CONTRIBUTIONS TO WAVELET METHODS IN NONPARAMETRIC STATISTICS

By

Athanasia Petsa

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF CYPRUS
NICOSIA, CYPRUS
JULY 2009

UNIVERSITY OF CYPRUS
DEPARTMENT OF
MATHEMATICS AND STATISTICS

The undersigned hereby certify that they have read and recommend to the senate for acceptance a thesis entitled "**Contributions to Wavelet Methods in Nonparametric Statistics**" by **Athanasia Petsa** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: July 2009

Examining Committee:

Kostas Fokianos, Chairman
Associate Professor, University of Cyprus

Theofanis Sapatinas, Research Advisor
Associate Professor, University of Cyprus

Anestis Antoniadis
Professor, University Joseph Fourier, France

Efstathios Paparoditis
Professor, University of Cyprus

Yves Rozenholc
Associate Professor, University of Paris 5

UNIVERSITY OF CYPRUS

Date: **July 2009**

Author: **Athanasia Petsa**

Title: **Contributions to Wavelet Methods in Nonparametric Statistics**

Department: **Mathematics and Statistics**

Degree: **Ph.D.** Convocation: **February** Year: **2010**

Signature of Author

To my parents Andreas and Zoe and to my twin sister Fani.

Table of Contents

List of Tables

List of Figures

Abstract

We first consider the problem of estimating the integral of the square of a probability density function f on the basis of a random sample from a weighted distribution. Specifically, using model selection via a penalized criterion, an adaptive estimator for $\int f^2$ based on weighted data is proposed, for probability density functions which are uniformly bounded and belong to certain Besov bodies. We show that the proposed estimator attains the minimax rate of convergence that is optimal in the case of direct data. Additionally, we obtain the information bound for the problem of estimating $\int f^2$ when weighted data are available and compare it with the information bound for the case of direct data. A small simulation study is conducted to illustrate the usefulness of the proposed estimator in practical situations.

We then consider the problem of estimating the unknown response function in the standard Gaussian white noise model. We first utilize the recently developed maximum *a posteriori* (MAP) testimation procedure for recovering an unknown high-dimensional Gaussian mean vector. The existing results for its upper error bounds over various sparse l_p-balls are extended to more general settings and compared with other well-known threshold estimators. The MAP testimation procedure is then applied in a wavelet context to derive adaptively optimal global and level-wise MAP wavelet testimators of the unknown response function in the standard Gaussian white noise model over a wide range of Besov balls. These results are also extended to the estimation of derivatives of the response function. Simulated examples are conducted to illustrate the performance of the proposed adaptive level-wise MAP wavelet testimator, and to compare it with three proposed adaptive empirical Bayes estimation procedures that attain the optimal convergence rate, and one block wavelet thresholding estimator that is near optimal (up to a logarithmic factor). An application to real data is also considered.

Finally, we extend the minimax results obtained in the functional deconvolution model by Pensky & Sapatinas (2009a) under the L^2-risk to the case of L^p-risk, $1 \leq p < \infty$. Lower bounds are given for an arbitrary estimator of the unknown response function when the latter is assumed to belong to a Besov ball and under appropriate smoothness assumptions on the blurring function, including both regular-smooth and super-smooth convolutions. Furthermore, we investigate the asymptotic minimax properties of an adaptive wavelet estimator over a wide range of Besov balls. Box-car convolutions in the multichannel deconvolution model are also considered and the results of Pensky & Sapatinas (2009b) under the L^2-risk are extended to the case of L^p-risk, $1 \leq p < \infty$. A simulation study is conducted to show that the proposed adaptive wavelet estimator performs well in finite sample situations.

Acknowledgements

I would like to thank Dr Theofanis Sapatinas, my supervisor, for his many suggestions, constant support and friendly encouragement during this research. I am also thankful to Dr Felix Abramovich and Dr Vadim Grinsthein for their help and suggestions on Chapter 3 of my thesis.

Dr Cristophe Chesneau expressed his interest in my work and supplied me with the preprints of some of his recent work. Dr Claudia Angelini shared with me her knowledge of computational statistics, using MATLAB and R. Theodoros Papagiannis helped me to discover latex secrets and make latex a useful tool for my work.

Of course, I am grateful to my parents, Andreas and Zoe, and to my twin sister Fani for their patience and *love*. Without them this work would never have come into existence (literally).

Finally, I wish to thank the following: Georgia (for her friendship); Ilias (he knows why); my sister Fani (for all the good and bad times we had together).

Athanasia Petsa

April 8, 2009

Introduction

Chapter 1 is devoted to an overview of the mathematical tools that will be used in the main Chapters 2-4 of the Thesis.

We first consider the problem of estimation of the integral of the square of a probability density function (p.d.f) f, i.e., $\int f^2$, given a sample of weighted data. Laurent (2005) recently proposed an adaptive and optimal (in the minimax sense) estimator of $\int f^2$ for direct data. In Chapter 2, we construct an adaptive estimator of $\int f^2$, for p.d.f.'s which are uniformly bounded and belong to certain Besov bodies. We show that the proposed estimator attains the minimax rate of convergece that is optimal in the case of direct data. A small simulation study is conducted in order to assess the performance of the proposed estimator in practice. Using the theory of Ibramigov & Khasminski (1991), we also derive the information bound for the problem of estimating $\int f^2$ when weighted data are available. A comparison with the information bound given for the case of direct data (see, e.g., Pfanzagl, 1982) is presented.

Abramovich, Grinshtein & Pensky (2007) have recently proposed a Bayesian testimation procedure in order to recover a high-dimensional Gaussian mean $\mu = (\mu_1, \ldots, \mu_n)'$ with independent terms and common variance, under the assumption that μ is sparse. Optimality of the proposed MAP (maximum *a posteriori*) testimator $\hat{\mu}^*$ for μ belonging to strong l_p-balls and weak m_p-balls, $0 \leq p < 2$, was established in Abramovich, Grinshtein & Pensky (2007). In Chapter 3, we first consider the Gaussian sequence model and generalize the results of Abramovich, Grinshtein & Pensky (2007) for the MAP testimator in several directions. We then consider the problem of estimating the unknown function f on the basis of observations from the standard Gaussian white noise model and show that, under mild conditions on the prior distribution, the global wavelet MAP testimator is asymptotically nearly-minimax over the entire range of Besov balls $B^s_{p,q}(C)$, $s > \frac{1}{p}$, $0 < p, q \leq \infty$ and $C > 0$. Then, we show that an adaptive level-wise MAP testimator $\hat{f}$ is asymptotically optimal (in the minimax sense), as the sample size increases, for the same class of functions. These results are extended to the estimation of derivatives of f. Moreover, we demonstrate that the discretization of the data does not affect the order of magnitude of the accuracy of the MAP wavelet testimator, under the sample data model. A simulation study is conducted in order to assess the performance of the proposed level-wise wavelet testimator in finite sample situations. Additionally, a real data set is analyzed using five different methods.

Finally, we consider the estimation problem of an unknown function f based on observations from the functional deconvolution model proposed by Pensky & Sapatinas (2009a). In Chapter 4, we extend the results of Pensky & Sapatinas (2009a) to the case of L_p-risk, $1 \leq p < \infty$. In particular, lower bounds are derived for the L_p-risk, $1 \leq p < \infty$, of an estimator of f for both the functional deconvolution model and its discrete counterpart, under appropriate regularity assumptions on both f and the blurring function $g(\cdot, \cdot)$. Additionally, an adaptive thresholding estimator of f, which is a generalized version of the estimator proposed by Pensky & Sapatinas (2009a), is shown to be asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, as $n \to \infty$, in a wide range of Besov balls. Minimax lower and upper bounds are also given under the L^p-risk, $1 \leq p < \infty$ for the multichannel deconvolution problem with box-car convolutions under appropriate assumptions on the number of channels M and the selection points u_l, $l = 1, 2, \ldots, M$.

This work is concluded with a Future Research plan.

Περίληψη

Στο πρώτο κεφάλαιο γίνεται μια ανασκόπηση των βασικών μαθηματικών εργαλείων που χρησιμοποιούνται στα κυρίως κεφάλαια 2-4 της διδακτορικής διατριβής.

Το πρώτο θέμα της διδακτορικής μου διατριβής είναι η εκτίμηση της ποσότητας $\int f^2$ όταν ο στατιστικός έχει στη διάθεση του δεδομένα με βάρος. Το αντίστοιχο πρόβλημα για δεδομένα χωρίς βάρος είχε μελετηθεί απο την Laurent (2005). Μια εκτίμηση αυτής της ποσότητας μπορεί να χρησιμοποιηθεί όταν στατιστκές διαδικασίες που έχουν μελετηθεί για δεδομένα χωρίς βάρος γενικεύονται στην περίπτωση δεδομένων με βάρος. Στο δεύτερο κεφάλαιο της διδακτορικής μου διατριβής, τροποποιώντας τη μέθοδο που είχε εισαχθεί από τη Laurent (2005) για δεδομένα χωρίς βάρος, κατασκευάζουμε ένα εκτιμητή του οποίου η ταχύτητα σύγκλισης είναι βέλτιστη για δεδομένα χωρίς βάρος, για την κλάση των συναρτήσεων πυκνότητας που είναι ομοιόμορφα φραγμένες και ανήκουν σε Besov χώρους. Επίσης έχουν γίνει κάποιες προσομοιώσεις οι οποίες δείχνουν τη πρακτική σημασία της εκτιμήτριας που προτείνουμε όταν υπάρχουν στη διάθεση μας δεδομένα με βάρος. Χρησιμοποιώντας τη θεωρία των Ibragimov and Khasmiinski (1991), δίνουμε το πληροφοριακό φράγμα για το πρόβλημα της εκτίμησης της ποσότητας $\int f^2$ όταν δεδομένα με βάρος είναι διαθέσιμα και το σύγκρινουμε με το πληροφοριακό φράγμα στην περίπτωση δεδομένων χωρίς βάρος.

Οι Abramovich, Grinshtein & Pensky (2007) πρότειναν τη μέθοδο της εκ των υστέρων μεγιστοποίησης για την εκτίμηση ενός Κανονικού (Gaussian) πολυδιάστατου μέσου $\mu = (\mu_1, \mu_2, \ldots, \mu_n)'$ με ανεξάρτητους όρους και κοινή διασπορά, και απέδειξαν ότι επιτυγχάνει τη βέλτιστη ταχύτητα σύγκλισης όταν το διάνυσμα μ ανήκει σε ισχυρές και ασθενείς l_p-μπάλες, $0 \leq p < 2$. Στο τρίτο κεφάλαιο γενικεύουμε τα αποτελέσματα των Abramovich, Grinshtein & Pensky (2007). Ακολούθως, θεωρούμε το πρόβλημα της εκτίμησης της άγνωστης συνάρτησης f με βάση τις παρατηρήσεις από το Κανονικό μοντέλο λευκού θορύβου. Μελετούμε τη

συμπεριφορά μιας εκτιμήτριας της f η οποία βασίζεται στη MAP μέθοδο και δεν εξαρτάται από τις άγνωστες παραμέτρους της κλάσης. Δείχνουμε ότι η προτεινόμενη εκτιμήτρια επιτυγχάνει ασυμπτωτικά τη βέλτιστη ταχύτητα συγκλίσης για συναρτήσεις f των οποίων οι wavelet συντελεστές ανήκουν σε μια Besov μπάλα $B^s_{p,q}(C)$ με $s > \frac{1}{p}$, $0 < p, q \leq \infty$, $C > 0$. Τα αποτελέσματα αυτά επεκτείνονται και στην εκτίμηση των παραγώγων της f. Επίσης δείχνουμε ότι η χρήση διακριτών δεδομένων δεν επηρεάζει την ταχύτητα σύγκλισης της προτεινόμενης εκτιμήτριας κάτω από το μοντέλο αυτό. Μια αριθμητική μελέτη χρησιμοποιείται για να δείξει την απόδοση της προτεινόμενης εκτιμήτριας στην πράξη. Επίσης, περιλαμβάνεται η ανάλυση ενός συνόλου πραγματικών δεδομένων με τη χρήση της προτεινόμενης εκτιμήτριας και άλλων εκτιμητριών.

Τέλος, θεωρούμε το προβλημα της εκτίμησης της άγνωστης συνάρτησης f όταν οι παρατηρήσεις προέρχονται από το μοντέλο συναρτησιακής συνέλιξης που είχε εισαχθεί από τους Pensky & Sapatinas (2009a). Στο τέταρτο κεφάλαιο επεκτείνουμε τα αποτελέσματα των Pensky & Sapatinas (2009a) στην περίπτωση του L^p σφάλματος, $1 \leq p < \infty$. Συγκεκριμμένα, δίνουμε τα κάτω φράγματα για το L^p σφάλμα, $1 \leq p < \infty$, μιας εκτιμήτριας της f στο μοντέλο συναρτησιακής συνέλιξης και στο αντίστοιχο διακριτό μοντέλο, κάτω από κατάλληλες συνθήκες πάνω στην f και τη συνάρτηση συνέλιξης $g(\cdot, \cdot)$. Στη συνέχεια προτείνουμε μια εκτιμήτρια της f, η οποία είναι μια γενικευμένη έκδοση της εκτιμήτριας των Pensky & Sapatinas (2009a), και δείχνουμε ότι είναι ασυμπτωτικά βέλτιστη, ή σχεδόν-βέλτιστη, εκτός από ένα λογαριθμικό παράγοντα, για μια ευρεία κλάση από Besov μπάλες.

Τέλος, γίνονται κάποιες εισηγήσεις για θέματα συναφή με τη διδακτορική μου διατριβή τα οποία θα μπορούσαν να μελετηθούν στο μέλλον.

Chapter 1

Overview on Wavelets and other statistical techniques

This chapter is an overview on wavelets and other statistical techniques which are going to be used in the following three chapters of the Thesis. For a detailed review of wavelets in various statistical applications and appropriate software see, e.g., Antoniadis (1999), Abramovich, Bailey & Sapatinas (2000) and Antoniadis, Bigot & Sapatinas (2001).

1.1 Wavelets

Wavelets consist an orthonormal basis with local properties in both frequency and time. For this reason they are called a local basis. In this section we will briefly discuss multiresolution analysis (MRA), Haar and Meyer wavelets, the discrete wavelet transformation, the sample data model and boundary wavelets.

1.1.1 Multiresolution Analysis

A multiresolution analysis (MRA) of $L^2(\mathbb{R})$ can be used to define wavelets. A MRA of $L^2(\mathbb{R})$ is a sequence of closed subspaces V_j of $L^2(\mathbb{R})$, $j \in \mathbb{Z}$ with the following properties.

1. Nesting property: $V_j \subset V_{j+1}$.

2. Density property: $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.

3. Separation property: $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

4. Scaling relation: $f(t) \in V_j \Longleftrightarrow f(2t) \in V_{j+1}$.

5. There is a scaling function $\phi \in V_0$ such that $\{\phi(t-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0.

Using the nesting property, we can obtain W_j, the orthogonal complement of V_j in V_{j+1}. The density property leads to

$$L^2(\mathbb{R}) = \overline{\oplus_{j\in\mathbb{Z}} W_j} = \overline{V_0 \oplus_{j\geq 0} W_j}.$$

Using properties (4) and (5), it is easy to see that $\{\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), k \in \mathbb{Z}\}$ forms an orthonormal basis for V_j. Since the scaling function $\phi(x) \in V_0 \subset V_1$, ϕ can be written as a linear combination of bases of V_1

$$\phi(x) = \sum_k \alpha(k)\phi(2x-k),$$

where $\alpha(k) = 2\int \phi(x)\overline{\phi(2x-k)}dx$ and $\sum_k |\alpha(k)|^2 = 2$. The wavelet space W_0 is the orthogonal complement of V_0 in V_1 and $W_0 \subset V_0$. Therefore, the mother wavelet $\psi(x)$ satisfies the following relation

$$\psi(x) = \sum_k b(k)\phi(2x-k),$$

where $b(k) = (-1)^k\overline{\alpha}(1-k)$, so that $\psi(x)$ is orthogonal to $\phi(x)$. It is easy to see that $\{\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), k \in \mathbb{Z}\}$ is an orthonormal basis for W_j and $\{\psi_{jk}(x), j,k \in \mathbb{Z}\}$ is an orthonormal basis, namely 'wavelets', of $L^2(\mathbb{R})$. $L^2(\mathbb{R})$ can then be expressed as follows

$$L^2(\mathbb{R}) = \overline{\oplus_{j\in\mathbb{Z}} W_j} = \overline{V_0 \oplus_{j\geq 0} W_j} = \overline{V_{j_0} \oplus_{j\geq j_0} W_j},$$

where j_0 is some integer. Therefore, any function $f \in L^2(\mathbb{R})$ can be written in the following ways

$$\begin{aligned} f(x) = \sum_{j,k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) &= \sum_{k\in\mathbb{Z}} c_{0k}\phi_{0k}(x) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) \\ &= \sum_{k\in\mathbb{Z}} c_{j_0k}\phi_{j_0k}(x) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x), \end{aligned}$$

where $c_{jk} = \int f(x)\phi_{jk}(x)dx$ and $d_{jk} = \int f(x)\psi_{jk}(x)dx$.

1.1.2 Different wavelet bases

There are several examples of wavelets. The Haar basis is known since 1910. Strang (1993) and Vidaković & Müller (1994) start to explain wavelets by the Haar wavelet which is simple. The Haar wavelet function is defined by

$$\phi(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

and is called Haar father wavelet. The Haar mother wavelet is defined by

$$\psi(x) = \begin{cases} 1, & \text{if } 0 \leq x < \frac{1}{2}, \\ -1, & \text{if } \frac{1}{2} \leq x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Haar wavelets have good properties such as simplicity, orthogonality and compact support. However, they are discontinuous and cannot be used to approximate smooth functions. Meyer (1985) developed orthonormal wavelet bases with infinite support and exponential decay. A key development was the work of Daubechies (1988, 1992, Chapters 6 and 7) who derived two families of orthonormal wavelet bases (the so-called *extremal phase* and *least asymmetric* families) which combine compact support with various degrees of smoothness and numbers of vanishing moments. *Coiflets* (Daubechies, 1993) and *spline wavelets* (Chui, 1992) are other examples of wavelet bases that are used in practice and several additional wavelet families (orthogonal, biorthogonal and semiorthogonal) have also been developed during the last decade.

1.1.3 Meyer Wavelets

Let

$$\hat{f}(\xi) = \int_0^1 f(x) e^{-2\pi i x \xi} dx$$

for $f \in L^2[0,1]$. We define the 'mother' Meyer wavelet ψ, in the frequency domain as

$$\hat{\psi}(\xi) = \begin{cases} e^{-i\pi\xi} \sin(\frac{\pi}{2}\nu(3|\xi| - 1)), & \text{if } \frac{1}{3} < |\xi| \leq \frac{2}{3}, \\ e^{-i\pi\xi} \cos(\frac{\pi}{2}\nu(\frac{3}{2}|\xi| - 1)), & \text{if } \frac{2}{3} < |\xi| \leq \frac{4}{3}, \\ 0, & \text{otherwise,} \end{cases}$$

where $\nu(\cdot)$ is a smooth function such that

$$\nu(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1, & \text{if } x > 1 \end{cases}$$

and $\nu(x) + \nu(1-x) = 1$.

We define

$$\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), \quad x \in \mathbb{R}.$$

We can write

$$\hat{\psi}(\xi) = u_{jk}(\xi) - iv_{jk}(\xi),$$

where $i^2 = -1$. Let f be a periodic function in $L^2[0,1]$ and let

$$\psi^0_{jk}(x) = \sum_{l\in\mathbb{Z}} \psi_{jk}(x+l)$$

for $j \in \mathbb{Z}^+$ and $k = 0, 1, \ldots, 2^j - 1$ be the periodized Meyer wavelet on $[0,1]$. It is easy to see that

$$\hat{\psi}^o_{jk}(l) = \hat{\psi}_{jk}(l), \quad l \in \mathbb{Z}.$$

Hence,

$$\hat{\psi}^o_{jk}(l) = u_{jk}(l) - iv_{jk}(l)$$

and

$$\alpha^o_{jk} = \int_0^1 f(x)\psi^o_{jk}(x)dx = \sum_{l\in\mathbb{Z}} \Re\hat{f}(l)u_{jk}(l) - \sum_{l\in\mathbb{Z}} \Im\hat{f}(l)v_{jk}(l),$$

where $\Re(x)$ and $\Im(x)$ are the real and imaginary parts of x respectively.

The 'father' Meyer wavelet ϕ is written in the frequency domain as

$$\hat{\phi}(\xi) = \begin{cases} 1, & \text{if } |\xi| \leq \frac{1}{3}, \\ \cos\left(\frac{\pi}{2}\nu(3|\xi| - 1)\right), & \text{if } \frac{1}{3} < |\xi| \leq \frac{2}{3}, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\phi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), \quad x \in \mathbb{R}$$

and

$$\hat{\phi}^o_{jk}(x) = \sum_{l\in\mathbb{Z}} \phi_{jk}(x+l)$$

for $j \in \mathbb{Z}^+$ and $k = 0, \ldots, 2^j - 1$ be the 'father' Meyer wavelet on $[0, 1]$. It is easy to see that

$$\hat{\phi}^o_{jk}(l) = \hat{\phi}_{jk}(l), \quad l \in \mathbb{Z}.$$

1.1.4 Discrete Wavelet Transformation

The discrete wavelet transform (DWT) (see, e.g., Nason & Silverman (1994) and Edwards (1992)) is an orthogonal transform applied on discrete data. DWT requires input data with the sample size of a power of 2, a high-pass filter and a low-pass filter. At each stage, a sequence of smoothed c and detailed coefficients d is produced. Low-pass filter returns smoothed data, while high-pass filter returns detailed data. The low-pass filter h is a convolution followed by dyadic decimation, as in Mallat (1989):

$$c_k^{j-1} = \sum_{n \in \mathbb{Z}} h(n - 2k) c_n^j = \sum_{m \in \mathbb{Z}} h(m) c_{m+2k}^j = \sum_{m=0}^{N_h - 1} h(m) c_{m+2k}^j.$$

The high-pass filter g acts in the following way in order to return detailed coefficients

$$d_k^{j-1} = \sum_{n \in \mathbb{Z}} g(n - 2k) c_n^j,$$

where $c_0^M, c_1^M, \ldots, c_{N-1}^M$ is a set of $N = 2^M$ data. The superscript M means that we have the original data. Wavelet coefficients are obtained by applying two filters and a downsampling filter. By applying two filters and an upsampling filter on the wavelet coefficients, we obtain the original data. Computation of the forward and inverse DWT would be expected to require $O(n^2)$ operations. However, due to its construction, it only requires $O(n)$ operations and it is faster than the fast Fourier transform (FFT) which requires $O(n \log(n))$ operations.

Let $N = 2^M$ be the sample size. Donoho & Jonhstone (1995) used the matrix form of DWT, i.e., $w = Wy$, where the vector w contains the wavelet coefficients of a vector y of size N. Due to the orthogonality of the transform matrix W, the input data y can be reconstructed as follows $y = W^T w$, or equivalently $y_i = \sum_{jk} w_{jk} W_{jk}(i)$. Since $\sqrt{N} W_{j,k}(i)$ approximates $2^{\frac{j}{2}} \psi(2^j \frac{i}{N} - k)$, the data $y_i, \quad i = 0, 1, \ldots, N-1$, can be expressed as

$$y_i = \sum_{j,k} d_{jk} \psi_{jk}(\frac{i}{N}), \tag{1.1}$$

where $d_{jk} \approx w_{jk}/\sqrt{N}$ (see, e.g., Vidakovic (1999)).

In Chapter 4, we consider the fast $O(n \log_2^2(n))$ forward and inverse discrete, periodized Meyer wavelet transforms due to Kolaczyk (1994). These algorithms take place in the frequency domain and associate each level of coefficients with a projection of the signal onto a frequency band.

1.1.5 Minimax Risk

The minimax risk associated with a statistical model $\{P_\theta, \theta \in \Theta\}$ and with a semi-distance d is defined by

$$\mathcal{R}_n^* = \inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\hat{\theta}_n, \theta)],$$

where the infimum is taken over all estimators (see, e.g., Tsybakov (2009), p. 78). Consider the following inequalities.

$$\limsup_{n \to \infty} \psi_n^{-2} \mathcal{R}_n^* \leq C, \tag{1.2}$$

$$\liminf_{n \to \infty} \psi_n^{-2} \mathcal{R}_n^* \geq c, \tag{1.3}$$

where c, C are positive constants, independent of n. A positive, decreasing sequence $\{\psi_n\}_{n=1}^\infty$ is called an optimal rate of convergence of estimators on (Θ, d) if (1.2) and (1.3) hold. An estimator θ_n^* satisfying

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\theta_n^*, \theta)] \leq C' \psi_n^2,$$

where $\{\psi_n\}_{n=1}^\infty$ is the optimal rate of convergence and $C' < \infty$ is a constant, independent of n, is called a rate optimal estimator on (Θ, d).

1.1.6 Sampled data model and boundary wavelets

The theory of nonparametric function estimation is usually based on the standard Gaussian white noise model, i.e.,

$$dY(t) = f(t)dt + \frac{1}{\sqrt{N}} dW(t), \quad t \in [0, 1], \tag{1.4}$$

where W is the standard Wiener process. Carrying out a wavelet decomposition of the function $f(t)+N^{-\frac{1}{2}}dW(t)$, we obtain independent observations $Y_{jk} \sim N(\theta_{jk}, \frac{1}{N})$. In practice, however, instrumentally acquired data that is digitally processed is typically discrete. Such settings can be represented by the sampled data model, that is,

$$Y_i = f(\frac{i}{N}) + \epsilon_i, \tag{1.5}$$

where ϵ_i are independent $N(0,1)$ random variables. The discrete wavelet transform of $N^{-\frac{1}{2}}Y$ yields $\tilde{Y}_{jk} \sim N(\tilde{\theta}_{jk}, \frac{1}{N})$. In much of the existing literature the difference between Y_{jk} and $\tilde{Y}_{jk}$ is ignored. Estimators are usually motivated, derived and analyzed in the standard Gaussian white noise model (1.4) and are applied to discrete wavelet transform data, in practice. An interesting problem is to investigate the risk bounds of an estimator based on observation $\tilde{Y}_{jk}$ from model (1.5). Jonhstone & Silverman (2004) used boundary-modified coiflets to show that the discrete wavelet transform of finite data from (1.5) asymptotically provides a close approximation to the wavelet transform of the standard Gaussian white noise model (1.4). These results were used in Johnstone & Silverman (2005) to prove that discretization of the data does not affect the asymptotic convergence rates of the upper risk bounds of the proposed empirical Bayes estimators.

Consider a scaling function ϕ with vanishing moments of order $1, 2 \ldots, R-1$, and R continuous derivatives, for some integer R and a mother wavelet ψ which is orthogonal to all polynomials of degree $R-1$ or less. Additionally, both ϕ and ψ should be supported on the interval $[-S+1, S]$ for some $S > R$. Coiflets satisfy these properties (see, e.g. Daubechies, 1992, Section 8.2). Let

$$\phi_k^B, \quad k = -R, -R+1, \ldots, R-2, R-1$$

and

$$\psi_k^B, \quad k = -S+1, -S+2, \ldots, S-2, S-1$$

be the boundary scaling functions and boundary wavelets, respectively. The support of these functions is contained in $[0, 2S-2]$ for $k \geq 0$ and in $[-(2S-2), 0]$ for $k < 0$. The coarse resolution level L should satisfy $6S-6<2^L$. Let

$$\phi_{jk}(x)=\begin{cases}2^{\frac{j}{2}}\phi_k^B(2^jx), & \text{for } k\in 0:(R-1),\\ 2^{\frac{j}{2}}\phi(2^jx-k), & \text{for } k\in (S-1):(2^j-S),\\ 2^{\frac{j}{2}}\phi_{k-2^j}^B(2^j(x-1)), & \text{for } k\in (2^j-R):(2^j-1),\end{cases}$$

and

$$\psi_{jk}(x)=\begin{cases}2^{\frac{j}{2}}\psi_k^B(2^jx), & \text{for } k\in 0:(S-2),\\ 2^{\frac{j}{2}}\psi(2^jx-k), & \text{for } k\in (S-1):(2^j-S),\\ 2^{\frac{j}{2}}\psi_{k-2^j}^B(2^j(x-1)), & \text{for } k\in (2^j-S+1):(2^j-1),\end{cases}$$

be the scaling functions and wavelets, respectively.

1.1.7 Constructing wavelet coefficients from discrete data

Let W be a $R\times R$ matrix and U be a $S\times R$ matrix defined by

$$W_{kl}=\int_0^\infty x^l\phi_k^B(x)dx \quad k=1,2,\dots,R; l=0,1,\dots,R-1$$

and

$$U_{jl}=j^l,\quad j=1,2,\dots,S; l=0,1,\dots,R-1.$$

Since U is of full rank, A^L can be constructed to be an $R\times S$ matrix satisfying $A^LU=W$. Similarly, the matrix A^R satisfies $A^R\overline{U}=\overline{W}$, where

$$\overline{W}_{kl}=\int_{-\infty}^0 x^l\phi_{-k}^B(x)dx,\quad k=1,2,\dots,R; l=0,1,\dots,R-1,$$

$$\overline{U}_{jl}=(-1)^lj^l,\quad j=1,2,\dots,S; l=0,1,\dots,R-1.$$

For a given sequence $Y_0,Y_1,\dots,Y_{N-1}$, we define the preconditioned sequence P_jY by

$$(P_jY)_k=\begin{cases}\sum_{i=0}^{S-2}A_{ki}^LY_i, & \text{for } k\in 0:(R-1),\\ Y_k, & \text{for } k\in (S-1):(N-S),\\ \sum_{i=1}^{S-1}A_{N-k,i}^RY_{N-i}, & \text{for } k\in (N-R):(N-1).\end{cases}$$

Let $\tilde{Y}$ be the boundary corrected discrete wavelet transform of $N^{-\frac{1}{2}}P_JY$ and c_A be the maximum of the eigenvalues of $A^L(A^L)^T$ and $A^R(A^R)^T$. Due to the orthogonality of the boundary-corrected discrete wavelet transform, the variance of the elements of $\tilde{Y}$ is bounded by $\frac{c_A}{N}$. Let $\tilde{Y}^I$ ($\tilde{Y}_{jk} : L \leq j < J, S-1 \leq k \leq 2^j - S$) be the array of interior coefficients. It is an uncorrelated array of variables with variance $\frac{1}{N}$. For a detailed review of sampled data model and boundary coiflets, see, e.g., Antoniadis (1994) and Johnstone & Silverman (2004a).

1.1.8 Wavelet Series Estimator

Suppose that we have the data

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, n \tag{1.6}$$

where f is a function on [0,1] and $\epsilon_i's$ are i.i.d. random variables with mean zero and common variance σ^2. Unlike parametic models, when a nonparametric regression model is considered, we assume that f belongs to some infinite dimensional collection of functions.

Assume that we have data from model (1.6). By using (1.1) and thresholding small wavelet coefficients, a smooth curve is obtained. Hence, we can construct a wavelet series estimator as follows

$$\hat{f}^W(x) = \sum_{j,k} \delta(\hat{d}_{jk})\psi_{jk}(x),$$

where $\delta(\hat{d}_{jk})$ is a hard or soft thresholding rule given respectively by

$$\delta(\hat{d}_{jk}) = \begin{cases} 0 & \text{if} \quad |d_{jk}| \leq \lambda \\ \hat{d}_{jk} & \text{if} \quad |d_{jk}| > \lambda, \end{cases}$$

$$\delta(\hat{d}_{jk}) = \begin{cases} 0 & \text{if} \quad |d_{jk}| \leq \lambda \\ \hat{d}_{jk} - \lambda & \text{if} \quad |d_{jk}| > \lambda \\ \hat{d}_{jk} + \lambda & \text{if} \quad |d_{jk}| < -\lambda, \end{cases}$$

where $\hat{d}_{jk} = \frac{1}{n}\sum_{i=1}^{n} y_i \psi_{jk}(x_i)$ and $\lambda > 0$. Such a wavelet series estimator was proposed by Donoho & Johnstone (1994a)

1. By applying DWT on the data $y = (y_1, \ldots, y_n)$ with $n = 2^M$ (w=Wy), obtain wavelet coefficients w.

2. Remove wavelet coefficients smaller than a chosen threshold (λ) and 'keep' or 'shrink' the other wavelet coefficients in order to construct new wavelet coefficients w^* from w.

3. Obtain $\hat{f}^W$, by reconstructing f from w^*, using the inverse discrete wavelet transformation (IDWT) ($\hat{f}^W = W^T w^*$).

For a smooth function, wavelet coefficients at the coarsest resolution level should not be removed. Hence, we have

$$\hat{f}^W(x) = \sum_k \hat{c}_{j_0k}\phi_{j_0k}(x) + \sum_{j\geq j_0}\sum_k \delta(\hat{d}_{jk})\psi_{jk}(x),$$

where $\hat{c}_{j_0k} = \frac{1}{n}\sum_{i=1}^n y_i\phi_{j_0k}(x_i)$. Under this scheme we obtain thresholded wavelet coefficients using either the hard or soft thresholding rule. Thresholding allows the data itself to decide which wavelet coefficients are significant. Hard thresholding (a discontinuous function) is a 'keep' or 'kill' rule, while soft thresholding (a continuous function) is a 'shrink' or 'kill' rule. Donoho & Jonhstone (1996) showed that this simple nonlinear method using hard thresholding achieves a risk within the logarithmic factor of the optimal minimax risk for either global or pointwise estimation. For more details on soft, hard or other types of thresholding in wavelet estimation see, e.g., Antoniadis, Bigot & Sapatinas (2001).

1.2 Penalized Model Selection

The basic idea of model selection is to assume that the unknown parameter may be well approximated by some family of models and estimate it under this assumption, although we know that this might not be the case. Suppose that we have at hand a family of models. The risk (or risk bound) corresponding to a given model is the sum of two components: a variance component which is proportional to the dimension of the model and a bias term that it is equal to the square of the distance between the true parameter and the model and results from the fact that we use an approximate model. If we knew the parameters, the optimal model would be the one that minimizes the risk or risk bound. However, in practice, we should develop a statistical procedure $\hat{m}(Y)$ or $\hat{D}(Y_i)_{i\geq 1}$ in order to choose the model from the data that has a risk which is close to the optimal risk. *Model selection* actually consists of two steps:

1. Choose a family of models S_m with $m \in M$, where M is an appropriate set of indices, and a collection of estimators $\hat{s}_m$ with values in S_m.

2. Choose a value $\hat{m}$ and set $\hat{s}_{\hat{m}}$ as the proposed estimator. The goal of *model selection* is to choose an estimator with a risk that is as close as possible to the minimal risk of the estimators $\hat{s}_m$, $m \in M$.

Model selection has various statistical applications in a wide range of models, see, e.g., Barron, Birgé & Massart (1999), Birgé & Massart (2001) and Massart (2007).

1.3 l_p-balls and Besov spaces

1.3.1 Minimaxity over weak and strong l_p-balls

Suppose that $\mu = (\mu_1, \ldots, \mu_n)$. Let $\|\mu\|_0 = \sharp\{i : \mu_i \neq 0, i = 1, \ldots, n\}$. An l_0-ball is defined by

$$l_0(\eta) = \{\mu \in \mathbb{R}^n : \|\mu\|_0 \leq \eta n\}.$$

An l_0-ball can be described as the set of vectors where the proportion of non-zero entries is bounded by η. A weak l_p-ball, $m_p(\eta)$, $0 < p < \infty$, with radius η is given by

$$m_p(\eta) = \{\mu \in \mathbb{R}^n : |\mu|_{(i)} \leq \eta(\frac{n}{i})^{\frac{1}{p}}, i = 1, \ldots, n\},$$

where $|\mu|_{(i)}$ is the $i-$th largest absolute value of the components of μ. Finally, a strong l_p-ball $l_p(\eta)$, $0 < p \leq \infty$, with radius η is given by

$$l_p(\eta) = \begin{cases} \{\mu \in \mathbb{R}^n : \frac{1}{n}\sum_{i=1}^n |\mu_i|^p \leq \eta^p\}, & 0 < p < \infty \\ \{\mu \in \mathbb{R}^n : |\mu_i| \leq n\eta \quad \forall i = 1, 2, \ldots, n\}, & p = \infty. \end{cases}$$

The above sets represent different ways to measure sparsity. An important relationship between weak and strong l_p-balls is as follows

$$l_p(\eta) \subseteq m_p(\eta) \nsubseteq l_{p'}(\eta), \quad p^{'} > p,$$

(see, e.g., Abramovich, Grinshtein & Pensky, 2007).

Let

$$y_i = \mu_i + \epsilon_i, \tag{1.7}$$

where $\epsilon_i \backsim N(0, \sigma^2)$. Define

$$R_n(l_p(\eta)) = \inf_{\hat{\mu}} \sup_{\mu \in l_p(\eta)} \mathbb{E}\|\hat{\mu} - \mu\|_2^2,$$

where the infimum is taken over all estimators (i.e. measurable functions) $\hat{\mu}$ of μ based on observations from (1.7). Donoho, Johnstone, Hoch & Stern (1992), Donoho & Johnstone (1994b, 1996) and Johnstone (1994) gave rates of convergence $R_n(F(\eta))$, where

$$F(\eta) = \begin{cases} l_0, & \\ l_p, \quad 0 < p \leq \infty, & \text{or} \\ m_p, \quad 0 < p < \infty. & \end{cases}$$

These rates will be used in the main chapters 2-4 of the Thesis.

1.3.2 Besov spaces and Besov balls

We define $\Delta_\epsilon f(x) = f(x+\epsilon) - f(x)$, $\Delta_\epsilon^2 f(x) = \Delta_\epsilon(\Delta_\epsilon f)(x)$ and similarly $\Delta_\epsilon^R(x)$ for positive integer R. Let

$$\rho^R(t, f, \pi) = \sup_{|\epsilon| \leq t} \left\{ \int_0^1 \left|\Delta_\epsilon^R f(u)\right|^\pi du \right\}^{\frac{1}{\pi}}.$$

Then, for $R > s$, we define

$$B_{\pi,r}^s[0,1] = \left\{ f \text{ periodic} : \left[\int_0^1 \left(\frac{\rho^R(t, f, \pi)}{t^s} \right)^r \frac{dt}{t} \right]^{\frac{1}{r}} < \infty \right\},$$

with the integral replaced by the sum when $r = \infty$ and/or $\pi = \infty$. In particular, for $f \in L^p[0,1]$,

$$f = \sum_{j,k} \beta_{jk} \Psi_{jk} \in B_{\pi,r}^s[0,1] \Longleftrightarrow \sum_{j \geq 0} 2^{j(s+1/2-1/\pi)r} \Big(\sum_{0 \leq k \leq 2^j} |\beta_{jk}|^\pi \Big)^{r/\pi} < \infty.$$

The parameter s measures the degree of smoothness of the function, while the integration parameters π and r indicate the type of norms used to measure smoothness.

Chapter 2

Adaptive Quadratic Functional Estimation of a weighted density by model selection

This chapter of the thesis consists of two main theorems and their proofs. Theorem 2.3.1 shows that the minimax rate of convergence, that is optimal in the case of direct data, can be also attained by the proposed estimator for $\theta = \int f^2$ in the case of weighted data. Theorem 2.5.1 derives the information bound for the problem of estimating $\theta = \int f^2$, when weighted data are available.

2.1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f.) F and probability density function (p.d.f) f with respect to the Lebesgue measure on the real line $\mathbb{R} = (-\infty, \infty)$. In practice, it sometimes happens that such direct data are not available. There are several settings that lead to weighted data sets. Weighted distributions are used in statistics to model sampling in the presence of selection bias. Observations which do not have an equal chance of being selected lead to this sampling scheme which can be described in the following way: let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by

$$g_w(y) = \frac{w(y) f(y)}{\mu_w}, \tag{2.1}$$

where the weight function w is known, satisfies $w(y) > 0$ for all y and $\mu_w = \mathbb{E}(w(X)) < \infty$ (see, e.g., Patil, Rao & Zelen, 1988). The restriction $w(y) > 0$ for all y is necessary for identifiability reasons. This constraint guarantees that g_w is indeed a p.d.f. (see, e.g., Vardi, 1982, 1985).

When the probability that an observation is selected is proportional to its size, i.e., when $w(y) = y$, length-biased data arise. Meta-analysis, the visibility bias in aerial survey techniques, line transect sampling, and sampling from queues or telephone networks are some examples of settings where weighted data arise (see, e.g., Cox, 1969; Vardi, 1982).

Cox (1969) proposed an estimator of F given by

$$\tilde{F}(y) = n^{-1}\hat{\mu}_w \sum_{i=1}^{n} w^{-1}(Y_i)\mathbb{I}_{(-\infty,y]}(Y_i),$$

where $\hat{\mu}_w = n\left(\sum_{i=i}^{n} w^{-1}(Y_i)\right)^{-1}$ and $\mathbb{I}_A(y) = 1$ if $y \in A$ and 0 otherwise. Hence, this estimator can be interpreted as the empirical distribution function for weighted data. Vardi (1982, 1985) showed that $\tilde{F}$ is the nonparametric maximum likelihood estimator of F for this case, and that $\hat{\mu}_w$ is a $\sqrt{n}$-consistent estimator of μ_w.

Kernel estimators of f for weighted data from model (2.1) were proposed by Bhattacharyya, Franklin and Richardson (1988) and Jones (1991), while their multivariate extensions were considered in Ahmad (1995). Asymptotic properties of these estimators were considered in Wu (1995) and Wu & Mao (1996), for a Hölder class of p.d.f.'s. A Fourier series estimator of f for weighted data from model (2.1) was proposed by Jones & Karunamuni (1997), while a transformation-based estimator was suggested by El Barmi & Simonoff (2000). Efromovich (2004a) suggested a blockwise shrinkage estimator of f for weighted data from model (2.1) and showed that this estimator is sharp minimax, that is, the proposed estimator attains the optimal constant and rate of convergence, for a Sobolev class of p.d.f.'s. Additionally a second-order sharp minimax estimator for F, via a projection on trigonometric bases, and of f by differentiaton, for an analytic class of c.d.f.'s, was derived in Efromovich (2004b).

Let X be a random variable with c.d.f. F and p.d.f f with respect to Lebesgue measure on the real line $\mathbb{R}$, and let $f \in L^2(\mathbb{R})$ (the space of squared-integrable functions on $\mathbb{R}$). We consider the problem of estimation of $\int f^2$, assuming f belongs to some smooth class of p.d.f.'s. This functional appears, e.g., in the Pitman efficacy of the Wilcoxon signed-rank statistic, in rank tests based on residuals in the linear model and in the asymptotic variance of the Hodges-Lehmann estimator (see, e.g., Aubuchon & Hettmansperger, 1984; Draper, 1988; Ritov & Bickel, 1990). Additionally, an estimate of this quantity can be used in test statistics based on the L^2-distance (see, e.g., Fromont & Laurent, 2006; Butucea 2007).

If direct data are available, then optimal solutions to this problem are well known. Bickel & Ritov (1988) proposed an estimator of $\int \left(f^{(k)}\right)^2$, where $f^{(k)}$ is the k-th derivative of f, for p.d.f.'s satisfying the Hölder condition on $f^{(m)}$ with smoothness parameter α. Although their estimator is asymptotically efficient when $m + \alpha > 2k + 1/4$ and rate optimal for $k < m + \alpha \leq 2k + 1/4$, it is non-adaptive since it depends on unknown parameters. Birgé & Massart (1995) proposed non-adaptive, $\sqrt{n}$-consistent estimators for functionals of the form $\int \phi(f, f', \ldots, f^{(k)}, \cdot)$, for f belonging to some smooth class of p.d.f.'s with smoothness parameter s satisfying $s \geq 2k + \frac{1}{4}$, and proved that $\int \phi(f, f', \ldots, f^{(k)}, \cdot)$ cannot be estimated at a rate faster than $n^{-\frac{4(s-k)}{4s+1}}$ if $s < 2k + \frac{1}{4}$. Laurent (1996, 1997) extended these results and built non-adaptive and asymptotically efficient estimators of more general functionals.

Finally, Laurent (2005) constructed an adaptive and asymptotically optimal (in the minimax sense) estimator of $\int f^2$, for p.d.f's belonging to some smooth class of densities. We construct an adaptive estimator of $\int f^2$, when weighted data are available. An estimate of this functional could be used when statistical procedures developed for direct data (e.g., tests based on L^2-distance) are adapted to weighted data. The proposed estimator is shown to attain the minimax rate of convergence that is optimal in the case of direct data for the same class of p.d.f.'s, under the assumption that the biasing function $w(y)$ is bounded away from 0 and ∞. The information bound for the problem of estimating $\int f^2$ when weighted data are available is also derived and compared with the information bound for the case of direct data (see, e.g., Pfanzagl, 1982; Bickel & Ritov, 1988; Laurent, 1996). A small simulation study is conducted in order to illustrate the usefulness of the proposed estimator in finite sample situations.

2.2 Estimation of $\int f^2$ using weighted data by model selection

We consider below the problem of estimating $\theta = \int f^2$ based on a weighted sample from model (2.1), for p.d.f.'s which are uniformly bounded and belong to a certain Besov body. The approach adapted, modifies the method used by Laurent (2005) for the case of direct data, borrowing also ideas from Jones (1991), Jones & Karunamuni(1997) and Brunel, Comte & Guilloux (2007) in order to take into account the selection bias. We project f onto the space generated by the constant piecewise functions on the intervals $\left(\frac{k}{D}, \frac{k+1}{D}\right]$, where $k \in \mathbb{Z}$ and D is a natural number. The projection of f onto this space is given by

$$f_D = \sum_{k\in\mathbb{Z}} \alpha_{k,D}\, p_{k,D},$$

where $p_{k,D} = \sqrt{D}\mathbb{I}_{\left(\frac{k}{D}, \frac{k+1}{D}\right]}$ and $\alpha_{k,D} = \int f p_{k,D}$. It is easy to see that

$$\tilde{\theta}_D = \frac{{\mu_w}^2}{n(n-1)} \sum_{\substack{1\le i,j\le n \\ i\neq j}} \sum_{k\in\mathbb{Z}} \frac{p_{k,D}(Y_i)}{w(Y_i)} \frac{p_{k,D}(Y_j)}{w(Y_j)} \tag{2.2}$$

is an unbiased estimator for $\theta_D = \int f_D^2$. *Assumption 1.* Let w be a real-valued function satisfying $0 < w_1 \le w(y) \le w_2 < \infty$ for all $y \in \mathbb{R}$. Under Assumption 1, and for uniformly bounded densities f, i.e., $\|f\|_\infty = \sup_{y\in\mathbb{R}} |f(y)| \le M$ for some finite constant $M > 0$, it is easy to check that

$$\mathbb{E}(\tilde{\theta}_D - \theta)^2 \le \left\{(\theta_D - \theta)^2 + C(M,w)\left(\frac{D}{n^2} + \frac{1}{n}\right)\right\},$$

where $C(M,w)$ is an absolute constant depending on M, w_1 and w_2. Under Assumption 1 and for uniformly bounded densities, it is easy to check that

$$\mathbb{E}(\tilde{\theta}_D - \theta)^2 \le \{(\theta_D - \theta)^2 + C(M,w)\left(\frac{D}{n^2} + \frac{1}{n}\right)\},$$

where $C(M,w)$ is a positive constant depending on M, w_1 and w_2 only with $M > 0$ and $\|f\|_\infty \le M$.

According to the ideas presented in Laurent & Massart (2000), the optimal choice of D should minimize the quantity $\theta - \theta_D + \frac{\sqrt{D}}{n}$ or, equivalently, maximize the quantity $\theta_D - \frac{\sqrt{D}}{n}$.

Therefore, we consider the following estimator

$$\tilde{\theta} = \sup_{D \in \mathcal{D}_n} \left(\tilde{\theta}_D - \text{pen}(D) \right), \tag{2.3}$$

where pen(D) is given by

$$\text{pen}(D) = \frac{\kappa}{n} \sqrt{\left(\tilde{\theta}_D + 1 \right) D \log(D+1)}, \tag{2.4}$$

for some constant $\kappa > 0$. However, μ_w is unknown in practice and therefore $\tilde{\theta}_D$ should be replaced by

$$\hat{\theta}_D = \frac{\hat{\mu}_w^2}{n(n-1)} \sum_{\substack{1 \leq i,j \leq n \\ i \neq j}} \sum_{k \in \mathbb{Z}} \frac{p_{k,D}(Y_i)}{w(Y_i)} \frac{p_{k,D}(Y_j)}{w(Y_j)}, \tag{2.5}$$

where $\hat{\mu}_w$ is a $\sqrt{n}$-consistent estimator of μ_w (see Section 4.32). Therefore, a natural adaptive estimator for $\theta = \int f^2$ is given by

$$\hat{\theta} = \sup_{D \in \mathcal{D}_n} \left(\hat{\theta}_D - \text{pen}_u(D) \right), \tag{2.6}$$

where pen$_u(D)$ is given by

$$\text{pen}_u(D) = \frac{\kappa}{n} \sqrt{\left(\hat{\theta}_D + 1 \right) D \log(D+1)}. \tag{2.7}$$

In the sequel, the notation C is used for absolute constants whose values may vary from one line to another. The dependency of a constant on some parameter or the bounds of the weight function is implied in the following way: For example $C(\alpha, R, M)$ denotes an absolute constant depending on α, R and M, while $C(w)$ denotes a constant depending on w_1 and w_2.

2.3 Upper risk bounds

Let $\phi(x) = \mathbb{I}_{(0,1]}(x)$ and $\psi(x) = \mathbb{I}_{[0,\frac{1}{2}]}(x) - \mathbb{I}_{(\frac{1}{2},1]}(x)$, and for any $j \in \mathbb{N}$, $k \in \mathbb{Z}$, let

$$\phi_{jk}(x) = 2^{j/2} \mathbb{I}_{(0,1]} \left(2^j x - k \right) \quad \text{and} \quad \psi_{jk}(x) = 2^{j/2} \left[\mathbb{I}_{[0,\frac{1}{2}]} \left(2^j x - k \right) - \mathbb{I}_{(\frac{1}{2},1]} \left(2^j x - k \right) \right].$$

Then, the functions $\{\phi_{Jk}, \psi_{jk} : \; j \geq J, \, k \in \mathbb{Z}\}$ forms an orthonormal basis for $L^2(\mathbb{R})$, which is the well-known Haar basis of $L^2(\mathbb{R})$. Therefore, any f can be represented (in the L^2-sense) by a Haar series as

$$f = \sum_{k \in \mathbb{Z}} \alpha_{Jk}(f) \phi_{Jk} + \sum_{j=J}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{jk}(f) \psi_{jk},$$

where $\alpha_{jk}(f) = \int f\phi_{jk}$ and $\beta_{jk}(f) = \int f\psi_{jk}$. Let now $\mathcal{F}(\alpha, R, M)$ be the class of p.d.f.'s f which are uniformly bounded by some finite constant $M > 0$ and the sequence of coefficients onto the Haar basis belongs to the following Besov body

$$\mathcal{B}_{\alpha,2,\infty}(R) = \{f \mid \beta(f) = (\beta_{jk})_{j\geq J, k\in\mathbb{Z}}, \sum_{k\in\mathbb{Z}} \beta_{jk}^2 \leq R^2 2^{-2j\alpha}, \forall j \geq J\},$$

for some finite constants $\alpha, R > 0$, that is, we consider the class of p.d.f.'s

$$\mathcal{F}(\alpha, R, M) = \{f \mid \beta(f) \in \mathcal{B}_{\alpha,2,\infty}(R), \|f\|_\infty \leq M\}. \tag{2.8}$$

Theorem 2.3.1 below shows that the proposed adaptive estimator of $\theta = \int f^2$ based on weighted data converges at the rate which is optimal in the direct data case, uniformly over the class of p.d.f.'s $\mathcal{F}(\alpha, R, M)$ given by (2.8) for all $\alpha, R, M > 0$.

Theorem 2.3.1. *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by* (2.1)*, with weight function w being continuous and satisfying Assumption 1. Consider the class of p.d.f.'s $\mathcal{F}(\alpha, R, M)$ defined by* (2.8) *with $\alpha > 0$, and let*

$$\mathcal{D}_n = \left\{D \mid D \in \mathbb{N},\, D \leq n^2/\log^3(n)\right\}. \tag{2.9}$$

There exists some constant $\kappa_0 > 0$ such that if $\mathrm{pen}_u(D)$ is given by (2.7) for all $D \in \mathcal{D}_n$ with $\kappa \geq \kappa_0$, then, there exists some $n_0 := n_0(\alpha, R, M, w)$ such that $\hat{\theta}$, given by (2.6)*, satisfies the following inequalities*

- *For $\alpha > 1/4$,*
$$\sup_{f\in\mathcal{F}(\alpha,R,M)} \mathbb{E}(\hat{\theta} - \theta)^2 \leq \frac{C(\alpha, R, M, w)}{n},$$
- *for $0 < \alpha \leq \frac{1}{4}$*
$$\sup_{f\in\mathcal{F}(\alpha,R,M)} \mathbb{E}(\hat{\theta} - \theta)^2 \leq C(\alpha, R, M, w)\left(\frac{\sqrt{\log(nR^2)}}{n}\right)^{8\alpha/(1+4\alpha)}.$$

Remark 2.3.1 Theorem 2.3.1 gives a uniform bound of the mean squared error of $\hat{\theta}$, leading to the conclusion that $\hat{\theta}$ is an adaptive and $\sqrt{n}$-consistent estimator of $\theta = \int f^2$, uniformly over $\mathcal{F}(\alpha, R, M)$ with $\alpha > \frac{1}{4}$, and it also achieves the minimax rate of convergence $(\frac{\sqrt{\log n}}{n})^{4\alpha/(1+4\alpha)}$ which is optimal (in the minimax sense) in the case of direct data when $0 < \alpha \leq \frac{1}{4}$. The fact that the minimax rate of convergence that is optimal in the case of direct data can be also attained in the presence of selection bias is consistent with analogous results for density estimation (see, e.g., Wu, 1995; Wu & Mao, 1996; Efromovich, 2004a) and distribution estimation (see, e.g., Efromovich, 2004b).

Remark 2.3.2 The estimator $\hat{\theta}$ can be used in tests of the null hypothesis $H_0 : f = f_0$, based on the L^2-distance in order to estimate $\int (f - f_0)^2$, in the case of weighted data. More precisely, under the assumptions of Theorem 2.3.1, the L^2-distance of f and f_0 can be estimated by $\hat{\theta} - \int f_0^2 - 2\hat{\mu}_w \sum_{i=1}^{n} \frac{f_0(Y_i)}{w(Y_i)}$ at the minimax rate of convergence that is optimal in the case of direct data.

Remark 2.3.3 If $\beta(f) \in \mathcal{B}_{\alpha,2,\infty}(R)$ with $\alpha > 1/2$, then f is uniformly bounded and the restriction $\|f\|_\infty \leq M$ is not needed in the definition of $\mathcal{F}(\alpha, R, M)$ (see, e.g., inequality (8.15) of Proposition 8.3 in Härdle, Kerkyacharian, Picard & Tsybakov, 1998).

Remark 2.3.4 The simple projection estimator $\hat{\theta}_n$ (i.e, $\hat{\theta}_D$ given in (2.5) with $D = n$) can be shown to be uniformly $\sqrt{n}$-consistent for all $f \in \mathcal{F}(\alpha, R, M)$ with $\alpha \geq \frac{1}{4}$. On the other hand, in addition, the penalized estimator given by (2.6) also attains the minimax rate of convergence, that is optimal in the case of direct data, for all $f \in \mathcal{F}(\alpha, R, M)$ with $0 < \alpha \leq \frac{1}{4}$. Furthermore, it performs better in finite sample situations, as it will be illustrated in Section 2.4.

Remark 2.3.5 The assumption $0 < w_1 \leq w(y) \leq w_2 < \infty$, $y \in [0, 1]$, is very common in density estimation for weighted data (see, e.g., Efromovich, 2004a; Brunel, Comte & Guilloux, 2005; Wu & Mao, 1996). The only difference with Assumption 1 considered above is that we require $0 < w_1 \leq w(y) \leq w_2 < \infty$ for all y, in order to cover the case of densities with non-compact support. Below, we report some examples that arise in practical settings leading to weighted data with weight function satisfying Assumption 1.

(i) Let $1 - w(y)$ be a proportion of the frequency of the variable X that is missing (see, e.g., Efromovich, 2004a). Then, weighted data from model (2.1) arise. Let $w(y) = 0.1$ for $y < 0$, $w(y) = 0.9y + 0.1$ for $0 \leq y \leq 1$ and $w(y) = 1$ for $y \geq 1$. The missing proportion is constant for $y < 0$, decreases in the interval $[0, 1]$ and remains 0 for $y > 1$. A generalization of this weight function is $w(y) = b$ for $y < 0$, $w(y) = cy + b$, for $0 < y < \frac{1-b}{c}$ and $w(y) = 1$ for $y > \frac{1-b}{c}$, $0 < c, b < 1$, where the missing proportion is constant for $y < 0$, decreases in the interval $[0, \frac{1-b}{c}]$ and remains 0 for $y > \frac{1-b}{c}$.

(ii) Line transect sampling is another example where weighted data arise (see, e.g., Efromovich, 2004a). If we are interested to estimate the abundance of plants or animals of a particular species in a given region, we can use line transects. This essentially means that an observer moves along fixed paths and includes the sighted clusters of objects of interest in the sample. It is obvious that larger clusters have a larger probability to be included in the sample. An appropriate weight function would be $w(y) = cy + b$, for $0 < y < \frac{1-b}{c}$ and $w(y) = 1$ for $y > \frac{1-b}{c}$, $0 < c, b < 1$.

(iii) The purpose of a photographic survey described by Patil (2002) was to estimate the abundance of the deep-sea red crab. The data can be analyzed using the composite weight function of the form $w(y) = (a + by)v(y, \theta)$ for $0 < y < c$ and $w(y) = (a + bc)v(y, \theta)$, for $y > c$, where $a, b > 0$, c is a large positive constant and the sighting function $v(y, \theta)$ represents the sighting-distance bias that is usually bounded away from zero.

(iv) Meta-analysis studies the publication-selection bias and the heterogeneity that might exist among different studies. Appropriate weight functions that have been found include, (a) half-normal model $w(y) = \exp[-\beta p^2(y)]$ and (b) negative exponential model $w(y) = \exp[-\beta p(y)]$, where $\beta > 0$ and $p(y)$ is the P-value when the resulting test statistic takes the value y (see, e.g., Patil & Taillie, 1989).

2.4 Simulations

We present a small simulation study to illustrate the usefulness of the proposed estimator in finite sample situations. We use the weight function

$$w(y) = \begin{cases} 1.10^{-40} & \text{if } y < 1.10^{-40}, \\ y & \text{if } 1.10^{-40} \leq y \leq 40, \\ 40 & \text{if } y > 40, \end{cases}$$

and five different distributions, i.e., (I) χ^2-distribution with 3 degrees of freedom, (II) Beta distribution with parameters $\alpha = 3$ and $\beta = 1$, (III) Beta distribution with parameters $\alpha = 5$ and $\beta = 4$, (IV) Beta distribution with parameters $\alpha = 5$ and $\beta = 2$, and (V) Gamma

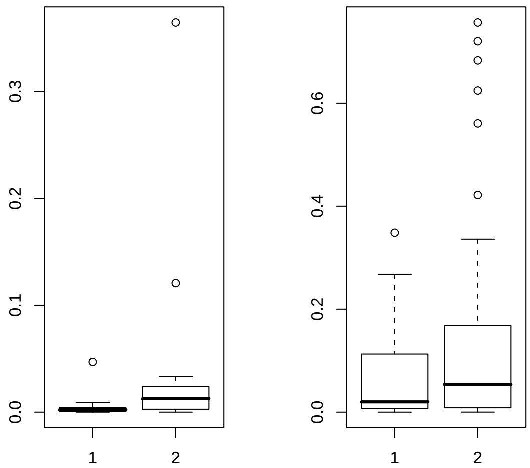

Figure 2.1: MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases I and II.

distribution with parameters $\alpha = 3$ and $\lambda = 1$. In each case, M=50 samples of size $n = 50$ and 100 were used in order to construct the boxplots of the mean squared error in Figures 2.1-2.3. We just present the boxplots of MSE for $n = 50$. For the proposed estimator we set $\kappa = 2$. In Figures 2.1-2.2, we compare the proposed estimator with a simple projection estimator $\hat{\theta}_n$ described in Remark 2.3.4. The boxplot on the right represents the MSE of the projection estimator, while the boxplot on the left represents the MSE of the proposed estimator. Obviously, the proposed estimator outperforms the projection estimator in all cases. Although not reported here, the proposed estimator is still better than the projection estimator for larger sample sizes. In Figure 2.3, we compare the MSE of pseudoestimator (2.3) with the MSE of estimator (2.6), in cases (II) and (IV). The boxplot on the right represents the MSE of the pseudoestimator, while the boxplot on the left represents the MSE of the proposed estimator. It is evident that the estimation of μ_w deteriorates the quality of the estimator.

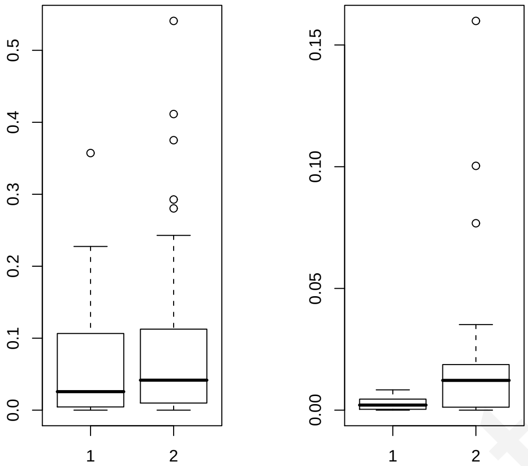

Figure 2.2: MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases III and V.

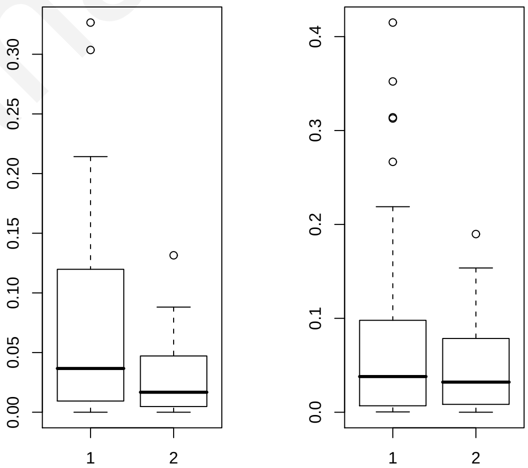

Figure 2.3: MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases II and IV.

2.5 The information bound for estimating $\int f^2$ using weighted data

Theorem 2.5.1 below provides the information bound for the problem of estimating $\theta = \int f^2$, when weighted data are available. For some finite $M > 0$ let $\mathcal{H}$ be a class of p.d.f's defined by

$$\mathcal{H} = \{f | f \in L^2(\mathbb{R}), \|f\|_\infty \leq M\}.$$

Theorem 2.5.1. *Let f be a member of $\mathcal{H}$. Then, the information bound, $I_w(f)$, for the estimation of $\theta = \int f^2$, using a weighted sample with weight function w satisfying Assumption 1, is given by*

$$I_w(f) = 4\,\mu_w \int \frac{f^3}{w} - 4\left(\int f^2\right)^2.$$

Remark 2.5.1 The information bound, $I_d(f)$, for the estimation of $\theta = \int f^2$ based on a direct sample (see, e.g., Pfanzagl, 1982; Bickel & Ritov, 1988; Laurent, 1996), equals

$$I_d(f) = 4\int f^3 - 4\left(\int f^2\right)^2.$$

It is easy to see that for any uniform distribution $U(a,b)$, with $a < b$, $I_w(f)$ is no smaller than $I_d(f)$ since

$$\mu_w \int \frac{f^3}{w} = \frac{1}{(b-a)^2} d(f,w) = d(f,w) \int f^3 \geq \int f^3,$$

where $d(f,w) = \mu_w \int \frac{f}{w} \geq 1$, by Jensen's inequality, with equality if and only if $w \equiv 1$ (see, e.g., Efromovich, 2004b). However, there are cases where $I_w(f)$ is (strictly) smaller than $I_d(f)$. For example, let $w(x) = 1 - 0.9x$ for all $x \in (0,1)$. Let f be the p.d.f of a Beta distribution with parameters $\alpha = 1$ and $\beta = 3$. Then, using numerical integration (performed in `R`, version 2.4.0), or by direct calculations, we can compute

$$\mu_w \int \frac{f^3}{w} = 3.4209404 \quad \text{and} \quad \int f^3 = \frac{27}{7},$$

thus concluding that $I_w(f)$ is smaller than $I_d(f)$. The above observations lead to the conclusion that model sampling in the presence of selection bias can either improve or worsen the information bound in the problem of estimating $\theta = \int f^2$. Analogous conclusions regarding density estimation based on weighted data can be found in Cox (1969) and

Efromovich (2004b).

Remark 2.5.2 Theorem 2.5.1 has the following implication. If an estimator, say T_n, of θ based on a weighted sample given by (2.1), with w satisfying Assumption 1 and f belonging to $\mathcal{H}$, satisfies

$$\sqrt{n}(T_n - \theta) \longrightarrow N\big(0, I_w(f)\big), \quad \text{in distribution,}$$

and

$$\lim_{n \longrightarrow \infty} n\mathbb{E}(T_n - \theta)^2 = I_w(f),$$

then T_n is asymptotically efficient (see, e.g., Laurent, 1996).

2.6 Appendix: Proofs

2.6.1 Proof of Theorem 2.3.1

The proof of Theorem 2.3.1 is broken into several parts. We first prove a lemma and three propositions which are used in the proof of Theorem 2.3.1.

Let

$$U_n(H_D) = \frac{{\mu_w}^2}{n(n-1)} \sum\sum_{\substack{1 \le i,j \le n \\ i \ne j}} \frac{H_D(Y_i, Y_j)}{w(Y_i)w(Y_j)} \tag{2.10}$$

and

$$H_D(x,y) = \sum_{k \in \mathbb{Z}} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w} \right) \left(p_{k,D}(y) - \frac{\alpha_{k,D} w(y)}{\mu_w} \right). \tag{2.11}$$

Lemma 2.6.1. *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by (2.1), where f belongs to $L^2(\mathbb{R})$ and the weight function w is continuous and satisfies Assumption 1. There exist some positive constants κ_1, κ_2 and κ_3 for which the following inequality holds*

$$\mathbb{P}\left\{ \left|U_n(H_D)\right| > \frac{1}{n-1}\left[\kappa_1 \sqrt{D\theta_D t} + \kappa_2 \|f\|_\infty t + \frac{\kappa_3 D t^2}{n} \right] \right\} \le 5.6 \exp(-t).$$

Proof of Lemma 2.6.1. Let

$$g(x,y) = \frac{H_D(x,y)\mu_w^2}{w(x)w(y)},$$

and

$$A_1^2 = n(n-1)\mathbb{E}(g(Y_1,Y_2)^2),$$

$$A_2 = \sup\left\{\left|\mathbb{E}(\sum_{i=2}^{n}\sum_{j=1}^{i-1} g(Y_i,Y_j)\alpha_i(Y_i)b_j(Y_j))\right| : \sum_{i=2}^{n}\mathbb{E}(\alpha_i^2(Y_i)) \leq 1, \sum_{j=1}^{n-1}\mathbb{E}(b_j^2(Y_j)) \leq 1\right\},$$

$$A_3^2 = n\sup_x\left[\mathbb{E}\{g^2(Y,x)\}\right],$$

$$A_4 = \sup_{x,y}|g(x,y)|.$$

Consider now the quantities B,C and Δ, defined in Theorem 3.4 of Houdré & Reynaud (2003). It is easy to see that the following inequalities hold

$$B^2 = \max\left\{\sup_{x,i}\left[\sum_{j=1}^{i-1}\mathbb{E}(g^2(x,Y_1))\right],\ \sup_{x,j}\left[\sum_{i=j+1}^{n}\mathbb{E}(g^2(Y_1,x))\right]\right\} \leq n\sup_x[\mathbb{E}(g^2(Y,x))] = A_3^2,$$

$$C^2 \leq n(n-1)\mathbb{E}(g^2(Y_1,Y_2)) = A_1^2,$$

$$\begin{aligned}\Delta &\leq \sup\left\{\left|\mathbb{E}(\sum_{i=2}^{n}\sum_{j=1}^{i-1} g(Y_i,Y_j)\alpha_i(Y_i)b_j(Y_j))\right| : \sum_{i=2}^{n}\mathbb{E}\left(\alpha_i^2(Y_i)\right) \leq 1, \sum_{j=1}^{n-1}\mathbb{E}\left(b_j^2(Y_j)\right) \leq 1\right\}\\ &= A_2.\end{aligned}$$

Evaluation of A_1

Note that

$$\begin{aligned} g^2(Y_1,Y_2) &= \frac{{\mu_w}^4}{w^2(Y_1)w^2(Y_2)}\left[\sum_{k\in\mathbb{Z}}\left(p_{k,D}(Y_1) - \frac{\alpha_{k,D}w(Y_1)}{\mu_w}\right)^2\left(p_{k,D}(Y_2) - \frac{\alpha_{k,D}w(Y_2)}{\mu_w}\right)^2\right.\\ &+ \sum_{\substack{k\in\mathbb{Z}\\ k'\neq k}}\sum\left(p_{k,D}(Y_1) - \frac{\alpha_{k,D}w(Y_1)}{\mu_w}\right)\left(p_{k,D}(Y_2) - \frac{\alpha_{k,D}w(Y_2)}{\mu_w}\right)\\ &\times \left.\left(p_{k',D}(Y_1) - \frac{\alpha_{k',D}w(Y_1)}{\mu_w}\right)\left(p_{k',D}(Y_2) - \frac{\alpha_{k',D}w(Y_2)}{\mu_w}\right)\right].\end{aligned}$$

Hence, it is easy to see that

$$
\begin{aligned}
\mathbb{E} \quad & \left\{g^2(Y_1,Y_2)\right\} = \\
= \quad & \left[\sum_{k\in\mathbb{Z}} \mathbb{E}\left(\frac{\mu_w^2}{w^2(Y_1)}\left(p_{k,D}(Y_1) - \frac{\alpha_{k,D}w(Y_1)}{\mu_w}\right)^2\right)\right. \\
\times \quad & \mathbb{E}\left(\frac{\mu_w^2}{w^2(Y_2)}\left(p_{k,D}(Y_2) - \frac{\alpha_{k,D}w(Y_2)}{\mu_w}\right)^2\right) \\
+ \quad & \sum_{k\in\mathbb{Z}}\sum_{k^{'}\neq k} \mathbb{E}\left(\frac{\mu_w^2}{w^2(Y_1)}\left(p_{k,D}(Y_1) - \frac{\alpha_{k,D}w(Y_1)}{\mu_w}\right)\left(p_{k',D}(Y_1) - \frac{\alpha_{k',D}w(Y_1)}{\mu_w}\right)\right) \\
\times \quad & \left.\mathbb{E}\left(\frac{\mu_w^2}{w^2(Y_2)}\left(p_{k,D}(Y_2) - \frac{\alpha_{k,D}w(Y_2)}{\mu_w}\right)\left(p_{k',D}(Y_2) - \frac{\alpha_{k',D}w(Y_2)}{\mu_w}\right)\right)\right] \\
= \quad & \sum_{k\in\mathbb{Z}}\left[\int (p_{k,D} - \frac{\alpha_{k,D}w}{\mu_w})^2 \frac{f\mu_w}{w}\right]^2 \\
+ \quad & \sum_{k\in\mathbb{Z}}\sum_{k^{'}\neq k}\left[\int \left(p_{k,D} - \frac{\alpha_{k,D}w}{\mu_w}\right)\left(p_{k',D} - \frac{\alpha_{k',D}w}{\mu_w}\right)\frac{f\mu_w}{w}\right]^2 \\
= \quad & \sum_{k\in\mathbb{Z}}\left[\int \frac{f\mu_w p_{k,D}^2}{w} - \alpha_{k,D}^2\right]^2 + \sum_{k\in\mathbb{Z}}\sum_{k^{'}\neq k}\left[-\alpha_{k,D}\alpha_{k',D}\right]^2.
\end{aligned}
$$

Additionally, one can see that

$$
\sum_{k\in\mathbb{Z}}\left[\int \frac{f\mu_w p_{k,D}^2}{w} - \alpha_{k,D}^2\right]^2 \leq 2\sum_{k\in\mathbb{Z}}\left[\int \frac{f\mu_w p_{k,D}^2}{w}\right]^2 + 2\left[\sum_{k\in\mathbb{Z}}\alpha_{k,D}^4\right].
$$

Using the above inequality and $\sum_{k\in\mathbb{Z}}\alpha_{k,D}{}^2 \leq D$, one arrives at

$$
\begin{aligned}
\mathbb{E}\left(g^2(Y_1,Y_2)\right) \quad & \leq \quad 2\sum_{k\in\mathbb{Z}}\left(\int \frac{f\mu_w p_{k,D}^2}{w}\right)^2 + 2\left(\sum_{k\in\mathbb{Z}}\alpha_{k,D}{}^2\right)^2 \\
& \leq \quad 2D\theta_D + 2\frac{{w_2}^2}{{w_1}^2}D\left(\sum_{k\in\mathbb{Z}}\alpha_{k,D}{}^2\right) = C_1(w)D\theta_D.
\end{aligned}
$$

Therefore, we get

$$
A_1 \leq C_1(w)\sqrt{n(n-1)D\theta_D}, \tag{2.12}
$$

where

$$
C_1(w) = \sqrt{2(1+\frac{w_2^2}{w_1^2})}.
$$

Evaluation of A_2

It is easy to check that the following inequality holds

$$\begin{aligned}
A_2 \;\leq\; & \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}}\left\{\mathbb{E}\left[\frac{{\mu_w}^2|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)w(Y_j)}p_{k,D}(Y_i)p_{k,D}(Y_j)\right]+\alpha_{k,D}^2\mathbb{E}\left(|\alpha_i(Y_i)||b_j(Y_j)|\right)\right.\\
+\; & \left.\mathbb{E}\left[\frac{\mu_w\alpha_{k,D}p_{k,D}(Y_i)|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)}\right]+\mathbb{E}\left[\frac{\mu_w\alpha_{k,D}p_{k,D}(Y_j)|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_j)}\right]\right\}.
\end{aligned}$$

Evaluation of the first term of A_2

Using repeatedly the Cauchy-Schwartz inequality, one arrives at

$$\begin{aligned}
& \sum_{k\in\mathbb{Z}}\mathbb{E}\left(\frac{\mu_w^2|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)w(Y_j)}p_{k,D}(Y_i)p_{k,D}(Y_j)\right)\leq\\
\leq\; & \left[\sum_{k\in\mathbb{Z}}\mathbb{E}^2\left(\frac{\mu_w p_{k,D}(Y_i)|\alpha_i(Y_i)|}{w(Y_i)}\right)\right]^{\frac{1}{2}}\left[\sum_{k\in\mathbb{Z}}\mathbb{E}^2\left(\frac{\mu_w p_{k,D}(Y_j)|b_i(Y_j)|}{w(Y_j)}\right)\right]^{\frac{1}{2}}\\
\leq\; & \frac{\|f\|_\infty w_2^2}{w_1^2}\left[\mathbb{E}\left(b_j^2(Y_j)\right)\right]^{\frac{1}{2}}\left[\mathbb{E}\left(\alpha_i^2(Y_i)\right)\right]^{\frac{1}{2}}.
\end{aligned}$$

Since

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left\{\left[\mathbb{E}\left(\alpha_i^2(Y_i)\right)\mathbb{E}\left(b_j^2(Y_j)\right)\right]\right\}^{\frac{1}{2}}\leq n\left[\sum_{i=2}^{n}\sum_{j=1}^{n-1}\mathbb{E}\left(\alpha_i^2(Y_i)\right)\mathbb{E}\left(b_j^2(Y_j)\right)\right]^{\frac{1}{2}},$$

one obtains

$$\begin{aligned}
& \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}}\mathbb{E}\left(\frac{\mu_w^2 p_{k,D}(Y_i)p_{k,D}(Y_j)|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)w(Y_j)}\right)\leq\\
\leq\; & \frac{\|f\|_\infty w_2^2}{w_1^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left[\mathbb{E}\left(\alpha_i^2(Y_i)\right)\mathbb{E}\left(b_j^2(Y_j)\right)\right]^{\frac{1}{2}}\leq\frac{\|f\|_\infty n w_2^2}{w_1^2}.
\end{aligned}$$

Evaluation of the second term of A_2

The following inequalities hold

$$\begin{aligned}
\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}}\alpha_{k,D}^2\mathbb{E}(|\alpha_i(Y_i)||b_j(Y_j)|) \;\leq\; & \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}}\alpha_{k,D}^2\left[\mathbb{E}(\alpha_i^2(Y_i))\mathbb{E}(b_j^2(Y_j))\right]^{\frac{1}{2}}\\
\leq\; & n\theta_D\\
\leq\; & n\|f\|_\infty.
\end{aligned}$$

Evaluation of the third term of A_2

Using the Cauchy-Schwartz inequality, we get

$$\begin{aligned}
\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}} \quad & \mathbb{E} \quad \left(\frac{\mu_w \alpha_{k,D} p_{k,D}(Y_i)|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)}\right) \\
& \le \quad \sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathbb{E}\left(\frac{f_D(Y_i)\mu_w|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_i)}\right) \\
& \le \quad \frac{w_2\|f_D\|_\infty}{w_1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathbb{E}^{\frac{1}{2}}(\alpha_i^2(Y_i))\mathbb{E}^{\frac{1}{2}}(b_j^2(Y_j)) \\
& \le \quad n\frac{w_2}{w_1}\|f\|_\infty.
\end{aligned}$$

Similarly, we can see that

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k\in\mathbb{Z}}\mathbb{E}\left(\frac{\mu_w \alpha_{k,D} p_{k,D}(Y_j)|\alpha_i(Y_i)||b_j(Y_j)|}{w(Y_j)}\right) \le \frac{n w_2\|f\|_\infty}{w_1}.$$

Hence, we get

$$A_2 \le C_2(w)\|f\|_\infty n, \tag{2.13}$$

where

$$C_2(w) = \frac{w_2^2}{w_1^2} + 1 + 2\frac{w_2}{w_1}.$$

Evaluation of the term A_3

It is easy to check that

$$
\begin{aligned}
\mathbb{E} & \left(g^2(x, Y^2)\right) = \\
= & \sum_{k\in\mathbb{Z}} \frac{\mu_w^2}{w^2(x)} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right)^2 \mathbb{E}\left(\left(p_{k,D}(Y) - \frac{\alpha_{k,D} w(Y)}{\mu_w}\right)^2 \frac{\mu_w^2}{w^2(Y)}\right) \\
+ & \sum_{k\in\mathbb{Z}} \sum_{\substack{k\in\mathbb{Z} \\ k' \neq k\in\mathbb{Z}}} \frac{\mu_w^2}{w^2(x)} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right) \left(p_{k',D}(x) - \frac{\alpha_{k',D} w(x)}{\mu_w}\right) \\
\times & \ \mathbb{E}\left(\frac{\mu_w^2}{w^2(Y)} \left(p_{k,D}(Y) - \frac{\alpha_{k,D} w(Y)}{\mu_w}\right) \left(p_{k',D}(Y) - \frac{\alpha_{k',D} w(Y)}{\mu_w}\right)\right) \\
= & \sum_{k\in\mathbb{Z}} \frac{\mu_w^2}{w^2(x)} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right)^2 \left(\int \left(p_{k,D} - \frac{\alpha_{k,D} w}{\mu_w}\right)^2 \frac{\mu_w^2}{w^2} \frac{fw}{\mu_w}\right) \\
+ & \sum_{k\in\mathbb{Z}} \sum_{\substack{k\in\mathbb{Z} \\ k' \neq k\in\mathbb{Z}}} \frac{\mu_w^2}{w^2(x)} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right) \left(p_{k',D}(x) - \frac{\alpha_{k',D} w(x)}{\mu_w}\right) \\
\times & \left(\int \frac{f\mu_w}{w} \left(p_{k,D} - \frac{\alpha_{k,D} w}{\mu_w}\right) \left(p_{k',D} - \frac{\alpha_{k',D} w}{\mu_w}\right)\right) \\
= & \sum_{k\in\mathbb{Z}} \frac{\mu_w^2}{w^2(x)} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right)^2 \left(\int \frac{p_{k,D}^2 f \mu_w}{w}\right) \\
- & \left(\sum_{k\in\mathbb{Z}} \frac{\mu_w}{w(x)} \alpha_{k,D} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w}\right)\right)^2 \\
\leq & \frac{w_2^3}{w_1^3} \sum_{k\in\mathbb{Z}} p_{k,D}^2(x) \left(\int_{\frac{k}{D}}^{\frac{k+1}{D}} Df\right) + \sum_{k\in\mathbb{Z}} \frac{\alpha_{k,D}^2 w_2}{w_1} \left(\int_{\frac{k}{D}}^{\frac{k+1}{D}} Df\right) \\
\leq & \frac{w_2^3}{w_1^3} \|f\|_\infty \left(\sum_{k\in\mathbb{Z}} p_{k,D}^2(x)\right) + \frac{w_2 \|f\|_\infty}{w_1} \left(\sum_{k\in\mathbb{Z}} \alpha_{k,D}^2\right).
\end{aligned}
$$

Hence, we get

$$
A_3 \leq C_3(w) \sqrt{n \|f\|_\infty D}, \tag{2.14}
$$

where

$$
C_3(w) = \sqrt{\frac{w_2^3}{w_1^3} + \frac{w_2}{w_1}}.
$$

Evaluation of the term A_4

For all x and y

$$\begin{aligned} |g(x,y)| &\leq \sum_{k\in\mathbb{Z}} \frac{p_{k,D}(x)p_{k,D}(y)\mu_w^2}{w(x)w(y)} + \sum_{k\in\mathbb{Z}} \alpha_{k,D}^2 + \sum_{k\in\mathbb{Z}} \frac{\alpha_{k,D}p_{k,D}(x)\mu_w}{w(x)} \\ &+ \sum_{k\in\mathbb{Z}} \frac{\alpha_{k,D}p_{k,D}(y)\mu_w}{w(y)} \leq \frac{2w_2^2 D}{w_1^2} + D + \frac{4w_2 D}{w_1}\int f. \end{aligned}$$

Hence, we can give an upper bound for A_4

$$A_4 \leq C_4(w)D, \tag{2.15}$$

where

$$C_4(w) = 1 + \frac{w_2^2}{w_1^2} + 2\frac{w_2}{w_1}.$$

Using inequalities (2.12)-(2.15), we can deduce from Theorem 3.4 of Houdré & Reynaud-Bouret (2003) that

$$\mathbb{P}\Big\{\big|U_n(H_D)\big| > \frac{1}{n-1}\Big[\kappa_1\sqrt{D\theta_D t} + \kappa_2\|f\|_\infty t + \frac{\kappa_3 D t^2}{n}\Big]\Big\} \leq 5.6\exp(-t), \tag{2.16}$$

where $\kappa_1 = C_1(\epsilon_0, w)$, $\kappa_2 = C_3(\epsilon_0, w) + C_2(\epsilon_0, w)$, $\kappa_3 = C_3(\epsilon_0, w) + C_4(\epsilon_0, w)$, $C_1(\epsilon_0, w) = 4(1+\epsilon_0)^{\frac{3}{2}}C_1(w)$, $C_2(\epsilon_0, w) = 2n(\epsilon_0)C_2(w)$, $C_3(\epsilon_0, w) = 2\beta(\epsilon_0)C_3(w)$ and $C_4(\epsilon_0, w) = 2\gamma(\epsilon_0)C_4(w)$, where ϵ_0 is a fixed positive number. This completes the proof Lemma 2.6.1. □

Proposition 2.6.1 gives a non-asymptotic risk bound for the pseudo-estimator $\tilde{\theta}$.

Proposition 2.6.1. *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by* (2.1)*, with weight function w being continuous and satisfying Assumption 1. Consider the class of functions satisfying $\|f\|_\infty \leq M$ with M known, and let $\theta = \int f^2$. For all $D \in \mathcal{D}$, where $\mathcal{D}$ is a subset of $\mathbb{N}$, let $\tilde{\theta}_D$ be defined by* (2.2)*. There exists some $\kappa > 0$ such that if we set for all $D \in \mathcal{D}$*

$$\text{pen}(D) = \frac{\kappa}{n}[\sqrt{MD\log(D+1)} + M\log(D+1) + \frac{D\log^2(D+1)}{n}], \tag{2.17}$$

then $\tilde{\theta}$, given by (2.3), satisfies the following inequality for all $n \geq 2$

$$\mathbb{E}\left\{\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}\left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta\right)\right\}^2 \leq C(w)\inf_{D\in\mathcal{D}_n}\big[\|f_D - f\|_2^4 + \text{pen}^2(D)\big].$$

Proof Proposition 2.6.1

Let

$$P_n(h_D) = \frac{1}{n}\sum_{i=1}^{n} \frac{h_D(Y_i)\mu_w}{w(Y_i)} - \int h_D f = \frac{1}{n}\sum_{i=1}^{n} \mu_w \frac{2(f_D(Y_i) - f(Y_i))}{w(Y_i)} - \int 2(f_D - f)f$$

and $H_D(x,y)$ and $U_n(H_D)$ be defined by (2.10) and (2.11) respectively.

The following decomposition holds

$$U_n(H_D) + P_n(h_D) - \int (f - f_D)^2 = \tilde{\theta}_D - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{\mu_w f(Y_i)}{w(Y_i)} - \theta).$$

Let

$$V_D = U_n(H_D) + P_n(h_D) - \int (f - f_D)^2 - \text{pen}(D).$$

Hence, it is easy to check that

$$\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{\mu_w f(Y_i)}{w(Y_i)} - \theta) = \sup_{D\in\mathcal{D}}(V_D).$$

$\mathcal{D}$ is a subset of $\mathbb{N}$ and pen(D) is given by (2.17). The following inequalities hold

$$|\sup_{D\in\mathcal{D}}(V_D)| \leq [\sup_{D\in\mathcal{D}}(V_D)_+] \bigvee \inf_{D\in\mathcal{D}}(V_D)_-,$$

$$\mathbb{E}\left(\sup_{D\in\mathcal{D}}(V_D)\right)^2 \leq \sum_{D\in\mathcal{D}} \mathbb{E}\big((V_D)_+^2\big) + \mathbb{E}\left(\inf_{D\in\mathcal{D}}(V_D)_-^2\right).$$

Control of $U_n(H_D)$

Let

$$u_D(t) = \frac{1}{n-1}\left[\kappa_1\sqrt{D\theta_D t} + \kappa_2\|f\|_\infty t + \frac{\kappa_3 D t^2}{n}\right].$$

Using Lemma 2.6.1 and the following inequality

$$u_D(\frac{t_1 + t_2}{\sqrt{2}}) \leq u_D(t_1) + u_D(t_2),$$

one obtains that

$$\mathbb{P}\Big\{|U_n(H_D)| > u_D(\sqrt{2}y_D) + u_D(\sqrt{2}t)\Big\} \leq 5.6\exp(-t - y_D). \tag{2.18}$$

Control of $P_n(h_D) - \|f - f_D\|_2^2$

According to Lemma 8 of Birgé and Massart (1998), if $U_1, U_2, \ldots, U_n$ are n independent random variables such that $|U_i| \leq b$ and $\mathbb{E}(U_i^2) \leq \delta^2$ for $i = 1, \ldots, n$, then the following inequality holds

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(U_i - \mathbb{E}(U_i)) > \delta\sqrt{\frac{2t}{n}} + \frac{bt}{3n}\right\} \leq \exp\left(-t\right). \tag{2.19}$$

If

$$U_i = \frac{2\mu_w[f_D(Y_i) - f(Y_i)]}{w(Y_i)},$$

then it is easy to check that

$$|U_i| \leq \frac{4w_2\|f\|_\infty}{w_1}$$

and

$$\mathbb{E}(U_i^2) \leq \frac{4w_2\|f_D - f\|_2^2\|f\|_\infty}{w_1}.$$

Therefore, using (2.19), one gets that

$$\mathbb{P}\left\{P_n(h_D) > \frac{\sqrt{8tw_2\|f\|_\infty}}{\sqrt{nw_1}}\|f_D - f\|_2 + \frac{4w_2\|f\|_\infty t}{3w_1 n}\right\} \leq \exp(-t), \tag{2.20}$$

and using the elementary inequality

$$2\sqrt{\frac{2tw_2\|f\|_\infty}{w_1 n}}\|f_D - f\|_2 \leq \|f_D - f\|_2^2 + 2\frac{tw_2\|f\|_\infty}{w_1 n}, \tag{2.21}$$

one obtains

$$\mathbb{P}\left\{P_n(h_D) - \|f - f_D\|_2^2 > \frac{10}{3}(t)\frac{w_2\|f\|_\infty}{w_1 n}\right\} \leq \exp(-t) \tag{2.22}$$

and

$$\mathbb{P}\left\{P_n(h_D) - \|f - f_D\|_2^2 > \frac{10}{3}(t + y_D)\frac{w_2\|f\|_\infty}{w_1 n}\right\} \leq \exp\left(-t - y_D\right). \tag{2.23}$$

Let $x_D = \log\left(D+1\right)$ and $\kappa = \max(C_1, C_2, C_3)$, where $C_1 = 4\sqrt{\sqrt{2}}\kappa_1$, $C_2 = 8\sqrt{2}\kappa_2 + \frac{10w_2}{3w_1}$ and $C_3 = 64\kappa_3$. The following inequalities hold

$$\begin{aligned} u_D(\sqrt{2}y_D) + \frac{10}{3}\frac{My_D w_2}{nw_1} &\leq \frac{1}{n}\left(C_1\sqrt{Dx_D\theta_D} + C_2 M x_D + \frac{C_3 D x_D^2}{n}\right) \\ &\leq \operatorname{pen}(D). \end{aligned} \tag{2.24}$$

Using (2.18), (2.23) and (2.24), one arrives at

$$\mathbb{P}\left\{V_D > u_D(\sqrt{2}t) + \frac{10Mtw_2}{3nw_1}\right\} \leq \mathbb{P}\left\{U_n(H_D) > u_D(\sqrt{2}t) + u_D(\sqrt{2}y_D)\right\} +$$

$$+\mathbb{P}\left\{P_n(h_D) - \|f - f_D\|_2^2 > \frac{10Mw_2(t + y_D)}{3nw_1}\right\} \leq 6.6\exp(-t - y_D). \tag{2.25}$$

The following identity

$$\mathbb{E}\left[(V_D)_+^2\right] = 2\int_0^\infty t\mathbb{P}(V_D > t)dt,$$

(2.25) and

$$u_D(\sqrt{2}t_0) + \frac{10Mt_0w_2}{3nw_1} \leq t,$$

where

$$t_0 = \inf\left\{\frac{t^2n^2}{36D\theta_D\kappa_1^2}, \frac{tn}{3M(2\kappa_2 + \frac{10w_2}{3w_1})}, n\sqrt{\frac{t}{6\kappa_3 D}}\right\},$$

lead to the following inequality

$$\mathbb{E}\left[(V_D)_+^2)\right] \leq 6.6C(w)\left\{\frac{DM}{n^2} + \frac{M^2}{n^2} + \frac{D^2}{n^4}\right\}\exp\left(-y_D\right).$$

Hence, using the inequality

$$\sum_{D\in\mathcal{D}} D^2\exp\left(-y_D\right) \leq \sum_{D\geq 1}\frac{1}{D^2},$$

we obtain

$$\sum_{D\in\mathcal{D}} \mathbb{E}(V_D)_+^2 \leq C(w)\left\{\frac{M}{n^2} + \frac{M^2}{n^2} + \frac{1}{n^4}\right\}.$$

Now, we give an upper bound for $\mathbb{E}\left[(V_D)_-^2)\right]$. The following inequality holds

$$\mathbb{E}\left[(V_D)_-^2)\right] \leq 4\mathbb{E}(U_n^2(H_D)) + 4\mathbb{E}(P_n^2(h_D)) + 4\|f - f_D\|_2^4 + 4(\mathrm{pen}(D))^2. \tag{2.26}$$

Using inequality (2.16) and $u_D(f_0) \leq t$, where $f_0 = \inf\left\{\frac{t^2n^2}{36D\theta_D\kappa_1^2}, \frac{tn}{6M\kappa_2}, n\sqrt{\frac{t}{6\kappa_3 D}}\right\}$, we arrive, using similar steps as before, at

$$\mathbb{E}(U_n^2(H_D)) \leq C(w)\{\frac{DM}{n^2} + \frac{M^2}{n^2} + \frac{D^2}{n^4}\}. \tag{2.27}$$

It is also easy to check that

$$\mathrm{pen}^2(D) \geq \frac{DM}{n^2} + \frac{M^2}{n^2} + \frac{D^2}{n^4}.$$

Hence, the following inequality holds

$$\mathbb{E}(U_n^2(H_D)) \leq C(w)\mathrm{pen}^2(D) \quad \forall D \in \mathcal{D}.$$

Now, we obtain an upper bound for $\mathbb{E}(P_n^2(h_D))$. Using (2.20), the inequality $u(y_0) \leq y$, where

$$y_0 = \inf\left\{\frac{y^2 w_1 n}{32 M w_2 \|f - f_D\|_2^2}, \frac{w_1 y n}{8 w_2 M}\right\}$$

and

$$u(y) = 2\sqrt{\frac{2yMw_2}{nw_1}}\|f_D - f\|_2 + M\frac{4yw_2}{3w_1 n},$$

one obtains

$$\mathbb{E}(P_n^2(h_D)) \leq C(w)\left\{\|f_D - f\|_2^4 + \frac{M^2}{n^2}\right\}. \tag{2.28}$$

Using (2.26)-(2.28) we have

$$\mathbb{E}\left[(V_D^2)_-\right] \leq C(w)[\|f_D - f\|_2^4 + \mathrm{pen}^2(D)].$$

Collecting the above inequalities together we arrive at

$$\mathbb{E}\left(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)\right)^2 \leq C(w)\inf_{D\in\mathcal{D}}\left[\|f_D - f\|_2^4 + \mathrm{pen}^2(D)\right],$$

thus concluding the proof of Proposition 2.6.1.

In order to prove Propositon 2.6.1, we used an exponential inequality of order 2 obtained by Houdré-Bouret & Reynaud (2003).

Proposition 2.6.2. *Let $Y_1, Y_2, \ldots, Y_n$ be defined as in Proposition 2.6.1. Consider the class of functions satisfying $\|f\|_\infty \leq M$ with M unknown. Let $\theta = \int f^2$ and $\mathcal{D}_n$ be defined by (2.9). There exists some constant $\kappa_0 > 0$ such that if $\mathrm{pen}(D)$ is given by (2.4) for all $D \in \mathcal{D}_n$, then there exists some $n^* := n^*(\alpha, R, M, w)$ such that $\tilde{\theta}$, given by* (2.3), *satisfies the following inequality for all $n \geq n^*$ and for all $\kappa \geq \kappa_0$*

$$\mathbb{E}\left\{\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}\left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta\right)\right\}^2 \leq$$

$$\leq C(w)\inf_{D\in\mathcal{D}_n}\left[\|f_D - f\|_2^4 + \frac{D(M+1)\log(D+1)}{n^2}\right] + \frac{C(M,w)}{n^2}.$$

Proof of Proposition 2.6.2

Let $A = \{\omega \in \omega : \tilde{\theta}_D + \frac{1}{2} \geq \theta_D \forall D \in \mathcal{D}_n\}$. We obtain an upper bound for

$$\mathbb{E}\Big[\Big(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)\Big)\mathbb{I}_A\Big]^2.$$

The following inequality holds

$$\mathbb{E}\left[|\sup_{D \in \mathcal{D}_n} V_D|^2 \mathbb{I}_A\right] \leq \sum_{D \in \mathcal{D}_n} \mathbb{E}\left[(V_D)_+^2 \mathbb{I}_A\right] + \inf_{D \in \mathcal{D}_n} \mathbb{E}\big[(V_D)_-^2\big].$$

Let

$$C_0(M) = \inf\{D \in \mathbb{N} : \frac{D}{\log(D+1)} \geq M^2\}.$$

If

$$\text{pen}(D) \geq u_D(\sqrt{2}y_D) + \frac{10My_Dw_2}{3nw_1} \tag{2.29}$$

holds on A, then the following inequality holds

$$\mathbb{P}\left\{\Big(V_D > u_D(\sqrt{2}t) + \frac{10}{3}\frac{Mtw_2}{nw_1}\Big)\bigcap A\right\} \leq 6.6\exp(-y_D - t).$$

Using the previous inequality and the identity

$$\mathbb{E}\big[(V_D)_+^2\mathbb{I}_A\big] = 2\int_0^\infty t\mathbb{P}\Big\{(V_D > t)\bigcap A\Big\}dt,$$

one obtains

$$\mathbb{E}\big[(V_D)_+^2\mathbb{I}_A\big] \leq C(M,w)\exp(-y_D)\frac{D^2}{n^2}.$$

Additionally, it is easy to check that the following inequality holds on A

$$\begin{aligned}
\text{pen}(D) &\geq \frac{\kappa}{n}\sqrt{(\theta_D + \frac{1}{2})D\log(D+1)} \geq \frac{\kappa}{\sqrt{2}}\frac{\sqrt{\theta_D D x_D}}{n} + \frac{\kappa}{2}\frac{\sqrt{Dx_D}}{n} \\
&\geq \frac{\kappa}{\sqrt{2}}\frac{\sqrt{\theta_D D x_D}}{n} + \frac{\kappa M x_D}{4n} + \frac{\kappa}{4}\frac{\sqrt{Dx_D}}{n} \quad \forall D \in \mathcal{D}_n \quad \text{such that} \quad D \geq C_0(M).
\end{aligned}$$

Now, if $\kappa_0 = \max(\sqrt{2}C_1, 4C_2, 4C_1^{'}C_3)$, where $C_1^{'}$ is a positive constant satisfying $\frac{Dx_D^2}{n\sqrt{Dx_D}} \leq C_1^{'}$, then

$$\text{pen}(D) \geq C_1\frac{\sqrt{\theta_D D x_D}}{n} + C_2\frac{Mx_D}{n} + C_3\frac{Dx_D^2}{n^2} \geq u_D(\sqrt{2}y_D) + \frac{10My_Dw_2}{3nw_1}$$

$\forall D \in \mathcal{D}_n$ and $\forall \kappa \geq \kappa_0$. Therefore, (2.29) holds on A $\forall D \in \mathcal{D}_n$ such that $D \geq C_0(M)$. On the other hand, if $D \leq C_0(M)$, then

$$\mathbb{E}\big[(V_D)_+^2\mathbb{I}_A\big] \leq 2\mathbb{E}(U_n^2(H_D)) + 2\mathbb{E}((P_n(h_D) - \|f - f_D\|)_+^2).$$

Using (2.22) and the well-known identity $\mathbb{E}(X^2) = \int_0^\infty t\mathbb{P}(|X| > t)dt$ we arrive at

$$\mathbb{E}\big((P_n(h_D) - \|f - f_D\|)_+^2\big) \leq \frac{C(M,w)}{n^2}.$$

The previous inequality and (2.27) imply that

$$\mathbb{E}((V_D)_+^2 \mathbb{I}_A) \leq \frac{C(M,w)}{n^2} \quad \forall D \in \mathcal{D}_n \quad \text{such that} \quad D \leq C_0(M).$$

Since $|\mathcal{D}_1| = |\{D | D \in \mathcal{D}_n, D \leq C_0(M)\}| \leq C_0(M)$, the following inequality holds

$$\sum_{D \in \mathcal{D}_n} \mathbb{E}((V_D)_+^2 \mathbb{I}_A) \leq \sum_{D \in \mathcal{D}_n, D \leq C_0(M)} \mathbb{E}((V_D)_+^2 \mathbb{I}_A) + \sum_{D \in \mathcal{D}_n, D \geq C_0(M)} \mathbb{E}((V_D)_+^2 \mathbb{I}_A) \leq \frac{C(M,w)}{n^2}.$$

Now, we obtain an upper bound for $\mathbb{E}\left[(V_D)_-^2\right]$. Note that

$$\mathbb{E}(V_D)_-^2) \leq 4\mathbb{E}(P_n^2(h_D)) + 4\mathbb{E}(U_n^2(H_D)) + 4\mathbb{E}(\text{pen}^2(D)) + 4\|f - f_D\|_2^4.$$

For all $D \in \mathcal{D}_n$, it is easy to see that

$$\mathbb{E}(\text{pen}^2(D)) = \frac{\kappa^2(\theta_D + 1)Dx_D}{n^2} \leq \frac{C(w)(M+1)Dx_D}{n^2}.$$

Using (2.27) and (2.28) to control $\mathbb{E}(P_n^2(h_D))$ and $\mathbb{E}(U_n^2(H_D))$, we arrive at

$$\mathbb{E}((V_D)_-^2) \leq C(w)\left[\|f_D - f\|_2^4 + \frac{M^2}{n^2} + \frac{(M+1)Dx_D}{n^2}\right].$$

Now, it remains to find an upper bound for $\mathbb{E}\left[\left(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^n\left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta\right)\right)^2 \mathbb{I}_{A^c}\right]$. We now obtain an upper bound for $P(A^c)$. We can easily check that the following inequality holds

$$\mathbb{P}(A^c) = \sum_{D \in \mathcal{D}_n} \mathbb{P}(\tilde{\theta}_D + \frac{1}{2} \leq \theta_D) \leq \sum_{D \in \mathcal{D}_n} \mathbb{P}(|\tilde{\theta_D} - \theta_D| > \frac{1}{2}).$$

Moreover, the following decomposition holds

$$\tilde{\theta_D} - \theta_D = U_n(H_D) + P_n(2f_D).$$

Hence, we arrive at

$$\mathbb{P}(A^c) \leq \mathbb{P}\left(|U_n(H_D)| > \frac{1}{4}\right) + \mathbb{P}\left(|P_n(2f_D)| > \frac{1}{4}\right).$$

We now obtain an upper bound for $\mathbb{P}\left(|U_n(H_D)| > \frac{1}{4}\right)$. Using the inequality

$$u_D(\beta_0) \leq \frac{2}{n}\left[\kappa_1\sqrt{DM\beta_0} + \kappa_2 M\beta_0 + \kappa_3\frac{D\beta_0^2}{n}\right] \leq \frac{1}{4},$$

where

$$\beta_0(n, M, w) = \inf\left\{\frac{(\log n)^3}{(24\kappa_1)^2 M}, \frac{n}{24\kappa_2 M}, \frac{(\log n)^{\frac{3}{2}}}{\sqrt{24\kappa_3}}\right\},$$

one obtains

$$\mathbb{P}\left(|U_n(H_D)| > \frac{1}{4}\right) \leq 5.6 \exp(-\beta_0).$$

Since $\exp(-\beta_0) \leq \frac{C(M,w)}{n^8}$ $\forall n \geq n_{01}$, one arrives at

$$\mathbb{P}\left(|U_n(H_D)| > \frac{1}{4}\right) \leq \frac{C(M,w)}{n^8} \quad \forall n \geq n_{01}. \tag{2.30}$$

Now, we will give an upper bound for $\mathbb{P}\left(|P_n(2f_D)| > \frac{1}{4}\right)$. Let $U_i = \frac{2f_D(Y_i)\mu_w}{w(Y_i)}$ satisfying $|U_i| \leq \frac{2Mw_2}{w_1}$ and $\mathbb{E}(U_i^2) \leq \frac{16M^2 w_2}{w_1}$. Using Lemma 8 of Birgé & Massart (1998), we obtain

$$\mathbb{P}(|P_n(2f_D)| > 4M\sqrt{\frac{2w_1 y_0}{nw_1}} + \frac{2Mw_2 y_0}{3nw_1}) \leq 2\exp(-y_0), \tag{2.31}$$

where $y_0(n, M, w) = \inf\left\{\frac{nw_1}{2^{11}M^2 w_2}, \frac{3nw_1}{16Mw_2}\right\}$. It is easy to see that

$$4M\sqrt{\frac{2w_2 y_0}{nw_1}} + \frac{2Mw_2 y_0}{3nw_1} \leq \frac{1}{4}. \tag{2.32}$$

The following inequality holds $\forall n \geq n_{02}$

$$\exp(-y_0(n, M, w)) \leq \frac{C(M,w)}{n^8}. \tag{2.33}$$

Inequalities (2.31)-(2.33) lead to

$$\mathbb{P}(|P_n(2f_D)| > \frac{1}{4}) \leq \frac{C(M,w)}{n^8}, \forall n \geq n_{02}. \tag{2.34}$$

By using (2.30), (2.34) and the fact that $|\mathcal{D}_n| \leq \frac{n^2}{\log^3 n}$, one obtains that

$$\mathbb{P}(A^c) \leq |\mathcal{D}_n|\frac{C(M,w)}{n^8} \leq \frac{C(M,w)}{n^6}, \forall n \geq n' = \max(n_{01}, n_{02}).$$

It is easy to see that the following inequalities hold

$$0 \leq \tilde{\theta}_D \leq \frac{2Dw_2^2}{w_1^2}, \quad \mathrm{pen}(D) \leq C(w)n, \forall n \geq 3,$$

$$|\tilde{\theta}| \leq C(w)n^2, \forall n \geq 3, \quad \theta = \int f^2 \leq \|f\|_\infty \int f \leq M,$$

$$|\frac{2}{n}\sum_{i=1}^n(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)| \leq 2M(1 + \frac{w_2}{w_1}), \quad \left(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^n(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)\right)^2 \leq C(M,w)n^4.$$

Therefore, we arrive at

$$\mathbb{E}\Bigg[\Bigg(\tilde{\theta}-\theta-\frac{2}{n}\sum_{i=1}^{n}(\frac{f(Y_i)\mu_w}{w(Y_i)}-\theta)\Bigg)^2\mathbb{I}_{A^c}\Bigg]\leq C(M,w)n^4\mathbb{P}(A^c)\leq\frac{C(M,w)}{n^2}$$

$\forall n\geq n^*=\max(n',3)$, thus concluding the proof of 2.6.2.

Proposition 2.6.3. *Let $Y_1,Y_2,\dots,Y_n$ be defined as in Proposition 2.6.1. Consider the smooth class of p.d.f's $\mathcal{F}(\alpha,R,M)$ defined by (2.8). Let $\tilde{\theta}$ be defined as in Proposition 2.6.2. For any $\alpha>0$, $R>0$ and $M>0$, there exists some $\kappa_0>0$ and some integer $n^{**}:=n^{**}(\alpha,R,M)$ such that the following inequality holds for all $n\geq n^{**}$ and all $\kappa\geq\kappa_0$*

$$\sup_{f\in\mathcal{F}(\alpha,R,M)}\mathbb{E}\Bigg\{\tilde{\theta}-\theta-\frac{2}{n}\sum_{i=1}^{n}\Bigg(\frac{f(Y_i)\mu_w}{w(Y_i)}-\theta\Bigg)\Bigg\}^2\leq C(\alpha,w)(RM^\alpha)^{\frac{4}{1+4\alpha}}\Bigg(\frac{\sqrt{\log(nR^2)}}{n}\Bigg)^{\frac{8\alpha}{1+4\alpha}}.$$

Furthermore, for all $\alpha>0$, $R>0$ and $M>0$, there exists some integer $n_1:=n_1(\alpha,R,M)$ such that the following inequality holds for all $n\geq n_1$

$$\sup_{f\in\mathcal{F}(\alpha,R,M)}\mathbb{E}(\tilde{\theta}-\theta)^2\leq C(\alpha,w)(RM^\alpha)^{\frac{4}{1+4\alpha}}\Bigg(\frac{\sqrt{\log(nR^2)}}{n}\Bigg)^{\frac{8\alpha}{1+4\alpha}}+C(w)\frac{M^2}{n},$$

where $C(w)$ is some constant depending on w_1 and w_2.

Proof of Proposition 2.6.3

Let

$$J_n=\lfloor\log_2\Bigg(\frac{n^2R^4}{(M+1)\log(nR^2)}\Bigg)^{\frac{1}{1+4\alpha}}\rfloor+1$$

and

$$\mathcal{D}_n=\Bigg\{D\in\mathbb{N}:D\leq\frac{n^2}{\log^3 n}\Bigg\}.$$

We will show that

$$\exists\ n_1\in\mathbb{N}:2^{J_n}\in\mathcal{D}_n\quad\text{for all } n\geq n_1.$$

Since, obviously,

$$\exists\ n_{11}\in\mathbb{N}:2^{J_n}\geq 1\quad\text{for all } n\geq n_{11},$$

it only remains to show that

$$\exists\ n_{12}\in\mathbb{N}:2^{J_n}\leq\frac{n^2}{\log^3 n}\quad\text{for all } n\geq n_{12}.$$

The following inequality holds

$$2^{J_n}\leq\frac{C(\alpha,R,M)n^{\frac{2}{1+4\alpha}}}{(\log n)^{\frac{1}{1+4\alpha}}}\quad\text{for all } n\geq n_{12}.$$

It is easy to check that for all $\alpha > 0$

$$\frac{C(\alpha, R, M) n^{\frac{2}{1+4\alpha}}}{(\log n)^{\frac{1}{1+4\alpha}}} \leq \frac{n^2}{\log^3 n} \quad \text{for all} \quad n \geq n_{12}.$$

From Proposition 2.6.2, one obtains that

$$\mathbb{E}(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}[\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta])^2 \leq$$

$$\leq \frac{C(M, w)}{n^2} + C(w) \inf_{D \in \mathcal{D}_n}\{\|f_D - f\|_2^4 + \frac{(M+1)Dx_D}{n^2}\} \tag{2.35}$$

$\forall n \geq n^*$ and $\forall f : \|f\|_\infty \leq M$.

Additionally,

$$\|f - f_{2^J}\|_2^2 = \sum_{j \geq J}\sum_{k \in \mathbb{Z}} {\beta_{j,k}}^2 \leq \sum_{j \geq J} R^2 2^{-2j\alpha} = C(\alpha) R^2 2^{-2J\alpha}, \tag{2.36}$$

$\forall f : \beta(f) \in \mathcal{B}_{\alpha,2,\infty}(R)$.

Using (2.35), one arrives at

$$\begin{aligned}\sup_{f : f \in \mathcal{F}(\alpha,R,M)} \mathbb{E}\left[\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)\right]^2 \leq& \frac{C(M, w)}{n^2} + C(w)\|f_{2^{J_n}} - f\|_2^4 \\ +& \frac{C(w)(M+1)2^{J_n}\log(2^{J_n}+1)}{n^2},\end{aligned}$$

$\forall n \geq n^{**}$, where $n^{**} = \max\{n^*, n_{12}\}$.

Combining the above inequality and (2.36), it is easy to see that the following inequality holds

$$\sup_{f : f \in \mathcal{F}(\alpha,R,M)} \mathbb{E}\left[\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta)\right]^2 \leq$$

$$\leq C(\alpha, w)\{R(M+1)^\alpha\}^{\frac{4}{1+4\alpha}}\{\frac{\sqrt{\log(nR^2)}}{n^2}\}^{\frac{8\alpha}{1+4\alpha}}, \forall n \geq n^{**}. \tag{2.37}$$

Now, we will give an upper bound for $\mathbb{E}(\tilde{\theta} - \theta)^2$.

$$\mathbb{E}(\tilde{\theta} - \theta)^2 \leq 2\mathbb{E}(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^{n}[\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta])^2 + 2\mathbb{E}(\frac{2}{n}\sum_{i=1}^{n}[\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta])^2 \tag{2.38}$$

and

$$\mathbb{E}(\frac{2}{n}\sum_{i=1}^{n}[\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta])^2 \leq \frac{C(w)M^2}{n}. \tag{2.39}$$

Inequalities (2.37)-(2.39) lead to

$$\sup_{f: f\in\mathcal{F}(\alpha,R,M)} \mathbb{E}(\tilde{\theta}-\theta)^2 \leq \frac{C(w)M^2}{n} + C(\alpha,w)\{R(M+1)^{\alpha}\}^{\frac{4}{1+4\alpha}}\{\frac{\sqrt{\log(nR^2)}}{n^2}\}^{\frac{8\alpha}{1+4\alpha}},$$

$\forall n \geq n_1 = \max\{n_{11}, n_{12}\}$, which completes the proof of proposition 2.6.3.

Proposition 2.6.4 shows that $\mathbb{E}\Big(\sup_{D\in\mathcal{D}_n}(\hat{\theta}_D - 2\theta_D)\Big)^4$ is uniformly bounded for all $f \in \mathcal{F}(\alpha, R, M)$.

Proposition 2.6.4. *Let $Y_1, Y_2, \ldots, Y_n$ and the smooth class of p.d.f.'s $\mathcal{F}(\alpha, R, M)$ be defined as in Proposition 2.6.3. $\tilde{\theta}_D$ is given by (2.2). Then, the following inequality holds*

$$\mathbb{E}\Big(\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)\Big)^4 \leq C(M, w, R, \alpha).$$

Proof of Proposition 2.6.4

It is easy to see that the following inequality holds

$$\begin{aligned}
\mathbb{E}\Big(\big|\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)\big|^4\Big) &= \mathbb{E}\Big\{\big|\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)\big|^4\mathbb{I}_F\Big\} + \mathbb{E}\Big\{\big|\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)\big|^4\mathbb{I}_{F^c}\Big\} \\
&\leq c\sum_{D\in\mathcal{D}_n}\mathbb{E}\big((V_{D2})_+^4\big) + c\inf_{D\in\mathcal{D}_n}\mathbb{E}\big((V_{D2})_-^4\big) + \inf_{D\in\mathcal{D}_n}\mathbb{E}\big((V_{D1})_-^4\big) \\
&+ c\Big\{\sup_{D\in\mathcal{D}_n}[u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3w_1n}]\Big\}^4,
\end{aligned}$$

where

$$\begin{aligned}
F &= \{\tilde{\theta}_D - 2\theta_D \geq 0 \quad \text{for some} \quad D \in D_n\}, \quad V_{D1} = \tilde{\theta}_D - 2\theta_D, \\
V_{D2} &= \tilde{\theta}_D - 2\theta_D - u_D(\sqrt{2}y_D) - \frac{8My_Dw_2}{3w_1n} \quad \text{and} \quad y_D = 4\log(D+1).
\end{aligned}$$

Additionally, using Lemma 8 of Birgé & Massart (1998) we have

$$\mathbb{P}\Big(P_n(2f_D) - \|f_D\|^2 > \frac{8Mw_2(t+y_D)}{3w_1n}\Big) \leq \exp(-t-y_D),$$

which together with inequality (2.18) leads to

$$\mathbb{P}\Big(V_{D2} > u_D(\sqrt{2}t) + \frac{8Mw_2t}{3w_1n}\Big) \leq 6.6\exp(-t-y_D). \tag{2.40}$$

Combining the identity

$$\mathbb{E}(X^4) = \int_0^{\infty} t^3\mathbb{P}(|X| > t)dt \tag{2.41}$$

and (2.40), we get

$$\mathbb{E}\{(V_{D2})_+^4\} \leq C(M,w)\exp(-y_D)[\frac{D^4}{n^8} + \frac{D^2}{n^4} + \frac{1}{n^4}].$$

Since $|D_n| \leq \frac{n^2}{\log^3 n}$, we arrive at

$$\sum_{D \in D_n} \mathbb{E}\{(V_{D2})_+^4\} \leq C(M,w). \tag{2.42}$$

Now,

$$\begin{aligned}
\mathbb{E}\left((V_{D1})_-^4\right) &\leq c\mathbb{E}\left(U_n^4(H_D)\right) + c\mathbb{E}\left(P_n^4(2f_D)\right) + c\|f_D\|_2^8, \\
\mathbb{E}\left((V_{D2})_-^4\right) &\leq c\mathbb{E}\left(U_n^4(H_D)\right) + c\mathbb{E}\left(P_n^4(2f_D)\right) + c\|f_D\|_2^8 \\
&+ c(u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3nw_1})^4.
\end{aligned}$$

Using inequalities (2.16) and (2.31) and the identity (2.41), we get

$$\begin{aligned}
&\mathbb{E}\left(U_n^4(H_D)\right) \leq C(M,w)[\frac{D^2}{n^4} + \frac{D^4}{n^8} + \frac{1}{n^4}], \quad \mathbb{E}\left(P_n^4(2f_D)\right) \leq \frac{C(M,w)}{n^2}, \\
&\|f_D\|_2^8 \leq M^8, \quad \text{and} \quad \left\{\sup_{D \in \mathcal{D}_n} [u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3w_1n}]\right\}^4 \leq C(M,w),
\end{aligned}$$

which lead to

$$\inf_{D \in \mathcal{D}_n} \mathbb{E}\big((V_{D1})_-^4\big) + c \inf_{D \in \mathcal{D}_n} \mathbb{E}\big((V_{D2})_-^4\big) \leq C(M,w). \tag{2.43}$$

Finally, inequalities (2.42)-(2.43) complete the proof of Proposition 2.6.4.

We now are in the position to prove Theorem 2.3.1. It is easy to check that

$$
\begin{aligned}
\mathbb{E}(\tilde{\theta}-\hat{\theta})^2 &= \mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)-\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big]^2\Big\} \\
&= \mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)-\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big]_+ \\
&- \Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)-\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big]_-\Big\}^2 \\
&\leq 2\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)-\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big]_+^2\Big\} \\
&+ 2\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)-\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big]_-^2\Big\} \\
&\leq 2\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)-\hat{\theta}_D+\mathrm{pen}_u(D)\big)\Big]^2\mathbb{I}(B)\Big\} \\
&+ 2\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)-\tilde{\theta}_D+\mathrm{pen}(D)\big)\Big]^2\mathbb{I}(B^c)\Big\} \\
&\leq 4\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\hat{\theta}_D\big)\Big]^2\Big\}+4\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\mathrm{pen}_u(D)-\mathrm{pen}(D)\big)\Big]^2\Big\} \\
&+ 4\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\tilde{\theta}_D\big)\Big]^2\Big\}+4\mathbb{E}\Big\{\Big[\sup_{D\in\mathcal{D}_n}\big(\mathrm{pen}(D)-\mathrm{pen}_u(D)\big)\Big]^2\Big\},
\end{aligned}
$$

where

$$
B=\Big\{\omega \mid \sup_{D\in\mathcal{D}_n}\big(\tilde{\theta}_D-\mathrm{pen}(D)\big)\geq\sup_{D\in\mathcal{D}_n}\big(\hat{\theta}_D-\mathrm{pen}_u(D)\big)\Big\}.
$$

Using the Cauchy-Schwartz inequality, it is easy to see that the following inequalities hold

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{D\in\mathcal{D}_n}(\hat{\theta}_D-\tilde{\theta}_D)\Big]^2 &= \mathbb{E}\Big\{\frac{(\hat{\mu}_w^2-\mu_w^2)^2}{\mu_w^4}\Big[\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D)\Big]^2\Big\} \\
&\leq C(w)\mathbb{E}\Big\{\big(\hat{\mu}_w-\mu_w\big)^2\Big[\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D-2\theta_D)\Big]^2\Big\} \\
&+ C(w)[\sup_{D\in\mathcal{D}_n}(2\theta_D)]^2\mathbb{E}\left(\hat{\mu}_w-\mu_w\right)^2 \\
&\leq C(w)\sqrt{\mathbb{E}(\hat{\mu}_w-\mu_w)^4}\sqrt{\mathbb{E}\Big\{\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D-2\theta_D)\Big\}^4} \\
&+ C(w,M)\mathbb{E}\left(\hat{\mu}_w-\mu_w\right)^2. \qquad (2.44)
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D-\hat{\theta}_D)\Big]^2 &\leq C(w)\sqrt{\mathbb{E}(\hat{\mu}_w-\mu_w)^4}\sqrt{\mathbb{E}\Big\{\sup_{D\in\mathcal{D}_n}(\tilde{\theta}_D-2\theta_D)\Big\}^4} \\
&+ C(w,M)\mathbb{E}\left(\hat{\mu}_w-\mu_w\right)^2. \qquad (2.45)
\end{aligned}
$$

Moreover, one can easily see that

$$\mathbb{E}\left[\sup_{D\in\mathcal{D}_n}\left(\mathrm{pe}n(D)-\mathrm{pe}n_u(D)\right)\right]^2 \le C\kappa^2\mathbb{E}\left[\sup_{D\in\mathcal{D}_n}\left(\tilde{\theta}_D-\hat{\theta}_D\right)\right]^2 \tag{2.46}$$

and

$$\mathbb{E}\left[\sup_{D\in\mathcal{D}_n}\left(\mathrm{pe}n_u(D)-\mathrm{pe}n(D)\right)\right]^2 \le C\kappa^2\mathbb{E}\left[\sup_{D\in\mathcal{D}_n}(\hat{\theta}_D-\tilde{\theta}_D)\right]^2. \tag{2.47}$$

It is also easy to check that

$$\mathbb{E}(\hat{\mu}_w-\mu_w)^2 \le \frac{C(w)}{n} \quad \text{and} \quad \mathbb{E}(\hat{\mu}_w-\mu_w)^4 \le \frac{C(w)}{n^2}, \tag{2.48}$$

(see, e.g., Efromovich, 2004b; Brunel, Comte & Guilloux, 2005). Inequalities (2.44)-(2.48), together with Proposition 2.6.4, lead to

$$\mathbb{E}(\hat{\theta}-\tilde{\theta})^2 \le \frac{C(w)}{n} \quad \text{for all } n \ge n_0 := \max(n^*, 3)$$

which, together with Proposition 2.6.3, completes the proof of Theorem 2.3.1. □

2.6.2 Proof of Theorem 2.5.1

Let g_ν be a sequence of p.d.f's such that $\|g_\nu - g_0\|_2 \longrightarrow 0$ as $n \longrightarrow \infty$, where $g_0 = \frac{wf_0}{\mu_0}$ and $\mu_0 = \int f_0 w$. Let $\mu_w = \int fw$. We are going to determine the Fréchet derivative of the functional $\theta = \int \frac{g^2\mu_w^2}{w^2}$ at a point g_0, where $g_0 = \frac{wf_0}{\mu_0}$ with f_0 belonging to the class of p.d.f's $\mathcal{H}$. It is easy to see that the following equalities hold

$$\theta(g_\nu) = \int \frac{\mu_0^2 g_\nu^2}{w^2} = \int \frac{\mu_0^2 g_0^2}{w^2} + \int \frac{\mu_0^2(g_\nu-g_0)^2}{w^2} + 2\int \frac{\mu_0^2 g_0(g_\nu-g_0)}{w^2} \tag{2.49}$$

and

$$\int \frac{\mu_0^2 g_0(g_\nu-g_0)}{w^2} = \int \frac{\mu_0^2 g_0 g_\nu}{w^2} - \theta(g_0) = \int g_\nu\left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right]. \tag{2.50}$$

Additionally, one observes that the following inequality holds

$$\int \frac{\mu_0^2(g_\nu-g_0)^2}{w^2} \quad \le \quad \frac{w_2^2}{w_1^2}\|g_\nu-g_0\|_2^2 = o(\|g_\nu-g_0\|_2). \tag{2.51}$$

Using (2.49)-(2.51), one obtains that

$$\begin{aligned}
\theta(g_\nu) \quad &= \quad \theta(g_0) + 2\int g_\nu\left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] + o(\|g_\nu-g_0\|_2) \\
&= \quad \theta(g_0) + 2\int (g_\nu-g_0)\left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] + o(\|g_\nu-g_0\|_2).
\end{aligned}$$

Therefore, the Fréchet derivative is given by $\theta'(g_0) = 2\big[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\big]$. In the sequel, $< \cdot, \cdot >$ denotes the scalar product in $L^2(\mathbb{R})$. Following Ibragimov & Khasminskii (1991), we consider the space orthogonal to the square root of the likelihood $s_0 = \sqrt{g_0}$, i.e.,

$$H = \left\{ k \in L^2(\mathbb{R}) : \int k s_0 = 0 \right\},$$

and the projection operator onto this space, i.e.,

$$P_H(t) = t - \left(\int t s_0 \right) s_0.$$

Since $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables, the family $\{P_{g_0}^n\}$ is locally asymptotically Gaussian at all points $g_0 = \frac{w f_0}{\mu_0}$ with f_0 belonging to $\mathcal{H}$, in the direction $H(g_0)$ with normalizing factor $A_n(g_0)$, where $A_n(t) = \big(\frac{1}{\sqrt{n}}\big)(\sqrt{g_0})t$ (see, e.g., Example 2.2 of Ibragimov & Khasminskii, 1991). Let $K_n = \sqrt{n}\theta'(g_0) A_n P_{H(g_0)}$, where $\theta'(g_0) = 2\big[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\big]$. Then

$$\begin{aligned} K_n(k) = K(k) &= \int 2 s_0 k \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] - 2 \int k s_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] \\ &= \int k \left\{ 2 s_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] - 2 s_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] \right\}. \end{aligned}$$

Therefore, $K_n(k) \longrightarrow K(k)$ weakly, where $K(k) =< h, k >$ and

$$h = 2 s_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] - 2 s_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right].$$

According to Theorem 4.1 of Ibragimov & Khasminskii (1991), for any estimator of $\theta(g_0)$, say T_n, and for any family of vicinities of g_0, say $\{V(g_0)\}$, we have

$$\inf_{\{V(g_0)\}} \liminf_{n \longrightarrow \infty} \sup_{g \in V(g_0)} n\, \mathbb{E}(T_n - \theta(g))^2 \geq \|h\|_2^2.$$

Hence, the information bound is given by

$$\begin{aligned} I_w(f_0) := \|h\|_2^2 &= 4 \int g_0 \left\{ \frac{\mu_0^2 g_0}{w^2} - \theta(g_0) - \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) \right] \right\}^2 \\ &= 4 \int \frac{f_0^3 \mu_0}{w} + 4\theta^2(g_0) - 8\theta(g_0) \int \frac{g_0^2 \mu_0^2}{w^2} \\ &= 4\,\mu_0 \int \frac{f_0^3}{w} - 4 \left(\int f_0^2 \right)^2, \end{aligned}$$

thus completing the proof of Theorem 2.5.1. $\square$

Chapter 3

Estimation of a signal using the maximum a posteriori method

In this chapter, the optimality results presented in Abramovich, Grinshtein & Pensky (2007), for estimating a high dimensional Gaussian mean vector, are generalized, providing adaptive conditions and a wider range of strong and weak $l_p(\eta)$ balls under which optimality of the maximum *a posteriori* (MAP) testimator is proved. The standard Gaussian white noise model is then considered and MAP testimation procedure is applied in a wavelet context, in order to construct an adaptive estimator of f which attains the optimal rate under this model. Using the boundary-modified coiflets of Johnstone & Silverman (2004a), it is also shown that discretization of the data does not affect the rate of convergence of the proposed MAP testimator. The optimality results are extended to the estimation of derivatives of f. Finally, a simulation study is conducted in order to illustrate the performance of the proposed estimator in practice.

3.1 Introduction

We consider the problem of estimating the unknown response function in the Gaussian white noise model, where one observes Gaussian processes $Y_n(t)$ as follows

$$dY_n(t) = f(t)dt + \frac{\sigma}{\sqrt{n}} dW(t), \quad t \in [0, 1]. \tag{3.1}$$

The noise parameter $\sigma > 0$ is assumed to be known, W is a standard Wiener process, and $f \in L^2[0, 1]$ is the unknown response function. Under some smoothness constraints on f, such a model is asymptotically equivalent (in Le Cam sense) to the standard nonparametric regression setting (see, e.g., Brown & Low, 1996).

If f possesses some smoothness properties, we can deal with a consistent estimation theory. We assume that f belongs to a Besov ball $B^s_{p,q}(M)$ of a radius $M > 0$, where $0 < p, q \leq \infty$ and $s > \max(0, \frac{1}{p} - \frac{1}{2})$. The latter restriction ensures that the corresponding Besov spaces are embedded in $L^2[0,1]$. The parameter s measures degree of smoothness while p and q specify the type of norm used to measure the smoothness. Besov classes contain various traditional smoothness spaces such as Hölder and Sobolev spaces as special cases. However, they also include different types of spatially inhomogeneous functions (see, e.g., Meyer, 1992, Chapter 6).

Wavelet series constitute unconditional bases for Besov spaces that has made various wavelet-based estimation procedures to be widely used for estimating the unknown response $f \in B^s_{p,q}(M)$ in the model (3.1). The standard wavelet approach to the estimation of f is based on finding the empirical wavelet coefficients of the data and to further denoising them, usually by some type of a thresholding rule. Transforming back to the function space yields then the resulting estimate. The main statistical challenge in such an approach is a proper choice of a thresholding rule. A series of various wavelet thresholds originated by different ideas has been proposed in the literature during the last decade, e.g., the universal threshold (see Donoho & Johnstone, 1994a), SURE threshold (see Donoho & Johnstone, 1995), FDR threshold (see Abramovich & Benjamini, 1996), cross-validation threshold (see Nason, 1996), Bayesian threshold (see Abramovich, Sapatinas & Silverman, 1998), empirical Bayes threshold (see Johnstone & Silverman, 2005).

Abramovich & Benjamini (1996) demonstrated that thresholding can be viewed as a multiple hypothesis testing procedure, where one first simultaneously tests the wavelet coefficients of the unknown response function for significance. The coefficients concluded to be significant are then estimated by the corresponding empirical wavelet coefficients of the data, while others are discarded. Such a *testimation* procedure evidently mimics a hard thresholding rule. Various choices for adjustment to multiplicity on the testing step lead to different thresholds. In particular, the universal threshold of Donoho & Johnstone

(1994a) and the FDR threshold of Abramovich & Benjamini (1996) fall within a testimation framework corresponding to Bonferroni and FDR multiplicity corrections, respectively.

We continue to go along the lines of testimation approach, where we utilize the recently developed maximum *a posteriori* (MAP) Bayesian multiple testing procedure of Abramovich & Angelini (2006). A hierarchical prior model used in their approach is based on imposing a prior distribution on the number of false null hypotheses. Abramovich, Grinshtein & Pensky (2007) applied it to estimating a high-dimensional Gaussian mean vector and showed the optimality (in the minimax sense) of such MAP testimation approach, where the unknown mean vector was assumed to be sparse.

We first extend the results of Abramovich, Grinshtein & Pensky (2007) to more general settings. Consider the problem of estimating an unknown high-dimensional Gaussian mean vector, where one observes Gaussian data y_i governed by

$$y_i = \mu_i + \sigma_n z_i, \quad i = 1, 2, \ldots, n. \tag{3.2}$$

The variance $\sigma_n^2 > 0$, that may depend on n, is assumed to be known, z_i are independent $N(0,1)$ random variables, and the unknown mean vector $\mu = (\mu_1, \ldots, \mu_n)'$ is assumed to lie in a strong l_p-ball $l_p[\eta_n]$, $0 < p \leq \infty$, of a standardized radius η_n, that is, $||\mu||_p \leq C_n$, where $C_n = n^{1/p}\sigma_n\eta_n$. Abramovich, Grinshtein & Pensky (2007) considered the Gaussian sequence model (3.2) with $\sigma_n^2 = \sigma^2$, and obtained upper error bounds of an *adaptive* MAP testimator of μ in the sparse case, where $0 < p < 2$ and $\eta_n \to 0$ as $n \to \infty$. We extend their results for all combinations of p and η_n, and for variance in the Gaussian sequence model (3.2) that may depend on n. We show, in particular, that for a properly chosen prior distribution on the number of non-zero entries of μ, the corresponding adaptive MAP testimator of μ is asymptotically minimax (up to a constant factor) for almost all strong l_p-balls including both sparse and dense cases.

We then apply the MAP testimation framework to the wavelet thresholding estimation in the standard Gaussian white noise model (3.1). We show that, under mild conditions on the prior distribution on the number of non-zero wavelet coefficients, a global MAP wavelet testimator of f, where the MAP testimation procedure is applied to the entire set of wavelet coefficients at all resolution levels, is adaptive and asymptotically nearly-minimax (up to an additional logarithmic factor) over the entire range of Besov balls. Furthermore, we demonstrate that performing the MAP testimation procedure at each resolution level separately allows one to remove the extra logarithmic factor. Moreover, these results can be also extended to the estimation of derivatives of f. In a way, these results complement recent adaptively optimal estimators obtained in an empirical Bayes context (see Johnstone & Silverman, 2005).

In what follows, we review the MAP testimation methodology proposed by Abramovich, Grinshtein & Pensky (2007) for estimating the unknown high-dimensional mean vector μ in the Gaussian sequence model (3.1). Their upper error bounds are generalized to a wider, not necessarily sparse, range of strong l_p-balls, $0 < p \leq \infty$, while for a properly chosen prior, the resulting MAP testimator is asymptotically minimax (up to a constant factor) over a wide range of sparse and dense strong l_p-balls. Analogous results can be obtained for l_0-balls and weak l_p-balls, $0 < p < 2$. Adaptive global and level-wise MAP wavelet testimators of the unknown response function f in the Gaussian white noise model (3.1) are proposed, and their asymptotic optimality (in the minimax sense) under the L^2-risk is established in a wide range of Besov balls. These results can also be extended to the estimation of derivatives of f. Additionally, using the boundary-modified coiflets of Johnstone & Silverman (2004a), it is shown that the order of magnitude of the accuracy of the proposed level-wise MAP wavelet testimator is not affected when the sampled data model is used. Finally, we illustrate the performance of the proposed adaptive level-wise MAP wavelet testimator on several simulated examples, and compare it with three adaptive empirical Bayes estimation procedures and one block wavelet thresholding estimator that have recently been shown to attain the optimal (or near-optimal) convergence rates and to perform well in finite sample situations. An application to a dataset collected in an anaesthesiological study is also presented.

3.2 MAP testimation in the Gaussian sequence model

3.2.1 MAP testimation procedure

We start with reviewing the MAP testimation procedure for the Gaussian sequence model (3.2) developed by Abramovich, Grinshtein & Pensky (2007).

For this model, consider the multiple hypothesis testing problem, where we wish to simultaneously test

$$H_{0i} : \mu_i = 0 \quad \text{versus} \quad H_{1i} : \mu_i \neq 0, \quad i = 1, 2, \dots, n.$$

A configuration of true and false null hypotheses is uniquely defined by the indicator vector $x = (x_1, \dots, x_n)'$, where $x_i = \mathbb{I}(\mu_i \neq 0),\ i = 1, 2, \dots, n$. (Here, $\mathbb{I}(A)$ denotes the indicator function of the set A.) Let $\kappa = x_1 + ... + x_n = ||\mu||_0$ be the number of non-zero μ_i (false nulls), i.e., $||\mu||_0 = \#\{i : \mu_i \neq 0\}$. Assume some prior distribution π_n on κ with $\pi_n(\kappa) > 0,\ \kappa = 0, 1, \dots, n$. For a given κ, there are $\binom{n}{\kappa}$ different vectors x. Assume all of them to be equally likely *a priori*, that is, conditionally on κ,

$$\mathbb{P}\left(x \,\Big|\, \sum_{i=1}^{n} x_i = \kappa \right) = \binom{n}{\kappa}^{-1}.$$

Naturally, $(\mu_i \,|\, x_i = 0) \sim \delta_0$, where δ_0 is a probability atom at zero. To complete the prior specification, we assume that $(\mu_i \,|\, x_i = 1) \sim N(0, \tau_n^2)$.

For the proposed hierarchical prior, the posterior probability of a given vector x with κ non-zero entries is

$$\pi_n(x, \kappa \mid y) \propto \binom{n}{\kappa}^{-1} \pi_n(\kappa)\, \mathbb{I}\left(\sum_{i=1}^{n} x_i = \kappa \right) \prod_{i=1}^{n} (B_i^{-1})^{x_i}, \tag{3.3}$$

where the Bayes factor B_i of H_{0i} is

$$B_i = \sqrt{1 + \gamma_n}\ \exp\left\{ -\frac{y_i^2}{2\sigma_n^2(1 + 1/\gamma_n)} \right\} \tag{3.4}$$

and $\gamma_n = \tau_n^2 / \sigma_n^2$ is the variance ratio (see Abramovich & Angelini, 2006).

Given the posterior distribution $\pi_n(x, \kappa \mid y)$, we apply a MAP rule to choose the most likely indicator vector. Generally, to find the posterior mode of $\pi_n(x, \kappa \mid y)$, one should look through all 2^n possible sequences of zeroes and ones. However, for the proposed model, the number of candidates for a mode is, in fact, reduced to $n+1$ only. Indeed, let $\hat{x}(\kappa)$ be a maximizer of (3.3) for a fixed κ that indicates the most plausible vector x with κ non-zero entries. From (3.3), it follows immediately that $\hat{x}_i(\kappa) = 1$ at the κ entries corresponding to the smallest Bayes factors B_i and zeroes otherwise. Due to the monotonicity of B_i in $|y|_i$ (see (3.4)), it is equivalent to $\hat{x}_i(\kappa) = 1$ corresponding to the κ largest $|y|_i$ and zeroes for others. The Bayesian MAP multiple testing procedure then leads to finding $\hat{\kappa}$ that maximizes

$$\log \pi_n(\hat{x}(\kappa), \kappa \mid y) = const + \sum_{i=1}^{\kappa} y_{(i)}^2 + 2\sigma_n^2(1+1/\gamma_n) \log \left\{ \binom{n}{\kappa}^{-1} \pi_n(\kappa)(1+\gamma_n)^{-\frac{\kappa}{2}} \right\},$$

or, equivalently, minimizes

$$\sum_{i=\kappa+1}^{n} y_{(i)}^2 + 2\sigma_n^2(1+1/\gamma_n) \log \left\{ \binom{n}{\kappa} \pi_n^{-1}(\kappa)(1+\gamma_n)^{\frac{\kappa}{2}} \right\},$$

where $|y|_{(1)} \geq \ldots \geq |y|_{(n)}$. The $\hat{\kappa}$ null hypotheses corresponding to $|y|_{(1)}, \ldots, |y|_{(\hat{\kappa})}$ are rejected. The resulting Bayesian testimation yields a hard thresholding with a threshold $\hat{\lambda}_{MAP} = |y|_{(\hat{\kappa})}$, i.e.,

$$\hat{\mu}_i = \begin{cases} y_i, & |y_i| \geq \hat{\lambda}_{MAP}, \\ 0, & otherwise. \end{cases}$$

(If $\hat{\kappa} = 0$, then all y_i, $i = 1, 2, \ldots, n$, are thresholded and $\hat{\mu} \equiv 0$.)

Various thresholding rules can be considered as penalized likelihood estimators minimizing

$$\|y - \mu\|_2^2 + P(\mu) \tag{3.5}$$

for the corresponding penalties $P(\mu)$. Complexity type penalties are placed on the number of nonzero μ_i. Let $\|\mu_0\| = \sharp\{i : \mu_i \neq 0\}$. For a general complexity type penalty $P_n(\|\mu_0\|)$, the corresponding penalized estimator $\hat{\mu}^*$ is a hard thresholding rule with the data-dependent threshold $\hat{\lambda} = |y|_{(\hat{\kappa})}$, where $\hat{\kappa}$ is the minimizer of

$$\sum_{i=\kappa+1}^{n} y_{(i)}^2 + P_n(\kappa).$$

From a frequentist viewpoint, the above MAP *testimator* $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)'$ can be thought as a penalized likelihood estimator with the complexity penalty (see, e.g., Birgé & Massart, 2001)

$$P_n(\kappa) = 2\sigma_n^2(1+1/\gamma_n) \log \left\{ \binom{n}{\kappa} \pi_n^{-1}(\kappa)(1+\gamma_n)^{\frac{\kappa}{2}} \right\}. \tag{3.6}$$

3.2.2 Upper error bounds for MAP testimation

Abramovich, Grinshtein & Pensky (2007, Theorem 6) obtained upper error bounds for the l^2-risk of the resulting MAP testimator in the Gaussian sequence model (3.2) (with $\sigma_n^2 = \sigma^2$) for sparse strong $l_p[\eta_n]$-balls, where $0 < p < 2$ and $\eta_n \to 0$ as $n \to \infty$. We extend now these results to more general settings.

Fix a prior distribution $\pi_n(\kappa) > 0,\ \kappa = 0, 1, \ldots, n$, on the number of non-zero entries of μ, and let $\gamma_n = \tau_n^2/\sigma_n^2$ be the variance ratio.

Proposition 3.2.1. *Let $\hat{\mu}$ be the MAP testimator of μ in the Gaussian sequence model (3.2), let $\mu \in l_p[\eta_n]$, $0 < p \le \infty$. Assume that there exist positive constants $\underline{\gamma}$ and $\bar{\gamma}$ such that $\underline{\gamma} \le \gamma_n \le \bar{\gamma}$. Define $c(\gamma_n) = 8(\gamma_n + 3/4)^2 > 9/2$.*

1. *Let $0 < p \le \infty$. Assume $e^{-\tilde{c}(\gamma_n)n} \le \pi_n(n) \le e^{-c(\gamma_n)n}$, where $c(\gamma_n) \le \tilde{c}(\gamma_n) \le c_0$, and $c_0 > 0$. Then,*

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(\sigma_n^2 n) \quad \textit{as} \quad n \to \infty.$$

2. *Let $2 \le p \le \infty$. Assume that there exists $\beta \ge 0$ such that $\pi_n(0) \ge n^{-c_1 n^{-\beta}}$ for some $c_1 > 0$. Then,*

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(\sigma_n^2 n \eta_n^2) + O(\sigma_n^2 n^{-\beta} \log n) \quad \textit{as} \quad n \to \infty.$$

3. *Let $0 < p < 2$. Assume $\pi_n(\kappa) \ge (\kappa/n)^{c_2 \kappa}$ for all $\kappa = 1, 2, \ldots, \alpha_n n$, where $n^{-1}(2\log n)^{p/2} \le \alpha_n \le e^{-c(\gamma_n)}$ and $c_2 > 0$. Then,*

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O\big(\sigma_n^2 n \eta_n^p (2\log \eta_n^{-p})^{1-p/2}\big) \quad \textit{as} \quad n \to \infty$$

for all $n^{-1}(2\log n)^{p/2} \le \eta_n^p \le \alpha_n$.

4. *Let $0 < p < 2$. Assume that there exists $\beta \ge 0$ such that $\pi_n(0) \ge n^{-c_1 n^{-\beta}}$ for some $c_1 > 0$. Then,*

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(\sigma_n^2 n^{2/p} \eta_n^2) + O(\sigma_n^2 n^{-\beta} \log n) \quad \textit{as} \quad n \to \infty$$

for all $\eta_n^p < n^{-1}(2\log n)^{p/2}$.

Note first that since the prior assumptions in Proposition 3.2.1 do not depend on the parameters p and η_n of the strong l_p-ball, the resulting MAP testimator is inherently adaptive. The condition on $\pi_n(n)$ guarantees that the risk of the MAP testimator is always bounded by an order of $n\sigma_n^2$, corresponding to the risk of the maximum likelihood estimator, $\hat{\mu}_i^{MLE} = y_i$, in the Gaussian sequence model (3.2).

The following corollary of Proposition 3.2.1 essentially defines two zones (dense and sparse) for $p \geq 2$ and three zones (dense, sparse and super-sparse), for $0 < p < 2$, of different behavior for the quadratic risk of the MAP testimator. To evaluate the accuracy of the MAP testimator, we also compare the resulting risks with the corresponding minimax risks $R(l_p[\eta_n]) = \inf_{\tilde{\mu}} \sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\tilde{\mu} - \mu||_2^2$ that can be found, e.g., in Donoho & Johnstone (1994b). (Here, $g_1(n) \asymp g_2(n)$ denotes $0 < \liminf(g_1(n)/g_2(n)) \leq \limsup(g_1(n)/g_2(n)) < \infty$ as $n \to \infty$.)

Collorary 3.2.1. *Let $\hat{\mu}$ be the MAP testimator of μ in the Gaussian sequence model (3.2), where $\mu \in l_p[\eta_n]$, $0 < p \leq \infty$. Assume that there exist positive constants $\underline{\gamma}$ and $\bar{\gamma}$ such that $\underline{\gamma} \leq \gamma_n \leq \bar{\gamma}$. Define $c(\gamma_n) = 8(\gamma_n + 3/4)^2 > 9/2$ and let the prior π_n satisfy the following conditions*

1. *$\pi_n(0) \geq n^{-c_1 n^{-\beta}}$ for some $\beta \geq 0$ and $c_1 > 0$;*

2. *$\pi_n(\kappa) \geq (\kappa/n)^{c_2 \kappa}$ for all $\kappa = 1, 2, \ldots, \alpha n$, where $\alpha = e^{-9/2}$ (or, $\alpha = e^{-c(\underline{\gamma})}$ if $\underline{\gamma}$ is known) and $c_2 > 0$;*

3. *there exists $c(\gamma_n) \leq \tilde{c}(\gamma_n) \leq c_0$, where $c_0 > 0$, such that $e^{-\tilde{c}(\gamma_n)n} \leq \pi_n(n) \leq e^{-c(\gamma_n)n}$.*

Then, as $n \to \infty$, depending on p and η_n, one has:

Case 1, $0 < p \leq \infty$, $\eta_n^p > \alpha$.

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(n\sigma_n^2), \quad \textit{while} \quad R(l_p[\eta_n]) \asymp n\sigma_n^2.$$

Case 2, $p \geq 2$, $\eta_n^p \leq \alpha$.

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(\sigma_n^2 n \eta_n^2) + O(\sigma_n^2 n^{-\beta} \log n), \quad \textit{while} \quad R(l_p[\eta_n]) \asymp \sigma_n^2 n \eta_n^2.$$

Case 3, $0 < p < 2,\ n^{-1}(2\log n)^{p/2} \le \eta_n^p \le \alpha$.

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O\big(\sigma_n^2 n \eta_n^p (2\log \eta_n^{-p})^{1-p/2}\big),$$
$$\textit{while} \quad R(l_p[\eta_n]) \asymp \sigma_n^2 n \eta_n^p (2\log \eta_n^{-p})^{1-p/2}.$$

Case 4, $0 < p < 2,\ \eta_n^p < n^{-1}(2\log n)^{p/2}$.

$$\sup_{\mu \in l_p[\eta_n]} \mathbb{E}||\hat{\mu} - \mu||_2^2 = O(\sigma_n^2 n^{2/p}\eta_n^2) + O(\sigma_n^2 n^{-\beta}\log n), \quad \textit{while} \quad R(l_p[\eta_n]) \asymp \sigma_n^2 n^{2/p}\eta_n^2.$$

The impact of Corollary 3.2.1 is that, up to a constant factor, the MAP testimator is adaptively minimax for almost all strong l_p-balls, except those with very small standardized radiuses, where $\eta_n^2 = o(n^{-(\beta+2/\min(p,2))}\log n)$. Hence, while the optimality of the most existing threshold estimators (e.g., universal, SURE, FDR) has been established over various sparse settings, the MAP testimator is appropriate for both sparse and dense cases. To the best of our knowledge, such a wide adaptivity range can be compared only with the penalized likelihood estimators of Birgé & Massart (2001) and the empirical Bayes threshold estimators of Johnstone & Silverman (2004b, 2005). Additionally, the penalized likelihood estimators of Birgé & Massart (2001) have not been studied in practice, while the range of optimality of the empirical Bayes threshold estimators of Johnstone & Silverman (2004b, 2005) is somewhat smaller than the range of optimality of the MAP testimator.

In fact, as we have mentioned, there are interesting asymptotic relations between the MAP testimator and the penalized likelihood estimator of Birgé & Massart (2001) that may explain their similar behavior. For estimating the normal mean vector in (3.2) within strong l_p-balls Birgé & Massart (2001) considered a penalized likelihood estimator with a specific complexity penalty

$$\tilde{P}_n(\kappa) = C\sigma_n^2\kappa(1+\sqrt{2L_\kappa})^2, \tag{3.7}$$

where $L_\kappa = \log(n/\kappa) + (1+\theta)(1+\log(n)/\kappa)$ for fixed $C > 1$ and $\theta > 0$ (see their Section 6.3). Note that, for large n and $\kappa < n/e$, this penalty is approximately of the following form:

$$\tilde{P}_n(\kappa) \sim 2\sigma_n^2 c\kappa L_\kappa \sim 2\sigma_n^2\tilde{c}_1\left(\log\binom{n}{\kappa} + \tilde{c}_2\kappa\right) \tag{3.8}$$

for some positive constants $c,\ \tilde{c}_1,\ \tilde{c}_2 > 1$ (see also Lemma 3.5.1 in Section 3.4). Thus, within this range, $\tilde{P}_n$ in (3.7)-(3.8) behaves similar to a particular case of the MAP penalty P_n

in (3.6) corresponding to the geometric type prior $\pi_n(\kappa) \propto (1/\tilde{c}_2)^\kappa$. Note that this prior satisfies the second condition on π_n of Corollary 3.2.1. Such a Bayesian interpretation can be also helpful in providing some intuition behind the penalty $\tilde{P}_n$ motivated in Birgé & Massart (2001) mostly by technical reasons. In addition, under the conditions of Corollary 3.2.1, $P_n(n) \sim \tilde{P}_n(n) \sim cn$.

Furthermore, for $\kappa \ll n$ (sparse cases), under the conditions on the prior π_n of Corollary 3.2.1, both the MAP penalty P_n and the penalty $\tilde{P}_n$ of Birgé & Massart (2001) are of the same so-called $2\kappa \log(n/\kappa)$-type penalties of the form $2\sigma_n^2 \zeta\kappa(\log(n/\kappa) + c_{\kappa,n})$, where $\zeta > 1$ and $c_{\kappa,n}$ is negligible relative to $\log(n/\kappa)$. Within different frameworks, such type of penalties arose in a series of recent works on estimation and model selection (e.g., Foster & Stine, 1999; George & Foster, 2000; Birgé & Massart, 2001; Johnstone, 2002, Chapter 13; Abramovich, Benjamini, Donoho & Johnstone, 2006; Abramovich, Grinshtein & Pensky, 2007).

The following proposition extends the results of Abramovich, Grinshtein & Pensky (2007) showing that the MAP testimator attains the rate which is optimal as $\eta_n \to 0$ under l_0-balls for any $\frac{1}{n} \leq \eta_n \leq \frac{e^{-c(\gamma)}}{2}$. Recall that $R(l_0(\eta_n)) \sim \sigma_n^2 n\eta_n(2\log(\eta_n^{-1}))$ (see Donoho, Jonhstone, Hoch & Stern (1992)). (Below and in the remaining of the thesis, $\lfloor \cdot \rfloor$ denotes the floor function. Also, $g_1(n) \sim g_2(n)$ denotes $g_1(n)/g_2(n) \to 1$ as $n \to \infty$.)

Proposition 3.2.2. *Define $k_n^* = \lfloor n\eta_n \rfloor$. Let $\eta_n \in [\frac{1}{n}, \frac{e^{-c(\gamma)}}{2}]$. If there exists a constant $c_0 > 0$ such that $\pi_n(k_n^*) \geq (k_n^*/n)^{c_0 k_n^*}$, then MAP testimator $\hat{\mu}^*$ satisfies*

$$\sup_{\mu \in l_0[\eta_n]} E||\hat{\mu}^* - \mu||^2 = O(n\eta_n(2\log \eta_n^{-1})).$$

Remark 3.2.1 The proof of Proposition 3.2.1 which is given in Section 3.4 shows that the proposed estimator is also adaptively minimax over weak l_p-balls, $0 < p < 2$, for sparse and partially dense cases. Note that neither of the estimators mentioned after Collorary 3.2.1 has been shown to be adaptively minimax over weak l_p-balls, $0 < p < 2$.

3.3 MAP wavelet testimation in the Gaussian white noise and sampled data models

In this section, we apply the results of Section 3.2 on MAP testimation in the Gaussian sequence model (3.2) to wavelet estimation of the unknown response function f in the standard Gaussian white noise model (3.1).

Given a compactly supported scaling function ϕ of regularity $r > s$ and the corresponding mother wavelet ψ, one can generate an orthonormal wavelet basis on the unit interval from a finite number C_{j_0} of scaling functions $\phi_{j_0 k}$ at a primary resolution level j_0 and wavelets ψ_{jk} at resolution levels $j \geq j_0$ and scales $k = 0, 1, \ldots, 2^j - 1$ (see, e.g., Cohen, Daubechies & Vial, 1993; Johnstone & Silverman, 2004a). For clarity of exposition, we use the same notation for interior and edge wavelets, and in what follows denote $\phi_{j_0 k}$ by $\psi_{j_0-1,k}$.

Then, f is expanded in the orthonormal wavelet series on $[0, 1]$ as

$$f(t) = \sum_{j=j_0-1}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk} \psi_{jk}(t),$$

where $\theta_{jk} = \int_0^1 f(t)\psi_{jk}(t)dt$. In the wavelet domain, the Gaussian white noise model (3.1) becomes

$$Y_{jk} = \theta_{jk} + \epsilon_{jk}, \quad j \geq j_0 - 1,\ k = 0, 1, \ldots, 2^j - 1,$$

where the empirical wavelet coefficients Y_{jk} are given by $Y_{jk} = \int_0^1 \psi_{jk}(t)dY(t)$ and ϵ_{jk} are independent $N(0, \sigma^2/N)$ random variables.

Define $J = \log_2 N$. Estimate wavelet coefficients θ_{jk} at different resolution levels j by the following scheme:

1. set $\hat{\theta}_{j_0-1,k} = Y_{j_0-1,k}$;

2. apply the MAP testimation procedure of Abramovich, Grinshtein & Pensky (2007) described in Section 3.2 to estimate θ_{jk} at resolution levels $j_0 \leq j < J$ by the corresponding $\hat{\theta}_{j,k}$;

3. set $\hat{\theta}_{jk} = 0,\ j \geq J$.

The resulting MAP wavelet testimator $\hat{f}_N$ of f is then defined as

$$\hat{f}_N(t) = \sum_{k=0}^{C_{j_0}-1} Y_{j_0-1,k}\psi_{j_0-1,k}(t) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\psi_{jk}(t).$$

In what follows, we show that, under mild conditions on the prior π_N, the resulting global MAP wavelet testimator of f, where the MAP testimation procedure is applied to the entire set of wavelet coefficients at all resolution levels $j_0 \le j < J$, is asymptotically nearly-minimax (up to an additional logarithmic factor) over the entire range of Besov classes (see Theorem 3.3.1). Furthermore, we demonstrate that performing the MAP testimation procedure at each resolution level separately allows one to remove the extra logarithmic factor (see Theorem 3.3.2). Moreover, a level-wise version of the MAP testimation procedure allows one to estimate the derivatives of f at optimal convergence rates as well (see Theorem 3.3.3).

3.3.1 Global MAP wavelet testimator

The number of wavelet coefficients at all resolution levels up to J is $\tilde{N} = 2^J - 2^{j_0} \sim N$ for large N. Let $\pi_N(\kappa) > 0,\ \kappa = 0, 1, \dots, \tilde{N}$, be a prior distribution on the number of non-zero wavelet coefficients of f at all resolution levels $j_0 \le j < J$, and let the prior variance of non-zero coefficients be τ^2/N. The corresponding variance ratio is $\gamma = \tau^2/\sigma^2$.

It is well-known (see Donoho & Johnstone, 1998) that the minimax convergence rate for the L^2-risk of estimating the unknown response function f in the Gaussian white noise model (3.1) over Besov balls $B^s_{p,q}(M)$, where $s > \max(0, 1/p - 1/2)$, $0 < p, q \le \infty$ and $M > 0$, is given by

$$\inf_{\tilde{f}_N} \sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\tilde{f}_N - f||_2^2 \asymp N^{-2s/(2s+1)} \quad \text{as} \quad N \to \infty,$$

where the infimum is taken over all estimators (i.e., measurable functions) $\tilde{f}_N$ of f based on observations from (3.1).

Theorem 3.3.1. *Let ψ be a mother wavelet of regularity r and let $\hat{f}_N$ be the corresponding global MAP wavelet testimator of f in the Gaussian white noise model (3.1), where $f \in B^s_{p,q}(M)$, $0 < p, q \le \infty$, $\frac{1}{p} < s < r$ and $M > 0$. Assume that there exist positive constants $\underline{\gamma}$ and $\bar{\gamma}$ such that $\underline{\gamma} \le \gamma_j \le \bar{\gamma}$ for all $j = j_0, j_0+1, \dots, J-1$. Let the prior π_N satisfy*

$\pi_N(\kappa) \geq (\kappa/N)^{c\kappa}$ *for all* $\kappa = 1, 2, \ldots, \lfloor e^{-9/2}N \rfloor$. *Then,*

$$\sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\hat{f}_N - f||_2^2 = O\left(\left(\frac{\log N}{N}\right)^{\frac{2s}{2s+1}}\right) \quad as \quad N \to \infty. \tag{3.9}$$

The proof of Theorem 3.3.1 is based on the well-known relation between the smoothness conditions on functions within Besov spaces and the conditions on their wavelet coefficients. Namely, if $f \in B^s_{p,q}(M)$, then the sequence of its wavelet coefficients $\{\theta_{jk},\ k = 0, 1, \ldots, 2^j - 1,\ j = j_0, j_0 + 1, \ldots, J - 1\}$ belongs to a weak $l_{2/(2s+1)}$-ball of a radius aM, where the constant a depends only on a chosen wavelet basis (see Lemma 2 in Donoho, 1993). One can then apply the corresponding results of Abramovich, Grinshtein & Pensky (2007) for MAP testimation over weak l_p-balls, $0 < p < \infty$. Details of the proof of Theorem 3.3.1 are given in Section 3.4.

The resulting global MAP wavelet testimator does not rely on the knowledge of the parameters s, p, q and M of a specific Besov ball and it is, therefore, inherently adaptive. Theorem 3.3.1 establishes the upper bound for its L^2-risk and shows that the resulting adaptive global MAP wavelet testimator is asymptotically nearly-minimax within the entire range of Besov balls. In fact, the additional logarithmic factor in (3.9) is the unavoidable minimal price for adaptivity for any global wavelet threshold estimator (see, e.g., Donoho, Johnstone, Kerkyacharian & Picard, 1995; Cai, 1999), and in this sense, the upper bound for the convergence rates in (3.9) is sharp. To remove this logarithmic factor one should consider *level-wise* thresholding (see Section 3.3.2).

3.3.2 Level-wise MAP wavelet testimator

Consider now the MAP wavelet testimation applied separately at each resolution level j. The number of wavelet coefficients at the j-th resolution level is $n_j = 2^j$. Let $\pi_j(\kappa) > 0,\ \kappa = 0, 1, \ldots, 2^j$, be the prior distribution on the number of non-zero wavelet coefficients, and let τ_j^2/N be their prior variance, $j_0 \leq j < J$. The corresponding level-wise variance ratios are $\gamma_j = \tau_j^2/\sigma^2$.

Theorem 3.3.2. *Let* ψ *be a mother wavelet of regularity* r *and let* $\hat{f}_N(\cdot)$ *be the corresponding level-wise MAP wavelet testimator of* f *in the Gaussian white noise model (3.1), where*

$f \in B^s_{p,q}(M)$, $0 < p, q \leq \infty$, $\frac{1}{p} < s < r$ and $M > 0$. Assume that there exist positive constants $\underline{\gamma}$ and $\bar{\gamma}$ such that $\underline{\gamma} \leq \gamma_j \leq \bar{\gamma}$ for all $j = j_0, j_0 + 1, \ldots, J - 1$. Let the priors π_j satisfy the following conditions for all $j = j_0, j_0 + 1, \ldots, J - 1$:

1. *$\pi_j(0) \geq 2^{-c_1 j}$ for some $c_1 > 0$;*

2. *$\pi_j(\kappa) \geq (\kappa 2^{-j})^{c_2 \kappa}$ for all $\kappa = 1, 2, \ldots, \alpha_j 2^j$, where $c_2 > 0$ and $0 < c_\alpha \leq \alpha_j \leq e^{-c(\gamma_j)}$ for some constant $c_\alpha > 0$ and $c(\gamma_j) = 8(\gamma_j + 3/4)^2$;*

3. *there exists $c(\gamma_j) \leq \tilde{c}(\gamma_j) \leq c_0$, where $c_0 > 0$ such that $e^{-\tilde{c}(\gamma_j)2^j} \leq \pi_j(2^j) \leq e^{-c(\gamma_j)2^j}$.*

Then,

$$\sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\hat{f}_N - f||_2^2 = O\Big(N^{-\frac{2s}{2s+1}}\Big) \quad \textit{as} \quad N \to \infty.$$

For $f \in B^s_{p,q}(M)$, the sequence of its wavelet coefficients at the j-th resolution level belongs to $l_p[\eta_j]$, where $\eta_j = C_0 N^{1/2} 2^{-j(s+1/2)}$ for some $C_0 > 0$ (see, e.g., Meyer, 1992, Section 6.10). The conditions on the prior in Theorem 3.3.2 ensure that all the four statements of the Proposition 3.2.1 simultaneously hold at all resolution levels $j_0 \leq j < J$ with $\beta = 0$, and one can exploit any of them at each resolution level. It is necessary for *adaptivity* of the resulting level-wise MAP wavelet testimator.

The assumptions of Theorem 3.3.2 are, in fact, not too restrictive. For example, one can easily verify that all three conditions of Theorem 3.3.2 hold for the truncated geometric prior TrGeom$(1 - q_j)$ with the probability of success $p_j = 1 - q_j$ given by $\pi_j(\kappa) = (1 - q_j)q_j^\kappa/(1 - q_j^{2^j+1})$, $\kappa = 0, 1, \ldots, 2^j$, and $q_j \sim e^{-c(\gamma_j)}$. On the other hand, no binomial prior, Bin$(2^j, p_j)$ can "kill three birds with one stone". The requirement $\pi_j(0) = (1 - p_j)^{2^j} \geq 2^{-c_1 j}$, necessarily implies $p_j \to 0$ as $j \to \infty$. However, to satisfy $\pi_j(2^j) = p_j^{2^j} \sim e^{-c(\gamma_j)2^j}$, one needs $p_j \sim e^{-c(\gamma_j)}$, which is bounded away from zero.

It turns out that requiring a slightly more stringent condition on $\pi_j(0)$, allows one also to estimate derivatives of f by the corresponding derivatives of its level-wise MAP wavelet testimator $\hat{f}_N$ at the optimal convergence rates. Such a plug-in estimation of $f^{(m)}$ by $\hat{f}_N^{(m)}$ is, in fact, along the lines of the vaguelette-wavelet decomposition approach of Abramovich & Silverman (1998).

Recall that the minimax convergence rate for the L^2-risk of estimating an m-th derivative ($m \geq 0$) of the unknown response function f in the model (3.1) over Besov balls $B^s_{p,q}(M)$, where $m < \min(s, (s+1/2-1/p)p/2)$, $0 < p, q \leq \infty$ and $M > 0$, is given by

$$\inf_{\tilde{f}_N^{(m)}} \sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\tilde{f}_N^{(m)} - f^{(m)}||_2^2 \asymp N^{-2(s-m)/(2s+1)} \quad \text{as} \quad N \to \infty,$$

where the infimum is taken over all estimators (i.e., measurable functions) $\tilde{f}_N^{(m)}$ of $f^{(m)}$ based on observations from (3.1), (see, e.g., Donoho, Johnstone, Kerkyacharian & Picard, 1997; Johnstone & Silverman, 2005).

The following Theorem 3.3.3 is a generalization of Theorem 3.3.2 for simultaneous level-wise MAP wavelet testimation of f and its derivatives.

Theorem 3.3.3. *Let ψ be a mother wavelet of regularity r and let $\hat{f}_N$ be the level-wise MAP wavelet testimator of f in the Gaussian white noise model (3.1), where $f \in B^s_{p,q}(M)$, $0 < p, q \leq \infty$, $1/p < s < r$ and $M > 0$. Assume that there exist positive constants $\underline{\gamma}$ and $\bar{\gamma}$ such that $\underline{\gamma} \leq \gamma_j \leq \bar{\gamma}$ for all $j = j_0, j_0+1, \ldots, J-1$. Let the priors π_j satisfy the following conditions for all $j = j_0, j_0+1, \ldots, J-1$:*

1. *$\pi_j(0) \geq 2^{-c_1 j 2^{-\beta j}}$ for some $\beta \geq 0$ and $c_1 > 0$;*

2. *$\pi_j(\kappa) \geq (\kappa 2^{-j})^{c_2 \kappa}$ for all $\kappa = 1, 2, \ldots, \alpha_j 2^j$, where $c_2 > 0$ and $0 < c_\alpha \leq \alpha_j \leq e^{-c(\gamma_j)}$ for some constant $c_\alpha > 0$, and $c(\gamma_j) = 8(\gamma_j + 3/4)^2$;*

3. *there exists $c(\gamma_j) \leq \tilde{c}(\gamma_j) \leq c_0$, where $c_0 > 0$ such that $e^{-\tilde{c}(\gamma_j)2^j} \leq \pi_j(2^j) \leq e^{-c(\gamma_j)2^j}$.*

Then, for all m-th derivatives $f^{(m)}$ of f, where $0 \leq m \leq \beta/2$ and $m < \min(s, (s+1/2-1/p)p/2)$,

$$\sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\hat{f}_N^{(m)} - f^{(m)}||_2^2 = O\Big(N^{-\frac{2(s-m)}{2s+1}}\Big) \quad \textit{as} \quad N \to \infty.$$

Theorem 3.3.2 is evidently a particular case of Theorem 3.3.3 corresponding to the case $m = 0$, for $\beta = 0$ in the condition on $\pi_j(0)$. The range of derivatives is the same as that for the empirical Bayes shrinkage and threshold estimators appearing in Theorem 1 of Johnstone & Silverman (2005). The proof of Theorem 3.3.3 is given in the Section 3.5

3.3.3 MAP testimation procedure under the sampled data model

The following theorem considers the sample data model introduced by Johnstone & Silverman (2004a). The sampled data model is given by

$$Y_i = f(\frac{i}{N}) + \epsilon_i \quad i = 1, \dots, N,$$

where ϵ_i are independent and identically distributed random variables from $N(0,1)$. Assume that for $N = 2^J$, we have sufficient observations to evaluate the preconditioned sequence $P_J Y$ defined in Chapter 1. Let R be the number of continuous derivatives of the scaling function ϕ. Suppose that the wavelets and scaling functions are modified by the boundary construction described in Chapter 1. Let $\tilde{Y}$ be the boundary corrected discrete wavelet transform of $N^{-\frac{1}{2}} P_J Y$. Define the estimated coefficient array $\hat{\theta}$ as follows

1. Set $\hat{\theta}_{j_0-1} = \tilde{Y}_{j_0-1}$.

2. Estimate the interior coefficients $\hat{\theta}^I$ by applying the MAP procedure on $\tilde{Y}^I$ for each $j_0 \leq j < J$ under the following assumptions on the prior

   (a) $\pi_j(0) \geq (2^j - 2S + 2)^{-c_1}$ for some $c_1 > 0$, where $S > R$.

   (b) $\pi_j(\kappa) \geq (\frac{\kappa}{2^j - 2S + 2})^{c_2 \kappa}$ for all $\kappa = 1, ..., \alpha_j(2^j - 2S + 2)$,

   where $c_2 > 0$ and $0 < c_\alpha \leq \alpha_j \leq e^{-c(\gamma_j)}$ for some constant $c_\alpha > 0$.

   (c) there exists $c(\gamma_j) \leq \tilde{c}(\gamma_j) \leq c_0$, where $c_0 > 0$ such that

   $e^{-\tilde{c}(\gamma_j)(2^j - 2S + 2)} \leq \pi_j(2^j - 2S + 2) \leq e^{-c(\gamma_j)(2^j - 2S + 2)}$.

3. Threshold the boundary coefficients separately. At level j, use a hard threshold of $\tau(\frac{j}{N})^{\frac{1}{2}}$, where $\tau > 0$, so that for each $k \in K_j^B$, $\hat{\theta}_{jk} = \tilde{Y}_{jk} \mathbb{I}(|\tilde{Y}_{jk}| > \tau(\frac{j}{N})^{\frac{1}{2}})$.

4. For unobserved levels $j \geq J$ set $\hat{\theta}_{jk} = 0$.

Theorem 3.3.4. *Assume that the scaling function ϕ and the mother wavelet ψ have R continuous derivatives and support [-S+1,S] for some integer S and that $\int x^m \phi(x) dx = 0$ for $m = 1, 2, \dots, R-1$. Assume that the wavelets and scaling functions are modified by the boundary construction described in Chapter 1. Then, the construction of the estimator is set* *out as above. Suppose that $0 < s < R$, $0 < p \leq \infty$ and either i) $s > \frac{1}{p}$ or ii) $s = p = 1$. Let $\mathcal{F}(C)$ be the set of functions f whose wavelet coefficients fall in $B^s_{p,\infty}(C)$. Then, there is a constant c independent of C and N such that*

$$\sup_{f\in\mathcal{F}(C)} R^*_N(f) \leq c\{\frac{C^{2s/(2s+1)}}{N^{2s/(2s+1)}} + o(\frac{1}{N^{2s/(2s+1)}})\},$$

where

$$R^*_N(f) = \mathbb{E}(\|\hat{\theta}_{j_0-1} - \theta_{j_0-1}\|_2^2) + \sum_{j=j_0}^{\infty} \mathbb{E}(\|\hat{\theta}_j - \theta_j\|_2^2).$$

3.4 Numerical Study

In this section, we present a simulation study to illustrate the performance of the developed level-wise MAP wavelet testimator and compare it with three empirical Bayes wavelet estimation procedures and one block wavelet thresholding estimation method, namely, the posterior mean (PostMean) and posterior median (PostMed) wavelet estimators proposed in Johnstone & Silverman (2005), the Bayes Factor (BF) wavelet estimator proposed in Pensky & Sapatinas (2007) and the NeighBlock (Block) wavelet thresholding estimator proposed in Cai & Silverman (2001). We note that all estimators are adaptive to the unknown smoothness and attain the optimal convergence rate, except for the Block estimator that is near optimal (up to a logarithmic factor).

The computational algorithms related to wavelet analysis were performed using the WaveLab software (`http://www-stat.stanford.edu/software/software.html`) and the EbayesThresh software
(`http://www-lmc.imag.fr/lmc-sms/Anestis.Antoniadis/EBayesThresh`). The entire study was carried out using the Matlab programming environment.

3.4.1 Estimation of parameters

To apply the level-wise MAP wavelet testimator one should specify the priors π_j, the noise variance σ^2 and the prior variances τ_j^2 or, equivalently, the variance ratios $\gamma_j = \tau_j^2/\sigma^2$. We used the truncated geometric priors TrGeom$(1 - q_j)$ discussed in Section 3.3.2. Since the parameters σ^2, q_j and γ_j are rarely known *a priori* in practice, they should be estimated from the data in the spirit of empirical Bayes.

The unknown σ was robustly estimated by the median of the absolute deviation of the empirical wavelet coefficients at the finest resolution level $J-1$, divided by 0.6745 as suggested by Donoho & Johnstone (1994), and usually applied in practice. For a given σ, we then estimate q_j and γ_j by the conditional likelihood approach of Clyde & George (1999).

Consider the prior model described in Section 3.2.1. The corresponding marginal likelihood of the observed empirical wavelet coefficients, say Y_{jk}, at the j-th resolution level is then given by

$$L(q_j, \gamma_j; Y_j) \propto \sum_{\kappa=0}^{2^j} \pi_j(\kappa) \binom{2^j}{\kappa}^{-1} (1+\gamma_j)^{-\frac{\kappa}{2}} \sum_{\{x_i : \sum_k x_{ik} = \kappa\}} \exp\left\{ \frac{\gamma_j \sum_k x_{ik} Y_{jk}^2}{2\sigma^2(1+\gamma_j)} \right\},$$

where $\pi_j(\kappa) = (1-q_j)q_j^\kappa/(1-q_j^{2^j+1})$ and x_i are indicator vectors. Instead of direct maximization of $L(q_j, \gamma_j; Y_j)$ with respect to q_j and γ_j, regard the indicator vector x as a latent variable and consider the corresponding log-likelihood for the "augmented" data (Y_j, x), i.e.,

$$l(q_j, \gamma_j; Y_j, x) = const + \log \pi_j(\kappa) - \log \binom{2^j}{\kappa} - \frac{\kappa}{2} \log(1+\gamma_j) + \frac{\gamma_j \sum_k x_{ik} Y_{jk}^2}{2\sigma^2(1+\gamma_j)}. \tag{3.10}$$

The EM-algorithm iteratively alternates between computation of the expectation of $l(q_j, \gamma_j; Y_j, x)$ in (3.10) with respect to the distribution of x given Y_j evaluated using the current estimates for current values of parameters (E-step), and updating the parameters by maximizing it with respect to q_j and γ_j (M-step). However, for a general prior distribution π_n (and for the truncated geometric prior, in particular), the EM-algorithm does not allow one to get analytic expressions on the E-step. Instead, we apply the conditional likelihood estimation approach originated by George & Foster (2000) and adapted to the wavelet estimation context by Clyde & George (1999). The approach is based on evaluating the augmented log-likelihood (3.10) at the *mode* for the indicator vector x at the E-step rather than using the *mean* as in the original EM-algorithm (see, e.g., Abramovich & Angelini, 2006).

For a fixed number κ of its non-zero entries, it is evidently from (3.10) that the most likely vector $\hat{x}(\kappa)$ is $\hat{x}_i(\kappa) = 1$ for the κ largest $|Y_{jk}|$ and zero otherwise. For the given κ, maximizing

(3.10) with respect to γ_j after some algebra one has $\hat{\gamma}_j(\kappa) = \max\Big(0, \sum_{k=1}^{\kappa} Y_{(k)}^2/(\kappa\sigma^2) - 1\Big)$. To simplify maximization with respect to q_j, approximate the truncated geometric distribution π_j in (3.10) by a non-truncated one. This approximation does not strongly affect the results, especially at sufficiently high resolution levels, and allows one to get analytic solutions for $\hat{q}_j$, i.e., $\hat{q}_j(\kappa) = \kappa/(\kappa+1)$. It is now straightforward to find $\hat{\kappa}$ that maximizes (3.10) and the corresponding $\hat{\gamma}_j(\hat{\kappa})$ and $\hat{q}_j(\hat{\kappa})$. The above conditional likelihood approach results thus in rapidly computable estimates for γ_j and q_j in closed forms.

3.4.2 Simulation study

We now present and discuss the results of the simulation study. For PostMean, PostMed and BF wavelet estimators we used the Double-exponential prior, where the corresponding prior parameters were estimated level-by-level by marginal likelihood maximization, as described in Johnstone & Silverman (2005). The prior parameters for the level-wise MAP wavelet testimator were estimated by conditional likelihood maximization described in Section 3.4.1 above. For the Block wavelet estimator, the lengths of the overlapping and non-overlapping blocks and the value of the thresholding coefficient, associated with the method, were selected as suggested by Cai & Silverman (2001). Finally, for all competing methods, σ was estimated by the median of the absolute value of the empirical wavelet coefficients at the finest resolution level divided by 0.6745 as we have discussed in Section 3.4.1.

In the simulation study, we evaluated the above five estimators for a series of test functions. We present the results for the *Wave, Peak, Bumps* and *HeaviSine* test functions defined on $[0, 1]$.

For each test function, $M = 100$ samples were generated by adding independent Gaussian noise $\varepsilon \sim N(0, \sigma^2)$ to $n = 1024$ equally spaced points on [0,1]. The value of the (root) signal-to-noise ratio (SNR) was taken to be 3 (high noise level), 5 (moderate noise level) and 7 (low noise level), where

$$SNR(f, \sigma) = \sigma^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} (f(t_i) - \bar{f})^2 \right)^{1/2} \quad and \quad \bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(t_i).$$

The goodness-of-fit for an estimator $\hat{f}$ of f in a single replication was measured by its mean

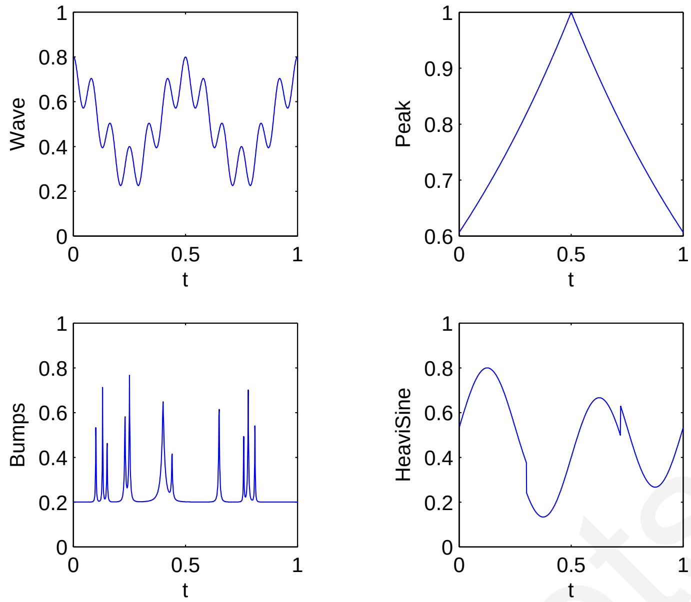

Figure 3.1: Wave, Peak, Bumps and HeaviSine test functions sampled at $n = 1024$ points.

squared error (MSE), defined as

$$MSE(f, \hat{f}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}(t_i) - f(t_i))^2.$$

We report the results for $n = 1024$ using the compactly supported mother wavelet *Coiflet 3* (see Daubechies, 1992, p.258) and the primary resolution level $j_0 = 4$ (different choices of wavelet functions and resolution levels yielded basically similar results in magnitude). The sample distributions of MSE over replications for all estimators in simulation studies were typically highly asymmetrical and affected by outliers. Therefore, we preferred the sampled medians of MSEs rather than means to gauge the estimators' goodness-of-fit. For each estimator, test function and noise level, we calculated the median MSE over all 100 replications. To quantify the comparison between estimators over various test functions and noise levels, for each considered model we found the best estimator among the five ones, i.e., the one achieving the minimum median MSE, and evaluated the *relative* median MSE of the i-th estimator defined as $\min_{1 \leq j \leq 5}\{\mathrm{Median}(\mathrm{MSE}_j)\}/\mathrm{Median}(\mathrm{MSE}_i)$, $i = 1, 2, \ldots, 5$ (see Table 3.1). As expected, Table 3.1 shows that there is no the "uniformly best" estimator. The relative performance of each estimator depends on a specific test function and the noise level. Thus, the Block estimator, for example, is clearly the best for the *Peak* function but the worst for the *Wave*. PostMed and MAP are overall favorites among Bayesian estimators. However, the MAP testimator results in the highest *minimal* relative median MSE over all cases among the considered five estimators (see the bold numbers in Table 3.1). The minimal relative median MSE of an estimator reflects its inefficiency at the most challenging combination of a test function and SNR level and is a natural measure of its robustness. Additionally, we compared the competing estimators in terms of sparsity, measured by the total number of non-zero wavelet coefficients (averaged over 100 replications) surviving after thresholding. These results are given in Table 3.2 below. The proposed method is sparser than the empirical Bayes estimators (note that PostMean is not included in this comparison since is a non-linear shrinkage, hence all wavelet coefficients survive). The sparsity of the Neighblock thresholding estimator depends on the signal.

n	signal	SNR	MAP	BF	Postmed	Postmean	Block
1024	Wave	3	1	0.4701	0.9729	0.9041	**0.4948**
		5	1	0.4194	0.9819	0.9002	0.5773
		7	1	0.5525	0.9908	0.9017	0.7827
1024	Peak	3	0.7981	**0.1403**	**0.7464**	**0.5997**	1
		5	0.8013	0.1681	0.7835	0.6427	1
		7	0.7988	0.1824	0.881	0.7039	1
1024	Bumps	3	**0.7833**	0.8252	0.9084	1	0.6222
		5	0.7892	0.8543	0.8867	1	0.6985
		7	0.8037	0.8635	0.8885	1	0.7973
1024	HeaviSine	3	0.8488	0.3486	1	0.942	0.7348
		5	0.9113	0.3092	1	0.9698	0.6677
		7	0.8121	0.4806	1	0.987	0.69

Table 3.1: Relative median MSE for the Wave, Peak, Bumps and HeaviSine test functions sampled at $n = 1024$ data points and using three values of SNR (3, 5 and 7), for the various wavelet estimators. The minimal relative median MSE for each estimator is bold.

3.4.3 Inductance plethysmography data

We now consider a dataset from anaesthesiology collected by inductance plethysmography to illustrate the performance of the level-wise MAP wavelet testimator, and compare it with the PostMean, PostMed, BF and Block wavelet estimators. The recordings were made by

n	signal	SNR	MAP	BF	Postmed	Block
1024	Wave	3	116.64	217.61	131.24	52.97
		5	104.38	206.73	134.62	64.4
		7	115.94	195.79	127.61	64.9
1024	Peak	3	61.21	152.26	89.08	16.45
		5	80.97	188.43	105.88	16.6
		7	50.13	149.55	87.32	16.78
1024	Bumps	3	103.29	114.55	109.67	170.4
		5	114.12	129.02	127.49	216.41
		7	129.9	141.96	141.13	240.17
1024	HeaviSine	3	58.65	164.05	85.13	20.41
		5	79.31	200.16	120.86	32.02
		7	64.21	152.21	97.38	41.17

Table 3.2: Sparsity, averaged over 100 simulations, of the various wavelet methods for the Wave, Peak, Bumps and HeaviSine functions, sampled at $n = 1024$ data points and using three values of SNR (3, 5 and 7).

the Department of Anaesthesia at the Bristol Royal Infirmary and measure the flow of air during breathing (see, e.g., Nason, 1996).

Figure 3.2 shows a section of the inductance plethysmography recording lasting approximately 80 seconds ($n = 4096$ signal points). The two main sets of regular oscillations correspond to normal breathing. The disturbed behavior in the center of the plot, where the normal breathing pattern disappears, corresponds to the patient vomiting. Figure 3.3 contains various reconstructions of the inductance plethysmography recording displayed in Figure 3.3, obtained by the competing wavelet estimators. The PostMean, PostMed and BF wavelet estimators were used with double-exponential prior and normal error models, as suggested by Johnstone and Silverman (2005, p. 1718) and Pensky & Sapatinas (2007, p. 618). All the prior parameters for these latter methods were estimated level-by-level by marginal maximum likelihood from the data. The level-wise MAP wavelet testimator was used with truncated geometric prior distributions, and the prior parameters were estimated separately at each level by conditional likelihood maximization, as described in Section 3.4.1. The various wavelet estimators were evaluated using Daubechies's compactly supported wavelets *Symmlet 8* (see Daubechies, 1992, p. 198) and *Coiflet 3* (see Daubechies, 1992, p. 258). For all methods, the primary resolution level was set equal to $j_0 = 4$.

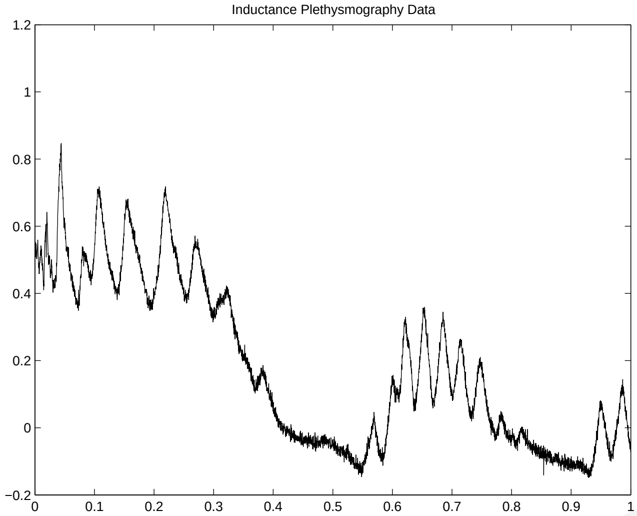

Figure 3.2: Section of the inductance plethysmograph recording lasting approximately 80 seconds ($n = 4096$ points).

As in Johnstone & Silverman (2005) and Pensky & Sapatinas (2007), we judged the efficacy of the various estimation methods in preserving peak heights simply by looking at the maximum of the various estimates, the height of the first peak in the inductance plethysmography curve. The numerical findings are displayed in Table 4.1. Similarly to Johnstone & Silverman (2005) and Pensky & Sapatinas (2007), we further quantified the efficacy of the various estimation methods in dealing with the rapid variation near the point 0.85 (on the x-axis) by the range of the estimated curves over a small interval at this point. The numerical findings are displayed in Table 4.2. Although we do not reproduce them here, similar results in magnitude are also true by increasing or decreasing the value of the primary resolution level j_0.

As observed in Tables 4.1 and 4.2, the level-wise MAP wavelet testimator and Bayes factor estimator are essentially the best among the competitors in preserving the peak height without any substantial cost of inferior treatment of presumably spurious variation elsewhere.

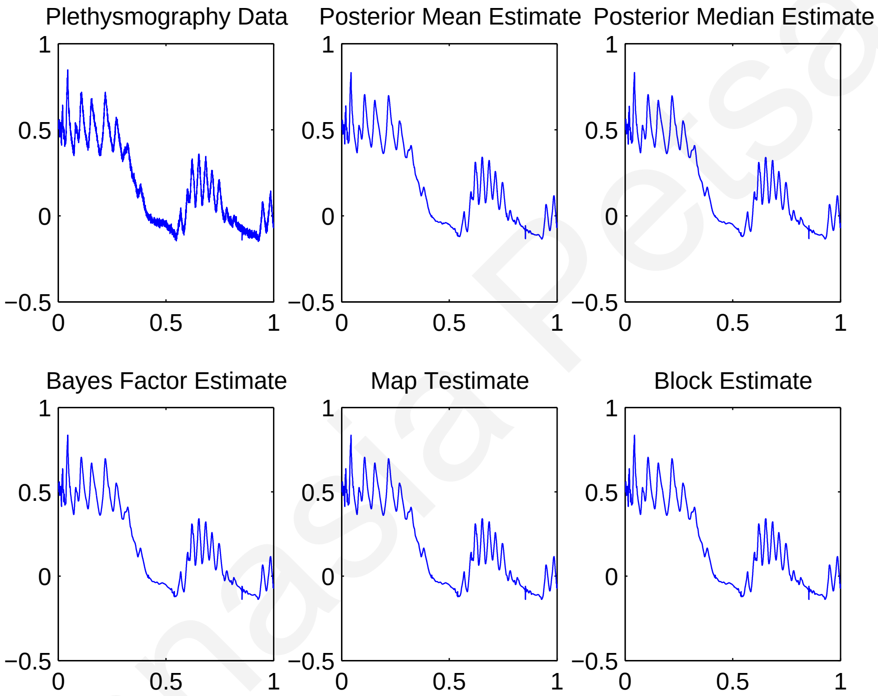

Figure 3.3: The various reconstructions of the inductance plethysmography recording displayed in Figure 3.2 Analysis of the inductance plethysmography recording using, from left to right, (top) PostMean and PostMed (bottom) BF, Level-wise MAP and Block.

Estimate	Data	PostMean	PostMed	BF	MAP	Block
Peak (*Symmlet 8*)	0.8472	0.8422	0.8424	0.8456	0.8458	0.8248
Peak (*Coiflet 3*)	0.8472	0.8311	0.8315	0.8351	0.8322	0.8231

Table 3.3: Height of the first peak of the inductance plethysmography data using PostMean, PostMed, BF, Level-wise MAP and Block estimators.

Estimate	Data	PostMean	PostMed	BF	MAP	Block
Range (*Symmlet 8*)	0.0879	0.0766	0.0768	0.0795	0.0796	0.0084
Range (*Coiflet 3*)	0.0879	0.0742	0.0741	0.0771	0.0766	0.0093

Table 3.4: Range of Spurious Variation at point 0.85 (on the x-axis) of the inductance plethysmography data using PostMean, PostMed, BF, Level-wise MAP and Block estimators.

3.5 Appendix: Proofs

3.5.1 Proof of Proposition 3.2.1

For the proof of Proposition 3.2.1 we need the following Lemma

Lemma 3.5.1.

$$\log \begin{pmatrix} n \\ \kappa \end{pmatrix} \leq c^{'} \kappa \log(\frac{n}{\kappa})$$

for $\kappa = 0, 1, \ldots, \frac{n}{e}$ *and some* $c^{'} \geq 2$.

Proof of Lemma 3.5.1

For $\kappa = 0$ the above inequality holds trivially. Using Stirling's formula, one gets

$$\begin{pmatrix} n \\ \kappa \end{pmatrix} \leq \left(\frac{n}{e}\right)^{n} \left(\frac{e}{n-\kappa}\right)^{n-\kappa} \left(\frac{e}{\kappa}\right)^{\kappa} = \left(\frac{n}{\kappa}\right)^{\kappa} \left(\frac{n}{n-\kappa}\right)^{n-\kappa}$$

for $n \geq 2$ and $\kappa = 1, 2, \ldots, n-1$.

Hence,

$$\begin{aligned} \log \begin{pmatrix} n \\ \kappa \end{pmatrix} &\leq \kappa \log\left(\frac{n}{\kappa}\right) + \left(n-\kappa\right) \log\left(\frac{n}{n-\kappa}\right) \\ &= \kappa \left[\log\left(\frac{n}{\kappa}\right) + \left(\frac{n}{\kappa} - 1\right) \log\left(\frac{\frac{n}{\kappa}}{\frac{n}{\kappa} - 1}\right)\right]. \end{aligned}$$

Let $t = \frac{n}{\kappa} - 1$ and $c \geq 1$. For $\kappa \leq \frac{n}{e}$, $t \geq e - 1$. Now, define

$$f(t) = c \log(t+1) - t \log(\frac{t+1}{t}).$$

In order to prove Lemma 3.5.1, we need to show that

$$\left(\frac{n}{\kappa} - 1\right) \log\left(\frac{\frac{n}{\kappa}}{\frac{n}{\kappa} - 1}\right) \leq c \log\left(\frac{n}{\kappa}\right) \quad \text{or equivalently} \quad f(t) \geq 0 \quad \text{for all} \quad t \geq e - 1.$$

Since

$$\log(1+x) \leq x \quad \text{for} \quad x \geq 0,$$

$$(e-1)\log(\frac{e}{e-1}) \leq 1 \leq c.$$

Hence,

$$f(e-1) = c\log(e) - (e-1)\log(\frac{e}{e-1}) \geq 0. \tag{3.11}$$

Now,

$$f'(t) = \frac{c+1}{t+1} - \log(1+\frac{1}{t}) \geq \frac{ct-1}{t(t+1)} > 0, \quad \text{for} \quad t \geq e-1 \quad \text{and} \quad c \geq 1. \tag{3.12}$$

Using (3.11) and (3.12), we get

$$f(t) \geq 0 \quad \text{for} \quad t \geq e-1.$$

$$\text{Hence,} \quad \log\begin{pmatrix} n \\ \kappa \end{pmatrix} \leq \kappa\Big[\log\Big(\frac{n}{\kappa}\Big) + c\log\Big(\frac{n}{\kappa}\Big)\Big], \quad \text{for} \quad c \geq 1 \quad \text{and} \quad \kappa = 1, 2, \ldots, \frac{n}{e}.$$

This completes the proof of Lemma 3.5.1.

We now are in the position to prove Proposition 3.2.1.

Case 1.

Under the condition $\pi_n(n) \geq e^{-\tilde{c}(\gamma)n}$, where $\tilde{c}(\gamma) \leq c_0$ and $c_0 > 0$, Definition 3.5 immediately implies $\|y-\hat{\mu}\|_2^2 \leq \|y-\hat{\mu}\|_2^2 + P_n(\hat{\kappa}) \leq P_n(n) \leq cn\sigma_n^2$. Thus,

$$\mathbb{E}\|\hat{\mu}-\mu\|_2^2 \leq 2\mathbb{E}\|\hat{\mu}-y\|_2^2 + 2\mathbb{E}\|y-\mu\|_2^2 = O(\sigma_n^2 n).$$

Case 2. We consider $2 \leq p \leq \infty$. We need to maximize $\sum_{i=1}^n {\mu_i}^2$ subject to $\|\mu\|_p = \sum_{i=1}^n |\mu_i|^p \leq \sigma_n^p n\eta_n^p$, if $0 < p < \infty$ and $\|\mu\|_\infty = \max_{1\leq i\leq n}\{|\mu_i|\} \leq \sigma_n\eta_n$. Using Lagrange multipliers, it is easy to see that the solution of the maximization problem is $\mu_1 = \mu_2 = \ldots = \mu_n = \sigma_n\eta_n$, for $0 < p < \infty$, yielding $\sum_{i=1}^n {\mu_i}^2 \leq \sigma_n^2 n\eta_n^2$. The case $p = \infty$ gives trivially the result. Under the assumption of Case 2 in Proposition 3.2.1 we have

$$\log(\pi_n^{-1}(0)) \leq c_1\frac{\log(n)}{n^\beta}.$$

Since γ_n is bounded, we apply Collorary 1 of Abramovich, Grinshtein & Pensky (2007) for $\kappa = 0$. The last term $c_1(\gamma_n)(1-\pi_n(0))\sigma_n^2$ in the upper error bound for the l^2-risk can be shown to be of order $O(\frac{\sigma_n^2\log n}{n^\beta})$. Thus, we get

$$\begin{aligned} \mathbb{E}||\hat{\mu}-\mu||_2^2 &\leq c_0(\gamma_n)\left(\sigma_n^2\eta_n^2 n + 2(1+1/\gamma_n)\sigma_n^2\log\pi_n^{-1}(0)\right) + c_1(\gamma_n)(1-\pi_n(0))\sigma_n^2 \\ &= O(\sigma_n^2 n\eta_n^2) + O(\sigma_n^2\frac{\log n}{n^\beta}). \end{aligned}$$

Case 3. Let

$$k_n^* = \lfloor n\eta_n^p[\log(\eta_n)^{-p}]^{-p/2}\rfloor + 1.$$

It is easy to see that $n\eta_n^p[\log(\eta_n^p)]^{-p/2} > [\frac{1}{1-\frac{p\log(\log(n))}{2\log(n)}}]^{p/2} \geq 1$.

Additionally, under the conditions of the proposition, $\eta_n^p \leq \alpha_n \leq e^{-c(\gamma_n)} < \frac{1}{e}$ and, therefore, $1 \leq k_n^* < ne^{-c(\gamma_n)}$.

Using Lemma 3.5.1, we have

$$\log\binom{n}{k_n^*} \leq (c+1)k_n^* \log(\frac{n}{k_n^*}), \tag{3.13}$$

where $c \geq 1$. Under assumption $\pi_n(\kappa) \geq (\frac{\kappa}{n})^{c_2 k}$ for $\kappa = 1, 2, \ldots, \alpha_n n$, we have

$$\log(\pi_n^{-1}(k_n^*)) \leq c_2 k_n^* \log(\frac{n}{k_n^*}). \tag{3.14}$$

It is easy to see that

$$ck_n^* \log(\frac{n}{k_n^*}) \leq cn\eta_n^p[\log(\eta_n^{-p})]^{1-\frac{p}{2}}. \tag{3.15}$$

Additionally,

$$\frac{k_n^*}{2}\log(1+\gamma) \leq cn\eta_n^p[\log(\eta_n^{-p})]^{1-\frac{p}{2}}, \tag{3.16}$$

since $\log(\eta_n{}^{-p}) \geq 1$. Now, we need to maximize $\sum_{i=k+1}^n \mu_{(i)}{}^2$ subject to

$$\sum_{i=1}^n |\mu_i|^p \leq n\eta_n^p\sigma_n^p \quad \text{for} \quad 0 < p < 2.$$

Define the least favorable sequence $\mu_i = \sigma_n\eta_n(n/i)^{1/p}, i = 1, \ldots, n$ that maximizes $\sum_{i=\kappa+1}^n \mu_i^2$ over $\mu \in m_p(\eta_n)$ for any $\kappa = 0, \ldots, n-1$. For $\kappa \geq 1$,

$$\sum_{i=\kappa+1}^n \mu_i^2 \leq \sigma_n^2\eta_n^2 n^{2/p}\int_\kappa^\infty x^{-2/p}dx = \frac{p}{2-p}\sigma_n^2\eta_n^2 n^{2/p}\kappa^{1-2/p},$$

while for $\kappa = 0$

$$\sum_{i=1}^n \mu_i^2 \leq \sigma_n^2\eta_n^2 n^{2/p}\zeta(2/p), \tag{3.17}$$

where $\zeta(\cdot) < \infty$ is the Riemann Zeta-function. Therefore,

$$\sum_{i=k_n^*+1}^n \mu_i^2 \leq c\sigma_n^2 n^{\frac{2}{p}}\eta_n^2(k_n^*)^{1-\frac{2}{p}} \leq c\sigma_n^2 n\eta_n^p[\log(\eta_n{}^{-p})]^{1-\frac{p}{2}}. \tag{3.18}$$

Thus, using Corollary 1 of Abramovich, Grinshtein & Pensky (2007) for $\kappa = k_n^*$, (3.13)-(3.18) and the boundeness of γ_n we arrive at

$$\begin{aligned}\mathbb{E}||\hat{\mu}-\mu||_2^2 &\leq c_0(\gamma_n)\left(\sum_{i=k_n^*+1}^{n}\mu_{(i)}^2 + 2(1+1/\gamma_n)\sigma_n^2\log\binom{n}{k_n^*}\right)\\ &+ 2c_0(\gamma_n)(1+1/\gamma_n)\sigma_n^2\left(\log\pi_n^{-1}(k_n^*) + \frac{k_n^*}{2}\log(1+\gamma_n)\right)\\ &+ c_1(\gamma_n)\sigma_n^2 = O(\sigma_n^2 n\eta_n^p(\log(\eta_n{}^{-p}))^{1-p/2}).\end{aligned}$$

Case 4.

We now consider the super-sparse case for $0 < p < 2$ and $\eta_n \leq \frac{\sqrt{2\log(n)}}{n^{\frac{1}{p}}}$. Under assumptions of Proposition 3.2.1 we have

$$\log(\pi_n^{-1}(0)) \leq c_1\frac{\log(n)}{n^\beta}. \tag{3.19}$$

Hence, using Collorary 1 of Abramovich, Grinshtein & Pensky (2007) for $\kappa = 0$, (3.17), (3.19) and the assumption of boundeness of γ_n, we get

$$\begin{aligned}\mathbb{E}||\hat{\mu}-\mu||_2^2 &\leq c_0(\gamma_n)\left(\sigma_n^2\eta_n^2 n^{2/p} + 2(1+1/\gamma_n)\sigma_n^2\log\pi_n^{-1}(0)\right) + c_1(\gamma_n)(1-\pi_n(0))\sigma_n^2\\ &= O(\sigma_n^2\frac{\log n}{n^\beta}) + O(\sigma_n^2 n^{\frac{2}{p}}\eta_n^2).\end{aligned}$$

This completes the proof of Proposition 3.2.1.

3.5.2 Proof of Proposition 3.2.2

Evidently, for any $\mu \in l_0[\eta_n]$, $\mu_{(i)} = 0$, $i > k^* = \lfloor n\eta_n\rfloor + 1$. Since $1 \leq k^* \leq ne^{-c(\gamma)}$, from the general upper bound for the risk established in Corollary 1 of Abramovich, Grinshtein & Pensky (2007) it follows that

$$E||\hat{\mu}^* - \mu||^2 \leq c_0(\gamma)2\sigma_n^2(1+1/\gamma)\left(\log\left\{\binom{n}{k^*}\pi_n^{-1}(k^*)\right\} + \frac{k^*}{2}\log(1+\gamma)\right) + c_1(\gamma)\sigma_n^2.$$

From the first part of Lemma 2 of Abramovich, Grinshtein & Pensky (2007)

$$\log\left\{\binom{n}{k^*}\pi_n^{-1}(k^*)\right\} \geq \log\binom{n}{k^*} \geq k^*\log\frac{\eta_n^{-1}}{2} \gg k^*\log(1+\gamma)$$

when $\eta_n \leq \frac{e^{-c(\gamma)}}{2}$. On the other hand, under the conditions of Proposition 3.2.2, Lemma 3.5.1 implies

$$\log\left\{\binom{n}{n\eta_n}\pi_n^{-1}(n\eta_n)\right\} \leq \tilde{c}n\eta_n\log\eta_n^{-1}$$

for sufficiently large n. Summarizing, one has

$$\mathbb{E}\|\hat{\mu}^* - \mu\|^2 \leq \tilde{c}_2(\gamma)\sigma_n^2 n\eta_n \log \eta_n^{-1}.$$

This completes the proof of Proposition 3.2.2.

3.5.3 Proof of Theorem 3.3.1

Let $R_j = \sum_{k=0}^{2^j-1} E(\hat{\theta}_{jk} - \theta_{jk})^2,\ j \geq j_0 - 1$ be a quadratic risk of a global wavelet MAP estimator on a j-th level. Due to the Parseval relation,

$$E||\hat{f}_N - f||^2 = \sum_{j \geq j_0 - 1} R_j.$$

Scaling coefficients are not thresholded and therefore

$$R_{j_0-1} = C_{j_0}\sigma^2 N^{-1} = o(N^{-2s/(2s+1)}).$$

On very high resolution levels, where $j \geq J$, all coefficients $\hat{\theta}_{jk}$ are set to zero and, therefore

$$\sum_{j=J}^{\infty} R_j = \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1} \theta_{jk}^2 = O(N^{-2s'}) = o(N^{-2s/(2s+1)}),$$

where $s' = s + 1/2 - 1/\min(p, 2)$ (e.g., Johnstone, 2002, Proposition 15.4).

Consider now $\sum_{j=j_0}^{J-1} R_j$. The set of wavelet coefficients $\{\theta_{jk},\ j = j_0, ..., J-1\}$ of a function $f \in B_{p,q}^s(M)$ lies within a weak l_q-ball of a radius aM with $q = 2/(2s+1)$, where the constant a depends only on a chosen wavelet basis: $m_q[\eta_N] = \{\theta : |\theta|_{(i)} \leq (aM)i^{-1/q}\}$ (e.g., Donoho & Johnstone, 1996). The corresponding standardized radius $\eta_N = (\sigma/\sqrt{N})^{-1}\tilde{N}^{-1/q}aM = O(N^{-s})$, where $\tilde{N} = N - 2^{j_0} \sim N$ for large N.

Under the conditions of the theorem, one can apply Theorem 6 of Abramovich, Grinshtein & Pensky (2007) for $m_q[\eta_N]$ to get

$$\sum_{j=j_0}^{J-1} R_j \leq \sup_{\theta \in m_q[\eta_N]} \mathbb{E}||\hat{\theta} - \theta||_2^2 = O\left(\eta_N^q (2\log\eta_N^{-q})^{1-q/2}\right) = O\left(\left(\frac{\log N}{N}\right)^{2s/(2s+1)}\right).$$

This completes the proof of Theorem 3.3.1.

3.5.4 Proof of Theorem 3.3.2

Let

$$\begin{aligned} R_N(\theta) &= \sum_{j=j_0-1}^{j_1} \mathbb{E}(\|\hat{\theta}_j - \theta_j\|^2) + \sum_{j=j_1+1}^{J-1} \mathbb{E}(\|\hat{\theta}_j - \theta_j\|^2) + \sum_{j=J}^{\infty} \mathbb{E}(\|\hat{\theta}_j - \theta_j\|^2) \\ &= R_{lo} + R_{mid} + R_{hi}. \end{aligned}$$

Let $R_j = \sum_{k=0}^{2^j-1} E(\hat{\theta}_{jk} - \theta_{jk})^2,\ j \geq j_0 - 1$ be now a quadratic risk of a level-wise wavelet MAP estimator on a j-th level. For $f \in B^s_{p,q}(M)$, the sequence of its wavelet coefficients on a j-th level belongs to an l_p-ball of a standardized radius $\eta_j = CN^{1/2}2^{-j(s+1/2)}$ for some $C > 0$. Define j_1 as the largest integer satisfying $\eta_j^p \geq c_\alpha$ and note that $\eta_j^p \geq c_\alpha$ for $j \leq j_1$ and $\eta_j^p < c_\alpha$ for $j > j_1$.

Applying the first statement of Proposition 3.2.1 for each level with $n_j = 2^j$ and $\sigma_N^2 = \frac{\sigma^2}{N}$ we have

<u>Risk of scaling coefficients and Risk at low levels</u>

$$\begin{aligned} R_{j_0-1} &= \mathbb{E}\|\hat{\theta}_{j_0-1} - \theta_{j_0-1}\|_2^2 \leq 2^{j_0}\frac{\sigma^2}{N} \\ R_{lo} &= \sum_{j=j_0}^{j_1} \mathbb{E}(\|\hat{\theta}_j - \theta_j\|_2^2) = \sum_{j=j_0}^{j_1} \frac{c2^j}{N} \leq c\frac{2^{j_1}}{N} \leq \frac{cC^{2/(2s+1)}}{N^{2s/(2s+1)}}. \end{aligned}$$

<u>Risk at medium levels</u>

If $j_1 \geq J-1$, $R_{mid} = 0$. Otherwise,

$$R_{mid} \leq C\{S_1 + S_2\},$$

where $S_2 = \sum_{j=j_1+1}^{J-1}\{\frac{\log(2^j)}{N} + \frac{1}{N}\}$ and S_1 will be defined separately for $0 < p < 2$ and $p \geq 2$.

Case 1. $0 < p < 2$. Define j_2 as the largest integer for which $\eta_j^p \geq 2^{-j}(2\log 2^j)^{p/2}$. Let $S_1 = S_{11} + S_{12}$. We first consider the case $j_1 < j_2 < J-1$. Using the monotonicity arguments, $\eta_j^p \geq 2^{-j}(2\log 2^j)^{p/2}$ for all $j_1 < j \leq j_2$. For $0 < p < 2$

$$S_{11} = \sum_{j=j_1+1}^{j_2} C^p 2^{-jp(s+1/2-1/p)}\epsilon_1^{2-p},$$

where $\epsilon_1 = N^{-1/2}\sqrt{\log(C^{-p}N^{-p/2}2^{jp(s+1/2)})}$.

It is easy to see that

$$\frac{2s-(s+1/2-1/p)p}{2s+1}=1-p/2.$$

For $j>j_1$, $2^{-j}<c(C^2N)^{-1/(2s+1)} \Rightarrow 2^{-j(2s-(s+1/2-1/p)p)}<c(C^2N)^{-1+p/2}$ which leads to

$$\begin{aligned}S_{11} &\leq c\sum_{j=j_1+1}^{j_2} C^pN^{-1+\frac{p}{2}}2^{-p(s+1/2-1/p)j}\{\log(2^{(2s+1)j}N^{-1}C^{-2})\}^{1-p/2}\\ &\leq cC^pN^{-1+p/2}(C^2N)^{-\frac{p(s+1/2-1/p)}{2s+1}}=cC^{2/(2s+1)}N^{-\frac{2s}{2s+1}}.\end{aligned}$$

Additionally, applying the fourth statement of Proposition 3.2.1

$$S_{12}\leq\sum_{j=j_2+1}^{J-1}\mathbb{E}\|\hat{\theta}_j-\theta_j\|_2^2\leq c\sum_{j=j_2+1}^{J-1}\frac{j}{N}\leq cJ^2N^{-1}\leq c\frac{\log_2^2(N)}{N}\leq cN^{\frac{-2s}{2s+1}}.$$

Now, if $j_1<J-1\leq j_2$, $S_{12}=0$ and

$$\begin{aligned}S_{11} &\leq c\sum_{j=j_1+1}^{J-1} C^pN^{-1+\frac{p}{2}}2^{-p(s+1/2-1/p)j}\{\log(2^{(2s+1)j}N^{-1}C^{-2})\}^{1-p/2}\\ &\leq cC^{2/(2s+1)}N^{-\frac{2s}{2s+1}}.\end{aligned}$$

Hence,

$$S_1\leq S_{11}+S_{12}\leq cN^{-\frac{2s}{2s+1}}. \tag{3.20}$$

Case 2. For $2\leq p\leq\infty$

$$S_1=c\sum_{j=j_1+1}^{J-1}C^22^{j(1-2/p)}2^{-2j(s+1/2-1/p)}.$$

Since $1-2/p-2(s+1/2-1/p)=-2s$,

$$\begin{aligned}S_1=c\sum_{j=j_1+1}^{J-1}C^22^{-2sj} &= cC^22^{-2s(j_1+1)}\frac{(1-2^{-2s(J-j_1-1)})}{1-2^{-2s}}\\ &\leq cC^{2/(2s+1)}N^{-2s/(2s+1)}. \end{aligned} \tag{3.21}$$

Additionally, for $0<p\leq\infty$

$$S_2=c\sum_{j=j_1+1}^{J-1}\{\frac{\log(2^j)}{N}+\frac{1}{N}\}\leq\sum_{j=j_1+1}^{J-1}c\{\frac{j}{N}+\frac{1}{N}\}\leq c\{\frac{J}{N}+\frac{J^2}{N}\}\leq\frac{c\log^2(N)}{N}. \tag{3.22}$$

Hence, using (3.20)-(3.22),

$$R_{mid}\leq c\{\frac{C^{\frac{2}{2s+1}}}{N^{\frac{2s}{2s+1}}}+\frac{\log^2(N)}{N}\}.$$

Risk at very high levels

Let $\Delta = (1/2 - 1/p)_+$, $r'' = s - (1/p - 1/2)_+$. Hence, $r'' = s - 1/p + 1/2$ for $0 < p < 2$ and $r'' = s$ for $2 \leq p \leq \infty$. Therefore, we can write $r'' = s + 1/2 - 1/p - \Delta$. The following inequality holds as a consequence of the equivalence of norms in finite dimensional spaces.

$$\|\theta_j\|_2 \leq 2^{\Delta j}\|\theta_j\|_p \quad \text{for any} \quad p.$$

Since $r'' = s$ for $2 \leq p \leq \infty$ and $r'' = s + 1/2 - 1/p \geq 1/2$ for $0 < p < 2$, we obtain

$$\begin{aligned} R_{hi} &= \sum_{j=J-1}^{\infty} \mathbb{E}(\|\theta_j\|_2^2) = \sum_{j=J}^{\infty} 2^{2j\Delta}\|\theta_j\|_p^2 \leq cC^2 \sum_{j=J}^{\infty} 2^{-2j(s+1/2-1/p)} 2^{2j\Delta} \\ &= cC^2 2^{-2J(s+1/2-1/p-\Delta)} \leq cC^2 N^{-2r''} = o(\frac{1}{N^{2s/(2s+1)}}). \end{aligned}$$

Therefore, we arrive at the desired result. This completes the proof of Theorem 3.3.2.

3.5.5 Proof of Theorem 3.3.3

Let $R_j = \sum_{k=0}^{2^j-1} E(\hat{\theta}_{jk} - \theta_{jk})^2$, $j \geq j_0 - 1$ be now a quadratic risk of a level-wise wavelet MAP estimator on a j-th level. Johnstone & Silverman (2005, Section 5.6) showed that $E||\hat{f}_N^{(m)} - f^{(m)}||^2 \asymp \sum_{j \geq j_0 - 1} 2^{2mj} R_j$.

For $f \in B_{p,q}^s(M)$, the sequence of its wavelet coefficients on a j-th level belongs to an l_p-ball of a standardized radius $\eta_j = C_0 N^{1/2} 2^{-j(s+1/2)}$ for some $C_0 > 0$ (e.g, Meyer, 1992, Section 6.10). Define j_1 as the largest integer satisfying $\eta_j^p \geq c_\alpha$ and note that $\eta_j^p \geq c_\alpha$ for $j \leq j_1$ and $\eta_j^p < c_\alpha$ for $j > j_1$ (with obvious modifications for $p = \infty$). Consider the following cases.

1. Scaling coefficients, $j = j_0 - 1$. Similarly to the global wavelet estimator, for a fixed j_0, $2^{2m(j_0-1)} R_{j_0-1} = O(N^{-1}) = o(N^{-2(s-m)/(2s+1)})$ as $n \to \infty$.

2. Coarse levels, $j_0 \leq j \leq j_1$. Applying the first statement of Proposition 3.2.1 for each level one has

$$\sum_{j=j_0}^{j_1} 2^{2mj} R_j \leq C \sum_{j=j_0}^{j_1} 2^{2mj} N^{-1}\sigma^2 n_j \leq CN^{-1} \sum_{j=j_0}^{j_1} 2^{(2m+1)j} = O(N^{-2(s-m)/(2s+1)}) \quad \text{as} \quad n \to \infty$$

3. Mid+high levels, $j_1 < j < J$. Consider separately a) $2 \le p \le \infty$ and b) $0 < p < 2$.

a) $2 \le p \le \infty$. Under the conditions of the theorem, the second statement of Proposition 3.2.1 on a j-th level yields

$$R_j \le CN^{-1}(n_j\eta_j^2 + n_j^{-\beta}\log n_j) \le C(2^{-2js} + N^{-1}2^{-\beta j}j)$$

and, hence,

$$\begin{aligned}\sum_{j=j_1+1}^{J-1} 2^{2mj}R_j &\le C(2^{-2j_1(s-m)} + N^{-1}J^2) \le C(N^{-2(s-m)/(2s+1)} + N^{-1}\log_2^2 N)\\ &= O(N^{-2(s-m)/(2s+1)}) \quad \text{as} \quad n \to \infty.\end{aligned}$$

b) $0 < p < 2$. Let j_2 be the largest integer for which $\eta_j^p \ge n_j^{-1}(2\log n_j)^{p/2}$. One can easily verify that $j_1 < j_2 < J$.

Using the monotonicity arguments, $\eta_j^p \ge n_j^{-1}(2\log n_j)^{p/2}$ for all $j_1 < j \le j_2$ (mid-levels). One can then apply the third statement of Proposition 3.2.1 and after some algebra to get for $0 \le m < (s+1/2-1/p)p/2$

$$\begin{aligned}\sum_{j=j_1+1}^{j_2} 2^{2mj}R_j &\le CN^{-1}\sum_{j=j_1+1}^{j_2} 2^{(2m+1)j}N^{p/2}2^{-jp(s+1/2)}\left(\log(N^{-p/2}2^{jp(s+1/2)})\right)^{1-p/2}\\ &\le CN^{-(1-p/2)}2^{-j_1p(s+1/2-(2m+1)/p)}\log(N^{-p/2}2^{j_1p(s+1/2)})\\ &= O(N^{-2(s-m)/(2s+1)}) \quad \text{as} \quad n \to \infty\end{aligned}$$

On high levels $j_2 < j < J$, $\eta_j^p < n_j^{-1}(2\log n_j)^{p/2}$ and the fourth statement of Proposition 3.2.1 implies

$$R_j \le C(2^{-2j(s+1/2-1/p)} + N^{-1}2^{-j\beta}j).$$

Hence, for $0 \le m \le \beta/2$ and $m < \min(s, (s+1/2-1/p)p/2)$, one has

$$\sum_{j=j_2+1}^{J-1} 2^{2mj}R_j \le C(2^{-2(j_2+1)(s+1/2-1/p-m)} + N^{-1}J^2) = S_1 + S_2,$$

where, evidently, $S_2 = O(N^{-1}\log_2^2 N) = o(N^{-2(s-m)/(2s+1)})$ as $n \to \infty$. From the definition of j_2, $2^{(j_2+1)(s+1/2-1/p)} > \sqrt{NC/(j_2+1)} > \sqrt{NC/\log_2 N}$ that after some algebra yields $S_1 = o(N^{-2(s-m)/(2s+1)})$ as $n \to \infty$.

4. Very high levels, $j \geq J$. Using the results of Johnstone & Silverman (2005), the tailed sum

$$\sum_{j \geq J} 2^{2mj} R_j = O(N^{-2(s'-m)}) = o(N^{-2(s-m)/(2s+1)}),$$

where $s' = s + 1/2 - 1/\min(p, 2)$.

Summarizing, $\sum_{j \geq j_0 - 1} 2^{2mj} R_j = O(N^{-2(s-m)/(2s+1)})$ as $n \to \infty$.

This completes the proof of Theorem 3.3.3.

3.5.6 Proof of Theorem 3.3.4

Set $\tilde{a} = s - (\frac{1}{p} - 1)_+ > \frac{1}{2}$. Using Proposition 5 of Johnstone & Silverman (2004a) we obtain

$$2^{j(s+\frac{1}{2}-\frac{1}{p})} \|\theta_j - \tilde{\theta}_j\|_p \leq cC2^{-\tilde{a}(J-j)},$$

for all j such that $j_0 - 1 \leq j < J$ and for all $p > 0$.

Hence,

$$\|\tilde{\theta}_j\|_p \leq \|\tilde{\theta}_j - \theta_j\|_p + \|\theta_j\|_p \leq cC2^{-\tilde{a}(J-j)} 2^{-j(s+\frac{1}{2}-\frac{1}{p})} + C2^{-j(s+\frac{1}{2}-\frac{1}{p})} = cC2^{-j(s+\frac{1}{2}-\frac{1}{p})} \quad (3.23)$$

Using the above inequality, it is easy to see that

$$\|\tilde{\theta}_j^I\|_p \leq cC2^{-j(s+\frac{1}{2})}(2^j - 2S - 2)^{\frac{1}{p}}. \quad (3.24)$$

Therefore, the interior 'discretized' coefficients obey (up to a constant) the same Besov sequence bounds as the true interior coefficients. $\tilde{Y}_{jk}$ each has expected value $\tilde{\theta}_{jk}$ and the interior coefficients are independent normals with variance $\frac{\sigma^2}{N}$. Because of the bound (3.24), using the same arguments as in Theorem 3.3.2, we arrive at

$$\sum_{j=j_0}^{J-1} \mathbb{E}(\|\hat{\theta}_j^I - \tilde{\theta}_j^I\|^2) + \sum_{j=J}^{\infty} \|\theta_j\|_2^2 \leq c\{\frac{C^{2/(2s+1)}}{N^{2s/(2s+1)}} + o(\frac{1}{N^{2s/(2s+1)}})\}. \quad (3.25)$$

Coarse scale error

Consider first the coarse level scaling coefficient θ_{j_0-1}. $\hat{\theta}_{j_0-1} = \tilde{Y}_{j_0-1}$ has variance bounded by $\frac{\sigma^2 c_A}{N}$.

Therefore,

$$\mathbb{E}(\|\hat{\theta}_{j_0-1} - \tilde{\theta}_{j_0-1}\|_2^2) \leq \frac{2^{j_0} c_A \sigma^2}{N} \leq \frac{c}{N}. \tag{3.26}$$

Boundary coefficients

The elements of the array $\tilde{Y}^B$ are normally distributed with expected values $\tilde{\theta}^B$ and variances bounded by $\frac{\sigma^2 c_A}{N}$. We obtain $\hat{\theta}^B_{jk}$ by individually thresholding $\tilde{Y}^B_{jk}$ with threshold $\tau(\frac{j}{N})^{\frac{1}{2}}$. So, by standard properties of 2-norm and thresholding we get

$$\begin{aligned}
\mathbb{E}(|\hat{\theta}_{jk} - \tilde{\theta}_{jk}|^2) &\leq c\mathbb{E}(|\hat{\theta}_{jk} - \tilde{Y}_{jk}|^2) + c\mathbb{E}(|\tilde{Y}_{jk} - \tilde{\theta}_{jk}|^2) \leq \\
&\leq c\{\frac{j}{N} + \frac{\sigma^2 c_A}{N}\} \leq c\frac{j}{N}. && (3.27)
\end{aligned}$$

$$\begin{aligned}
R_{lo+high} &= \sum_{j=j_0}^{J-1} \mathbb{E}(\|\hat{\theta}^B_j - \tilde{\theta}^B_j\|_2^2) \leq \sum_{j=j_0}^{J-1} K^B_j c\frac{j}{N} \\
&\leq \frac{cJ^2}{N} \leq c\frac{(\log N)^2}{N}, && (3.28)
\end{aligned}$$

since $|K^B_j| = 2(S-1)$ for any level j.

Discretization bias

Let $\Delta = (\frac{1}{2} - \frac{1}{p})_+$ and $r^{''} = s - (\frac{1}{p} - \frac{1}{2})_+$. For $2 \leq p \leq \infty$, $\frac{1}{p} - \frac{1}{2} < 0$, $\Longrightarrow r^{''} = s$. For $0 < p < 2$, $r^{''} = s - \frac{1}{p} + \frac{1}{2}$. Hence, for any $0 < p \leq \infty$,

$$r^{''} = s - \frac{1}{p} + \frac{1}{2} - \Delta.$$

Using the equivalence of norms in finite-dimensional spaces, i.e

$$\|\tilde{\theta}_j - \theta_j\|_2 = 2^{j\Delta}\|\tilde{\theta}_j - \theta_j\|_p,$$

and inequality (3.23), we have

$$\begin{aligned}
R_D &= \sum_{j=j_0-1}^{J-1} \|\tilde{\theta}_j - \theta_j\|_2^2 = \sum_{j=j_0-1}^{J-1} 2^{2\Delta j}\|\tilde{\theta}_j - \theta_j\|_p^2 \leq \sum_{j=j_0-1}^{J-1} C^2 2^{2\Delta j} 2^{-2\tilde{a}(J-j)} 2^{-2j(s+\frac{1}{2}-\frac{1}{p})} \\
&= cC^2 2^{-2\tilde{\alpha}J} \sum_{j=j_0-1}^{J-1} 2^{2j(\tilde{\alpha}-r^{''})}.
\end{aligned}$$

- If $\tilde{a} = r^{''}$, $R_D \leq cC^2 2^{-2\tilde{a}J} J$.

- If $\tilde{a} < r^{''}$,

$$\begin{aligned} R_D &\leq cC^2 2^{-2\tilde{a}J} \sum_{j=j_0-1}^{J-1} 2^{2j(\tilde{a}-r^{''})} = cC^2 2^{-2\tilde{a}J} 2^{(j_0-1)(\tilde{a}-r^{''})} [\frac{1-2^{2(\tilde{a}-r^{''})(J-j_0+1)}}{1-2^{2(\tilde{a}-r^{''})}}] \\ &\leq cC^2 2^{-2\tilde{a}J}. \end{aligned}$$

Therefore,

$$R_D \leq cC^2 J^{\lambda^{'}} 2^{-2\tilde{a}J} \quad \text{for} \quad a^{'} \leq r^{''},$$

with $\lambda^{'} = 1$ if and only if $\tilde{a} = r^{''}$ and $\lambda^{'} = 0$ otherwise.

- If $\tilde{a} > r^{''}$,

$$\begin{aligned} R_D &= cC^2 2^{-2\tilde{a}J} \sum_{j=j_0-1}^{J-1} 2^{2j(\tilde{a}-r^{''})} = cC^2 2^{-2\tilde{a}J} 2^{2(\tilde{a}-r^{''})(j_0-1)} [2^{2(\tilde{a}-r^{''})(J-j_0+1)} - 1] \leq \\ &\leq cC^2 2^{-2\tilde{a}J} 2^{2(\tilde{a}-r^{''})J} \leq cC^2 2^{-2Jr^{''}} = \frac{cC^2}{N^{2r^{''}}}. \end{aligned}$$

Let $r^{'''} = \min\{r^{''}, \tilde{a}\}$. Hence,

$$\sum_{j=j_0-1}^{J-1} \|\tilde{\theta}_j - \theta_j\|_2^2 \leq cC^2 N^{-2r^{'''}} (\log N)^{\lambda^{'}}, \tag{3.29}$$

with $\lambda^{'} = 1$ if and only if $\tilde{a} = r^{''}$ and $\lambda^{'} = 0$ otherwise. Using inequalities (3.25)-(3.29), we have

$$\begin{aligned} R_N^*(f) &\leq cC^2 N^{-2r^{'''}} (\log N)^{\lambda^{'}} + c\frac{(\log N)^2}{N} + c\{C^{2/(2s+1)} N^{-2s/(2s+1)} \\ &+ o(\frac{1}{N^{2s/(2s+1)}})\}. \end{aligned} \tag{3.30}$$

- For $0 < p < 2$, $r^{''} = s - \frac{1}{p} + \frac{1}{2} \Longrightarrow r^{''} \neq \tilde{a}, \Longrightarrow \lambda^{'} = 0$.

- For $2 \leq p \leq \infty$, $\tilde{a} = s$ and $r^{''} = s \Longrightarrow r^{''} = \tilde{a}$ and $\tilde{a} = s < s + \frac{1}{2} - \frac{1}{p} \Longrightarrow \lambda^{'} = 1$.

1st Case

For $0 < p < 2$, the first term of (3.30) takes the form $cC^2 N^{-2r^{'''}}$.

- If $r^{'''} = \tilde{a} = s$, $cC^2 N^{-2r^{'''}} = cC^2 N^{-2s} = o(\frac{1}{N^{2s/(2s+1)}})$.

- If $r''' = r''$, $cC^2N^{-2r'''} = cC^2N^{-2r'''} \leq cC^2N^{-2(s+\frac{1}{2}-\frac{1}{p})} \leq \frac{cC^2}{N}$.

2nd Case

For $2 \leq p \leq \infty$, the first term of (3.30) takes the form $cC^2(\log N)N^{-2r'''}$.

- If $r''' = \tilde{a} = s$, $cC^2N^{-2r'''}(\log N)^{\lambda'} \leq \frac{cC^2 \log N}{N^{2s}} = o(\frac{1}{N^{2s/(2s+1)}})$.
- If $r''' = r''$, $cC^2N^{-2r'''}(\log N)^{\lambda'} \leq \frac{cC^2 \log(N)}{N^{2r''}} = \frac{cC^2 \log(N)}{N^{2s}} = o(\frac{1}{N^{2s/(2s+1)}})$.

Hence, we arrive at

$$R_N^*(f) \leq c\{\frac{C^{2s/(2s+1)}}{N^{2s/(2s+1)}} + o(\frac{1}{N^{2s/(2s+1)}})\}.$$

This completes the proof of Theorem 3.3.4.

Chapter 4

Minimax convergence rates under the L_p-risk in the functional deconvolution model

We derive minimax results in the functional deconvolution model under the L^p-risk, $1 \leq p < \infty$. Lower bounds are given when the unknown response function is assumed to belong to a Besov ball and under appropriate smoothness assumptions on the blurring function, including both regular-smooth and super-smooth convolutions. Furthermore, we investigate the asymptotic minimax properties of an adaptive wavelet estimator over a wide range of Besov balls. The new findings extend recently obtained results under the L^2-risk. As an illustration, we discuss particular examples for both continuous and discrete settings. Additionally, we show that when the number of channels tends to infinity, functional deconvolution with a box-car type blurring function in the discrete model can provide estimators with the same asymptotical minimax rates of convergence for L^p-risk as in the continuous model. A small simulation study shows that the proposed estimator performs well in finite sample situations.

4.1 Introduction

In the past decades, the standard deconvolution model was studied by many researchers who tried to find optimal solutions to this problem. Amongst them, Donoho (1995), Abramovich & Silverman (1998), Jonhstone, Kerkyacharian, Picard & Raimondo (2004) and Chenseau (2008) proposed various wavelet thresholding estimators of the unknown response function in this model that achieve optimal (in the minimax or the maxiset sense), or near-optimal within a logarithmic factor, convergence rates over a wide range of Besov balls and for a range of L^p-loss functions defining the risk.

On the one hand, there are several cases when one needs to recover initial or boundary conditions on the basis of observations of a noisy solution of a partial differential equation. The estimation problem of the initial condition in the heat conductivity equation was initiated by Lattes & Lions (1967). This problem and the problem of recovering the boundary condition for elliptic equations based on observations in an internal domain were considered in a minimax setting by Golubev & Khasminskii (1999), and sharp asymptotics for the L^2-risk over a range of Sobolev balls were obtained. On the other hand, Casey & Walnut (1994) and De Canditiis & Pensky (2004, 2006) considered the multichannel deconvolution model which arises in signal and image processing, e.g., in LIDAR (Light Detection and Ranging) remote sensing and reconstructions of blurred images (see, e.g., Park, Dho & Kong (1997)). Using the maxiset approach, De Canditiis & Pensky (2006) derived upper bounds for the L^p-risk, $1 < p < \infty$, over a wide range of Besov balls, of an adaptive term-by-term thresholding wavelet estimator for a *fixed* target function $f(\cdot)$. However, the minimax properties of their estimator and the case when the number of channels increases with the number of points at which $f(\cdot)$ is observed were not considered by De Canditiis & Pensky (2006).

Recently, Pensky & Sapatinas (2009a) showed that all the above described problems are special cases of the functional deconvolution model given by

$$y(u,t) = \int_T f(x)g(u,t-x)dx + \frac{1}{\sqrt{n}}z(u,t), \quad t \in T = [0,1], \quad u \in U = [a,b], \tag{4.1}$$

with $-\infty < a \leq b < \infty$. Here, the kernel or blurring function $g(\cdot,\cdot)$ is assumed to be known, and $z(u,t)$ is assumed to be a two-dimensional Gaussian white noise, i.e., a generalized two-dimensional Gaussian field with covariance function

$$\mathbb{E}(z(u_1,t_1)z(u_2,t_2)) = \delta(u_1-u_2)\delta(t_1-t_2),$$

where $\delta(\cdot)$ denotes the Dirac δ-function. The analogous discrete model, when $y(u,t)$ is observed at $n = NM$ points (u_l,t_i), $l = 1,2,\ldots,M$ and $i = 1,2,\ldots,N$, is given by

$$y(u_l,t_i) = \int_T f(x)g(u_l,t_i-x)dx + \epsilon_{li}, \quad t_i = \frac{i}{N} \in T = [0,1], \quad u_l \in U = [a,b], \tag{4.2}$$

where ϵ_{li} are standard Gaussian random variables, independent for different l and i.

Pensky & Sapatinas (2009a) obtained minimax lower bounds and proposed an adaptive (linear or block thresholding) wavelet estimator, for both the functional deconvolution model (4.1) and its discrete version (4.2), that is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, under the L^2-risk over a wide range of Besov balls.

The aim of this chapter is to provide the analogous statements of the above mentioned minimax results obtained by Pensky & Sapatinas (2009a) under the L^2-risk for the case of L^p-risk, $1 \leq p < \infty$. More specifically, we first obtain lower bounds for the L^p-risk, $1 \leq p < \infty$, when the unknown response function $f(\cdot)$ in functional deconvolution model (4.1) and its discrete version (4.2) are assumed to belong to a Besov ball and the blurring function $g(\cdot, \cdot)$ is assumed to possess some smoothness properties, including both regular-smooth and super-smooth convolutions. Furthermore, we investigate the asymptotic optimal (in the minimax sense) properties of an adaptive (linear or block thresholding) wavelet estimator under the L^p-risk, $1 \leq p < \infty$, over a wide range of Besov balls. As an illustration, we discuss particular examples for both continuous and discrete settings.

The results under the L^p-risk are also extended to the multichannel deconvolution model with box-car convolutions. Moreover, we verify the practical importance of block thresholding wavelet estimators in the functional deconvolution model by conducting a simulation study.

In what follows, as in Pensky & Sapatinas (2009a), we assume that both $f(\cdot)$ and, for a fixed $u \in [a, b]$, $g(u, \cdot)$ are periodic functions with period on the unit interval $T = [0, 1]$; this assumption appears naturally in the above mentioned special models which (4.1) and (4.2) generalize.

4.2 Meyer wavelets and Besov balls

Let $\phi^*(\cdot)$ and $\psi^*(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively (see, e.g., Meyer (1992), Chapter 3). As usual,

$$\phi^*_{jk}(x) = 2^{j/2}\phi^*(2^j x - k), \quad \psi^*_{jk}(x) = 2^{j/2}\psi^*(2^j x - k), \quad j, k \in \mathbb{Z},$$

are, respectively, the dilated and translated Meyer scaling and wavelet functions at resolution level j and scale position $k/2^j$. (Here, and in what follows, $\mathbb{Z}$ refers to the set of integers.) Similarly to Section 2.3 in Johnstone, Kerkyacharian, Picard & Raimondo (2004), we obtain a periodized version of Meyer wavelet basis by periodizing the basis functions $\{\phi^*(\cdot), \psi^*(\cdot)\}$, i.e.,

$$\phi_{jk}(x) = \sum_{i\in\mathbb{Z}} 2^{j/2}\phi^*(2^j(x+i)-k), \quad \psi_{jk}(x) = \sum_{i\in\mathbb{Z}} 2^{j/2}\psi^*(2^j(x+i)-k).$$

Note that, for any $j_0 \geq 0$ and any $j \geq j_0$, any $f \in L^2(T)$ can be written as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(t),$$

where $\alpha_{j_0k} = \int_T f(t)\phi_{j_0k}(t)$ and $\beta_{jk} = \int_T f(t)\psi_{jk}(t)$. It is well known that the Meyer wavelet basis satisfies the following three properties (see, e.g., Johnstone, Kerkyacharian, Picard & Raimondo, 2004):

1. **Property of concentration** Let $p \in [1,\infty)$ and $h \in \{\phi,\psi\}$. For any integer $j \in \{\tau,\ldots,\infty\}$ and any sequence $u = (u_{j,k})_{j,k}$, there exists a constant $c > 0$ such that

$$\|\sum_{k=0}^{2^j-1} u_{j,k}h_{j,k}\|_p^p \leq c2^{j(\frac{p}{2}-1)}\sum_{k=0}^{2^j-1}|u_{j,k}|^p. \tag{4.3}$$

2. **Property of unconditionality.** Let $p \in (1,\infty)$. Let us set $\psi_{\tau-1,k} = \phi_{\tau,k}$. For any sequence $u = (u_{jk})_{j,k}$, we have

$$\|\sum_{j=\tau-1}^{\infty}\sum_{k=0}^{2^j-1} u_{jk}\psi_{jk}\|_p^p \asymp \|(\sum_{j=\tau-1}^{\infty}\sum_{k=0}^{2^j-1}|u_{jk}\psi_{jk}|^2)^{\frac{1}{2}}\|_p^p.$$

(Here, and in what follows, the notation $a \asymp b$ means there exist two positive constants c_1 and c_2 such that $c_1 b \leq a \leq c_2 b$.)

3. **Temlyakov property**. Let $\sigma \in [0,\infty)$. Let $\psi_{\tau-1,k} = \phi_{\tau,k}$. For any subset $A \subseteq \{\tau-1,\ldots,\infty\}$ and for any subset $\Omega \subseteq \{0,\ldots,2^j-1\}$, we have

$$\|(\sum_{j\in A}\sum_{k\in\Omega}|2^{\sigma_j}\psi_{jk}|^2)^{\frac{1}{2}}\|_p^p \asymp \sum_{j\in A}\sum_{k\in\Omega} 2^{jp\sigma}\|\psi_{jk}\|_p^p.$$

Remark 4.2.1. The property of concentration is used in the proof of Theorem 4.4.2, in the case of super-smooth convolutions. The property of unconditionality and Temlyakov property are indirectly used in the proof of Theorem 4.4.2, since they are used in the proofs of some auxiliary results (i.e., Theorems 5.4.1 and 5.4.2 in Chesneau, 2006).

Now, let us give the definition of Besov balls, the main function spaces used in our study. Let $M \in (0, \infty)$, $s \in (0, R)$, $\rho \in [1, \infty]$ and $r \in [1, \infty]$. (Here, R refers to the number of vanishing moments and continuous derivatives of the mother wavelet function $\psi^*(\cdot)$; note that, for the Meyer wavelet basis, $R = \infty$.) Let $\beta_{\tau-1,k} = \alpha_{\tau,k}$. We say that a function f belongs to the Besov ball $B^s_{\rho,r}(M)$ if and only if the associated wavelet coefficients β_{jk}, when $\rho \in [1, \infty)$ and $r \in [1, \infty)$, satisfy

$$\Big(\sum_{j=\tau-1}^{\infty} \big[2^{j(s+\frac{1}{2}-\frac{1}{\rho})} \big(\sum_{k=0}^{2^j-1} |\beta_{jk}|^{\rho} \big)^{\frac{1}{\rho}} \big]^r \Big)^{\frac{1}{r}} \leq M,$$

with the usual convention when $\rho = \infty$ and/or $r = \infty$.

4.3 Construction of the wavelet estimator

Let $e_m(t) = e^{i2\pi mt}$, $m \in \mathbb{Z}$, and for any $j_0 \geq 0$ and any $j \geq j_0$, let

$$\phi_{mj_0k} = \langle e_m, \phi_{j_0k} \rangle, \quad \psi_{mjk} = \langle e_m, \psi_{jk} \rangle, \quad f_m = \langle e_m, f \rangle$$

be the Fourier coefficients of $\phi_{j_0k}(\cdot)$, $\psi_{jk}(\cdot)$ and $f(\cdot)$, respectively. Moreover, let

$$h(u,t) = \int f(x) g(u, t-x) dx, \quad t \in T = [0,1], \quad u \in U = [a,b], \tag{4.4}$$

and let the functional Fourier coefficients of $h(u, \cdot)$, $y(u, \cdot)$, $g(u, \cdot)$ and $z(u, \cdot)$ be given, respectively, by

$$h_m(u) = \langle e_m, h(u, \cdot) \rangle, \quad y_m(u) = \langle e_m, y(u, \cdot) \rangle,$$

$$g_m(u) = \langle e_m, g(u, \cdot) \rangle, \quad z_m(u) = \langle e_m, z(u, \cdot) \rangle.$$

Using the properties of the Fourier transform, then for each $u \in U$, for the continuous model (4.1), we have

$$y_m(u) = g_m(u) f_m + \frac{1}{\sqrt{n}} z_m(u),$$

where $g_m(u) = h_m(u)/f_m$ and $z_m(u)$ are generalized one-dimensional Gaussian processes satisfying

$$\mathbb{E}(z_{m_1}(u_1)z_{m_2}(u_2)) = \delta_{m_1,m_2}\delta(u_1 - u_2),$$

where δ_{ml} is the Kronecker's delta. For the discrete version (4.2), using properties of the discrete Fourier transform, for each $l = 1, 2, \ldots, M$, we have

$$y_m(u_l) = g_m(u_l)f_m + \frac{1}{\sqrt{N}} z_{ml},$$

where z_{ml} are standard Gaussian random variables, independent for different m and l, i.e.,

$$\mathbb{E}(z_{m_1,l_1}z_{m_2,l_2}) = \delta_{m_1,m_2}\delta_{l_1,l_2}.$$

A natural estimator of f_m is given by

$$\hat{f}_m = \begin{cases} \frac{\int_a^b \overline{g_m(u)} y_m(u) du}{\int_a^b |g_m(u)|^2 du}, & \text{in the continuous case,} \\ \frac{\sum_{l=1}^M \overline{g_m(u_l)} y_m(u_l)}{\sum_{l=1}^m |g_m(u_l)|^2}, & \text{in the discrete case.} \end{cases}$$

(Here, and in what follows, $\overline{h}$ denotes the conjugate of a complex number or a complex function h; h is real if and only if $\overline{h} = h$.) Consider also the following assumptions on the blurring function $g(\cdot, \cdot)$. Define

$$\tau_1(m) = \begin{cases} \int_a^b |g_m(u)|^2 du, & \text{in the continuous case,} \\ \frac{1}{M}\sum_{l=1}^M |g_m(u_l)|^2, & \text{in the discrete case,} \end{cases}$$

and suppose that, for some constants $\nu \in \mathbb{R}$, $\alpha \geq 0$, $\beta > 0$ and some constants K_1 and K_2, independent of m, the choice of M and the selection of points u_l, $l = 1, 2, \ldots, M$, with $0 < K_1 \leq K_2$,

$$\tau_1(m) \leq K_2 |m|^{-2\nu} \exp(-\alpha|m|^\beta), \quad \nu > 0 \quad \text{if} \quad \alpha = 0 \tag{4.5}$$

and

$$\tau_1(m) \geq K_1 |m|^{-2\nu} \exp(-\alpha|m|^\beta), \quad \nu > 0 \quad \text{if} \quad \alpha = 0. \tag{4.6}$$

Define also

$$\begin{aligned} j_0 = J = \lfloor \log_2 (\tfrac{3}{8\pi} \big(\tfrac{\log(n)}{2\alpha}\big)^{\frac{1}{\beta}}) \rfloor, \quad \alpha > 0, \\ 2^{j_0} \asymp \log(n)^{\frac{p}{2} \vee 1}, \quad 2^J \asymp n^\delta, \qquad \alpha = 0, \end{aligned} \tag{4.7}$$

where $\delta \in (0, (2\nu+1)^{-1}]$. Here, and in what follows $[x]$ denotes the integer part of x and $a \vee b = \max(a, b)$.

By Plancherel's formula, the scaling coefficients, α_{j_0k}, and the wavelet coefficients, β_{jk}, can be represented as

$$\alpha_{j_0k} = \sum_{m \in C^*_{j_0}} f_m \overline{\phi_{mj_0k}}, \quad \beta_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}},$$

where $C^*_{j_0} = \{m : \phi_{mj_0k} \neq 0\}$ and, for all $j \geq j_0$, $C_j = \{m : \psi_{mjk} \neq 0\}$, both subsets of $2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]$, due to the fact that Meyer wavelets are band limited (see, e.g., Johnstone, Kerkyacharian, Picard & Raimondo, 2004, Section 3.1). Hence, α_{j_0k} and β_{jk}, are naturally estimated by

$$\hat{\alpha}_{j_0k} = \sum_{m \in C^*_{j_0}} \hat{f}_m \overline{\phi_{mj_0k}} \quad \hat{\beta}_{jk} = \sum_{m \in C_j} \hat{f}_m \overline{\psi_{mjk}}. \tag{4.8}$$

We now construct a block thresholding wavelet estimator of $f(\cdot)$. For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length l_j. More specifically, let the following set of indices

$$A_j = \{1, 2, \ldots, 2^j/l_j\}, \quad U_{jt} = \{k = 0, 1, \ldots, 2^j - 1 \mid (t-1)l_j \leq k \leq tl_j - 1\}$$

and let

$$l_j \asymp (\log(n))^{(\frac{p}{2} \vee 1)}, \quad \hat{B}_{jt} = \Big(\sum_{k \in U_{jt}} |\hat{\beta}_{jk}|^p / l_j \Big)^{\frac{1}{p}}.$$

For any $j_0 \geq 0$, we finally reconstruct $f(\cdot)$ as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0k} \phi_{j_0k}(t) + \sum_{j=j_0}^{J} \sum_{t \in A_j} \sum_{k \in U_{jt}} \hat{\beta}_{jk} \mathbb{I}\big(|\hat{B}_{jt}| \geq d2^{j\nu} n^{-\frac{1}{2}}\big) \psi_{jk}(t), \tag{4.9}$$

where $\mathbb{I}(A)$ is the indicator function of the set A.

Remark 4.3.1. The estimator of $f(\cdot)$ given by (4.9) is similar to the estimator introduced by Chesneau (2008) for the regular-smooth case (i.e., $\alpha = 0$ in (4.5) and (4.6)) in the standard deconvolution model (i.e., when $a = b$ in the functional deconvolution model (4.1)). Here, we consider the estimator (4.9), with $\hat{\alpha}_{j_0k}$ and $\hat{\beta}_{jk}$ given by (4.8), and prove its optimality (in the minimax sense) under both the functional deconvolution model (4.1) and its discrete version (4.2), for both regular-smooth and super-smooth (i.e., $\alpha > 0$ in (4.5) and (4.6)) convolutions.

Remark 4.3.2. Note that the proposed estimator (4.9) of $f(\cdot)$ is adaptive with respect to s, ρ, r and M, i.e., with respect to the parameters of the Besov ball $B^s_{\rho,r}(M)$, that are usually unknown in practical situations.

4.4 Minimax study under the L^p-risk

We construct below minimax lower bounds for the L^p-risk, $1 \le p < \infty$, both for the continuous model (4.1) and the discrete model (4.2). For this purpose, we define the minimax L^p-risk, $1 \le p < \infty$, over the set Ω as

$$R_n(\Omega) = \inf_{\tilde{f}_n} \sup_{f \in \Omega} \mathbb{E}\|\tilde{f}_n - f\|_p^p,$$

where $\|g\|_p$ is the L^p-norm, $1 \le p < \infty$, of a function $g(\cdot)$ and the infimum is taken over all possible estimators $\tilde{f}_n(\cdot)$ (measurable functions) of $f(\cdot)$, based on observations either from the continuous model (4.1) or the discrete model (4.2).

The following theorem provides the minimax lower bounds for the L^p-risk, $1 \le p < \infty$, under assumption (4.5).

Theorem 4.4.1. *Let $\{\phi_{j_0k}(\cdot), \psi_{jk}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 4.2. Let $s > 1/\rho$, $1 \le \rho \le \infty$, $1 \le r \le \infty$ and $M > 0$. Then, under the assumption (4.5), as $n \to \infty$, there exists some constant $C > 0$ such that,*

$$R_n(B^s_{\rho,r}(M)) \ge \begin{cases} C(\log n)^{-\frac{ps^*}{\beta}}, & if \quad \alpha > 0, \\ Cn^{-p\alpha_1}, & if \quad \alpha = 0, \ \epsilon > 0, \\ C(\frac{\log n}{n})^{\alpha_2 p}, & if \quad \alpha = 0, \ \epsilon \le 0, \end{cases}$$

where

$$\begin{aligned} \epsilon &= s\rho + \frac{2\nu+1}{2}(\rho - p), \quad \alpha_1 = \frac{s}{2(s+\nu)+1} \quad \alpha_2 = \frac{s - \frac{1}{\rho} + \frac{1}{p}}{2(s - \frac{1}{\rho} + \nu) + 1}, \\ s^* &= s + \frac{1}{p} - \frac{1}{\min(p,\rho)}. \end{aligned}$$

Remark 4.4.1. The two different lower bounds for $\alpha = 0$ in Theorem 4.4.1 refer to the dense case ($\epsilon > 0$) when the worst functions $f(\cdot)$ (i.e., the hardest functions to estimate) are spread uniformly over the unit interval T, and the sparse case ($\epsilon \leq 0$) when the worst functions $f(\cdot)$ have only one non-vanishing wavelet coefficient. Also, the restriction $s > \frac{1}{\rho}$, $1 \leq \rho \leq \infty$, $1 \leq r \leq \infty$, that appears in the statement of Theorem 4.4.1, ensures that the corresponding Besov spaces are embedded in the space of continuous functions defined on T, (and, hence, belong to $L^p(T)$, for $1 \leq p < \infty$).

The next theorem provides the upper bounds for the block thresholding wavelet estimator given by (4.9), under the assumption (4.6).

Theorem 4.4.2. *Let $\hat{f}_n(\cdot)$ be the wavelet estimator defined by (4.9), with j_0, J and δ given by (4.7). Let $s > \frac{1}{\rho} - \frac{1}{2} + \frac{1}{2\delta} - \nu$ when $\alpha = 0$ and $s > \frac{1}{\rho}$ when $\alpha = 0$, $1 \leq \rho \leq \infty$, $1 \leq r \leq \infty$ and $M > 0$. Then, under assumption (4.6), as $n \to \infty$, there exists some constant $C > 0$ such that,*

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}(\|\hat{f} - f\|^p_p) \leq \begin{cases} C(\log n)^{-\frac{ps^*}{\beta}}, & \alpha > 0, \\ Cn^{-\alpha_1 p}(\log n)^{\alpha_1 p \mathbb{I}_{\{p>\rho\}}}, & \alpha = 0,\ \epsilon > 0 \\ C(\frac{\log n}{n})^{\alpha_2 p}(\log n)^{(p-\frac{\rho}{r}) + \mathbb{I}_{\{\epsilon=0\}}}, & \alpha = 0,\ \epsilon \leq 0, \end{cases}$$

where α_1, α_2, ϵ and s^* are defined as in Theorem 4.4.1

Remark 4.4.2. Theorems 4.4.1 and 4.4.2 imply that, for the L^p-risk, $1 \leq p < \infty$, the estimator $\hat{f}_n(\cdot)$ defined by (4.9) is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, over a wide range of Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > \frac{1}{\rho} - \frac{1}{2} + \frac{1}{2\delta} - \nu$, when $\alpha = 0$ and $s > \frac{1}{\rho}$ when $\alpha > 0$, $1 \leq \rho \leq \infty$ and $1 \leq r \leq \infty$. In particular, the estimator (4.9) is asymptotically optimal, except for $\epsilon = 0$, $p > \frac{\rho}{r}$ and $\epsilon > 0$, $\quad p > \rho$, in which cases the estimator $\hat{f}_n(\cdot)$ defined by (4.9) is asymptotically near-optimal within a logarithmic factor, i.e.,

$$R_n(B^s_{\rho,r}(M)) \asymp \begin{cases} n^{-p\alpha_1}, & \text{if } \ \alpha = 0,\ \epsilon > 0,\ \rho \geq p, \\ (\frac{\log n}{n})^{\alpha_2 p}, & \text{if } \ \alpha = 0,\ \epsilon < 0, \\ & \text{or } \ \alpha = 0,\ \epsilon = 0,\ p \leq \frac{\rho}{r}, \\ (\log n)^{-\frac{ps^*}{\beta}}, & \text{if } \ \alpha > 0, \end{cases}$$

and

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}\|\hat{f}_n - f\|^p_p \leq \begin{array}{ll} Cn^{-\alpha_1 p}(\log n)^{p\alpha_1}, & \text{if } \ \alpha = 0,\ \epsilon > 0,\ p > \rho, \\ C(\frac{\ln n}{n})^{\alpha_2 p}(\log n)^{(p-\frac{\rho}{r})}, & \text{if } \ \alpha = 0,\ \epsilon = 0,\ p > \frac{\rho}{r}. \end{array}$$

Remark 4.4.3. For the L^p-risk, $1 \leq p < \infty$, the upper bounds obtained in Theorem 4.4.2 are the same as those obtained by Chesneau (2008) for the regular-smooth case (i.e., $\alpha = 0$ in (4.5) and (4.6)) in the standard deconvolution model (i.e., when $a = b$ in the functional deconvolution model (4.1)).

Remark 4.4.4. Following the steps of the proof of Theorem 4.4.2, it is easy to see that if the threshold in (4.9) takes the form $\frac{d\sqrt{\Delta_1(j)}}{\sqrt{n}}$, where $\Delta_1(j) = \frac{1}{|C_j|}\sum_{m \in C_j} \tau_1^{-1}(m)$, Theorem 4.4.2 still holds.

4.5 Examples

In this section, we briefly present inverse problems discussed in Section 4.1 which can be seen as applications of the functional deconvolution model (4.1) or its discrete version (4.2). The optimality (in the minimax sense), or near-optimality within a logarithmic factor, for the L^2-risk over a wide range of Besov balls in the Examples 1-5 below have been discussed in Pensky & Sapatinas (2009a) (see their Examples 4, 1, 2, 3 and 5, respectively); here, we use the methodology presented in Sections 4.3 and 4.4 to check that the corresponding estimators are also optimal or near optimal under the L^p-risk ($1 \leq p < \infty$).

Example 1. ***Estimation of the speed of a wave on a finite interval.*** Let $h(t, x)$ be a solution of the initial–boundary value problem for the wave equation

$$\begin{aligned} \frac{\partial^2 h(t,x)}{\partial t^2} &= \frac{\partial^2 h(t,x)}{\partial x^2} \quad \text{with} \quad h(0,x) = 0, \\ \left.\frac{\partial h(t,x)}{\partial t}\right|_{t=0} &= f(x), \quad h(t,0) = h(t,1) = 0. \end{aligned} \tag{4.10}$$

Here, $f(\cdot)$ is a function defined on the unit interval $[0, 1]$ and $t \in [a, b]$, $a > 0$, $b < 1$.

We assume that a noisy solution $y(t, x) = h(t, x)+n^{-1/2}z(t, x)$ is observed, where $z(t, x)$ is a generalized two-dimensional Gaussian field with covariance function $\mathbb{E}[z(t_1, x_1)z(t_2, x_2)] = \delta(t_1 - t_2)\delta(x_1 - x_2)$, and the goal is to recover the unknown speed of a wave $f(\cdot)$ on the basis of observations $y(t, x)$.

Extending $f(\cdot)$ periodically over the real line, it is well-known (see, e.g., Strauss (1992),

p. 61) that the solution $h(t,x)$ can then be recovered as

$$h(t,x) = \frac{1}{2} \int_0^1 \mathbb{I}(|x-z|<t) f(z) dz, \tag{4.11}$$

so that (4.11) takes the form (4.4) with $g(u,x) = 0.5\,\mathbb{I}(|x|<u)$, where u in (4.4) is replaced by t in (4.11). It is easily seen that the functional Fourier coefficients $g_m(\cdot)$ satisfy (4.5) and (4.6) with $\nu = 1$ and $\alpha = 0$.

Hence, according to Theorem 4.4.1 and Theorem 4.4.2 , the adaptive block thresholding wavelet estimator given by (4.9) achieves the following minimax upper bounds (in the L^p-risk, $1 \le p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \le \begin{cases} n^{-\frac{sp}{2s+3}} \; (\ln n)^{\frac{sp}{2s+3}}, & \text{if} \quad s > \frac{3}{2}(1-p/\rho), \\ \left(\frac{\ln n}{n}\right)^{\frac{p(s-1/\rho+1/p)}{2s-2/\rho+3}} \; (\ln n)^{\max(0,p-\rho/r)\mathbb{I}(\epsilon=0)}, & \text{if} \quad s \le \frac{3}{2}(1-p/\rho), \end{cases}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M>0$ with $s > 1/\rho - 1/2 - 1/(2\delta) + \nu$, $1 \le \rho \le \infty$ and $1 \le r \le \infty$. (The minimax lower bounds (in the L^p-risk, $1 \le p < \infty$) have the same form without the extra logarithmic factor.)

Example 2. ***Estimation of the initial condition in the heat conductivity equation.*** Let $h(t,x)$ be a solution of the heat conductivity equation

$$\frac{\partial h(t,x)}{\partial t} = \frac{\partial^2 h(t,x)}{\partial x^2}, \quad x \in [0,1], \quad t \in [a,b], \quad a>0, \; b<\infty,$$

with initial condition $h(0,x) = f(x)$ and periodic boundary conditions

$$h(t,0) = h(t,1), \quad \left.\frac{\partial h(t,x)}{\partial x}\right|_{x=0} = \left.\frac{\partial h(t,x)}{\partial x}\right|_{x=1}.$$

It is well-known (see, e.g., Strauss, 1992, p. 48) that, under the assumption of periodicity, the solution $h(t,x)$ is given by

$$h(t,x) = (4\pi t)^{-1/2} \int_0^1 \sum_{k\in\mathbb{Z}} \exp\left\{-\frac{(x+k-z)^2}{4t}\right\} f(z) dz,$$

which coincides with (4.4) when t and x are replaced by u and t, respectively. It is easily seen that the functional Fourier coefficients $g_m(\cdot)$ satisfy (4.5) and (4.6) with $\nu = 1$, $\alpha = 8\pi^2 a$ and $\beta = 2$.

Hence, according to Theorem 4.4.1 and Theorem 4.4.2, the adaptive wavelet estimator given by (4.9) achieves the following minimax convergence rate (in the L^p-risk, $1 \leq p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \asymp (\ln n)^{-\frac{p\left(s+\frac{1}{p}-\frac{1}{\min(p,\rho)}\right)}{2}}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho$, $1 \leq \rho \leq \infty$ and $1 \leq r \leq \infty$.

Example 3. ***Estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle.*** Let $h(x, w)$ be a solution of the Dirichlet problem of the Laplacian on a region D on the plane

$$\frac{\partial^2 h(x,w)}{\partial x^2} + \frac{\partial^2 h(x,w)}{\partial w^2} = 0, \quad (x,w) \in D, \tag{4.12}$$

with a boundary ∂D and boundary condition

$$h(x,w)\Big|_{\partial D} = F(x,w). \tag{4.13}$$

Consider the situation when D is the unit circle. It is well-known (see, e.g., Strauss (1992), p. 161) that the solution $h(u,t)$ is given by

$$h(u,t) = \frac{(1-u^2)}{2\pi} \int_0^{2\pi} \frac{f(x)}{1 - 2u\cos(t-x) + u^2}\, dx.$$

$h(x,w) = h(u,t)$, where $u \in [0,1]$ is the polar radius and $t \in [0, 2\pi]$ is the polar angle. Observations are available only on the interior of the unit circle with $u \in [0, r_0]$, $r_0 < 1$, i.e., $a = 0,\ b = r_0 < 1$. It is easily seen that the functional Fourier coefficients satisfy (4.5) and (4.6) with $\nu = 0$, $\alpha = 2\ln(1/r_0)$ and $\beta = 1$.

Hence, according to Theorem 4.4.1 and Theorem 4.4.2, the adaptive wavelet estimator given by (4.9) achieves the following minimax convergence rate (in the L^p-risk, $1 \leq p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \asymp (\ln n)^{-p\left(s+\frac{1}{p}-\frac{1}{\min(p,\rho)}\right)}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho$, $1 \leq \rho \leq \infty$ and $1 \leq r \leq \infty$.

Example 4. ***Estimation of the boundary condition for the Dirichlet problem of the Laplacian on a rectangle.*** Consider the problem 4.12-4.13 in the Example 3 above, with the region D being now a rectangle, i.e., $(x, w) \in [0, 1] \times [a, b]$, $a > 0$, $b < \infty$, and periodic boundary conditions

$$h(x, 0) = f(x), \quad h(0, w) = h(1, w).$$

It is well-known (see, e.g., Strauss (1992), p. 188, p. 407) that, in a periodic setting, the solution $h(x, w)$ can be written as

$$h(x, w) = \pi^{-1} \int_0^1 \sum_{k \in \mathbb{Z}} \frac{w}{w^2 + (x + k - z)^2} f(z) dz.$$

It is easily seen that the functional Fourier coefficients $g_m(\cdot)$ satisfy (4.5) and (4.6), with $\nu = 1/2$, $\alpha = 4\pi a$ and $\beta = 1$.

Hence, according to Theorem 4.4.1 and Theorem 4.4.2, the adaptive wavelet estimator given by (4.9) achieves the following minimax convergence rate (in the L^p-risk, $1 \le p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \asymp (\ln n)^{-p\left(s + \frac{1}{p} - \frac{1}{\min(p,\rho)}\right)}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho$, $1 \le \rho \le \infty$ and $1 \le r \le \infty$.

Example 5. ***Estimation in the multichannel deconvolution problem.*** Consider the problem of estimating $f(\cdot) \in L^p(T)$ on the basis of the following model

$$Y_l(dt) = f * g_l(t)dt + \frac{\sigma_l}{\sqrt{n}} W_l(dt), \quad t \in T = [0, 1], \;\; l = 1, 2, \dots, M, \tag{4.14}$$

where $g_l(\cdot)$ are known blurring functions, σ_l are known positive constants and $W_l(t)$ are independent standard Wiener processes. Note that a discretization of (4.14) (with $\sigma_l = 1$ for $l = 1, \dots, M$) leads to the discrete setup (4.2).

Adaptive term-by-term wavelet thresholding estimators for the model (4.14) were constructed in De Canditiis & Pensky (2006) for regular-smooth convolutions (i.e., $\alpha = 0$ in (4.5) and (4.6)). However, minimax lower and upper bounds were not obtained by these authors who concentrate instead on upper bounds (in the L^p-risk, $1 < p < \infty$) for the error, for a fixed target function. Moreover, the case of super-smooth convolutions (i.e., $\alpha > 0$ in (4.5) and (4.6)) and the case when M can increase together with N have not been treated in De Canditiis & Pensky (2006).

Consider now the adaptive block thresholding wavelet estimator $\hat{f}_n(\cdot)$ defined by (4.9) for the continuous model (4.1) or the discrete model (4.2). Then, under the assumption (4.6), the corresponding minimax lower bounds are given by Theorem 4.4.1, while, under the assumption (4.5), the corresponding minimax upper bounds are given by Theorem 4.4.2. Thus, the proposed functional deconvolution methodology significantly expands on the theoretical findings in De Canditiis & Pensky (2006), and extends the results obtained by Pensky & Sapatinas (2009a) under the L^2-risk to the case of L^p-risk, $1 \leq p < \infty$.

However, the theoretical results obtained in Theorems 4.4.1 and 4.4.2 cannot be blindly applied to the case of boxcar-like convolutions $g(u_l, t) = \frac{1}{2u_l}\mathbb{I}(|t| \leq u_l)$, (i.e. boxcar convolution for each fixed u_l), since $g_m(u) = \sin(2\pi m u)/(2\pi m u)$ and $|g_m(u_*)|^2 = 0$, where $u^* = \arg\max_u g_m(u)$, $u_*, u^* \in [a, b]$. This is an example where a careful choice of u_l, $l = 1, 2, \ldots, M$, can make a difference. For example, if one takes $M = 1$ and $u = u_1$ as a rational number, then $\tau_1(m)$ will vanish for some m large enough and the algorithm will fail to deliver the answer. For this case, we need to use Lemma 1 of Pensky & Sapatinas (2009b) in order establish Theorem 4.5.1.

Lemma 4.5.1. ***(Pensky & Sapatinas (2009b), Lemma 1.)*** *Consider* $g(u,t) = \frac{\gamma(u)}{2}\mathbb{I}(|t| < u), u \in U, t \in T$, *with* $\gamma(\cdot)$ *satisfying* $0 < \gamma_1 \leq \gamma(u) \leq \gamma_2 < \infty, u \in U$ *and* $0 < a < b < \infty$. *Let* $m \in A_j$, *where* $|A_j| = c2^j$, *for some* $c > 0$, *with* $(\ln n)^\delta \leq 2^j \leq n^{1/3}$, $j \geq j_0$, *for some* $\delta > 0$ *and* $j_0 \geq 0$. *Take* $u_l = a + (b-a)l/M$, $l = 1, 2, \ldots, M$. *If* $M \geq M_{0n} = (32\pi/3)(b-a)n^{1/3}$, *then, for* n *and* $|m|$ *large enough,*

$$\tau_1^d(m, \underline{u}, M) \geq K_8 m^{-2},$$

for some constant $K_8 > 0$ *independent of* m, *the choice of* M *and the selection points* u_l, $l = 1, 2, \ldots, M$.

Lemma 1 of Pensky & Sapatinas (2009b) can be applied if $M = M_n \geq c_0 n^{\frac{1}{3}}$ for some constant $c_0 > 0$, independent of n, and $u_l = a + l\Delta/M$, where $\Delta = \min(3c_0/(32\pi), b-a)$.

Theorem 4.5.1. *Let* $\{\phi_{j_0 k}(\cdot), \psi_{jk}\}$ *be the periodic Meyer wavelet basis. Consider the functional deconvolution model (4.2) with* $g(u,t) = \frac{\gamma(u)}{2}\mathbb{I}(|t| < u), u \in U, t \in T$, *where* $\gamma(\cdot)$

is some positive function, satisfying $0 < \gamma_1 \le \gamma(u) \le \gamma_2 < \infty, u \in U$ *and* $0 < a < b < \infty$. *Suppose that* $s > 1/\rho$, $1 \le \rho \le \infty$, $1 \le r \le \infty$ *and* $M > 0$. *Then, as* $n \to \infty$,

$$R_n(B^s_{\rho,r}(M)) \ge \begin{cases} Cn^{-\alpha_1 p}, & \text{if } \epsilon > 0, \\ C(\frac{\log n}{n})^{\alpha_2 p} & \text{if } \epsilon \le 0, \end{cases}$$

where $\alpha_1 = \frac{s}{2s+3}$, $\alpha_2 = \frac{s-\frac{1}{\rho}+\frac{1}{p}}{2(s-\frac{1}{\rho})+3}$ *and* $\epsilon = s\rho + \frac{3}{2}(\rho - p)$.

(Upper bounds) Let $s \in (\frac{1}{\rho} - \frac{1}{2} + \frac{1}{2\delta} - \nu, R)$ $1 \le \rho \le \infty$, $1 \le r \le \infty$ *and* $M > 0$. *Set* $\nu = 1$ *and assume that* $M = M_n \ge c_0 n^{1/3}$ *for some constant* $c_0 > 0$, *independent of* n. *Let* $\hat{f}_n(\underline{u}, M, \cdot)$ *be the wavelet estimator defined by (4.9), evaluated at the points* $u_l = a + l\Delta/M$, $l = 1, 2, \dots, M$, *where* $\Delta = \min(3c_0/(32\pi), b - a)$ *and* j_0 *and* J *are given by (4.7).*

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}(\|\hat{f}_n - f\|_p^p) \le \begin{cases} Cn^{-\alpha_1 p}(\log n)^{\alpha_1 p \mathbb{I}_{\{p>\rho\}}}, & \alpha = 0,\ \epsilon > 0 \\ C(\frac{\log n}{n})^{\alpha_2 p}(\log n)^{(p-\frac{\rho}{r})+\mathbb{I}_{\{\epsilon=0\}}}, & \alpha = 0,\ \epsilon \le 0, \end{cases}$$

where $\alpha_1 = \frac{s}{2s+3}$, $\alpha_2 = \frac{s-\frac{1}{\rho}+\frac{1}{p}}{2(s-\frac{1}{\rho})+3}$ *and* $\epsilon = s\rho + \frac{3}{2}(\rho - p)$.

4.6 Simulation study

Here, we present a small simulation study in the multichannel box-car deconvolution model. We assess the performance of the suggested block thresholding wavelet estimator (BT) given by (4.9), with equispaced selected points u_l, $l = 1, 2, \dots, M$, and compare it to the term-by-term thresholding wavelet estimator (TT) proposed by De Canditiis & Pensky (2006), where the points u_l, $l = 1, 2, \dots, M$, were selected such that one of the u_l's is a BA irrational number, and $u_1, u_2, \dots, u_M$ is a BA irrational tuple.

Specifically, we assume that we observe

$$y(u_l, t) = \int_T f(x) g(u_l, t - x) dx + \frac{\sigma(u_l)}{\sqrt{n}} z(u_l, t), \quad u_l \in U^* \quad l = 1, \dots, M \quad t \in T = [0, 1], \tag{4.15}$$

where

$$g(u_l, t) = \frac{1}{2u_l} \mathbb{I}(|t| < u_l), \quad u_l \in U^*,$$

U^* is a compact subset of $U = [0, 1]$ (bounded away from zero), $\sigma(\cdot)$ is a known function, and $z(u_l, t)$ is a generalized two-dimensional Gaussian field with covariance function

$\mathbb{E}[z(u_1,t_1)z(u_2,t_2)] = \delta(t_1 - t_2)\delta(u_1 - u_2)$. However, in reality, we usually observe a discretization of the functional deconvolution model (4.15) when $y(u,t)$ is observed at $n = NM$ points (u_l, t_i), $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, i.e.,

$$y(u_l, t_i) = \int_T f(x)g(u_l, t_i - x)dx + \sigma_l\, \varepsilon_{li}, \quad u_l \in U^*, \ \ t_i = i/N, \tag{4.16}$$

where ε_{li} are standard Gaussian random variables, independent for different l and i. For simplicity, we assume that $\sigma_l = \sigma$ for all $l = 1, 2, \ldots, M$.

The proposed methodology consists of the following steps:

1. Generate M different equispaced sequences, y_{li} $(= y(u_l, i/N))$, $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, following model (4.16).

2. Generate functions $g(u_l, \cdot)$, $y(u_l, \cdot)$ $\phi_{j_0k}(\cdot)$ and $\psi_{jk}(\cdot)$, $j = j_0, j_0 + 1, \ldots, J - 1$, $k = 0, 1, \ldots, 2^j - 1$, at the same equispaced points, $t_i = i/N$, $i = 1, 2, \ldots, N$.

3. Apply the discrete Fourier transform (FFT) on g_l, y_l, ϕ_{j_0k} and ψ_{jk}, $j = j_0, j_0+1, \ldots, J-1$ $k = 0, 1, \ldots, 2^j - 1$.

4. Estimate α_{j_0k} and β_{jk} by, respectively, $\hat{\alpha}_{j_0k}$ and $\hat{\beta}_{jk}$, given by (4.8)

5. Compute $\hat{B}_{jt} = (\sum_{k\in U_{jt}} |\hat{\beta}_{jk}|^2/l_l)^{\frac{1}{2}}$.

6. Compute the threshold (see Remark 4.4.4)

$$\lambda_j = \hat{\sigma} d^* \frac{\sqrt{\Delta_1(j)}}{\sqrt{n}}, \quad j \geq j_0,$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{M(N-2)} \sum_{l=1}^{M} \sum_{i=2}^{N-1} \left(\frac{y_{l,i-1}}{\sqrt{6}} - \frac{2y_{li}}{\sqrt{6}} + \frac{y_{l,i+1}}{\sqrt{6}} \right)^2}$$

(see, e.g., Müller & Stadmüller, 1987),

$$\Delta_1(j) = \frac{1}{|C_j|} \sum_{m \in C_j} \tau_1^{-1}(m) \quad \text{and} \quad d^* = 1.$$

7. Threshold the wavelet coefficients belonging to blocks with $|\hat{B}_{jt}| < \lambda_j$.

8. Apply the inverse wavelet transform to obtain $\hat{f}_n(\cdot)$ given by (4.9).

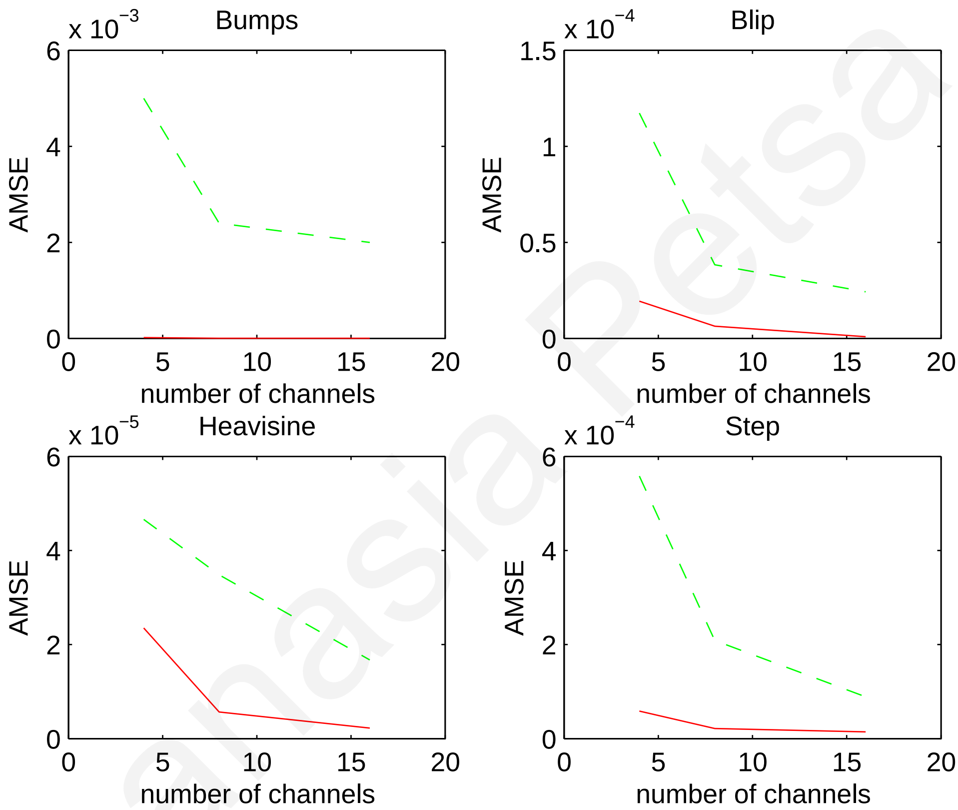

Figure 4.1: AMSE for the *Bumps*, *Blip*, *Heavisine* and *Step* functions sampled at a fixed number of $N = 128$ points, based on rsnr=1, as the number of channels M (and hence the sample size n) increases. Solid line: BT wavelet estimator; Dash line: TT wavelet estimator.

In our numerical analysis, we used the test functions “Bumps”, “Blip”, “Heavisine” and “Step”, and j_0 was set equal to 3. For a fixed value of the (root) signal-to-noise ratio (rsnr=1), we generated $S = 100$ samples of size $n = NM$ from model (4.16) in order to calculate the average mean-squared error (AMSE) given by

$$\frac{1}{S}\sum_{m=1}^{S}\sum_{i=1}^{N}\frac{(\hat{f}_m(t_i) - f(t_i))^2}{\sum_{i=1}^{N} f^2(t_i)}, \quad t_i = i/N.$$

In Figure 4.1, for a fixed number of data points $N = 2^7$, we evaluate the AMSE as the number of channels M, and hence the sample size n, increases for the four signals mentioned above. Obviously, both BT and TT wavelet estimators improve their performances, as n increases, and the BT wavelet estimator appears to be better than the TT wavelet estimator in all cases.

Although not reported here, we also evaluated the performance of the suggested BT wavelet estimator for a wide variety of other test functions (see the list of test functions in Appendix I of Antoniadis, Bigot & Sapatinas, 2001), with very good performances. This numerical study confirms that under the multichannel box-car deconvolution model, block thresholding wavelet estimators with equispaced selected points u_l, $l = 1, 2, \ldots, M$, are quite useful in order to produce accurate estimates of $f(\cdot)$, in finite sample situations.

4.7 Appendix: Proofs

4.7.1 Proof of Theorem 4.4.1

For the proof of Theorem 4.4.1 we are going to use the following lemma.

Lemma 4.7.1. ***(Härdle, Kerkyacharian, Picard & Tsybakov (1998), Lemma 10.1).*** *Let V be a functional space and $d(\cdot,\cdot)$ a distance on V. For f, g on V denote by $\Lambda_n(f,g)$ the likelihood ratio $\Lambda_n(f,g) = \frac{dP_{Y_n}(f)}{dP_{Y_n}(g)}$, where $dP_{Y_n}(h)$ is the probability distribution of the process Y_n if h is the true function.*

If V contains the functions $f_0, f_1, \ldots, f_K$ such that

1. *$d(f_k, f_{k'}) \geq \delta > 0$ for $k = 0, 1, \ldots K$, $k \neq k'$,*

2. $K \geq e^{\lambda_n}$ *for* $\lambda_n > 0$,

3. $\Lambda_n(f_0, f_k) = e^{z_{nk} - v_{nk}}$, *where* v_{nk} *are constants and* z_{nk} *are random variables such that* $\mathbb{P}(z_{nk} > 0) \geq \pi_0$ *for some* $\pi_0 > 0$ *independent of* n *and* k,

4. $\sup_k v_{nk} \leq \lambda_n$.

Then $\sup_{f \in V} P_{Y_n(f)}(d(\hat{f}_n, f) \geq \frac{\delta}{2}) \geq \frac{\pi_0}{2}$.

Sparse case. Consider the continuous model (4.1). Let the functions f_{jk} be of the form $f_{jk} = \gamma_j \psi_{jk}$ and let $f_0 \equiv 0$. Note that it is sufficient to set $\gamma_j = c2^{-j(s+\frac{1}{2}-\frac{1}{\rho})}$, where c is a positive constant such that $c < A$, in order $f_{jk} \in B^s_{\rho,r}(M)$. We then apply Lemma 4.7.1 We consider the class of functions $V = \{f_{jk} : 0 \leq k \leq 2^j - 1\}$ so that $K = 2^j$. We choose $d(f, g) = \|f - g\|_p$, where $\|\cdot\|_p$ is the L^p-norm on the unit interval T. Using the properties of the functions ψ_{jk}, it is easy to see that $d(f_{jk}, f_{jk'}) \asymp \gamma_j 2^{j(\frac{1}{2}-\frac{1}{p})} = \delta$. Set $K = 2^j$, $\lambda_n = j_n \log 2$, $v_{nk} \equiv \lambda_n$ and $z_{nk} = \log(\Lambda_n(f_0, f_{jk})) + \lambda_n$. In order to apply Lemma 4.7.1, we need to show

$$\mathbb{P}_{f_{jk}}(z_{nk} > 0) = \mathbb{P}(\log(\Lambda_n(f_0, f_{jk}) > -j \log 2) \geq \pi_0 > 0$$

uniformly for all f_{jk}. Using the Markov inequality, it is easy to see that we need to find a uniform upper bound for $\mathbb{E}_{f_{jk}}|\log(\Lambda_n(f_0, f_{jk})|$. Set $U = [a, b]$ and $T = [0, 1]$. Let $W(u, t)$ and $\tilde{W}(u, t)$ be Wiener sheets on $U \times T$. Let

$$d\tilde{W}(u, t) = \sqrt{n} f_{jk} * g(u, t) + dW(u, t),$$

where $W(u, t)$ and $\tilde{W}(u, t)$ are the primitives of $dW(u, t)$ and $d\tilde{W}(u, t)$, respectively. Let Q and P be the probability measures associated with $\tilde{W}$ and W, respectively. Using multiparameter Girsanov formula (see, e.g., Dozzi, 1989, p.89), under the assumption, $\int_T \int_U (g * f_{jk})^2(u, t) du dt < \infty$ we arrive at

$$\begin{aligned}
\frac{dQ}{dP} &= \exp\{-\int_T \int_U \sqrt{n}(f_{jk} * g)(u, t) dW(u, t) + \frac{1}{2} n \int_T \int_U (f_{jk} * g)^2(u, t) du dt\} \\
&= \Lambda_n(f_{jk}, f_0).
\end{aligned}$$

Therefore, it is easy to see that

$$\mathbb{E}_{f_{jk}} |\log(\Lambda_n(f_0, f_{jk})| \leq A_n + B_n,$$

where

$$\begin{aligned} A_n &= \sqrt{n}\,\gamma_j\ \mathbb{E}\Big|\int_T\int_U(\psi_{jk}*g)(u,t)\ dW(u,t)\Big|, \\ B_n &= 0.5\,n\,\gamma_j^2\ \int_T\int_U(\psi_{jk}*g)^2(u,t)\ dudt. \end{aligned}$$

Jensen's inequality leads to $A_n \le \sqrt{2B_n}$. Therefore, we only need to construct an upper bound for B_n. Let $\psi_m =< e_m(\cdot),\psi(\cdot) >$. Using the fact that $|\psi_{mjk}| \le 2^{-j/2}$ (see, e.g., Johnstone, Kerkyacharian, Picard & Raimondo, 2004, p.565) and the properties of the Fourier transform we arrive at

$$\begin{aligned} D_j &= \int_0^1\int_a^b(\psi_{jk}*g)(u,t)dudt = \int_a^b\int_0^1|\hat{\psi}_{jk}(\omega)|^2|\hat{g}(u,\omega)|^2 d\omega du \\ &\le \int_a^b 2^{-j}\int_0^1|\hat{g}(u,\omega)|^2 d\omega du = 2^{-j}\int_a^b\sum_{m\in C_j}|\hat{g}_m(u)|^2 du, \end{aligned}$$

where $\hat{\psi}_{jk}(\omega) =< \psi_{jk}(\cdot), e_\omega(\cdot) >$ and $\hat{g}(u,\omega) =< g(u,.), e_\omega(\cdot) >$, where $e_\omega(t) = e^{2\pi i\omega t}$. This leads to

$$B_n \le\ 2^{-j}\frac{n\gamma_j^2}{2}\sum_{m\in C_j}\int_U|g_m(u)|^2 du.$$

We need to choose j_n satisfying

$$\frac{B_n+\sqrt{2B_n}}{j_n\log 2} \le \frac{1}{2}. \tag{4.17}$$

For the case $\alpha > 0$, using assumption (4.5) we have

$$\begin{aligned} \sum_{m\in C_j}\int_a^b|g_m(u)|^2du &\le \sum_{|m|=\frac{2\pi 2^j}{3}}^{\frac{2^j 8\pi}{3}} C|m|^{-2\nu}\exp\{-\alpha|m|^\beta\} \le C\int_{\alpha(\frac{2^j2\pi}{3})^\beta}^{\alpha(\frac{2^j 8\pi}{3})^\beta} z^{-\frac{2\nu}{\beta}}e^{-z}z^{\frac{1}{\beta}-1}dz \\ &\le C2^{j\beta(\frac{1-2\nu-\beta}{\beta})}\exp(-\alpha(\frac{2\pi}{3})^\beta 2^{j\beta}). \end{aligned}$$

For the case $\alpha = 0$, under assumption (4.5)

$$\sum_{m\in C_j}\int_a^b|g_m(u)|^2du \le \sum_{|m|=\frac{2\pi 2^j}{3}}^{\frac{2^j 8\pi}{3}} C|m|^{-2\nu} \le C2^{j(-2\nu+1)}.$$

Hence,

$$\sum_{m\in C_j}\int_U|g_m(u)|^2 \le \begin{cases} C2^{-j(2\nu-1)}, & \text{if } \alpha = 0, \\ \\ C2^{-j(2\nu+\beta-1)}\exp\left(-\alpha(2\pi/3)^\beta 2^{j\beta}\right), & \text{if } \alpha > 0. \end{cases}$$

The smallest j_n satisfying (4.17) satisfies $2^{j_n} \asymp (n/\log n)^{1/(2s+2\nu+1-\frac{2}{\rho})}$ if $\alpha = 0$ and $2^{j_n} \asymp (\log n)^{1/\beta}$ if $\alpha > 0$. Then, Lemma 4.7.1 and Markov inequality lead to

$$\inf_{\tilde{f}} \sup_{f \in B^s_{p,q}} \mathbb{E}\|\tilde{f} - f\|^2 \geq \begin{cases} C(\log n/n)^{\frac{p(s+\frac{1}{p}-\frac{1}{\rho})}{2s+2\nu+1-\frac{2}{\rho}}}, & \text{if } \alpha = 0, \\ \\ C(\log n)^{-\frac{p(s+\frac{1}{p}-\frac{1}{\rho})}{\beta}}, & \text{if } \alpha > 0. \end{cases} \tag{4.18}$$

For the discrete model (4.2) the likelihood ratio is given by

$$\begin{aligned} -\log \Lambda_n(f_0, f_{jk}) &= 0.5 \sum_{i=1}^{N} \sum_{l=1}^{M} \left\{ [y(u_l, t_i) - \gamma_j (\psi_{jk} * g)]^2 (u_l, t_i) - y^2(u_l, t_i) \right\} \\ &= -v_{jk} - u_{jk}, \end{aligned}$$

where

$$\begin{aligned} u_{jk} &= \gamma_j \sum_{i=1}^{N} \sum_{l=1}^{M} (\psi_{jk} * g)(u_l, t_i) \varepsilon_{li}, \\ v_{jk} &= 0.5\, \gamma_j^2 \sum_{i=1}^{N} \sum_{l=1}^{M} (\psi_{jk} * g)^2 (u_l, t_i). \end{aligned}$$

Using Jensen's inequality, it is easy to see that

$$\mathbb{E}(|u_{jk}|) \leq \gamma_j \sqrt{\mathbb{E}(\sum_{i=1}^{N} \sum_{l=1}^{M} (\psi_{jk} * g)^2 (t_i, u_l) \epsilon_{il}^2)} = c\sqrt{v}_{jk}.$$

Therefore, we only need to find an upper bound for v_{jk}. Using the properties of Fourier transform, we arrive at

$$\begin{aligned} |v_{jk}| &\leq 0.5\gamma_j^2 \sum_{l=1}^{M} N \frac{1}{N} \sum_{i=1}^{N} |(\psi_{jk} * g)(t_i, u_l)|^2 \\ &\leq 0.5\gamma_j^2 N \sum_{l=1}^{M} \sum_{m \in C_j} |\psi_{mjk}|^2 |g_m(u_l)|^2 \\ &\leq c\gamma_j^2 NM \frac{1}{|C_j|} \sum_{m \in C_j} \frac{1}{M} \sum_{l=1}^{M} |g_m(u_l)|^2. \end{aligned}$$

Working along the same lines as in the continuous case, with the integral replaced by the sum, we obtain (4.18).

Dense case Consider the continuous model (4.1). Let $\eta = (\eta_0, \eta_1, \ldots, \eta_{2^j-1})$ be the vector with components $\eta_k = \pm 1$, $k = 0, 1, \ldots, 2^j - 1$, set Ξ the set of all possible vectors η

and let $f_{j\eta} = \gamma_j \sum_{k=0}^{2^j-1} \eta_k \psi_{jk}$. Let also η^m be the vector with components $\eta_k^m = (-1)^{\mathbb{I}(m=k)}\eta_k$ for $m, k = 0, 1, \ldots, 2^j - 1$. It is sufficient to set $\gamma_j \leq A2^{-j(s+1/2)}$, in order $f_{j\eta} \in B^s_{\rho,r}(M)$. Additionally we set $\gamma_j = c_\star 2^{-j(s+1/2)}$, where $c_\star$ is a positive constant such that $c_\star < A$, and apply the following lemma on lower bounds.

Lemma 4.7.2. ***(Willer (2005), Lemma 2).*** *Let $\Lambda_n(f, g)$ be defined as in Lemma 4.7.1, and let η and $f_{j\eta}$ be as described above. Suppose that, for some positive constants λ and π_0, we have*

$$\mathbb{P}_{f_{j\eta}}(-\log \Lambda_n(f_{j\eta^m}, f_{j\eta}) \leq \lambda) \geq \pi_0,$$

uniformly for all $f_{j\eta}$ and all $m = 0, \ldots, 2^j - 1$. Then, for any arbitrary estimator $\tilde{f}_n$ and for some positive constant C,

$$\max_{\eta \in \Xi} \mathbb{E}_{f_{j\eta}} \|\tilde{f}_n - f_{j\eta}\| \geq C\pi_0 e^{-\lambda}\, 2^{j/2}\gamma_j.$$

We now need to establish

$$\mathbb{P}_{f_{j,\eta}}(\Lambda_n(f_{j\eta}, f_{j\eta^m}) \geq e^{-\lambda}) \geq \pi^* > 0.$$

Using the same arguments as in the sparse case, it is enough to show

$$\mathbb{E}_{f_{j\eta}} |\log \Lambda_n(f_{j\eta^m}, f_{j\eta})| \leq \lambda_1,$$

for sufficiently small positive λ_1. Using the multiparameter Girsanov formula (see, e.g., Dozzi, 1989, p. 89), under the assumption $\int_T \int_U n(g * f_{jk})^2(u,t)dudt < \infty$, and $|f_{j\eta^m} - f_{j\eta}| = 2\gamma_j|\psi_{jm}|$ we arrive at

$$\begin{aligned}
\log \Lambda_n(f_{j\eta^m}, f_{j\eta}) \;&=\; \sqrt{n} \int_T \int_U ((f_{j\eta^m} - f_{j\eta}) * g)(u,t)dW(u,t) \\
&-\; \frac{n}{2} \int_T \int_U \Big((f_{j\eta^m} - f_{j\eta}) * g\Big)^2 (u,t)dudt \\
&=\; 2\gamma_j\sqrt{n}, \int_T \int_U (\psi_{jm} * g)(u,t)dW(u,t) \\
&-\; 2n{\gamma_j}^2 \int_T \int_U (\psi_{jm} * g)^2(u,t)dudt.
\end{aligned}$$

Therefore, we get

$$\mathbb{E}_{f_{j\eta}} |\log \Lambda_n(f_{j\eta^m}, f_{j\eta})| \leq A_n + B_n,$$

where

$$A_n = 2\sqrt{n}\,\gamma_j\ \mathbb{E}\Big|\int_T\int_U(\psi_{jm}*g)(u,t)\ dW(u,t)\Big|,\quad B_n = 2\,n\,\gamma_j^2\ \int_T\int_U(\psi_{jm}*g)^2(u,t)\ dudt.$$

Due to Jensen's inequality, we have $A_n \leq \sqrt{2}\sqrt{B_n}$. Similarly to the sparse case, it is easy to see that

$$B_n = \begin{cases} O(n2^{-j(2\nu+2s+1)}), & \text{if } \ \alpha = 0, \\ \\ O\Big(n2^{-j(2\nu+\beta+2s+1)}\exp(-\alpha(\frac{2\pi}{3})^\beta 2^{j\beta})\Big), & \text{if } \ \alpha > 0. \end{cases}$$

The smallest j_n satisfying $B_n \leq c$ is given by

$$2^{j_n} \asymp \begin{cases} n^{\frac{1}{2s+2\nu+1}}, & \text{if } \ \alpha = 0, \\ \\ (\log n)^{\frac{1}{\beta}}, & \text{if } \ \alpha > 0, \end{cases}$$

yielding

$$\inf_{\tilde{f}_n}\sup_{f\in B^s_{\rho,r}} \mathbb{E}(\|\tilde{f}_n - f\|_p^p) \geq \begin{cases} Cn^{-\frac{ps}{2s+2\nu+1}}, & \text{if } \ \alpha = 0, \\ \\ C(\log n)^{\frac{-ps}{\beta}}, & \text{if } \ \alpha > 0. \end{cases} \tag{4.19}$$

In the discrete model the log-likelihood ratio is given by

$$\log\Lambda_n(f_{j,\eta^m}, f_{j,\eta}) =$$

$$\begin{aligned} &+ \ 0.5\sum_{i=1}^N\sum_{l=1}^M\Big(y_{il} - \gamma_j(f_{j\eta}*g)(t_i,u_l)\Big)^2 - 0.5\sum_{i=1}^N\sum_{l=1}^M\Big(y_{il} - \gamma_j(f_{j\eta^m}*g)(t_i,u_l)\Big)^2 \\ &= \ \gamma_j\sum_{i=1}^N\sum_{l=1}^M y_{il}[f_{j\eta^m}*g - f_{j\eta}*g](t_i,u_l) + 0.5\gamma_j^2\sum_{i=1}^N\sum_{l=1}^M(f_{j\eta}*g)^2(t_i,u_l) \\ &- \ 0.5\gamma_j^2\sum_{i=1}^N\sum_{l=1}^M(f_{j\eta^m}*g)^2(t_i,u_l) = v^*_{njm} + z^*_{njm}, \end{aligned}$$

where $v^*_{njm} = \gamma_j^2\sum_{i=1}^N\sum_{l=1}^M[(f_{j\eta}*g)(f_{j\eta^m}*g) - 0.5(f_{j\eta}*g)^2 - 0.5(f_{j\eta^m}*g)^2](t_i,u_l)$,

$$\text{and}\quad z^*_{njm} = \gamma_j\sum_{i=1}^N\sum_{l=1}^M\epsilon_{il}[(g*f_{j\eta^m}) - (g*f_{j\eta})](t_i,u_l).$$

It is easy to see that

$$|z^*_{njm}| \leq 2\gamma_j\sum_{i=1}^N\sum_{l=1}^M|\epsilon_{il}(\psi_{jm}*g)(t_i,u_l)|$$

and

$$v^*_{njm} = \gamma_j^2 \sum_{i=1}^{N} \sum_{l=1}^{M} [g * \psi_{jm}]^2(t_i, u_l).$$

Now, we can replace B_n by v_{njm} in the proof of the continuous case. Following the same steps, we obtain (4.19). Obviously, if we have two lower bounds in the same space, the larger one is the true one. We say that we have a dense case if the lower bound obtained with $f_{j\eta}$ holds. Otherwise, we have a sparse case. Hence, it remains to see, for what value of parameters we have sparse and dense case. Using elementary calculus, it is easy to see that we have a dense case for $\epsilon > 0$ and a sparse case for $\epsilon \leq 0$. This completes the proof of Theorem 4.4.1.

4.7.2 Proof of Theorem 4.4.2

For the proof of Theorem 4.4.2, we are going to use two theorems of Chesneau (2006). We first consider the following assumptions

(F1) Let us set $\hat{\beta}_{j_0-1,k} = \hat{\alpha}_{j_0k}$. There exists some constant $C > 0$ such that, for all $j \in \{j_0 - 1, \ldots, J\}$, $k \in \{0, \ldots, 2^j - 1\}$ and n sufficiently large:

$$\mathbb{E}^n_f(|\hat{\beta}_{jk} - \beta_{jk}|^{2p}) \leq C 2^{2jp\nu} n^{-p},$$

(F2) There exist two constants $d > 0$ and $C > 0$ such that, for $j \in \{j_0, \ldots, J\}$, $t \in A_j$ and n sufficiently large

$$\mathbb{P}\Bigg(\Bigg\{\frac{1}{l_j} \sum_{k \in U_{j,t}} |\hat{\beta}_{jk} - \beta_{jk}|^p\Bigg\}^{\frac{1}{p}} \geq 2^{-1} d 2^{j\nu} n^{-\frac{1}{2}}) \leq C n^{-p}.$$

Theorem 4.7.1. ***(Chesneau (2006), Theorem 5.4.1).*** *Let $p \in [1, \infty)$. We assume to have a sequence of models Γ_n in which we are able to produce estimates of the wavelet coefficients α_{jk} and β_{jk} of f on the basis ζ. The corresponding estimators are denoted by $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$. We suppose that assumptions (F1) and (F2) are satisfied. Let $\hat{f}_n$ be defined by*

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{J} \sum_{t \in A_j} \sum_{k \in U_{jt}} \hat{\beta}_{jk}\psi_{jk}(t)\mathbb{I}_{\left\{|\hat{B}_{j,t}| \geq d 2^{j\nu} n^{-\frac{1}{2}}\right\}},$$

with l_j, j_0, J, $\hat{B}_{jt}$, A_j *and* U_{jt} *defined as in Section 4.3 for* $\alpha = 0$. *Then, there exists some constant* $C > 0$ *such that for all* $\rho \in [p, \infty]$, $s \in \left(\frac{1}{\rho} - (\frac{1}{2} - \frac{1}{2\delta} + \nu), R\right)$, $r \in [1, \infty]$ *and* n *sufficiently large, we have*

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}(\|\hat{f}_n - f\|_p^p) \leq C n^{-\alpha_1 p}, \tag{4.20}$$

where $\alpha_1 = \frac{s}{2(s+\nu)+1}$.

Theorem 4.7.2. ***(Chesneau (2006), Theorem 5.4.2).*** *Let* $p \in (1, \infty)$. *We assume to have a sequence of models* Γ_n *in which we are able to produce estimates of the wavelet coefficients* α_{jk} *and* β_{jk} *of* f *on the basis* ζ. *The corresponding estimators are denoted by* $\hat{\alpha}_{jk}$ *and* $\hat{\beta}_{jk}$. *We suppose that assumptions (F1) and (F2) are satisfied. Let* $\hat{f}_n$ *be defined as in Theorem 4.7.1. Then, there exists some constant* $C > 0$ *such that for all* $\rho \in [1, p)$, $s \in \left(\frac{1}{\rho} - (\frac{1}{2} - \frac{1}{2\delta} + \nu), R\right)$, $r \in [1, \infty]$ *and* n *sufficiently large, we have*

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}^n_f(\|\hat{f}_n - f\|_p^p) \leq C \phi_n, \tag{4.21}$$

where

$$\phi_n = \begin{cases} (\frac{\log n}{n})^{\alpha_1 p}, & \epsilon > 0, \\ \\ (\frac{\log n}{n})^{\alpha_2 p} (\log n)^{(p - \frac{\rho}{r}) + \mathbb{I}_{\{\epsilon = 0\}}}, & \epsilon \leq 0, \end{cases}$$

$\alpha_1 = \frac{s}{2(s+\nu)+1}$, $\alpha_2 = \frac{(s - \frac{1}{\rho} + \frac{1}{p})}{2(s - \frac{1}{\rho} + \nu) + 1}$ *and* $\epsilon = s\rho + (\nu + \frac{1}{2})(\rho - p)$.

We will show that Assumptions (F1) and (F2) hold in order to apply Theorems 4.7.1 and 4.7.2 for the case $a = 0$.

<u>Assumption F1</u>

$$\hat{\alpha}_{j_0 k} - \alpha_{j_0 k} = \sum_{m \in C_{j_0}} \phi_{m j_0 k} (\hat{f}_m - f_m) = \int_a^b \int_0^1 h(u,t) \epsilon(u,t) dt du,$$

where $h(u,t) = \sum_{m \in C_{j_0}} \phi_{m j_0 k} \rho_m(u) e^{2imt\pi}$ and $\rho_m(u) = \frac{\overline{g_m(u)}}{\int_a^b |g_m(u)|^2}$. Using the theory of generalized random fields, it is easy to see that $\hat{\alpha}_{j_0 k} - \alpha_{j_0 k}$ is a centered Gaussian random variable. Under the continuous model

$$\begin{aligned}
\operatorname{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k}) &= \frac{1}{n}\int_a^b \int_0^1 |h(u,t)|^2 du dt \\
&= \frac{1}{n}\sum_{m\in C_{j_0}} |\phi_{mj_0k}|^2 \int_a^b |\rho_m(u)|^2 du \int_0^1 |e^{2imt\pi}|^2 dt \\
&+ \frac{1}{n}\sum_{m\neq m'} \phi_{mj_0k}\overline{\phi}_{m'j_0k} \int_a^b \rho_m(u)\rho_{m'}(u) du \int_0^1 e^{2imt\pi} e^{-2im't\pi} dt \\
&= \frac{1}{n}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \int_a^b |\rho_m(u)|^2 du.
\end{aligned}$$

Under the discrete model, if we replace the integral by the sum and use similar arguments, it is easy to see that

$$\operatorname{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k}) = \frac{1}{NM}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \left(\frac{\sum_{l=1}^M |g_m(u_l)|^2}{M}\right)^{-1}.$$

Therefore, $\hat{\alpha}_{j_0k} - \alpha_{j_0k}$ is a centered Gaussian random variable with

$$\operatorname{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k}) = \begin{cases} \frac{1}{n}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \left(\int_a^b |g_m(u)|^2 du\right)^{-1}, & \text{for the continuous model,} \\ \\ \frac{1}{NM}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \left(\frac{\sum_{l=1}^M |g_m(u_l)|^2}{M}\right)^{-1}, & \text{for the discrete model.} \end{cases}$$

Using (4.6), it is easy to see that

$$\begin{aligned}
\operatorname{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k}) &\leq \sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \frac{\tau_1^{-1}(m)}{n} \leq \frac{K_1}{n}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 |m|^{2\nu} \leq \frac{c}{n} 2^{-j_0} 2^{j_0} 2^{2\nu j_0} \\
&\leq \frac{c 2^{2j_0\nu}}{n}
\end{aligned}$$

under both the discrete and continuous model. Similar arguments lead to the conclusion that $\hat{\beta}_{jk} - \beta_{jk}$ for $j \geq j_0$ are also centered Gaussian with variance

$$\operatorname{Var}\left(\hat{\beta}_{jk} - \beta_{jk}\right) \leq \frac{c 2^{2j\nu}}{n}.$$

If a random variable $Z \sim N(0, \sigma^2)$, then

$$\mathbb{E}(|Z|^{2p}) \leq C_{2p}\sigma^{2p}, \quad p > 0,$$

where $C_{2p} = \mathbb{E}(|U|^{2p})$, $U \sim N(0,1)$. Therefore, the following inequalities hold

$$\mathbb{E}(|\hat{\alpha}_{j_0k} - \alpha_{j_0k}|^p) \leq c_p \left[\operatorname{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k})\right]^{\frac{p}{2}} \leq \frac{c 2^{j_0\nu p}}{n^{\frac{p}{2}}}, \tag{4.22}$$

$$\mathbb{E}(|\hat{\beta}_{jk} - \beta_{jk}|^{2p}) \leq c_p[\mathrm{Var}(\hat{\beta}_{jk} - \beta_{jk})]^p \leq \frac{c2^{2p\nu j}}{n^p}.$$

Note that in case $\hat{\alpha}_{j_0k}$ or $\hat{\beta}_{jk}$ are complex numbers, we just take the real part of the above quantities. In this case $\Re\{\hat{\alpha}_{j_0k} - \alpha_{j_0k}\}$ and $\Re\{\hat{\beta}_{jk} - \beta_{jk}\}$ are centered Gaussian with $\mathrm{Var}\bigg(\Re\{\hat{\alpha}_{j_0k} - \alpha_{j_0k}\}\bigg) \leq \mathrm{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k})$ and $\mathrm{Var}\bigg(\Re\{\hat{\beta}_{j_k} - \beta_{jk}\}\bigg) \leq \mathrm{Var}(\hat{\beta}_{jk} - \beta_{jk})$. Using the same arguments as before, we can show that (F1) holds.

Assumption F2

We will first show that F2 holds for $p \geq 2$. It is sufficient to show that

$$\mathbb{P}\left(\left\{\frac{1}{l_j}\sum_{k\in U_{j,t}} |\hat{\beta}_{jk} - \beta_{jk}|^p\right\}^{\frac{1}{p}} \geq \frac{d2^{j\nu}n^{-\frac{1}{2}}}{2}\right) \leq cn^{-p}. \tag{4.23}$$

Consider the centered Gaussian process

$$Z_{jt} = \sum_{k\in U_{jt}} v_k(\hat{\beta}_{jk} - \beta_{jk}),$$

where $v_k \in \Omega_q = \bigg\{v_k : k \in U_{jt} \quad \text{and} \quad \sum_{k\in U_{jt}} |v_k|^q \leq 1\bigg\}$, $\frac{1}{p} + \frac{1}{q} = 1$. Let $\hat{\theta}_{jk} = \hat{\beta}_{jk} - \beta_{jk}$. For any sequence v_{jk} of l_p the Hölder inequality yields

$$\sum_{k\in U_{jt}} v_{jk}\hat{\theta}_{jk} \leq \sum_{k\in U_{jt}} |v_{jk}||\hat{\theta}_{jk}| \leq \{\sum_{k\in U_{jt}} |v_{jk}|^q\}^{\frac{1}{q}}\{\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\}^{1/p}.$$

Therefore, by definition of Ω_q we have

$$\sup_{v\in\Omega_q}\sum_{k\in U_{jt}} v_{jk}\hat{\theta}_{jk} \leq \sup_{v\in\Omega_q}\{\sum_{k\in U_{jt}} |v_{jk}|^q\}^{\frac{1}{q}}\{\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\}^{1/p} \leq \{\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\}^{\frac{1}{p}}. \tag{4.24}$$

Now, let us consider $v^* = (v^*_{jk})_{j,k}$ with $v^*_{jk} = |\hat{\theta}_{jk}|^p\hat{\theta}^{-1}_{jk}\bigg(\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\bigg)^{-\frac{1}{q}}$.

Hence, v^* satisfies

$$|v^*_{jk}| = |\hat{\theta}_{jk}|^{p-1}\left(\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\right)^{-q}$$

and

$$|v^*_{jk}|^q = |\hat{\theta}_{jk}|^p\left(\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p\right)^{-1}.$$

Therefore, $\sum_{k\in U_{jt}} |v^*_{jk}|^q = 1$ and

$$\begin{aligned}
\sup_{v\in\Omega_q} \sum_{k\in U_{jt}} v_{jk}\hat{\theta}_{jk} &\geq \sum_{k\in U_{jt}} v^*_{jk}\hat{\theta}_{jk} = \sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p \hat{\theta}^{-1}_{jk} (\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p)^{-\frac{1}{q}} \hat{\theta}_{jk} = (\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p)^{1-\frac{1}{q}} \\
&= (\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p)^{\frac{1}{p}}. \qquad (4.25)
\end{aligned}$$

If we combine (4.24) and (4.25), we obtain the desired equality, i.e.,

$$\sup_{v\in\Omega_q} Z_{jt}(v) = (\sum_{k\in U_{jt}} |\hat{\theta}_{jk}|^p)^{\frac{1}{p}}. \qquad (4.26)$$

Additionally, Jensen's inequality, (4.26) and (F1) lead to

$$\mathbb{E}(\sup_{v\in\Omega_q} Z_{jt}(v)) = \mathbb{E}\Big[\sum_{k\in U_{jt}} |\hat{\beta}_{jk} - \beta_{jk}|^p\Big]^{\frac{1}{p}} \leq \Big[\sum_{k\in U_{jt}} \mathbb{E}\Big(|\hat{\beta}_{jk} - \beta_{jk}|^p\Big)\Big]^{\frac{1}{p}} \leq C l_j^{\frac{1}{p}} n^{-\frac{1}{2}} 2^{\nu j} = N.$$

Under assumption (4.6) we have

$$\begin{aligned}
&\mathbb{E}\Big((\hat{\beta}_{jk} - \beta_{jk})(\overline{\hat{\beta}_{jk'} - \beta_{jk'}})\Big) = \\
&= \int_a^b \int_0^1 \Big(\sum_{m\in C_j} \psi_{mjk}\rho_m(u) e^{2imt\pi}\Big)\Big(\sum_{m'\in C_j} \overline{\psi}_{m'jk'}\overline{\rho}_{m'}(u)\overline{e^{2imt\pi}}\Big) dtdu \\
&= \sum_{m\in C_j} \int_a^b \int_0^1 \psi_{mjk}\overline{\psi}_{mjk'} |\rho_m(u)|^2 |e^{2imt\pi}|^2 dtdu + \sum_{m\neq m'} \psi_{mjk}\overline{\psi}_{m'jk'} \times \\
&\times \int_a^b \int_0^1 \rho_m(u)\rho_{m'}(u) e^{2imt\pi} e^{-2imt\pi} dtdu = \sum_{m\in C_j} \psi_{mjk}\overline{\psi}_{mjk'} \frac{\tau_1^{-1}(m)}{n}. \qquad (4.27)
\end{aligned}$$

Hence, using (4.27) we get

$$\begin{aligned}
\sup_{v\in\Omega_q} \mathrm{Var}\Big(Z_{jt}(v)\Big) &= \sup_{v\in\Omega_q} \sum_{k\in U_{jt}} \sum_{k'\in U_{jt}} v_k\overline{v}_{k'} \mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})(\overline{\hat{\beta}_{jk'} - \beta_{jk'}})) \qquad (4.28) \\
&= \sup_{v\in\Omega_q} \sum_{k\in U_{jt}} \sum_{k'\in U_{jt}} v_k\overline{v}_{k'} \sum_{m\in C_j} \psi_{mjk}\overline{\psi}_{mjk'} \frac{\tau_1^{-1}(m)}{n} \\
&\leq \frac{K_1 2^{2j\nu}}{n} \sup_{v\in\Omega_q} \sum_{k\in U_{jt}} \sum_{k'\in U_{jt}} v_k\overline{v}_{k'} \mathbb{I}_{\{k=k'\}} \qquad (4.29) \\
&\leq \frac{C 2^{2j\nu}}{n} \sup_{v\in\Omega_q} \sum_{k\in U_{jt}} |v_k|^2 = \frac{K_1 2^{2j\nu}}{n} = V. \qquad (4.30)
\end{aligned}$$

Now, we are going to use the following lemma.

Lemma 4.7.3. ***(Cirelson, Ibragimov & Sudakov (1976)).*** *Let D be a subset of $\mathbb{R} = (-\infty, \infty)$, and let $(\xi_t)_{t \in D}$ be a centered Gaussian process. If $\mathbb{E}(\sup_{t\in D}\ \xi_t) \leq N$ and $\sup_{t\in D}\ Var(\xi_t) \leq V$, then, for all $x > 0$, we have*

$$\mathbb{P}\big(\sup_{t\in D}\xi_t \geq x + N\big) \leq \exp\big(-x^2/(2V)\big). \tag{4.31}$$

Applying Lemma 4.7.3 with $x = \frac{dn^{-\frac{1}{2}}l_j^{\frac{1}{p}}2^{\nu j}}{4}$, $V = K_1\frac{2^{2\nu j}}{n}$, $N = Cl_j^{\frac{1}{p}}n^{-\frac{1}{2}}2^{\nu j}$ and d sufficiently large we have

$$\mathbb{P}\Bigg(\Big\{\frac{1}{l_j}\sum_{k\in U_{jt}}|\hat{\beta}_{jk} - \beta_{jk}|^p\Big\}^{\frac{1}{p}} \geq 2^{-1}2^{\nu j}dn^{-\frac{1}{2}}\Bigg) = \mathbb{P}\Bigg(\sup_{v\in\Omega_q} Z(v) \geq l_j^{\frac{1}{p}}2^{-1}2^{j\nu}dn^{-\frac{1}{2}}\Bigg) \leq$$

$$\leq \mathbb{P}\Bigg(\sup_{v\in\Omega_q} Z(v)) \geq x + N\Bigg) \leq \exp(-\frac{x^2}{2V}) \leq \exp(-cd^2\log n) \leq Cn^{-p}.$$

We will show that assumption (F2) holds for $1 \leq p < 2$. It is easy to see that the following inequalities hold

$$\mathbb{P}\Bigg(\Big\{\frac{1}{l_j}\sum_{k\in U_{jt}}|\hat{\beta}_{jk} - \beta_{jk}|^p\Big\}^{\frac{1}{p}} \geq 0.5d2^{j\nu}n^{-\frac{1}{2}}\Bigg) \leq$$

$$\leq \mathbb{P}\Bigg(\Big\{\frac{1}{l_j}\sum_{k\in U_{jt}}|\hat{\beta}_{jk} - \beta_{jk}|^2\Big\}^{\frac{1}{2}} \geq 0.5d2^{j\nu}n^{-\frac{1}{2}}\Bigg) \leq Cn^{-2} \leq Cn^{-p}. \tag{4.32}$$

Hence, we have shown that (F1) and (F2) are satisfied for all $1 \leq p < \infty$. Applying Theorems 4.7.1 and 4.7.2, we obtain the upper bounds in (4.20) and (4.21) for $\alpha = 0$.

For the case $\alpha > 0$, the estimator is given by $\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1}\hat{\alpha}_{j_0k}\phi_{j_0k}(t)$. Minkowski's inequality leads to

$$\mathbb{E}(\|\hat{f}_n - f\|_p^p) \leq 2^{p-1}\mathbb{E}(\|\sum_{k=0}^{2^{j_0}}(\hat{\alpha}_{j_0k} - \alpha_{j_0k})\phi_{j_0k}\|_p^p) + 2^{p-1}\|\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j-1}\beta_{jk}\psi_{jk}\|_p^p. \tag{4.33}$$

Additionally, using the property of concentration (4.3), inequality (4.22) and the definition of j_0, we have

$$\mathbb{E}(\|\sum_{k=0}^{2^{j_0}-1}(\hat{\alpha}_{j_0k} - \alpha_{j_0k})\phi_{j_0k}\|_p^p) \leq C2^{j_0(\frac{p}{2}-1)}\sum_{k=0}^{2^{j_0}-1}\mathbb{E}(|\hat{\alpha}_{j_0k} - \alpha_{j_0k}|^p) \leq \tag{4.34}$$

$$= (\log n)^{p(\nu+\frac{1}{2})/\beta}n^{-\frac{p}{4}} = o(\log n^{-\frac{ps^*}{\beta}}), \tag{4.35}$$

and

$$\begin{aligned}
\| \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} \|_p^p &\leq \{ \sum_{j=j_0}^{\infty} \| \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} \|_p \}^p \leq \left\{ \sum_{j=j_0}^{\infty} C 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^p \right)^{\frac{1}{p}} \right\}^p \\
&\leq \{ \sum_{j=j_0}^{\infty} C 2^{-j(s+\frac{1}{p}-\frac{1}{\min\{p,\rho\}})} \}^p = C(\log n)^{-\frac{p}{\beta}(s+\frac{1}{p}-\frac{1}{\min\{\rho,p\}})}. \qquad (4.36)
\end{aligned}$$

Inequalities (4.33)-(4.36) lead to the optimal rate of convergence for $\alpha > 0$. This completes the proof of Theorem 4.4.2.

4.7.3 Proof of Theorem 4.5.1

Since $\gamma_1 \leq \gamma(u) \leq \gamma_2$, $u \in [a, b]$, for some $0 < \gamma_1 \leq \gamma_2 < \infty$, it is easily seen that

$$\tau_1^d(m, \underline{u}, M) = \frac{1}{M} \sum_{l=1}^{M} \frac{\gamma^2(u_l)\sin^2(2\pi m u_l)}{4\pi^2 m^2} \asymp \frac{1}{m^2 M} \sum_{l=1}^{M} \sin^2(2\pi m u_l). \qquad (4.37)$$

It follows from (4.37) that for any choice of M and any selection of points $\underline{u}$, we have

$$\tau_1^d(m, \underline{u}, M) \leq K_9 m^{-2}. \qquad (4.38)$$

Using Lemma 4.5.1 and (4.38), it is easy to see that assumptions (4.5) and (4.6) hold. Hence, we can apply Theorems 4.4.1 and 4.4.2, leading to the proof of Theorem 4.5.1.

Future Work

In Chapter 2, we considered the problem of estimation of $\int f^2$ given a sample of weighted data and we constructed an adaptive estimator of $\int f^2$, attaining the minimax rate of convergence that is optimal in the case of direct data for a smooth class of p.d.f.'s. Moreover, using the theory of Ibragimov & Khasminski (1991), we derived the information bound for the problem of estimating $\int f^2$ when weighted data are available.

We now discuss some related questions which remain open. As we mentioned in Chapter 2, the estimation of μ_w (which is obviously unknown in practice) by $\hat{\mu}_w$ prevents us from proving that $\hat{\theta}$ is also asymptotically efficient. We conjecture that this is a general problem for any estimator of $\theta = \int f^2$ based on a weighted sample for the considered class of p.d.f.'s. However, one may be able to propose an asymptotically efficient estimator of $\int f^2$ based on weighted data for a smoother class of p.d.f's, (e.g. Sobolev or Hölder classes).

Moreover, one can investigate whether Assumption 1 (p. 19) can be relaxed. This assumption is very common in estimation for weighted samples but it is interesting to see whether milder assumptions, covering also the case of length biased data, $w(y) = y$, can lead to optimal theoretical results.

As we mentioned in Chapter 2, the estimator $\hat{\theta}$ can be used in order to estimate the L^2-distance of f and f_0, appearing in hypothesis testing for weighted data. Optimality in hypothesis testing is evaluated by other criteria but such a procedure was beyond the scope of Chapter 2. Hence, our work can be used as the intermediate step for hypothesis testing when weighted data are available. Additionally, several statistical procedures using the unknown quantity $\int f^2$ have not been generalized to the case of weighted data, yet. However, there are several settings that lead to weighted data sets and our work can be used when procedures which have been developed for direct data (see, e.g., estimation of the Pitman efficacy of the Wilcoxon signed-rank statistic, rank tests based on residuals in the linear model and estimation of the asymptotic variance of the Hodges-Lehmann estimator) are adapted to weighted data.

In Chapter 3 we considered the problem of estimating the unknown response function in the standard Gaussian white noise model. To deal with this problem, we first utilized the recently developed maximum *a posteriori* (MAP) testimation procedure for recovering an unknown high-dimensional Gaussian mean vector. The existing results for its upper error bounds over various sparse l_p-balls were extended to more general settings and compared with other well-known threshold estimators. These results are of independent interest.

We then applied the MAP testimation procedure in a wavelet context to derive adaptively optimal global and level-wise MAP wavelet testimators of the unknown response function in the standard Gaussian white noise model over a wide range of Besov balls. These results were also extended to the estimation of derivatives of the response function. The efficacy of the proposed level-wise MAP wavelet testimator in finite sample situations was illustrated with a simulation study.

Although we considered only quadratic losses in our exposition, we believe that the obtained results can be extended to more general global losses similar to those in Donoho & Johnstone (1994b) and Johnstone & Silverman (2004b, 2005). Furthermore, the proposed methodology can be adapted to derive pointwise optimal level-wise MAP wavelet testimators of the unknown response function and its derivatives in the standard Gaussian white noise model, as in Cai (2003). Moreover, in Chapter 3, we have shown that the assumptions of Theorem 3.3.2 are satisfied by the truncated geometric prior. One can investigate whether there are other simple parametric priors satisfying all the assumptions of Theorem 3.3.2. Additionally, more explanations on the construction of an appropriate prior π_n for $\beta > 0$ are required, since π_n for $\beta > 0$ does not belong to the class of "standard known" distributions. Extensions of these results to more general inverse problems can also be considered. Moreover, a problem that has not been addressed in Chapter 3 is how the estimation of the standard deviation of the error affects the asymptotic convergence rates.

Appropriate adjustments are needed for each specific problem at hand, and we hope to address these issues elsewhere.

In Chapter 4, we considered the problem of estimation of f under the functional deconvolution model and presented a minimax study under the L^p-risk, $1 \leq p < \infty$. We now present some possible extensions of our work.

A plaussible question is whether the developed theory can be extended to the case of L_∞-risk.

We studied deconvolution with a box-car type blurring function. This important model occurs, e.g., in the problem of estimation of the speed of a wave on a finite interval. It turned out that if $M = M_n \geq c_0 n^{1/3}$ for some constant $c_0 > 0$, independent of n, and the points u_l, $l = 1, 2, \ldots, M$, were selected to be equispaced, then the asymptotical minimax rates of convergence in the discrete model with a box-car type blurring function coincide with the asymptotical minimax rates of convergence in the continuous model.

However, the question remains: if $M = M_n \to \infty$ as $n \to \infty$, but at a rate slower than $O(n^{1/3})$, can one select points u_l, $l = 1, 2, \ldots, M$, such that the asymptotical minimax rates of convergence, in the discrete model coincide with the corresponding asymptotical minimax rates of convergence obtained in the continuous model? And, if for some such $M = M_n$, the asymptotical minimax rates of convrgence in the discrete and the continuous models are not the same, what are the best asymptotical minimax rates of convergence that can be attained and the best selection of points u_l, $l = 1, 2, \ldots, M$? Recent work of Pensky & Sapatinas (2009c) has given answers to some of these questions under the L^2-risk and it will be worth to investigate extending these results under the L^p-risks, $1 \leq p < \infty$.

Bibliography

1. Abramovich, F. & Angelini, C (2006). Bayesian maximum a posteriori multiple testing procedure. *Sankhya*, **68**, 436–460.

2. Abramovich, F. & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Computational Statistics and Data Analysis,* **22**, 351–361.

3. Abramovich, F. & Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

4. Abramovich, F., Bailey, T. C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1-29.

5. Abramovich, F., Grinshtein V. & Pensky M. (2007). On Optimality of Bayesian Testimation in the normal means problem. *Annals of Statistics*, **22**, 121–129.

6. Abramovich, F., Sapatinas, T. & Silverman, B.W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society*, Ser. B, **60**, 725–749.

7. Abramovich, F., Benjamini, Y., Donoho, D.L & Johnstone, I.M. (2006). Adapting to unknown sparsity by controlling the False discovery rate. *Annals of Statistics*, **34**, 584–653.

8. Ahmad, I.A. (1995). On multivariate kernel estimation for samples from weighted distributions. *Statistics and Probability Letters*, **22**, 121–129.

9. Antoniadis, A. (1994). Smoothing noisy data with coiflets. *Statistica Sinica*, **4**, 651-678.

10. Antoniadis, A. (1999) Wavelets in Statistics: A Review, Invited discussion paper. *Journal of the Italian Statistical Society*, **6**.

11. Antoniadis, A., Bigot, J. & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software,* **6**, Article **6**.

12. Aubuchon, J.C. & Hettmansperger, T.P. (1984). A note on the estimation of the integral $f^2(x)$. *Journal of Statistical Planning and Inference*, **9**, 321–331.

13. Barron, A.R., Birgé, L. & Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, **113**, 301-415.

14. Bhattacharyya, B.B., Franklin, L.A. & Richardson, G.D. (1988). A comparison of nonparametric unweighted and length-biased density estimation of fibres. *Communications in Statistics*, Series A, **17**, 3629–3644.

15. Bickel, P.J. & Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya*, Series A, **50**, 381–393.

16. Birgé, L. & Massart, P. (1995). Estimation of integral functionals of a density. *Annals of Statistics*, **23**, 11–29.

17. Birgé, L. & Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, **4**, 329–375.

18. Birgé, L., & Massart, P. (2001). Gaussian model selection. *Journal of European Mathematical Society,* **3**, 203-268.

19. Brown, L.D. & Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics,* **24**, 2384–2398.

20. Brunel, E., Comte, F. & Guilloux, A. (2007). Nonparametric density estimation in the presence of bias and censoring. *Test*, to appear (DOI 10.1007/s11749-007-0075-5).

21. Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Annals of Statistics*, **35**, 1907–1930.

22. Cai, T.T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of Statistics,* **27**, 898–924.

23. Cai, T.T. (2003). Rates of convergence and adaptation over Besov spaces under pointwise risk. *Statistica Sinica* **13**, 881–902.

24. Cai, T.T. & Silverman, B.W. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhyā*, Ser. B, **63**, 127–148.

25. Casey, S.D. & Walnut, D.F. (1994). Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms. *SIAM Review*, **36**, 537–577.

26. Chesneau, C. (2006). Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes. *PhD Thesis*, Université Pierre et Marie Curie - Paris VI, France.

27. Chesneau, C. (2008). Wavelet estimation via block thresholding: a minimax study under the L_p-risk. *Statistica Sinica*, **18**, 1007-1024.

28. Chui, C.K. (1992). *An introduction to Wavelets,* New York: Academic Press.

29. Cirelson, B.S., Ibragimov, I.A. & Sudakov, V.N. (1976). Norm of Gaussian sample function. In *Proceedings of the 3rd Japan-U.S.S.R. Symposium on Probability Theory*, Lecture Notes in Mathematics, Vol. **550**, pp. 20–41, Berlin: Springer-Verlag.

30. Clyde, M.A. & George, E.I. (1999). Empirical Bayes estimation in wavelet nonparametric regression. In *Bayesian Inference in Wavelet-Based Models* (P. Müller and B. Vidakovic, eds), Lecture Notes in Statistics, **141**, pp. 309–322, New York, Springer-Verlag.

31. Cohen, A., Daubechies, I. & Vial, P. (1993). Wavelet bases on the interval and fast algorithms. *Applied and Computational Harmonic Analysis,* **1**, 54–81.

32. Cox, D.R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling*, (Eds. N.L. Johnson & H. Smith), pp. 506–527, New York: Wiley.

33. Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics*, **41**, 909–996.

34. Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia.

35. Daubechies, I. (1993). Orthonormal bases of compactly supported wavelets: II, Variations on a theme. *SIAM Journal on Mathematical Analysis,* **24**, 499-519.

36. De Canditiis, D. & Pensky, M. (2004). Discussion on the meeting on "Statistical Approaches to Inverse Problems". *Journal of the Royal Statistical Society*, Series B, **66**, 638–640.

37. De Canditiis, D. & Pensky, M. (2006). Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics*, **33**, 293–306.

38. Donoho, D.L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis,* **1**, 100–115.

39. Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, **2**, 101–126.

40. Donoho, D.L. & Johnstone, I.M. (1994a). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**, 425–455.

41. Donoho, D.L. & Johnstone, I.M. (1994b). Minimax risk over l_p-balls for l_q-error. *Probability Theory and Related Fields*, **99**, 277–303.

42. Donoho, D.L. & Johnstone, I.M. (1995). Adapting to Unkown Smoothness via Wavelet Shrinkage. *Journal of American Statistical Associaton,* **90**, 1200–1224.

43. Donoho, D.L. & Johnstone, I.M. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli*, **2**, 39-62.

44. Donoho, D.L. & Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics,* **26**, 879–921.

45. Donoho, D.L. & Raimondo, M. (2004) Translation invariant deconvolution in a periodic setting. *International Journal of Wavelets, Multiresolution and Information Processing*, **14**, 415–432.

46. Donoho, D.L., Johnstone, I.M., Hoch, J.C. & Stern, A.C. (1992). Maximum entropy and the nearly black object (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **54**, 41-81.

47. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptotia? (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **57**, 301–369.

48. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1997). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G.L. Yang, eds), pp. 183 – 218, Berlin, Springer-Verlag.

49. Dozzi, M. (1989). *Stochastic Processes with a Multidimensional Parameter.* New York: Longman Scientific & Technical.

50. Draper, D.C. (1988) Rank-based robust analysis of linear models (with discussion). *Statistical Science*, **3**, 239–271

51. Edwards, T. (1992). *Discrete wavelet transforms: Theory and implementations.* Preprint, Mathsoft Wavelet Resources (http://www.mathsoft.com wavelets.html).

52. Efromovich, S. (2004a). Density estimation for biased data. *Annals of Statistics*, **32**, 1137–1161.

53. Efromovich, S. (2004b). Distribution estimation for biased data. *Journal of Statistical Planning and Inference*, **124**, 1–43.

54. El Barmi, H. & Simonoff, J.S. (2000). Transformation-based density estimation for weighted distributions. *Journal of Nonparametric Statistics*, **12**, 861–878.

55. Foster, D.P. and Stine, R. (1999). Local asymptotic coding. *IEEE Transactions on Information Theory*, **45**, 1289–1293.

56. Fromont, M. & Laurent, B. (2006). Adaptive goodness of fit tests in a density model. *Annals of Statistics*, **34**, 680-720

57. George, E.I. & Foster, D.P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, **87**, 731–747.

58. Golubev, G.K. & Khasminskii, R.Z. (1999). A statistical approach to some inverse problems for partial differential equations. *Problems of Information Transmission*, **35**, 136–149.

59. Härdle, W., Kerkyacharian, G., Picard, D. & Tsybakov A. (1998). Wavelets, Approximations and Statistical Applications. *Lecture Notes in Statistics*, **129**, New York: Springer-Verlag.

60. Harsdorf, S. & Reuter, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

61. Houdré, C. & Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*, (Eds. E. Giné, C. Houdré & D. Nualart), pp. 55–69, Basel: Birkhäuser.

62. Ibragimov, I.A. & Khasminskii, R.Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Annals of Statistics*, **19**, 1681–1724.

63. Johnstone, I.M. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. In *Stochastic Decision Theory and Related topics, V*, (Eds. S.S. Gupta & J.O. Berger), pp. 303-326, New York: Springer Verlag.

64. Johnstone, I.M. (2002). Function Estimation and Gaussian Sequence Models, *Unpublished manuscript.*

65. Johnstone, I.M. & Silverman, S.W. (2004a). Boundary coiflets for wavelet shrinkage in function estimation. *Journal of Applied Probability,* **41A**, 81–98.

66. Johnstone, I.M. & Silverman, B.W. (2004b). Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics,* **32**, 1594–1649.

67. Johnstone I.M. & Silverman W.B. (2005). Empirical Bayes Selection of Wavelet thresholds *Annals of Statistics*, **33**, 1700–1752.

68. Johnstone, I.M., Kerkyacharian, G., Picard, D., & Raimondo, M., (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, **66**, 547–573.

69. Jones, M.C. (1991). Kernel density estimation for length biased data. *Biometrika*, **78**, 511–519.

70. Jones, M.C. & Karunamuni, R.J. (1997). Fourier series estimation for length biased data. *Australian Journal of Statistics*, **39**, 57–68.

71. Kalifa, J. & Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, **31**, 58–109.

72. Kolaczyk, E.D. (1994). Wavelet Methods for the Inversion of Certain Homogeneous Linear Operators in the Presence of Noisy Data. *Ph.D. Disseration,* Department of Statistics, Stanford University, USA.

73. Lattes, R. & Lions, J.L. (1967). *Methode de Quasi-Reversibilite et Applications.* Travoux et Recherche Mathematiques, **15**, Paris: Dunod.

74. Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Annals of Statistics*, **24**, 659–681.

75. Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli*, **3**, 181–211.

76. Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM: Probability and Statistics*, **9**, 1–18.

77. Laurent, B. & Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, **28**, 1302–1338.

78. Mallat, S. G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11(7)**, 674-693.

79. Massart, P. (2007) Concentration Inequalities and Model Selection. *Lecture Notes in Mathematics,* **1896**, Berlin, Springer-Verlag.

80. Meyer, Y. (1985). Principe d'incertitude, bases hilbertiennes et algébres d' opérateurs *Sém. Bourbaki, 1985/86,* **662**.

81. Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.

82. Müller, H.G. & Stadmüller, U. (1987) Variable bandwidth kernel estimators of regression curves. *Annals of Statistics*, **15**, 182–201.

83. Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society*, Ser. B, **58**, 463–479.

84. Nason, G.P. & Silverman, B.W. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, **3**, 163—191.

85. Park, Y.J., Dho, S.W. and Kong, H.J. (1997). Deconvolution of long-pulse lidar signals with matrix formulation. *Applied Optics* **36** 5158–5161.

86. Patil, G.P. (2002). Weighted distributions. In *Encyclopedia of Environmetrics*, (Eds. A.H. El-Shaarawi & W.W. Piegorsch), Vol. **4**, pp. 2369–2377, Chichester, John Wiley & Sons.

87. Patil, G.P. & Taillie, C. (1989). Probing encountered data, meta analysis and weighted distributions methods. In *Statistical Data Analysis and Inference*, (Ed. Y. Dodge), pp. 317–345, Amsterdam: Elsevier.

88. Patil, G.P., Rao, C.R. & Zelen, M. (1988). Weighted distributions. In *Encyclopedia of Statistical Sciences*, (Eds. S. Kotz & N.L. Johnson), Vol. **6**, pp. 664–667, New York: Wiley. *Contributions to a General Asymptotic Statistical Theory.* Lecture Notes in Statistics, Vol. **13**, New York: Springer-Verlag.

89. Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory.* Lecture Notes in Statistics, Vol. **13**, New York: Springer-Verlag.

90. Pensky, M. & Sapatinas, T. (2007). Frequentist optimality of Bayes factor estimators in wavelet regression models. *Statistica Sinica*, **17**, 599-633.

91. Pensky, M. & Sapatinas, T. (2009a). Functional deconvolution in a periodic setting: uniform case. *Annals of Statistics*, **37**, 73-104.

92. Pensky, M. & Sapatinas, T. (2009b). On convergence rates equivalency and sampling strategies in functional deconvolution models. *Preprint.*

93. Pensky, M. & Sapatinas, T. (2009c). Diophantine approximations and the problem of estimation of the initial speed of a wave on a finite interval in a stochastic setting. *Preprint.*

94. Ritov, Y. & Bickel P.J. (1990). Achieving information bounds in non and semiparametric models. *Annals of Statistics*, **18**, 925–938.

95. Strang, G. (1993). Wavelet transforms versus fourier transforms. *Bulletin of the American Mathematical Society*, **28**, 288-305.

96. Strauss, W.A. (1992). *Partial Differential Equations: An Introduction.* New York: John Wiley & Sons.

97. Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation.* New York: Springer.

98. Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Annals of Statistics*, **10**, 616–620.

99. Vardi, Y. (1985). Empirical distribution in selection bias model (with discussion). *Annals of Statistics*, **13**, 178–205.

100. Vidakovic, B. & Müller, P. (1994). *Wavelet for kids: Tutorial introduction. Discussion paper 94-13.* Institude of Statistics and Decision Sciences, Duke University, Durham, NC.

101. Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* New York: Wiley.

102. Wu, C.O. (1995). Minimax kernel density estimators with length biased data. *Mathematical Methods of Statistics*, **4**, 56–80.

103. Wu, C.O. & Mao, A.Q. (1996). Minimax kernels for density estimation with biased data. *Annals of the Institute of Statistical Mathematics*, **48**, 451–467.