and $\mathbf{n} = (n_1, n_2, \ldots n_d) \in \mathbb{N}_0^d$ with $n_1 = 0$ or 1, define

$$Y(\mathbf{n}) = g_{n_1,n_2}(x_1, x_2) \prod_{j=3}^{d} (x_1^2 + \cdots + x_j^2)^{n_j/2} C_{n_j}^{(\lambda_j)}(x_j (x_1^2 + \cdots + x_j^2)^{-1/2}) \quad (2.5)$$

with

$$\lambda_j = \lambda_j(n_1, \ldots, n_{j-1}) = \sum_{i=1}^{j-1} n_i + \frac{j-2}{2}.$$

Then $Y(\mathbf{n})$ is homogeneneous of degree $|\boldsymbol{n}|$ and $\{Y(\mathbf{n}) : |\mathbf{n}| = n$ with $n_1 = 0$ or $1\}$ is an orthogonal basis of $\mathcal{H}_n^d$. This result, stated in the hyperspherical coordinates, is presented in [Dunkl and Xu, 2014, p. 116, Theorem 4.1.4]. The presented formulation is taken from [Piñar and Xu, 2017].

Let $B_d = \{\boldsymbol{x} \in \mathbb{R}^d : ||\boldsymbol{x}||_2 \leq 1\}$ be the unit ball in $\mathbb{R}^d$. For $\mu > -1/2$, the classical orthogonal polynomials over B_d are defined w.r.t. the weight function $W_\mu^{B_d}$ with the corresponding normalization constant $c_\mu^{B_d}$,

$$W_\mu^{B_d}(\boldsymbol{x}) = (1 - ||\boldsymbol{x}||_2)^{\mu - 1/2}, \quad c_\mu^{B_d} = \left[\int_{B^d} W_\mu^{B_d}(\boldsymbol{x})\, d\boldsymbol{x} \right]^{-1} = \frac{\Gamma(\mu + \frac{d+1}{2})}{\pi^{d/2} \Gamma(\mu + \frac{1}{2})}.$$

The following lemma gives a tool for constructing the orthogonal polynomials over the unit ball in dimension $d = 3$ and higher [Dunkl and Xu, 2014, p. 142, Proposition 5.2.1].

Lemma 2.1. *For $n \in \mathbb{N}_0$, $0 \leq j \in \mathbb{N}_0 \leq n/2$, let $\{Y_\nu^{n-2j} : 1 \leq \nu \leq \dim \mathcal{H}_{n-2j}^d\}$ be an orthonormal basis of $\mathcal{H}_{n-2j}^d$, the polynomials*

$$P_{j,\nu}^n(\boldsymbol{x}) = (h_{j,n}^\mu)^{-1} P_j^{(\mu - \frac{1}{2}, n-2j+\frac{d-2}{2})}(2||\boldsymbol{x}||_2^2 - 1) \cdot Y_\nu^{n-2j}(\boldsymbol{x}) \quad (2.6)$$

form an orthonormal basis of $\mathcal{V}_n^d(W_\mu^B)$; the constant is given as

$$(h_{j,n}^\mu)^2 = (c_\mu^{B_d})^{-1} \frac{(\mu + \frac{1}{2})_j (\frac{d}{2})_{n-j} (n - j + \mu + \frac{d-1}{2})}{j! (\mu + \frac{d+1}{2})_{n-j} (n + \mu + \frac{d-1}{2})}$$

where $(x)_n = \Gamma(x+n)/\Gamma(x)$, $n \geq 0$ is the Pochhammer symbol.

We have adapted the specification of the normalizing constant, because in [Dunkl and Xu, 2014] they have them specified as $h_n = c \int P_n(\boldsymbol{x}) P_n(\boldsymbol{x}) w(\boldsymbol{x})\, d\boldsymbol{x}$ (see Section 1.4 or check the proof of the lemma) and we consider the specification $h_n = \int P_n(\boldsymbol{x}) P_n(\boldsymbol{x}) w(\boldsymbol{x})\, d\boldsymbol{x}$.

In Chapter 4, we are interested in products $P^n_{j,\nu}(\mathbf{0}) P^n_{j,\nu}(\boldsymbol{x})$. Let us show that these products are non-zero only in limited cases. Since Y_ν^{n-2j} is a homogeneous polynomial, one has $Y_\nu^{n-2j}(\mathbf{0}) = 0$, unless its degree is zero, i.e., unless $j = n/2$. Consequently, $P^n_{j,\nu}(\mathbf{0}) = 0$ unless $j = n/2$. Therefore if n is odd, all products $P^n_{j,\nu}(\mathbf{0}) P^n_{j,\nu}(\boldsymbol{x})$ are zero. If n is even, then only the products related to $\mathcal{H}_0^d$ space are non-zero. In what follows, we will consider even n, indexed by j, i.e., $n = 2j$ for $j \in \mathbb{N}_0$.

If $n = 2j$, $j \in \mathbb{N}_0$, then we work with $\mathcal{H}_0^d$ space in Lemma 2.6. $\dim \mathcal{H}_0^d$=1 and $Y(\boldsymbol{n} = \mathbf{0}) = \{Y_1^0(\boldsymbol{x})\}$ from (2.5). Because $C_0^{(\lambda)} = 1$ for any $\lambda > -1/2$, one has $Y_1^0(\boldsymbol{x}) = 1$. Hence, the non-zero $P^{2j}_{j,\nu}(\mathbf{0}) P^{2j}_{j,\nu}(\boldsymbol{x})$ products write

$$P^{2j}_{j,1}(\mathbf{0}) P^{2j}_{j,1}(\boldsymbol{x}) = (h^\mu_{j,2j})^{-2} P_j^{(\mu-\frac{1}{2},\frac{d-2}{2})}(-1) \cdot P_j^{(\mu-\frac{1}{2},\frac{d-2}{2})}(2||\boldsymbol{x}||^2 - 1), \;\; j \in \mathbb{N}_0$$

where on the right-hand side there is the respective Jacobi polynomial.

2.2 Radial and Bessel functions

This section recalls the concept of radial functions which are basically the functions that depend on norms of their arguments. Specifically, the radial functions are invariant to rotations when the Euclidean norm is considered. Bessel functions come in play when dealing with the Fourier transform of the radial functions.

2.2.1 Radial functions

Through the thesis, we will work exclusively with the radial functions using the Euclidean norm on $\mathbb{R}^d$. The formal definition reads as follows.

Definition 2.1. *A function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called radial if there exists a univariate function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\Phi(\boldsymbol{x}) = \varphi(||\boldsymbol{x}||_2)$ where $||\cdot||_2$ is the Euclidean norm on $\mathbb{R}^d$.*

Remark that starting with rotation invariance, i.e., if $g(\boldsymbol{x}) = g(\mathbf{O}\boldsymbol{x})$ for every orthogonal matrix $\mathbf{O}$, then g is radial if and only if there exists a univariate function φ such that the above definition applies, i.e., iff $g(\boldsymbol{x}) = \varphi(||\boldsymbol{x}||_2)$, see [Sasvári, 2013, Lemma 3.6.2. on p. 160].

2.2.2 Bessel functions

Bessel functions originate in solutions of spherical tasks specified in terms of differential equations. There are several types of these functions. We only recall the Bessel functions and the spherical Bessel functions of the first kind J_ν and j_ν, respectively. For the properties of these functions see [Weisstein, 2002, pp. 198 and 2779, resp.] or [Abramowitz and Stegun, 1964, Sec. 9, 10].

Definition 2.2. *For $\nu > -1$ and $x \geq 0$ the Bessel function of the first kind $J_\nu : [0, \infty) \to \mathbb{R}$ is specified by its series expansion as*

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(k+\nu+1)} \left(\frac{x}{2}\right)^{2k}. \tag{2.7}$$

In fact, the above power series converges in the entire complex plane and for $x > 0$ J_ν can be extended recursively on all $\nu \in \mathbb{R}$ using differentiation.

Definition 2.3. *The spherical Bessel function of the first kind $j_\nu : [0, \infty) \to \mathbb{R}$ is derived from J_ν as*

$$j_\nu(x) = \sqrt{\frac{\pi}{2x}} J_{\nu+1/2}(x) = x^\nu \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,(2k+2\nu+1)!!} \left(\frac{x^2}{2}\right)^k. \tag{2.8}$$

Remark that $J_{1/2}(x) = \sqrt{\frac{2}{\pi}} \cdot \sin(x)/x = \sqrt{\frac{2}{\pi}} \cdot \mathrm{sinc}(x)$; and $\lim_{x\to 0^+} x^{-\nu} J_\nu(x) = \frac{1}{2^\nu \Gamma(\nu+1)}$, which can be easily seen from the series expansion. Especially, for $\nu = 1$ one has $\lim_{x\to 0^+} x^{-1} J_1(x) = \frac{1}{2}$.

For the spherical Bessel function of the first kind one gets from the series expansion $\lim_{x\to 0^+} x^{-\nu} j_\nu(x) = \frac{1}{(2(\nu+1/2))!!}$. For double factorial, it holds $(2n)!! = 2^n n! \sqrt{2/\pi}$, hence the limit writes

$$\lim_{x\to 0^+} x^{-\nu} j_\nu(x) = \frac{1}{(2(\nu+1/2))!!} = \sqrt{\frac{\pi}{2}} \frac{1}{2^{\nu+1/2}(\nu+1/2)!} \overset{\nu=(d-2)/2}{=} \frac{\sqrt{\pi}}{2^{d/2}\Gamma(\frac{d+1}{2})} \tag{2.9}$$

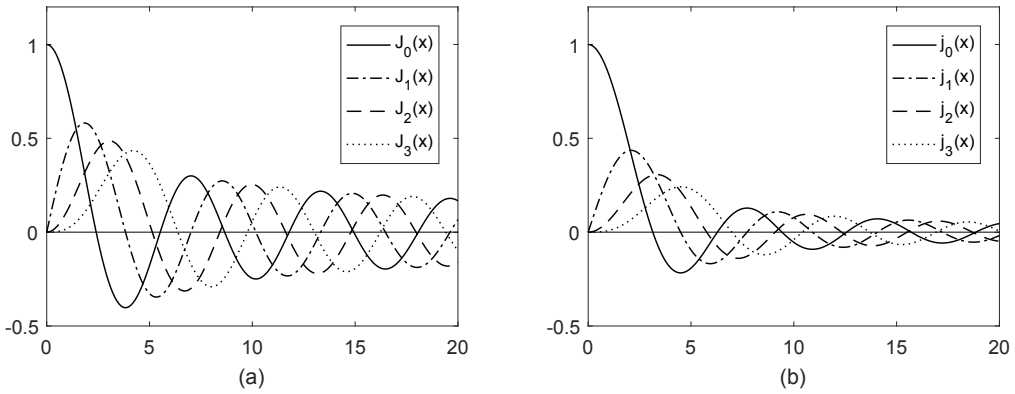

Figure 2.1: (a) Bessel function of the first kind J_ν; (b) spherical version j_ν.

2.3 Fourier and Hankel transforms

Within the thesis we work intensively with the Fourier transform. We also consider the Hankel transform as a tool for dealing with the multivariate Fourier transform of the radial functions. The Hankel transform calls for using Bessel fuctions of the first kind recalled above.

2.3.1 Fourier transform

In the thesis we consider the multivariate Fourier transform as specified in Definition 2.4 below that differs from the mainstream version which uses the negative exponent and scaling by $(2\pi)^{-d/2}$ factor [Rudin, 1991]. The reason is that we want the Fourier transform to correspond to the specification of the characteristic functions when applied on densities of probability distributions. Non-symmetry in the Plancherel's formula is the price we pay for not using scaling.

The Fourier transform applies on functions from $L^1(\mathbb{R}^d)$ with the standard extension to $L^2(\mathbb{R}^d)$ space.

Definition 2.4. *Let $f \in L^1(\mathbb{R}^d)$. Its multivariate Fourier transform is specified as*

$$\mathcal{F}[f](\boldsymbol{\omega}) = \int_{\mathbb{R}^d} e^{\mathrm{i}\langle \boldsymbol{x}, \boldsymbol{\omega} \rangle} f(\boldsymbol{\omega})\, d\boldsymbol{x}\,. \tag{2.10}$$

The basic properties of the Fourier transform are well known [Pinsky, 2008, p. 91]. Let $f, g \in L^1(\mathbb{R}^d)$, $D^\alpha f \in L^1(\mathbb{R}^d)$ for a multi-index $\alpha \in \mathbb{N}_0^d$, then the following properties of the Fourier transform are relevant to our research:

- boundedness: $|\mathcal{F}[f](\boldsymbol{\omega})| \leq 1$, for f being a density
- 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: $\mathcal{F}[f(\boldsymbol{x} - \boldsymbol{s})](\boldsymbol{\omega}) = e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{s}\rangle}\mathcal{F}[f](\boldsymbol{\omega}), \quad \boldsymbol{s} \in \mathbb{R}^d$
- scaling: $\mathcal{F}[f(\boldsymbol{x}/h)/h^d](\boldsymbol{\omega}) = \mathcal{F}[f](h\boldsymbol{\omega}), \quad h > 0$
- shifting & scaling: $\mathcal{F}[f((\boldsymbol{x} - \boldsymbol{s})/h)/h^d] = e^{i\langle\boldsymbol{\omega},\boldsymbol{s}\rangle}\mathcal{F}[f](h\boldsymbol{\omega}), \quad \boldsymbol{s} \in \mathbb{R}^d$
- complex conjugate: $\overline{\mathcal{F}[f](\boldsymbol{\omega})} = \mathcal{F}[f](-\boldsymbol{\omega})$
- convolution: $\mathcal{F}[f * g](\boldsymbol{\omega}) = \mathcal{F}[f](\boldsymbol{\omega})\mathcal{F}[g](\boldsymbol{\omega})$
- differentiation: if $D^\alpha f$ exists and $D^\alpha f \in L^1(\mathbb{R}^d)$, then $\mathcal{F}[D^\alpha f](\boldsymbol{\omega}) = (+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1}\omega_2^{\alpha_2} \cdot \cdots \cdot \omega_d^{\alpha_d})\mathcal{F}[f](\boldsymbol{\omega})$, for a multi-index $\alpha \in \mathbb{N}_0^d$
- symmetry: if $f(-\boldsymbol{x}) = f(\boldsymbol{x})$, then $\mathcal{F}[f](-\boldsymbol{\omega}) = \mathcal{F}[f](\boldsymbol{\omega})$
- isometry, provided by the Plancheler's formula for $f \in L^2(\mathbb{R}^d)$:

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

2.3.2 Hankel transform

The Hankel transform is an integral transform that use Bessel functions as its kernel. The transform is useful when dealing with problems that show circular symmetry. The definition follows.

Definition 2.5. *The Hankel transform of order ν of a function $\varphi : [0, \infty) \to \mathbb{R}$ is defined as*

$$\mathscr{H}_\nu\{\varphi(r)\}(s) = \int_0^\infty \varphi(r) J_\nu(sr) r\, dr.$$

where J_ν is the Bessel function of the first kind of order $\nu > -\frac{1}{2}$.

Of the properties of the Hankel transform, we metion that it is self-inverse, i.e.,

$$\mathscr{H}_\nu\{\mathscr{H}_\nu\{\varphi(r)\}\} = \varphi(r);$$

and its role when computing the multivariate Fourier transform of the radial functions as the Fourier transform of (Euclidean) radial functions are again radial functions [Pinsky, 2008, A. D. Poularikas (Ed.), 2000].

The proof of the following theorem (adjusted for our definition of the Fourier transform) can be found in [Stein and Weiss, 1971, Theorem 3.3].

Theorem 2.1. *Let $\Phi \in L^1(\mathbb{R}^d)$ be continuous and radial, i.e., $\Phi(\boldsymbol{x}) = \varphi(||\boldsymbol{x}||_2)$. Then its Fourier transform $\mathcal{F}[\Phi](\boldsymbol{\omega})$ is also radial $\mathcal{F}[\Phi](\boldsymbol{\omega}) = \varphi^{\mathscr{H}}(||\boldsymbol{\omega}||_2)$ where*

$$\varphi^{\mathscr{H}}(s) = \frac{(2\pi)^{d/2}}{\sqrt{s^{d-2}}} \int_0^\infty \varphi(r)\, r^{\frac{d}{2}} J_{(d-2)/2}(sr)\, dr = (2\pi)^{d/2} s^{-\nu} \mathscr{H}_\nu\{\varphi(r) \cdot r^\nu\}(s)$$

for $\nu = (d-2)/2$, i.e., $\nu = -\frac{1}{2}, 0, \frac{1}{2}, 1, \ldots$ for $d = 1, 2, 3, 4 \ldots$.

2.4 Particle filtering

Particle filtering is a methodology for solving the *filtering problem* in generally nonlinear/non-Gaussian settings which typically do not allow for analytical solutions. The filtering problem is casted in different ways [Fristedt et al., 2007, Särkkä, 2013]. In the thesis, we consider the variant with discrete time and continuous state space. A detailed description is provided, for example, in [Doucet and Johansen, 2011, Doucet et al., 2001]. Here we present the review of the filtering problem in the context of particle filtering as given in [Coufal, 2016, Section 2].

2.4.1 Filtering problem

The filtering problem is a problem of determining the optimal estimate of an inaccessible state of a stochastic process using accessible observations. The observations constitute a stochastic process called the *observation process*. The observation process is assumed to be interconnected with the principal stochastic process called the *signal process*. States of the signal process are the subject of estimation. The mathematical formulation follows.

Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$, $\{\boldsymbol{Y}_t\}_{t=1}^{\infty}$ be two stochastic processes specified on a common probabilistic space $(\Omega, \mathcal{A}, P)$ such that

- $\boldsymbol{X}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_x}, \mathcal{B}(\mathbb{R}^{d_x}))$, $t \in \mathbb{N}_0$, $d_x \in \mathbb{N}$,
- $\boldsymbol{Y}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_y}, \mathcal{B}(\mathbb{R}^{d_y}))$, $t \in \mathbb{N}$, $d_y \in \mathbb{N}$.

The first process $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ is the *signal process*. It represents generally an inhomogeneous Markov chain with a continuous state space $\mathbb{R}^{d_x}$ endowed with its standard Borel σ-algebra $\mathcal{B}(\mathbb{R}^{d_x})$. The probabilistic behavior of the chain is determined by the initial distribution $\pi_0(d\boldsymbol{x}_0)$ of $\boldsymbol{X}_0$ and 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 induced by the transition kernel K_{t-1} for $\boldsymbol{x}_{t-1} \in \mathbb{R}^{d_x}$ being fixed.

The second process $\{\boldsymbol{Y}_t\}_{t=1}^{\infty}$ is the *observation process*. $\mathbb{R}^{d_y}$ is the continuous state space of the process and $\mathcal{B}(\mathbb{R}^{d_y})$ its Borel σ-algebra. The observation process is derived from the signal process using the transformation

$$\boldsymbol{Y}_t = h_t(\boldsymbol{X}_t) + \boldsymbol{V}_t, \quad t \in \mathbb{N} \tag{2.11}$$

where $h_t : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$, $t \in \mathbb{N}$ are Borel functions and $\boldsymbol{V}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_y}, \mathcal{B}(\mathbb{R}^{d_y}))$ are i.i.d. random variables that are independent from $\boldsymbol{X}_{0:t} = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_t)$ for all $t \in \mathbb{N}$. Due to (2.11) and the Markov character of $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$, the independence of $\boldsymbol{V}_t$ transfers on observations as

$$P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t|\boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t-1}) = P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t|\boldsymbol{X}_t). \tag{2.12}$$

For $t=1$, the left-hand side reads as $P(\boldsymbol{Y}_1 \in d\boldsymbol{y}_1|\boldsymbol{X}_{0:1})$.

Within the thesis, the colon is used to denote finite sequences. That is, we use for example $\boldsymbol{Y}_{1:t-1} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1})$ or $\boldsymbol{y}_{1:t} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$, etc.

The ultimate purpose of filtering is to provide as wide information as possible on the current state $\boldsymbol{X}_t$ of the signal process using the current and past observations $\boldsymbol{y}_{1:t} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$. The well known practical example of filtering is target tracking when one is interested in estimating the current position of a moving object (e.g., a plane or a missile) based on some indirect measurements (radar signals) [Zhao, 2006].

2.4.2 Filtering distribution and filtering density

In terms of probability theory the filtering problem reads as the task for specifying the conditional distribution $P(\boldsymbol{X}_t|\boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t})$. This distribution is called the *filtering distribution*.

Assuming L^2 integrability of $\boldsymbol{X}_t$, a restricted version of the filtering problem ask only for the specification of $\mathbb{E}[\boldsymbol{X}_t|\boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t}]$, i.e., of the conditonal expectation of the filtering distribution. It is a classical result that provided the L^2 integrability of $\boldsymbol{X}_t$, the L^2-optimal estimate of $\boldsymbol{X}_t$ based on $\boldsymbol{Y}_{1:t}$ corresponds just to $\mathbb{E}[\boldsymbol{X}_t|\boldsymbol{Y}_{1:t}]$.

If the filtering distribution has a density with respect to the corresponding (in the sense of the dimension) Lebesgue measure, then this density is called the *filtering density*. The specification of the filtering density reads as another version of the filtering problem.

To induce the existence of the filtering density, it is assumed that all the involved distributions have *bounded densities* with respect to the corresponding Lebesgue measures. Namely, it is assumed that

- $\pi_0(d\boldsymbol{x}_0) = p_0(\boldsymbol{x}_0)\, d\boldsymbol{x}_0$
- $K_{t-1}(d\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, d\boldsymbol{x}_t$
- $P(\boldsymbol{V}_t \in d\boldsymbol{v}_t) = g_t^v(\boldsymbol{v}_t)\, d\boldsymbol{v}_t$

where $p_0 : \mathbb{R}^{d_x} \to [0, \infty)$ is the density of the initial distribution $\pi_0(d\boldsymbol{x}_0)$. $K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) : \mathbb{R}^{d_x} \to [0, \infty)$ is the conditional density of $K_{t-1}(d\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ and $g_t^v(\boldsymbol{v}_t)$ is the density of the distribution of the noise variable $\boldsymbol{V}_t$. This density is considered not only bounded but also *strictly positive*, i.e., $g_t^v(\boldsymbol{v}_t) > 0$ for all $\boldsymbol{v}_t \in \mathbb{R}^{d_y}$ and $t \in \mathbb{N}$.

The assumption on existence of densities enables the specification of the conditional densities of our interest. The density of $P(\boldsymbol{Y}_t \in d\boldsymbol{y}_t|\boldsymbol{X}_t = \boldsymbol{x}_t)$ will be denoted $g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)$ and using (2.11) it writes

$$g_t(\boldsymbol{y}_t|\boldsymbol{x}_t) = g_t^v(\boldsymbol{y}_t - h_t(\boldsymbol{x}_t)). \tag{2.13}$$

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

$$p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t}) = p_0(\boldsymbol{x}_0) \prod_{k=1}^{t} g_k(\boldsymbol{y}_k|\boldsymbol{x}_k)\, K_{k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}), \tag{2.14}$$

which is induced by the conditional independence of observations (2.12) and by the standard theory of Markov chains with a continuous state space.

The filtering density at time $t \in \mathbb{N}$ is denoted by $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$. Using (2.14), it reads as

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{x}_t, \boldsymbol{y}_{1:t})}{p(\boldsymbol{y}_{1:t})} = \frac{\int p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t})\, d\boldsymbol{x}_{0:t-1}}{\int p(\boldsymbol{x}_{0:t}, \boldsymbol{y}_{1:t})\, d\boldsymbol{x}_{0:t}}. \tag{2.15}$$

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*.

2.4.3 Filtering equations

The filtering equations describe recursively development of the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ over time. They consist of the *prediction formula* (2.16) and the *update formula* (2.17).

The prediction formula gives the expression for the so-called prediction density which is the density of $P(\boldsymbol{X}_t \in d\boldsymbol{x}_t|\boldsymbol{Y}_{1:t} = \boldsymbol{y}_{1:t-1})$. The update formula then gives the specification of the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$.

Lemma 2.2. *Let the joint density of* $(\boldsymbol{X}_{0:t}, \boldsymbol{Y}_{1:t})$ *be given by formula* (2.14)*, then*

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int K_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_{t-1}, \tag{2.16}$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})}{\int g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_t}, \quad t \in \mathbb{N} \tag{2.17}$$

with $p(\boldsymbol{x}_1|\boldsymbol{y}_{1:0})$ *understood as* $p(\boldsymbol{x}_1)$ *and* $p(\boldsymbol{x}_0|\boldsymbol{y}_{1:0})$ *as* $p(\boldsymbol{x}_0)$*.*

Proof. A basic proof can be found, for example, in [Särkkä, 2013], see Theorem 4.1 on page 54; or in Section 2.6.2 of [Doucet et al., 2001] where it is presented in a more general form. □

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, is updated to the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ on the basis of the current observation $\boldsymbol{y}_t \in \mathbb{R}^{d_y}$.

Speaking in the language of distributions, the filtering distribution is usually denoted by π_t, i. e., $\pi_t(d\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})\, d\boldsymbol{x}_t$. The filtering distribution 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., $\overline{\pi}_t(d\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_t$.

2.4.4 Particle filtering

Time development 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 a particle filter operation because the filter alternately generates empirical prediction and update measures.

In particle filtering, 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, i. e., $\boldsymbol{X}_i \sim \mu$, and constructing the empirical measure $\delta_n(d\boldsymbol{x})$ as

$$\delta_n(d\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{x}_i}(d\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{X}_i}(d\boldsymbol{x}) \tag{2.18}$$

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 (2.18), the last 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.

The construction of empirical measures proceeds sequentially. The particles generated in the previous cycle of operation are employed in the current cycle. In each cycle, a stochastic update of particles and their weights is performed. The weights are updated on the basis of the current observation. The procedure is in fact an instance of the sequential Monte Carlo methods applied in the context of the filtering problem [Doucet et al., 2001]; 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 [Doucet et al., 2001], as $t \in \mathbb{N}$ increases, the distribution of weights becomes more and more skewed and after a few time steps only a single 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 draws on the idea of discarding particles with low weights (with respect to $1/n$) and promote those with high weights. Practically, this is done by sampling with replacement from the set of original particles with the probabilities of selection given by the original particles' weights. It means that the resampled particles might be duplicated. In fact, the numbers of duplicates correspond to a sample from the multinomial distribution $\mathcal{M}(n, \widetilde{w}(\overline{\boldsymbol{x}}_t^1), \ldots, \widetilde{w}(\overline{\boldsymbol{x}}_t^n))$. Let us stress here that the resampled particles *does not constitute an i.i.d. sample.*

The particle filter's operation is presented in Algorithm 1.

Algorithm 1: Particle filtering algorithm

0. declarations

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

1. initialization

$t = 0$,
sample $\{\overline{\boldsymbol{x}}_0^i \sim p_0\}_{i=1}^n$,
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 $\{\overline{\boldsymbol{x}}_t^i \sim K_{t-1}(\cdot\,|\boldsymbol{x}_{t-1}^i)\}_{i=1}^n$,
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 $\mathcal{M}(n, \widetilde{w}(\overline{\boldsymbol{x}}_t^1), \dots, \widetilde{w}(\overline{\boldsymbol{x}}_t^n))$, resample $\{\boldsymbol{x}_t^i\}_{i=1}^n$ from $\{\overline{\boldsymbol{x}}_t^i\}_{i=1}^n$
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.

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 developments of the empirical measures and the theoretical distributions is presented in Figure 2.2.

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

Figure 2.2: Development of distributions in particle filtering.

2.4.5 Convergence results

In the particle filter, it is known that the empirical measures $\overline{\pi}_t^n$ and π_t^n converge weakly a.s. (they are the 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 real bounded functions on $\mathbb{R}^{d_x}$ by $B(\mathbb{R}^{d_x})$, the class of real bounded and continuous functions on $\mathbb{R}^{d_x}$ by $\mathcal{C}_b(\mathbb{R}^{d_x})$, the supremum norm of a function $f : \mathbb{R}^{d_x} \to \mathbb{R}$ by $||f||_\infty$ 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 $g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)$ of (2.13), $t \in \mathbb{N}$ are bounded, continuous and strictly positive functions.

Theorem 2.2. *Let $\{\overline{\pi}_t^n\}_{t=1}^T$ and $\{\pi_t^n\}_{t=1}^T$ be the sequences of empirical measures generated from particle filtering 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 $f \in B(\mathbb{R}^{d_x})$,*

$$\lim_{n\to\infty} |\overline{\pi}_t^n f - \overline{\pi}_t f| = 0 \ \ a.s., \ \ \lim_{n\to\infty} |\pi_t^n f - \pi_t f| = 0 \ \ a.s.$$

Proof. See [Doucet et al., 2001], Chapter 2 for a broader discussion of the convergence theorems. Other source is [Crisan and Doucet, 2002, Section IV]. Paper [Crisan and Míguez, 2014] has even the proof of the a.s. convergence for certain unbounded functions, see Proposition 2.1(b). □

Theorem 2.3. *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 $t \in \{1, \ldots, T\}$ *and* $f \in B(\mathbb{R}^{d_x})$,

$$\mathbb{E}[|\pi_t^n f - \pi_t f|^2] \leq \frac{c_t^2 ||f||_\infty^2}{n} \tag{2.19}$$

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

Proof. In this formulation, the theorem is presented in [Crisan and Doucet, 2002, Section V], (the authors use c_t instead ours c_t^2). □

Corollary. Theorem 2.3 holds also if $f \in B^{\mathbb{C}}(\mathbb{R}^{d_x})$, i. e., if f is a bounded complex function of real variables on $\mathbb{R}^{d_x}$.

Proof. If $f \in B^{\mathbb{C}}(\mathbb{R}^{d_x})$, then $f(\boldsymbol{x}) = h(\boldsymbol{x}) + \mathrm{i}g(\boldsymbol{x})$, where i denotes the imaginary unit; and $f, g \in B(\mathbb{R}^{d_x})$. Inequality (2.19) holds because for the squared modulus of $\pi_t^n f - \pi_t f$ one has $|\pi_t^n f - \pi_t f|^2 = (\pi_t^n h - \pi_t h)^2 + (\pi_t^n g - \pi_t g)^2$. □

Remark that the L^1 version of Theorem 2.3, i. e., $\mathbb{E}[|\pi_t^n f - \pi_t f|]$, is treated in [Doucet et al., 2001, Theorem 2.4.1]. The theorem is further mentioned for general L^p norm, $p \geq 1$, in [Crisan and Míguez, 2014, Proposition 2.1(a)].

2.5 Kernel estimates in Fourier domain

This section reviews theory of kernel density estimates in the Fourier domain as presented in [Coufal, 2016, Section 3]. It delivers the multivariate counterparts of results presented in [Tsybakov, 2009, Chapter 1].

2.5.1 Basics of kernel methods

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $n \in \mathbb{N}$ be a set of independent random variables identically distributed as the real random variable $\boldsymbol{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)), d \in \mathbb{N}$. Let the distribution of $\boldsymbol{X}$ have 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 $K : \mathbb{R}^d \to \mathbb{R}$ that 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

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\Big(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\Big) = \frac{1}{nh^d} \sum_{i=1}^{n} K\Big(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h}\Big). \tag{2.20}$$

In the formula, the last 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 relevant the question on consistency and unbiasedness of the estimate. In the univariate case, the classical result by Parzen [Parzen, 1962] (see also [Silverman, 1986, Sec. 3.7.1]) states the conditions under which the estimate is consistent. The result extends to the multivariate case, see e. g. [Givens, 1995]. Certain conditions are imposed on the properties of the kernel function and on development of the bandwidth h as a function of the sample size $n \in \mathbb{N}$. We mention only that h is required to develop 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* (MSE). 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

$$\mathrm{MSE}_{\boldsymbol{x}}(\hat{f}_n) = (\mathbb{E}[\hat{f}_n(\boldsymbol{x})] - f(\boldsymbol{x}))^2 + var[\hat{f}_n(\boldsymbol{x})] = (b\hat{f}_n(\boldsymbol{x}))^2 + \sigma^2\hat{f}_n(\boldsymbol{x}) \tag{2.21}$$

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 $\mathrm{MSE}_{\boldsymbol{x}}(\hat{f}_n)$ is the local measure of the quality of the estimate. It is desirable to introduce also a corresponding global measure. Expectedly, such 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.

Using (2.21) and the Fubini's theorem, the MISE of the kernel density estimate $\hat{f}_n$ is expressed as

$$\mathrm{MISE}(\hat{f}_n) = \mathbb{E} \int (\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}))^2 \, d\boldsymbol{x} = \int (b\hat{f}_n(\boldsymbol{x}))^2 \, d\boldsymbol{x} + \int \sigma^2\hat{f}_n(\boldsymbol{x}) \, d\boldsymbol{x}. \tag{2.22}$$

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

Because $\mathrm{MISE}(\hat{f}_n)$ represents the global error of the estimate, one tries to minimize it by localizing the minimizer h^*_{MISE} of (2.22). Analytical solution to this task is known only in some specific cases, e. g., when the estimated density corresponds to a convex sum of normal densities, see [Silverman, 1986, p. 37] or [Wand and Jones, 1995, Sec. 2.6] for exact formulas. To deal with the minimization problem generally, the widely used approach is to investigate the asymptotic behavior of the $\mathrm{MISE}(\hat{f}_n)$ with respect to the sample size $n \in \mathbb{N}$ going to infinity. This is called AMISE analysis and leads to the specification of the asymptotic minimizer h^*_{AMISE}.

However, in Section 1.2.4 of his book [Tsybakov, 2009], 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 the fixed density f, it is possible to construct such a kernel estimate that the $\mathrm{MISE}(\hat{f}_n)$ 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 [Tsybakov, 2009]. The Sobolev class is treated in Definition 2.7 below.

Based on this criticism, Tsybakov presents a different approach to the MISE analysis in Section 1.3 of [Tsybakov, 2009]. The approach relies on Fourier analysis.

2.5.2 Fourier analysis of kernel estimates

In the probability theory, Fourier analysis is intimately connected 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 $\phi_{\boldsymbol{X}}(\boldsymbol{\omega}) : \mathbb{R}^d \to \mathbb{C}$ of $\boldsymbol{X}$ is defined as the integral transform

$$\phi_{\boldsymbol{X}}(\boldsymbol{\omega}) = \mathbb{E}[e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{X}\rangle}] = \int e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle}\, \mu(d\boldsymbol{x}), \quad \boldsymbol{\omega} \in \mathbb{R}^d \tag{2.23}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard dot product in $\mathbb{R}^d$. 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 $\boldsymbol{X}$ or the distribution μ.

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

$$\mathcal{F}[f](\boldsymbol{\omega}) = \int e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle} f(\boldsymbol{x})\, d\boldsymbol{x}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{2.24}$$

Formula (2.24) can be treated as the special case of formula (2.23) when the distribution of $\boldsymbol{X}$ is absolutely continuous with respect to the d-dimensional Lebesgue measure and has the density f, i. e., if $\mu(d\boldsymbol{x}) = f(\boldsymbol{x})\, d\boldsymbol{x}$. On the other hand, in (2.24) f need not be necessarily a density.

The uniformly weighted sum of Dirac measures $\delta_n(d\boldsymbol{x})$ introduced in formula (2.18) represents a probability distribution which does not have a density with respect to the corresponding Lebesgue measure. Its characteristic function is denoted $\phi_n(\boldsymbol{\omega})$ and specified as

$$\phi_n(\boldsymbol{\omega}) = \int e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle} \delta_n(d\boldsymbol{x}) = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{X}_j\rangle}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{2.25}$$

Note that $\phi_n(\boldsymbol{\omega})$ constitutes a random variable for every $\boldsymbol{\omega} \in \mathbb{R}^d$ 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 kernel density estimate (2.20). Using the linearity and the shifting & scaling property of the Fourier transform, $\mathcal{F}[\hat{f}_n](\boldsymbol{\omega})$ is specified by formula

$$\mathcal{F}[\hat{f}_n](\boldsymbol{\omega}) = \frac{1}{n} \sum_{j=1}^{n} \mathcal{F}\left[\frac{1}{h^d} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_j}{h}\right)\right] = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{X}_j\rangle} \mathcal{F}[K](h\boldsymbol{\omega}). \tag{2.26}$$

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

$$\mathcal{F}[\hat{f}_n](\boldsymbol{\omega}) = \phi_n(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega}). \tag{2.27}$$

This shows that the standard kernel estimate, 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.

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 (2.27), we get for the MISE of (2.22) the expression

$$\mathrm{MISE}(\hat{f}_n) = \frac{1}{(2\pi)^d}\, \mathbb{E} \int |\phi_n(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega}) - \phi(\boldsymbol{\omega})|^2\, d\boldsymbol{\omega}. \tag{2.28}$$

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

Theorem 2.4. *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 kernel estimator $\hat{f}_n$ of (2.20) has the form*

$$\begin{aligned} \mathrm{MISE}(\hat{f}_n) &= \frac{1}{(2\pi)^d}\left[\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} + \frac{1}{n}\int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}\right] \\ &\quad - \frac{1}{(2\pi)^d}\frac{1}{n}\int |\phi(\boldsymbol{\omega})|^2 |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}. \end{aligned} \tag{2.29}$$

Proof. The proof is just a copy of the original univariate Tsybakov's proof, see [Tsybakov, 2009, p. 22] (generally, we do not need the symmetry of the kernel here). It rests on developing the formula (2.28) using the facts that $|z|^2 = z\overline{z}$ for $z \in \mathbb{C}$ and $\mathbb{E}[\phi_n(\boldsymbol{\omega})] = \phi(\boldsymbol{\omega})$. □

Now, we are going to discuss the individual terms in the Fourier MISE formula (2.29). We start with the notion of the *order of a kernel*. To do so we consider the differential operator D^α with the multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, $\alpha \in \mathbb{N}_0^d$. For a suitably differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, one has $D^\alpha f = \partial^{|\alpha|} f / \partial^{\alpha_1} x_1 \cdots \partial^{\alpha_d} x_d$ with $|\alpha| = \sum_{i=1}^d \alpha_i$ being the order of the derivative.

Definition 2.6. *Let $\ell \geq 1$ be an integer. We say that the kernel K: $\mathbb{R}^d \to \mathbb{R}$ is of order ℓ, if K is $L^1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ integrable, its Fourier transform $K_{\mathcal{F}}(\boldsymbol{\omega}) = \mathcal{F}[K](\boldsymbol{\omega})$ is real, satisfies $K_{\mathcal{F}}(\mathbf{0}) = 1$ and has all partial derivatives $D^\alpha K_{\mathcal{F}}$ up to the ℓ-th order such that $D^\alpha K_{\mathcal{F}}(\mathbf{0}) = 0$ for all $|\alpha| = 1, \ldots, \ell$.*

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

- $\int K(\boldsymbol{u}) \, d\boldsymbol{u} = 1$,
- $\int u_1^{\alpha_1} \cdots u_d^{\alpha_d} K(\boldsymbol{u}) \, d\boldsymbol{u} = 0$ for $|\alpha| = 1, \ldots, \ell$.

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 α-th partial derivative, one has

$$D^\alpha K_{\mathcal{F}}(\boldsymbol{\omega}) = \int (\mathrm{i}u_1)^{\alpha_1} \cdots (\mathrm{i}u_d)^{\alpha_d} \, e^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{u}\rangle} K(\boldsymbol{u}) \, d\boldsymbol{u}.$$

Thus, $0 = D^\alpha K_{\mathcal{F}}(\mathbf{0}) = (+\mathrm{i})^{|\alpha|} \int u_1^{\alpha_1} \cdots u_d^{\alpha_d} K(\boldsymbol{u}) \, d\boldsymbol{u}$.

As an example, mention that for the standard multivariate Gaussian kernel $K(\boldsymbol{u}) = (2/\pi)^{-d/2} \exp(-\frac{1}{2}||\boldsymbol{u}||_2^2)$, one has $K_{\mathcal{F}}(\boldsymbol{\omega}) = \exp(-\frac{1}{2}||\boldsymbol{\omega}||_2^2)$ and the kernel is of order $\ell = 1$.

The first term. For the first term in the Fourier MISE formula (2.29), we are able to say something more specific if we consider the order of the kernel involved in the estimate.

Lemma 2.3. *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*

$$\sup_{\boldsymbol{\omega}\in\mathbb{R}^d\setminus\{\mathbf{0}\}} \frac{|1 - K_{\mathcal{F}}(\boldsymbol{\omega})|}{||\boldsymbol{\omega}||_2^\ell} \leq A \tag{2.30}$$

and

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \leq A^2 h^{2\ell} \int ||\boldsymbol{\omega}||_2^{2\ell} |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \tag{2.31}$$

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

Proof. We employ the multidimensional Taylor's theorem [Brabec and Hruza, 1986]. Because the kernel K is of order $\ell \geq 1$, its Fourier transform $K_{\mathcal{F}}(\boldsymbol{\omega})$ is real and we have by the Taylor's theorem

$$K_{\mathcal{F}}(\boldsymbol{\omega}) = K_{\mathcal{F}}(\mathbf{0}) + \sum_{1\leq|\alpha|\leq\ell} \frac{D^\alpha K_{\mathcal{F}}(\mathbf{0})}{\alpha!}\, \boldsymbol{\omega}^\alpha + R_\ell(\boldsymbol{\omega})$$

with $\lim_{\omega\to\mathbf{0}} R_\ell(\boldsymbol{\omega})/||\boldsymbol{\omega}||_2^\ell = 0$ for the reminder, i. e., $R_\ell(\boldsymbol{\omega}) = o(||\boldsymbol{\omega}||_2^\ell)$.

As the partial derivatives vanish at origin, the remainder writes $R_\ell(\boldsymbol{\omega}) = K_{\mathcal{F}}(\boldsymbol{\omega}) - K_{\mathcal{F}}(\mathbf{0}) = K_{\mathcal{F}}(\boldsymbol{\omega}) - 1$ and $\lim_{\omega\to\mathbf{0}} |1 - K_{\mathcal{F}}(\boldsymbol{\omega})|/||\boldsymbol{\omega}||_2^\ell = 0$ by the Taylor's theorem.

Let us define $A_\ell(\boldsymbol{\omega}) = |1 - K_{\mathcal{F}}(\boldsymbol{\omega})|/||\boldsymbol{\omega}||_2^\ell$ for $\boldsymbol{\omega} \neq \mathbf{0}$, and $A_\ell(\mathbf{0}) = 0$. The function $A_\ell : \mathbb{R}^d \to [0, \infty)$ is continuous on $\mathbb{R}^d$ and attains its maximum on the unit ball $||\boldsymbol{\omega}||_2 \leq 1$. Let $M_1 = \max_{\{\boldsymbol{\omega}:||\boldsymbol{\omega}||_2\leq 1\}}\{A_\ell(\boldsymbol{\omega})\}$. 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, $|1 - K_{\mathcal{F}}(\boldsymbol{\omega})|/||\boldsymbol{\omega}||_2^\ell \leq 1 + M_2$ for $||\boldsymbol{\omega}||_2 > 1$. 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 (2.31) is implied by (2.30) as follows:

$$\begin{aligned}
\sup_{\omega\in\mathbb{R}^d\setminus\{\mathbf{0}\}} |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|/||h\boldsymbol{\omega}||_2^\ell &\leq A, \\
|1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 &\leq A^2 ||h\boldsymbol{\omega}||_2^{2\ell}, \\
\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} &\leq A^2 h^{2\ell} \int ||\boldsymbol{\omega}||_2^{2\ell} |\phi(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}.
\end{aligned}$$

This concludes the proof. □

The other terms in formula (2.29) refer to the properties of the kernel and density under considerations. We mention only two straightforward observations.

The second term. The second term can be translated from the frequency to the "time" domain using the Plancherel's theorem and the scaling property of the Fourier transform. Change of variables gives the final result:

$$\frac{1}{n}\int |K_{\mathcal{F}}(h\boldsymbol{\omega})|^2\, d\boldsymbol{\omega} = \frac{(2\pi)^d}{nh^{2d}}\int K^2(\boldsymbol{x}/h)\, d\boldsymbol{x} = \frac{(2\pi)^d}{nh^d}\int K^2(\boldsymbol{u})\, d\boldsymbol{u}. \tag{2.32}$$

The third term. The third term is actually the correction term. We have the following inequality for it:

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

2.5.3 Upper bound on the Fourier MISE formula

Concerning an upper bound on the Fourier MISE formula (2.29), we sum up the above results. First of all, to obtain the upper bound we can omit the correction (the third) term in (2.29). The second term is solely determined by the properties of the kernel, which is expressed by formula (2.32). 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 and functions.

Definition 2.7. *Let $\beta \geq 1$ be an integer and $L > 0$ a real. The Sobolev class of functions $\mathcal{P}^{\infty}_{S(\beta,L)}$ consists of $L_1(\mathbb{R}^d)$ integrable functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying*

$$\int ||\boldsymbol{\omega}||_2^{2\beta} |\mathcal{F}[f](\boldsymbol{\omega})|^2\, d\boldsymbol{\omega} \leq (2\pi)^d L^2. \tag{2.33}$$

The function $f \in L_1(\mathbb{R}^d)$ is called β-Sobolev if $f \in \mathcal{P}^{\infty}_{S(\beta,L)}$. If $f \in L^1(\mathbb{R}^d)$ is known to be a density we use the notation $f \in \mathcal{P}_{S(\beta,L)}$.

The condition (2.33) is related to integrability 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 (2.33) holds for $\beta = 1$ and $L = ||(L_1, \ldots, L_d)||_2$. Furthermore, if $f \in \mathcal{P}_S(\beta, L)$, for some $\beta \in \mathbb{N}$ and $L > 0$, then $f \in L^2(\mathbb{R}^d)$.

The announced matching is provided by 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 [Tsybakov, 2009], provides the final result.

Theorem 2.5. *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 $L > 0$. Let K be a kernel of order β. Assume that the inequality* (2.30) *holds for some constant $A > 0$. Fix $a > 0$ and set $h(n) = an^{-\frac{1}{2\beta+d}}$. Then for any $n \geq 1$ the kernel density estimate $\hat{f}_n$ satisfies*

$$\sup_{f \in \mathcal{P}_S(\beta,L)} \mathbb{E} \int (\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}))^2 \, d\boldsymbol{x} \leq C \cdot n^{-\frac{2\beta}{2\beta+d}} \tag{2.34}$$

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

Proof. By Lemma 2.3 and from the definition of the Sobolev class of densities, one has

$$\int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 |\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq A^2 h^{2\beta} \int ||\boldsymbol{\omega}||_2^{2\beta} |\phi(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \leq (2\pi)^d A^2 h^{2\beta} L^2.$$

Plugging this into the Fourier MISE decomposition formula (2.29) and using $\frac{1}{(2\pi)^d n} \int |K(h\boldsymbol{\omega})|^2 d\boldsymbol{\omega} = \frac{1}{nh^d} \int K^2(\boldsymbol{u}) \, d\boldsymbol{u}$ one gets for $h = an^{-\frac{1}{2\beta+d}}$,

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

and

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

□

The theorem provides the upper bound on the MISE of the multivariate kernel density estimate (2.20), if the order of the employed kernel fits to the Sobolev character of the density of the distribution the data are sampled from.

3. Kernel density estimates in particle filtering

This chapter delivers the upper and lower bounds on the kernel density estimates of the filtering densities in particle filtering. The upper bounds originally proved in [Coufal, 2016] are extended to the bounds on the partial derivatives of the filtering densities using tools of Fourier analysis. In contrast, derivation of the lower bounds draws on notions and techniques of information theory. As the lower bounds meet the upper ones, the standard kernel density estimates are in some sense optimal in the context of particle filtering.

We start by unifying the notations of the preceding sections and we further denote the dimensionality of the signal process state space by d instead of former d_x. Recall that particle filtering generates at each time step $t = 1, \ldots, T$, $T \in \mathbb{N}$ the empirical measure $\pi_t^n(d\boldsymbol{x}_t) = \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{x}_t^i}(d\boldsymbol{x}_t)$. The π_t^n 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 $\{\boldsymbol{x}_t^i \in \mathbb{R}^d\}_{i=1}^n$, $n \in \mathbb{N}$. 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 $\{\boldsymbol{x}_t^i\}_{i=1}^n$, the selected kernel K and bandwidth h, i. e., the estimate

$$\hat{p}_t^n(\boldsymbol{x}_t) = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{\boldsymbol{x}_t - \boldsymbol{x}_t^i}{h}\right), \tag{3.1}$$

we ask if $\hat{p}_t^n$ still converges in the MISE to the filtering density p_t, provided that the number of particles goes to infinity. More broadly, we are also interested in estimating partial derivatives of the filtering density using the corresponding partial derivatives of the kernel estimate (3.1).

3.1 Upper bounds

The next theorem provides the upper bound on the MISE of kernel density estimate of $D^\alpha p_t$. Clearly, if $|\alpha| = 0$, then no differentiation is applied and the result corresponds to that of Theorem 4.1 in [Coufal, 2016].

Theorem 3.1. *In the filtering problem, let* $\{\pi_t\}_{t=0}^T$, $\{D^\alpha p_t\}_{t=0}^T$, $T \in \mathbb{N}$ *be the sequences of the filtering distributions and partial derivatives of the corresponding filtering densities for some multi-index* $\alpha = (\alpha_1, \ldots, \alpha_d)$, $|\alpha| \in \mathbb{N}_0$. *Let* $D^\alpha p_t$, $t \in \{0, \ldots, T\}$ *be* β*-Sobolev for some* $\beta \in \mathbb{N}$ *and* $L_{t,\alpha} > 0$, *i.e.,* $D^\alpha p_t \in \mathcal{P}^{\propto}_{S(\beta, L_{t,\alpha})}$. *Let* $\{\pi_t^n\}_{t=1}^T$, $\{D^\alpha \hat{p}_t^n\}_{t=1}^T$, $n \in \mathbb{N}$ *be the sequences of the empirical measures generated from particle filtering and the partial derivatives of the related kernel density estimates* (3.1) *with the bandwidth varying as* $h(n) = an^{-\frac{1}{2\beta+d+2|\alpha|}}$ *for some* $a > 0$. *Let the kernel* K *employed in the estimates be of order* β *and* $||D^\alpha K||_2^2 < \infty$. *Then we have the following upper bounds on the MISE of* $D^\alpha \hat{p}_t^n$ *for* $t \in \{1, \ldots, T\}$:

$$\mathbb{E}\left[\int (D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right] \leq C_{t,\alpha}^2 \cdot n^{-\frac{2\beta}{2\beta+d+2|\alpha|}} \tag{3.2}$$

where

$$C_{t,\alpha} = AL_{t,\alpha}a^\beta + c_t a^{-(d/2+|\alpha|)}||D^\alpha K||_2. \tag{3.3}$$

In (3.3), *A is the constant of Lemma 2.3,* c_t, $t \in \{1, \ldots, T\}$ *are the constants of Theorem 2.3 and* $||D^\alpha K||_2$ *is the* L^2 *norm of the* α*-th partial derivative of the kernel* K.

Proof. Remind that for any sufficiently differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ and its α-th partial derivative $D^\alpha f : \mathbb{R}^d \to \mathbb{R}$, both assumed in $L^1(\mathbb{R}^d)$, one has for their Fourier transforms $\mathcal{F}[f](\boldsymbol{\omega})$ and $\mathcal{F}[D^\alpha f](\boldsymbol{\omega})$, respectively, the equality

$$\mathcal{F}[D^\alpha f](\boldsymbol{\omega}) = (+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1}\omega_2^{\alpha_2} \cdot \cdots \cdot \omega_d^{\alpha_d})\mathcal{F}[f](\boldsymbol{\omega}). \tag{3.4}$$

The Fourier transform of (3.1) read as $\mathcal{F}[\hat{p}_t^n](\boldsymbol{\omega}) = \psi_t^n(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega})$, where ψ_t^n is the characteristic function of π_t^n and $K_{\mathcal{F}}$ is the Fourier transform of the kernel K. Moreover, considering the convolution $p_t^* = p_t * (h^{-d}K(\cdot/h))$, one has $\mathcal{F}[p_t^*](\boldsymbol{\omega}) = \psi_t(\boldsymbol{\omega})K_{\mathcal{F}}(h\boldsymbol{\omega})$ with ψ_t being the characteristic function of π_t, see Section 2.5.2 for details.

In order to prove the theorem, we employ the complex exponential in (2.19). Let $f(\boldsymbol{x}_t) = e^{\mathrm{i}\langle\boldsymbol{\omega}, \boldsymbol{x}_t\rangle}$, then $||f||_\infty = 1$. Employing the characteristic functions

of π_t^n and π_t, the inequality (2.19) writes

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

$$|(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \mathbb{E}[|\psi_t^n(\boldsymbol{\omega}) - \psi_t(\boldsymbol{\omega})|^2] \leq\; |(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n},$$

$$\mathbb{E}\left[|(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d})(\psi_t^n(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega})) - (+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d})(\psi_t(\boldsymbol{\omega}) K_{\mathcal{F}}(h\boldsymbol{\omega}))|^2\right] \leq\; |(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d}) K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \cdot \frac{c_t^2}{n}.$$

Using (3.4) and integrating, we get

$$\begin{aligned}
\int \mathbb{E}\left[|\mathcal{F}[D^\alpha \hat{p}_t^n](\boldsymbol{\omega}) - \mathcal{F}[D^\alpha p_t^*](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}\right] &\leq \frac{c_t^2}{n} \int |\mathcal{F}[D^\alpha h^{-d} K(\boldsymbol{x}/h)]|^2 \, d\boldsymbol{\omega}, \\
\mathbb{E}\left[\int |\mathcal{F}[D^\alpha \hat{p}_t^n](\boldsymbol{\omega}) - \mathcal{F}[D^\alpha p_t^*](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}\right] &\leq \frac{c_t^2}{nh^{2d}} \int |\mathcal{F}[D^\alpha K(\boldsymbol{x}/h)]|^2 \, d\boldsymbol{\omega}, \\
\mathbb{E}\left[\int ((D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t^*(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right] &\leq \frac{c_t^2}{nh^{2d}} \int (D^\alpha K(\boldsymbol{x}/h))^2 \, d\boldsymbol{x}, \\
&\leq \frac{c_t^2}{nh^{2d+2|\alpha|}} \int (D^\alpha K(\boldsymbol{u}) \left|_{\boldsymbol{u}=\boldsymbol{x}/h}\right.)^2 \; d\boldsymbol{x}, \\
&\leq \frac{c_t^2}{nh^{d+2|\alpha|}} ||D^\alpha K||_2^2.
\end{aligned}$$

We assume that $D^\alpha p_t$ is β-Sobolev in terms of validity of (2.33). As kernel K is assumed to be of order β, we can again apply the Plancherel theorem, formula (3.4) and use Lemma 2.3 to get

$$\begin{aligned}
\int (D^\alpha p_t^*(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t &= (2\pi)^{-d} \int |\mathcal{F}[D^\alpha p_t^*](\boldsymbol{\omega}) - \mathcal{F}[D^\alpha p_t](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \\
&= (2\pi)^{-d} \int |1 - K_{\mathcal{F}}(h\boldsymbol{\omega})|^2 \, |\mathcal{F}[D^\alpha p_t](\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega} \\
&\leq A^2 h^{2\beta} L_{t,\alpha}^2.
\end{aligned}$$

As there is nothing random here, the inequality remains valid if we apply expectation on the left side. Considering the product measure $\lambda^d \otimes P$ (P is the probability measure the expectation $\mathbb{E}$ is taken w.r.t.) with the corresponding L^2 norm $||\cdot||_{\lambda^d \otimes P} = [\int\int |\cdot|^2 d(\lambda^d \otimes P)]^{1/2}$, the triangle inequality gives

$$||D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t)||_{\lambda^d \otimes P} \leq A h^\beta L_{t,\alpha} + \frac{c_t}{(nh^{d+2|\alpha|})^{-1/2}} ||D^\alpha K||_2.$$

The bandwidth h develops with n as $h(n) = an^{-\frac{1}{2\beta+d+2|\alpha|}}$ for some $a > 0$. So one has $h^\beta = a^\beta n^{-\frac{\beta}{2\beta+d+2|\alpha|}}$. Further, $(nh^{d+2|\alpha|})^{-1} = n^{-1}a^{-(d+2|\alpha|)}n^{\frac{d+2|\alpha|}{2\beta+d+2|\alpha|}} = a^{-(d+2|\alpha|)}n^{-\frac{2\beta}{2\beta+d+2|\alpha|}}$ and therefore $(nh^{d+2|\alpha|})^{-1/2} = a^{-(d+2|\alpha|)/2}n^{-\frac{\beta}{2\beta+d+2|\alpha|}}$. The above inequality then reads

$$||D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t)||_{\lambda^d \otimes P} \leq (AL_{t,\alpha}a^\beta + c_t a^{-(d+2|\alpha|)/2}||D^\alpha K||_2) \cdot n^{-\frac{2\beta}{2\beta+d+2|\alpha|}}.$$

Squaring to obtain the MISE, we get the statement of the theorem

$$\mathbb{E}\int (D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t \leq C_{t,\alpha}^2 \cdot n^{-\frac{2\beta}{2\beta+d+2|\alpha|}} \tag{3.5}$$

for $C_{t,\alpha} = AL_{t,\alpha}a^\beta + c_t a^{-(d+2|\alpha|)/2}||D^\alpha K||_2$. □

Let us discuss the theorem.

1) First of all, the theorem is proved without any assumption on the i.i.d. character of 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 (3.5) that the MISE of kernel estimates goes to zero as the number of particles increases and the bandwidth decreases accordingly, i. e.,

$$\lim_{n\to\infty} \mathbb{E}\int (D^\alpha \hat{p}_t^n(\boldsymbol{x}_t) - D^\alpha p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t = 0.$$

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

4) The dimension matters. One has $n^{-\frac{2\beta}{2\beta+d_1+|\alpha|}} < n^{-\frac{2\beta}{2\beta+d_2+|\alpha|}}$ 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, $n^{-\frac{2\beta_1}{2\beta_1+d+|\alpha|}} > n^{-\frac{2\beta_2}{2\beta_2+d+|\alpha|}}$ 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}. In Chapter 4 we present techniques for constructing the kernels of arbitrary orders, however, the order of the used kernel is primarily driven by the Sobolev character of the filtering density.

6) The theorem assumes that $D^\alpha p_t$ are β-Sobolev functions over time steps $t \in \{0, \dots, T\}$, $T \in \mathbb{N}$ with $\beta \in \mathbb{N}$ being constant over time. It is the question when this assumption holds. In Section 3.3, we show that the Sobolev character of $D^\alpha p_t$ is retained over time, if a certain condition holds on the transition kernels of the signal process.

3.2 Lower bounds

This section provides the lower bounds on the kernel density estimates in particle filtering. To establish the bounds we follow the approach referred in Chapter 2 of [Tsybakov, 2009] that provides the general framework for specifying minimax lower bounds in the nonparametric context. The theorem reads as follows:

Theorem 3.2. *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, $t \in \{0, \dots, T\}$ *be* β*-Sobolev for some* $\beta \in \mathbb{N}$ *and* $L_t > 0$. *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 from particle filtering and related kernel density estimates* (3.1) *with the bandwidth varying as* $h(n) = an^{-\frac{1}{2\beta+d}}$ *for some* $a > 0$. *Let the kernel* K *employed in the estimates be of order* β. *Then we have the following minimax lower bounds on the MISE of* $\hat{p}_t^n$ *for* $t \in \{1, \dots, T\}$:

$$\liminf_{n\to\infty} \sup_{p_t \in \mathcal{S}(\beta, L_t)} \mathbb{E}\left[\int (\hat{p}_t^n(\boldsymbol{x}_t) - p_t(\boldsymbol{x}_t))^2 \, d\boldsymbol{x}_t\right] \geq c_t^2 \cdot n^{-\frac{2\beta}{2\beta+d}} \tag{3.6}$$

where c_t *is constant with respect to* n.

Proof. The proof builds on the technique used for proving Theorem 2.8 in Tsybakov [2009]. It provides the lower bounds as a consequence of the following three conditions which are assumed to hold simultaneously:

1. there exists a suitable set of hypotheses, which is actually the set of $M+1$, $M \in \mathbb{N}$ distributions with appropriate Sobolev densities;

2. the hypotheses are well separated,

but they are

3. not too far from the certain basic hypothesis in terms of the KL divergence.

To ensure and check these conditions, we follow the Tsybakov's line of proof, see Section 2.6.1 of Tsybakov [2009], while changing the original univariate setting to the multivariate one, adapting to density estimation and using some suggestions by Chigansky[1].

We start with the univariate compactly supported C^∞ bump function

$$K_0(u) = \exp\left(-\frac{1}{1-u^2}\right) \cdot 1_{|u|\le 1}.$$

Asymptotically, the Fourier transform of this function decays faster than any polynomial [Johnson, 2015], so K_0 satisfies the inequality (2.33) for any $\beta \in \mathbb{N}$.

Define function K_1 by combining two shifted and scaled basic bump functions: $K_1(u) = K_0(4u+1) - K_0(4u-1)$, $u \in \mathbb{R}$. K_1 is a compactly supported "wave" function on the interval $[-1/2, 1/2]$. It is positive on $(-1/2, 0)$ and negative on $(0, 1/2)$. By construction, $\int K_1(u)\, du = 0$.

A multivariate version of K_1 writes $K(\boldsymbol{u}) = \prod_{i=1}^d K_1(u_i)$ for $\boldsymbol{u} \in \mathbb{R}^d$. Clearly, $\int K(\boldsymbol{u})\, d\boldsymbol{u} = 0$ and $||K||_\infty < \infty$.

K is β-Sobolev for any $\beta \in \mathbb{N}$. Indeed, K_1 is β-Sobolev because K_0 is β-Sobolev. The modulus of the Fourier transform of K reads as $|\mathcal{F}[K]| = \prod_{i=1}^d |\mathcal{F}[K_1](\omega_i)|$. Integral (2.33) then writes $\int (\sum_i \omega_i^2)^\beta \prod_{i=1}^d |\mathcal{F}[K_1](\omega_i)|^2\, d\boldsymbol{\omega}$. Employing the multinomial theorem and the fact that each $\mathcal{F}[K_1](\omega_i)$ is β-Sobolev for any $\beta \in \mathbb{N}$ shows that the integral can be decomposed into the β-Sobolev summands.

Construction of hypotheses. For a real number c_0, which will be specified below, and given number of particles $n \ge 1$ set

$$m = \lceil c_0 n^{\frac{1}{2\beta+d}} \rceil, \quad h_n = \frac{1}{m}$$

where $\lceil z \rceil$ is the smallest integer which is strictly greater than $z \in \mathbb{R}$.
Further, for $\boldsymbol{k} \in \{1, \dots, m\}^d$ let

$$\boldsymbol{x}_{\boldsymbol{k}} = \frac{\boldsymbol{k} - 1/2}{m},$$

i.e., $\boldsymbol{x}_{\boldsymbol{k}} \in [0,1]^d$. And finally, for $L > 0$ define the functions

$$g_{\boldsymbol{k}}(\boldsymbol{x}) = L h^\beta K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_{\boldsymbol{k}}}{h_n}\right),\ \boldsymbol{k} \in \{1, \dots, m\}^d.$$

Remark that $\int g_{\boldsymbol{k}}(\boldsymbol{x})\, d\boldsymbol{x} = 0$ for all $\boldsymbol{k}$.

[1] `http://pluto.huji.ac.il/~pchiga/teaching/Nonparametric/TsybakovSolutions.pdf`

As in [Tsybakov, 2009] we consider the set of all binary sequences of length $m^d : \Omega = \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{m^d}),\ \omega_i \in \{0,1\}\} = \{0,1\}^{m^d}$ and select[2] the $M+1$, $M \in \mathbb{N}$ well separated hypotheses p_{jn}, $j = 0, \ldots, M$ from the set

$$\mathcal{E} = \left\{ p_{\boldsymbol{\omega}}(\boldsymbol{x}) = \varphi(\boldsymbol{x}) + \sum_{\boldsymbol{k} \in \{1,\ldots,m\}^d} \omega_{\boldsymbol{k}} g_{\boldsymbol{k}}(\boldsymbol{x}),\ \boldsymbol{\omega} \in \Omega \right\}$$

where $\varphi(\boldsymbol{x})$ is the density of the d-variate standard normal distribution. Using the lexicographic order of the Cartesian product, each $\boldsymbol{\omega} \in \Omega$ maps uniquely to the sequence $\{\omega_{\boldsymbol{k}} \in \{0,1\}, \boldsymbol{k} \in \{1,\ldots,m\}^d\}$. If $\omega_{\boldsymbol{k}} = 1$, then the multi-dimensional "wave" $K_{\boldsymbol{k}}(\boldsymbol{x}) = K((\boldsymbol{x} - \boldsymbol{x}_{\boldsymbol{k}})/h_n)$ centered at $\boldsymbol{k}$ is present in $p_{\boldsymbol{\omega}}$ and absent if $\omega_{\boldsymbol{k}} = 0$. The map further allows for treating the sum $\sum_{\boldsymbol{k} \in \{1,\ldots,m\}^d}$ as the sum $\sum_{k=1}^{m^d}$, which will be used in the following text. Finally, note that $\int p_{\boldsymbol{\omega}}(\boldsymbol{x})\, d\boldsymbol{x} = 1$ and $p_{\boldsymbol{\omega}}$ are positive for sufficiently large n because $||K||_\infty < \infty$.

For any two $\boldsymbol{\omega}, \boldsymbol{\omega'} \in \Omega$, one has $||p_{\boldsymbol{\omega}} - p_{\boldsymbol{\omega'}}||_2^2 = \int_{[0,1]^d} (p_{\boldsymbol{\omega}} - p_{\boldsymbol{\omega'}})^2\, d\boldsymbol{x} = \sum_{k=1}^{m^d} (\omega_k - \omega'_k)^2 \int_{\Delta_{\boldsymbol{k}}^d} g_{\boldsymbol{k}}^2(\boldsymbol{x})\, d\boldsymbol{x} = L^2 h_n^{2\beta+d} ||K||_2^2\, \rho(\boldsymbol{\omega}, \boldsymbol{\omega'})$ where $\Delta_{\boldsymbol{k}}^d$ is the support of $K_{\boldsymbol{k}}$ and $\rho(\boldsymbol{\omega}, \boldsymbol{\omega'})$ is the Hamming distance between $\boldsymbol{\omega}$ and $\boldsymbol{\omega'}$, $\rho(\boldsymbol{\omega}, \boldsymbol{\omega'}) = \sum_{k=1}^{m^d} I(\omega_k \neq \omega'_k)$.

Separation of hypotheses. Using the VG bound, we want to show that $||p_{jn} - p_{kn}||_2 \geq 2s$ for s specified below, $j, k \in \{0, \ldots, M\}$, $j \neq k$.

We have

$$\begin{aligned} ||p_{jn} - p_{kn}||_2^2 &= ||p_{\boldsymbol{\omega}^{(j)}} - p_{\boldsymbol{\omega}^{(k)}}||_2^2 \\ &= L^2 h_n^{2\beta+d} ||K||_2^2\, \rho(\boldsymbol{\omega}^{(j)}, \boldsymbol{\omega}^{(k)}) \\ &\geq L^2 h_n^{2\beta+d} ||K||_2^2\, \frac{m^d}{2 \cdot 8} = (L/4)^2 h_n^{2\beta+d} ||K||_2^2\, m^d \\ &= (L/4)^2 h_n^{2\beta+d-d} ||K||_2^2 = (L/4)^2 ||K||_2^2\, m^{-2\beta} \end{aligned}$$

for $m^d \geq 8$. Now, for $n \geq n_*$ where $n_* = (7/c_0)^{2\beta+d}$, one has $m \geq 8$ hence also $m^d \geq 8$; and further $m^{2\beta} \leq (1 + 1/7)^{2\beta} c_0^{2\beta} n^{\frac{2\beta}{2\beta+d}} \leq (2c_0)^{2\beta} n^{\frac{2\beta}{2\beta+d}}$, i.e., $m^{-2\beta} \geq (2c_0)^{-2\beta} n^{-\frac{2\beta}{2\beta+d}}$, implying $||p_{jn} - p_{kn}||_2 \geq 2s$ with

$$s = An^{-\frac{\beta}{2\beta+d}}, \qquad A = \frac{L}{8} ||K||_2 (2c_0)^{-\beta}.$$

KL divergence. Because $1 + z \leq e^z$ for $z \in \mathbb{R}$, it holds also $\log(1+z) \leq z$, $z \in \mathbb{R}$. From the construction of the VG bound, $\boldsymbol{\omega}^{(0)} = \mathbf{0}$, and therefore

[2] This is possible due to the Varshamov-Gilbert bound, see [Tsybakov, 2009, Lemma 2.9, p. 104], and replace m by m^d.

$p_{0n} = \varphi$. We have

$$\begin{aligned} KL(p_{jn}, p_{j0}) &= \int p_{jn} \log \frac{p_{jn}(\boldsymbol{x})}{p_{0n}(\boldsymbol{x})} d\boldsymbol{x} = \int p_{jn} \left(\log \frac{\varphi(\boldsymbol{x}) + \sum_{k=1}^{m^d} \omega_k^{(j)} g_{\boldsymbol{k}}(\boldsymbol{x})}{\varphi(\boldsymbol{x})} \right) d\boldsymbol{x} \\ &\leq \int \left(\varphi(\boldsymbol{x}) + \sum_{k=1}^{m^d} \omega_k^{(j)} g_{\boldsymbol{k}}(\boldsymbol{x}) \right) \frac{\sum_{k=1}^{m^d} \omega_k^{(j)} g_{\boldsymbol{k}}(\boldsymbol{x})}{\varphi(\boldsymbol{x})} \, d\boldsymbol{x} \; = \\ &\int_{[0,1]^d} \frac{(\sum_{k=1}^{m^d} \omega_k^{(j)} g_{\boldsymbol{k}}(\boldsymbol{x}))^2}{\varphi(\boldsymbol{x})} \, d\boldsymbol{x} \; \leq \; \varphi^{-1}(1) \, ||g_{\mathbf{1}}||_2^2 \sum_{k=1}^{m^d} \omega_k^{(j)} \\ &\leq \underbrace{\varphi^{-1}(1) \, L^2 ||K||_2^2}_{=C} \, h_n^{2\beta+d} \, m^d = C \, h_n^{2\beta+d} \, m^d \leq C \, n h_n^{2\beta+d} \, m^d. \end{aligned}$$

As $nh^{2\beta+d} \leq c_0^{-(2\beta+d)}$, it follows that $KL(p_{jn}, p_{j0}) \leq C \, c_0^{-(2\beta+d)} \, m^d$. The VG bound ensures that $M \geq 2^{m^d/8}$, i.e., $m^d \leq 8 \log M / \log 2$. Hence setting

$$c_0 = \left(\frac{8C}{\alpha \log 2} \right)^{\frac{1}{2\beta+d}} = \left(\frac{8 \, \varphi^{-1}(1) \, L^2 ||K||_2^2}{\alpha \log 2} \right)^{\frac{1}{2\beta+d}}$$

provides that $KL(p_{jn}, p_{j0}) \leq \alpha \log M$. Specifically, c_0 can be chosen in such a way that $\alpha \in (0, 1/8)$.

Now, the statement of the theorem follows as the consequence of Theorem 2.5 (or Theorem 2.7) in [Tsybakov, 2009]. □

3.3 Sobolev character of filtering density

Here we deal with the substantial assumption in Theorem 3.1 that requires $D^\alpha p_t \in \mathcal{P}^\infty_{S(\beta, L_{t,\alpha})}$, i.e., that the derivatives of the filtering density are β-Sobolev in terms of inequality (2.33). This might be directly verified for p_0, but a straightforward verification for higher time instants $t > 1$ is inconvenient. Here we present a tool for making this more comfortable. It corresponds to a condition on the Fourier transform of the transition kernels in the signal process. This condition then ensures persistence of the Sobolev character as required. The theorem bellow is an extension of Theorem 5.2 in Coufal [2016].

To present the theorem let us recall the prediction and update formulas (2.16) and (2.17), respectively, describing the evolution of the filtering density over time. The equations can be written in a more concise form as

$$\overline{p}_t(\boldsymbol{x}_t) = \int K_{t-1}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) p_{t-1}(\boldsymbol{x}_{t-1}) \, d\boldsymbol{x}_{t-1}, \quad p_t(\boldsymbol{x}_t) = \frac{g_t(\boldsymbol{x}_t) \overline{p}_t(\boldsymbol{x}_t)}{\overline{\pi}_t g_t}.$$

In the formulas, $\overline{p}_t(\boldsymbol{x}_t)$ is the abbreviation for the density of the prediction distribution, i.e., $\overline{p}_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$, $g_t(\boldsymbol{x}_t)$ is the shortcut for the conditional density $g_t(\boldsymbol{y}_t|\boldsymbol{x}_t)$ and $\overline{\pi}_t g_t = \int g_t(\boldsymbol{x}_t)\overline{p}_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t$.

Theorem 3.3. *In the filtering problem, let $D^\alpha p_0 \in \mathcal{P}^\infty_{S(\beta,L_{0,\alpha})}$. Let $\{K_{t-1}, t \in \mathbb{N}\}$ be the set of the transition kernels, and $\{D^\alpha K_{t-1}, t \in \mathbb{N}\}$ be the set of its partial derivatives. Let $\{\mathcal{F}[D^\alpha K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1}), t \in \mathbb{N}\}$ be the set of the corresponding conditional Fourier transforms, i.e.,*

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

For all $t \in \mathbb{N}$, let $\mathcal{F}[D^\alpha K_{t-1}]$ be bounded by some function $K_b^\alpha : \mathbb{R}^d \to \mathbb{C}$ in such a way that for any $\boldsymbol{x}_{t-1} \in \mathbb{R}^d$ and $\boldsymbol{\omega} \in \mathbb{R}^d$,

$$|\mathcal{F}[D^\alpha K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})| \le |K_b^\alpha(\boldsymbol{\omega})|.$$

Let K_b^α be β-Sobolev for some $L_{K_b^\alpha} > 0$, i.e., $K_b^\alpha \in \mathcal{P}^\infty_{S(\beta,L_{K_b^\alpha})}$. Then $D^\alpha p_t \in \mathcal{P}^\infty_{S(\beta,L_{t,\alpha})}$, $t \in \mathbb{N}$ with the recurrence for $L_{t,\alpha}$ written as $L_{t,\alpha} = ||g_t||_\infty L_{K_b^\alpha}/\overline{\pi}_t g_t$.

Proof. The theorem holds for $D^\alpha p_0$ by the assumption. From the prediction formula, multiplying both sides of the prediction formula by the complex exponential gives

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

By integration, the left-hand side translates to 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

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

Multiplying by $(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d})$ we move both sides to the Fourier transforms of the corresponding partial derivatives. That is,

$$(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d})\overline{\psi}_t(\boldsymbol{\omega}) = (+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1} \cdot \cdots \cdot \omega_d^{\alpha_d})\mathcal{F}[\overline{p}_t] = \mathcal{F}[D^\alpha \overline{p}_t]$$

and

$$\int p_{t-1}(\boldsymbol{x}_{t-1})(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1}\cdot\dots\cdot\omega_d^{\alpha_d})\mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1} = \\ = \int p_{t-1}(\boldsymbol{x}_{t-1})\mathcal{F}[D^\alpha K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}.$$

Multiplication by the corresponding complex conjugates gives the expression

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

The Jensen's inequality and the boundedness of $\mathcal{F}[D^\alpha K_{t-1}]$ further gives

$$\begin{aligned} |\mathcal{F}[D^\alpha \overline{p}_t]|^2 &\leq \left(\int |D^\alpha \mathcal{F}[K_{t-1}](\boldsymbol{\omega}|\boldsymbol{x}_{t-1})|\, p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2 \\ &\leq \left(|K_b^\alpha(\boldsymbol{\omega})|\int p_{t-1}(\boldsymbol{x}_{t-1})\,d\boldsymbol{x}_{t-1}\right)^2 = |K_b^\alpha(\boldsymbol{\omega})|^2. \end{aligned}$$

Thus,

$$\int ||\boldsymbol{\omega}||^{2\beta}|D^\alpha \mathcal{F}[\overline{p}_t]|^2\,d\boldsymbol{\omega} \leq \int ||\boldsymbol{\omega}||^{2\beta}|K_b^\alpha(\boldsymbol{\omega})|^2 \leq (2\pi)^d L^2_{K_b^\alpha}.$$

The above formula shows that $D^\alpha \overline{p}_t \in \mathcal{P}^\infty_{S(\beta, L_{K_b^\alpha})}$ for any $t \in \mathbb{N}$. We proceed with specifying the Sobolev constant $L_{t,\alpha}$ for the partial derivative $D^\alpha p_t$.

The g_t function reads $g_t(\boldsymbol{x}_t) = g_t(\boldsymbol{y}_t|\boldsymbol{x}_t) = g_t^v(\boldsymbol{y}_t - h(\boldsymbol{x}_t))$. As the densities of the noise terms g_t^v are considered bounded and strictly positive in Section 2.4.1, we have $\sup_{\boldsymbol{x}_t,\boldsymbol{y}_t}\{g_t^v(\boldsymbol{y}_t - h(\boldsymbol{x}_t))\} = ||g_t||_\infty < \infty$ and $0 < \overline{\pi}_t g_t < \infty$.

Again, multiplying the update formula by the complex exponential, integrating, multiplying by $(+\mathrm{i})^{|\alpha|}(\omega_{i_1}^{\alpha_1}\cdot\dots\cdot\omega_{i_d}^{\alpha_d})$ and the respective conjugates we move to the Fourier transforms of the partial derivatives and get

$$\begin{aligned} (\overline{\pi}_t g_t)\, p_t(\boldsymbol{x}_t) &= g_t(\boldsymbol{x}_t)\,\overline{p}_t(\boldsymbol{x}_t), \\ (\overline{\pi}_t g_t)\int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle} p_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t &= \int e^{\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{x}_t\rangle} g_t(\boldsymbol{x}_t)\,\overline{p}_t(\boldsymbol{x}_t)\,d\boldsymbol{x}_t, \\ (\overline{\pi}_t g_t)\,(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1}\cdot\dots\cdot\omega_d^{\alpha_d})\,\psi_t(\boldsymbol{\omega}) &\leq ||g_t||_\infty\,(+\mathrm{i})^{|\alpha|}(\omega_1^{\alpha_1}\cdot\dots\cdot\omega_d^{\alpha_d})\,\overline{\psi}_t(\boldsymbol{\omega}), \\ ||\boldsymbol{\omega}||^{2\beta}\,(\overline{\pi}_t g_t)^2\,|\mathcal{F}[D^\alpha p_t]|^2 &\leq ||\boldsymbol{\omega}||^{2\beta}\,||g_t||^2_\infty\,|\mathcal{F}[D^\alpha \overline{p}_t]|^2, \\ (2\pi)^{-d}\int ||\boldsymbol{\omega}||^{2\beta}|\mathcal{F}[D^\alpha p_t]|^2\,d\boldsymbol{\omega} &\leq \frac{||g_t||^2_\infty L^2_{K_{b,\alpha}}}{(\overline{\pi}_t g_t)^2} = L^2_{t,\alpha}, \end{aligned}$$

which concludes the proof. □

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

4. Designing kernels

When using kernel estimates in particle filtering, selecting the kernel and the number of particles is left to the user. This chapter deals with the first option as computing power of used hardware affect mostly the second. Theorem 3.1 tells us that one should match the order of kernel with the Sobolev character of the filtering density to enjoy the presented convergence rates. The Sobolev character can be assured using Theorem 3.2. So the natural question arise - "How to design/select a kernel of the given order in the given dimension?"

The chapter consists of two sections. The first section addresses kernels from the dimensionality point of view. That is, we deal with the dimension d in specification of the kernel, $K : \mathbb{R}^d \to \mathbb{R}$. In the context of particle filtering, this is the dimension of the signal process state space. There are various one-dimensional kernels proposed to use for kernel estimation [Silverman, 1986, Wand and Jones, 1995, Tsybakov, 2009]. When moving into multiple dimensions there are two common approaches used for designing multivariate kernels on the basis of one-dimensional functions. They lead to constructing the *product* and *radial kernels*, respectively [Wand and Jones, 1995]. We discuss how to compute their Fourier transform as this is important for investigating the orders of these kernels.

The second section covers designing the kernels of the given order in the given dimension. We discuss designing of univariate kernels in both unbounded and bounded domains as well as an extension to the multivariate case. Namely, we will deal with the kernels supported on the unit disc and unit ball in the multiple dimensions. The corresponding Fourier transforms are discussed and presented as well.

4.1 Multivariate kernels

In this section we mention several basic univariate kernels which are commonly used in kernel density estimation. Further, we discuss constructing the product and radial kernels together with computing their multivariate Fourier transforms.

4.1.1 Product kernels

Various univariate kernels for use in kernel density estimation have been proposed and discussed in literature [Silverman, 1986, Tsybakov, 2009, p. 43, p. 3, resp.]. Several of them are listed in Tab. 4.1 together with their univariate Fourier transforms. The listed kernels are even functions centered at origin. Hence the related Fourier transforms are real.

kernel	$K(u)$	Fourier transform $\mathcal{F}[K](\omega)$
Gaussian	$K(u) = \exp(-u^2)$	$\mathcal{F}[K](\omega) = \sqrt{\pi}\exp\left[-\frac{\omega^2}{4}\right]$
Epanechnikov	$K(u) = \frac{3}{4}(1-u^2)_+$	$\mathcal{F}[K](\omega) = \frac{3[\sin(\omega)-\omega\cos(\omega)]}{\omega^3}$, $\mathcal{F}[K](0) = 1$
biweight	$K(u) = \frac{15}{16}(1-u^2)_+^2$	$\mathcal{F}[K](\omega) = \frac{15\,[(3-\omega^2)\sin(\omega)-3\,\omega\cos(\omega)]}{\omega^5}$, $\mathcal{F}[K](0) = 1$
rectangular	$K(u) = \begin{cases} 1 & 0 \le \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$	$\mathcal{F}[K](\omega) = 2\,\mathrm{sinc}(\omega)$
sinc	$K(u) = \mathrm{sinc}(x)$	$\mathcal{F}[K](\omega) = \pi \cdot 1_{[\lvert x\rvert \le 1]}$

Table 4.1: Univariate kernels and their Fourier transforms.

In the context of particle filtering, one needs to have an adjustable width of kernels. This is achieved by scaling the argument by the width parameter $b > 0$. That is, we consider the kernels K_b with their appropriate Fourier transforms, both written as

$$K_b(u) = K(u/b), \quad \mathcal{F}[K_b](\omega) = b\mathcal{F}[K](b\omega).$$

The product kernels are constructed by multiplying copies of univariate kernels along the individual dimensions of the input space. That is

$$K_b(\boldsymbol{u}) = \prod_{i=1}^{d} K_b(u_i).$$

The multiplicative form induces the separability of variables that makes the multivariate Fourier transform being the product of the one-dimensional Fourier transforms of the original univariate functions. Therefore the properties of the Fourier transform for multivariate product kernels are derived from the univariate counterparts, i.e.,

$$\mathcal{F}[K_b](\boldsymbol{\omega}) = \prod_{i=1}^{d} \mathcal{F}[K_b](\omega_i).$$

It is a basic fact that the product of univariate Gaussians gives a multivariate Gaussian, which is a radial function invariant w.r.t. rotations. This is due to the fact that the multivariate Gaussian's formula incorporate the Euclidean norm that transforms multidimensional inputs to a single dimension. This property is a natural choice for specifying general radial kernels.

4.1.2 Radial kernels

The radial kernels have the form of a radial function. That is

$$K(\boldsymbol{u}) = \varphi(||\boldsymbol{u}||_2)$$

where $\varphi : [0, \infty) \to \mathbb{R}$ is an univariate function, which usually has the property $\lim_{r\to\infty} \varphi(r) = 0$, and $||\cdot||_2$ is the Euclidean norm. Generally, other norms can be considered, but in the context of this thesis the Euclidean norm is exclusive because of the characterization of the radial function using the Euclidean norm in terms of the Fourier transform as presented in Section 2.3.

For the reader's convenience we recall the respective theorem stating that the Fourier transform of a d-variate radial function can be computed using the Hankel transform of the related univariate function.

That is, for a continuous radial function $\Phi(\boldsymbol{x}) = \varphi(||\boldsymbol{x}||_2) \in L^1(\mathbb{R}^d)$ its Fourier transform writes $\mathcal{F}[\Phi](\boldsymbol{\omega}) = \varphi^{\mathscr{H}}(||\boldsymbol{\omega}||_2)$ where

$$\varphi^{\mathscr{H}}(s) = \frac{(2\pi)^{d/2}}{\sqrt{s^{d-2}}} \int_0^\infty \varphi(r) r^{\frac{d}{2}} J_{(d-2)/2}(sr)\, dr = (2\pi)^{d/2} s^{-\nu} \mathscr{H}_\nu\{\varphi(r)\cdot r^\nu\}(s) \tag{4.1}$$

for $\nu = (d-2)/2$, i.e., $\nu = -\frac{1}{2}, 0, \frac{1}{2}, 1, \ldots$ for $d = 1, 2, 3, 4 \ldots$.

The formula gives a convenient tool for specifying the Fourier transforms of the multivariate radial kernels via the Hankel transforms of the related univariate functions multiplied by term r^ν.

In Table 4.2, there are presented several univariate functions with their Hankel transforms. They contain special functions, namely the Bessel functions of the first kind J_ν. Several sources providing tables of the Hankel transform pairs are available, such as [A. D. Poularikas (Ed.), 2000] or [Debnath and Bhatta, 2007]. The comprehensive source is [Bateman, 1954]. However, in [Bateman, 1954], the alternative definition of the Hankel transform is used: $h_\nu\{f(x)\}(y) = \int_0^\infty f(x) J_\nu(xy)(xy)^{1/2}\, dx$. The relation between both versions writes $\mathscr{H}_\nu\{f(x)\, x^\nu\}(s) = s^{-1/2} h_\nu\{x^{\nu+1/2} f(x)\}(s)$.

function name	univariate function $\varphi(r)$	Hankel transform $\mathscr{H}_\nu\{\varphi(r)\, r^\nu\}$	Btm. Vol. II
Gaussian	$\varphi(r) = \exp(-a^2 r^2)$	$H_\nu(s) = \frac{s^\nu}{(2a^2)^{\nu+1}} \exp(-\frac{s^2}{4a^2})$	p. 29 (10)
cut power	$\varphi(r) = (a^2 - r^2)_+^\mu$	$H_\nu(s) = \frac{2^\mu \Gamma(\mu+1) a^{\nu+\mu+1}}{s^{\mu+1}} J_{\nu+\mu+1}(as)$	p. 26 (33)
rectangular pulse	$\varphi(r) = \begin{cases} 1 & 0 \le r \le 1 \\ 0 & \text{otherwise} \end{cases}$	$H_\nu(s) = s^{-1} J_{\nu+1}(s)$	p. 22 (6)
jinc_ν	$\varphi(r) = r^{-(\nu+1)} J_{\nu+1}(ar)$	$H_\nu(s) = \frac{s^\nu}{a^{\nu+1}} \cdot 1_{[0,a]}(s)$	p. 48 (7)

Table 4.2: Hankel transforms of univariate functions $\varphi(r)\, r^\nu$.

function name	scaled function $\varphi_b(r) = \varphi(r/b)$	Hankel transform $\mathscr{H}_\nu\{\varphi_b(r)\, r^\nu\}$
Gaussian	$\varphi_b(r) = \exp(-(r/b)^2)$	$H_\nu(s) = \frac{b^{2(\nu+1)} s^\nu}{2^{\nu+1}} \exp(-\frac{1}{4}(bs)^2)$
cut power	$\varphi_b(r) = (a^2 - (r/b)^2)_+^\mu$	$H_\nu(s) = \frac{b^{\nu-\mu+1} 2^\mu \Gamma(\mu+1) a^{\nu+\mu+1}}{s^{\mu+1}} \cdot \cdot J_{\nu+\mu+1}(abs)$
rectangular pulse	$\varphi_b(r) = \begin{cases} 1 & 0 \le r/b \le 1 \\ 0 & \text{otherwise} \end{cases}$	$H_\nu(s) = b^{\nu+1} s^{-1} J_{\nu+1}(bs)$
jinc_ν	$\varphi_b(r) = (r/b)^{-(\nu+1)} J_{\nu+1}(r/b)$	$H_\nu(s) = b^{2(\nu+1)} s^\nu \cdot 1_{[0,1]}(bs)$

Table 4.3: Hankel transforms of scaled univariate functions $\varphi(r/b)\, r^\nu$.

It is known that the Hankel transform is its self-inverse. Let us check this for the rectangular and the generalized jinc functions [Weisstein, 2002, p. 1579]. If one is interested in $\mathscr{H}_\nu\{r^{-1} J_{\nu+1}(r)\}$, then the last row of Tab. 4.2 gives $\mathscr{H}_\nu\{r^{-(\nu+1)} J_{\nu+1}(r)\, r^\nu\} = 1_{[0,1]}(r) \cdot r^\nu$ for $a = 1$, which is indeed $\varphi(r)$ of the third row multiplied by r^ν. Note that the standard jinc function reads as $\text{jinc}(x) = x^{-1} J_1(x)$.

To compute the Fourier transforms of the radial kernels with adjustable widths we scale the argument of functions in Table 4.2 by the parameter $b > 0$ and update the presented Hankel transforms accordingly. To do so note that $\mathscr{H}_\nu\{f(r/b) r^\nu\}(s) = b^{\nu+2} \mathscr{H}_\nu\{f(r) r^\nu\}(bs)$, which can be directly derived from the definition formula of the Hankel transform. Moreover, $b^{2(\nu+1)} = b^d$ and $b^{\nu+1} = b^{d/2}$. The updated formulas are presented in Table 4.3.

In Table 4.4, we have computed the Fourier transforms of the scaled multivariate radial functions $\Phi(\boldsymbol{x}) = \varphi_b(||\boldsymbol{x}||_2) = \varphi(||\boldsymbol{x}||_2/b)$ for the univariate functions φ_b of Table 4.3. The Fourier transforms were computed following formula (4.1), i.e., by multiplying by factor $(2\pi)^{d/2} s^{-\nu}$, with the order of the Hankel transform set to $\nu = (d-2)/2 = d/2 - 1$ where $d \in \mathbb{N}$ is the dimension.

To check the computations of transforms in Table 4.4, we computed their univariate variants for the specific setting of parameters μ, a, b so that we match the transforms in Table 4.1. The results are presented in Table 4.5.

function name	multivariable expression	d-variate Fourier transform
Gaussian	$\varphi_b(\|\|\boldsymbol{x}\|\|_2) = \exp(-\|\|\boldsymbol{x}\|\|_2^2/b^2)$	$\mathcal{F}[\varphi_b](\boldsymbol{\omega}) = b^d\pi^{d/2}\exp(-\frac{b^2}{4}\|\|\boldsymbol{\omega}\|\|_2^2)$
cut power	$\varphi_b(\|\|\boldsymbol{x}\|\|_2) = (a^2 - \|\|\boldsymbol{x}\|\|_2^2/b^2)_+^\mu$	$\mathcal{F}[\varphi_b](\boldsymbol{\omega}) = \frac{(2\pi b)^{d/2}(2/b)^\mu\Gamma(\mu+1)\,a^{d/2+\mu}}{\|\|\boldsymbol{\omega}\|\|_2^{d/2+\mu}} \cdot$ $\cdot\, J_{d/2+\mu}(a\|\|\boldsymbol{\omega}\|\|_2)$
b - ball in $\mathbb{R}^d$	$\varphi_b(\|\|\boldsymbol{x}\|\|_2) = \begin{cases} 1 & 0 \le \|\|\boldsymbol{x}\|\|_2/b \le 1 \\ 0 & \text{otherwise} \end{cases}$	$\mathcal{F}[\varphi_b](\boldsymbol{\omega}) = (2\pi b/\|\|\boldsymbol{\omega}\|\|_2)^{d/2}J_{d/2}(b\|\|\boldsymbol{\omega}\|\|_2)$
jinc_ν	$\varphi_b(\|\|\boldsymbol{x}\|\|_2) = (\|\|\boldsymbol{x}\|\|_2/b)^{-(\nu+1)} \cdot$ $\cdot\, J_{\nu+1}(\|\|\boldsymbol{\omega}\|\|_2/b)$	$\mathcal{F}[\varphi_b](\boldsymbol{\omega}) = b^d(2\pi)^{d/2} \cdot 1_{[0,1]}(b\|\|\boldsymbol{\omega}\|\|_2)$

Table 4.4: Scaled multivariable radial functions and their Fourier transforms.

function name	univariate expression	1D-Fourier transform
Gaussian	$\varphi_{b=1}(\|x\|) = \exp(-x^2)$	$\mathcal{F}[\varphi_1](\omega) = \pi^{1/2}\exp(-\frac{\omega^2}{4})$
cut power - Epanechnikov	$\varphi_{b=1}(\|x\|) = \frac{3}{4}(1-\|x\|^2)_+^1$	$\mathcal{F}[\varphi_1](\omega) = \frac{3}{4}\frac{(2\pi)^{1/2}2^1\Gamma(2)}{\|\omega\|^{3/2}} \cdot J_{3/2}(\|\omega\|)$
cut power - biweight	$\varphi_{b=1}(\|x\|) = \frac{15}{16}(1-\|x\|^2)_+^2$	$\mathcal{F}[\varphi_1](\omega) = \frac{15}{16}\frac{(2\pi)^{1/2}2^2\Gamma(3)}{\|\omega\|^{5/2}} \cdot J_{5/2}(\|\omega\|)$
rectangular pulse	$\varphi_{b=1}(\|x\|) = \begin{cases} 1 & 0 \le \|x\| \le 1 \\ 0 & \text{otherwise} \end{cases}$	$\mathcal{F}[\varphi_b](\omega) = (2\pi/\|\omega\|)^{1/2} \cdot J_{1/2}(\|\omega\|)$
sinc	$\text{sinc}(x) = (\pi/2)^{1/2}(\|x\|)^{-1/2} \cdot$ $\cdot\, J_{1/2}(\|x\|)$	$\mathcal{F}[\varphi_1](\omega) = \pi \cdot 1_{[0,1]}(\|\omega\|)$

Table 4.5: Univariate radial functions and their Fourier transforms.

We see that the formula for Gaussian in Tab. 4.5 matches the one presented in Tab 4.1. Situation gets a bit complicated for cut power. The Epanechnikov kernel corresponds to the choice $\mu = 1$, $a = 1$, $b = 1$ and multiplication by factor $3/4$. From linearity of the Fourier transform and Table 4.4 one gets for $d = 1$ the expression $\mathcal{F}[\text{epa}](\omega) = \frac{3}{2}\frac{\sqrt{2\pi}}{|\omega|^{3/2}}J_{3/2}(|\omega|)$. Note that $\Gamma(2) = 1$. The Bessel function of the first kind $J_{3/2}$ admits the closed representation $J_{3/2}(|s|) = \sqrt{\frac{2}{\pi}}\frac{\sin(|\omega|)-|\omega|\cos(|\omega|)}{|\omega|^{3/2}}$, which together gives $\mathcal{F}[\text{epa}](\omega) = \frac{3}{\sqrt{2\pi}}\frac{\sin(|\omega|)-|\omega|\cos(|\omega|)}{|\omega|^3}$. Additionally, one can rid off the absolute value because the Fourier transform in discussion is an even function, i.e., $\mathcal{F}[\text{epa}](|\omega|) = \mathcal{F}[\text{epa}](\omega)$. An analogous discussion holds for the biweight kernel.

The univariate jinc function reads

$$\text{jinc}_{\nu=-1/2}(|x|) = |x|^{-1/2}J_{1/2}(|x|) = (\pi/2)^{-1/2}j_0(|x|)$$

The spherical Bessel function of order $\nu = 0$ corresponds to the sinc function, i.e., $j_0(|x|) = \sin(x)/x$, $j_0(0) = 1$. Hence the Fourier image of $(\pi/2)^{1/2}\text{jinc}_{\nu=-1/2}$ gives the Fourier transform of the sinc function and vice-versa.

Speaking about the orders of the presented kernels, a direct evaluation of the Fourier transforms and their partial derivatives at origin could give us the answer, however, it is inconvenient to do so directly. In the following section, we present an easy to use criterion for distinguishing the first order kernels from the higher order ones and how to design kernels of the given order in the given dimension.

4.2 Higher order kernels

Higher order kernels enjoy faster convergence rates when estimating the higher order Sobolev filtering densities. The announced criterion corresponds to the fact that for a kernel K of order ℓ one has

$$\int \boldsymbol{u}^{\alpha} K(\boldsymbol{u})\, d\boldsymbol{u} = \int u_1^{\alpha_1} \cdots u_d^{\alpha_d} K(\boldsymbol{u})\, d\boldsymbol{u} = 0$$

for any multi-index $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq \ell$. Thus for $\ell = 2$ one gets expressions such as $\int u_i^2 K(\boldsymbol{u}) = 0$, $i = 1, \ldots, d$ which hold if and only if $K(\boldsymbol{u}) \not\equiv 0$ takes also negative values on a set of positive Lebesgue measure. So this gives us immediately that all the presented kernels are (after proper scaling) of order $\ell = 1$, except the generalized jinc function.

At first sight, it may looks strange to use kernels which are negative at certain parts of its domain with no guarantee that kernel estimate takes only positive values as we are estimating densities. This issue is addressed in [Tsybakov, 2009, p. 10]. Let us cite from the related paragraph.

The estimators based on higher order kernels can also take negative values. This property is sometimes emphasized as a drawback of estimators with higher order kernels, since the density p itself is nonnegative. However, this remark is of minor importance because we can always use the positive part estimator $\hat{p}^+ = \max\{0, \hat{p}(x)\}$ whose risk is smaller than or equal to the risk of $\hat{p}$, i.e., (adapted for our case)

$$\mathbb{E} \int (\hat{p}_n^+(\boldsymbol{x}) - p(\boldsymbol{x}))^2\, d\boldsymbol{x} \;\leq\; \mathbb{E} \int (\hat{p}_n(\boldsymbol{x}) - p(\boldsymbol{x}))^2\, d\boldsymbol{x}.$$

Tsybakov also suggests the technique for constructing kernels of specific order $\ell \in \mathbb{N}$. It draws on orthogonal and consequently (after proper scaling) *orthonormal* polynomials. Inspecting Section 1.2.2 of [Tsybakov, 2009], one can state the following lemma that holds for multivariate kernels.

Lemma 4.1. *Let* $\{P_{\alpha^{(j)}}^{d,n} : |\alpha| = n\}_{j=0}^{N^{d,n}}$, *be an orthonormal basis of the space of orthogonal polynomials of degree* $n \in \mathbb{N}_0$ *with respect to the weight function* W_μ *on* $\mathbb{R}^d$, *i.e., the basis of* $\mathcal{V}_n^d(W_\mu)$. *Then for any* $\ell \in \mathbb{N}$ *the function* $K_\ell^d : \mathbb{R}^d \to \mathbb{R}$ *defined by formula*

$$K_\ell^d(\boldsymbol{u}) = \sum_{j=0}^{N^{d,\ell}} P_{\alpha^{(j)}}^{d,\ell}(\mathbf{0}) P_{\alpha^{(j)}}^{d,\ell}(\boldsymbol{u}) W_\mu(\boldsymbol{u}) \tag{4.2}$$

is a d-variate kernel of order ℓ.

Proof. We just recast the univariate version of the proof of Proposition 1.3 in [Tsybakov, 2009]. As $\{P_{\alpha^{(j)}}^{d,n} : |\alpha| = n\}_{j=0}^{N^{d,n}}$ is some basis of $\mathcal{V}_n^d(W_\mu)$, one has from the properties of the $S_n(W_\mu, \boldsymbol{u}^\alpha)$ operator, see Section 2.1.8, that any monomial $\boldsymbol{u}^\alpha$ for $\alpha \in \mathbb{N}_0^d$ that $|\alpha| = n$, $n \leq \ell$ can be expressed as

$$\boldsymbol{u}^\alpha = \sum_{j=0}^{N_{d,n}} b_j^\alpha P_{\alpha^{(j)}}^{d,n}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d. \tag{4.3}$$

Clearly, b_j^α are the coefficients of the linear combination for expressing $\boldsymbol{u}^\alpha$ in the basis $\{P_{\alpha^{(j)}}^{d,n} : |\alpha| = n\}_{j=0}^{N^{d,n}}$.

Now, considering K_ℓ^d of (4.2) and decomposition (4.3) one gets

$$\begin{aligned}
\int \boldsymbol{u}^\alpha K_\ell^d(\boldsymbol{u})\, d\boldsymbol{u} &= \sum_{j=0}^{N^{d,n}} \sum_{k=0}^{N^{d,\ell}} \int b_j^\alpha P_{\alpha^{(j)}}^{d,n}(\boldsymbol{u}) P_{\alpha^{(k)}}^{d,\ell}(\mathbf{0}) P_{\alpha^{(k)}}^{d,\ell}(\boldsymbol{u}) W_\mu(\boldsymbol{u})\, d\boldsymbol{u} \\
&= \begin{cases} 0 & \text{for} \quad n < \ell, \\ \sum_{j=0}^{N_{d,\ell}} b_j^\alpha P_{\alpha^{(j)}}^{d,0}(\mathbf{0}) = 1 & \text{for} \quad n = 0, \\ \sum_{j=0}^{N_{d,\ell}} b_j^\alpha P_{\alpha^{(j)}}^{d,\ell=n}(\mathbf{0}) = 0 & \text{for} \quad n = \ell > 0, \end{cases}
\end{aligned}$$

Indeed, if $n = |\alpha| = 0$, then the left side of (4.3) writes $\boldsymbol{u}^0 = 1$, and for $n = |\alpha| > 0$ one has $\mathbf{0}^\alpha = 0$. Using the above result, we have

$$D^\alpha \mathcal{F}[K](\mathbf{0}) = \int \boldsymbol{u}^\alpha K(\boldsymbol{u})\, d\boldsymbol{u} = \begin{cases} 1 & \text{for} \quad |\alpha| = 0, \\ 0 & \text{for} \quad |\alpha| = 1, \ldots, \ell, \end{cases}$$

which proves that K is the d-variate kernel of order ℓ. □

In what follows, we use the different families of the orthogonal polynomials recalled in Chapter 2 to design the multivariate kernels following Lemma 4.1.

4.2.1 Hermite kernels

Here we explicitly construct higer order multivariate kernels for the Hermite basis. We start with the higher order 1D kernels. As for the Hermite orthogonal polynomials one has $H_n(0) = 0$ for n even, it is enough to consider ℓ even in formula 4.3. Let us explicitly constuct 1D-kernels of the second and fourth order.

$H_n(x)$	$H_n(0)$	$h_n = \sqrt{\pi}\, 2^n n!$
$H_0(x) = 1$	$H_0(0) = 1$	$h_0 = \sqrt{\pi}$
$H_1(x) = 2x$	$H_1(0) = 0$	
$H_2(x) = 4x^2 - 2$	$H_2(0) = -2$	$h_2 = \sqrt{\pi}\, 8$
$H_3(x) = 8x^3 - 12x$	$H_3(0) = 0$	
$H_4(x) = 16x^4 + 8x^2 + 12$	$H_4(0) = 12$	$h_4 = \sqrt{\pi}\, 384$

Table 4.6: Hermite orthogonal polynomials.

For the second order one has

$$\begin{aligned} K_2^1(u) &= \left[\frac{1}{h_0} H_0(0)H_0(u) + \frac{1}{h_2} H_2(0)H_2(u)\right] \cdot e^{-u^2} \\ &= \frac{1}{\sqrt{\pi}} \left[1 - \frac{2}{8}(4u^2 - 2)\right] \cdot e^{-u^2} \\ &= \frac{1}{2\sqrt{\pi}} \left[3 - 2u^2\right] \cdot e^{-u^2} \end{aligned}$$

and for the fourth one

$$\begin{aligned} K_4^1(u) &= \left[\frac{1}{h_0} H_0(0)H_0(u) + \frac{1}{h_2} H_2(0)H_2(u) + \frac{1}{h_4} H_4(0)H_4(u)\right] \cdot e^{-u^2} \\ &= \frac{1}{\sqrt{\pi}} \left[1 - \frac{2}{8}(4u^2 - 2) + \frac{12}{384}(16u^4 - 48u^2 + 12)\right] \cdot e^{-u^2}. \\ &= \frac{1}{8\sqrt{\pi}} \left[4u^4 - 20u^2 + 15\right] \cdot e^{-u^2}. \end{aligned}$$

As Gaussians are eingenvalues of the Fourier transform, the Fourier transform of K_2^1 writes

$$\mathcal{F}[K_2^1](u) = \frac{1}{4} \left(\omega^2 + 4\right) \cdot e^{-\frac{\omega^2}{4}}$$

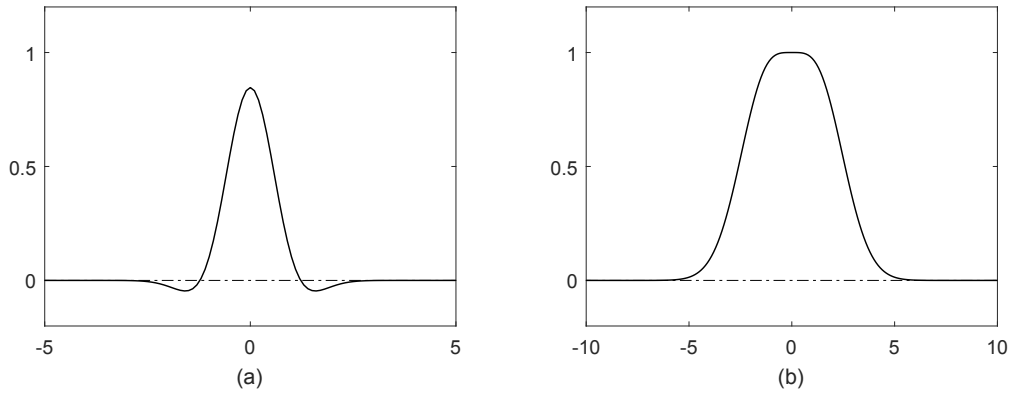

Figure 4.1: (a) 1D Hermite kernel of order $\ell = 2$ - K_2^1; (b) FT of K_2^1.

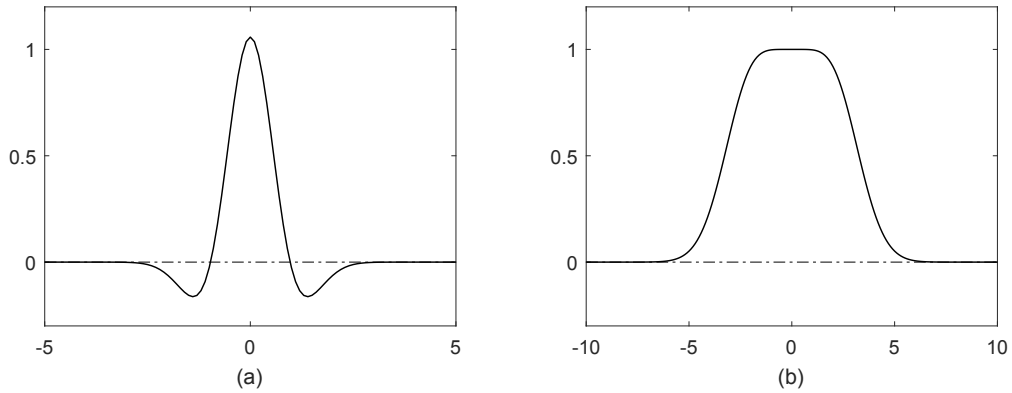

Figure 4.2: (a) 1D Hermite kernel of order $\ell = 4$ - K_4^1; (b) FT of K_4^1.

with $\frac{\mathrm{d}}{\mathrm{d}\omega}\mathcal{F}[K_2^1] = -\frac{1}{8}u^3\, e^{-\frac{\omega^2}{4}}$ and $\frac{\mathrm{d}^2}{\mathrm{d}^2\omega}\mathcal{F}[K_2^1] = \frac{1}{16}u^2(u^2-6)\, e^{-\frac{\omega^2}{4}}$. Hence, $\mathcal{F}[K_2^1](0) = 1$, $\mathcal{F}[K_2^1]'(0) = 0$ and $\mathcal{F}[K_2^1]''(0) = 0$, so K_1^2 is really of order $\ell = 2$. It is even of order $\ell = 3$, but one has $\mathcal{F}[K_2^1]^{(4)}(0) = -3/4$.

The Fourier transform of K_4^1 writes

$$\mathcal{F}[K_4^1](\omega) = \frac{1}{32}(\omega^4 + 8\omega^2 + 32) \cdot e^{-\frac{\omega^2}{4}}.$$

Clearly, $\mathcal{F}[K_4^1](0) = 1$ and $\mathcal{F}[K_4^1]^{(\ell)}(0) = 0$ up to $\ell = 4$. Thus K_4^1 is the kernel of 4-th order (in fact it is of 5-th order). Graphs of both kernels and their Fourier transforms are presented in Fig. 4.1 and Fig. 4.2, respectively.

Now let as switch to the multivariate versions, we construct the corresponding 2D-kernels.

One has

$$\begin{aligned} K_2^2(u_1,u_2) &= \frac{1}{2\sqrt{\pi}}\left[3-2u_1^2\right]\cdot e^{-u_1^2}\frac{1}{2\sqrt{\pi}}\left[3-2u_2^2\right]\cdot e^{-u_2^2} \\ &= \frac{1}{4\pi}\left[4u_1^2u_2^2-6(u_1^2+u_2^2)+9\right]\cdot e^{-||(u_1,u_2)||_2^2}. \end{aligned}$$

The corresponding Fourier transform reads as

$$\mathcal{F}[K_1^2](\omega_1,\omega_2)=\frac{1}{16}(\omega_1^2+4)(\omega_2^2+4)\cdot e^{-\frac{||(\omega_1,\omega_2)||_2^2}{4}}.$$

For the fourth order, the respecive formulas write

$$\begin{aligned} K_4^2(\boldsymbol{u}) &= \frac{1}{8^2}\left[\prod_{i=1}^{2}(4u_i^4-20u_i^2+15)\right]e^{-||\boldsymbol{u}||_2^2}, \\ \mathcal{F}[K_4^2](\boldsymbol{\omega}) &= \frac{1}{32^2}\left[\prod_{i=1}^{2}(\omega_i^4+8\omega_i^2+32)\right]e^{-\frac{||\boldsymbol{\omega}||_2^2}{4}}. \end{aligned}$$

4.2.2 Legendre kernels

The Legendre kernels are based on the Legendre polynomials. The Legendre polynomials form an orthogonal and after scaling corresponding orthonormal basis of $L^2([-1,1])$ space. Their associated weight function is the characteristic function of $[-1,1]$ interval.

$P_n(x)$	$P_n(0)$	$h_n=\frac{2}{2n+1}$
$P_0(x)=1$	$P_0(0)=1$	$h_0=2$
$P_1(x)=x$	$P_1(0)=0$	
$P_2(x)=\frac{1}{2}(3x^2-1)$	$P_2(0)=-\frac{1}{2}$	$h_2=\frac{2}{5}$
$P_3(x)=\frac{1}{2}(5x^3-3x)$	$P_3(0)=0$	
$P_4(x)=\frac{1}{8}(35x^4-30x^2+3)$	$P_4(0)=\frac{3}{8}$	$h_4=\frac{2}{9}$

Table 4.7: Legendre orthogonal polynomials.

Following the same approach as for the Hermite polynomials gives us the explict forms of the corresponding higher order univariate kernels.

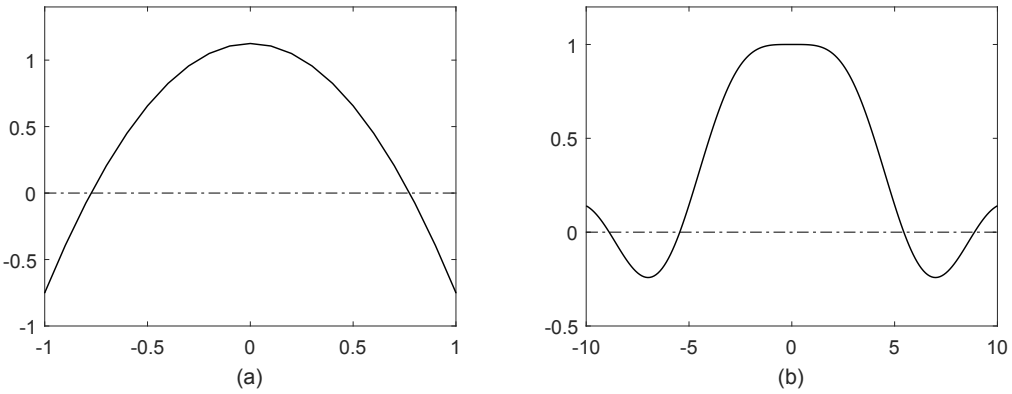

Figure 4.3: (a) 1D Legendre kernel of order $\ell = 2$ - K_2^1; (b) FT of K_2^1.

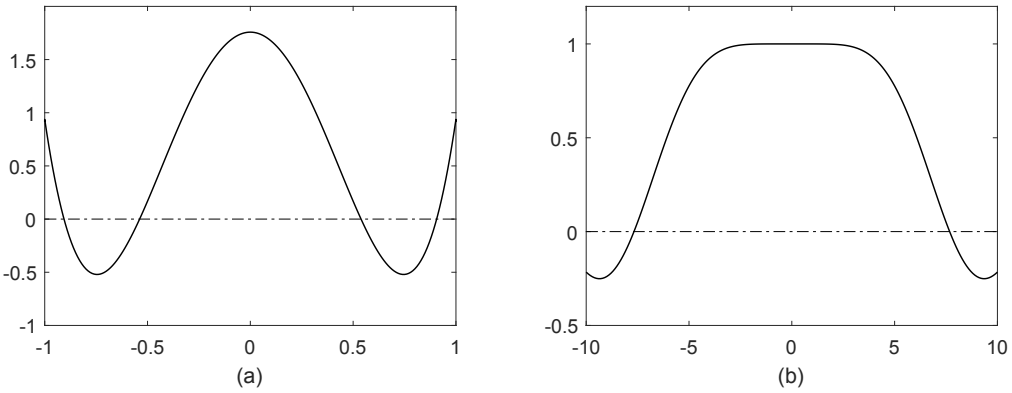

Figure 4.4: (a) 1D Legendre kernel of order $\ell = 4$ - K_4^1; (b) FT of K_4^1.

$$\begin{aligned} K_2^1(u) &= \left[\frac{1}{2} - \frac{1}{2}\frac{5}{2}\frac{1}{2}(3u^2 - 1)\right] \cdot 1_{[-1,1]}(u) \\ &= \left[\frac{9}{8} - \frac{15}{8}u^2\right] \cdot 1_{[-1,1]}(u), \end{aligned}$$

$$\begin{aligned} K_4^1(u) &= \left[\frac{1}{2} - \frac{1}{2}\frac{5}{2}\frac{1}{2}(3u^2 - 1) + \frac{3}{8}\frac{9}{2}\frac{1}{8}(35x^4 - 30x^2 + 3)\right] \cdot 1_{[-1,1]}(u) \\ &= \left[\frac{15}{128}(63x^4 - 70x^2 + 15)\right] \cdot 1_{[-1,1]}(u). \end{aligned}$$

Graphically, both kernels together with their Fourier transforms are presented in Fig. 4.3 and Fig. 4.4, respectively.

The Fourier transform of K_2^1 writes

$$\begin{aligned}\mathcal{F}[K_2^1](\omega) &= \sqrt{\frac{2\pi}{\omega}}\left[\frac{1}{2}J_{1/2}(\omega)+\frac{5}{4}J_{5/2}(\omega)\right] = \sqrt{\frac{\pi}{2\omega}}\left[J_{1/2}(\omega)+\frac{5}{2}J_{5/2}(\omega)\right] \\ &= j_0(\omega)+\frac{5}{2}j_2(\omega) = \frac{-3[(\omega^2-5)\sin(\omega)+5\omega\cos(\omega)]}{2\omega^3}\end{aligned}$$

and one has

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\omega}\mathcal{F}[K_2^1](\omega) &= \frac{9(2\omega^2-5)\sin(\omega)-3\omega(\omega^2-15)\cos(\omega)}{2\omega^4}, \\ \frac{\mathrm{d}^2}{\mathrm{d}\omega^2}\mathcal{F}[K_2^1](\omega) &= \frac{3[x(7x^2-60)\cos(\omega)+(\omega^4-27\omega^2+60)\sin(\omega)]}{2\omega^5}.\end{aligned}$$

Considering limits at $u=0$ and $\omega=0$, the above formulas admit a continuous extension and $K_2^1(0)=1$, $\mathcal{F}'[K_2^1](0)=0$, $\mathcal{F}''[K_2^1](0)=0$. Thus K_2^1 is indeed the kernel of order $\ell=2$.

For the fourth order kernel the Fourier transform writes

$$\begin{aligned}\mathcal{F}[K_4^1](u) &= \sqrt{\frac{2\pi}{\omega}}\left[\frac{1}{2}J_{1/2}(\omega)+\frac{5}{4}J_{5/2}(\omega)+\frac{27}{16}J_{9/2}(\omega)\right] \\ &= \sqrt{\frac{\pi}{2\omega}}\left[J_{1/2}(\omega)+\frac{5}{2}J_{5/2}(\omega)+\frac{27}{8}J_{9/2}(\omega)\right] \\ &= j_0(\omega)+\frac{5}{2}j_2(\omega)+\frac{27}{8}j_4(\omega).\end{aligned}$$

Again $K_4^1(0)=1$ and it can be checked that $\frac{\mathrm{d}^j}{\mathrm{d}\omega^j}\mathcal{F}[K_4^1](0)=0$ for $j=1,\ldots,4$.

Considering the multivariate product Legendre kernels, there is no special property we can use to simplify the product of univariate parts. That is why we do not present these kernels explicitly here and move directly to the radial kernels.

4.2.3 Zernike kernels

The Zernike kernels are derived from the Zernike polynomials which are orthogonal over the unit disc as described in Section 2.1.7. In order to use Lemma 4.1 we are interested in the orthonormalized products

$$\frac{1}{h_n}Z_n^m(0,0)Z_n^m(\rho,\varphi) = \frac{1}{h_n}R_n^m(0)Z_n^m(\rho,\varphi) \quad n,m\in\mathbb{N}_0, m\leq n$$

and recall, that the non-zero radial polynomial up to the fourth order read as

$R_n^m(x)$	$R_n^m(0)$	$h_n = \frac{\epsilon_m \pi}{2(n+1)}$
$R_0^0(\rho) = 1$	$R_0^0(0) = +1$	$h_0 = \pi$
$R_1^1(\rho) = \rho$	$R_1^1(0) = 0$	
$R_2^0(\rho) = 2\rho^2 - 1$	$R_2^0(0) = -1$	$h_2 = \frac{\pi}{3}$
$R_2^2(\rho) = \rho^2$	$R_2^2(0) = 0$	
$R_3^1(\rho) = 3\rho^3 - 2\rho$	$R_3^1(0) = 0$	
$R_3^3(\rho) = \rho^3$	$R_3^3(0) = 0$	
$R_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1$	$R_4^0(0) = +1$	$h_4 = \frac{\pi}{5}$
$R_4^2(\rho) = 4\rho^4 - 3\rho^2$	$R_4^2(0) = 0$	
$R_4^4(\rho) = \rho^4$	$R_4^4(0) = 0$	

Table 4.8: The radial parts of the Zernike orthogonal polynomials.

We see that only R_0^0, R_2^0 and R_4^0 are non-zero at the origin. As in all these cases $m = 0$, one has $Z_{2j}^0(\rho, \varphi) = R_{2j}^0(\rho)$ for $j = 0, 1, 2$ and also $h_n = \frac{\pi}{n+1}$ because $\epsilon_0 = 2$ in (2.4). Inspecting the representation (2.3),

$$R_n^m(\rho) = (-1)^{(n-m)/2} \rho^m P_{(n-m)/2}^{(m,0)}(1 - 2\rho^2), \quad n - m \text{ even},$$

we see that $R_n^m(0)$ is positive only for $m = 0$ with $n = 2j$, because the terminal values of Jacobi polynomials are positive. Hence the sums of orthonormalized products write

$$\sum_{n=0}^{\ell=2k} \sum_{m:|m|\le n}^{n} \frac{1}{h_n} R_n^m(0) Z_n^m(\rho, \varphi) = \sum_{j=0}^{k} \frac{2j+1}{\pi} R_{2j}^0(\rho), \quad k \in \mathbb{N}. \tag{4.4}$$

Further, in this case

$$R_{2j}^0(\rho) = (-1)^j P_j^{(0,0)}(1 - 2\rho^2) = (-1)^j P_j(1 - 2\rho^2), \quad R_{2j}^0(0) = (-1)^j P_j(1)$$

where P_j are the Legendre orthogonal polynomials with $P_j(1) = 1$ for $j \in \mathbb{N}_0$.

Using Lemma 4.1, the Zernike kernels of order $\ell = 2k$, $k \in \mathbb{N}$ have the form

$$K_{\ell=2k}^2(||\boldsymbol{u}||_2) = \frac{1}{\pi} \left[\sum_{j=0}^{k} (2j+1) P_j(1 - 2\rho^2) \right] \cdot 1_{||\boldsymbol{u}||_2 \le 1}, \quad \boldsymbol{u} \in \mathbb{R}^2.$$

Because they are radial functions we can compute their Fourier transforms using the Hankel transform. Inspecting [Bateman, 1954, p. 13 (1)], one gets the Hankel transform of order zero ($d = 2$, i.e., $\nu = 0$) of terms $P_j(1 - 2\rho^2)$ as

$$\mathscr{H}_0(P_j(1 - 2\rho^2)) = s^{-1} J_{2j+1}(s).$$

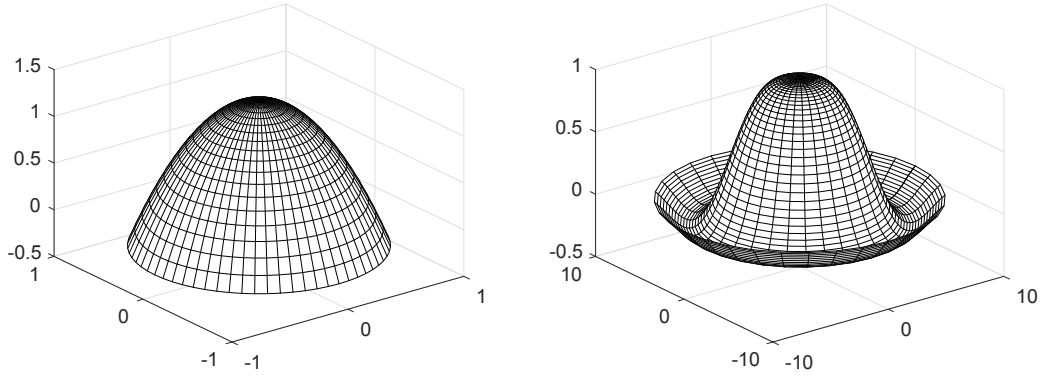

Figure 4.5: (a) 2D Zernike kernel of order $\ell = 2$ - K_2^2; (b) FT of K_2^2.

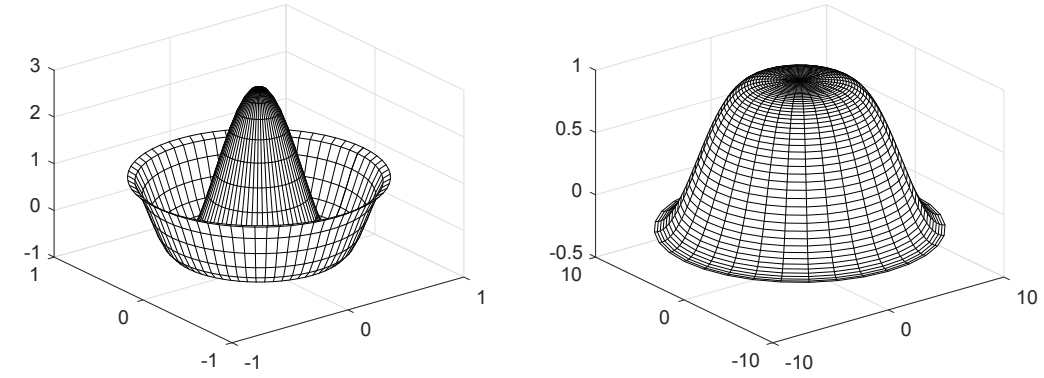

Figure 4.6: (a) 2D Zernike kernel of order $\ell = 4$ - K_4^2; (b) FT of K_4^2.

Therefore, following Theorem 2.1, the respective Fourier transform is obtained by multiplying with the factor $(2\pi)^{d/2}s^{-\nu} = 2\pi$, which gives

$$\mathcal{F}[K^2_{\ell=2k}](\boldsymbol{\omega}) = 2(||\boldsymbol{\omega}||_2)^{-1}\sum_{j=0}^{k}(2j+1)J_{2j+1}(||\boldsymbol{\omega}||_2), \quad \boldsymbol{\omega} \in \mathbb{R}^2.$$

Let us concretize the above formulas for $\ell = 2, 4$. We have

$$\begin{aligned} K_2^2(\rho) &= \frac{1}{\pi}\Big[1 - 3(2\rho^2 - 1)\Big] \cdot 1_{[0,1]}(\rho), \\ K_2^4(\rho) &= \frac{1}{\pi}\Big[1 - 3(2\rho_2^2 - 1) + 5(6\rho_2^4 - 6\rho_2^2 + 1)\Big] \cdot 1_{[0,1]}(\rho), \end{aligned}$$

which simplifies to

$$\begin{aligned} K_2^2(\rho) &= \frac{1}{\pi}(4-6\rho^2)\cdot 1_{[0,1]}(\rho), \\ K_2^4(\rho) &= \frac{1}{\pi}(30\rho^4-36\rho^2+9)\cdot 1_{[0,1]}(\rho). \end{aligned}$$

In Cartesian coordinates this clearly writes

$$\begin{aligned} K_2^2(||\boldsymbol{u}||_2) &= \frac{1}{\pi}(4-6||\boldsymbol{u}||_2^2)\cdot 1_{||\boldsymbol{u}||_2\leq 1}, \\ K_2^4(||\boldsymbol{u}||_2) &= \frac{1}{\pi}(30||\boldsymbol{u}||_2^4-36||\boldsymbol{u}||_2^2+9)\cdot 1_{||\boldsymbol{u}||_2\leq 1}. \end{aligned}$$

The 2D-Fourier transforms have form

$$\begin{aligned} \mathcal{F}[K_2^2](||\boldsymbol{\omega}||_2) &= 2(||\boldsymbol{\omega}||_2)^{-1}\left[J_1(||\boldsymbol{\omega}||_2)+3J_3(||\boldsymbol{\omega}||_2)\right], \\ \mathcal{F}[K_4^2](||\boldsymbol{\omega}||_2) &= 2(||\boldsymbol{\omega}||_2)^{-1}\left[J_1(||\boldsymbol{\omega}||_2)+3J_3(||\boldsymbol{\omega}||_2)+5J_5(||\boldsymbol{\omega}||_2)\right]. \end{aligned}$$

Both kernels together with their Fourier transforms are presented in Fig. 4.5 and Fig. 4.6, respectively. Note that for the Bessel function of the first kind one has $\lim_{r\to 0^+} r^{-1}J_1(r) = 1/2$ and $\lim_{r\to 0^+} r^{-1}J_{1+2j}(r) = 0$ for $j > 0$. Hence $F[K_2^2](\mathbf{0}) = \mathcal{F}[K_4^2](\mathbf{0}) = 0$. Checking the partial derivatives is left to the interested reader. It is of further interest to design the kernels supported on the d-dimensional unit ball for general $d \geq 3$. These kernels are called the spherical kernels.

4.2.4 Spherical kernels

To design the spherical kernels we are interested in the products introduced at the end of Section 2.1.8 that write

$$P_{j,1}^{2j}(\mathbf{0})P_{j,1}^{2j}(\boldsymbol{x}) = (h_{j,2j}^{\mu})^{-2}P_j^{(\mu-\frac{1}{2},\frac{d-2}{2})}(-1)\cdot P_j^{(\mu-\frac{1}{2},\frac{d-2}{2})}(2||\boldsymbol{x}||_2^2-1),\ \ j\in\mathbb{N}_0.$$

As we will consider the classical weight function $W_{\mu=0}^{B_d}(\boldsymbol{x}) = (1-||\boldsymbol{x}||_2)^{-1/2}$ we use $\mu = 0$ in the formula and the above then writes

$$P_{j,1}^{2j}(\mathbf{0})P_{j,1}^{2j}(\boldsymbol{x}) = (h_{j,2j}^{0})^{-2}P_j^{(-\frac{1}{2},\frac{d-2}{2})}(-1)\cdot P_j^{(-\frac{1}{2},\frac{d-2}{2})}(2||\boldsymbol{x}||_2^2-1),\ \ j\in\mathbb{N}_0$$

with

$$(h_{j,2j}^{0})^{-2} = c_0^{B_d}\frac{j!(\frac{d+1}{2})_j(2j+\frac{d-1}{2})}{(\frac{1}{2})_j(\frac{d}{2})_j(j+\frac{d-1}{2})} \quad \text{and} \quad c_0^{B_d} = \frac{\Gamma(\frac{d+1}{2})}{\pi^{d/2}\sqrt{\pi}}.$$

Using the symmetry relation in (2.1) one gets

$$P_{j,1}^{2j}(\mathbf{0})P_{j,1}^{2j}(\boldsymbol{x}) = (h_{j,2j}^{0})^{-2}P_j^{(\frac{d-2}{2},-\frac{1}{2})}(1)\cdot P_j^{(\frac{d-2}{2},-\frac{1}{2})}(1-2||\boldsymbol{x}||_2^2),\ \ j\in\mathbb{N}_0,$$

and further from (2.1) and (2.2), respectively,

$$\begin{aligned} P_j^{(\frac{d-2}{2},-\frac{1}{2})}(1) &= \binom{j+\frac{d-2}{2}}{j} = \frac{\Gamma(j+d/2)}{\Gamma(d/2)\Gamma(j+1)} = \frac{(\frac{d}{2})_j}{j!}\,, \\ P_j^{(\frac{d-2}{2},-\frac{1}{2})}(1-2||\boldsymbol{x}||_2^2) &= \frac{(\frac{1}{2})_j}{(\frac{d-1}{2})_j}\,C_{2j}^{(1/2+(d-2)/2)}((1-||\boldsymbol{x}||_2^2)^{1/2}). \end{aligned}$$

Putting all together and denoting $\nu=(d-2)/2$ one gets

$$\begin{aligned} P_{j,1}^{2j}(\mathbf{0})P_{j,1}^{2j}(\boldsymbol{x}) &= (h_{j,2j}^{0,d})^{-2}\frac{(\frac{d}{2})_j(\frac{1}{2})_j}{j!(\frac{d-1}{2})_j}\cdot C_{2j}^{(\nu+1/2)}((1-||\boldsymbol{x}||_2^2)^{1/2}) \\ &= c_0^{B_d}\frac{j!(\frac{d+1}{2})_j(2j+\frac{d-1}{2})}{(\frac{1}{2})_j(\frac{d}{2})_j(j+\frac{d-1}{2})}\frac{(\frac{d}{2})_j(\frac{1}{2})_j}{j!(\frac{d-1}{2})_j}\cdot C_{2j}^{(\nu+1/2)}((1-||\boldsymbol{x}||_2^2)^{1/2}) \\ &= c_0^{B_d}\frac{(\frac{d+1}{2})_j(2j+\frac{d-1}{2})}{(\frac{d-1}{2})_j(j+\frac{d-1}{2})}\cdot C_{2j}^{(\nu+1/2)}((1-||\boldsymbol{x}||_2^2)^{1/2}) \\ &= c_0^{B_d}\Big(\frac{4j}{d-1}+1\Big)\cdot C_{2j}^{(\nu+1/2)}((1-||\boldsymbol{x}||_2^2)^{1/2}). \end{aligned}$$

Finally, applying Lemma 4.1 for $W_{\mu=0}^{B}(\boldsymbol{x}) = (1-||\boldsymbol{x}||_2)^{-1/2}$ we get the following formula for specification of the spherical kernels of order $\ell=2k$, $k\in\mathbb{N}$ in $d\geq 3$ dimensions:

$$K_{\ell=2k}^{d}(||\boldsymbol{x}||_2) = c_0^{B_d}\sum_{j=0}^{\ell}\Big(\frac{4j}{d-1}+1\Big)\,C_{2j}^{((d-1)/2)}((1-||\boldsymbol{x}||_2^2)^{1/2})\cdot(1-||\boldsymbol{x}||_2^2)^{-1/2}\cdot 1_{[||\boldsymbol{x}||_2\leq 1]} \tag{4.5}$$

The kernels are clearly radial functions, hence their Fourier transform can be computed using the Hankel transform. The appropriate function for using [Bateman, 1954] writes

$$\varphi(r) = r^{\nu+1/2}\cdot(1-r^2)^{-1/2}\cdot C_{2j}^{\nu+1/2}((1-r^2)^{1/2})\cdot 1_{[0,1]}(r).$$

Entry (13) on page 44 of [Bateman, 1954] gives for $\alpha=0$ the transform (after multiplication by the $s^{-1/2}$ factor)

$$\begin{aligned} \mathscr{H}\{\varphi\}(s) &= (-1)^j\,2^{-1/2}\pi^{1/2}s^{-1/2}\cdot C_{2j}^{(\nu+1/2)}(0)\cdot J_{\nu+1/2+2j}(s) \\ &= (-1)^j C_{2j}^{(\nu+1/2)}(0)\cdot j_{\nu+2j}(s). \end{aligned}$$

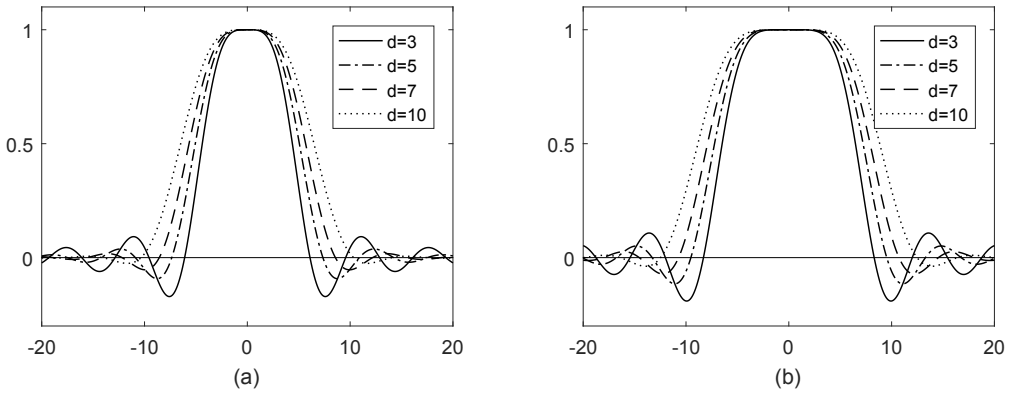

Figure 4.7: Spherical kernels of order (a) $\ell = 2$ and (b) $\ell = 4$.

Hence the d-variate Fourier transform of (4.5) reads as

$$\begin{aligned}\mathcal{F}[K^d_{\ell=2k}](\boldsymbol{\omega}) &= \frac{(2\pi)^{d/2} c_0^{B_d}}{||\boldsymbol{\omega}||_2^{\nu}} \sum_{j=0}^{\ell} \frac{(-1)^j (4j+d-1)}{d-1} C_{2j}^{(\nu+1/2)}(0) \cdot j_{\nu+2j}(||\boldsymbol{\omega}||_2) \\ &= \frac{2^{d/2}\Gamma(\frac{d+1}{2})}{\sqrt{\pi}||\boldsymbol{\omega}||_2^{(d-2)/2}} \sum_{j=0}^{\ell} \frac{(-1)^j (4j+d-1)}{d-1} C_{2j}^{((d-1)/2)}(0) \cdot j_{(d-2)/2+2j}(||\boldsymbol{\omega}||_2).\end{aligned}$$

To check the result, using (2.9) one has

$$\mathcal{F}[K^d_{\ell=2k}](\mathbf{0}) = \frac{2^{d/2}\Gamma(\frac{d+1}{2})}{\sqrt{\pi}} \lim_{r\to 0^+} r^{-\nu} j_\nu(r) = 1.$$

Fig. 4.7 shows the above Fourier transform formula for a univariate argument, i.e., $\mathcal{F}[K^d_{\ell=2k}](|s|), s \in \mathbb{R}$, for increasing d and $\ell = 2, 4$ to see its dependence on the dimension.

5. Discussion

In accordance with its assignment, the thesis delivers three theoretical results related to using kernel methods in particle filtering including a more practical discussion on designing suitable kernels. Namely, in Chapter 3, we have presented

- The upper bounds on the kernel density estimates of the filtering densities and their partial derivatives.
- The lower bounds on the kernel density estimates of the filtering densities.
- The condition for checking persistence of the Sobolev character of the filtering densities.

Chapter 4 then focuses on working with kernels from a more practical point of view.

To assess originality of the thesis contribution, we compare our results with those presented in [Crisan and Míguez, 2014]. This paper is highly relevant to this purpose as its topic significantly overlaps with the one presented in the thesis. We discuss explicitly differences between the two groups of results.

The results of Crisan and Míguez [2014] in Section 4 split into two groups. The first comprises a.s. convergence results – Theorems CM-4.1, CM-4.2, CM-4.3 and CM-4.5; all drawing on Lemma CM-4.1[1]. The second group comprises the results for integrated versions (w.r.t. the probability and Lebesgue measures) – Theorems CM-4.4 and CM-4.6. Note that Theorem CM-4.5 falls into the first group as integration is provided only w.r.t. the Lebesgue measure. Results of Section 5 are aimed on applications and will not be discussed here.

Our theorems relate mainly to the second group. In fact, we have presented a stronger version of Theorem CM-4.4 due to the different assumption on the estimated density - the Sobolev character instead of the Lipschitz continuity.

Theorems CM-4.4 and CM-4.6 are restricted to densities and kernels compactly supported on $\mathcal{K} \subseteq \mathbb{R}^d$. Thus, for example, they do not cover the basic Gaussian case. It is clear that their constants $c_{\alpha,\mathcal{K},t}$ and $c_{\mathcal{K},t}$ grow to infinity as the volume of $\mathcal{K}$ does. The reason for introducing the compact support requirement is that the inequality (4.13) of [Crisan and Míguez, 2014] cannot be simply integrated w.r.t. Lebesgue measure on $\mathbb{R}^d$ as the right-hand side would turn to an uninformative unlimited upper bound. Our Theorem DC-3.1 is not

[1] To make a clear distinction between the theorems of the paper and the thesis, we denote the theorems of [Crisan and Míguez, 2014] as CM-x.x and ours as DC-x.x.

restricted by these limitations. In our approach, the behavior of $D^\alpha p_t$ with respect to integration over $\mathbb{R}^d$ is induced by the requirement on its Sobolev character. Similarly, this is also the case for the used kernel when its behavior is determined by its order.

Further, for $\beta = 1$, the bound in our Theorem DC-3.1 is tighter than that of Theorem CM-4.4 and equals the one presented in Theorem CM-4.6. Indeed, in Theorem CM-4.6 it is required that the filtering density has bounded partial derivatives up to order 2, which implies that the density is 1-Sobolev. The bound in Theorem CM-4.6 writes $b_2 = n^{-4/2(d_x+2)}$; see the discussion in paragraph 4.4 of [Crisan and Míguez, 2014] for transforming k to the number of particles n. Our bound for $\beta = 1$ then writes $b_1 = n^{-2/(2+d)}$, so $b_2 = b_1$ (in both cases the constants are omitted). Remark that Theorem CM-4.6 applies to $|\alpha|$=0.

The lower bounds on kernel density estimates are not discussed in [Crisan and Míguez, 2014] at all. In fact, up to our knowledge our result is the first case of introducing these bounds into the context of particle filtering.

Theorem DC-3.3 on persistence of the Sobolev character corresponds to Remark CM-3.4 of [Crisan and Míguez, 2014]. The difference is that we are more specific. In Remark CM-3.4, it is required that $g_t^{y_t}$ (g_t^v in our notation) is bounded similarly as in our case, but we do not have any requirement on derivatives of g_t^v. Speaking about the transition kernels, our requirement is that they are uniformly bounded by a common β-Sobolev function, which is a simpler condition than that of the remark.

To sum up, due to our assumptions we are able to obtain stronger results for MISE in terms of a general integration domain. Moreover, in [Crisan and Míguez, 2014] the transition to a.s. versions comes from the integrated versions via Lemma CM-4.1. Thus, using this lemma we might obtain the a.s. version for ISE (the counterpart of Theorem CM-4.5) without further restrictions on supports of the filtering densities and employed estimation kernel.

Chapter 4 is geared towards a more practical aspects of estimating the filtering densities. Namely, on how to design multivariate kernels of the given orders in the given dimension. The ability to do so make possible to enjoy higher convergence rates in estimation. Hence, from a practitioner point of view, it is an important complement to the presented theory.

In conclusion, we express the conviction that the thesis brings original material and ideas related to using kernel density estimation in particle filtering and adds its share to the research in the respective field.

Bibliography

A. D. Poularikas (Ed.). *The Transforms and Applications Handbook, Second Edition*. Electrical Engineering Handbook. CRC Press LLC, 2000.

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 1964.

H. Bateman. *Tables of Integral Transforms* [*Volumes* I *&* II]. McGraw-Hill Book Company, New York, 1954. URL `http://authors.library.caltech.edu/43489`.

J. Brabec and B. Hruza. *Matematická analýza II (Mathematical Analysis* II*, in Czech)*. SNTL/ALFA, 1986.

D. Coufal. On Convergence of Kernel Density Estimates in Particle Filtering. *Kybernetika*, 52(5):735–756, 2016.

D. Coufal. Upper and Lower Bounds on Kernel Density Estimates in Particle Filtering. *Submitted to Statistics & Probability Letters*, 2018a.

D. Coufal. On persistence of convergence of kernel density estimates in particle filtering. In P. Kulczycki *et al.*, editor, *Proceedings of 3rd Conference on Information Technology, Systems Research and Computational Physics*, Advances in Intelligent Systems and Computing. Springer, 2018b.

D. Crisan and A. Doucet. A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.

D. Crisan and J. Míguez. Particle-kernel estimation of the filter density in state-space models. *Bernoulli*, 20(4):1879–1929, 2014. URL `https://arxiv.org/pdf/1111.5866.pdf`.

D. Crisan and B. Rozovskii. *The Oxford Handbook of Nonlinear Filtering*. Oxford Handbooks in Mathematics. OUP Oxford, 2011.

L. Debnath and D. Bhatta. *Integral Transforms and Their Applications*. Chapman & Hall/CRC, 2007.

A. Doucet and A. M. Johansen. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. In D. Crisan and B. Rozovskii, editors, *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.

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

Charles F. Dunkl and Yuan Xu. *Orthogonal Polynomials of Several Variables*. Cambridge University Press, 2nd edition, 2014.

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

G. H. Givens. Consistency of the Local Kernel Density Estimator. *Statistics & Probability Letters*, 25:55–61, 1995.

M. Hürzeler and H. R. Künsch. Monte Carlo Approximations for General State-Space Models. *Journal of Computational and Graphical Statistics*, 7 (2):175–193, 1998.

S. G. Johnson. Saddle-Point Integration of C_∞ "bump" Functions. *arXiv:1508.04376*, 2015.

H. R. Künsch. Recursive Monte Carlo Filters: Algorithms and Theoretical Bounds. *The Annals of Statistics*, 33(5):1983–2021, 2005.

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*, 14(1):144–187, 2004.

C. Musso, N. Oudjane, and F. Le Gland. Improving Regularised Particle Filters. In A. Doucet, N. Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 12, pages 247–272. Springer, 2001.

E. Parzen. On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

Miguel A. Piñar and Yuan Xu. Best Polynomial Approximation on the Unit Ball. *arXiv:1609.05515v2*, 2017.

M. A. Pinsky. *Introduction to Fourier Analysis and Wavelets*. American Mathematical Society, 2008.

W. Rudin. *Modern Analysis*. Dover, 1991.

S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

Z. Sasvári. *Multivariate Characteristic and Correlation Functions*. De Gruyter, 2013.

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

E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.

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

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

Eric W. Weisstein. *CRC Concise Encyclopedia of Mathematics, Second Edition*. CRC Press, 2002.

Z. Zhao. Comparison of several ballistic target tracking filters. In *2006 American Control Conference*, pages 2197–2202, 2006.

List of Figures

List of Tables

List of publications

D. Coufal. On Convergence of Kernel Density Estimates in Particle Filtering. *Kybernetika*, 52(5):735–756, 2016.

D. Coufal. Upper and Lower Bounds on Kernel Density Estimates in Particle Filtering. *Submitted.*

D. Coufal. On persistence of convergence of kernel density estimates in particle filtering. In P. Kulczycki *et al.*, editor, *Proceedings of 3rd Conference on Information Technology, Systems Research and Computational Physics*, Advances in Intelligent Systems and Computing. Springer, 2018.