2013

Nonparametric And Empirical Bayes Estimation Methods

Rida Benhaddou
University of Central Florida

NONPARAMETRIC AND EMPIRICAL BAYES ESTIMATION METHODS

by

RIDA BENHADDOU
M.S. University of Central Florida, 2007

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
in the College of Science
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Professor: Marianna Pensky

ABSTRACT

In the present dissertation, we investigate two different nonparametric models; empirical Bayes model and functional deconvolution model.

In the case of the nonparametric empirical Bayes estimation, we carried out a complete minimax study. In particular, we derive minimax lower bounds for the risk of the nonparametric empirical Bayes estimator for a general conditional distribution. This result has never been obtained previously. In order to attain optimal convergence rates, we use a wavelet series based empirical Bayes estimator constructed in Pensky and Alotaibi (2005). We propose an adaptive version of this estimator using Lepski's method and show that the estimator attains optimal convergence rates. The theory is supplemented by numerous examples.

Our study of the functional deconvolution model expands results of Pensky and Sapatinas (2009, 2010, 2011) to the case of estimating an $(r+1)$-dimensional function or dependent errors. In both cases, we derive minimax lower bounds for the integrated square risk over a wide set of Besov balls and construct adaptive wavelet estimators that attain those optimal convergence rates.

In particular, in the case of estimating a periodic $(r+1)$-dimensional function, we show that by choosing Besov balls of mixed smoothness, we can avoid the "curse of dimensionality" and, hence, obtain higher than usual convergence rates when r is large. The study of deconvolution of a multivariate function is motivated by seismic inversion which can be reduced to solution of noisy two-dimensional convolution equations that allow to draw inference on underground layer structures along the chosen profiles. The common practice in seismology is to recover layer structures separately for each profile and then to combine the derived estimates into a two-dimensional function. By studying the two-dimensional version of the model, we demonstrate that this strategy usually leads to estimators which are less accurate than the ones obtained as two-dimensional functional deconvolutions.

Finally, we consider a multichannel deconvolution model with long-range dependent Gaussian errors. We do not limit our consideration to a specific type of long-range dependence, rather we assume that the eigenvalues of the covariance matrix of the errors are bounded above and below. We show that convergence rates of the estimators depend on a balance between the smoothness parameters of the response function, the

smoothness of the blurring function, the long memory parameters of the errors, and how the total number of observations is distributed among the channels.

ACKNOWLEDGMENTS

In memory of my father L'Amine Benhaddou whose words of encouragement and push for tenacity still linger deep in my ears. I remember the last conversation we had and the advice he gave me to help other people whenever is possible, and the last thing a person should say if somebody ask for help, is no. A special feeling of gratitude to my loving mother Anissa, to my beloved wife Zainab, to Aboubakre, Adil, Anouar, Simo and Wafae. Thank you very much for all your support and sacrifice throughout these years. To the families Benhaddou and Tagmouti in Morocco and abroad and to anyone who cares about them.

A special thanks to Dr. Marianna Pensky, my advisor for her kindness and generosity, for her countless hours of explaining, reading, correcting, encouraging and most of all her patience throughout the whole process. I would not have gone this far without a lot of her time and sincere effort. Thank you Dr. Daguang Han, Dr. Liqiang Ni, and Dr. Jason Swanson for agreeing to serve on my committee and for your willingness to provide help whenever I ask.

I would like to acknowledge and thank the chair of the mathematics department Dr. Piotr Mikusinski and the graduate coordinators Dr. Joseph Brennan and Dr. Xin Li for their countless help throughout these five years. Special thanks goes to the staff members of the mathematics department, namely, Janice Burns, Norma Robles, Linda Perez-Rodriguez and Jeffrey Brody. Special thanks also goes to Dr. Ram Mohapatra and Dr. David Nickerson, for their suggestions and encouragements.

Finally, I would like to thank my friends, Hatim Boustique, Kamran Sadiq, Li Pan and Saliha Pehlivan for their help throughout these five years.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

In this dissertation, we investigate two different topics in nonparametric estimation: empirical Bayes EB methods and functional deconvolution models. For the EB estimation, we propose an adaptive wavelet-based EB estimator using Lepski's method. For the functional deconvolution model, we study two different problems: the problem of estimating an *anisotropic* multivariate periodic function, and the problem of estimating a periodic function under Long-Range dependent (LRD) errors assumption.

Empirical Bayes methods EB are estimation techniques in which the prior distribution, in the standard Bayesian sense, is estimated from the data. They are powerful tools, in particular, when data are generated by repeated execution of the same type of experiment. The EB are directly related to the standard Bayes models but there is difference in perspective between the two: in the standard Bayesian approach, the prior distribution, say $g(\theta)$, is assumed to be fixed before any data are observed, whereas in the EB setting the prior distribution, in some way or another, is estimated from the observed data.

In a typical EB set up, observed data $X = \{X_1, X_2, X_3, \cdots, X_n\}$ are assumed to be generated from an unobserved set of parameters $\{\theta_1, \theta_2, \cdots, \theta_n\}$ according to a probability density function (pdf), $q(x \mid \theta)$. Here, θ is also a random variable but not enough information about its distribution, $g(\theta)$, is available. The idea is the following: an observation X is made characterized by a parameter θ, a realization of Θ, and X is to be used in making a decision about θ. At the time of making that particular observation, denote it by X_{n+1}, there are other observations available, $\{X_1, X_2, X_3, \cdots, X_n\}$ associated with independent realizations $\{\theta_1, \theta_2, \cdots, \theta_n\}$ of Θ. In such a setting, every x_i is a realization of X_i and the $X_i's$ are mutually independent. The goal is to estimate θ_{n+1}, the parameter associated with x_{n+1}, based on the data at hand.

In particular, one has the following setting. One observes independent two-dimensional random vectors $(X_1, \theta_1), \cdots, (X_n, \theta_n)$, where each θ_i is distributed according to some unknown prior pdf g and, given $\theta_i = \theta$ the observation X_i has the known conditional density function $q(x \mid \theta)$. In each pair the first component is observable, but the second is not. After the $(n+1)$-th observation $y \equiv X_{n+1}$ is taken, the goal is to estimate $t \equiv \theta_{n+1}$.

The main contributions of this dissertation to the EB estimation methods are in two ways. The first one is to derive lower bounds for the posterior risk of a nonparametric empirical Bayes estimator. The second one is to provide an adaptive version of the wavelet EB estimator developed in Pensky and Alotaibi (2005), and to construct, in parallel to the minimax lower bounds for the posterior risk, the corresponding upper bounds for the posterior risk of the suggested estimator, in order to justify the asymptotic optimality of such estimator. In particular, we preserve the structure of the linear structure of the estimator. However, since expansion over scaling functions at the resolution level m leads to excessive variance when resolution level m is too high and disproportionately large bias when m is too small, we choose the resolution level using Lepski method introduced in Lepski (1991) and further developed in Lepski, Mammen and Spokony (1997). The resulting estimator is adaptive and attains optimal convergence rates (within a logarithmic factor of n). In addition, it has an advantage of computational efficiency since it is based on the solution of low-dimensional sparse system of linear equations the matrix of which tends to a scalar multiple of an identity matrix as the scale m grows. The theory is supplemented by numerous examples that demonstrate how the estimator can be implemented for various types of distribution families.

Functional deconvolution model deals with the estimation of an unknown function based on observations from its noisy convolution. It has a multitude of applications, in particular, it can be used in a number of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation. For instance, the problem of recovering the initial condition for parabolic equations based on observations in a fixed-time trip was first investigated in Lattes and Lions (1967), and the problem of recovering the boundary condition for elliptic equations based on observations in an interval domain was studied in Golubev and Khasminski (1999) and Golubev (2004).

In this sense, the study is related to a multitude of papers which offered wavelet solutions to deconvolution problems (see, e.g., Donoho (1995), Abramovich and Silverman (1998), Pensky and Vidakovic (1999), Walter and Shen (1999), Fan and Koo (2002), Kalifa and Mallat (2003), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Donoho and Raimondo (2004), Johnstone and Raimondo (2004), Neelamani, Choi and Baraniuk (2004) and Kerkyacharian, Picard and Raimondo (2007)).

A special case of the functional deconvolution model is the standard deconvolution model. In this sense, the study is related to a multitude of papers which offered wavelet solutions to deconvolution problems (see, e.g., Donoho (1995), Abramovich and Silverman (1998), Pensky and Vidakovic (1999), Walter and Shen

(1999), Fan and Koo (2002), Kalifa and Mallat (2003), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Donoho and Raimondo (2004), Johnstone and Raimondo (2004), Neelamani, Choi and Baraniuk (2004) and Kerkyacharian, Picard and Raimondo (2007)).

The main contribution of this dissertation to the analysis the functional deconvolution model is in two different ways. In particular, we consider two problems. The first problem considers the estimation of a periodic $(r+1)$-dimensional function f based on observations from its noisy convolution. An adaptive wavelet-based estimator is constructed here also, and minimax lower bounds for the L^2-risk are derived when f belongs to a Besov ball of mixed smoothness. Furthermore, our estimator proves to be asymptotically near-optimal, in the minimax sense, within a logarithmic factor, in a wide range of Besov balls. We prove in particular that choosing this type of mixed smoothness leads to convergence rates that are free of dimension. Models of these types are very useful, for example, in geophysical explorations, in particular, the ones which rely on inversions of seismic signals. The problem studied in the dissertation is related to seismic inversion which can be reduced to solution of noisy convolution equations which deliver underground layer structures along the chosen profiles. The common practice in seismology is to recover the layer's structure separately for each profile and then to combine them together using interpolation techniques. This, however, is usually not the best strategy and leads to estimators which are inferior to the ones obtained as two-dimensional functional deconvolutions. Indeed, as it is shown in the two-dimensional case, unless function f is very smooth in the direction of the profiles, very spatially inhomogeneous along another dimension and the number of profiles is very limited, functional deconvolution solution has precision superior to combination of M solutions of separate convolution equations.

The second problem looks considers the multichannel deconvolution model from a minimax point of view in the case when errors are not independent but exhibit long-range dependence (LRD). We do not limit our consideration to a specific type of long-range dependence; rather we assume that the errors satisfy a general assumption in terms of the smallest and largest eigenvalues of their covariance matrices. We derive minimax lower bounds for the L^2-risk in the proposed multichannel deconvolution model when the response function is assumed to belong to a Besov ball and the blurring function is assumed to possess some smoothness properties, including both regular-smooth and super-smooth convolutions. Furthermore, we propose an adaptive wavelet estimator of the response function that is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls. It is shown that the optimal convergence rates depend on the balance between the smoothness parameter of the response function, the kernel parameters of the blurring function, the long memory parameters of the errors, and how the total number of observations is distributed among the total number of channels. Some examples of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation are used to illustrate the application of the theory we developed. The optimal convergence rates and the adaptive estimators we consider extend the ones studied by Pensky and Sapatinas (2009, 2010) for independent and identically distributed Gaussian errors to the case of long-range dependent Gaussian errors.

The rest of the dissertation is organized as follows. In Chapter 2 we go over some background information regarding EB estimation as well as some wavelet theory. In Chapter 3 we present our construction of an adaptive wavelet-based EB estimator, discuss the asymptotic optimality of the proposed methodology, and then illustrate the theory with some examples from different families of distributions. Chapter 4 is devoted to our first contribution in functional deconvolution model, in particular, the problem of estimating a periodic $(r + 1)$-dimensional function f based on observations from its noisy convolution. We will also discuss in this chapter its application to geophysical exploration, and provide the argument when the proposed model outperforms old practices in geophysics. Chapter 5 studies our last contribution which deals with another type of functional deconvolution model, multichannel deconvolution model, with the long-range dependent LRD errors. The theory is supplemented by examples of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation to illustrate the application of the theory we developed, before we conclude with a discussion. Finally, in Chapter 6 we give a discussion of our contributions and describe possible future work.

CHAPTER 2: BACKGROUND INFORMATION

2.1 Empirical Bayes Estimation

Empirical Bayes methods (EBM) are estimation techniques in which the prior distribution, in the standard Bayesian sense, is estimated from the data. They are powerful tools in particular when data are generated by repeated execution of the same type of experiment. The EBM are directly related to the standard Bayes models but there is difference in perspective between the two in the sense that in the standard Bayesian approach the prior distribution, say $g(\theta)$, is assumed to be fixed before any data are observed, whereas in the EB setting the prior distribution is, in some way or another, estimated from the observed data.

In a typical EB set up, observed data $X = (X_1, X_2, X_3, \cdots, X_n)$ are assumed to be generated from an unobserved set of parameters $\{\theta_1, \theta_2, \cdots, \theta_n\}$ according to a probability density function (pdf), $q(x \mid \theta)$. Here, θ is also a random variable but not enough information about its distribution, $g(\theta)$, is available. The idea is the following: an observation X is made characterized by a parameter θ, a realization of Θ, and X is to be used in making a decision about θ. At the time of making that particular observation, denote it by X_{n+1}, there are other observations available, $\{X_1, X_2, X_3, \cdots, X_n\}$ associated with independent realizations $\{\theta_1, \theta_2, \cdots, \theta_n\}$ of Θ. In such a setting, every X_i is a realization of x_i and the $x_i's$ are mutually independent. The goal is to estimate θ_{n+1}, the parameter associated with X_{n+1}, based on the data at hand.

2.1.1 Prior Distribution and Identifiability

In the standard Bayesian approach, the conditional expectation of θ given the observed data is given by

$$\mathbf{E}(\theta \mid x) = t(x) = \frac{\int_{-\infty}^{\infty} \theta q(x \mid \theta) g(\theta) d\theta}{p(x)}, \quad where\ p(x) = \int_{-\infty}^{\infty} q(x \mid \theta) g(\theta) d\theta. \tag{2.1.1}$$

Depending on our assumptions about the prior $g(\theta)$, it could be known to belong to a particular family of distributions but no information about its parameters is available or it could be completely unknown for us(unparametrized). This distinction leads to two different types of EBM ; Parametric empirical Bayes methods and nonparametric empirical Bayes methods. Parametric EB methods use information available about the prior distribution at hand and collected data to estimate the values of the population parameters associated with that particular prior by implementing empirical techniques such as the maximum likelihood method and the method of moments. In practice, it is rarely the case that information about the prior distribution would be available for the experimenter; nonparametric EB methods are constructed in such a way that the only thing one has to provide is the sampling distribution, $q(x \mid \theta)$, based on one's own belief, which would depend on the particular experiment at hand.

The possibility of obtaining an estimate of an unparametrized prior g arises from the last expression of the marginal,

$$p(x) = \int_{-\infty}^{\infty} q(x \mid \theta) g(\theta) d\theta \tag{2.1.2}$$

in the sense that the left hand side can be estimated empirically using the observations, and $q(x \mid \theta)$ is known. In terms of distribution functions, the empirical cumulative distribution function obtained from the data, say $P_n(x)$, is an estimate of $P(x)$, the cumulative distribution function associated with $p(x)$, such that $P_n(x) \longrightarrow P(x)(P), as\ n \longrightarrow \infty$. This leads to

$$P_n(x) \approx \int_{-\infty}^{\infty} F(x \mid \theta) g(\theta) d\theta, \tag{2.1.3}$$

where $F(x \mid \theta)$ is cdf of $q(x \mid \theta)$. Robbins (1955) was the first to investigate the possibility of solving for g, his paper is discussed later.

The question is to whether solutions exist, or even unique solution exists. The answer would depend on the nature of $q(x \mid \theta)$, that is, if its corresponding parameter θ is identifiable. In other words, if the sampling distribution, $q(x \mid \theta)$, has the property that different values of its parameters must generate distinct probability distributions. The next couple of examples will clarify the concept of identifiability.

Example 1. Consider an example of the family of normal distributions:

$$P = \left\{ f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mid \theta = (\mu, \sigma) : \mu \in R, \sigma > 0 \right\}$$

Then

$$
\begin{aligned}
f_{\theta_1} &= f_{\theta_2} \\
\Longrightarrow \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2} \\
\Longrightarrow \frac{1}{\sigma_1^2}(x-\mu_1^2)^2 + \log\sigma_1^2 &= \frac{1}{\sigma_2^2}(x-\mu_2^2)^2 + \log\sigma_2^2
\end{aligned}
$$

which implies

$$
x^2\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_1^2}\right) - 2x\left(\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}\right) + \left(\frac{\mu_1^2}{\sigma_1^2}-\frac{\mu_2^2}{\sigma_2^2}+\log\sigma_1^2-\log\sigma_2^2\right) = 0
$$

In fact, in order for the above expression to be equal to zero for almost all values of x we must have $\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}=0$, $\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}=0$ and $\frac{\mu_1^2}{\sigma_1^2}-\frac{\mu_2^2}{\sigma_2^2}+\log\sigma_1^2-\log\sigma_2^2=0$. Consequently, since $\sigma>0$, we must have $\sigma_1=\sigma_2$ and $\mu_1=\mu_2$. Hence $f_{\theta_1}=f_{\theta_2}$ if and only if $\theta_1=\theta_2$, and therefore the parameters of the normal distribution are identifiable.

Example 2. Another interesting example is the standard linear regression model. Indeed

$$
y = \beta' x + \varepsilon, \ \mathbf{E}[\varepsilon \mid x] = 0
$$

Then

$$
\begin{aligned}
y_{\beta_1} &= y_{\beta_2} \\
\beta_1' x + \varepsilon &= \beta_2' x + \eta \\
\Longrightarrow (\beta_1' - \beta_2')x + (\varepsilon - \eta) &= 0
\end{aligned}
$$

Now right multiply both sides by x' and take expectation we obtain

$$
\begin{aligned}
\mathbf{E}[(\beta_1'-\beta_2')xx'] + \mathbf{E}[(\varepsilon-\eta)x'] &= 0 \\
(\beta_1'-\beta_2')\mathbf{E}[xx'] + 0 &= 0
\end{aligned}
$$

This implies that in this case the parameter β is identifiable if and only if $\mathbf{E}[xx']$ is invertible.

Indeed, going back to our previous discussion $P(x)$ is estimated using the data and $g(\theta)$ is approximated by , say $\widehat{g}(\theta)$, the solution of the above integral equation. Another interesting question would be whether such approximation also holds the property that $\widehat{g}(\theta) \longrightarrow g(\theta)(P), as\ n \longrightarrow \infty$. After an estimate of $g(\theta)$ is found we plug it back in the original formula of the posterior mean, this will result in an empirical Bayes decision rule

$$\widehat{t}(x) = \frac{\int_{-\infty}^{\infty} \theta q(x \mid \theta)\widehat{g}(\theta)d\theta}{\int_{-\infty}^{\infty} q(x \mid \theta)\widehat{g}(\theta)d\theta}. \tag{2.1.4}$$

Later we will find out that solving the integral equation with $P(x)$ replaced by $P_n(x)$ will not be an easy task, in particular when $F(x \mid \theta)$ is a continuous *cdf* in the sense that $P_n(x)$ can only be a step function, so solving the integral equation would be impossible. However, provided that some conditions on $F(x \mid \theta)$ are met, it is possible then to solve the integral equation by replacing $P_n(x)$ itself by some $P_n^*(x)$.

2.1.2 Parametric Empirical Bayes Methods (PEBM)

If both the likelihood, $q(x \mid \theta)$, and its prior are assumed to belong to some specified parametric *pdf*'s , that is $g(\theta \mid \psi)$, such as the case of a one or two-dimensional likelihood functions with simple conjugate priors, then the empirical Bayes problem reduces to estimating the marginal, say $p(x \mid \psi)$ and the parameter ψ using empirical methods in the sense that in this parametric set-up the marginal is

$$p(x) = \int_{-\infty}^{\infty} q(x \mid \theta)g(\theta \mid \psi)d\theta = p(x \mid \psi). \tag{2.1.5}$$

For instance, one approach is to approximate the marginal, $p(x \mid \psi)$, by replacing ψ by its empirical counterpart using one of the classical methods of estimation such as the method of maximum likelihood, or the method of moments. This also allows one to replace parameters associated with the prior (population mean and/or variance) by empirical quantities. Keep in mind that whether or not a conjugate family is the right choice for a particular problem is the experimenter's own responsibility, it is a very subjective matter.

There is a great deal of PEBM which includes; the Poisson-Gamma model, Beta-Binomial model, the Gaussian-Gaussian model, the Bayesian linear regression model and the Bayesian multivariate linear regression model. More sophisticated models include hierarchical Bayes models and the Bayesian mixture models. The following example illustrate how this works.

Example 3. Consider the Poisson-Gamma model, where the likelihood function, $q(x \mid \theta)$ is $Poisson(\theta)$ and the prior is $Gamma(\theta \mid \alpha, \beta)$. Suppose we had only one observation x, then the posterior would also be

$Gamma(x+\alpha,, \frac{\beta}{\beta+1})$. In addition, the marginal is

$$\begin{aligned}
p(x) &= \int_0^\infty \frac{\theta^x e^{-\theta}}{x!}\frac{\theta^{\alpha-1}e^{-\frac{\theta}{\beta}}}{\beta^\alpha\Gamma(\alpha)}d\theta, \\
&= \frac{1}{x!}\frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{\theta^{x+\alpha-1}}{\beta^{x+\alpha}}\frac{e^{-\theta(\frac{1+\beta}{\beta})}(1+\beta)^{1+\alpha}}{\Gamma(x+\alpha)}d\theta\frac{\beta^x}{(1+\beta)^{x+\alpha}}\Gamma(x+\alpha), \\
&= \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)\Gamma(x+1)}\left(\frac{\beta}{1+\beta}\right)^x\left(\frac{1}{1+\beta}\right)^\alpha, \\
&= p(x \mid \alpha, \beta). \qquad (2.1.6)
\end{aligned}$$

Therefore, the marginal distribution of X_G would be a *negative binomial*(α, β). An empirical Bayes will be carried out as follows: Estimate the parameters, α and β, of the marginal using one of the empirical methods, either the method of maximum likelihood or the method of moments. In fact we use whichever is easier by hand, that is, since the ML method will require optimizing with respect to two parameters; differentiating with respect to β will lead to $\widehat{\alpha\beta} = \bar{X}$, but differentiating with respect to α is much more complicated and it can be solved only numerically, we prefer then to use the method of moments which relies on matching the first two moments with their empirical counterparts. Consequently,

$$\begin{aligned}
\alpha\beta &= \bar{X}, \qquad (2.1.7) \\
\alpha\beta(1+\beta) + \alpha^2\beta^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 \qquad (2.1.8)
\end{aligned}$$

Then, replacing $\alpha\beta$ by its estimate $\bar{X}$ in (2.1.8), we obtain

$$\begin{aligned}
\alpha\beta(1+\beta) + \bar{X}^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 \\
\Longrightarrow \alpha\beta(1+\beta) &= S^2, \\
\bar{X}(1+\beta) &= S^2. \qquad (2.1.9)
\end{aligned}$$

where $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Which leads to the straightforward method of moments estimates of the parameters α and β,

$$\widehat{\beta} = \frac{S^2}{\bar{X}} - 1, \tag{2.1.10}$$

$$\widehat{\alpha} = \frac{\bar{X}}{\widehat{\beta}}. \tag{2.1.11}$$

The method of moments estimates of the parameters of the marginal turn out to be just what we needed to estimate the prior, $g(\theta \mid \alpha, \beta)$. The second step is to approximate the posterior mean, indeed, using the n data points, the posterior mean in this case,

$$\mathbf{E}(\theta/x) = t(x) = \frac{n\beta\bar{X}}{1+n\beta} + \frac{\alpha\beta}{1+n\beta}, \tag{2.1.12}$$

can be replaced by the empirical quantity

$$\widehat{t}(x) = \frac{n\bar{X}\widehat{\beta}}{1+n\widehat{\beta}} + \frac{\widehat{\alpha}\widehat{\beta}}{1+n\widehat{\beta}}, \tag{2.1.13}$$

or more precisely,

$$\widehat{t}(x) = \frac{n\bar{X}(S^2 - \bar{X}) + \bar{X}^2}{\bar{X} + n(S^2 - \bar{X})}. \tag{2.1.14}$$

Notice that in this example the prior $g(\theta \mid \alpha, \beta)$ was the conjugate of the likelihood function $q(x \mid \theta)$. The empirical reasoning can be carried out the same way in the presence of other conjugate distributions; find the marginal of x_G based on one observation, it should take the form $p(x \mid \psi)$, compute the empirical values of ψ, based on the n data points we have, and then use them to determine the empirical posterior mean and/or variance.

2.1.3 Non-Parametric Empirical Bayes Methods (NPEBM)

In general, information about the prior distribution will not be available but data may be used to obtain approximations to the Bayesian decision rule, $t(x)$. Robbins, who was the first to use the term empirical Bayes estimation (EBE), looked into finding an estimator using a squared error loss function of the value of θ, a realization of the random variable Θ, associated with X whose *pdf* is $q(x \mid \theta)$, known, but the prior of Θ is unknown. Working with several families of discrete probability density functions, Robbins was able to conclude that provided that the available observations are *iid* having an unconditional distribution $p(x)$, the empirical Bayes estimator according to the data $X = (X_1, X_2, X_3, \cdots, X_n)$ at hand can be found and converges with probability 1 to the Bayes estimator as $n \longrightarrow \infty$, for any prior distribution of Θ.

In a non-parametric EB setting we use data available to approximate the decision rule $t(x)$ without requiring the knowledge that the prior distribution takes on any specific parametric form. EB decision rule can be constructed according to two main approaches; either by solving the integral equation

$$p(x) = \int_{-\infty}^{\infty} q(x \mid \theta) g(\theta) d\theta, \tag{2.1.15}$$

or by noticing that the predictive

$$\theta(X_{n+1}) = t(X_{n+1}) = \frac{\int_{-\infty}^{\infty} X_{n+1} q(x \mid \theta) g(\theta) d\theta}{p(X_{n+1})}, \tag{2.1.16}$$

can be manipulated a bit. In fact, denote X_{n+1} by y, then

$$\begin{aligned} \theta(y) = t(y) &= \frac{\int X_{n+1} q(x \mid \theta) g(\theta) d\theta}{p(X_{n+1})}, \\ &= \int y P(\theta \mid y) d\theta, \\ &= \frac{\Psi(y)}{p(y)}. \end{aligned} \tag{2.1.17}$$

One approach is to estimate $\Psi(y)$ and $p(y)$ separately and then compute the ratio. For the other approach, notice that in the case when $q(x \mid \theta)$ is continuous the expression for the marginal $p(x)$ becomes a Fredholm integral equation of type I. As mentioned above the marginal density $p(x)$ can be estimated empirically; so if the integral equation admits a unique solution, then that solution is used to find the EB estimator by plugging the solution in the original formula.

2.1.4 Empirical Bayes Estimators in a Closed Form

One of the first attempts in this regard and due to Robbins (1955). Note that in some cases the Bayes decision rule takes a very closed form in that expressions involving

$$\int_{-\infty}^{\infty} \theta q(x \mid \theta) g(\theta) d\theta \tag{2.1.18}$$

can be rewritten merely in terms of the marginal of X_G and some functions of x; this is the case when some particular family of $pdf's$ is involved. The next couple of examples illustrate the situation.

Example 4. Suppose that $q(x \mid \theta)$ is a $Poisson(\theta)$. Then the posterior mean becomes

$$
\begin{aligned}
t(x) = \mathbf{E}(\theta \mid x) &= \frac{\int_{-\infty}^{\infty} \frac{\theta e^{-\theta}\theta^x g(\theta)}{x!} d\theta}{\int_{-\infty}^{\infty} \frac{e^{-\theta}\theta^x g(\theta)}{x!} d\theta}, \\
&= \frac{(x+1)\int_{-\infty}^{\infty} \frac{\theta^{x+1} e^{-\theta}}{(x+1)!} g(\theta) d\theta}{\int_{-\infty}^{\infty} \frac{e^{-\theta}\theta^x g(\theta)}{x!} d\theta}, \\
&= \frac{(x+1)p(x+1)}{p(x)}. \qquad (2.1.19)
\end{aligned}
$$

Example 5. If the *pdf* is geometric, $q(x \mid \theta) = (1-\theta)\theta^x$. then the same calculations lead to

$$t(x) = \frac{p(x+1)}{p(x)} \qquad (2.1.20)$$

In the above examples we were able to express $t_n(x)$ in terms of the marginal probabilities of the random variable X_G, that is,

$$t(x) = C(x)\frac{p(x+1)}{p(x)} \qquad (2.1.21)$$

The catch is to take advantage of this property and use it in the EB estimation without having to deal with $g(\theta)$. It turns out that this property is unique to the members of discrete exponential family of distributions. In the case of the continuous exponential family, it is more appropriate to use differentiation rather than differencing(discrete). The next couple of examples will illustrate this situation.

Example 6. Suppose that $q(x \mid \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}(x-\theta)^2\}$. Then take a log and differentiate with respect of x to obtain

$$
\begin{aligned}
\frac{d \ln q(x \mid \theta)}{dx} &= 0 - \frac{(x-\theta)}{\sigma^2}, \\
\frac{1}{q(x \mid \theta)}\frac{dq(x \mid \theta)}{dx} &= -\frac{(x-\theta)}{\sigma^2},
\end{aligned}
$$

Which leads to

$$\theta = x + \frac{\sigma^2}{q(x \mid \theta)}\frac{dq(x \mid \theta)}{dx}. \qquad (2.1.22)$$

Now replace the last expression in the definition of $t(x)$ we obtain

$$t(x) = x + \sigma^2 \frac{p'(x)}{p(x)}. \qquad (2.1.23)$$

Example 7. Consider $q(x \mid \theta) = \theta e^{-\theta x}$, for $x \geq 0$, $\theta > 0$. Then the posterior mean becomes

$$
\begin{aligned}
t(x) &= \frac{\int_{-\infty}^{\infty} \theta.\theta e^{-\theta x} g(\theta) d\theta}{\int_{-\infty}^{\infty} \theta e^{-\theta x} g(\theta) d\theta}, \\
&= \frac{\int_{-\infty}^{\infty} \left\{ -\frac{d\theta e^{-\theta x}}{dx} \right\} g(\theta) d\theta}{\int_{-\infty}^{\infty} \theta e^{-\theta x} g(\theta) d\theta}, \\
&= \frac{-\frac{d \int_{-\infty}^{\infty} \theta e^{-\theta x} g(\theta) d\theta}{dx}}{\int_{-\infty}^{\infty} \theta e^{-\theta x} g(\theta) d\theta}, \\
&= -\frac{p'(x)}{p(x)}. \qquad (2.1.24)
\end{aligned}
$$

provided that conditions that allow interchanging integration and differentiation are met. In the last two examples the Bayes decision rule is expressed in terms of the marginal probability density function of X_G and its derivative.

This interesting property which characterizes the exponential family of distributions was first explored by Robbins (1955), the father of the EBM. Robbins was able to device an innovative approach in estimating $t(x)$. Robbins, studying a number of discrete probability mass functions of the exponential family, suggested estimating the marginal frequencies empirically.

Consider the case of compound sampling, where the probability of x_i given θ_i is Poisson distribution, while the prior on θ, $g(\theta)$, is not specified but θ_i are assumed to be i.i.d. Compound sampling comes into play as a modeling tool in a great array of problems, such as accident rates and clinical trials. The goal is to find point prediction θ_i given all x_i, $t(x_i) = E(\theta_i \mid x_i)$, but without having to deal with estimating $g(\theta)$. He did the following: derive the posterior mean in a closed form, since $q(x_i \mid \theta_i)$ is Poisson(θ_i), the posterior mean takes the form

$$t(x_i) = (x_i + 1)\frac{p(x_i + 1)}{p(x_i)} \qquad (2.1.25)$$

then replace the right hand side by

$$\mathbf{E}(\theta_i \mid x_i) \approx (x_i + 1)\frac{N(X_j = x_i + 1)}{N(X_j = x_i)} = \widehat{t}(x_i) \qquad (2.1.26)$$

where x_i takes on the values $x_i = 0, 1, 2, 3, \cdots$, and $N(X_j = x_i)$ is the number of observations in the sample which take the particular value x_i. Robbins concluded that regardless of the unknown prior g we have for any fixed x

$$\widehat{t}(x) \longrightarrow t(x)(P),\ as\ n \longrightarrow \infty \tag{2.1.27}$$

However, Robbins did not attempt to investigate the question to whether $W(\widehat{t}(x)) \longrightarrow W(t(x))(P),\ as\ n \longrightarrow \infty$, where $W(t(x))$ is the Bayesian risk, nor even whether his approach represents the best possible choice amongst other approaches.

In his paper, Robbins also looked into the problem of approximating some functional of the unknown prior g, in particular g itself. He considered the general case in which X is not restricted to discrete values but has a distribution function

$$F(x \mid \theta) = \Pr(X \leq x \mid \Theta = \theta) \tag{2.1.28}$$

The marginal, unconditional, distribution function of X_G is then given by

$$P(x) = \int F(x \mid \theta)g(\theta)d\theta \tag{2.1.29}$$

Now let the infinite sequence $\{x_1, x_2, x_3, \cdots\}$ be iid random variables with common marginal distribution function $P(x)$. Robbins argued that the empirical marginal distribution function defined by

$$P_n(x) = \frac{number\ of\ X_i\ in\ the\ sample\ \{X_1, X_2, X_3, \cdots X_n\} \leq x}{n} \tag{2.1.30}$$

converges uniformly to $P(x)$ with probability 1 as $n \longrightarrow \infty$.

The question is whether one can find, based on $P_n(x)$, a distribution function $G_n(\theta)$ such that $G_n(\theta) \longrightarrow G(\theta), as\ n \longrightarrow \infty$. Let $\mathcal{G}$ denote some class of distribution functions the unknown G is assumed to belong to. Then

$$\int F(x \mid \theta)g(\theta)d\theta \tag{2.1.31}$$

maps $\mathcal{G}$ onto some class of distribution functions, say $\mathcal{F}$. Assume that $F(x \mid \theta)$ is such that the above mentioned mapping is one-to-one. Therefore since $P(x)$ is estimated by $P_n(x)$, then solving

$$P_n(x) = \int F(x \mid \theta)g(\theta)d\theta \tag{2.1.32}$$

would make a lot of sense. However, there is no guarantee that $P_n(x)$ will belong to the class $\mathcal{F}$. For instance, in the case of a continuous $F(x \mid \theta)$ all the elements of $\mathcal{F}$ will be continuous, whereas $P_n(x)$ can only be a step function. To get by this problem, Robbins suggested replacing $P_n(x)$ by some $P_n^*(x) \in \mathcal{F}$ such that the distance between P_n^* and P_n is within ε_n, with $\varepsilon_n \longrightarrow 0$ as $n \longrightarrow \infty$, of the minimum distance of P_n from $\mathcal{F}$. Thus, if G_n is the solution of

$$P_n^*(x) = \int F(x \mid \theta) g(\theta) d\theta \tag{2.1.33}$$

then $P_n^* \longrightarrow P$ in the maximum difference metric, and under suitable conditions on $F(x/\theta)$, we get $G_n(\theta) \longrightarrow G(\theta)$.

2.1.5 Linear Empirical Bayes Estimation

Another interesting nonparametric approach, called the linear EB estimator, is also due to Robbins (1983). Consider, within the typical empirical Bayes framework, that we seek to estimate the parameter θ by some function $t = t(x)$, linear in x. Such linear Bayes estimator is constructed as follows:

Take

$$t(w_0, w_1, x) = w_0 + w_1 x, \tag{2.1.34}$$

Then the goal is to find w_0, w_1 which minimize the Bayes risk (mean squared error)

$$W(t(w_0, w_1, x)) = \int \int (w_0 + w_1 x - \theta)^2 q(x/\theta) dx g(\theta) d\theta, \tag{2.1.35}$$

Indeed, the first order necessary condition of finding a minimum is

$$\frac{\partial W}{\partial w_0} = 2 \int \int (w_0 + w_1 x - \theta) q(x/\theta) dx g(\theta) d\theta = 0, \tag{2.1.36}$$

$$\frac{\partial W}{\partial w_1} = 2 \int \int x (w_0 + w_1 x - \theta) q(x/\theta) dx g(\theta) d\theta = 0, \tag{2.1.37}$$

Which reduces to the two conditions

$$w_0 + w_1 \ \mathbf{E}(x) = \int \theta g(\theta) d\theta, \tag{2.1.38}$$

$$w_0 \ \mathbf{E}(x) + w_1 \ \mathbf{E}(x^2) = \int \theta \ \mathbf{E}(x/\theta) g(\theta) d\theta. \tag{2.1.39}$$

Consequently, multiplying the first equation by $\mathbf{E}(x)$ and subtracting from the second we obtain

$$\begin{aligned} w_1\left(\mathbf{E}(x^2) - \mathbf{E}^2(x)\right) &= \mathbf{E}(\theta x) - \mathbf{E}(\theta)\mathbf{E}(x), & (2.1.40)\\ w_0 + w_1\ \mathbf{E}(x) &= \mathbf{E}(\theta). & (2.1.41)\end{aligned}$$

this implies that

$$w_0 = \mathbf{E}(\theta) - \frac{\text{Cov}(\theta, x)}{\text{Var}(x)}\ \mathbf{E}(x), \tag{2.1.42}$$

and

$$w_1 = \frac{\text{Cov}(\theta, x)}{\text{Var}(x)}, \tag{2.1.43}$$

One advantage to working with linear empirical Bayes over the general approach(not linear), as described by Robbins, is that we only have to deal with the quantities $\mathbf{E}(x)$, $Var(x)$, $\mathbf{E}(\theta)$ and $\text{Cov}(\theta, x)$, which for some families of $f(x \mid \theta)$ are easy to estimate using the data. An example will illustrate this procedure.

Example 8. Let x be exponential with mean θ. Then $E(x \mid \theta) = \theta$ and $E(x^2 \mid \theta) = 2\theta^2$. Thus, using the system of equations above we obtain the solution w_0 and w_1

$$w_1 = \frac{\text{Var}(\theta)}{\text{Var}(\theta) + E(\theta^2)},\ \textit{and}\ w_0 = \frac{E(\theta)E(\theta^2)}{\text{Var}(\theta) + E(\theta^2)}. \tag{2.1.44}$$

The linear Bayes estimators are of great importance because, besides being computationally efficient and easy to maneuver, they can also be useful in finding estimates for the first two moments of the random variable Θ, which are certainly needed to estimate w_0 and w_1. Just take a look at the integral

$$\int \mathbf{E}(x \mid \theta) g(\theta) d\theta \tag{2.1.45}$$

taken from the system of equations above; it is equal to the expectation over the marginal distribution of x_G, or in general we have

$$\mathbf{E}(x_G^r) = \int \mathbf{E}(x^r \mid \theta) g(\theta) d\theta,\ r = 1, 2 \tag{2.1.46}$$

Here again, $E(x_G^r)$ can be estimated using the data. Thus, empirical estimation of the first two moments will be carried out in the following way: First, estimate the first two moments of the marginal density using data, this will give us $\overline{X}$ and $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, then the first two empirical moments of the distribution of Θ

are found by matching. That is, using results from Example 8 we obtain

$$\int \mathbf{E}(x^2 \mid \theta) g(\theta) d\theta = \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \overline{X^2}, \tag{2.1.47}$$

$$\int \mathbf{E}(x \mid \theta) g(\theta) d\theta = \overline{X}, \tag{2.1.48}$$

or

$$\int 2\theta^2 g(\theta) d\theta = \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \overline{X^2}, \tag{2.1.49}$$

$$\int \theta g(\theta) d\theta = \overline{X}. \tag{2.1.50}$$

which leads to the empirical first two moments of the distribution of Θ

$$\int \theta^2 g(\theta) d\theta = \frac{1}{2} \overline{X^2} = \widehat{\mathbf{E}}(\theta^2), \tag{2.1.51}$$

$$\int \theta g(\theta) d\theta = \overline{X} = \widehat{\mathbf{E}}(\theta). \tag{2.1.52}$$

The second step is to replace w_0 and w_1 by $\widehat{w}_0$ and $\widehat{w}_1$ to obtain an empirical quantity for $t(w_0, w_1, x)$. Indeed,

$$\widehat{t}(w_0, w_1, x) = \widehat{w}_0 + \widehat{w}_1 x, \tag{2.1.53}$$

with

$$\widehat{w}_0 = \frac{\overline{X}\ \overline{X^2}}{2(\overline{X^2} - \overline{X}^2)} = \frac{\overline{X}(S^2 + \overline{X}^2)}{2S^2}, \tag{2.1.54}$$

$$\widehat{w}_1 = \frac{\overline{X^2} - 2\overline{X}^2}{2(\overline{X^2} - \overline{X}^2)} = \frac{S^2 - \overline{X}^2}{2S^2}. \tag{2.1.55}$$

where $S^2 = \overline{X^2} - \overline{X}^2$.

Linear EB estimators are simple to implement, and can be very efficient computationally, but they may not do a good job approximating the Bayesian estimator. The comprehensive list of references as well as numerous examples of applications of EB techniques can be found in Carlin and Louis (2000) or Maritz and Lwin (1989).

2.1.6 Measuring Precision of the EB Procedures

When the Bayes decision rule is estimated from the data its corresponding Bayes risk $W(t)$ will also be empirical. So the goodness of the empirical Bayes estimator could be evaluated by $W(\widehat{t})$. However, this latter is itself a random variable since it is derived from the data. Thus to be able to evaluate the overall quality of the EBE often we need to investigate the distribution of $W(\widehat{t})$. In this case a better measurement of the performance of an EB method would be the sample average of $W(\widehat{t})$, $E_n W(\widehat{t})$. Whether $W(\widehat{t})$ or $E_n W(\widehat{t})$ is the better choice depends on the situation, and it is left to experimenter's own belief. In my first research paper we propose an adaptive nonparametric Bayes estimation using wavelet series, and since the wavelets expansion allows us to better represent local behavior of a function we feel that working with a local measure of risk, $W(\widehat{t})$, would make more sense. If the EB estimator is powerful enough then it is said to be *asymptotically optimal*. That would happen if

$$\mathbf{E}_n W(\widehat{t}) \longrightarrow W(t)(E), \tag{2.1.56}$$

as $n \longrightarrow \infty$ (convergence in mean). This does not mean that $E_n W(\widehat{t})$ is very close to the Bayes risk $W(t)$; it could be even greater than the risk, in the mean squared error sense, of some non-Bayesian (classical) estimators. Asymptotic optimality can also be measured by convergence in probability, that is,

$$W(\widehat{t}) \longrightarrow W(t)(P), \tag{2.1.57}$$

as $n \longrightarrow \infty$. There is another measurement of the performance that can be very powerful in deciding between an EB procedure and a classical technique. Indeed, an EB estimator would also be asymptotically optimal if

$$P_n\left(W(\widehat{t}) < W(T_n)\right) > 1 - \varepsilon \tag{2.1.58}$$

for n large enough, where T_n is some classical estimator.

2.2 Wavelets Theory

2.2.1 Wavelets

Consider the space $L^2(\mathbb{R})$ of square integrable functions(sometimes, for theoretical reasons, and to satisfy some analytical requirements the space of measurable functions that are absolutely and square integrable is preferred). Hence functions must decay rapidly to zero.

The idea is to consider dilations and translations of a wavelet function, say ψ, in order to cover $\mathbb{R}$ entirely. That is, we consider the functions $\psi_{j,k}(x)$, where

$$\psi_{j,k}(x) = \psi(2^j x - k), \quad j,k \in \mathbb{Z}. \tag{2.2.1}$$

The functions $\{\psi_{j,k}(x), j,k \in \mathbb{Z}\}$ form a basis that is not necessarily orthogonal. However, there is need to work with orthogonal bases in that they have the property of allowing the perfect reconstruction of a signal from the coefficients of the transform.

The possibility of obtaining orthogonal bases allows us to expand any function $f \in L^2(\mathbb{R})$ as a wavelet series. Therefore, consider an orthonormal basis

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j,k \in \mathbb{Z}, \tag{2.2.2}$$

Then the wavelet series expansion of a function $f \in L^2(\mathbb{R})$ is given by

$$f(x) = \sum_{j,k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(x) \tag{2.2.3}$$

where $c_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx$, which is convergent in norm.

2.2.2 Properties of the Wavelet ψ

In order for ψ to be a wavelet for continuous wavelet transform, it must satisfy what is called admissibility condition so that we obtain a stably invertible transform. The properties of the wavelet functions are

i. $\int_{-\infty}^{\infty} \psi(x)\, dx = 0$ (Admissibility condition)

This condition is equivalent to $\widehat{\psi}(0) = 0$ where $\widehat{\psi}(\omega)$ is the Fourier transform since $\widehat{\psi}(0) = \int_{-\infty}^{\infty} \psi(x)\, dx$. This is also equivalent to

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|}\, d\omega < \infty. \tag{2.2.4}$$

Remark. If $\widehat{\psi}(0) = 0$ and $\widehat{\psi}(\omega)$ is continuously differentiable, then the admissibility condition holds. In addition, a sufficient time decay guarantees that $\widehat{\psi}(\omega)$ is continuously differentiable. Namely, the condition

$\int_{-\infty}^{\infty}(1+|x|)\,|\psi(x)|\;dx < \infty$ is a sufficient condition to guarantee continuous differentiability of $\widehat{\psi}(\omega)$, and thus the admissibility of $\psi(x)$.

For wavelet functions from the space $L^1(\mathbb{R})\bigcap L^2(\mathbb{R})$ one can also restrict ψ to satisfy the condition

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx = 1. \tag{2.2.5}$$

ii. The first $r-1$ moments of ψ vanish.

Sometimes it is useful to restrict ψ to be a continuous function with a higher number, $r-1$, of vanishing moments. that is,

$$\int_{-\infty}^{\infty} x^j \psi(x)\, dx = 0 \quad for\ j = 0, 1, \cdots, r-1, \tag{2.2.6}$$

where $r \geq 1$. Also

$$\int_{-\infty}^{\infty} |x^r \psi(x)|\; dx < \infty. \tag{2.2.7}$$

Note that the level of r is an indicator of how smooth ψ is: the larger r is, the smoother is ψ.

For the discrete wavelet transform, we need at least the condition that the wavelet series is a representation of the identity in $L^2(\mathbb{R})$. Generally, multi-resolution analysis is utilized to construct such transforms.

2.2.3 Properties of the Scaling Function φ

Wavelets are characterized by the wavelet function ψ and a scaling function φ. ψ is a band-pass filter and at each level of scales its band-width is halved. This leads to the problem that it would take an infinite number of levels to be able to cover the entire spectrum. Here when the scaling function comes in to play; it filters the lowest level of the transform and ensures that all the spectrum is covered.

The scaling function, φ, can be obtained from the mother wavelet, ψ, by the relation

$$|\widehat{\varphi}(\omega)|^2 = \int_{\omega}^{\infty} \frac{\left|\widehat{\psi}(r)\right|^2}{r}\, dr. \tag{2.2.8}$$

This function generates an orthonormal family of $L^2(\mathbb{R})$,

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) \quad j, k \in \mathbb{Z}. \tag{2.2.9}$$

Now consider the orthonormal system

$$\{\psi_{j,k}(x), \varphi_{j,k}(x), j, k \in \mathbb{Z}\}_{j \geq j_0, k}, \tag{2.2.10}$$

Then the wavelet series for $f \in L^2(\mathbb{R})$ becomes

$$f(x) = \sum_{k=-\infty}^{\infty} c_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j \geq j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(x), \tag{2.2.11}$$

where $c_{j_0,k} = \int_{-\infty}^{\infty} f(x)\varphi_{j_0,k}(x)dx$, and $d_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx$.

2.2.4 Multi-resolution Analysis

We can construct wavelets ψ such that the family

$$\{\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)\}_{(j,k)\in\mathbf{Z}} \tag{2.2.12}$$

forms an orthonormal basis of $L^2(\mathbb{R})$. A sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is said to be a multiresolution approximation if the following properties are met:

a. $f(t) \in V_j \leftrightarrow f(t - 2^j k) \in V_j$, for all $(j, k) \in \mathbb{Z}$.

b. For all $j \in \mathbb{Z}$, $V_{j+1} \subset V_j$.

c. $f(t) \in V_j \leftrightarrow f(t/2) \in V_{j+1}$, for all $j \in \mathbb{Z}$.

d. $\bigcap_j V_j = \{0\}$

e. $L^2(\mathbb{R}) = \overline{\bigcup_j V_j}$

f. There exists υ such that $\{\upsilon(t-n)\}_{n\in\mathbf{Z}}$ is a Riesz basis of V_0.

In fact, such a sequence $\{V_j\}_{j\in Z}$ satisfies certain self-similarity relation in time and scale, as well as completeness and regularity relations. Self-similarity in time requires that each subspace V_k is invariant to any shifts proportional to the scale 2^k, this is represented by property **a**. Self-similarity in scale requires that all subspaces $V_l \subset V_k$, where $k < l$, are time-scaled versions of each other in the sense that for each $f \in V_k$, there is a $g \in V_l$ with $x \in \mathbb{R}$ such that $g(x) = f(2^{k-l}x)$, this is represented by property **c**. Property **b** can be thought of as a causality property in that approximations at a resolution 2^{-j} provide whatever necessary information we need to compute approximations at a resolution 2^{-j-1}. Properties **d** and **e** summarize the concept of completeness, which requires that those nested subspaces fill in the whole $L^2(\mathbb{R})$ space in the sense that their union should be dense in $L^2(\mathbb{R})$ and their intersection should only contain the zero element. Finally, the existence of a Riesz basis $\{\upsilon(t-n)\}_{n\in\mathbf{Z}}$ is important to guarantee that signal expansions over $\{\upsilon(t-n)\}_{n\in\mathbf{Z}}$ are stable.

2.2.5 The Wavelet Transform

Theorem 1. For any $f \in L^2(\mathbb{R})$, the continuous wavelet transform (CWT) with respect to ψ is defined by:

$$Wf(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-b}{a}\right) dt, \quad \text{for } a,b \in \mathbb{R} \textit{ and } a \neq 0, \tag{2.2.13}$$

and its inverse transform is

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} Wf(a,b)\, \frac{1}{\sqrt{|a|}}\, \psi\left(\frac{t-a}{a}\right) db\, \frac{da}{a^2}, \tag{2.2.14}$$

where $C_\psi = \int_{-\infty}^{\infty} \frac{\left|\widehat{\psi}(w)\right|^2}{|w|}\, dw$.

In addition

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |Wf(a,b)|^2\, db \frac{da}{a^2}. \tag{2.2.15}$$

Before we proceed to the proof let us first state a couple of powerful results from Fourier transform theory.

Lemma 1. If f and h are functions in $L^1(\mathbb{R}) \bigcap L^2(\mathbb{R})$, then

$$\int_{-\infty}^{\infty} f(t)h^*(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\omega)\widehat{h}^*(\omega)d\omega, \ (\textit{Parseval relation}) \tag{2.2.16}$$

In the particular case when $h = f$ the above relation becomes

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\widehat{f}(\omega)\right|^2 d\omega, \ (\textit{Plancherel relation}) \tag{2.2.17}$$

Proof of Theorem 1. Recall that

$$\begin{aligned} Wf(a,b) &= \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)\,dt, \\ &= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{|a|}}\psi^*\left(-\left[\frac{b-t}{a}\right]\right)\,dt. \end{aligned} \tag{2.2.18}$$

Note that this transform can be viewed as a convolution. That is, if we denote $\overline{\psi}_a(t) = \frac{1}{\sqrt{|a|}}\psi^*(-\frac{t}{a})$, then $Wf(a,b) = f * \overline{\psi}_a(b)$. Now to verify the inversion formula we denote the right-hand side by

$$g(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty} Wf(a,b)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-a}{a}\right)\,db\frac{da}{a^2}, \tag{2.2.19}$$

which can be rewritten in terms of convolution as

$$\begin{aligned} g(t) &= \frac{1}{C_\psi}\int_{-\infty}^{\infty} (f * \overline{\psi}_a) * \overline{\psi}_a(t)\,\frac{da}{a^2}, \\ &= \frac{1}{C_\psi}\int_{-\infty}^{\infty} f * \overline{\psi}_a * \overline{\psi}_a(t)\,\frac{da}{a^2}. \end{aligned} \tag{2.2.20}$$

To prove that $f(t) = g(t)$, it suffices to show that their corresponding Fourier transforms are equal. Consequently, taking the Fourier transform of both sides we get

$$\widehat{g}(\omega) = \frac{1}{C_\psi}\int_{-\infty}^{\infty} \widehat{f}(\omega)\widehat{\overline{\psi}}_a(\omega)\widehat{\psi}_a(\omega)\,\frac{da}{a^2}. \tag{2.2.21}$$

Now recall that $\overline{\psi}_a(t) = \frac{1}{\sqrt{|a|}}\psi^*(-\frac{t}{a})$, so taking the Fourier transform we obtain

$$\widehat{\overline{\psi}}_a(\omega) = \sqrt{|a|}\widehat{\psi}^*(a\omega), \tag{2.2.22}$$

and

$$\widehat{\psi}_a(\omega) = \sqrt{|a|}\widehat{\psi}(a\omega). \tag{2.2.23}$$

Consequently

$$\begin{aligned}
\widehat{g}(\omega) &= \frac{1}{C_\psi}\int_{-\infty}^{\infty}\widehat{f}(\omega)\sqrt{|a|}\widehat{\psi}^*(a\omega)\sqrt{|a|}\widehat{\psi}(a\omega)\,\frac{da}{a^2},\\
&= \frac{1}{C_\psi}\int_{-\infty}^{\infty}\widehat{f}(\omega)\widehat{\psi}^*(a\omega)\widehat{\psi}(a\omega)\,\frac{da}{|a|},\\
&= \frac{\widehat{f}(\omega)}{C_\psi}\int_{-\infty}^{\infty}\left|\widehat{\psi}(a\omega)\right|^2\,\frac{da}{|a|}.
\end{aligned} \tag{2.2.24}$$

Now, make the substitution $u = a\omega$ to obtain

$$\widehat{g}(\omega) = \widehat{f}(\omega). \tag{2.2.25}$$

For the second relation we apply the Parseval theorem to the right-hand side. Consequently,

$$\begin{aligned}
\frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|Wf(a,b)|^2\,db\frac{da}{a^2} &= \frac{1}{C_\psi}\int_{-\infty}^{\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\widehat{f}(\omega)\widehat{\psi}_a(\omega)\right|^2\,d\omega\,\frac{da}{a^2},\\
&= \frac{1}{C_\psi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\widehat{f}(\omega)\right|^2\left|\sqrt{|a|}\widehat{\psi}(a\omega)\right|^2\,d\omega\,\frac{da}{a^2}.
\end{aligned} \tag{2.2.26}$$

Now we interchange the order of integration, and make the substitution $u = a\omega$, to get

$$\begin{aligned}
\frac{1}{C_\psi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\widehat{f}(\omega)\right|^2\left|\sqrt{|a|}\widehat{\psi}(a\omega)\right|^2\,d\omega\,\frac{da}{a^2} &= \frac{1}{C_\psi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\widehat{f}(\omega)\right|^2\int_{-\infty}^{\infty}\frac{\left|\widehat{\psi}(u)\right|^2}{|u|}\,du\,d\omega,\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\widehat{f}(\omega)\right|^2\,d\omega.
\end{aligned} \tag{2.2.27}$$

Applying the Parseval theorem one more time yields the left-hand side of the formula, $\int_{-\infty}^{\infty}|f(t)|^2\,dt$.

Remark 1. When $Wf(a,b)$ is only known for some $a < a_0$, in order then to invert the transform one needs to include the scaling function, φ to be able to cover the part where $a > a_0$. In such a situation the inversion formula becomes

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{a_0}Wf(.,a)*\psi_a(t)\,\frac{da}{a^2} + \frac{1}{C_\psi}\frac{1}{|a_0|}Lf(.,a_0)*\phi_{a_0}(t). \tag{2.2.28}$$

where $Lf(.,a) = \langle f(t), \frac{1}{\sqrt{|a|}}\phi(\frac{t-b}{a})\rangle$.

Let us now prove this version of the inversion formula. In the same fashion, denote the right-hand side of the formula by $g(t)$, then it suffices to show that the corresponding Fourier transforms are equal. Indeed,

$$\begin{aligned} \widehat{g}(\omega) &= \frac{1}{C_\psi}\int_{-\infty}^{a_0}\widehat{f}(\omega)\widehat{\overline{\psi}}_a(\omega)\widehat{\psi}_a(\omega)\,\frac{da}{a^2}+\frac{1}{C_\psi}\frac{1}{|a_0|}\widehat{f}(\omega)|a_0|\widehat{\overline{\phi}}_{a_0}(\omega)\widehat{\phi}_{a_0}(\omega), \\ &= \frac{1}{C_\psi}\int_{-\infty}^{a_0}\widehat{f}(\omega)\widehat{\psi}^*(a\omega)\widehat{\psi}(a\omega)\,\frac{da}{|a|}+\frac{1}{C_\psi}\frac{1}{|a_0|}\widehat{f}(\omega)|a_0|\left|\widehat{\phi}(a_0\omega)\right|^2, \\ &= \frac{\widehat{f}(\omega)}{C_\psi}\int_{-\infty}^{a_0}\left|\widehat{\psi}(a\omega)\right|^2\frac{da}{|a|}+\frac{\widehat{f}(\omega)}{C_\psi}\left|\widehat{\phi}(a_0\omega)\right|^2. \end{aligned} \tag{2.2.29}$$

Finally, recalling the formula

$$\left|\widehat{\phi}(\omega)\right|^2 = \int_\omega^\infty \frac{\left|\widehat{\psi}(r)\right|^2}{r}\,dr, \tag{2.2.30}$$

yields

$$\begin{aligned} \widehat{g}(\omega) &= \frac{\widehat{f}(\omega)}{C_\psi}\int_{-\infty}^{a_0\omega}\left|\widehat{\psi}(r)\right|^2\frac{dr}{|r|}+\frac{\widehat{f}(\omega)}{C_\psi}\left|\widehat{\phi}(a_0\omega)\right|^2, \\ &= \widehat{f}(\omega). \end{aligned} \tag{2.2.31}$$

2.2.6 Discrete Wavelet Transform

Suppose that we have observations $X = (x_0, x_1, x_2, \cdots, x_{N-1})$ that may be *i.i.d.*, and take $N = 2^M$. Then the discrete wavelet transform of x with respect to ψ is

$$d_{j,k}^{(\psi)} = \sum_{t=0}^{N-1} x_t\psi_{j,k}(t) \tag{2.2.32}$$

This transform is computed for $j = 0, 1, 2, \cdots, M-1$ and $k = 0, 1, 2, \cdots, 2^j-1$. Since we have N observations and only $N-1$ coefficients we need one more, denote it $d_{-1,0}$.

The data X may be associated to a function f on the interval $[0, 1)$ such that

$$f(t) = \sum_{k=0}^{N-1} x_k \mathbf{1}_{\{\frac{k}{2^M} \le t < \frac{k+1}{2^M}\}} \tag{2.2.33}$$

So the discrete wavelet expansion of f is given by

$$f(t) = d_{-1,0}\varphi(t) + \sum_{j=0}^{M-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{2.2.34}$$

In practice we do not consider all the resolution levels, M, but a number J, which corresponds to the coarsest scale, 2^{-J}, or the smooth part of the data. Thus, the the discrete wavelet expansion of f becomes

$$f(t) = \sum_{k=0}^{2^J-1} c_{J,k}\varphi_{J,k}(t) + \sum_{j=J}^{M-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{2.2.35}$$

Notice that $c_{J,k}$ capture the low frequency oscillations, whereas $d_{j,k}$ capture high frequency. In addition the coefficients $d_{M-1,k}$ represent fine scale, details, and $c_{J,k}$, $d_{J,k}$ correspond to the coarsest scale, or smoothness.

2.2.7 Wavelet Series Versus Fourier Series

The wavelet series are often compared to the Fourier series. Historically, there was need to device a better tool than Fourier series representation in the sense that it takes an infinite number of terms to represent a function but for a practical matter we can only use a finite number of terms. This Fourier series must then be truncated, and this truncation will produce an error. Thus we must try to balance between the number of terms to keep and how much error we are willing to tolerate. In order to achieve satisfactory results (accuracy) a greater number of terms is needed, and this will require more computer time and storage space. Another disadvantage with the Fourier series is that although it represents the frequency of a function well, it does a poor job preserving that function's localized properties. Mathematicians along with physicists had to wait until the 1980's to see their prayer answered when a new type of series called *wavelet* series was invented. The main difference between wavelets and Fourier series is that wavelets are localized in both time and frequency, whereas the standard Fourier transform is only localized in frequency. This can be explained by the fact that while the Fourier series depend on a single basis (sine/cosine) which represents frequencies well but whose support is not localized, wavelet series give us an infinite number of bases to choose from so we can pick the best basis for a particular function. As a result wavelets often give a better signal representation thanks to *multiresolution analysis*, with balanced resolution at any time and frequency. The closest type of Fourier series to the wavelet series is what is called the *short time Fourier series*, in the sense that it is localized in time as well. Another advantage of wavelet over Fourier series is their computational efficiency, taking only $O(N)$ compared to $O(N \log N)$ with the Fourier series, where N is the size of the signal.

CHAPTER 3: ADAPTIVE NONPARAMETRIC EMPIRICAL BAYES ESTIMATION VIA WAVELETS SERIES

3.1 Formulation of the Problem

Consider the following setting, one observes independent two-dimensional random vectors $(X_1, \theta_1), \cdots,$ (X_n, θ_n), where each θ_i is distributed according to some unknown prior pdf g and, given $\theta_i = \theta$ the observation X_i has the known conditional density function $q(x \mid \theta)$, so that each pair (X_i, θ_i) has an absolutely continuous distribution with the density function $q(x \mid \theta)g(\theta)$. In each pair the first component is observable, but the second is not. After the $(n+1)$-th observation $y \equiv X_{n+1}$ is taken, the goal is to estimate $t \equiv \theta_{n+1}$.

If the prior density $g(\theta)$ were known, then the standard Bayes estimator of θ_{n+1} would be given by the following equation

$$t(y) = \frac{\int_{-\infty}^{\infty} \theta q(y \mid \theta)g(\theta)d\theta}{\int_{-\infty}^{\infty} q(y \mid \theta)g(\theta)d\theta} \tag{3.1.1}$$

Since the prior density is unknown, an EB estimator $\widehat{t}(y; X_1, X_2, X_3, \cdots, X_n)$ is to be used.
Denote

$$p(y) = \int_{-\infty}^{\infty} q(y \mid \theta)g(\theta)d\theta, \tag{3.1.2}$$

$$\Psi(y) = \int_{-\infty}^{\infty} \theta q(y \mid \theta)g(\theta)d\theta. \tag{3.1.3}$$

Hence $t(y)$ can be rewritten as

$$t(y) = \Psi(y)/p(y). \tag{3.1.4}$$

There is a variety of methods which allow to estimate $t(y)$ on the basis of observations $y; X_1, \cdots, X_n$. After Robbins (1955, 1964) formulated EB estimation approach, many statisticians have been working on developing EB methods. The comprehensive list of references as well as numerous examples of applications of EB techniques can be found in Carlin and Louis (2000) or Maritz and Lwin (1989).

As stated in Chapter 2, in nonparametric EB estimation, prior distribution is completely unspecified. One of the approaches to nonparametric EB estimation is based on estimation of the numerator and the denominator in the ratio in (3.1.4). This approach was introduced by Robbins (1955, 1964) himself and later developed by a number of authors (see, e.g., Brown and Greenshtein (2009), Datta (1991, 2000), Ma and Balakrishnan (2000), Nogami (1988), Pensky (1997a,b), Raykar and Zhao (2011), Singh (1976, 1979) and Walter and Hamedani (1991) among others). The method provides estimators with good convergence rates, however, it requires relatively tedious three-step procedure: estimation of the top and the bottom of the fraction and then the fraction itself.

Wavelets provide an opportunity to construct adaptive wavelet–based EB estimators with better computational properties in this framework (see, e.g., Huang (1997) and Pensky (1998, 2000, 2002)) but the necessity of estimation of the ratio in (3.1.4) remains. Another nonparametric approach developed in Jiang and Zhang (2009), is based on application of nonparametric MLE technique which is computationally extremely demanding.

In 1983, Robbins introduced a much more simple, local nonparametric EB method, linear EB estimation. Robbins (1983) suggested to approximate Bayes estimator $t(y)$ locally by a linear function of y and to determine the coefficients of $t(y)$ by minimizing the expected squared difference between $t(y)$ and θ, with subsequent estimation of the coefficients on the basis of observations $\{X_1, \cdots, X_n\}$. The technique is extremely efficient computationally and was immediately put to practical use, for instance, for prediction of the finite population mean (see, e.g., Ghosh and Meeden (1986), Ghosh and Lahiri (1987) and Karunamuni and Zhang (2003)).

However, a linear EB estimator has a large bias since, due to its very simple form, it has a limited ability to approximate the Bayes estimator $t(y)$. For this reason, linear EB estimators are optimal only in the class of estimators *linear* in y. To overcome this defect, Pensky and Ni (2000) extended approach of Robbins (1983) to approximation of $t(y)$ by algebraic polynomials. However, although the polynomial-based EB estimation provides significant improvement in the convergence rates in comparison with the linear EB estimator, the system of linear equations resulting from the method is badly conditioned which leads to computational difficulties and loss of precision.

To overcome those difficulties, Pensky and Alotaibi (2005) proposed to replace polynomial approximation of the Bayes estimator $t(y)$ by its approximation via wavelets, in particular, by expansion over scaling functions at the resolution level m. The method exploits de-correlating property of wavelets and leads to a low-dimensional well-posed sparse system of linear equations. The paper also treated the issue of local asymptotic optimality as $n \to \infty$: if the resolution level is selected correctly, in accordance with the smoothness of the Bayes estimator, then the suggested EB estimator attains optimal convergence rates.

However, smoothness of the Bayes estimator $t(y)$ is hard to assess. For this reason, the EB estimator of Pensky and Alotaibi (2005) is non-adaptive. One of the possible ways of achieving adaptivity would be to replace the linear scaling function based approximation by a traditional wavelet expansion with subsequent thresholding of wavelet coefficients. The deficiency of this approach, however, is that it yields the system of equations which is much less sparse and is growing in size with the number of observations n.

The present chapter has two main objectives. The first one is to derive lower bounds for the postrior risk of a nonparametric empirical Bayes estimator. In spite of a fifty years long history of empirical Bayes methods, general lower bounds for the risk of an empirical Bayes estimators have not been derived so far. In particular, Penskaya (1995) obtain lower bounds for the posterior risk of nonparametric empirical Bayes estimators of a location parameter. Li, Gupta and Liese (2005) obtained lower bounds for the risk of empirical Bayes estimators in the exponential families. However, since their lower bound is of the form C/n, practically no estimator can attain this lower bound. Construction of the lower bounds was attempted also in the empirical Bayes linear loss two-action problem in the case of continuous one-parameter exponential family. Karunamuni (1996) published the paper on the subject but his results were proved to be inaccurate, at least in the case of the normal distirbution, when Liang (2000) constructed an estimator with the convergence rates below the lower bound for the risk. Pensky (2003) derived lower bounds for the loss in the empirical Bayes two-action problem involving normal means. However, no general theory has ever been attempted so far. In what follows, we construct lower bounds for the risk of an empirical Bayes estimator under a general assumption that the marginal density $p(x)$ given by formula (3.1.2) is continuously differentiable in the neighborhood of y.

The second purpose of this chapter is to provide an adaptive version of the wavelet EB estimator developed in Pensky and Alotaibi (2005). In particular, we preserve the structure of the linear structure of the estimator. However, since expansion over scaling functions at the resolution level m leads to excessive variance when resolution level m is too high and disproportionately large bias when m is too small, we choose the resolution level using Lepski method introduced in Lepski (1991) and further developed in Lepski, Mammen and Spokony (1997). The resulting estimator is adaptive and attains optimal convergence rates (within a logarithmic factor of n). In addition, it has an advantage of computational efficiency since it is based on the solution of low-dimensional sparse system of linear equations the matrix of which tends to a scalar multiple of an identity matrix as the scale m grows. The theory is supplemented by numerous examples that demonstrate how the estimator can be implemented for various types of distribution families.

3.2 EB estimation algorithm

In order to construct an estimator of $t(y)$ defined in (3.1.4), choose twice continuously differentiable scaling function φ with bounded support and s vanishing moments, so that

$$\begin{aligned} \text{supp } \varphi &\in [M_1, M_2]. & (3.2.1) \\ \int_{-\infty}^{\infty} x^{\aleph} \sum_{k \in Z} \varphi(x-k)\varphi(z-k)dx &= z^{\aleph}, \quad 0 \leq \aleph \leq s-1, & (3.2.2) \end{aligned}$$

(see e.g., Walter and Shen (2001)).

Approximate $t(y)$ by a wavelet series, or some fixed $m \geq 0$,

$$t_m(y) = \sum_{k \in \mathbf{Z}} a_{m,k}\, \varphi_{m,k}(y) \tag{3.2.3}$$

where $\varphi_{m,k}(y) = 2^{m/2}\varphi(2^m y - k)$, and estimate coefficients of $t(y)$ by minimizing the global mean squared difference

$$\min_{a_{m,k}} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\sum_{k \in \mathbf{Z}} a_{m,k}\varphi_{m,k}(y) - z \right]^2 q(y|z)g(z)dz\ dy \right\}. \tag{3.2.4}$$

Taking derivatives of the last expression with respect to $a_{m,j}$ and equating them to zero, we obtain the system of linear equations

$$B_m a_m = c_m \tag{3.2.5}$$

with

$$\begin{aligned} (B_m)_{j,k} = B_{j,k} &= \int_{-\infty}^{\infty} \varphi_{m,k}(x)\varphi_{m,j}(x)p(x)dx = \mathbf{E}\left[\varphi_{m,k}(X)\varphi_{m,j}(X)\right], & (3.2.6) \\ c_j = &= \int_{-\infty}^{\infty} \varphi_{m,j}(x)\Psi(x)dx. & (3.2.7) \end{aligned}$$

Here and in what follows we use the symbol $\mathbf{E}$ for expectation over the distribution of $X_1, X_2, \cdots, X_n$. The expectations over any other distributions are represented in integral forms. Also, we supress index m in notations of matrix $B_m = B$ and vector $c_m = c$ unless this leads to a confusion.

System (3.2.5) is an infinite system of equations. However, since we are interested in estimating $t(x)$ locally at $x = y$, we shall keep only indices $k, j \in K_{m,y}$ where

$$K_{m,y} = \{k \in Z : \ 2^m y - M_2 - s(M_2 - M_1) \le k \le 2^m y - M_1 + s(M_2 - M_1)\} \tag{3.2.8}$$

where r will be determined later. Observe that really expansion (3.2.3) contains just coefficients $a_{m,k}$ with $2^m y - M_2 \le k \le 2^m y - M_1$, however, for evaluation of these coefficients we need to keep more terms in the system of equations (3.2.5) (see Lemma A.3 in Pensky and Alotaibi (2005) for more detail).

The entries (3.2.6) of the matrix B are unknown and can be estimated by sample means

$$\widehat{b}_{j,k} = n^{-1} \sum_{l=1}^{n} \left[\varphi_{m,k}(X_l)\varphi_{m,j}(X_l)\right]. \tag{3.2.9}$$

In order to estimate c_j, find functions $u_{m,j}(x)$ such that for any θ

$$\int_{-\infty}^{\infty} q(x \mid \theta) u_{m,j}(x) dx = \int_{-\infty}^{\infty} \theta q(x \mid \theta) \varphi_{m,j}(x) dx. \tag{3.2.10}$$

Then, multiplying both sides of (4.1.5) by $g(\theta)$ and integrating over θ, we obtain

$$\mathbf{E} u_{mj}(X) = \int_{-\infty}^{\infty} u_{mj}(x) p(x) dx = \int_{-\infty}^{\infty} \varphi_{m,j}(x) \Psi(x) dx = c_j. \tag{3.2.11}$$

Note that functions $u_{m,j}(x)$ are the same functions which appear in the wavelet estimator of the numerator $\Psi(y)$ of the EB estimator (3.1.4), therefore, the estimator considered herein can be constructed whenever wavelet EB estimation is possible (see e.g. Pensky (1997, 1998)). Solutions of equation (4.1.5) can be easily obtained, for example, when $q(x \mid \theta)$ is a location parameter family, scale parameter family, one-parameter exponential family or a family of uniform distributions (see Pensky (1998, 2002)). Later in this discussion we consider in detail the case when θ is a location parameter.

Once functions $u_{m,j}(x)$ are derived, coefficients c_j can be estimated by

$$\widehat{c_j} = n^{-1} \sum_{l=1}^{n} u_{m,j}(X_l) \tag{3.2.12}$$

and system (3.2.5) replaced by $\widehat{B}\widehat{a} = \widehat{c}$. However, though estimators $\hat{B}$ and $\hat{c}$ converge in mean squared sense to B and c, respectively, the estimator $\widehat{a} = \widehat{B}^{-1}\widehat{c}$ may not even have finite expectation. To understand this fact, note that both $\widehat{B}$ and $\widehat{c}$ are asymptotically normal. In one dimensional case, the ratio of two normal random variables has Cauchy distribution and hence does not have finite mean. In multivariate case the difficulty remains. To ensure that the estimator of a has finite expectation, we choose $\delta = \delta_m > 0$ and construct an estimator of a of the form

$$\widehat{a}_\delta = (\widehat{B} + \delta I)^{-1}\widehat{c} \tag{3.2.13}$$

where I is the identity matrix. Observe that matrix $\widehat{B}$ is nonnegative definite, so that $\widehat{B} + \delta I$ is a positive definite matrix and, hence, is nonsingular. Solution $\widehat{a}_\delta$ is used for construction of the EB estimator

$$\widehat{t}(y) = \sum_{k \in K_{m,y}} (\widehat{a}_\delta)_{m,k}\, \varphi_{m,k}(y). \tag{3.2.14}$$

3.3 The Prior and the Posterior Risks

An EB estimator $\widehat{t}(y)$ may be characterized by the posterior risk

$$R(y;\hat{t}) = (p(y))^{-1} E \int_{-\infty}^{\infty} (\hat{t}(y) - \theta)^2 q(y|\theta) g(\theta) d\theta \tag{3.3.1}$$

which can be partitioned into two components. The first component of this sum is

$$R(y;t(y)) = \inf_f R(y;f(y)) = (p(y))^{-1} \int_{-\infty}^{\infty} (t(y) - \theta)^2 q(y|\theta) g(\theta) d\theta, \tag{3.3.2}$$

which is independent of $\hat{t}(y)$ and represents the posterior risk of the Bayes estimator (3.1.1). Thus we shall judge EB estimator by the second component

$$\hat{R}_n(y) = E(\hat{t}(y) - t(y))^2. \tag{3.3.3}$$

It must be noted that often the quality of the EB estimator is described by

$$\mathbf{E}\hat{R}_n(y) = \int_{-\infty}^{\infty} \hat{R}_n(y) p(y) dy, \tag{3.3.4}$$

which is the difference between the prior risk

$$\mathbf{E} \int_{-\infty}^{\infty} R(y;\hat{t}(y)) p(y) dy \tag{3.3.5}$$

of the EB estimator $\hat{t}(y)$ and the prior risk

$$\int_{-\infty}^{\infty} R(y; t(y))p(y)dy = \inf_{f} \int_{-\infty}^{\infty} R(y; f(y))p(y)dy \tag{3.3.6}$$

of the orresponding Bayes estimator $t(y)$. However, the risk function (3.3.3) has several advantages compared with $\mathbf{E}\hat{R}_n(y)$.

First, $\hat{R}_n(y)$ enables one to calculate the mean squared error for the given observation y which is the quantity of interest. Note that the wavelet series (3.2.14) is local in a sense that coefficients $(\widehat{a}_\delta)_{m,k}$ change whenever y changes, hence, working with a local measure of the risk makes much more sense. Using the prior risk for the estimator which is local in nature prevents one from seeing advantages of this estimator. Second, by using the risk function (3.3.3) we eliminate the influence on the risk function of the observations having very low probabilities. So, the use of $\widehat{R}_n(y)$ provides a way of getting EB estimators with better convergence rates. Third, posterior risk allows one to assess optimality of EB estimators for majority of familiar distribution families via comparison of the convergence rate of the estimator with the lower bounds for the risk derived in Pensky (1997). Finally, one can pursue evaluation of the prior risk for the estimator (3.2.14). The derivation will require assumptions similar to the ones in Pensky (1998) and can be accomplished by standard methods.

The error (3.3.3) is dominated by the sum of two components

$$R_n(y) \le 2(R_1(y) + R_2(y)) \tag{3.3.7}$$

where the first component $R_1 = R_1(y)$ is due to replacement of the Bayes estimator $t(y)$ by its wavelet representation (3.2.3), while $R_2 = R_2(y)$ is due to replacement of vector $a = B^{-1}c$ by $\widehat{a}_\delta$ given by (3.2.13):

$$R_1(y) = (t_m(y) - t(y))^2, \tag{3.3.8}$$

$$R_2(y) = \mathbf{E}\left[\sum_{k \in K_{m,y}} ((\widehat{a}_\delta)_{m,k} - a_{m,k})\varphi_{m,k}(y)\right]^2. \tag{3.3.9}$$

We shall refer to R_1 and R_2 as the systematic and the random error components, respectively. Since we are using the posterior risk, from now on we treat y as a fixed quantity throughout this chapter.

3.4 Minimax Lower Bounds

Our goal in this section is to construct a lower bound for the minimax risk on $(\mathcal{G}_r, d)$ where $\mathcal{G}_r$ is a class of r times continuously differentiable functions in the neighborhood Ω_y of y, and where d is a distance between f and g in $\mathcal{G}_r$ at the fixed point y:

$$d(f, g) = |f(y) - g(y)|\,. \tag{3.4.1}$$

In order to construct minimax lower bounds for (3.3.3) we follow procedure developed in Tsybakov (2008), in particular, we use Theorem 2.7 which we reformulate here for the case of squared risk.

Lemma 2. [Tsybakov (2007), Theorem 2.7] Assume that Ξ contains elements $\xi_0, \xi_1, \cdots, \xi_\Upsilon$, $\Upsilon \geq 1$, such that

(i) $d(\xi_\iota, \xi_\zeta) \geq 2\chi, \quad$ for $0 \leq \iota < \zeta \leq \Upsilon$;

(ii) the Kullback divergences $K(P_\iota, P_0)$ between the measures P_ι and P_0, with $P_\iota << P_0$, for $\iota = 1, \ldots, \Upsilon$, satisfy the inequality

$$K(P_\iota, P_0) \leq C_\Upsilon \tag{3.4.2}$$

with C_Υ is a positive constant. Then, for some absolute positive constant C_3, one has

$$\inf_{\widehat{\xi}} \sup_{\xi \in \ \Xi} \mathbf{E}_\xi \left[d^2(\widehat{\xi}, \xi) \right] \geq C_3 \chi^2. \tag{3.4.3}$$

Let y be a fixed point. Consider an r-times continuously differentiable pdf $p_0(x)$, and an r-times continuously differentiable kernel $k(\cdot)$ with supp $k = (-1, 1)$ and such that $\int k(z)dz = 0$. Let $p_0(\cdot)$ and $k(\cdot)$ satisfy the following assumptions:

Assumption A1: There exists $g_0(\theta)$ such that for any x

$$p_0(x) \int q(x \mid \theta) g_0(\theta) d\theta \tag{3.4.4}$$

Assumption A2: There exists a function $\psi_{h,y}(\theta)$ such that for any x, y and $h > 0$

$$k\left(\frac{x - y}{h}\right) = \int q(x \mid \theta) \psi_{h,y}(\theta) d\theta. \tag{3.4.5}$$

Assumption A3: Density $p_0(x)$ is such that for any x and y such that $|x-y| \le h$ and any $0 < \zeta \le \zeta_0$

$$p_0(x) > 2\zeta \left|\left|k\right|\right|_\infty . \tag{3.4.6}$$

Denote

$$\begin{aligned}
\Psi_0(x) &= \int \theta q(x \mid \theta) g_0(\theta) d\theta, \quad w_{h,y}(x) = \int \theta q(x \mid \theta)\psi_{h,y}(\theta) d\theta, && (3.4.7)\\
\rho_r(h) &= \left[\max_{1\le j \le r} \left| \frac{d^j}{dx^j}[w_{h,y}(x)] \right|_{x=y} \right]^{-1}, && (3.4.8)
\end{aligned}$$

and consider

$$\begin{aligned}
p_1(x) &= p_0(x) + \zeta k\left(\frac{x-y}{h}\right) && (3.4.9)\\
\Psi_1(x) &= \Psi_0(x) + \zeta w_{h,y}(x) && (3.4.10)
\end{aligned}$$

Now, consider equations (3.4.9) and (3.4.10) and let the Assumptions $1-3$ hold. Choose ζ such that $p_1(x)$ and $\Psi_1(x) \in \mathcal{G}_r$. That is, the $r-th$ derivatives of $p_1(x)$ and $\Psi_1(x)$ are bounded above. This is achieved by taking $\zeta = \zeta_0 \min\{h^r, \rho_r(h)\}$ for ζ_0 some constant independent of j, where $\rho_r(h)$ is defined in (3.4.8). Then, calculating the distance $d(t_1,t_0)$ at the fixed point y and and since $p_1(y) \ge p_0(y)/2$, we obtain

$$\begin{aligned}
d(t_1,t_0) &= \left| \frac{\Psi_0(y) + \zeta w_{j,k}(y)}{p_0(y) + \zeta k(0)} - \frac{\Psi_0(y)}{p_0(y)} \right| \\
&= \zeta \left| \frac{w_{j,k}(y)p_0(y) - \Psi_0(y)k(0)}{p_0(y)[p_0(y) + \zeta k(0)]} \right| \\
&\ge \zeta/2 \left| \frac{w_{j,k}(y)p_0(y) - \Psi_0(y)k(0)}{p_0^2(y)} \right| && (3.4.11)
\end{aligned}$$

Hence,

$$d(t_1,t_0) \ge \begin{cases} C\zeta, & \text{if} \quad |w_{h,y}(y)| \le C_0, \\ C\zeta|w_{h,y}(y)|, & \text{if} \quad \lim_{h\to 0} |w_{h,y}(y)| = \infty. \end{cases} \tag{3.4.12}$$

Therefore,

$$2\chi = \begin{cases} C\zeta, & \text{if} \quad |w_{h,y}(y)| \le C_0, \\ C\zeta|w_{h,y}(y)|, & \text{if} \quad \lim_{h\to 0} |w_{h,y}(y)| = \infty. \end{cases} \tag{3.4.13}$$

In order to apply Lemma 14 , one needs to verify condition (ii). Observe that

$$p_0(x_1, \cdots, x_n) = \prod_{i=1}^n p_0(x_i) \tag{3.4.14}$$

$$p_1(x_1, \cdots, x_n) = \prod_{i=1}^n \left[p_0(x_i) + \zeta k\left(\frac{x_i - y}{h}\right)\right] \tag{3.4.15}$$

Then, recalling the fact that for $x > -1$, $\log(1+x) \le x$ the Kullback divergences between p_1 and p_0 is given by

$$\begin{aligned} K(p_1, p_0) &= \int \cdots \int \log\left\{\prod_{i=1}^n p_1(x_i)/p_0(x_i)\right\} \prod_{i=1}^n p_1(x_i) dx_i \\ &= \sum_{i=1}^n \int \log\left\{\frac{p_0(x) + \zeta k\left(\frac{x-y}{h}\right)}{p_0(x)}\right\}\left\{p_0(x) + \zeta k\left(\frac{x-y}{h}\right)\right\} dx \\ &\le \sum_{i=1}^n \int h_j \frac{k\left(\frac{x-y}{h}\right)}{p_0(x)}\left\{p_0(x) + \zeta k\left(\frac{x-y}{h}\right)\right\} dx \\ &= n\zeta^2 \int \frac{k^2\left(\frac{x-y}{h}\right)}{p_0(x)} dx \end{aligned} \tag{3.4.16}$$

Apply Lemma 14 with ζ and h such that

$$n\zeta^2 h \le C_6, \tag{3.4.17}$$

Now, assume that

$$\rho_r(h) \le Ch^{r_1}, \quad |w_{h,y}(x)| \le C_0 h^{-r_2} \tag{3.4.18}$$

Then,

$$h \asymp n^{\frac{1}{2\max\{r,r_1\}+1}}$$

obtain

$$\chi^2 = Cn^{-\frac{2\max\{r,r_1\}-2r_2}{2\max\{r,r_1\}+1}} \tag{3.4.19}$$

Hence, the lower bounds of the (3.3.3) is summarized in the following Theorem.

Theorem 2. Let functions $p(x)$ and $\Psi(x)$ be $r \in [1/2, s-1]$ times continuously differentiable in the neighborhood Ω_y of y such that $\Omega_{m,y} \subseteq \Omega_y$ where $\Omega_{m,y}$ is defined in (3.6.11). Let Assumptions A1-A3 hold and assume, for some $r_1 \geq 0$ and $r_2 \geq 0$,

$$\rho_r(h) \leq Ch^{r_1}, \quad |w_{h,y}(x)| \leq C_0 h^{-r_2} \tag{3.4.20}$$

Then, for any y such that $p(y) > 0$, as $n \to \infty$,

$$R_n(y) = \inf_{\hat{t}} \sup_{g \in \mathcal{G}_r} \mathbf{E}(\hat{t}_m(y) - t(y))^2 \geq C n^{-\frac{2\max\{r,r_1\}-2r_2}{2\max\{r,r_1\}+1}}, \tag{3.4.21}$$

where C is an absolute constant independent of n. In particular, if $r_1 \leq r$ and $r_2 = 0$, then

$$R_n(y) = \mathbf{E}(\hat{t}_m(y) - t(y))^2 \geq C n^{-\frac{2r}{2r+1}}, \quad m, n \to \infty, \tag{3.4.22}$$

Remark 2. If $r_1 \leq r$ and $r_2 = 0$, then convergence rates are defined by behavior of $p(x)$ in the neighborhood of y. Otherwise, the rates are defined by behavior of $\Psi(x)$ in the neighborhood of y.

3.5 Supplementary Lemmas

In future, we shall need the following supplementary results.

Lemma 3. Let $\widehat{B}$, B, $\widehat{c}$ and c be defined as in (3.2.9), (3.2.6), (3.2.12) and (3.2.7), respectively. Then,

$$\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^{2l} = O\left(n^{-l} 2^{ml}\right), \quad l = 1, 2, 4, \tag{3.5.1}$$

Also

$$\mathbf{E}\,||\widehat{c} - c||^2 = O\left(n^{-1}\gamma_m^2\right), \tag{3.5.2}$$

$$\mathbf{E}\,||\widehat{c} - c||^4 = O\left(n^{-3}\left|\left|\gamma^{(2)}(m)\right|\right|^2 + n^{-2}\gamma_m^4\right), \tag{3.5.3}$$

and

$$\begin{aligned} \mathbf{E}\,||\widehat{c} - c||^8 &= O\left(n^{-7}\left|\left|\gamma^{(4)}(m)\right|\right|^2 + n^{-6}\left|\left|\gamma^{(2)}(m)\right|\right|^4 + n^{-6}\gamma_m^2\left|\left|\gamma^{(3)}(m)\right|\right|^2\right) \\ &+ O\left(n^{-5}\gamma_m^4\left|\left|\gamma^{(2)}(m)\right|\right|^2 + n^{-4}\gamma_m^8\right). \end{aligned} \tag{3.5.4}$$

where γ_m^2 is defined in (3.8.2).

Proof of Lemma 3. Recall that $\hat{b}_{j,k} - b_{j,k} = \frac{1}{n}\sum_{t=1}^{n}\eta_t$ where $\eta_t = \varphi_{m,k}(X_t)\varphi_{m,j}(X_t),\ t = 1, \cdots, n.$ Indeed taking the second moment we obtain

$$\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\eta_t\right]^2 = \frac{\mathbf{E}[\eta_t^2]}{n}, \tag{3.5.5}$$

which implies that

$$\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^2 \quad \leq \quad M^2 n^{-1}\mathbf{E}[\eta_t^2] \leq n^{-1}M^2\left[2\,||p||_\infty\,||\varphi||_\infty^2\,2^m\right] = O\left(n^{-1}2^m\right). \tag{3.5.6}$$

For $l = 2$, we take the fourth moment and apply Jensen's inequality

$$\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\eta_t\right]^4 \quad = \quad \frac{n\mathbf{E}[\eta_t^4] + n(n-1)\mathbf{E}^2[\eta_t^2]}{n^4} \leq 2\frac{n^2\mathbf{E}[\eta_t^4]}{n^4} \tag{3.5.7}$$

Consequently,

$$\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^4 \quad \leq \quad 2M^2\frac{\mathbf{E}[\eta_t^4]}{n^2} \leq 4M^2\left[\frac{||p||_\infty\,||\varphi||_\infty^4\,2^{2m}}{n^2}\right] = O\left(n^{-2}2^{2m}\right). \tag{3.5.8}$$

For $l = 4$, taking the eighth moments and applying Jensen's inequality we obtain

$$\begin{aligned}\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\eta_t\right]^8 \quad &= \quad n^{-8}\left[n\mathbf{E}[\eta_t^8] + \binom{n}{2}\left(\mathbf{E}[\eta_t^2]\mathbf{E}[\eta_t^6] + \mathbf{E}[\eta_t^4]^2\right) + \binom{n}{3}\mathbf{E}[\eta_t^4]\mathbf{E}[\eta_t^2]^2 + \binom{n}{4}\mathbf{E}[\eta_t^2]^4\right] \\ &\leq \quad n^{-8}\left[n\mathbf{E}[\eta_t^8] + \binom{n}{2}2\mathbf{E}[\eta_t^8] + \binom{n}{3}\mathbf{E}[\eta_t^8] + \binom{n}{4}\mathbf{E}[\eta_t^8]\right] \leq 5\frac{\mathbf{E}[\eta_t^8]}{n^4}\end{aligned} \tag{3.5.9}$$

which yields

$$\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^8 \quad \leq \quad 5M^2\frac{\mathbf{E}[\eta_t^8]}{n^4} \leq 10M^2\left[\frac{||p||_\infty\,||\varphi||_\infty^8\,2^{4m}}{n^4}\right] = O\left(n^{-4}2^{4m}\right). \tag{3.5.10}$$

For $l = 8$, taking the sixteenth moment and applying the Jensen's inequality several times yields

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\eta_t\right]^{16} &= n^{-16}\left[n\mathbf{E}[\eta_t^{16}] + \binom{n}{2}\left(\mathbf{E}[\eta_t^{6}]\mathbf{E}[\eta_t^{10}] + \mathbf{E}^2[\eta_t^{8}]^2 + \mathbf{E}[\eta_t^{4}]\mathbf{E}[\eta_t^{12}] + \mathbf{E}[\eta_t^{2}]\mathbf{E}[\eta_t^{14}]\right)\right] \\
&+ n^{-16}\binom{n}{3}\left(\mathbf{E}[\eta_t^{4}]\mathbf{E}^2[\eta_t^{6}] + \mathbf{E}^2[\eta_t^{4}]\mathbf{E}[\eta_t^{8}] + \mathbf{E}[\eta_t^{12}]\mathbf{E}^2[\eta_t^{2}] + \mathbf{E}[\eta_t^{4}]\mathbf{E}[\eta_t^{2}]\mathbf{E}[\eta_t^{10}]\right) \\
&+ n^{-16}\binom{n}{3}\mathbf{E}[\eta_t^{6}]\mathbf{E}[\eta_t^{2}]\mathbf{E}[\eta_t^{8}] \\
&+ n^{-16}\binom{n}{4}\left(\mathbf{E}^4[\eta_t^{4}] + \mathbf{E}^3[\eta_t^{2}]\mathbf{E}[\eta_t^{10}] + \mathbf{E}[\eta_t^{2}]\mathbf{E}[\eta_t^{6}]\mathbf{E}^2[\eta_t^{4}] + \mathbf{E}^2[\eta_t^{2}]\mathbf{E}[\eta_t^{8}]\mathbf{E}^2[\eta_t^{4}]\right) \\
&+ n^{-16}\binom{n}{4}\mathbf{E}^2[\eta_t^{6}]\mathbf{E}^2[\eta_t^{2}] \\
&+ n^{-16}\binom{n}{5}\left(\mathbf{E}^4[\eta_t^{2}]\mathbf{E}[\eta_t^{8}] + \mathbf{E}^3[\eta_t^{4}]\mathbf{E}^2[\eta_t^{2}] + \mathbf{E}^3[\eta_t^{2}]\mathbf{E}[\eta_t^{6}]\mathbf{E}[\eta_t^{4}]\right) \\
&+ n^{-16}\left[\binom{n}{6}\left(\mathbf{E}^5[\eta_t^{2}]\mathbf{E}[\eta_t^{6}] + \mathbf{E}^2[\eta_t^{4}]\mathbf{E}^4[\eta_t^{2}]\right) + \binom{n}{7}\left(\mathbf{E}^6[\eta_t^{2}]\mathbf{E}[\eta_t^{4}]\right)\right] \\
&+ n^{-16}\binom{n}{8}\mathbf{E}^8[\eta_t^{2}] \le 22\frac{n^8\mathbf{E}[\eta_t^{16}]}{n^{16}} \qquad (3.5.11)
\end{aligned}$$

This leads to

$$\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^{16} \le 22M^2\frac{n^8\mathbf{E}[\eta_t^{16}]}{n^{16}} \le 44M^2\left[\frac{||p||_\infty\,||\varphi||_\infty^{16}\,2^{8m}}{n^8}\right] \le O\left(n^{-8}2^{8m}\right). \qquad (3.5.12)$$

Now to prove (3.5.2)–(3.5.4), recall $\widehat{c}_k - c_k = \frac{1}{n}\sum_{t=1}^{n}\xi_t$ where $\xi_t = u_{mj}(X_t)$, $t = 1, \cdot, n$. Thus, taking the second moment, we derive

$$\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\xi_t\right]^2 = \frac{\mathbf{E}[\xi_t^2]}{n} \qquad (3.5.13)$$

Consequently, using (3.8.1) and (3.8.2), we obtain

$$\mathbf{E}\,||\widehat{c} - c||^2 \le M\frac{\mathbf{E}[\xi_t^2]}{n} \le M\left[\frac{2\,||p||_\infty\,\gamma_m^2}{n}\right] \le O\left(n^{-1}\gamma_m^2\right) \qquad (3.5.14)$$

which proves (3.5.2). Now, taking the fourth moment and using Jensen's inequality we obtain

$$\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\xi_t\right]^4 = \frac{n\mathbf{E}[\xi_t^4] + n(n-1)\mathbf{E}^2[\xi_t^2]}{n^4} \le 2\frac{\mathbf{E}[\xi_t^4]}{n^2}. \qquad (3.5.15)$$

Thus, using (3.8.1) and (3.8.2), yields

$$\mathbf{E}\,||\widehat{c}-c||^4 \quad \le \quad 2M\frac{\mathbf{E}[\xi_t^4]}{n^2} \le M\left[\frac{2\,||p||_\infty\,\left|\left|\gamma^{(2)}(m)\right|\right|^2}{n^2}\right] = O\left(n^{-2}\left|\left|\gamma^{(2)}(m)\right|\right|^2\right). \tag{3.5.16}$$

Finally, taking the eighth moment and using the Jensen's inequality yields

$$\begin{aligned}\mathbf{E}\left[\frac{1}{n}\sum_{t=1}^{n}\xi_t\right]^8 &= n^{-8}\left[n\mathbf{E}[\xi_t^8] + \binom{n}{2}\left(\mathbf{E}[\xi_t^2]\mathbf{E}[\xi_t^6] + \mathbf{E}[\xi_t^4]^2\right) + \binom{n}{3}\mathbf{E}[\xi_t^4]\mathbf{E}[\xi_t^2]^2 + \binom{n}{4}\mathbf{E}[\xi_t^2]^4\right] \\ &\le 5\frac{\mathbf{E}[\xi_t^8]}{n^4} \end{aligned} \tag{3.5.17}$$

which yields

$$\mathbf{E}\,||\widehat{c}-c||^8 \quad \le \quad 5M\frac{\mathbf{E}[\xi_t^8]}{n^4} \le M\left[\frac{2\,||p||_\infty\,\left|\left|\gamma^{(4)}(m)\right|\right|^2}{n^4}\right] = O\left(n^{-4}\left|\left|\gamma^{(4)}(m)\right|\right|^2\right) \tag{3.5.18}$$

and completes the proof of Lemma 3.

Lemma 4. Let a and $\widehat{a}_\delta$ be defined by (3.2.5) and (3.2.13) respectively, and

$$\Omega_B = \left\{\omega : \left|\left|\widehat{B}-B\right|\right| > 0.5\,\left|\left|B^{-1}\right|\right|^{-1}\right\} \tag{3.5.19}$$

Then

$$\begin{aligned}||\widehat{a}_\delta - a|| &\le (2M+4\delta M^2)\,||\widehat{c}-c|| + \frac{2}{\delta}\,||c||\,\mathbf{1}\left(\Omega_B\right) \\ &+ 8M\,||\widehat{c}-c||\,\left|\left|\widehat{B}-B\right|\right| + \frac{2}{\delta}\,||\widehat{c}-c||\,\mathbf{1}\left(\Omega_B\right) \\ &+ 4\delta M^2\,||c|| + 8M^2\,||c||\,\left|\left|\widehat{B}-B\right|\right| \end{aligned} \tag{3.5.20}$$

Proof of Lemma 4. Recall that $a = B^{-1}c$ and $\widehat{a}_\delta = (\widehat{B}+\delta I)^{-1}\widehat{c}$. Then by the properties of the norm

$$||\widehat{a}_\delta - a|| \le \left|\left|B^{-1}\right|\right|\,||\widehat{c}-c|| + \left|\left|\widehat{B}_\delta^{-1} - B^{-1}\right|\right|\,||c|| + \left|\left|\widehat{B}_\delta^{-1} - B^{-1}\right|\right|\,||\widehat{c}-c|| \tag{3.5.21}$$

Also

$$\left|\left|\widehat{B}_\delta^{-1} - B^{-1}\right|\right| \le \left|\left|\widehat{B}_\delta^{-1} - B_\delta^{-1}\right|\right| + \left|\left|B_\delta^{-1} - B^{-1}\right|\right| \tag{3.5.22}$$

In addition, $B^{-1} = I + 2^{-m}V_1 + o(2^{-m})$, therefore taking the norm we obtain

$$\begin{aligned} \left|\left|B^{-1}\right|\right| &= \left|\left|p^{-1}(y)I + 2^{-m}V_1 + \cdots\right|\right| \\ &\leq \left|\left|p^{-1}(y)I\right|\right| + \left|\left|2^{-m}V_1\right|\right| \\ &\leq M(p^{-1}(y) + O(2^{-m})) \\ &\leq 2Mp^{-1}(y) \end{aligned} \tag{3.5.23}$$

or

$$\left|\left|B^{-1}\right|\right| \leq 2Mp^{-1}(y) \tag{3.5.24}$$

Now, for the first part of the right hand side in (3.5.22) we have

$$\begin{aligned} \left|\left|\widehat{B}_\delta^{-1} - B_\delta^{-1}\right|\right| &\leq 2\left|\left|B^{-1}\right|\right|^2 \left|\left|\widehat{B} - B\right|\right| \mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| < \frac{1}{2}\left|\left|B^{-1}\right|\right|^{-1}\right) \\ &+ \left|\left|\widehat{B}_\delta^{-1} - B_\delta^{-1}\right|\right| \mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{2}\left|\left|B^{-1}\right|\right|^{-1}\right) \\ &\leq 2\left|\left|B^{-1}\right|\right|^2 \left|\left|\widehat{B} - B\right|\right| + \frac{2}{\delta}\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{2}\left|\left|B^{-1}\right|\right|^{-1}\right) \end{aligned} \tag{3.5.25}$$

Consequently,

$$\left|\left|\widehat{B}_\delta^{-1} - B_\delta^{-1}\right|\right| \leq 8M^2[p^{-1}(y)]^2 \left|\left|\widehat{B} - B\right|\right| + \frac{2}{\delta}\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{2}\ \left|\left|B^{-1}\right|\right|^{-1}\right) \tag{3.5.26}$$

For the second part of (3.5.22), and using (3.5.24) we obtain

$$\left|\left|B_\delta^{-1} - B^{-1}\right|\right| \leq \delta\left|\left|B^{-1}\right|\right|^2 \leq 4\delta M^2[p^{-1}(y)]^2 \tag{3.5.27}$$

Finally, using results (3.5.27), and (3.5.26) in (3.5.21) we derive

$$
\begin{aligned}
||\widehat{a}_\delta - a|| &\leq \left|\left|B^{-1}\right|\right| ||\widehat{c} - c|| \\
&+ ||c|| \left(8M^2[p^{-1}(y)]^2 \left|\left|\widehat{B} - B\right|\right| + \frac{2}{\delta}\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{2}\left|\left|B^{-1}\right|\right|^{-1}\right) + 4\delta M^2\right) \\
&+ ||\widehat{c} - c|| \left(8M^2 \left|\left|\widehat{B} - B\right|\right| + \frac{2}{\delta}\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{2}\left|\left|B^{-1}\right|\right|^{-1}\right) + 4\delta M^2[p^{-1}(y)]^2\right) \\
&\leq 2Mp^{-1}(y)\,||\widehat{c} - c|| + 8M^2[p^{-1}(y)]^2\,||c|| \left|\left|\widehat{B} - B\right|\right| + \frac{2}{\delta}\,||c||\,\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{p(y)}{4M}\right) \\
&+ 8M^2[p^{-1}(y)]^2\,||\widehat{c} - c|| \left|\left|\widehat{B} - B\right|\right| \\
&+ \frac{2}{\delta}\,||\widehat{c} - c||\,\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{p(y)}{4M}\right) + 4\delta M^2[p^{-1}(y)]^2\,||c|| \\
&+ 4\delta M^2[p^{-1}(y)]^2\,||\widehat{c} - c|| \qquad (3.5.28)
\end{aligned}
$$

which completes the proof of lemma 4.

Furthermore, recall $\Omega_B = \left\{\omega : \left|\left|\widehat{B} - B\right|\right| > 0.5\,\left|\left|B^{-1}\right|\right|^{-1}\right\}$ from Lemma 4, then, squaring (3.5.20), taking expectation and applying the Cauchy-Schwartz inequality leads to the following corollary.

Corollary 1. Let a and $\widehat{a}_\delta$ be defined by (3.2.5) and (3.2.13) respectively. Then,

$$
\begin{aligned}
\mathbf{E}\,||\widehat{a}_\delta - a||^2 &\leq 32M^2[p^{-1}(y)]^2\mathbf{E}\,||\widehat{c} - c||^2 + 32\delta^2 M^4[p^{-1}(y)]^4\,||c||^2 \\
&+ 128M^4[p^{-1}(y)]^4\,||c||^2\,\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^2 + 128M^4[p^{-1}(y)]^4\sqrt{|||\widehat{c} - c||^4\,\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^4} \\
&+ \frac{8}{\delta^2}\,||c||^2\,\Pr(\Omega_B) \\
&+ \frac{8}{\delta^2}\sqrt{||\widehat{c} - c||^4\,\Pr(\Omega_B)} \qquad (3.5.29)
\end{aligned}
$$

In addition, taking the fourth power, taking the expectation and applying the Cauchy-Schwartz inequality leads to a second corollary

Corollary 2.

$$
\begin{aligned}
\mathbf{E}\,||\widehat{a}_\delta - a||^4 &\leq 1024M^4[p^{-1}(y)]^4\mathbf{E}\,||\widehat{c} - c||^4 + 4096M^8[p^{-1}(y)]^8\,||c||^4\,\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^4 \\
&+ 4096M^8[p^{-1}(y)]^8\sqrt{\mathbf{E}\,||\widehat{c} - c||^8\,\mathbf{E}\left|\left|\widehat{B} - B\right|\right|^8} + 256\delta^4 M^8[p^{-1}(y)]^8\,||c||^4 \\
&+ \frac{64}{\delta^4}\,||c||^4\,\Pr(\Omega_B) \\
&+ \frac{64}{\delta^4}\sqrt{||\widehat{c} - c||^8\,\Pr(\Omega_B)} \qquad (3.5.30)
\end{aligned}
$$

3.6 The systematic error component.

For evaluation of the systematic error component R_1, let us introduce matrices U_h and U_h^* and vectors D_h and D_h^* with components

$$(U_h)_{k,l} = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)\varphi(z + 2^m y - l)dz, \tag{3.6.1}$$

$$(U_h^*)_{k,l} = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)\varphi(z + 2^m y - l)R_{p,h}(z)dz, \tag{3.6.2}$$

$$(D_h)_k = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)dz. \tag{3.6.3}$$

$$(D_h^*)_k = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)R_{\Psi,h}(z)dz, \tag{3.6.4}$$

where $R_{p,h}$ and $R_{\Psi,h}$ are respectively the remainders in the Taylor series expansions of $p(y + 2^{-m}z)$ and $\Psi(y + 2^{-m}z)$.

Lemma 5. Let the matrices U_h and U_h^* and vectors D_h and D_h^* be defined as in (3.6.1)- (3.6.4) respectively. Then

$$\lim_{m \longrightarrow \infty} |U_r^*(m) - U_r(m)| = 0, \tag{3.6.5}$$

Also

$$\lim_{m \longrightarrow \infty} |D_r^*(m) - D_r(m)| = 0. \tag{3.6.6}$$

Proof of Lemma 5 Recall that $(U_h)_{k,l} = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)\varphi(z + 2^m y - l)dz$ and $(U_h^*)_{k,l} = \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)\varphi(z + 2^m y - l)R_{p,r}(z)dz$, with $R_{p,r}$ being the remainder in the Taylor expansion of $P(y + \frac{z}{2^m})$. That is, $R_{p,r} = \frac{1}{(r-1)!}(\frac{1}{2^m})^r(\frac{1}{2^m})^{r-1}\int_0^z (z - \xi)^{r-1}p^{(r)}(y + \frac{\xi}{2^m})d\xi$. Indeed,

$$\begin{aligned} |(U_h^*)_{k,l} - (U_h)_{k,l}| &= \left| \int_{-\infty}^{\infty} z^h \varphi(z + 2^m y - k)\varphi(z + 2^m y - l)(R_{p,r-1} - 1)dz \right| \\ &\leq \int_{-\infty}^{\infty} \left| z^h \varphi(z + 2^m y - k) \right| \left| \varphi(z + 2^m y - l) \right| \left| (R_{p,r-1} - 1) \right| dz \\ &\leq \int_{-\infty}^{\infty} \left| z^h \phi(z + 2^m y - k) \right| \left| \varphi(z + 2^m y - l) \right| dz \\ &\leq \max_z \left| \varphi(z + 2^m y - l) \right| \int_{-\infty}^{\infty} \left| z^h \varphi(z + 2^m y - k) \right| dz \end{aligned} \tag{3.6.7}$$

Notice that when $m \longrightarrow \infty$, $\left|z^h\varphi(z+2^my-k)\right| \longrightarrow 0$. So the dominated convergence theorem applies to $\left|z^h\varphi(z+2^my-k)\right|$, and therefore

$$\begin{aligned}\lim_{m\longrightarrow\infty}|(U_h^*)_{k,l}-(U_h)_{k,l}| &\leq \max_z|\varphi(z+2^my-l)|\lim_{m\longrightarrow\infty}\int_{-\infty}^{\infty}\left|z^h\varphi(z+2^my-k)\right|dz \\ &= 0 \end{aligned} \tag{3.6.8}$$

which completes the proof of (3.6.5) of lemma 5.

For the second part of the lemma recall $(D_h)_k = \int_{-\infty}^{\infty} z^h\varphi(z+2^my-k)dz$ and $(D_h^*)_k = \int_{-\infty}^{\infty} z^h\varphi(z+2^my-k)R_{\Psi,r}(z)dz$, where $R_{\Psi,r}(z) = \int_0^z \frac{(z-\xi)^{r-1}}{(r-1)!}2^{-(2r-1)m}\Psi^{(r)}(y+\frac{\xi}{2^m})d\xi$. So

$$\begin{aligned}|(D_h^*)_k-(D_h)_k| &= \left|\int_{-\infty}^{\infty} z^h\varphi(z+2^my-k)(R_{\Psi,r}-1)dz\right| \\ &\leq \int_{-\infty}^{\infty}\left|z^h\varphi(z+2^my-k)\right||(R_{\Psi,r}-1)|\,dz \\ &\leq \max_z|(R_{\Psi,r}-1)|\int_{-\infty}^{\infty}\left|z^h\varphi(z+2^my-k)\right|dz \end{aligned} \tag{3.6.9}$$

Since $\left|z^h\varphi(z+2^my-k)\right| \longrightarrow 0$ as $m \longrightarrow \infty$ the dominated convergence theorem applies to $\left|z^h\varphi(z+2^my-k)\right|$. Concequently

$$\begin{aligned}\lim_{m\longrightarrow\infty}|(D_h^*)_{k,l}-(D_h)_{k,l}| &\leq \max_z|(R_{\Psi,r}-1)|\lim_{m\longrightarrow\infty}\int_{-\infty}^{\infty}\left|z^h\varphi(z+2^my-k)\right|dz \\ &= 0 \end{aligned} \tag{3.6.10}$$

Which completes the proof of (3.6.6) and Lemma 5.

Observe that U_h and D_h are independent of unknown functions $p(x)$ and $\Psi(x)$, and that $U_0 = I$ where I is the identity matrix. Denote

$$\Omega_{m,y} = \left\{x:\ |x-y| \leq 2^{-m}s(M_2-M_1)\right\}. \tag{3.6.11}$$

Then

$$
\begin{aligned}
B_{j,k} &= 2^m \int_{-\infty}^{\infty} \varphi(2^m x - k)\varphi(2^m x - j)p(x)dx \\
&= \int_{-\infty}^{\infty} \varphi(z + 2^m y - k)\varphi(z + 2^m y - j)p(y + 2^{-m}z)dz \\
&= \sum_{h=0}^{r} 2^{-mh}(h!)^{-1}p^{(h)}(y)U_h + o(2^{-mr}), \qquad (3.6.12)
\end{aligned}
$$

where $U_0 = I$, the identity matrix. Deriving a similar representation for c_k, we obtain an asymptotic expansions of matrix B and vector c via matrices U_h and vectors D_h, respectively, as $m \to \infty$

$$
B = p(y)I + \sum_{h=1}^{r} 2^{-mh}(h!)^{-1}p^{(h)}(y)U_h + o(2^{-mr}), \qquad (3.6.13)
$$

$$
c = 2^{-m/2} \sum_{l=0}^{r} 2^{-ml}(l!)^{-1}\Psi^{(l)}(y)D_l + o(2^{-mr}), \qquad (3.6.14)
$$

where I is the identity matrix. Formula (3.6.13) establishes that, for large m, matrix B is close to $p(y)I$, so the system (3.2.5) is well-conditioned. Furthermore, if $m \to \infty$, vector a in (3.2.3) tends to $2^{-m/2}[\Psi(y)/p(y)]D_0$ where $2^{-m/2}\sum_k (D_0)_k \varphi_{m,k}(y) = 1$ for any y. The latter implies that the systematic error goes to zero as $m \to \infty$. at a rate $O\left(2^{-mr}\right)$ and has the following asymptotic upper bound.

In order to prove this fact, we shall need the following statement.

Lemma 6. Matrix B^{-1} can be represented as

$$
B^{-1} = \sum_{h=0}^{r} 2^{-mh}V_h + o(2^{-mr}) \qquad (3.6.15)
$$

with $V_j = \sum \alpha_{k_1,k_2,\cdots,k_l} U_{k_1}U_{k_2}\cdots U_{k_l}$. Here, $l \le j$ and coefficients $\alpha_{k_1,k_2,\cdots,k_l}$ are polynomial functions of the derivatives $p^{(k_h)}(y)$ with $\sum_{h=1}^{l} k_h = j$ divided by powers of $p(y)$.

Proof of Lemma 6. Write B^{-1} in the form (3.6.15) and find V_h, $0 \le h \le r$, and V_r^* such that $BB^{-1} = I + O\left(2^{-mr}\right)$. Multiplying (3.6.13) and (3.6.15), introducing a new parameter $i = l + h$ and eliminating $O(2^{-mr})$ terms, we obtain equation

$$
\sum_{i=0}^{2r} 2^{-mi} \sum_{h=\max(0,i-r)}^{\min(i,r)} \frac{p^{(i-h)}(y)}{(i-h)!}\, U_{i-h}V_h = I. \qquad (3.6.16)
$$

Equating matrix coefficients for various powers of 2^{-m}, we derive a system of linear equations

$$\sum_{h=0}^{i} U_{i-h} p^{(i-h)}(y)[(i-h)!]^{-1} V_h = I(i=0)I, \quad i=0,\cdots,r, \tag{3.6.17}$$

where $I(\cdot)$ is the indicator function. Formula (3.6.17) suggests a recursive procedure to calculate V_h, $0 \le h \le r$.

It is straightforward to see that $V_0 = [p(y)]^{-1}I$ which verifies (3.6.15) for $r=0$. Let us use mathematical induction to prove Lemma 6. Assuming that Lemma 6 is valid for j, we shall show that it remains valid for $j+1$. Since we can keep any number of terms in representation (3.6.15), we only need to prove that V_r can be represented in the form stated by Lemma 6. From (3.6.17) and induction assumption it follows that

$$\begin{aligned} V_{j+1} &= -\frac{1}{p(y)} \sum_{h=0}^{j} \frac{p^{(j+1-h)}(y)}{(j+1-h)!} V_h\, U_{j+1-h} \\ &= -\frac{1}{p(y)} \sum_{h=0}^{j} \frac{p^{(j+1-h)}(y)}{(j+1-h)!} V_h\, U_{j+1-h} \\ &= -\frac{1}{p(y)} \sum_{h=0}^{j} \frac{p^{(j+1-h)}(y)}{(j+1-h)!} U_{j+1-h} \sum_{k_1,\cdots,k_l} \alpha_{k_1,k_2,\cdots,k_l} U_{k_1} U_{k_2} \cdots U_{k_l} \\ &= \sum_{h=0}^{j} \sum_{k_1,\cdots,k_l} \alpha_{k_1,k_2,\cdots,k_l} \frac{\left[-p^{(j+1-h)}(y)\right]}{p(y)(j+1-h)!} U_{k_1} U_{k_2} \cdots U_{k_l} U_{j+1-h}. \end{aligned} \tag{3.6.18}$$

Here $k_1 + \cdots + k_l = h$ and $k_1 + \cdots + k_l + (j+1-h) = j+1$ which completes the proof.

Now, we can prove that the systematic error tends to zero as $m \to \infty$.

Lemma 7. Let functions $p(x)$ and $\Psi(x)$ be $r \le s-1$ times continuously differentiable in the neighborhood Ω_y of y and let $\Omega_{m,y} \subseteq \Omega_y$, with $\Omega_{m,y}$ defined in (3.6.11). Then, for R_1 defined in (3.3.8), as $m \to \infty$,

$$R_1(y) = (t_m(y) - t(y))^2 = o(2^{-2mr}). \tag{3.6.19}$$

Proof of Lemma 7. Let $Q(x,z) = \sum_{k \in Z} \varphi(x-k)\varphi(z-k)$. Recall that, by Theorem 3.2 in Walter and Shen (2001), one has

$$\int_{-\infty}^{\infty} z^j Q(2^m y + x, 2^m y + z) dz = x^j, \; 0 \le j \le s-1. \tag{3.6.20}$$

Now, let us show that for $1 \le v \le s-1$, $0 \le l_i \le s-1$, $l_1 + \cdots + l_t = h$ where $t \le s-1$, one has

$$S = \sum_{k \in K_{m,y}} (U_{l_1} U_{l_2} \cdots U_{l_t} D_{v-h})_k \ \varphi(2^m y - k) = 0. \tag{3.6.21}$$

Indeed,

$$\begin{aligned} S &= \int \cdots \int \sum_{k_1, \cdots, k_t, k \in K_{m,y}} z_1^{l_1} \varphi(2^m y + z_1 - k) \varphi(2^m y + z_1 - k_1) \cdots z_t^{l_t} \varphi(2^m y + z_t - k_{t-1}) \\ &\times \ \varphi(2^m y + z_t - k_t) \times z^{v-h} \varphi(2^m y + z - k_t) \varphi(2^m y - k) dz_1 \cdots dz_t dz. \end{aligned} \tag{3.6.22}$$

Now, note that the set $K_{m,y}$ is chosen so that all the sums over $k_1 \notin K_{m,y}, \cdots, k_t \notin K_{m,y}, k \notin K_{m,y}$ vanish. Hence, we can replace $K_{m,y}$ by the set of all integers Z in the summations. Using the definition of $Q(x,z)$ and (3.6.20), we derive

$$S = \int z_1^{v-h-l_1+\cdots+l_t} Q(2^m y, 2^m y + z_1) dz_1 = 0, \quad v \le r. \tag{3.6.23}$$

Therefore, by formulae (3.6.13) and (3.6.14), one has

$$\begin{aligned} a &= B^{-1} c = 2^{-\frac{m}{2}} \left(\sum_{h=0}^{r} \frac{2^{-mh}}{h!} p^{(h)}(y) U_h + o(2^{-mr}) \right) \left(\sum_{l=0}^{r} \frac{2^{-ml}}{l!} \Psi^{(l)}(y) D_l + o(2^{-mr}) \right) \\ &= 2^{-m/2} \sum_{v=0}^{2r} 2^{-mv} \sum_{h=\max(0,v-r)}^{\min(v,r)} [(v-h)!]^{-1} \Psi^{(v-h)}(y) V_h D_{v-h} + o(2^{-mr}). \end{aligned} \tag{3.6.24}$$

where the last relation is obtained by introducing a new parameter $v = l + h$, re-arranging the sums and combining the terms $O(2^{-mr})$. Recall that $V_0 = [p(y)]^{-1} I$ and all the sums are finite, so that

$$t_{m,y} = \sum_{k \in K_{m,y}} 2^{-m/2} \sum_{v=0}^{r} \sum_{h=0}^{v} 2^{-mv} [(v-h)!]^{-1} \Psi^{(v-h)}(y) \ (V_h D_{v-h})_k \ \varphi(2^m y - k) + o(2^{-mr}) \tag{3.6.25}$$

$$\begin{aligned} &= 2^{-m/2} \sum_{v=0}^{r} 2^{-mv} \sum_{h=0}^{v} \frac{\Psi^{(v-h)}(y)}{(v-h)!} \sum_{\substack{k_1, \cdots, k_t \\ \sum k_i = h}} \alpha_{k_1, k_2, \cdots, k_t} \underbrace{\sum_{k \in K_{m,y}} (U_{k_1} \cdots U_{k_t} D_{v-h})_k \, \varphi(2^m y - k)}_{=0 \text{ if } v \neq 0} \\ &+ \ o(2^{-mr}) = \Psi(y) \sum_{k \in K_{m,y}} (V_0 D_0)_k 2^{-m/2} \varphi(2^m y - k) + o(2^{-mr}) = t(y) + o(2^{-mr}) \end{aligned} \tag{3.6.26}$$

as $m \to \infty$.

3.7 Large Deviation Results

Application of Lepski method requires the use of the so called, large deviation results, which we prove in this section. The matrix norm used throughout this discussion is the spectral norm.

Lemma 8. Let $\widehat{B}$, B, $\widehat{c}$ and c be defined as in (3.2.9), (3.2.6), (3.2.12) and (3.2.7), respectively. Then, provided that m is such that $2^m(\gamma_m^2+1) \le \frac{n}{\log(n)^2}$, we have the following

$$\Pr\left(\left|\left|\widehat{B}-B\right|\right|^2 \ge \frac{M^2\tau^2 2^m}{n}\log n\right) \le 2M^2 n^{-\frac{\tau^2}{D_0}}, \tag{3.7.1}$$

where

$$D_0 = 8\,||\varphi||_\infty^2\,||p||_\infty. \tag{3.7.2}$$

In addition, provided that $||u_{m,k}||_\infty \le 2^{\frac{m}{2}}\gamma_m$, then

$$\Pr\left(||\widehat{c}-c||^2 \ge \frac{M\tau^2\gamma_m^2}{n}\log n\right) \le 2Mn^{-\frac{\tau^2}{D_1}}, \tag{3.7.3}$$

where

$$D_1 = 8\,||p||_\infty. \tag{3.7.4}$$

Recall that $\widehat{B}_{j,k}$ defined by (3.2.9) are the unbiased estimators of $B_{j,k}$ defined in (3.2.6). Denote $\eta_t = \varphi_{mj}(X_t)\varphi_{mk}(X_t) - B_{j,k}$, thus $\widehat{B}_{j,k} - B_{j,k} = \frac{1}{n}\sum_{t=1}^{n}\eta_t$, where η_t are iid.

Consequently, taking expectation we get, $\mathbf{E}\left(\widehat{B}_{j,k}-B_{j,k}\right) = \mathbf{E}(\eta_t) = 0$. Whereas for the variance, we have

$$\sigma_B^2 \quad = \quad \mathbf{E}\left(\eta_t^2\right) \le 2\int \varphi_{mj}^2(x)\varphi_{mk}^2(x)p(x)\,dx \le 2\,||p||_\infty\,||\varphi||_\infty^2\,2^m \tag{3.7.5}$$

Also, $||\eta_t||_\infty = 2.2^m\,||\varphi||_\infty^2$.

Since η_t, with $t = 1,2,3,\ldots,n$ are i.i.d with $\mathbf{E}(\eta_t) = 0$, $\mathbf{E}\left(\eta_t^2\right) = \sigma^2$ and $||\eta||_\infty < \infty$, Bernstein inequality applies. Thus, taking $z = \frac{\tau 2^{\frac{m}{2}} \sqrt{\log n}}{\sqrt{n}}$ and using the Bernstein inequality we obtain

$$\begin{aligned}
\Pr\left(\left|\frac{1}{n}\sum_{t=1}^{n}\eta_t\right| > \frac{\tau 2^{\frac{m}{2}}\sqrt{\log n}}{\sqrt{n}}\right) &\leq 2\exp\left(\frac{-\frac{n\tau^2 .2^m \log n}{n}}{2\left(2.2^m \|p\|_\infty \|\varphi\|_\infty^2 + 2.2^m \|\varphi\|_\infty^2 \frac{2^{\frac{m}{2}}\tau\sqrt{\log n}}{3\sqrt{n}}\right)}\right) \\
&= 2\exp\left(\frac{-\tau^2 \log n}{4\|\varphi\|_\infty^2 \left(\|p\|_\infty + 2^{\frac{m}{2}}\frac{\tau\sqrt{\log n}}{3\sqrt{n}}\right)}\right) \\
&\leq 2\exp\left(\frac{-\tau^2 \log n}{4\|\varphi\|_\infty^2 \left(\|p\|_\infty + \frac{\tau}{3\gamma_m\sqrt{\log(n)}}\right)}\right) \\
&\leq 2\exp\left(\frac{-\tau^2 \log n}{8\|\varphi\|_\infty^2 \|p\|_\infty}\right) \qquad (3.7.6)
\end{aligned}$$

Thus,

$$\Pr\left(\left|\widehat{B}_{j,k} - B_{j,k}\right| \geq \frac{\tau 2^{\frac{m}{2}}\sqrt{\log n}}{\sqrt{n}}\right) \leq 2n^{-\frac{\tau^2}{C_0}}. \qquad (3.7.7)$$

This is equivalent to

$$\Pr\left(\sum_{j,k}\left|\hat{B}_{j,k} - B_{j,k}\right|^2 \geq \frac{\tau^2 M^2 2^m \log n}{n}\right) \leq 2M^2 n^{-\frac{\tau^2}{C_0}}, \qquad (3.7.8)$$

Hence, (3.7.1) is valid.

Now, in order to prove (3.7.3), recall that $c_k = \int_{-\infty}^{\infty} u_{m,k}(x)p(x)dx$ has an unbiased estimator

$$\widehat{c}_k = \frac{1}{n}\sum_{t=1}^{n} u_{m,k}(X_t). \qquad (3.7.9)$$

Denote $\widehat{c}_k - c_k = \frac{1}{n}\sum_{t=1}^{n}\xi_t$, where ξ_t are i.i.d. Thus taking the expectation we get, $\mathbf{E}(\widehat{c}_k - c_k) = \mathbf{E}(\xi_t) = 0$. Also, for the variance one has

$$\sigma_c^2 = \mathbf{E}\left(\xi_t^2\right) \leq 2\int_{-\infty}^{\infty} u_{m,k}^2(x)p(x)\,dx \leq 2\|p\|_\infty \gamma_m^2. \qquad (3.7.10)$$

In addition,

$$\|\xi_t\|_\infty \leq 2\|u_{m,k}\|_\infty \leq 2\gamma_m 2^{\frac{m}{2}}. \qquad (3.7.11)$$

where γ_m^2 is defined in (3.8.2). Thus, taking $z = \frac{\tau\gamma_m}{\sqrt{n}}\sqrt{\log n}$ and applying the Bernstein inequality we obtain

$$\begin{aligned}
\Pr\left(\left|\frac{1}{n}\sum_{t=1}^{n}\xi_t\right| > \frac{\tau\gamma_m\sqrt{\log n}}{\sqrt{n}}\right) &\leq 2\exp\left(\frac{-\tau^2\gamma_m^2\log n}{2(2\,||p||_\infty\,\gamma_m^2 + 2\gamma_m 2^{\frac{m}{2}}\frac{\gamma_m\tau\sqrt{\log n}}{3\sqrt{n}})}\right) \\
&\leq 2\exp\left(\frac{-\tau^2\log n}{2(2\,||p||_\infty + 2\frac{2^{\frac{m}{2}}\tau\sqrt{\log n}}{3\sqrt{n}})}\right) \\
&\leq 2\exp\left(\frac{-\tau^2\log n}{4(||p||_\infty + \frac{\tau}{3\gamma_m\sqrt{\log n}})}\right) \\
&\leq 2\exp\left(\frac{-\tau^2\log n}{8\,||p||_\infty}\right) \qquad (3.7.12)
\end{aligned}$$

Consequently,

$$\Pr\left(\left|\frac{1}{n}\sum_{t=1}^{n}\xi_t\right| > \frac{\tau\gamma_m\sqrt{\log n}}{\sqrt{n}}\right) \leq 2n^{-\frac{\tau^2}{D_1}} \qquad (3.7.13)$$

This is equivalent to

$$\Pr\left(|\widehat{c}_k - c_k| > \frac{\tau\gamma_m\sqrt{\log n}}{\sqrt{n}}\right) \leq 2n^{-\frac{\tau^2}{D_1}}, \qquad (3.7.14)$$

or

$$\Pr\left(\sum_k |\widehat{c}_k - c_k|^2 > \frac{M\tau^2\gamma_m^2\log n}{n}\right) \leq 2Mn^{-\frac{\tau^2}{D_1}}. \qquad (3.7.15)$$

Hence,

$$\Pr\left(||\widehat{c} - c|| > \frac{\sqrt{M}\tau\gamma_m\sqrt{\log n}}{\sqrt{n}}\right) \leq 2Mn^{-\frac{\tau^2}{D_1}}. \qquad (3.7.16)$$

Lemma 9. Let the resolution level m be such that $m_1 \leq m \leq m_n$, where m_1 and m_n is such that $2^{m_1} = \log n$ and $2^{m_n}(\gamma_{m_n}^2 + 1) \asymp \frac{n}{\log^2 n}$. Denote,

$$\rho_{mn}^2 = 2^m n^{-1}(1 + \gamma_m^2)\log(n) \qquad (3.7.17)$$

where γ_m^2 is defined in (3.8.2). Then

$$\Pr\left(||\widehat{a}_\delta - a|| > \lambda\ \rho_{mn}\right) = O(n^{-\frac{\lambda^2}{D_2}}) \tag{3.7.18}$$

where

$$D_2 = 32\sqrt{2}M\sqrt{||p||_\infty}\max\{\sqrt{M}, M^2\,||\varphi||_\infty\,||\Psi(y)||_\infty\,||D_0||\} \tag{3.7.19}$$

Proof of Lemma We seek λ that makes the probability

$$\Pr\left(||\widehat{a}_\delta - a|| > \lambda\rho_{mn}\right) = O(n^{-1}) \tag{3.7.20}$$

Thus, recall (3.5.20), then using the properties of the probability we obtain for $\alpha + \beta + \nu + \rho + \epsilon = 1$,

$$\begin{aligned}
\Pr\left(||\widehat{a}_\delta - a|| > \lambda\rho_{mn}\right) &\leq \Pr\left((2M + 4\delta M^2)\,||\widehat{c} - c|| > \alpha\lambda\rho_{mn}\right) \\
&+ \Pr\left((\frac{2}{\delta}\,||\widehat{c} - c|| + \frac{2}{\delta}\,||c||\,\mathbf{1}\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4M}\right)) > \beta\lambda\rho_{mn}\right) \\
&+ \Pr\left(8M^2\,||\widehat{c} - c||\,\left|\left|\widehat{B} - B\right|\right| > \nu\lambda\rho_{mn}\right) + \Pr\left(8M^2\,||c||\,\left|\left|\widehat{B} - B\right|\right| > \rho\lambda\rho_{mn}\right) \\
&+ \Pr\left(4\delta M^2\,||c|| > (1 - \alpha - \beta - \nu - \rho)\lambda\rho_{mn}\right) \\
&\leq \Pr\left((2M + 4\delta M^2)\,||\widehat{c} - c|| > \alpha\lambda\rho_{mn}\right) \\
&+ \Pr\left((||\widehat{c} - c|| + ||c||)^2 > \frac{\delta}{2}\beta\lambda\rho_{mn}\right)\Pr\left(\mathbf{1}^2(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4M}) > \frac{\delta}{2}\beta\lambda\rho_{mn}\right) \\
&+ \Pr\left(||\widehat{c} - c||\,\left|\left|\widehat{B} - B\right|\right| > \frac{\nu}{8M^2}\lambda\rho_{mn}\right) + \Pr\left(\left|\left|\hat{B} - B\right|\right| > \frac{\rho}{8M^2\,||c||}\lambda\rho_{mn}\right)
\end{aligned} \tag{3.7.21}$$

Now as $n \longrightarrow \infty$, $\delta \longrightarrow 0$ and therefore this probability becomes

$$\begin{aligned}
\Pr\left(||\widehat{a}_\delta - a|| > \lambda\rho_{mn}\right) &\leq \Pr\left(||\widehat{c} - c|| > \frac{\alpha}{4M}\lambda\rho_{mn}\right) \\
&+ \Pr\left((||\widehat{c} - c|| + ||c||)^2 > \frac{\delta}{2}\beta\lambda\rho_{mn}\right)\Pr\left(\mathbf{1}^2(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4M}) > 0\right) \\
&+ \Pr\left(||\widehat{c} - c||^2 > \frac{\nu}{8M^2}\lambda\rho_{mn}\right) + \Pr\left(\left|\left|\widehat{B} - B\right|\right|^2 > \frac{\nu}{8M^2}\lambda\rho_{mn}\right) \\
&+ \Pr\left(\left|\left|\widehat{B} - B\right|\right| > \frac{\rho}{8M^2\,||c||}\lambda\rho_{mn}\right)
\end{aligned} \tag{3.7.22}$$

The application of the large deviation results to $\Pr\left(\mathbf{1}^2(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4\sqrt{M}}) > 0\right))$ leads to an infinitesimally small probability, so it makes sense to only evaluate the other parts of the right hand side. Now, before we go any further let us evaluate $||c||$. Indeed, recall that the vector c can be asymptotically expanded according to (3.6.14). Therefore it can be written as

$$c = 2^{-\frac{m}{2}}\Psi(y)D_0 + o(2^{-\frac{m}{2}}) \tag{3.7.23}$$

Consequently, taking the norm yields

$$||c|| \le 2^{-\frac{m}{2}} ||\Psi(y)||_\infty ||D_0|| \tag{3.7.24}$$

Therefore, it is of order $O(2^{-\frac{m}{2}})$. Thus, using large deviation results, we have

$$\begin{aligned}
\Pr\left(\left|\left|\widehat{B}-B\right|\right| > \frac{\rho}{8M^2\,||c||}\lambda\rho_{mn}\right) &\le \Pr\left(\left|\left|\widehat{B}-B\right|\right|^2 > \left(\frac{\rho}{8M^2\,||c||}\right)^2 \lambda^2 \frac{\gamma_m^2+1}{n}\log(n)\right) \\
&\le \Pr\left(\left|\left|\widehat{B}-B\right|\right|^2 > \left(\frac{\rho\lambda}{8M^2 2^{-\frac{m}{2}}\,||\Psi(y)||_\infty\,||D_0||}\right)^2 \frac{1}{n}\log(n)\right) \\
&\le \Pr\left(\left|\left|\widehat{B}-B\right|\right|^2 > \left(\frac{\rho\lambda}{8M^2\,||\Psi(y)||_\infty\,||D_0||}\right)^2 \frac{2^m}{n}\log(n)\right) \\
&\le 2M^2 n^{-\frac{1}{8||p||_\infty ||\varphi||_\infty^2}\left(\frac{\rho\lambda}{8M^3||\Psi(y)||_\infty ||D_0||}\right)^2}
\end{aligned} \tag{3.7.25}$$

Also

$$\begin{aligned}
\Pr\left(||\widehat{c}-c|| > \frac{\alpha}{4M}\lambda\rho_{mn}\right) &\le \Pr\left(||\widehat{c}-c||^2 > \left(\frac{\alpha}{4M}\right)^2 \lambda^2 \frac{\gamma_m^2+1}{n}\log(n)\right) \\
&\le \Pr\left(||\widehat{c}-c||^2 > \left(\frac{\alpha\lambda}{4M}\right)^2 \frac{\gamma_m^2}{n}\log(n)\right) \\
&\le 2Mn^{-\frac{1}{8||p||_\infty}\left(\frac{\alpha\lambda}{4\sqrt{M^3}}\right)^2}
\end{aligned} \tag{3.7.26}$$

Now, for the remaining term in (3.7.22), it can be shown that

$$\Pr\left(||\widehat{c}-c||^2 > \frac{\nu}{8M}\lambda\rho_{mn}\right) \le 2M\exp\left(-\frac{\nu\lambda\sqrt{n\log(n)}}{64M^2\gamma_m\,||p||_\infty}\right) \tag{3.7.27}$$

and

$$\Pr\left(\left|\left|\widehat{B}-B\right|\right|^2 > \frac{\nu}{8M}\lambda\rho_{mn}\right) \leq 2M^2\exp\left(-\frac{\nu\lambda\sqrt{n\log(n)}}{64M^3 2^{\frac{m}{2}}\,||\varphi||_\infty^2\,||p||_\infty}\right) \tag{3.7.28}$$

which are infinitesimally small probabilities, so it make more sense not to consider them in the derivation of λ. The idea now is to balance the former two probabilities so that $\Pr\left(||\widehat{a}_\delta - a|| > \lambda\rho_{mn}\right) = O(n^{-1})$. Thus, we need to choose λ such that

$$\Pr\left(||\widehat{a}_\delta - a|| > \lambda\rho_{mn}\right) = O(n^{-1}) \tag{3.7.29}$$

Indeed, we choose λ such that

$$\frac{\rho^2\ \lambda^2}{512M^6\,||p||_\infty\,||\varphi||_\infty^2\,||\Psi(y)||_\infty^2\,||D_0||^2} = 1 \tag{3.7.30}$$

or

$$\frac{\alpha^2\ \lambda^2}{128M^3\,||p||_\infty} = 1 \tag{3.7.31}$$

Now take $\alpha = \frac{1}{4}$ and $\rho = \frac{1}{2}$ and solving for λ we obtain

$$\lambda = 32\sqrt{2}M^3\sqrt{||p||_\infty}\,||\varphi||_\infty\,||\Psi(y)||_\infty\,||D_0|| \tag{3.7.32}$$

or

$$\lambda = 32\sqrt{2}\sqrt{M^3}\sqrt{||p||_\infty} \tag{3.7.33}$$

The final step is to choose the larger of the two quantities so that the larger of the probabilities is of order $O(n^{-1})$. That is,

$$D_2 = 32\sqrt{2}M\sqrt{||p||_\infty}\max\{\sqrt{M}, M^2\,||\varphi||_\infty\,||\Psi(y)||_\infty\,||D_0||\} \tag{3.7.34}$$

3.8 The Random Error Component

In order to calculate R_2, introduce vectors $\gamma^{(j)}(m),\ j=1,2$, with components

$$\gamma_k^{(\varrho)}(m) = \left[\int_{-\infty}^{\infty} u_{m,k}^{2\varrho}(x)dx\right]^{1/2}, \quad k \in K_{m,y},\ \varrho = 1,2,3,4. \tag{3.8.1}$$

where $u_{m,k}(x)$ are defined in (4.1.5). Denote

$$\gamma_m = \left|\left|\gamma^{(1)}(m)\right|\right|. \tag{3.8.2}$$

The following expression provides an asymptotic expression for the random error component as $m, n \to \infty$.

Lemma 10. Let $\delta^2 \sim n^{-1}2^m$. Then, under the assumptions of Lemma 7, as $m, n \to \infty$, the random error component R_2 defined in (3.3.9) is such that

$$R_2 = O\left(2^m n^{-1}\gamma_m^2 + 2^m n^{-1}\right), \quad m, n \to \infty, \tag{3.8.3}$$

provided m is such that $2^m n^{-1} \to 0$ and $\left|\left|\gamma^{(2)}(m)\right|\right|^2 2^{2m} = o(n^3)$ as $n \to \infty$. Here, $||z||$ is the Eucledean norm of the vector z.

Proof of Lemma 10. Recall equation (4.17), then we have

$$\begin{aligned} R_2 &= \mathbf{E}\left[\sum_{k \in K_{m,y}} ((\widehat{a}_\delta)_{m,k} - a_{m,k})\varphi_{m,k}(y)\right]^2 \\ &\leq C2^m \mathbf{E}\,||\widehat{a}_\delta - a||^2 \end{aligned} \tag{3.8.4}$$

Corollary 1 provides us with an upper bound for $\mathbf{E}\,||\widehat{a}_\delta - a||^2$. In addition, large deviation results can be applied to $\Pr\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4M}\right)$ in (3.5.29). Indeed, using (3.7.1) yields

$$\Pr\left(\left|\left|\widehat{B} - B\right|\right| > \frac{1}{4M}\right) = o(n^{-\alpha}) \tag{3.8.5}$$

for any $\alpha > 0$, which implies that this probability decays faster than any power of n as $n \to \infty$. Consequently, (3.5.29) reduces to

$$\begin{aligned}\mathbf{E}\,||\widehat{a}_\delta - a||^2 \;&\leq\; 32M^2\mathbf{E}\,||\widehat{c}-c||^2 + 32\delta^2 M^4\,||c||^2 \\ &+\; 128M^4\,||c||^2\,\mathbf{E}\left|\left|\widehat{B}-B\right|\right|^2 + 128M^4\sqrt{\mathbf{E}\,||\widehat{c}-c||^4\,\mathbf{E}\left|\left|\widehat{B}-B\right|\right|^4}\end{aligned} \tag{3.8.6}$$

Finally, using the assumptions of Lemma 10, and for $\delta^2 \sim n^{-1}2^m$ we conclude that

$$\mathbf{E}\,||\widehat{a}_\delta - a||^2 = O\left(\frac{\gamma_m^2}{n} + \frac{1}{n}\right) \tag{3.8.7}$$

The result of (3.8.3) follows directly by using equation (3.8.4) as $n \to \infty$.

Observe that the values of $\gamma_k^{(\varrho)}(m)$ are independent of the unknown density $g(\theta)$ and can be calculated explicitly. Later in this chapter, we shall bring examples of construction of functions $u_{m,k}(x)$ as well as the asymptotic expressions for $\gamma_k^{(\varrho)}(m)$, $\varrho = 1, 2$, for some common special cases (location parameter family, scale parameter family, one-parameter exponential family). In vast majority of situations, γ_m^2 is bounded above by the following expression

$$\gamma_m^2 \leq C_\gamma 2^{\alpha m}\exp\left(b2^{\beta m}\right), \quad b, \beta \geq 0,\; C_\gamma > 0,\; \alpha \in \mathbb{R}, \tag{3.8.8}$$

where α, b, β and C_γ are the absolute constant independent of m.

Lemma 7 shows that systematic error goes to zero at an optimal rate of $O\left(2^{-2mr}\right)$. Lemma 10 asserts that the random error component of the EB estimator is proportional to γ_m^2. In order to balance both errors choose

$$m_0 = \arg\min\left(n^{-1}2^m[\gamma_m^2 + 1] + 2^{-2mr}\right). \tag{3.8.9}$$

In particular, under assumption (3.8.8), as $n \to \infty$, m_0 is such that

$$2^{m_0} \asymp \begin{cases} C_m n^{\frac{1}{2r+\max(1,\alpha)}}, & \text{if } \; b = 0, \\ ((2b)^{-1}\log n)^{1/\beta}, & \text{if } \; b > 0. \end{cases} \tag{3.8.10}$$

Here, $a_n \asymp b_n$ for two sequences, $\{a_n\}$ and $\{b_n\}$, $n = 1, 2, \cdots$, of positive real numbers if there exist C_1 and C_2 independent of n such that $0 < C_1 < C_2 < \infty$ and $C_1 \leq \alpha_n/\beta_n \leq C_2$.

Then combining results of Lemmas 7 and 10 with results in (3.8.10), the following statement is true.

Theorem 3. Let twice continuously differentiable scaling function φ satify (3.2.1) and (3.2.2). Let functions

$p(x)$ and $\Psi(x)$ be r times continuously differentiable in the neighborhood Ω_y of y such that $\Omega_{m,y} \subseteq \Omega_y$, where $\Omega_{m,y}$ is defined in (3.6.11). Let $r \in [1/2, s-1]$. Choose m_0 according to (3.8.9) and let in (3.2.13) be such that $\delta \sim n^{-1} 2^{m_0}$. If wavelets possesses s vanishing moments, $s \geq r+1$, then, for any y such that $p(y) > 0$, as $n \to \infty$, $R_n(y)$ defined in (3.3.3) satisfies the following asymptotic relation

$$R_n(y) = \mathbf{E}(\hat{t}_m(y) - t(y))^2 = O\left(2^{-2m_0 r}\right), \quad m \to \infty, \tag{3.8.11}$$

provided $2^m n^{-1} \to 0$ and $\left|\left|\gamma^{(2)}(m_0)\right|\right|^2 2^{2m_0} = o(n^3)$ as $n \to \infty$. In particular, if assumption (3.8.8) holds, then, as $n \to \infty$, one has

$$R_n(y) = \begin{cases} O\left(n^{-\frac{2r}{2r+\max(1,\alpha)}}\right), & \text{if } \ b = 0, \\ ((2b)^{-1}\log n)^{-\frac{2r}{\beta}}, & \text{if } \ b > 0. \end{cases} \tag{3.8.12}$$

Note that when $b > 0$, the optimal resolution level m_0 is determined by the values of b and β which are completely known, so that the resulting EB estimator is *adaptive*, i.e. it attains the optimal convergence rate. However, if $b = 0$, the value of m_0 depends on the unknown smoothness of the functions $p(x)$ and $\Psi(x)$. In order to construct an adaptive estimator in this case, we shall apply Lepski method (see e.g., Lepski (1991) and Lepski *et al.* (1997)) for the optimal selection of the resolution level.

3.9 Adaptive choice of the resolution level using Lepski method

In order to construct an adaptive estimator, we apply Lepski method (see e.g., Lepski (1991) and Lepski *et al.* (1997)) for the optimal selection of the resolution level. The method suggests to choose

$$\widehat{m} = \min\{m \leq m_n : \left|\widehat{t}_m(y) - \widehat{t}_j(y)\right|^2 \leq \lambda^2 \left[\left|\left|\widehat{B}_{\delta m}^{-1}\right|\right|^2 + \left|\left|\widehat{B}_{\delta j}^{-1}\right|\right|^2\right]^2 n^{-1} 2^j (1+\gamma_j^2) \log n \ \text{ for any } \ j \geq m\}, \tag{3.9.1}$$

where, $m_n = \{m > 1 : 2^m(\gamma_m^2 + 1) \leq \frac{n}{\log^2(n)}\}$, γ_j is defined in (3.8.1) and (3.8.2), and λ is a constant independent of m. Recall that m_1 and m_n are defined such that

$$2^{m_1} = \log n, \quad 2^{m_n}(\gamma_{m_n}^2 + 1) \asymp (\log n)^{-2} n. \tag{3.9.2}$$

Also, observe that under assumption (3.8.8) with $b = 0$, m_n is such that

$$2^{m_n} = \left(\frac{n}{(C_\gamma + 1)\log^2 n}\right)^{\frac{1}{1+\max(\alpha,0)}}, \tag{3.9.3}$$

so that, for m_0 given by (3.8.10), one has $m_n/m_0 \to \infty$ as $n \to \infty$.

In order to see how the method works, note that the error can be decomposed according to the sets of the resolution levels as

$$\Delta = \mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = \Delta_1 + \Delta_2 \tag{3.9.4}$$

where m_0 is the optimal resolution level defined in formula (3.8.9) and

$$\Delta_1 = \mathbf{E}[\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 \mathbf{1}(\widehat{m} \le m_0)], \tag{3.9.5}$$

$$\Delta_2 = \mathbf{E}[\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 \mathbf{1}(\widehat{m} > m_0)]. \tag{3.9.6}$$

If $\widehat{m} \le m_0$, then, by definition of $\widehat{m}$, one has

$$\left|\widehat{t}_{\widehat{m}}(y) - \widehat{t}_{m_0}\right|^2 \le \lambda^2 \rho_{mn}^2 \left[\left|\left|\widehat{B}_{\delta m}^{-1}\right|\right|^2 + \left|\left|\widehat{B}_{\delta j}^{-1}\right|\right|^2\right]^2 = O(\rho_{m_0 n}^2), \tag{3.9.7}$$

so that

$$\Delta_1 \le 2\left[\mathbf{E}\left[\left|\widehat{t}_{\widehat{m}}(y) - \widehat{t}_{m_0}\right|^2 \mathbf{1}(\widehat{m} \le m_0)\right] + \mathbf{E}\left|\widehat{t}_{m_0}(y) - t(y)\right|^2\right]. \tag{3.9.8}$$

Here,

$$\mathbf{E}\left|\widehat{t}_{m_0}(y) - t(y)\right|^2 = O\left(\rho_{m_0 n}^2\right) \tag{3.9.9}$$

so that

$$\Delta_1 \le 2\left[\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - \widehat{t}_{m_0}(y)\right|^2 + \mathbf{E}\left|\widehat{t}_{m_0}(y) - t(y)\right|^2\right] = O\left(\rho_{m_0 n}^2\right). \tag{3.9.10}$$

by equation (3.8.9). For the other term in (4.4.34), observe that in norms we have

$$\begin{aligned}\left|\left|\widehat{B}_{\delta m}^{-1}\right|\right| &= \left|\left|\widehat{B}_{\delta m}^{-1} - B_{\delta m}^{-1}\right|\right| + \left|\left|B_{\delta m}^{-1}\right|\right| \\ &\le 2\left|\left|B^{-1}\right|\right|^2 \left|\left|\widehat{B} - B\right|\right| + 2\delta^{-1}\mathbf{1}(\Omega_B) + \left|\left|B^{-1}\right|\right| \end{aligned} \tag{3.9.11}$$

consequently, for any m, we have

$$\begin{aligned}\mathbf{E}\left|\left|\widehat{B}_{\delta m}^{-1}\right|\right|^4 &= O\left(\mathbf{E}\left|\left|\widehat{B}-B\right|\right|^4 + \delta^{-4}\Pr(\Omega_B) + \left|\left|B^{-1}\right|\right|^4\right) \\ &= O(1) \end{aligned} \tag{3.9.12}$$

and thus,

$$\mathbf{E}\left[\left|\widehat{t}_{\widehat{m}}(y)-\widehat{t}_{m_0}(y)\right|^2 \mathbf{1}(\widehat{m}\le m_0)\right] = \sum_{m=m_1}^{m_0}\mathbf{E}\left[\left|\widehat{t}_{\widehat{m}}(y)-\widehat{t}_{m_0}(y)\right|^2 \mid \widehat{m}=m\right]\Pr(\widehat{m}=m) \tag{3.9.13}$$

Finally, taking the expectation both sides of (3.9.7), using the results (3.9.12) and (3.9.9), and applying it to (3.9.8) we obtain (3.9.10).

Now, in the case when $\widehat{m} > m_0$, by (3.9.1), there exists an l such that $l > m_0$ and such that

$$|\widehat{t}_l(y)-\widehat{t}_{m_0}(y)|^2 \ge \lambda^2\rho_{ln}^2\left[\left|\left|\hat{B}_{\delta j}^{-1}\right|\right|^2 + \left|\left|\hat{B}_{\delta m_0}^{-1}\right|\right|^2\right]^2. \tag{3.9.14}$$

where ρ_{mn} is defined in (3.7.17). Define, for some positive constant λ, the set

$$\Theta_{l,m,\lambda} = \left\{\Theta : |\widehat{t}_l(y)-\widehat{t}_m(y)|^2 \ge \lambda^2\rho_{ln}^2\left[\left|\left|\hat{B}_{\delta j}^{-1}\right|\right|^2 + \left|\left|\hat{B}_{\delta m}^{-1}\right|\right|^2\right]^2\right\}. \tag{3.9.15}$$

It turns out that probability of such an event is very low.

Indeed,

$$\begin{aligned}\Delta_2 &= \mathbf{E}[\left|\widehat{t}_{\widehat{m}}(y)-t(y)\right|^2\mathbf{1}(\widehat{m}>m_0)] \\ &\le \sum_{l=m_0+1}^{m_n}\sqrt{\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y)-t(y)\right|^4\Pr(\Theta_{l,m_0,\lambda})} \\ &\le \sum_{l=m_0+1}^{m_n}\sqrt{\mathbf{E}\left|\widehat{t}_{m}(y)-t(y)\right|^4\mathbf{1}\left(m_0<\widehat{m}=m\le m_n\right)\Pr(\Theta_{l,m_0,\lambda})} \\ &\le \sum_{l=m_0+1}^{m_n}\sqrt{\mathbf{E}\left|\widehat{t}_{m}(y)-t(y)\right|^4\Pr(\Theta_{l,m_0,\lambda})} \end{aligned} \tag{3.9.16}$$

In order to evaluate Δ_2 we need to look into $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y)-t(y)\right|^4$ and $\Pr(\Theta_{l,m_0,\lambda})$ separately. The next couple of lemmas provide upper bounds for these two quantities.

Lemma 11. Let conditions of Theorem 4 hold. If resolution level m is such that $m_0 < m \le m_n$, then, as

$n \to \infty$,

$$\mathbb{P}\left(\Theta_{l,m_0,\lambda}\right) = O\left(n^{-2}\right). \tag{3.9.17}$$

where $\rho_{mn}^2 = 2^m(1+\gamma_m^2)n^{-1}\log n$, and $\Theta_{l,m,\lambda}$ is defined in (3.9.15).

Proof of Lemma 11. Denote $R_{mn}^2 = \left[\left|\left|\hat{B}_{\delta m}^{-1}\right|\right|^2 + \left|\left|\hat{B}_{\delta m_0}^{-1}\right|\right|^2\right]^2 \rho_{mn}^2$ and observe that

$$\begin{aligned} \mathbb{P}(|\widehat{t}_m(y) - \widehat{t}_{m_0}(y)| \geq \lambda R_{mn}) &\leq \mathbb{P}(|\widehat{t}_m(y) - t_m(y)| + |t_m(y) - t(y)| \geq 0.5\,\lambda R_{mn}) \\ &+ \mathbb{P}(|\widehat{t}_{m_0}(y) - t_{m_0}(y)| + |t_{m_0}(y) - t(y)| \geq 0.5\,\lambda R_{mn}). \end{aligned} \tag{3.9.18}$$

Since $m > m_0$ and R_{mn} is an increasing function of m, one has $|t_m(y) - t(y)| = o(2^{-mr})$ as $m \to \infty$ and $R_{mn} > R_{m_0 n}$. Therefore, it is sufficient to show that

$$\mathbb{P}\left(|\widehat{t}_m(y) - t_m(y)| \geq 0.5\ \lambda R_{mn} - o(2^{-mr})\right) = O(n^{-2}) \tag{3.9.19}$$

for any $m \geq m_0$. Taking into account that $|\widehat{t}_m(y) - t_m(y)| \leq 2^{m/2} C_\varphi \,||\hat{a}_m - a||$ and $2^{-mr}/R_{mn} \to 0$ as $m, n \to \infty$, it is sufficient to show that

$$\mathbb{P}\left(||\hat{a}_m - a|| \geq 2^{-m/2}(\lambda - 1)R_{mn}/(2C_\varphi)\right) = O(n^{-2}), \quad n \to \infty. \tag{3.9.20}$$

Recall that $\hat{B}_\delta^{-1} - B^{-1} = \hat{B}_\delta^{-1}(B - \hat{B}_\delta)B^{-1}$, so that, for any $\delta > 0$, one has

$$\left|\left|\hat{B}_\delta^{-1} - B^{-1}\right|\right| \leq ||B_\delta^{-1}||^2 \left(\left|\left|\hat{B} - B\right|\right| + \delta_m\right) + 2\left|\left|\hat{B}_\delta^{-1}\right|\right|^2 ||B^{-1}|| \left[\left|\left|\hat{B} - B\right|\right|^2 + \delta_m^2\right]. \tag{3.9.21}$$

and, also,

$$||\hat{a}_\delta - a|| \leq \left|\left|\hat{B}_\delta^{-1}\right|\right| ||\hat{c} - c|| + \left|\left|\hat{B}_\delta^{-1} - B^{-1}\right|\right| ||c||. \tag{3.9.22}$$

Consequently, probability in (3.9.20) can be partition into three terms:

$$\mathbb{P}\left(||\hat{a}_m - a|| \geq 2^{-m/2}(\lambda - 1)R_{mn}/(2C_\varphi)\right) \leq \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 \tag{3.9.23}$$

where

$$\mathbb{P}_1 = \mathbb{P}\left(\left|\left|\hat{B}_\delta^{-1}\right|\right| ||\hat{c} - c|| \geq \frac{\alpha_1 R_{mn}(\lambda - 1)}{2^{m/2} 2C_\varphi}\right) \tag{3.9.24}$$

$$\mathbb{P}_2 = \mathbb{P}\left(|||c|| \, [\left|\left|\hat{B} - B\right|\right| + \delta_m] \geq \frac{\alpha_2 \sqrt{1+\gamma_m^2}\sqrt{\log n}(\lambda - 1)}{2\sqrt{n} C_\varphi}\right) \tag{3.9.25}$$

$$\mathbb{P}_3 = \mathbb{P}\left(||B^{-1}|| \, [\left|\left|\hat{B} - B\right|\right|^2 + \delta_m^2] \geq \frac{\alpha_3 \sqrt{1+\gamma_m^2}\sqrt{\log n}(\lambda - 1)}{4\sqrt{n} C_\varphi \, ||c||}\right) \tag{3.9.26}$$

and α_1, α_2 and α_3 are positive constants such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Applying (3.7.3) and taking into account that $\left|\left|\hat{B}_\delta\right|\right| \leq 2 \, ||p||_\infty$, obtain

$$\mathbb{P}_1 \leq \mathbb{P}\left(||\widehat{c} - c||^2 \geq \alpha_1 \frac{\sqrt{1+\gamma_m^2}\sqrt{\log n}(\lambda - 1)}{4 \, ||p||_\infty \sqrt{n} C_\varphi}\right) \leq 2Mn^{-\tau_1} \tag{3.9.27}$$

where $\tau_1 = (128 M C_\varphi^2 \, ||p||_\infty^3 \, M)^{-1} \, \alpha_1^2(\lambda - 1)^2$. Recalling that $||c|| \leq 2M \, ||\Psi||_\infty \, 2^{-m/2}$, using formula (3.7.1) and taking into account that $1 - 4M \, ||\Psi||_\infty \, C_\varphi/(\alpha_2(\lambda - 1) > 1 - \nu_1$ for any small positive constant ν_1 as $n \to \infty$, we derive

$$\begin{aligned}
\mathbb{P}_2 &\leq \mathbb{P}\left(\left|\left|\hat{B} - B\right|\right| \geq \frac{(\alpha_2\lambda - 1)2^{m/2} \, \sqrt{1+\gamma_m^2}\sqrt{\log n}}{4M \, ||\Psi||_\infty \, C_\varphi \, \sqrt{n}} - \frac{2^{m/2}}{\sqrt{n}}\right) \\
&\leq \mathbb{P}\left(\left|\left|\hat{B} - B\right|\right| \geq \frac{M 2^{m/2}\sqrt{\log n}}{\sqrt{n}} \, \frac{\alpha_2(\lambda - 1)\sqrt{1 - \nu_1}}{4M^2}\right) \leq 2M^2 n^{-\tau_2}
\end{aligned} \tag{3.9.28}$$

where $\tau_2 = (128 M^4 C_\varphi^2 \, ||\Psi||_\infty^2 \, ||\varphi||_\infty^2 \, ||p||_\infty)^{-1} \, \alpha_2^2(1 - \delta_1)(1 - \lambda)^2$.

In order to find an upper bound for $\mathbb{P}_3$, recall that $\left|\left|B^{-1}\right|\right| \leq 2M/p(y)$ and $||c|| \leq 2M \, ||\Psi||_\infty \, 2^{-m/2}$. Also, note that $p(y) \geq (\log n)^{-1/2}$, for any fixed y, as $n \to \infty$. Therefore, applying (3.7.1) and taking into account that, due to (3.9.2) and $m \leq m_n$, one has $(2^{-m/2}\sqrt{n} - 1)/\log n > 1 - \nu_2$ for any small positive constant ν_2 as $n \to \infty$, derive

$$\begin{aligned}
\mathbb{P}_3 &\leq \mathbb{P}\left(\left|\left|\hat{B} - B\right|\right|^2 \geq \frac{\alpha_3(\lambda - 1)\sqrt{1+\gamma_m^2}\sqrt{\log n}}{2\sqrt{n} C_\varphi \, ||B^{-1}|| \, ||c||} - 2^m n^{-1}\right) \\
&\leq \mathbb{P}\left(\left|\left|\hat{B} - B\right|\right|^2 \leq \frac{2^m M^2 \log n}{n} \, \frac{\alpha_3(\lambda - 1)}{16 C_\varphi M^3 \, ||\Psi||_\infty}\right) \leq 2M^2 n^{-\tau_3}
\end{aligned} \tag{3.9.29}$$

where $\tau_3 = (128 M^3 C_\varphi \, ||\Psi||_\infty \, ||\varphi||_\infty^2 \, ||p||_\infty^2)^{-1} \, \alpha_3(\lambda - 1)(1 - \nu_2)$.

Now, in order to complete the proof, combine (3.9.23) – (3.9.29) and choose α_i, $i = 1, 2, 3$, such that

$\tau_i \geq 2$ for $i = 1, 2, 3$, and $\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3$ takes minimal value. This is achieved by choosing λ such that

$$\lambda = 16C_\varphi \, ||p||_\infty^{1/2} \sqrt{M} D + 1 \tag{3.9.30}$$

where D is defined by

$$D = ||p||_\infty + ||\Psi||_\infty \, ||\varphi||_\infty \, M\sqrt{M}[1-\nu_1]^{-1/2} + 16 \, ||\Psi||_\infty \, ||\varphi||_\infty^2 \, ||p||_\infty^{3/2} \, M^3\sqrt{M}[1-\nu_2]^{-1} \tag{3.9.31}$$

with ν_1 and ν_2 are small positive values, $\nu_1 + \nu_2 < 1$, M is the size of vector c and matrix B and $C_\varphi = \sum_k |\varphi(z-k)|$.

Lemma 12. Let $\delta^2 \sim n^{-1}2^m$ and assumptions (3.8.8) hold.. Then, under the assumptions of Lemma 10

$$\mathbf{E} \left| \widehat{t}_m(y) - t(y) \right|^4 = O\left(n^{-2}2^{2m}(\gamma_m^2 + 1)^2 + 2^{-4mr}\right). \tag{3.9.32}$$

as $n \to \infty$, provided that

$$||u_{m,k}||_\infty \leq 2^{\frac{m}{2}} \gamma_m \tag{3.9.33}$$

Proof of Lemma 12.

$$\begin{aligned} \mathbf{E} \left| \widehat{t}_m(y) - t(y) \right|^4 &= \mathbf{E} \left| \widehat{t}_m(y) - t_m(y) + t_m(y) - t(y) \right|^4 \\ &\leq 8\mathbf{E} \left| \widehat{t}_m(y) - t_m(y) \right|^4 + 8 \left| t_m(y) - t(y) \right|^4 \end{aligned} \tag{3.9.34}$$

Thus, by lemma 7 the second part of the right hand side is equal to $o(2^{-4mr})$, and therefore it is of order $o(2^{-4m_0 r})$ since $\widehat{m} > m_0$. For the other term, recall (3.5.30), then using results of lemma 3 we obtain

$$\begin{aligned} \mathbf{E} \, ||\widehat{a}_\delta - a||^4 &\leq 1024M^4 O\left(n^{-3} \left|\left|\gamma^{(2)}(m)\right|\right|^2 + \left(\gamma_m^2 n^{-1}\right)^2\right) \\ &+ O(2^{-2m}) O\left(2^{2m}n^{-2}\right) \\ &+ 4096M^8 \sqrt{\mathbf{E} \, ||\widehat{c} - c||^8 \, \mathbf{E} \left|\left|\widehat{B} - B\right|\right|^8} + O(2^{2m}n^{-2}) O(2^{-2m}) \end{aligned} \tag{3.9.35}$$

For the remaining term in the right hand side recall from lemma 7 that

$$
\begin{aligned}
\mathbf{E}\left|\left|\widehat{c}-c\right|\right|^{8}\mathbf{E}\left|\left|\widehat{B}-B\right|\right|^{8} &= O\left(2^{4m}n^{-4}\right)O\left(\frac{\gamma_m^4\left|\left|\gamma^{(2)}(m)\right|\right|^2}{n^5}+\frac{\gamma_m^8}{n^4}\right) \\
&+ O\left(2^{4m}n^{-4}\right)O\left(\frac{\left|\left|\gamma^{(4)}(m)\right|\right|^2}{n^7}+\frac{\left|\left|\gamma^{(2)}(m)\right|\right|^4+\gamma_m^2\left|\left|\gamma^{(3)}(m)\right|\right|^2}{n^6}\right) \\
&= O\left(2^{4m}n^{-4}\right)O\left(\left(\gamma_m^2 n^{-1}\right)^2\frac{\left|\left|\gamma^{(2)}(m)\right|\right|^2}{n^3}+\left(\gamma_m^2 n^{-1}\right)^4\right) \\
&+ O\left(2^{4m}n^{-4}\right)O\left(\frac{\left|\left|\gamma^{(4)}(m)\right|\right|^2}{n^7}+\frac{\left|\left|\gamma^{(2)}(m)\right|\right|^4+\gamma_m^2\left|\left|\gamma^{(3)}(m)\right|\right|^2}{n^6}\right) \\
&= O\left(n^{-4}\right)O\left(\left(2^m\gamma_m^2 n^{-1}\right)^2\frac{2^{2m}\left|\left|\gamma^{(2)}(m)\right|\right|^2}{n^3}+\left(2^m\gamma_m^2 n^{-1}\right)^4\right) \\
&+ O\left(n^{-4}\right)O\left(\frac{2^{4m}\left|\left|\gamma^{(4)}(m)\right|\right|^2}{n^7}+\frac{2^{4m}\left|\left|\gamma^{(2)}(m)\right|\right|^4+\left(2^m\gamma_m^2\right)2^{3m}\left|\left|\gamma^{(3)}(m)\right|\right|^2}{n^6}\right)
\end{aligned} \tag{3.9.36}
$$

Therefore using the condition $||u_{m,k}||_\infty \leq 2^{\frac{m}{2}}\gamma_m$, (3.9.35) reduces to

$$
\mathbf{E}\left|\left|\widehat{a}_\delta-a\right|\right|^4 = O\left(\left(\gamma_m^2+1\right)^2 n^{-2}\right) \tag{3.9.37}
$$

which completes the proof of (3.9.32). It remains now to evaluate the probability term in (3.9.16).

The last two lemmas lead to the following result about Δ_2.

Lemma 13. Let Δ_2 be defined as in (3.9.16). Then

$$
\Delta_2 = O\left(n^{-1}\right) \tag{3.9.38}
$$

Proof of Lemma 13. First, let us show that $\mathbf{E}\left[\left|\left|\hat{B}_{\delta m}^{-1}\right|\right|^2+\left|\left|\hat{B}_{\delta m_0}^{-1}\right|\right|^2\right]^2 = O(1)$ as $m, n \to \infty$, so that asymptotic relation (4.4.34) holds. Indeed, for $m_1 \leq m \leq m_0$ and any fixed y, one has

$$
\begin{aligned}
\left|\left|\hat{B}_{\delta m}^{-1}\right|\right| &\leq \left|\left|\hat{B}_{\delta m}^{-1}-B_{\delta m}^{-1}\right|\right|+\left|\left|B_{\delta m}^{-1}\right|\right| \\
&\leq 2\left|\left|B_{\delta m}^{-1}\right|\right|^2\left|\left|\hat{B}_{\delta m}-B_{\delta m}\right|\right|+2\delta_m^{-1}\mathbf{1}(\Omega_m)+\left|\left|B_m^{-1}\right|\right|
\end{aligned} \tag{3.9.39}
$$

where Ω_m is defined in Lemma 4. Then,

$$\mathbf{E}\left|\left|\hat{B}_{\delta m}^{-1}\right|\right|^4 = O\left(\mathbf{E}\left|\left|\hat{B}_{\delta m} - B_{\delta m}\right|\right|^4 + \delta_m^{-4}\mathbb{P}(\Omega_m) + \left|\left|B_m^{-1}\right|\right|^4\right) = O(1), \tag{3.9.40}$$

so that both (3.9.7) and (4.4.34) are valid.

Recall (3.9.16), then using Lemmas 12 and 11, we have

$$\begin{aligned}
\Delta_2 &\leq \sum_{l=m_0+1}^{m_n} \sqrt{\mathbf{E}\left|\widehat{t}_m(y) - t(y)\right|^4 \Pr(\Theta_{l,m_0,\lambda})} \\
&\leq (m_n - (m_0+1))\sqrt{\mathbf{E}\left|\widehat{t}_m(y) - t(y)\right|^4 \max_{m>m_0} \Pr(\Theta_{m,m_0,\lambda})} \qquad (3.9.41)
\end{aligned}$$

where $\Theta_{l,m,\lambda}$ is defined in (3.9.15). Now recall that m_n is of order $O(\log n)$, therefore Δ_2 becomes

$$\begin{aligned}
\Delta_2 &\leq (m_n - (m_0+1))\sqrt{O\left(n^{-2}2^{2m}(\gamma_m^2+1)^2 + 2^{-4mr}\right)O\left(n^{-2}\right)} \\
&= O\left(n^{-1}\log n\sqrt{O\left(n^{-2}2^{2m}(\gamma_m^2+1)^2 + 2^{-4mr}\right)}\right) \qquad (3.9.42)
\end{aligned}$$

so that $\Delta_2 = O(n^{-1}) = o\left(\rho_{m_0 n}^2\right)$ as $n \to \infty$. Which completes the proof of Lemma 13.

Therefore, the following statement is true.

Theorem 4. Let twice continuously differentiable scaling function φ satisfy (3.2.1) and (3.2.2). Let functions $p(x)$ and $\Psi(x)$ be $r \geq 1/2$ times continuously differentiable in the neighborhood Ω_y of y and let $\Omega_{m,y} \subseteq \Omega_y$ where $\Omega_{m,y}$ is defined in (3.6.11). Let γ_m satisfy inequality (3.8.8) with $b = 0$. Construct EB estimator of the form (3.2.14) and choose $\widehat{m}$ according to (3.9.1) with λ defined in (3.9.30) where D given by (3.9.31). If, for any $k \in K_{m,y}$,

$$||u_{m,k}||_\infty \leq C_u 2^{\frac{m}{2}}\gamma_m, \tag{3.9.43}$$

then

$$\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+1+\max(\alpha,0)}} \log n\right). \tag{3.9.44}$$

3.10 Examples

3.10.1 Location Parameter Family

In the case of the location parameter family of distributions the sampling distribution takes the form $q(x/\theta) = q(x-\theta)$, EB estimator $t(y)$ in (3.1.1) is of the form

$$t(y) = y - \frac{\int_{-\infty}^{\infty}(y-\theta)q(y/\theta)g(\theta)d\theta}{\int_{-\infty}^{\infty}q(y/\theta)g(\theta)d\theta} \tag{3.10.1}$$

Indeed, replacing this latter in equation (4.1.5) we obtain

$$\int_{-\infty}^{\infty} q(x-\theta)u_{m,j}(x)dx = \int_{-\infty}^{\infty}(x-\theta)q(x-\theta)\varphi_{m,j}(x)dx \tag{3.10.2}$$

Notice here that both sides of the above equation are expressed in terms of convolutions, the left hand side involves $q(x)$ and $u_{m,j}(x)$, and the right hand side involves $xq(x)$ and $\varphi_{m,j}(x)$. Now taking the Fourier transform of both sides and using its properties in regards of convolutions yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q(x-\theta)u_{m,j}(x)e^{i\omega\theta}dxd\theta \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\theta)q(x-\theta)\varphi_{m,j}(x)e^{i\omega\theta}dxd\theta \tag{3.10.3}$$

Now, introduce the substitution $u = x-\theta$, then, our integrals become

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{\infty}^{-\infty} q(u)e^{i\omega(x-u)}(-du)u_{m,j}(x)dx &= \int_{-\infty}^{\infty}\int_{\infty}^{-\infty} uq(u)e^{i\omega(x-u)}(-du)\varphi_{m,j}(x)dx \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q(u)e^{-iu\omega}due^{i\omega x}u_{m,j}(x)dx &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uq(u)e^{-i\omega u}due^{i\omega x}\varphi_{m,j}(x)dx \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q(u)e^{iu(-\omega)}due^{i\omega x}u_{m,j}(x)dx &= -\frac{1}{i}\frac{d}{d\omega}\left\{\int_{-\infty}^{\infty} q(u)e^{iu(-\omega)}du\right\}\int_{-\infty}^{\infty} e^{i\omega x}\varphi_{m,j}(x)dx \\
\hat{q}(-\omega)\int_{-\infty}^{\infty} u_{m,j}(x)e^{ix\omega}dx &= -\frac{1}{i}\frac{d}{d\omega}\left\{\hat{q}(-\omega)\right\}\int_{-\infty}^{\infty}\varphi_{m,j}(x)e^{ix\omega}dx \\
\hat{q}(-\omega)\hat{u}_{m,j}(\omega) &= i\frac{d}{d\omega}\left\{\hat{q}(-\omega)\right\}\hat{\varphi}_{m,j}(\omega)
\end{aligned} \tag{3.10.4}$$

Consequently,

$$\hat{u}_{m,j}(\omega) = i\left[\hat{q}(-\omega)\right]^{-1}\frac{d}{d\omega}\left\{\hat{q}(-\omega)\right\}\hat{\varphi}_{m,j}(\omega) \tag{3.10.5}$$

Now we need to relate $\hat{\varphi}_{m,j}(\omega)$ to $\hat{\varphi}(\omega)$. Indeed, using the definition $\varphi_{m,j}(x) = 2^{\frac{m}{2}}\varphi(2^m x - j)$ and taking the Fourier transform we derive

$$\hat{\varphi}_{m,j}(\omega) = 2^{\frac{m}{2}} \int_{-\infty}^{\infty} \varphi(2^m x - j) e^{ix\omega} dx \tag{3.10.6}$$

Now make the substitution $u = 2^m x - j$, then our integral becomes

$$\begin{aligned}
\hat{\varphi}_{m,j}(\omega) &= 2^{\frac{-m}{2}} \int_{-\infty}^{\infty} \varphi(u) e^{i2^{-m}\omega(u+j)} du \\
\hat{\varphi}_{m,j}(\omega) &= 2^{\frac{-m}{2}} e^{i\omega 2^{-m} j} \int_{-\infty}^{\infty} \varphi(u) e^{iu(2^{-m}\omega)} du \\
\hat{\varphi}_{m,j}(\omega) &= 2^{\frac{-m}{2}} e^{i\omega 2^{-m} j} \hat{\varphi}(2^{-m}\omega) \qquad (3.10.7)
\end{aligned}$$

In addition, define $u_{m,j}(x)$ as

$$u_{m,j}(x) = 2^{m/2} U_m(2^m x - j) \tag{3.10.8}$$

Then, applying Fourier transform and making the substitution $s = 2^m x - j$, we obtain

$$\begin{aligned}
\hat{u}_{m,j}(\omega) &= 2^{\frac{m}{2}} \int_{-\infty}^{\infty} U_m(2^m x - j) e^{ix\omega} dx \\
&= 2^{\frac{-m}{2}} \int_{-\infty}^{\infty} U_m(s) e^{i2^{-m}\omega(s+j)} ds \\
&= 2^{\frac{-m}{2}} e^{i\omega 2^{-m} j} \int_{-\infty}^{\infty} U_m(s) e^{is(2^{-m}\omega)} ds \\
&= 2^{\frac{-m}{2}} e^{i\omega 2^{-m} j} \hat{U}_m(2^{-m}\omega) \qquad (3.10.9)
\end{aligned}$$

Consequently, applying (3.10.7) and (3.10.9) to (5.68) yields Consequently, applying (3.10.7) and (3.10.9) to (5.68) yields

$$\begin{aligned}
\hat{u}_{m,j}(\omega) &= i2^{-m/2} 2^{i2^{-m}\omega j} [\hat{q}(-\omega)]^{-1} \frac{d}{d\omega} \{\hat{q}(-\omega)\} \hat{\varphi}(2^{-m}\omega) \\
2^{\frac{-m}{2}} e^{i\omega 2^{-m} j} \hat{U}_m(2^{-m}\omega) &= i2^{-m/2} 2^{i2^{-m}\omega j} [\hat{q}(-\omega)]^{-1} \frac{d}{d\omega} \{\hat{q}(-\omega)\} \hat{\varphi}(2^{-m}\omega) \\
\hat{U}_m(2^{-m}\omega) &= i[\hat{q}(-\omega)]^{-1} \hat{q}'(-\omega) \hat{\varphi}(2^{-m}\omega) \qquad (3.10.10)
\end{aligned}$$

Notice that it suffices to evaluate $\hat{U}_m(\omega)$ apply the inverse Fourier transform and then use (3.10.8), where $\hat{U}_m(\omega)$ is given by

$$\hat{U}_m(\omega) = i[\hat{q}(-2^m\omega)]^{-1} \hat{q}'(-2^m\omega) \hat{\varphi}(\omega) \tag{3.10.11}$$

To calculate $\hat{U}_m(\omega)$, it suffices to evaluate the Fourier transform $\hat{q}(\omega)$ of the particular sampling distribution $q(x/\theta)$, and then use formula (3.10.11). Finally, to derive an expression for $\gamma^2(m)$ we can calculate the norm of $u_{m,j}(x)$ by applying the Parseval identity and then take into account that $K_{m,y}$ has finite number of terms to derive

$$\gamma_m^2 \asymp \int_{-\infty}^{\infty} \left|[\hat{q}(-2^m\omega)]^{-1}\, \hat{q}'(-2^m\omega)\hat{\varphi}(\omega)\right|^2 d\omega \tag{3.10.12}$$

Also, the following relation allows to check the validity of condition (3.9.43):

$$||u_{m,k}(x)||_\infty \le 2^{m/2}\, ||U_m(x)||_\infty \tag{3.10.13}$$

Now, in order to calculate minimax lower bounds for the risk in the case of the location parameter family of distributions, we need to find $\psi_{h,y}(\theta)$ and $\omega_{h,y}(x)$. Let $\psi_{h,y}(\theta)$ be solution of equation (3.4.5). It is easy to show that $\psi_{h,y}(\theta)$ is of the form $\psi_{h,y}(\theta) = \psi_h((\theta - y)/h)$, where the Fourier transform $\widehat{\psi}_h(\omega)$ of $\psi_h(.)$ is of the form It is easy to show that $\psi_{h,y}(\theta)$ is of the form where the Fourier transform $\widehat{\psi}_h(\omega)$ of $\psi_h(.)$ is

$$\widehat{\psi}_h(\omega) = \frac{\widehat{k}(\omega)}{\widehat{q}(\omega/h)} \tag{3.10.14}$$

To obtain expression for $\omega_{h,y}(x)$ recall that in the case of location parameter family equation (3.4.5) can be rewritten as

$$\begin{aligned} \omega_{h,y}(x) &= x - \int (x-\theta)\, q(x-\theta)\psi_{h,y}(\theta)d\theta \\ &= x - \omega_h\left(\frac{x-y}{h}\right) \end{aligned} \tag{3.10.15}$$

where, $\omega_h(.)$ is the inverse Fourier transform of

$$\widehat{\omega}_h(\omega) = i^{-1}\widehat{k}(\omega)\frac{\widehat{q}'(\omega/h)}{\widehat{q}(\omega/h)} \tag{3.10.16}$$

In this situation, the quantity $\rho_r(h)$ defined in (3.4.20) and $w_{h,y}(y)$ are given by

$$\rho_r(h) = \left[\max_{1\le j\le r}\left(h^{-j}[w_h^{(j)}(0)]\right)\right]^{-1}, \quad w_{h,y}(y) = y - w_h(0) \tag{3.10.17}$$

Below, we consider some special cases.

Example 9. Double-exponential distribution

Let $q(x/\theta)$ be the *pdf* of the double-exponential distribution

$$q(x/\theta) = \frac{1}{2\sigma} e^{-\frac{|x-\theta|}{\sigma}} \tag{3.10.18}$$

where $\sigma > 0$ is known. Then, it suffices to evaluate the Fourier transform of $q(x)$. Hence, calculating the Fourier transform for the latter function, one has

$$\begin{aligned}
\hat{q}(\omega) &= 1/2\sigma \int_{-\infty}^{\infty} e^{-\frac{|x|}{\sigma}} e^{ix\omega} dx \\
&= 1/2\sigma \int_{-\infty}^{0} e^{x/\sigma} e^{ix\omega} dx + 1/2\sigma \int_{0}^{\infty} e^{-x/\sigma} e^{ix\omega} dx \\
&= 1/2\sigma \int_{-\infty}^{0} e^{\frac{x(1+i\omega\sigma)}{\sigma}} dx + 1/2\sigma \int_{0}^{\infty} e^{-\frac{x(1-i\omega\sigma)}{\sigma}} dx \\
&= \frac{1}{2\sigma} \frac{\sigma}{1+i\omega\sigma} + \frac{1}{2\sigma} \frac{\sigma}{1-i\omega\sigma} \\
&= \frac{1}{2} \frac{1-i\omega\sigma+1+i\omega\sigma}{(1+i\omega\sigma)(1-i\omega\sigma)} \\
&= \frac{1}{1+\omega^2\sigma^2}
\end{aligned} \tag{3.10.19}$$

or

$$\hat{q}(\omega) = \frac{1}{1+\omega^2\sigma^2} \tag{3.10.20}$$

Now let us plug (3.10.20) in equation (3.10.11) and apply the inverse Fourier transform to try to obtain expression for $u_{m,j}(x)$. Indeed, applying the convolution property of Fourier transform, we obtain

Now, making the substitution $s = 2^m\omega\sigma$ on the integral part and noticing that it can be expressed as a derivative with respect to x, yields

$$\begin{aligned}
\varphi(x) * \int_{-\infty}^{\infty} \frac{-i2s}{1+s^2} e^{-ixs/\sigma} 2^{-m} ds &= \varphi(x) * \sigma \frac{d}{dx} \int_{-\infty}^{\infty} \frac{2}{1+s^2} e^{-i2^{-m}xs/\sigma} ds \\
&= \varphi(x) * \sigma \frac{d}{dx} e^{-\frac{2^{-m}|x|}{\sigma}} \\
&= \varphi(x) * sign(x) 2^{-m} e^{-\frac{2^{-m}|x|}{\sigma}} \\
&= \int_{-\infty}^{\infty} \varphi(t) 2^{-m} sign(x-t) \exp\left(-2^{-m}|x-t|/\sigma\right) dt
\end{aligned} \tag{3.10.21}$$

Consequently,

$$U_m(x) = \int_{-\infty}^{\infty} \varphi(t) 2^{-m} sign(x-t) \exp\left(-2^{-m}|x-t|/\sigma\right) dt \tag{3.10.22}$$

Hence, (3.10.8) can be expressed as

$$\begin{aligned} u_{m,j}(x) &= 2^{m/2}\int_{-\infty}^{\infty}\varphi(2^m t-j)2^{-m}sign(2^m x-j-2^m t+j)\exp\left(-2^{-m}|2^m x-j-2^m t+j|/\sigma\right)dt \\ &= 2^{m/2}\int_{-\infty}^{\infty}\varphi(2^m t-j)2^{-m}sign(2^m x-2^m t)\exp\left(-2^{-m}|2^m x-2^m t|/\sigma\right)dt \\ &= 2^{m/2}\int_{-\infty}^{\infty}\varphi(2^m t-j)sign(x-t)\exp\left(-|x-t|/\sigma\right)dt \qquad (3.10.23) \end{aligned}$$

or

$$u_{m,j}(x)=\int_{-\infty}^{\infty}\varphi_{m,j}(t)sign(x-t)\exp\left(-|x-t|/\sigma\right)dt \qquad (3.10.24)$$

In order to calculate (3.10.12) and (3.10.13) recall that $(\omega\sigma)^2+1\geq 2\omega\sigma$, which implies that

$$\begin{aligned} \left|[\hat{q}(-2^m\omega)]^{-1}\,\hat{q}'(-2^m\omega)\right| &= \frac{2\omega\sigma^2}{1+(\omega\sigma)^2} \\ &\leq \sigma \qquad (3.10.25) \end{aligned}$$

Therefore,

$$\begin{aligned} \gamma_m^2 &\asymp \int_{-\infty}^{\infty}\left|[\hat{q}(-2^m\omega)]^{-1}\,\hat{q}'(-2^m\omega)\hat{\varphi}(\omega)\right|^2 d\omega \\ &\asymp \int_{-\infty}^{\infty}\left|\hat{\varphi}(\omega)\right|^2 d\omega = 1 \qquad (3.10.26) \end{aligned}$$

Notice here that γ_m^2 is bounded above according to (3.8.8) with $\alpha=0$ and $b=0$.

Let us now investigate convergence rates of the EB estimators. Indeed, using the definition of $R_n(y)$, as $n\to\infty$, one has

$$\begin{aligned} R_n(y) &= \frac{2^m}{n}\left(\gamma^2(m)+1\right)+2^{-2mr} \\ &\asymp \frac{2^m}{n}(1+1)+2^{-2mr} \qquad (3.10.27) \end{aligned}$$

Therefore one needs to select m_0 that minimizes $R_n(y)$. Hence, differentiating $R_n(y)$ with respect to m and equating to zero yields

$$\begin{aligned} 2\ln 2\frac{2^m}{n}-2r2^{-2mr}\ln 2 &= 0 \\ \frac{2^m}{n}-r2^{-2mr} &= 0 \qquad (3.10.28) \end{aligned}$$

rearranging terms yields

$$\frac{1}{nr} = 2^{-m(2r+1)} \tag{3.10.29}$$

which implies that $2^{m_0} \sim n^{\frac{1}{2r+1}}$, and therefore $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+1}} \log n\right)$ by *theorem* 4, provided assumptions of Lemma 7 are met. Now that convergence rates are derived, it remains to verify whether condition (3.9.43) of Lemma 8 is satisfied. Indeed, using (3.10.13), we obtain

$$|U_m(x)| \leq \int_{-\infty}^{\infty} |\hat{\varphi}(\omega)|\, d\omega \asymp 1 \tag{3.10.30}$$

This implies that that $|u_{m,j}(x)| \leq C_u 2^{m/2}$. Hence, our condition is satisfied.

Now, to verify the lower bounds we use (3.10.16) to obtain,

$$\widehat{\omega}_h(\omega) = -i^{-1}\widehat{k}(\omega)\frac{2\omega\sigma^2 h^{-1}}{\sigma^2\omega^2 h^{-2} + 1} \tag{3.10.31}$$

It can be shown that

$$\omega_{h,y}(x) \quad = \quad x - 2\int k\,(t)\,\text{sign}(x - y - ht)\exp(-|x - y - ht|/\sigma)dt \tag{3.10.32}$$

$$\rho_r(h) \quad = \quad h^{-1} \tag{3.10.33}$$

and

$$|\omega_{h,y}(y)| = y + 2h\left|\int_{-1}^{1} k(t)\exp\{-h/\sigma|t|\}dt\right| \tag{3.10.34}$$

Therefore, $r_1 = r$ and $r_2 = 0$ in (3.4.20), and application of Theorem 2 yields $R_n(y) \geq Cn^{-\frac{2r}{2r+1}}$, so that EB estimator is optimal up to a logarithmic factor.

Example 10. Normal Distribution (θ, σ), with σ known.

Let $q(x/\theta)$ be the *pdf* of the normal distribution

$$q(x/\theta) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\theta)^2}{2\sigma^2}} \tag{3.10.35}$$

where $\sigma > 0$ is known. Then, it suffices to evaluate the Fourier transform of $q(x)$. Hence, calculating the

Fourier transform for the latter function, one has

$$\begin{aligned}
\hat{q}(\omega) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} e^{ix\omega} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left\{-\frac{x^2 - 2(i\omega\sigma^2)x + (i\omega\sigma^2)^2 - (i\omega\sigma^2)^2}{2\sigma^2}\right\} dx \\
&= \exp\left\{\frac{(i\omega\sigma^2)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left\{-\frac{x^2 - 2(i\omega\sigma^2)x + (i\omega\sigma^2)^2}{2\sigma^2}\right\} dx \\
&= \exp\left\{\frac{-\omega^2\sigma^2}{2}\right\} \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left\{-\frac{\left(x - i\omega\sigma^2\right)^2}{2\sigma^2}\right\} dx \qquad (3.10.36)
\end{aligned}$$

Notice that the integrand in the integral expression

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left\{-\frac{\left(x - i\omega\sigma^2\right)^2}{2\sigma^2}\right\} dx \qquad (3.10.37)$$

represents the Kernel of a normal distribution with mean $i\omega\sigma^2$ and standard deviation σ, and hence it is equal to 1. Consequently,

$$\hat{q}(\omega) = \exp\left\{\frac{-\omega^2\sigma^2}{2}\right\} \qquad (3.10.38)$$

Now let us try to obtain expression for $u_{m,j}(x)$. Indeed, plugging (3.10.38) in equation (3.10.11), applying the inverse Fourier transform and using its convolution property, we obtain

$$\begin{aligned}
U_m(x) &= \int_{-\infty}^{\infty} -i\omega 2^m \sigma^2 \hat{\varphi}(\omega) e^{-ix\omega} d\omega \\
&= 2^m \sigma^2 \frac{d}{dx}\int_{-\infty}^{\infty} \hat{\varphi}(\omega) e^{-ix\omega} d\omega \\
&= 2^m \sigma^2 \frac{d}{dx}\varphi(x) \qquad (3.10.39)
\end{aligned}$$

Consequently,

$$U_m(x) = 2^m \sigma^2 \varphi'(x) \qquad (3.10.40)$$

or in terms of $u_{m,j}(x)$, we have

$$u_{m,j}(x) = 2^m \sigma^2 \varphi'_{m,j}(x) \qquad (3.10.41)$$

Let us now calculate (3.10.12) and (3.10.13). Indeed, using (3.10.12) for $\gamma^2(m)$ and applying the Parseval's identity, yields

$$
\begin{aligned}
\gamma_m^2 &\asymp \int_{-\infty}^{\infty} \left|[\hat{q}(-2^m\omega)]^{-1}\,\hat{q}'(-2^m\omega)\hat{\varphi}(\omega)\right|^2 d\omega \\
&\asymp \int_{-\infty}^{\infty} \left|-i\omega 2^m \sigma^2 \hat{\varphi}(\omega)\right|^2 d\omega \\
&\asymp 2^{2m} \int_{-\infty}^{\infty} \left|\omega\hat{\varphi}(\omega)\right|^2 d\omega \\
&\asymp 2^{2m} \int_{-\infty}^{\infty} \left|\varphi'(x)\right|^2 dx
\end{aligned} \tag{3.10.42}
$$

Since $\varphi'(x)$ is square integrable, it follows that $\gamma_m^2 \asymp 2^{2m}$. Notice here that $\gamma^2(m)$ is bounded above according to (3.8.8) with $\alpha = 2$ and $b = 0$. As for condition (3.9.43), we use (3.10.13). Thus, one has

$$
\begin{aligned}
||u_{m,k}(x)||_\infty &\leq 2^{m/2}\,||U_m(x)||_\infty \\
&= 2^{m/2}\,\left|\left|2^m\sigma^2\varphi'(x)\right|\right|_\infty \\
&= 2^{m/2}2^m\,\left|\left|\sigma^2\varphi'(x)\right|\right|_\infty \\
&\asymp 2^{m/2}2^m
\end{aligned} \tag{3.10.43}
$$

which implies that the condition holds.

Let us now investigate convergence rates of the *EB* estimators. Indeed, using the definition of $R_n(y)$, as $n \to \infty$, one has

$$
\begin{aligned}
R_n(y) &= \frac{2^m}{n}\left(\gamma^2(m)+1\right) + 2^{-2mr} \\
&\asymp \frac{2^m}{n}\left(2^{2m}+1\right) + 2^{-2mr} \\
&\asymp \frac{2^{3m}}{n} + 2^{-2mr}
\end{aligned} \tag{3.10.44}
$$

Therefore one needs to select m_0 according to definition (3.8.9). Hence, differentiating (3.10.44) with respect to m and equating to zero yields

$$
\begin{aligned}
3\ln 2\frac{2^{3m}}{n} - 2r2^{-2mr}\ln 2 &= 0 \\
\frac{2^{3m}}{n} - r2^{-2mr} &= 0 \\
\frac{2^{(2r+3)m}}{n} - \frac{2r}{3} &= 0
\end{aligned} \tag{3.10.45}
$$

which implies that $2^{m_0} \sim n^{\frac{1}{2r+3}}$, and therefore $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+3}}\ \log n\right)$ by *theorem* 4, provided assumptions of Lemma 7 are met.

In addition, to verify the lower bounds we use (3.10.16) to obtain,

$$\widehat{\omega}_h(\omega) = -i^{-1}\widehat{k}(\omega)\frac{\omega\sigma^2}{h} \tag{3.10.46}$$

consequently,

$$\begin{aligned} \omega_{h,y}(x) &= x + \sigma^2 h^{-1} k'\left(\frac{x-y}{h}\right) && (3.10.47)\\ \rho_r(h) &= h^{r+1} && (3.10.48) \end{aligned}$$

and

$$|\omega_{h,y}(y)| = y + h^{-1} \tag{3.10.49}$$

Hence, $r_1 = r+1$ and $r_2 = 1$ in (3.4.20), and application of Theorem 2 yields $R_n(y) \geq Cn^{-\frac{2r}{2r+3}}$, so that EB estimator is optimal up to a logarithmic factor.

3.10.2 One- Parameter exponential Family

Let the sampling distribution belong to the one-parameter exponential family. That is,

$$q(x|\theta) = h(\theta)f(x)e^{-x^\alpha\theta},\ x \in X,\ \theta \in \Theta \tag{3.10.50}$$

where $h(\theta) \geq 0$ and $f(x) \geq 0$. Also, $h(\theta)$ can not depend on x and $f(x)$ can not depend on θ. Then, $u_{m,j}(x)$ is a solution of the following equation

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)h(\theta)e^{-x^\alpha\theta}u_{m,j}(x)dx &= \int_{-\infty}^{\infty}\theta f(x)h(\theta)e^{-x^\alpha\theta}\varphi_{m,j}(x)dx \\ \int_{-\infty}^{\infty} f(x)e^{-x^\alpha\theta}u_{m,j}(x)dx &= \int_{-\infty}^{\infty}\theta f(x)e^{-x^\alpha\theta}\varphi_{m,j}(x)dx && (3.10.51) \end{aligned}$$

Recall now that we are using wavelets with bounded support. That is, supp $\varphi \in [M_1, M_2]$. Define $K_2 = \frac{M_2+j}{2^m}$, and $K_1 = \frac{M_1+j}{2^m}$. Then, rearranging and integrating by parts in the right hand side we obtain

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)e^{-x^\alpha\theta}u_{m,j}(x)dx &= \int_{-\infty}^{\infty} \theta x^{\alpha-1}/x^{\alpha-1} f(x)\varphi_{m,j}(x)e^{-x^\alpha\theta}dx \\
&= \int_{-\infty}^{\infty} \theta x^{\alpha-1}e^{-x^\alpha\theta}\{f(x)\varphi_{m,j}(x)\}/x^{\alpha-1}dx \\
&= \left|-e^{-x^\alpha\theta}/\alpha\{f(x)\varphi_{m,j}(x)\}/x^{\alpha-1}\right|_{K_1}^{K_2} \\
&+ \int_{-\infty}^{\infty} \frac{1}{\alpha}e^{-x^\alpha\theta}\ \frac{d}{dx}\left\{\frac{f(x)\varphi_{m,j}(x)}{x^{\alpha-1}}\right\}dx
\end{aligned} \tag{3.10.52}$$

Now, since we are using wavelets with bounded support, and provided φ is continuous, the first term in the right hand side vanishes, so we obtain the following expression

$$\int_{-\infty}^{\infty} \left\{f(x)u_{m,j}(x) - \frac{1}{\alpha}\ \frac{d}{dx}\left\{\frac{f(x)\varphi_{m,j}(x)}{x^{\alpha-1}}\right\}\right\} e^{-x^\alpha\theta}dx = 0 \tag{3.10.53}$$

So that

$$u_{m,j}(x) = \frac{1}{\alpha f(x)}\ \frac{d}{dx}\left\{\frac{f(x)\varphi_{m,j}(x)}{x^{\alpha-1}}\right\}. \tag{3.10.54}$$

Now, in order to calculate lower bounds for the risk in the case of Examples 11 and 12, we need to find $\psi_{h,y}(\theta)$ and $w_{h,y}(x)$. Let $\psi_{h,y}(\theta)$ be solutions of equation (3.4.5) and $w_{h,y}(x)$ be defined in (3.4.7). In Examples 11 and 12, $q(x \mid \theta)$ is of the form

$$q(x \mid \theta) = \alpha h(\theta)x^{\alpha-1}e^{-x^\alpha\theta}, \tag{3.10.55}$$

It is straightforward to verify that,

$$\begin{aligned}
w_{h,y}(x) &= -\frac{f(x)}{\alpha x^{\alpha-1}}\ \frac{d}{dx}\left[\frac{1}{f(x)}k\left(\frac{x-y}{h}\right)\right] \\
&= \frac{f'(x)}{\alpha x^{\alpha-1}f(x)}k\left(\frac{x-y}{h}\right) - \frac{1}{\alpha h x^{\alpha-1}}k'\left(\frac{x-y}{h}\right).
\end{aligned} \tag{3.10.56}$$

For a fixed value of $y > 0$, one has $r_1 = r+1$ and $r_2 = 1$ in (3.4.20). Hence, application of Theorem 2 yields

$$R_n(y) \geq Cn^{-\frac{2r}{2r+3}}. \tag{3.10.57}$$

Below, we consider some special cases.

Example 11. Gamma Distribution (α, θ), with α known.

In this case, $f(x) = x^{\alpha-1}$, $h(\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)}$. Then, replacing this latter in equation (4.1.5) we obtain

$$\int_{-\infty}^{\infty} x^{\alpha-1} e^{-x^\alpha \theta} u_{m,j}(x) dx = \int_{-\infty}^{\infty} \theta x^{\alpha-1} e^{-x^\alpha \theta} \varphi_{m,j}(x) dx \tag{3.10.58}$$

Now integrating by parts the right hand side, and noting that we are using wavelets with bounded support, (3.10.58) becomes

$$\begin{aligned} \int_{0}^{\infty} x^{\alpha-1} e^{-x^\alpha \theta} u_{m,j}(x) dx &= -e^{-x^\alpha \theta} x^{\alpha-1} \varphi_{m,j}(x) |_{K_1}^{K_2} \\ &+ \int_{0}^{\infty} e^{-x^\alpha \theta} \frac{d}{dx} \left\{ x^{\alpha-1} \varphi_{m,j}(x) \right\} dx \\ &= \int_{0}^{\infty} e^{-x^\alpha \theta} \frac{d}{dx} \left\{ x^{\alpha-1} \varphi_{m,j}(x) \right\} dx \end{aligned} \tag{3.10.59}$$

Note in the above calculation that φ is assumed to be continuous, and therefore vanishes at M_1 and M_2, the boundaries. Hence,

$$\begin{aligned} u_{m,j}(x) &= x^{-\alpha+1} \frac{d}{dx} \left\{ x^{\alpha-1} \varphi_{m,j}(x) \right\} \\ &= (\alpha - 1) x^{-1} \varphi_{m,j}(x) + \frac{d}{dx} \varphi_{m,j}(x) \end{aligned} \tag{3.10.60}$$

So that

$$u_{m,j}(x) = (\alpha - 1) x^{-1} 2^{m/2} \varphi(2^m x - j) + 2^{3m/2} \varphi'(2^m x - j) \tag{3.10.61}$$

Therefore, for any $j \neq 0$,

$$\begin{aligned}
\left[\gamma_j^{(1)}(m)\right]^2 &= \int_{K_1}^{K_2} u_{m,j}^2(x)dx \\
&= \int_{K_1}^{K_2} \left[(\alpha-1)x^{-1}2^{m/2}\varphi(2^m x - j) + 2^{3m/2}\varphi'(2^m x - j)\right]^2 dx \\
&= 2^{-m}\int_{M_1}^{M_2} \left[(\alpha-1)\left|\frac{z+j}{2^m}\right|^{-1} 2^{m/2}\varphi(z) + 2^{3m/2}\varphi'(z)\right]^2 dz \\
&= 2^{2m}\int_{M_1}^{M_2} \left[(\alpha-1)\left|z+j\right|^{-1}\varphi(z) + \varphi'(z)\right]^2 dz \\
&= 2^{2m}\int_{M_1}^{M_2} \left[(\alpha-1)^2\left|z+j\right|^{-2}\varphi^2(z) + (\varphi'(z))^2\right] dz \\
&+ 2(\alpha-1)2^{2m}\int_{M_1}^{M_2} \left|z+j\right|^{-1}\varphi(z)\varphi'(z)dz \qquad (3.10.62)
\end{aligned}$$

Now, integrating by parts the last integral term, and using the fact that φ is continuous and has bounded support, we obtain

$$\begin{aligned}
\left[\gamma_j^{(1)}(m)\right]^2 &= 2^{2m}\int_{M_1}^{M_2} \left[(\alpha-1)^2\left|z+j\right|^{-2}\varphi^2(z) + (\varphi'(z))^2\right] dz \\
&+ 2(\alpha-1)2^{2m}\left[\left|\left|z+j\right|^{-1} 1/2\varphi^2(z)\right|_{M_1}^{M_2} + \frac{1}{2}\int_{M_1}^{M_2}\left|z+j\right|^{-2}\varphi^2(z)dz\right] \\
&= 2^{2m}\int_{M_1}^{M_2} \left[\left\{(\alpha-1)^2 + (\alpha-1)\right\}\left|z+j\right|^{-2}\varphi^2(z) + (\varphi'(z))^2\right] dz \\
&= 2^{2m}\int_{M_1}^{M_2} \left[\alpha(\alpha-1)\left|z+j\right|^{-2}\varphi^2(z) + (\varphi'(z))^2\right] dz \qquad (3.10.63)
\end{aligned}$$

provided $\varphi'(x)$ is square integrable. Notice here that

$$2^{2m}\alpha(\alpha-1)\left|M_2+j\right|^{-2} \leq 2^{2m}\int_{M_1}^{M_2}\alpha(\alpha-1)\left|z+j\right|^{-2}\varphi^2(z)dz \leq 2^{2m}\alpha(\alpha-1)\left|M_1+j\right|^{-2} \qquad (3.10.64)$$

Now let the value of y be such that $c_1 \leq y \leq c_2$ for some $0 < c_1 < c_2 < \infty$. Then, it is easy to verify that if $j \in K_{m,y}$, then $j \asymp 2^m$, which implies that $2^{2m}\int_{M_1}^{M_2}\alpha(\alpha-1)\left|z+j\right|^{-2}\varphi^2(z)dz \asymp 1$. Hence, provided $\varphi'(x)$ is square integrable, one has

$$\gamma_m^2 \asymp 2^{2m} \qquad (3.10.65)$$

Notice here also that γ_m^2 is bounded above according to (3.8.8) with $\alpha = 2$ and $b = 0$. Again we need to verify whether the condition (3.9.43) of Lemma 8 is met. Thus,

$$
\begin{aligned}
|u_{m,j}(x)| &= \left|\varphi'_{m,j}(x) + (\alpha-1)\frac{\varphi_{m,j}(x)}{x}\right| \\
&\le \left|\varphi'_{m,j}(x)\right| + \max_x \left|\frac{\alpha-1}{x}\right| . \left|\varphi_{m,j}(x)\right| \\
&\le 2^{\frac{1}{2}m}\int_{-\infty}^{\infty} 2^m \left|\omega\hat{\varphi}(\omega)\right| d\omega + \max_x \left|\frac{\alpha-1}{x}\right| 2^{\frac{1}{2}m}\int_{-\infty}^{\infty} \left|\hat{\varphi}(\omega)\right| d\omega \\
&\le 2\ 2^{\frac{1}{2}m}\int_{-\infty}^{\infty} 2^m \left|\omega\hat{\varphi}(\omega)\right| d\omega \qquad (3.10.66)
\end{aligned}
$$

which implies that the condition is satisfied. Since $\gamma_m^2 \asymp 2^{2m}$, It follows that $2^{m_0} \sim n^{\frac{1}{2r+3}}$, and therefore $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+3}}\ \log n\right)$ by *theorem* 4, provided assumptions of Lemma 7 are met. Hence, the EB estimator is optimal within a log-factor of n due to (3.10.57).

Example 12. Weibull Distribution (α, θ), with α known.

If $q(x|\theta)$ is the pdf of the Weibull distribution

$$q(x|\theta) = \alpha\theta x^{\alpha-1}e^{-x^{\alpha}\theta},\ x \ge 0,\ \theta > 0,\ \alpha \ge 1. \qquad (3.10.67)$$

In this case, $f(x) = x^{\alpha-1}$ and $h(\theta) = \alpha\theta$ and, according to (3.10.54), $u_{m,j}(x)$ is of the form

$$u_{m,j}(x) = \frac{2^{3m/2}\varphi'(2^m x - j)}{\alpha x^{\alpha-1}}. \qquad (3.10.68)$$

Therefore, for any $j \ne 0$,

$$[\gamma_j^{(1)}(m)]^2 \quad = \quad 2^{2\alpha m}\alpha^{-2}\int_{M_1}^{M_2} |z+j|^{-(2\alpha-2)}[\varphi'(z)]^2 dz, \qquad (3.10.69)$$

so that

$$2^{2\alpha m}\alpha^{-2}(M_2 - M_1)|M_2 + j|^{-(2\alpha-2)} \le [\gamma_j^{(1)}(m)]^2 \le 2^{2\alpha m}\alpha^{-2}(M_2 - M_1)|M_1 + j|^{-(2\alpha-2)}. \qquad (3.10.70)$$

Let the value of y be such that $c_1 \le y \le c_2$ for some $0 < c_1 < c_2 < \infty$. Then, it is easy to show that if $j \in K_{m,y}$, then $j \asymp 2^m$. Since the set $K_{m,y}$ has a finite number of terms, it follows from (3.10.70) that $\gamma_m^2 \asymp 2^{2m}$. Finally, it remains to verify whether the condition (3.9.43) of Theorem 4 holds. Indeed, it follows from (3.10.68) that for $j \in K_{m,y}$, one has

$$\sup_x |u_{m,j}(x)| \leq \sup_z \left| \frac{2^{3m/2}\varphi'(z) 2^{m(\alpha-1)}}{\alpha[M_1+j]^{\alpha-1}} \right| \leq C 2^{m/2}\gamma(m), \tag{3.10.71}$$

so Theorem 4 can be applied. Hence, under assumptions of Lemma 7, one has $\alpha = 2$, $2^{m_0} \sim n^{\frac{1}{2r+3}}$, and, $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+3}} \log n\right)$, by Theorem 4. Therefore, the EB estimator is optimal within a log-factor of n due to (3.10.57).

3.10.3 Scale parameter family

If $q(x|\theta)$ is a scale parameter family, $q(x|\theta) = \frac{1}{\theta} q\left(\frac{x}{\theta}\right)$, it is difficult to pinpoint a general rule for finding $u_{m,j}(x)$, however, as it follows from Gamma distribution case many particular cases can be treated. Below, we consider one more example.

Example 13. Uniform Distribution *Let $q(x \mid \theta)$ be given by*

$$q(x \mid \theta) = \theta^{-1}\, \mathbf{1}\left(0 < x < \theta\right), \quad a \leq \theta \leq b. \tag{3.10.72}$$

Then, equation (4.1.5) is of the form

$$\int_0^\theta \theta^{-1} u_{m,j}(x) dx = \int_0^\theta \varphi_{m,j}(x) dx \tag{3.10.73}$$

Taking derivatives with respect to θ of both sides of (3.10.73) and replacing θ by x, we derive

$$u_{m,j}(x) = 2^{-m/2} \int_{M_1}^{2^m x - j} \varphi(z) dz + x 2^{m/2} \varphi(2^m x - j). \tag{3.10.74}$$

Since $a \leq \theta \leq b$, then also one has $a \leq x \leq b$, and it is easy to check that

$$\int_a^b x^2 2^m \varphi^2(2^m x - j) dx \asymp 1, \quad \int_a^b \left(2^{-m/2} \int_{M_1}^{2^m x - j} \varphi(z) dz \right)^2 dx = O(2^{-m}), \tag{3.10.75}$$

as $m \to \infty$. Then, $\gamma_m \asymp 1$, $\alpha = 0$ and condition (3.9.43) holds. Therefore, $2^{m_0} \sim n^{\frac{1}{2r+1}}$ and, by Theorem 4, $\mathbf{E}\left|\widehat{t}_{\widehat{m}}(y) - t(y)\right|^2 = O\left(n^{-\frac{2r}{2r+1}} \log n\right)$.

Now, in order to calculate lower bounds for the risk we need to find $\psi_{h,y}(\theta)$ and $w_{h,y}(x)$. Then, according

(3.4.5) and (3.4.7), functions $\psi_{h,y}(\theta)$ *and* $w_{h,y}(x)$ *satisfy equations*

$$\begin{aligned} \int_x^b \frac{1}{\theta}\psi_{h,y}(\theta)d\theta &= k\left(\frac{x-y}{h}\right), &(3.10.76)\\ \int_x^b \psi_{h,y}(\theta)d\theta &= \omega_{h,y}(x), &(3.10.77) \end{aligned}$$

Now, differentiating both sides of the first equation with respect to x *and solving for* $\psi_{h,y}(\theta)$*, we obtain*

$$\psi_{h,y}(\theta) = -\frac{\theta}{h}k'\left(\frac{\theta-y}{h}\right) \tag{3.10.78}$$

It can be shown that

$$w_{h,y}(x) = xk\left(\frac{x-y}{h}\right) + hK\left(\frac{x-y}{h}\right) \tag{3.10.79}$$

where, $K'(z) = k(z)$*. Notice that* $r_1 = r$ *and* $r_2 = 0$*, hence, applying Theorem 2, we obtain the following lower bounds for the risk* $R_n(y) \geq Cn^{-\frac{2r}{2r+1}}$*, so that the EB estimator is optimal, up to a logarithmic factor.*

CHAPTER 4: ANISOTROPIC DE-NOISING IN FUNCTIONAL DECONVOLUTION MODEL WITH DIMENSION-FREE CONVERGENCE RATES

4.1 Formulation of the Problem

Consider the problem of estimating a periodic $(r+1)$-dimensional function $f(\mathbf{u}, x)$ with $\mathbf{u} = (u_1, \cdots, u_r) \in [0,1]^r$ $x \in [0,1]$, based on observations from the following noisy convolution

$$y(\mathbf{u}, t) = \int_0^1 g(\mathbf{u}, t-x) f(\mathbf{u}, x) dx + \varepsilon z(\mathbf{u}, t), \quad \mathbf{u} \in [0,1]^r, t \in [0,1]. \tag{4.1.1}$$

Here, ε is a positive small parameter such that asymptotically $\varepsilon \to 0$, function $g(.,.)$ in (4.1.1) is assumed to be known and $z(\mathbf{u}, t)$ is an $r+1$-dimensional Gaussian white noise, i.e., a generalized $r+1$-dimensional Gaussian field with covariance function

$$\mathbf{E}[z(\mathbf{u}_1, t_1) z(\mathbf{u}_2, t_1)] \quad = \quad \delta(t_1 - t_2) \prod_{l=1}^{r} \delta(u_{1l} - \mathbf{u}_{2l}), \tag{4.1.2}$$

where $\delta(\cdot)$ denotes the Dirac δ-function and $\mathbf{u}_{il} = (u_{i1}, \cdots, u_{ir}) \in [0,1]^r$, $i = 1, 2$.

Denote

$$h(\mathbf{u}, t) = \int_0^1 g(\mathbf{u}, t-x) f(\mathbf{u}, x) dx. \tag{4.1.3}$$

Then, equation (4.1.5) can be rewritten as

$$y(\mathbf{u}, t) = h(\mathbf{u}, t) + \varepsilon z(\mathbf{u}, t) \tag{4.1.4}$$

In order to simplify the narrative, we start with the two dimensional version of equation (4.1.1)

$$y(u, t) = \int_0^1 g(u, t-x) f(u, x) dx + \varepsilon z(u, t), \quad u, t \in [0,1]. \tag{4.1.5}$$

The sampling version of problem (4.1.5) appears as

$$y(u_l, t_i) = \int_0^1 g(u_l, t_i - x) f(u_l, x) dx + \sigma \xi_{li}, \quad l = 1, \cdots, M, \ i = 1, \cdots, N, \tag{4.1.6}$$

where σ is a positive constant independent of N and M, $u_l = l/M$, $t_i = i/N$ and ξ_{li} are i.i.d normal variables with $\mathbf{E}(\xi_{li}) = 0$, and $\mathbf{E}(\xi_{l_1 i_1} \xi_{l_2 i_2}) = \delta(l_1 - l_2)\delta(i_1 - i_2)$.

Equation (5.1.1) seems to be equivalent to M separate convolution equations

$$y_l(t_i) = \int_0^1 f_l(x) g_l(t_i - x) dx + \sigma z_{li}, \quad l = 1, \cdots, M, \ i = 1, \cdots, N, \tag{4.1.7}$$

with $y_l(t_i) = y(u_l, t_i)$, $f_l(x) = f(u_l, x)$ and $g_l(t_i - x) = g(u_l, t_i - x)$. This is, however, not true since the solution of equation (5.1.1) is a **two-dimensional function** while solutions of equations (4.1.7) are M unrelated functions $f_i(t)$. In this sense, problem (4.1.5) and its sampling equivalent (5.1.1) are functional deconvolution problems.

Functional deconvolution problems have been introduced in Pensky and Sapatinas (2009) and further developed in Pensky and Sapatinas (2010, 2011). However, Pensky and Sapatinas (2009, 2010, 2011) considered a different version of the problem where $f(u,t)$ was a function of one variable, i.e. $f(u,t) \equiv f(t)$. Their interpretation of functional deconvolution problem was motivated by solution of inverse problems in mathematical physics and multichannel deconvolution in engineering practices. Functional deconvolution problem of types (4.1.5) and (5.1.1) are motivated by experiments where one needs to recover a two-dimensional function using observations of its convolutions along profiles $u = u_i$. This situation occurs, for example, in geophysical explorations, in particular, the ones which rely on inversions of seismic signals (see, e.g., monographs of Robinson *et al.* (1996) and Robinson (1999) and, e.g., papers of Wason *et al.* (1984), Berkhout (1986)and Heimer and Cohen (2008)).

In seismic exploration, a short duration seismic pulse is transmitted from the surface, reflected from boundaries between underground layers, and received by an array of sensors on the Earth surface. The signals are transmitted along straight lines called profiles. The received signals, called seismic traces, are analyzed to extract information about the underground structure of the layers along the profile. Subsequently, these traces can be modeled under simplifying assumptions as noisy outcomes of convolutions between reflectivity sequences which describe configuration of the layers and the short wave like function (called wavelet in geophysics) which corresponds to convolution kernel. The objective of seismic deconvolution is to estimate the reflectivity sequences from the measured traces. In the simple case of one layer and a single profile, the boundary will be described by an univariate function which is the solution of the convolution equation. The next step is usually to combine the recovered functions which are defined on the set of parallel planes passing through the profiles into a multivariate function which provides the exhaustive picture of the structure of the underground layers. This is usually accomplished by interpolation techniques. However, since the layers are intrinsically anisotropic (may have different structures in various directions) and spatially inhomogeneous (may experience, for example, sharp breaks), the former approach ignores the anisotropic and spatially inhomogeneous nature of the two-dimensional function describing the layer and loses precision by analyzing each profile separately.

This chapter will attempt to address three points:

i) Construction of a feasible procedure $\widehat{f}(\mathbf{u},t)$ for estimating the $(r+1)$-dimensional function $f(\mathbf{u},t)$ which achieves optimal rates of convergence (up to inessential logarithmic terms). We require $\widehat{f}(\mathbf{u},t)$ to be adaptive with respect to smoothness constraints on f. In this sense, this study is related to a multitude of papers which offered wavelet solutions to deconvolution problems (see, e.g., Donoho (1995), Abramovich and Silverman (1998), Pensky and Vidakovic (1999), Walter and Shen (1999), Fan and Koo (2002), Kalifa and Mallat (2003), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Donoho and Raimondo (2004), Johnstone and Raimondo (2004), Neelamani, Choi and Baraniuk (2004) and Kerkyacharian, Picard and Raimondo (2007)).

ii) Identification of the best achievable accuracy under smoothness constraints on f. We focus here on obtaining fast rates of convergence. In this context, we prove that considering multivariate functions with 'mixed' smoothness and hyperbolic wavelet bases allows to obtain rates which are free of dimension and, as a consequence, faster than the usual ones. In particular, the present study is related to anisotropic de-noising explored by, e.g., Kerkyacharian, Lepski and Picard (2001, 2008). We compare our functional classes as well as our rates with the results obtained there.

iii) Comparison of the two-dimensional version of the functional deconvolution procedure studied in the present chapter to the separate solutions of convolution equations. We show especially that our approach delivers estimators with higher precision. For this purpose, in Section 4.5, we consider a discrete version of functional deconvolution problem (5.1.1) (rather than the continuous equation (4.1.5)) and compare its solution with solutions of M separate convolution equations (4.1.7). We show that, unless the function f is very smooth in the direction of the profiles, very spatially inhomogeneous along the other direction and the number of profiles is very limited, functional deconvolution solution has a better precision than the combination of M solutions of separate convolution equations.

4.2 Estimation Algorithm

In what follows, $\langle \cdot, \cdot \rangle$ denotes the inner product in the Hilbert space $L^2([0,1])$ (the space of squared-integrable functions defined on the unit interval $[0,1]$), i.e., $\langle f, g \rangle = \int_0^1 f(t)\overline{g(t)}dt$ for $f, g \in L^2([0,1])$. We also denote the complex conjugate of a by $\bar{a}$. Let $e_m(t) = e^{i2\pi mt}$ be a Fourier basis on the interval $[0,1]$. Let $h_m(u) = \langle e_m, h(u,\cdot)\rangle$, $y_m(u) = \langle e_m, y(u,\cdot)\rangle$, $z_m(u) = \langle e_m, z(u,\cdot)\rangle$, $g_m(u) = \langle e_m, g(u,\cdot)\rangle$ and $f_m(u) = \langle e_m, f(u,\cdot)\rangle$ be functional Fourier coefficients of functions h, y, z, g and f respectively. Then, applying the Fourier transform to equation (4.1.4), one obtains for any $u \in [0,1]$

$$y_m(u) = g_m(u) f_m(u) + \varepsilon z_m(u) \tag{4.2.1}$$

and

$$h_m(u) = g_m(u) f_m(u). \tag{4.2.2}$$

Consider a bounded bandwidth periodized wavelet basis (e.g., Meyer-type) $\psi_{j,k}(t)$ and finitely supported periodized s_0-regular wavelet basis (e.g., Daubechies) $\eta_{j',k'}(u)$. The choice of the Meyer wavelet basis for t is motivated by the fact that it allows easy evaluation of the the wavelet coefficients in the Fourier domain while finitely supported wavelet basis gives more flexibility in recovering a function which is spatially inhomogeneous in u. Let m_0 and m_0' be the lowest resolution levels for the two bases and denote the scaling functions for the bounded bandwidth wavelet by $\psi_{m_0-1,k}(t)$ and the scaling functions for the finitely supported wavelet by $\eta_{m_0'-1,k'}(u)$. Then, $f(x,u)$ can be expanded into wavelet series as

$$f(u,x) = \sum_{j=m_0-1}^{\infty} \sum_{j'=m_0'-1}^{\infty} \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \beta_{j,k,j',k'} \psi_{j,k}(x) \eta_{j',k'}(u). \tag{4.2.3}$$

Denote $\beta_{j,k}(u) = \langle f, \psi_{j,k}\rangle$, then, $\beta_{j,k,j',k'} = \langle \beta_{j,k}(u), \eta_{j',k'}(u)\rangle$. If $\psi_{j,k,m} = \langle e_m, \psi_{j,k}\rangle$ are Fourier coefficients of $\psi_{j,k}$, then, by formula (5.3.4) and Plancherel's formula, one has

$$\beta_{j,k}(u) = \sum_{m \in W_j} f_m(u) \overline{\psi_{j,k,m}} = \sum_{m \in W_j} \frac{h_m(u)}{g_m(u)} \overline{\psi_{j,k,m}}, \tag{4.2.4}$$

where, for any $j \geq j_0$,

$$W_j = \{m : \psi_{jkm} \neq 0\} \subseteq 2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}], \tag{4.2.5}$$

due to the fact that Meyer wavelets are band-limited (see, e.g., Johnstone, Kerkyacharian, Picard & Raimondo (2004), Section 3.1). Therefore, $\beta_{j,k,j',k'}$ are of the form

$$\beta_{j,k,j',k'} = \sum_{m \in W_j} \overline{\psi_{j,k,m}} \int \frac{h_m(u)}{g_m(u)} \eta_{j',k'}(u) du, \tag{4.2.6}$$

and allow the unbiased estimator

$$\widetilde{\beta}_{j,k,j',k'} = \sum_{m \in W_j} \overline{\psi_{j,k,m}} \int \frac{y_m(u)}{g_m(u)} \eta_{j',k'}(u) du. \tag{4.2.7}$$

We now construct a hard thresholding estimator of $f(u,t)$ as

$$\widehat{f}(u,t) = \sum_{j=m_0-1}^{J-1} \sum_{j'=m_0'-1}^{J'-1} \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \widehat{\beta}_{jk,j'k'} \psi_{jk}(t) \eta_{j'k'}(u) \tag{4.2.8}$$

where

$$\widehat{\beta}_{j,k,j',k'} = \widetilde{\beta}_{j,k,j',k'} \mathbf{1} \left(\left| \widetilde{\beta}_{j,k,j',k'} \right| > \lambda_{j\varepsilon} \right). \tag{4.2.9}$$

and the values of J, J' and $\lambda_{j\varepsilon}$ will be defined later.

In what follows, we use the symbol C for a generic positive constant, independent of ε, which may take different values at different places.

4.3 Smoothness classes and minimax lower bounds

4.3.1 Smoothness classes

It is natural to consider *anisotropic* multivariate functions, i.e., functions whose smoothness is different in different directions. In order to construct Besov classes of mixed regularity, we choose $l \geq \max_j s_j$ and define

$$B_{p,\infty}^{s_1,\ldots,s_d} = \left\{ f \in \mathbb{L}_p, \sum_{e \subset \{1,\ldots,d\}} \sup_{t>0} \sup_{j \in e} t_j^{-s_j} \Omega^{l,e}(f, t^e)_p < \infty \right\}. \tag{4.3.1}$$

It is proved in, e.g., Heping (2004) that under appropriate (regularity) conditions which we are omitting here, classes (4.3.1) can be expressed in terms of hyperbolic-wavelet coefficients, thus, providing a convenient generalization of the one-dimensional Besov $B^s_{p,\infty}$ spaces. Furthermore, Heping (2004) considers more general Besov classes of mixed regularity $B^{s_1,\dots,s_d}_{p,q}$ that correspond to $q < \infty$ rather than $q = \infty$. In this discussion, we shall assume that the hyperbolic wavelet basis satisfies required regularity conditions and follow Heping (2004) definition of Besov spaces of mixed regularity

$$B^{s_1,\dots,s_d}_{p,q} = \left\{ f \in L^2(U) : \left(\sum_{j_1,\dots,j_d} 2^{(\sum_{i=1}^d j_i[s_i+\frac{1}{2}-\frac{1}{p}])q} \left(\sum_{k_1,\dots,k_d} |\beta_{j_1,k_1\dots,j_d k_d}|^p \right)^{\frac{q}{p}} \right)^{1/q} < \infty \right\}. \quad (4.3.2)$$

4.3.2 Lower bounds for the risk: two-dimensional case

Denote $U = [0,1] \times [0,1]$ and

$$s_i^* = s_i + 1/2 - 1/p, \quad s_i' = s_i + 1/2 - 1/p', \quad i = 1,2, \quad p' = \min\{p, 2\}. \quad (4.3.3)$$

In what follows, we assume that the function $f(u,t)$ belongs to a two-dimensional Besov ball as described above ($d = 2$), so that wavelet coefficients $\beta_{jk,j'k'}$ satisfy the following condition

$$B^{s_1,s_2}_{p,q}(A) = \left\{ f \in L^2(U) : \left(\sum_{j,j'} 2^{(js_1^*+j's_2^*)q} \left(\sum_{k,k'} |\beta_{jk,j'k'}|^p \right)^{\frac{q}{p}} \right)^{1/q} \le A \right\}. \quad (4.3.4)$$

Below, we construct minimax lower bounds for the L^2-risk. For this purpose, we define the minimax L^2-risk over the set V as

$$R_\varepsilon(V) = \inf_{\tilde{f}} \sup_{f \in V} \mathbf{E} \left|\left| \tilde{f} - f \right|\right|^2, \quad (4.3.5)$$

where $||g||$ is the L^2-norm of a function $g(\cdot)$ and the infimum is taken over all possible estimators $\tilde{f}(\cdot)$ (measurable functions taking their values in a set containing V) of $f(\cdot)$.

Assume that functional Fourier coefficients $g_m(u)$ of function $g(u,t)$ are uniformly bounded from above and below, that is, there exist positive constants ν, and C_1 and C_2, independent of m and u such that

$$C_1 \, |m|^{-2\nu} \le |g_m(u)|^2 \le C_2 \, |m|^{-2\nu}. \quad (4.3.6)$$

In order to construct lower bounds for the L^2-risk of any estimator $\tilde{f}_n$ of f, we consider two cases, the case when $f(u,t)$ is dense in both variables (the dense-dense case) and the case when $f(u,t)$ is dense in u

and sparse in t. The derivation is based on Lemma A.1 of Bunea, Tsybakov and Wegkamp (2007) which we reformulate here for the case of squared risk.

Lemma 14. [Bunea, Tsybakov, Wegkamp (2007), Lemma A.1] Let Ω be a set of functions of cardinality $\text{card}(\Omega) \geq 2$ such that

(i) $||f-g||^2 \geq 4\delta^2, \quad \text{for } f, g \in \Omega, \;\; f \neq g,$

(ii) the Kullback divergences $K(P_f, P_g)$ between the measures P_f and P_g satisfy the inequality $K(P_f, P_g) \leq \log(\text{card}(\Omega))/16, \quad \text{for} \quad f, g \in \Omega.$

Then, for some absolute positive constant C, one has

$$\inf_{T_n} \sup_{f \in \ \Omega} \mathbf{E}_f \, ||T_n - f||^2 \geq C\delta^2. \tag{4.3.7}$$

<u>The dense-dense case.</u> Let ω be the matrix with components $\omega_{k,k'} = \{0,1\}$, $k = 0, \cdots, 2^j - 1$, $k' = 0, \cdots, 2^{j'} - 1$. Denote the set of all possible values ω by Ω and let the functions $f_{j,j'}$ be of the form

$$f_{jj'}(t,u) = \gamma_{jj'} \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \omega_{k,k'} \psi_{jk}(t) \eta_{j'k'}(u). \tag{4.3.8}$$

Note that matrix ω has $N = 2^{j+j'}$ components, and, hence, cardinality of the set of such matrices is $\text{card}(\Omega) = 2^N$. Since $f_{jj'} \in B^{s_1 s_2}_{p,q}(A)$, direct calculations show that $\gamma_{jj'} \leq A 2^{-j(s_1+1/2)-j'(s_2+1/2)}$, so that we choose $\gamma_{jj'} = A 2^{-j(s_1+1/2)-j'(s_2+1/2)}$. If $\tilde{f}_{jj'}$ is of the form (4.3.8) with $\tilde{\omega}_{k,k'} \in \Omega$ instead of $\omega_{k,k'}$, then, the L^2-norm of the difference is of the form

$$\left|\left|\tilde{f}_{jj'} - f_{jj'}\right|\right|^2 \;\; = \;\; \gamma^2_{jj'} \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \mathbf{1}\left(\tilde{\omega}_{k,k'} \neq \omega_{k,k'}\right) = \gamma^2_{jj'} \rho(\tilde{\omega}, \omega) \tag{4.3.9}$$

where $\rho(\tilde{\omega}, \omega) = \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \mathbf{1}\left(\tilde{\omega}_{k,k'} \neq \omega_{k,k'}\right)$ is the Hamming distance between the binary sequences ω and $\tilde{\omega}$. In order to find a lower bound for the last expression, we apply the Varshamov-Gilbert lower bound (see Tsybakov (2008), page 104) which states that one can choose a subset Ω_1 of Ω, of cardinality at least $2^{N/8}$ such that $\rho(\tilde{\omega}, \omega) \geq N/8$ for any $\omega, \tilde{\omega} \in \Omega_1$. Hence, for any $\omega, \tilde{\omega} \in \Omega_1$ one has $\left|\left|\tilde{f}_{jj'} - f_{jj'}\right|\right|^2 \geq \gamma^2_{jj'} 2^{j+j'}/8$. Note that Kullback divergence can be written as

$$K(f, \tilde{f}) \;\; = \;\; (2\varepsilon^2)^{-1} \left|\left|(\tilde{f} - f) * g\right|\right|^2. \tag{4.3.10}$$

Since $|\omega_{jj'} - \tilde{\omega}_{jj'}| \le 1$, plugging f and $\tilde{f}$ into (4.3.10), using Plancherel's formula and recalling that $|\psi_{j,k,m}| \le 2^{-j/2}$, we derive

$$K(f, \tilde{f}) \quad \le \quad (2\varepsilon^2)^{-1} 2^{-j} \gamma_{jj'}^2 \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \sum_{m \in W_j} \int_0^1 \eta_{j'k'}^2(u)\, g_m^2(u)\, du. \tag{4.3.11}$$

Using (4.3.6), we obtain

$$2^{-j} \sum_{m \in W_j} \int_0^1 \eta_{j'k'}^2(u) g_m^2(u) du \quad \le \quad C_2 2^{-j} \sum_{m \in W_j} |m|^{-2\nu} \int_0^1 \eta_{j'k'}^2(u) du \le C_3 2^{-2\nu j}, \tag{4.3.12}$$

so that

$$K(f, \tilde{f}) \le C \varepsilon^{-2} \gamma_{jj'}^2 2^{j+j'} 2^{-2\nu j}. \tag{4.3.13}$$

Now, applying Lemma 14 with

$$\delta^2 = \gamma_{jj'}^2 2^{j+j'}/32 = A^2 2^{-2s_1 j - 2s_2 j'}/32 \tag{4.3.14}$$

one obtains constraint $2^{-j(2s_1+2\nu+1)-j'(2s_2+1)} \le C\varepsilon^2/A^2$ on j, j' and ε where C is an absolute constant. Denote

$$\tau_\varepsilon = \log_2(CA^2\varepsilon^{-2}). \tag{4.3.15}$$

Thus, we need to choose combination of j and j' which solves the following optimization problem

$$2js_1 + 2j's_2 \Rightarrow \min \;\; j(2s_1 + 2\nu + 1) + j'(2s_2 + 1) \ge \tau_\varepsilon, \quad j, j' \ge 0. \tag{4.3.16}$$

It is easy to check that solution of this linear constraint optimization problem is of the form $\{j, j'\} = \left\{(2s_1 + 2\nu + 1)^{-1}\tau_\varepsilon, 0\right\}$ if $s_2(2\nu + 1) > s_1$, and $\{j, j'\} = \left\{0, (2s_2 + 1)^{-1}\tau_\varepsilon\right\}$ if $s_2(2\nu + 1) \le s_1$. Plugging those values into (4.3.14), obtain

$$\delta^2 = \begin{cases} CA^2\,(\varepsilon^2/A^2)^{\frac{2s_2}{2s_2+1}}, & \text{if } \; s_1 > s_2(2\nu+1), \\ CA^2\,(\varepsilon^2/A^2)^{\frac{2s_1}{2s_1+2\nu+1}}, & \text{if } \; s_1 \le s_2(2\nu+1). \end{cases} \tag{4.3.17}$$

The sparse-dense case. Let ω be the vector with components $\omega_{k'} = \{0, 1\}$. Denote Ω the set of all possible ω and let the functions $f_{j,j'}$ be of the form

$$f_{jj'}(t,u) = \gamma_{jj'} \sum_{k'=0}^{2^{j'}-1} \omega_{k'} \psi_{jk}(t) \eta_{j'k'}(u) \tag{4.3.18}$$

Note that vector ω has $N = 2^{j'}$ components, and, hence, its cardinality is $\text{card}(\Omega) = 2^N$. Since $f_{jj'} \in B_{p,q}^{s_1 s_2}(A)$, direct calculations show that $\gamma_{jj'} \le A 2^{-js_1^* - j'(s_2+1/2)}$, so we choose $\gamma_{jj'} = A 2^{-js_1^* - j'(s_2+1/2)}$. If $\tilde{f}_{jj'}$ is of the form (4.3.18) with $\tilde{\omega}_{k,k'} \in \Omega$ instead of $\omega_{k,k'}$, then, calculating the L^2 norm of the difference similarly to dense-dense case, obtain

$$\left|\left|\tilde{f}_{jj'} - f_{jj'}\right|\right|^2 \;=\; \gamma_{jj'}^2 \sum_{k'=0}^{2^{j'}-1} \mathbf{1}\left(\tilde{\omega}_{k'} \ne \omega_{k'}\right) \ge \gamma_{jj'}^2 2^{j'}/8. \tag{4.3.19}$$

Similarly to dense-dense case, using formulae (4.3.6) and (4.3.10), Plancherel's formula and $|\psi_{j,k,m}| \le 2^{-j/2}$, we derive

$$K(f,\tilde{f}) \le (2\varepsilon^2)^{-1} \gamma_{jj'}^2 \sum_{k'=0}^{2^{j'}-1} 2^{-j} \sum_{m \in W_j} \int_0^1 \eta_{j'k'}^2(u) g_m^2(u) du \le C(2\varepsilon^2)^{-1} \gamma_{jj'}^2 2^{j'} 2^{-2\nu j}.$$

Now, applying Lemma 14 with

$$\delta^2 = \gamma_{jj'}^2 2^{j'}/32 = A^2 2^{-2s_1' j - 2s_2 j'}/32 \tag{4.3.20}$$

one obtains constraint $2^{-j(2s_1'+2\nu) - j'(2s_2+1)} \le C\varepsilon^2/A^2$ on j, j' and ε where C is an absolute constant. Thus, we need to choose combination of j and j' which delivers solution to the following linear optimization problem $\min\{2js_1 + 2j's_2\}$ subject to constraint

$$2js_1 + 2j's_2 \Rightarrow \min \quad s.t. \quad j(2s_1' + 2\nu) + j'(2s_2+1) \ge \tau_\varepsilon, \quad j, j' \ge 0. \tag{4.3.21}$$

It is easy to check that solution of this linear constraint optimization problem is of the form $\{j, j'\} = \left\{(2s_1' + 2\nu)^{-1}\tau_\varepsilon, 0\right\}$ if $2\nu s_2 > s_1'$, and $\{j, j'\} = \left\{0, (2s_2+1)^{-1}\tau_\varepsilon\right\}$ if $2\nu s_2 \le s_1'$. Plugging those values into (4.3.20), obtain

$$\delta^2 = \begin{cases} CA^2 \left(\varepsilon^2/A^2\right)^{\frac{2s_2}{2s_2+1}}, & \text{if } \; 2\nu s_2 \le s_1', \\ CA^2 \left(\varepsilon^2/A^2\right)^{\frac{2s_1'}{2s_1'+2\nu}}, & \text{if } \; 2\nu s_2 > s_1'. \end{cases} \tag{4.3.22}$$

Then, the following theorem gives the minimax lower bounds for the L^2-risk of any estimator $\tilde{f}_n$ of f.

Theorem 5. Let $\min\{s_1, s_2\} \geq \max\{1/p, 1/2\}$ with $1 \leq p, q \leq \infty$, let $A > 0$ and s_i', $i = 1, 2$, be defined in (5.4.1). Then, under assumption (4.3.6), as $\varepsilon \to 0$

$$R_\varepsilon(B_{p,q}^{s_1,s_2}(A)) \geq CA^2 \left(\frac{\varepsilon^2}{A^2}\right)^d \tag{4.3.23}$$

where

$$d = \min\left(\frac{2s_2}{2s_2+1}, \frac{2s_1}{2s_1+2\nu+1}, \frac{2s_1'}{2s_1'+2\nu}\right). \tag{4.3.24}$$

Note that the value of d in (4.3.24) can be re-written as

$$d = \begin{cases} \frac{2s_2}{2s_2+1}, & \text{if } \ s_1 > s_2(2\nu+1), \\ \frac{2s_1}{2s_1+2\nu+1}, & \text{if } \ (\frac{1}{p} - \frac{1}{2})(2\nu+1) \leq s_1 \leq s_2(2\nu+1), \\ \frac{2s_1'}{2s_1'+2\nu}, & \text{if } \ s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1). \end{cases} \tag{4.3.25}$$

Remark 3. Note that the rates obtained here are in fact the worst rate associated to the one dimensional problem in each direction, which is not surprising since a function of only one variable and constant in the other direction, e.g. $f(u_1, u_2) = h(u_1)$ belongs to $B_{p,q}^{s_1,s_2}(A)$ as soon as h belongs to a ball of the usual one dimensional Besov space $B_{p,q}^{s_1}$, for any s_2.

Also it is worthwhile to observe that the third rate (involving s_1') corresponds in dimension one to a 'sparse' rate. Hence we observe here the so-called 'elbow phenomenon' occurring only along the direction 2, because we are considering a L^2 loss and the problem has a degree of ill-posedness ν precisely in this direction.

4.4 Minimax upper bounds

Before deriving expressions for the minimax upper bounds for the risk, we formulate several useful lemmas which give some insight into the choice of the thresholds $\lambda_{j\varepsilon}$ and upper limits J and J' in the sums in (4.2.8).

Lemma 15. Let $\widetilde{\beta}_{j,k,j',k'}$ be defined in (4.2.7). Then, under assumption (4.3.6), one has

$$\text{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \asymp \varepsilon^2 2^{2j\nu}. \tag{4.4.1}$$

Proof of Lemma 15. Let us derive an expression for the upper bound of the variance of (4.2.7). Subtracting (4.2.6) from (4.2.7) we obtain

$$\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'} = \varepsilon \sum_{m \in W_j} \overline{\psi_{j,k,m}} \int_0^1 \frac{z_m(u)}{g_m(u)} \eta_{j',k'}(u) du. \tag{4.4.2}$$

Now, before we proceed to the derivation of the upper bound of the variance, let us first state a result that will be used in our calculation. Recall from stochastic calculus that for any function $F(t,u) \in L^2([0,1] \times [0,1])$, one has

$$\mathbf{E}\left[\int_0^1 \int_0^1 F(t,u) dz(t,u) du\right]^2 \quad = \quad \int_0^1 \int_0^1 F^2(t,u) dt du. \tag{4.4.3}$$

Hence, recalling that $z_m(u) = \int z(u,t) e_m(t) dt$, choosing

$$F(t,u) = \sum_{m \in W_j} \overline{\psi_{j,k,m}} \frac{e_m(t)}{g_m(u)} \eta_{j',k'}(u), \tag{4.4.4}$$

squaring both sides of (4.4.2), taking expectation and using the relation (4.4.3), we obtain

$$\begin{aligned}
\mathrm{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) &= \varepsilon^2 \, \mathbf{E} \left| \sum_{m \in W_j} \overline{\psi_{j,k,m}} \int_0^1 \int_0^1 \frac{\eta_{j',k'}(u)}{g_m(u)} e_m(t) dz(u,t) du \right|^2 \\
&= \varepsilon^2 \int_0^1 \int_0^1 \sum_m \sum_{m'} \frac{\overline{\psi_{j,k,m}} \psi_{j,k,m'}}{g_m(u) \overline{g_{m'}}(u)} \overline{e_m}(t) e_{m'}(t) |\eta_{j',k'}(u)|^2 dt du \\
&= \varepsilon^2 \sum_{m \in W_j} |\psi_{j,k,m}|^2 \int_0^1 \frac{|\eta_{j',k'}(u)|^2}{|g_m(u)|^2} du, \qquad (4.4.5)
\end{aligned}$$

since in the double summation above, all terms involving $m \neq m'$ vanish due to $\int_0^1 e_m(t) e_{m'}(t) dt = 0$. Consequently, Taking into account (4.2.5), (4.3.6) and the fact that $|\psi_{j,k,m}| \leq 2^{-j/2}$, obtain

$$\mathrm{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \asymp \varepsilon^2 \sum_{m \in W_j} |\psi_{j,k,m}|^2 |m|^{2\nu} \int_0^1 \left|\eta^2_{j',k'}(u)\right| du \asymp \varepsilon^2 2^{2j\nu} \tag{4.4.6}$$

so that (4.4.1) holds.

Lemma 15 suggests that thresholds $\lambda_{j\varepsilon}$ should be chosen as

$$\lambda_{j\varepsilon} = C_\beta \sqrt{\ln(1/\varepsilon)}\, 2^{j\nu}\, \varepsilon \tag{4.4.7}$$

where C_β is some positive constant independent of ε. We choose J and J' as

$$2^J = (\varepsilon^2)^{-\frac{1}{2\nu+1}}, \quad 2^{J'} = (\varepsilon^2)^{-1}. \tag{4.4.8}$$

Note that the choices of J, J' and $\lambda_{j\varepsilon}$ are independent of the parameters, s_1, s_2, p, q and A of the Besov ball $B^{s_1 s_2}_{p,q}(A)$, and therefore our estimator (4.2.8) is adaptive with respect of those parameters.

The next two lemmas provide upper bounds for the wavelet coefficients and the large deviation inequalities for their estimators.

Lemma 16. Under assumption (5.4.2), one has

$$\sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} |\beta_{j,k,j',k'}|^2 \le A^2 2^{-2(js'_1+j's'_2)} \tag{4.4.9}$$

for any $j, j' \ge 0$.

Proof of Lemma 16 First note that, under assumption (5.4.2), one has

$$\sum_{k,k'} |\beta_{j,k,j',k'}|^p \le A^p 2^{-p\left[(js_1+j's_2)+(\frac{1}{2}-\frac{1}{p})(j+j')\right]} \tag{4.4.10}$$

If $p \le 2$, one has $p' = p$, $s'_i = s_i + 1/2 - 1/p$, $i = 1, 2$, and

$$\sum_{k,k'} |\beta_{j,k,j',k'}|^2 \quad \le \quad \sum_{k,k'} |\beta_{j,k,j',k'}|^p \left\{ \max_{k,k'} |\beta_{j,k,j',k'}|^p \right\}^{(2-p)/p} \le A^2 2^{-2(js'_1+j's'_2)}. \tag{4.4.11}$$

If $p \ge 2$, then $p' = 2$, $s'_i = s_i$, $i = 1, 2$, and, applying the Cauchy-Schwarz inequality, one obtain

$$\sum_{k,k'} |\beta_{j,k,j',k'}|^2 \quad \le \quad \left(\sum_{k,k'} |\beta_{j,k,j',k'}|^p \right)^{2/p} \left(\sum_{k,k'} 1 \right)^{(1-2/p)} \le A^2 2^{-2\left[(js_1+j's_2)\right]}, \tag{4.4.12}$$

which completes the proof.

Lemma 17. Let $\widetilde{\beta}_{j,k,j',k'}$ and $\lambda_{j\varepsilon}$ be defined by formulae (4.2.7) and (4.4.7), respectively. Define, For some positive constant α, the set

$$\Theta_{jk,j'k',\alpha} = \{\Theta : \left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right| > \alpha\lambda_{j\varepsilon}\}. \tag{4.4.13}$$

Then, under assumption (4.3.6), as $\varepsilon \to 0$, one has

$$\Pr\left(\Theta_{jk,j'k',\alpha}\right) = O\left(\varepsilon^{\frac{\alpha^2 C_\beta^2}{2\sigma_0^2}} [\ln(1/\varepsilon)]^{-\frac{1}{2}}\right) \tag{4.4.14}$$

where $\sigma_0^2 = \left(\frac{8\pi}{3}\right)^{2\nu} \frac{1}{C_1}$ and C_1 is defined in (4.3.6).

Proof of Lemma 17 Observe that $\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}$ is a zero-mean Gaussian random variable with variance given by (4.4.6), so that

$$\text{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \le \varepsilon^2 \left(\frac{8\pi}{3}\right)^{2\nu} \frac{2^{2\nu j}}{C_1} = \sigma_0^2\varepsilon^2 2^{2\nu j} \tag{4.4.15}$$

Denoting by $\bar{\Phi}(x) = 1 - \Phi(x)$ where $\Phi(x)$ is the standard normal c.d.f. and recalling that $\bar{\Phi}(x) \le (x\sqrt{2\pi})^{-1}\exp(-x^2/2)$ if $x > 0$, we derive

$$\begin{aligned} \Pr\left(\Omega_{jk,j'k',\alpha}\right) &= \Pr\left(|\xi_{j,k,j',k'}| > \alpha\lambda_{j\varepsilon}\right) = 2\bar{\Phi}\left(\alpha\lambda_{j\varepsilon}(\sigma_0\varepsilon 2^{\nu j})^{-1}\right) \\ &\le 2\bar{\Phi}\left(\alpha C_\beta(\sigma_0)^{-1}\sqrt{\ln(1/\varepsilon)}\right) \le \frac{2\sigma_0}{\alpha C_\beta\sqrt{2\pi\ln(1/\varepsilon)}}\, \varepsilon^{\frac{\alpha^2 C_\beta^2}{2\sigma_0^2}} \end{aligned} \tag{4.4.16}$$

which completes the proof.

Using the statements above, we can derive upper bounds for the minimax risk of the estimator (4.2.8).

Now denote

$$\chi_{\varepsilon,A} = A^{-2}\varepsilon^2 \ln(1/\varepsilon), \tag{4.4.17}$$

$$2^{j_0} = (\chi_{\varepsilon,A})^{-\frac{d}{2s_1'}}, \quad 2^{j_0'} = (\chi_{\varepsilon,A})^{-\frac{d}{2s_2'}} \tag{4.4.18}$$

and observe that with J and J' given by (4.4.8), the estimation error can be decomposed into the sum of four components as follows

$$\mathbf{E}\left|\left|\widehat{f}_n - f\right|\right|^2 \le \sum_{j,k,j',k'} \mathbf{E}\left|\left|\widehat{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right|\right|^2 \le R_1 + R_2 + R_3 + R_4, \tag{4.4.19}$$

where

$$R_1 = \sum_{k=0}^{2^{m_0}-1} \sum_{k'=0}^{2^{m_0'}-1} \mathrm{Var}(\widetilde{\beta}_{m_0,k,m_0',k'}), \tag{4.4.20}$$

$$R_2 = \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} \mathbf{E}\left[\left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right|^2 \mathbf{1}\left(\left|\widetilde{\beta}_{j,k,j',k'}\right| > \lambda_{j\varepsilon}\right)\right], \tag{4.4.21}$$

$$R_3 = \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} |\beta_{j,k,j',k'}|^2 \Pr\left(\left|\widetilde{\beta}_{j,k,j',k'}\right| < \lambda_{j\varepsilon}\right), \tag{4.4.22}$$

$$R_4 = \left(\sum_{j=J}^{\infty} \sum_{j'=m_0'}^{J'-1} + \sum_{j=m_0}^{J-1} \sum_{j'=J'}^{\infty} + \sum_{j=J}^{\infty} \sum_{j'=J'}^{\infty}\right) \sum_{k,k'} |\beta_{j,k,j',k'}|^2 . \tag{4.4.23}$$

For R_1, using (4.4.1), derive, as $\varepsilon \to 0$,

$$R_1 \le C\varepsilon^2 = O\left(A^2\, \chi_{\varepsilon,A}^d\right). \tag{4.4.24}$$

To calculate R_4, we apply Lemma 16 and use (4.4.8) obtaining, as $\varepsilon \to 0$,

$$\begin{aligned} R_4 &= O\left(\left(\sum_{j \ge J} \sum_{j' \ge m_0'} + \sum_{j \ge m_0} \sum_{j' \ge J'}\right) A^2 2^{-2js_1' - 2j's_2'}\right) = O\left(A^2 2^{-2Js_1} + A^2 2^{-2J's_2}\right) \\ &= O\left(A^2 (\varepsilon^2)^{\frac{2s_1'}{2\nu+1}} + A^2 (\varepsilon^2)^{2s_2'}\right) = O\left(A^2 \chi_{\varepsilon,A}^d\right). \end{aligned} \tag{4.4.25}$$

Then, our objective is to prove that, as $\varepsilon \to 0$, one has $R_i = O\left(A^2 \chi_{\varepsilon,A}^d [\ln(1/\varepsilon)]^{d_1}\right)$.

Now, note that each R_2 and R_3 can be partitioned into the sum of two errors as follows

$$R_2 \le R_{21} + R_{22}, \quad R_3 \le R_{31} + R_{32}, \tag{4.4.26}$$

where

$$
\begin{aligned}
R_{21} &= \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} \mathbf{E}\left[\left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right|^2 \mathbf{1}\left(\left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right| > \frac{\lambda_{j\varepsilon}}{2}\right)\right] \quad (4.4.27)\\
R_{22} &= \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} \mathbf{E}\left[\left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right|^2 \mathbf{1}\left(|\beta_{j,k,j',k'}| > \frac{1}{2}\lambda_{j\varepsilon}\right)\right]. \quad (4.4.28)\\
R_{31} &= \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} |\beta_{j,k,j',k'}|^2 \Pr\left(\left|\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'}\right| > \frac{\lambda_{j\varepsilon}}{2}\right), \quad (4.4.29)\\
R_{32} &= \sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} |\beta_{j,k,j',k'}|^2 \mathbf{1}\left(|\beta_{j,k,j',k'}| \le \frac{3\lambda_{j\varepsilon}}{2}\right). \quad (4.4.30)
\end{aligned}
$$

Combining (4.4.27) and (4.4.29), and applying Cauchy-Schwarz inequality and Lemma 17 with $\alpha = 1/2$, one derives

$$
\begin{aligned}
R_{21} + R_{31} &= O\left(\sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} 2^{j+j'} \varepsilon^{\frac{C_\beta^2}{16\sigma_0^2}} [\ln(1/\varepsilon)]^{-\frac{1}{4}} \sqrt{\varepsilon^4 2^{4j\nu+j'}}\right)\\
&= O\left(2^{J(2\nu+1)} 2^{3J'/2} (\varepsilon)^{2+\frac{C_\beta^2}{16\sigma_0^2}}\right) = O\left((\varepsilon^2)^{\frac{C_\beta^2}{32\sigma_0^2}-\frac{3}{2}}\right). \quad (4.4.31)
\end{aligned}
$$

Hence, due to condition (4.4.45), one has, as $\varepsilon \to 0$,

$$
R_{21} + R_{31} \le C\varepsilon^2 = O\left(A^2 \chi_{\varepsilon,A}^d\right). \quad (4.4.32)
$$

For the sum of R_{22} and R_{32}, using (4.4.1) and (4.4.7), we obtain

$$
\Delta = R_{22} + R_{32} = O\left(\sum_{j=m_0}^{J-1} \sum_{j'=m_0'}^{J'-1} \sum_{k,k'} \min\left\{\beta_{j,k,j',k'}^2, \varepsilon^2 \ln(1/\varepsilon)\ 2^{2j\nu}\right\}\right). \quad (4.4.33)
$$

Then, Δ can be partitioned into the sum of three components Δ_1, Δ_2 and Δ_3 according to three different sets of indices:

$$\Delta_1 = O\left(\left\{\sum_{j=j_0+1}^{J-1}\sum_{j'=m_0'}^{J'-1} + \sum_{j=m_0}^{J-1}\sum_{j'=j_0'+1}^{J'-1}\right\} A^2 2^{-2js_1'-2j's_2'}\right), \tag{4.4.34}$$

$$\Delta_2 = O\left(\sum_{j=m_0}^{j_0}\sum_{j'=m_0'}^{j_0'} \varepsilon^2 \ln(1/\varepsilon)\, 2^{j(2\nu+1)+j'}\, \mathbf{1}\left(2^{j(2\nu+1)+j'} \le \chi_{\varepsilon,A}^{d-1}\right)\right), \tag{4.4.35}$$

$$\Delta_3 = O\left(\sum_{j=m_0}^{j_0}\sum_{j'=m_0'}^{j_0'} A^{p'} 2^{-p'js_1'-p'j's_2'}\left(\varepsilon^2 \ln(1/\varepsilon) 2^{2j\nu}\right)^{1-p'/2} \mathbf{1}\left(2^{j(2\nu+1)+j'} > \chi_{\varepsilon,A}^{d-1}\right)\right), \tag{4.4.36}$$

where d is defined in (4.3.24). It is easy to see that for Δ_1 given in (4.4.34) and j_0 and j_0' given by (4.4.18), as $\varepsilon \to 0$, one has

$$\Delta_1 = O\left(A^2\, \chi_{\varepsilon,A}^d\right), \tag{4.4.37}$$

For Δ_2 defined in (4.4.35), obtain

$$\Delta_2 = O\left(\varepsilon^2 \ln(1/\varepsilon) \chi_{\varepsilon,A}^{d-1}\right) = O\left(A^2\, \chi_{\varepsilon,A}^d\right), \quad \varepsilon \to 0. \tag{4.4.38}$$

In order to construct upper bounds for Δ_3 in (4.4.36), we need to consider three different cases.

Case 1: $s_1 \ge s_2(2\nu+1)$. In this case, $d = 2s_2/(2s_2+1)$ and

$$\begin{aligned}
\Delta_3 &\le CA^2(\chi_{\varepsilon,A})^{1-p'/2} \sum_{j=m_0}^{j_0} 2^{-j[p's_1'-2\nu(1-p'/2)]} \sum_{j'=m_0'}^{j_0'} 2^{-p'j's_2'}\, \mathbf{1}\left(2^{j'} > (\chi_{\varepsilon,A})^{d-1} 2^{-j(2\nu+1)}\right) \\
&\le CA^2(\chi_{\varepsilon,A})^{(1-p'/2)+p's_2'(1-d)} \sum_{j=m_0}^{j_0} 2^{-j[p's_1'-2\nu(1-p'/2)-p'(2\nu+1)s_2']} \\
&= CA^2(\chi_{\varepsilon,A})^d \sum_{j=m_0}^{j_0} 2^{-j[p's_1-p's_2(2\nu+1)]},
\end{aligned} \tag{4.4.39}$$

so that, as $\varepsilon \to 0$,

$$\Delta_3 = O\left(A^2\, \chi_{\varepsilon,A}^d\, [\ln(1/\varepsilon)]^{\mathbf{1}(s_1 = s_2(2\nu+1))}\right). \tag{4.4.40}$$

Case 2: $(\frac{1}{p}-\frac{1}{2})(2\nu+1) < s_1 < s_2(2\nu+1)$. In this case, $d = 2s_1/(2s_1+2\nu+1)$ and

$$
\begin{aligned}
\Delta_3 &\leq CA^2(\chi_{\varepsilon,A})^{1-p'/2} \sum_{j=m_0}^{j_0} 2^{-j[p's_1'-2\nu(1-p'/2)]} \sum_{j'=m_0'}^{j_0'} 2^{-p'j's_2'} \mathbf{1}\left(2^j > (\chi_{\varepsilon,A})^{\frac{d-1}{2\nu+1}} 2^{-\frac{j'}{2\nu+1}}\right) \\
&\leq CA^2(\chi_{\varepsilon,A})^{(1-p'/2)+p'\frac{(1-d)}{1+2\nu}(s_1-(2\nu+1)(1/p'-1/2)} \sum_{j'=m_0'}^{j_0'} 2^{-j'p'[s_2'-s_1/(2\nu+1)+(1/2-1/p')]} \\
&\leq CA^2(\chi_{\varepsilon,A})^{d} \sum_{j'=m_0'}^{j_0'} 2^{-j'p'[s_2-s_1/(2\nu+1)]},
\end{aligned}
\tag{4.4.41}
$$

so that, as $\varepsilon \to 0$,

$$
\Delta_3 = O\left(A^2\, \chi_{\varepsilon,A}^d\right). \tag{4.4.42}
$$

Case 3: $s_1 \leq (\frac{1}{p}-\frac{1}{2})(2\nu+1)$. In this case, $d = 2s_1'/(2s_1'+2\nu)$ and $p \leq 2$. Then, since $ps_1' - 2\nu(1-p/2) = p[s_1-(1/p-1/2)(2\nu+1)] \leq 0$, one has

$$
\begin{aligned}
\Delta_3 &\leq CA^2(\chi_{\varepsilon,A})^{1-p'/2} \sum_{j=m_0}^{j_0} 2^{-j[ps_1'-2\nu(1-p/2)]} \\
&\leq CA^2(\chi_{\varepsilon,A})^{1-p'/2}\, 2^{j_0 p[(1/p-1/2)(2\nu+1)-s_1]}\, [\ln(1/\varepsilon)]^{\mathbf{1}(s_1=(1/p-1/2)(2\nu+1))}.
\end{aligned}
\tag{4.4.43}
$$

Plugging in j_0 of the form (4.4.18), obtain as $\varepsilon \to 0$

$$
\Delta_3 = O\left(A^2\, \chi_{\varepsilon,A}^d\, [\ln(1/\varepsilon)]^{\mathbf{1}(s_1=(1/p-1/2)(2\nu+1))}\right). \tag{4.4.44}
$$

Now, combining formulae (4.4.19)–(4.4.44) leads to the next theorem which provides upper bounds results.

Theorem 6. Let $\widehat{f}(.,.)$ be the wavelet estimator defined in (4.2.8), with J and J' given by (4.4.8). Let condition (4.3.6) hold and $\min\{s_1, s_2\} \geq \max\{1/p, 1/2\}$, with $1 \leq p, q \leq \infty$. If C_β in (4.4.7) is such that

$$
C_\beta^2 \geq 80(C_1)^{-1}(2\pi/3)^{2\nu} \tag{4.4.45}
$$

where C_1 is defined in (4.3.6), then, as $\varepsilon \to 0$,

$$
\sup_{f \in B_{p,q}^{s_1,s_2}(A))} \mathbf{E}\left|\left|\widehat{f} - f\right|\right|^2 \leq CA^2 \left(\frac{\varepsilon^2 \ln(1/\varepsilon)}{A^2}\right)^d \ln\left(\frac{1}{\varepsilon}\right)^{d_1} \tag{4.4.46}
$$

where d is defined in (4.3.24) and

$$d_1 = \mathbf{1}(s_1 = s_2(2\nu+1)) + \mathbf{1}(s_1 = (2\nu+1)(1/p - 1/2)). \tag{4.4.47}$$

Remark 4. Looking at the previous results, we conclude that the rates obtained by the wavelet estimator defined in (4.2.8) are optimal, in the minimax sense, up to logarithmic factors. These factors are standard and coming from the thresholding procedure.

4.5 Sampling version and comparison with separate deconvolution recoveries

Consider now the sampling version (5.1.1) of the problem (4.1.5). In this case, the estimators of wavelet coefficients $\beta_{j,k,j',k'}$ can be constructed as

$$\widetilde{\beta}_{j,k,j',k'} = \frac{1}{M} \sum_{m \in W_j} \overline{\psi_{j,k,m}} \sum_{l=1}^{M} \frac{y_m(u_l)}{g_m(u_l)}\, \eta_{j',k'}(u_l). \tag{4.5.1}$$

In practice, $\widetilde{\beta}_{j,k,j',k'}$ are obtained simply by applying discrete wavelet transform to vectors $y_m(\cdot)/g_m(\cdot)$.

Recall that the continuous versions (4.2.7) of estimators (4.5.1) have $\text{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \asymp \varepsilon^2 2^{2j\nu}$ (see formula (4.4.1)). In order to show that equation (5.1.1) is the sampling version of (4.1.5) with $\varepsilon^2 = \sigma^2/(MN)$, one needs to show that, in the discrete case, $\text{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \asymp \sigma^2 (MN)^{-1} 2^{2j\nu}$. This indeed is accomplished by the following Lemma.

Lemma 18. Let $\widetilde{\beta}_{j,k,j',k'}$ be defined in (4.5.1). Then, under assumption (4.3.6), as $MN \to \infty$, one has

$$\text{Var}\left(\widetilde{\beta}_{j,k,j',k'}\right) \asymp \sigma^2 (MN)^{-1} 2^{2j\nu}. \tag{4.5.2}$$

Proof of Lemma 18. Subtracting $\beta_{j,k,j',k'}$ from (4.5.1), one obtains

$$\widetilde{\beta}_{j,k,j',k'} - \beta_{j,k,j',k'} = \frac{\sigma}{M} \sum_{m \in W_j} \overline{\psi_{j,k,m}} \sum_{l=1}^{M} \frac{z_m(u_l)}{g_m(u_l)}\, \eta_{j',k'}(u_l). \tag{4.5.3}$$

where $z_m(u_l) = y_m(u_l) - h_m(u_l)$. Since Fourier transform is an orthogonal transform, one has $\mathbf{E}[z_{m_1}(u_{l_1}) z_{m_2}(u_{l_2})] = 0$ if $l_1 \neq l_2$ and $\mathbf{E}[z_{m_1}(u_l) z_{m_2}(u_l)] = 0$, so that

$$\mathbf{E}[z_{m_1}(u_{l_1}) z_{m_2}(u_{l_2})] = \frac{\sigma^2}{N} \delta(m_1 - m_2)\delta(l_1 - l_2). \tag{4.5.4}$$

Therefore,

$$\begin{aligned}\mathrm{Var}(\widetilde{\beta}_{j,k,j',k'}) &= \frac{\sigma^2}{M^2N}\sum_{m\in W_j}|\psi_{j,k,m}|^2\sum_{l=1}^{M}\frac{1}{|g_m(u_l)|^2}|\eta_{j',k'}(u_l)|^2\\ &\asymp \frac{\sigma^2 2^{2j\nu}}{MN}\sum_{m\in W_j}|\psi_{j,k,m}|^2\,\frac{1}{M}\sum_{l=1}^{M}|\eta_{j',k'}(u_l)|^2 \asymp \frac{\sigma^2 2^{2j\nu}}{MN},\end{aligned} \tag{4.5.5}$$

which completes the proof.

Using tools developed in Pensky and Sapatinas (2009) and Lemma 18, it is easy to formulate the lower and the upper bounds for convergence rates of the estimator (4.2.8) with $\widehat{\beta}_{jk,j'k'}$ given by (4.2.9) and the values of $\lambda_{j\varepsilon}$ and J, J' defined in (4.4.7) and (4.4.8), respectively. In particular, we obtain the following statement.

Theorem 7. Let $\min\{s_1, s_2\} \geq \max\{1/p, 1/2\}$ with $1 \leq p, q \leq \infty$, let $A > 0$ and s_i^* be defined in (5.4.1). Then, under assumption (4.3.6), as $MN \to \infty$, for some absolute constant $C > 0$ one has

$$R_{(MN)}(B_{p,q}^{s_1,s_2}(A)) \geq C(\sigma^2(MN)^{-1})^d. \tag{4.5.6}$$

Moreover, if $\widehat{f}(.,.)$ is the wavelet estimator defined in (4.2.8), $\min\{s_1, s_2\} \geq \max\{1/p, 1/2\}$, and J and J' given by (4.4.8), then, under assumption (4.3.6), as $MN \to \infty$,

$$\sup_{f\in B_{p,q}^{s_1,s_2}(A))} \mathbf{E}\left|\left|\widehat{f} - f\right|\right|^2 \leq C(\sigma^2(MN)^{-1}\ln(MN))^d\,(\ln(MN))^{d_1}. \tag{4.5.7}$$

where d and d_1 are defined in (4.3.24) and (4.4.47), respectively.

Now, let us compare the rates in Theorem 7 with the rates obtained by recovering each deconvolution $f_l(t) = f(u_l, t)$, $u_l = l/M$, $l = 1, \cdots, M$, separately, using equations (4.1.7). In order to do this, we need to determine in which space functions $f_l(x)$ are contained. The following lemma provides the necessary conclusion.

Lemma 19. Let $f \in B_{p,q}^{s_1,s_2}(A)$ with $s_1 \geq \max\{1/p, 1/2\}$, $s_2 > \max\{1/p, 1/2\}$ and $1 \leq p, q \leq \infty$. Then, for any $l = 1, ..., M$, we have

$$f_l(t) = f(u_l, t) \in B_{p,q}^{s_1}(\tilde{A}). \tag{4.5.8}$$

Proof of Lemma 19. Recall that

$$f(u,t) = \sum_{j,k} \sum_{j',k'} \beta_{j,k,j',k'} \psi_{j,k}(t) \eta_{j',k'}(u) \quad \text{and} \quad f_l(t) = \sum_{j,k} b_{j,k}^{(l)} \psi_{j,k}(t) \eta_{j',k'}(u_l), \tag{4.5.9}$$

so that

$$b_{j,k}^{(l)} = \sum_{j'=0}^{\infty} \sum_{k' \in K_l} \beta_{j,k,j',k'} 2^{j'/2} \eta(2^{j'} u_l - k'), \tag{4.5.10}$$

where the set $K_l = \{k' : \ \eta(2^{j'} u_l - k') \neq 0\}$ is finite for any l due to finite support of η.

Thus, since $p \geq 1$, for any $\delta > 0$, one has

$$\begin{aligned} \sum_{k=0}^{2^j-1} |b_{j,k}^{(l)}|^p &\leq C \sum_{k=0}^{2^j-1} \left[\sum_{j'=0}^{\infty} \sum_{k' \in K_l} |\beta_{j,k,j',k'}| \, 2^{j'(1+\delta)/2} \, 2^{-j'\delta/2} \right]^2 \\ &\leq C \sum_{k=0}^{2^j-1} \left(\sum_{j'=0}^{\infty} \sum_{k' \in K_l} |\beta_{j,k,j',k'}|^p \, 2^{j'(1+\delta)p/2} \right) \left(\sum_{j'=0}^{\infty} \sum_{k' \in K_l} \left(2^{-j'\delta/2} \right)^{\frac{p}{p-1}} \right)^{p-1}. \end{aligned} \tag{4.5.11}$$

Then, for any $q \geq 1$, one has

$$B_j = \left(\sum_{j'=0}^{\infty} 2^{j'(1+\delta)p/2} \sum_{k,k'} |\beta_{j,k,j',k'}|^p \right)^{q/p}. \tag{4.5.12}$$

If $q/p \geq 1$, then, using Cauchy-Schwarz inequality again, it is straightforward to verify that

$$B_j \leq \tilde{C}_\delta \sum_{j'=0}^{\infty} \left[\sum_{k,k'} |\beta_{j,k,j',k'}|^p \right]^{q/p} 2^{j'(1+2\delta)q/2}. \tag{4.5.13}$$

Hence,

$$\sum_{j'=0}^{\infty} 2^{js_1'q} \left(\sum_{k=0}^{2^j-1} |b_{j,k}^{(l)}|^p \right)^{q/p} \leq \tilde{C}_\delta 2^{js_1'q} \sum_{j'=0}^{\infty} 2^{j'(1+2\delta)q/2} \left[\sum_{k,k'} |\beta_{j,k,j',k'}|^p \right]^{q/p} \leq \tilde{C}_\delta A^q = \tilde{A}^q \tag{4.5.14}$$

provided $s_2^* \geq (1+2\delta)/2$. Since $s_2 > \max\{1/2, 1/p\}$ implies $s_2 > 1/2$, choose $\delta = (s_2 - 1/2)/2$. If $q/p < 1$, then similar considerations yield

$$B_j \le \tilde{C}_\delta \sum_{j'=0}^{\infty} \left[\sum_{k,k'} |\beta_{j,k,j',k'}|^p \right]^{q/p} 2^{j'(1+\delta)q/2}, \tag{4.5.15}$$

so that the previous calculation holds with δ instead of 2δ, and the proof is complete.

Using Lemma 19 and standard arguments (see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004)), we obtain for each f_l

$$\sup_{f_l \in B^{s_1}_{p,q}(\tilde{A})} \mathbf{E} \left|\left| \tilde{f}_l - f_l \right|\right|^2 \asymp \begin{cases} CN^{-\frac{2s_1}{2s_1+2\nu+1}}, & \text{if } \ s_1 \ge (\frac{1}{p} - \frac{1}{2})(2\nu+1), \\ CN^{-\frac{2s_1'}{2s_1'+2\nu}}, & \text{if } \ s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1). \end{cases} \tag{4.5.16}$$

Now, consider estimator $\tilde{f}$ of f with $\tilde{f}(u_l, t_i) = f_l(t_i)$. If $f_u = \partial f / \partial u$ and $f_{uu} = \partial^2 f / \partial u^2$ exist and uniformly bounded for $u \in [0,1]$, then rectangle method for numerical integration yields

$$\mathbf{E} \left|\left| \tilde{f} - f \right|\right|^2 = M^{-1} \sum_{l=1}^{M} \mathbf{E} \left|\left| \tilde{f}_l - f_l \right|\right|^2 + R_M, \tag{4.5.17}$$

where

$$R_M \le (12M^2)^{-1} \left[\mathbf{E} \left|\left| \tilde{f}_u - f_u \right|\right|^2 + \sqrt{\mathbf{E} \left|\left| \tilde{f} - f \right|\right|^2 \ \mathbf{E} \left|\left| \tilde{f}_{uu} - f_{uu} \right|\right|^2} \right]. \tag{4.5.18}$$

If M is large enough, then $R_M = o\left(\mathbf{E} \left|\left| \tilde{f} - f \right|\right|^2 \right)$ as $M \to \infty$ and we derive

$$\mathbf{E} \left|\left| \tilde{f} - f \right|\right|^2 \asymp \begin{cases} CN^{-\frac{2s_1}{2s_1+2\nu+1}}, & \text{if } \ s_1 \ge (\frac{1}{p} - \frac{1}{2})(2\nu+1), \\ CN^{-\frac{2s_1'}{2s_1'+2\nu}}, & \text{if } \ s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1). \end{cases} \tag{4.5.19}$$

Recall that according to Theorem 7 the convergence rates due to simultaneous recoveries are represented by

$$\mathbf{E} \left|\left| \hat{f} - f \right|\right|^2 \asymp \begin{cases} (MN)^{-\frac{2s_2}{2s_2+1}}, & \text{if } \ s_1 > s_2(2\nu+1), \\ (MN)^{-\frac{2s_1}{2s_1+2\nu+1}}, & \text{if } \ (\frac{1}{p} - \frac{1}{2})(2\nu+1) \le s_1 \le s_2(2\nu+1), \\ (MN)^{-\frac{2s_1'}{2s_1'+2\nu}}, & \text{if } \ s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1). \end{cases} \tag{4.5.20}$$

Notice that comparing the two approaches we have two cases; the case when $s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1)$ and when $s_1 > s_2(2\nu+1)$. In the former one, simultaneous recoveries outperform separate recoveries without any additional conditions since dividing (4.5.20) by (4.5.19) yields $M^{-\frac{2s_1'}{2s_1'+2\nu}}$. By straightforward calculations, one can check that in the latter case $s_1 > s_2(2\nu+1)$ convergence rates of separate deconvolution recoveries may be better than that of the simultaneous estimator. To see this, notice that when $s_1 > s_2(2\nu+1)$, $s_1 > (\frac{1}{p}-\frac{1}{2})(2\nu+1)$. Indeed, in order for separate recoveries to outperform simultaneous recoveries, (4.5.19) divided by (4.5.20) must be less than one, so that comparing the rates, yields

$$\begin{aligned}
\frac{N^{-\frac{2s_1}{2s_1+2\nu+1}}}{(MN)^{-\frac{2s_2}{2s_2+1}}} &< 1 \\
M^{\frac{2s_2}{2s_2+1}} N^{-\frac{2s_1}{2s_1+2\nu+1}} N^{\frac{2s_2}{2s_2+1}} &< 1 \\
M^{\frac{2s_2}{2s_2+1}} N^{-\frac{2s_1-2s_2(2\nu+1)}{(2\nu+1)(2s_1+2\nu+1)}} &< 1
\end{aligned} \tag{4.5.21}$$

Hence, we derive the conclusion that simultaneous recovery delivers better precision than separate ones unless

$$\lim_{\substack{M\to\infty \\ N\to\infty}} MN^{-\frac{s_1-s_2(2\nu+1)}{s_2(2s_1+2\nu+1)}} < 1, \quad s_1 > s_2(2\nu+1). \tag{4.5.22}$$

It is easy to see that relation (4.5.22) holds only if s_1 is large, s_2 is small and M is relatively small in comparison with N.

4.6 Extension to the (r + 1)-dimensional case

In this section, we extend the results obtained above to the $(r+1)$-dimensional version of the model (4.1.1). In this case, expanding both sides of equation (4.1.1) over Fourier basis, as before, we obtain for any $\mathbf{u} \in [0,1]^r$

$$y_m(\mathbf{u}) = g_m(\mathbf{u}) f_m(\mathbf{u}) + \varepsilon z_m(\mathbf{u}). \tag{4.6.1}$$

Construction of the estimator follows the path of the two-dimensional case. With $\psi_{j,k}(t)$ and $\eta_{j',k'}(u)$ defined earlier, we consider vectors $\mathbf{j'} = (j_1', \cdots, j_r')$, $\mathbf{k'} = (k_1', \cdots, k_r')$, $\mathbf{m'} = (m_1', \cdots, m_r')$ and $\mathbf{J'} = (J_1', \cdots, J_r')$, and subsets $\Upsilon(\mathbf{m'}, \mathbf{J'})$ and $\mathcal{K}(\mathbf{j'})$ of the set of r-dimensional vectors with nonnegative integer components:

$$\Upsilon(\mathbf{m'}, \mathbf{J'}) = \{\mathbf{j'} : m_l' \le j_l' \le J_l',\ l = 1, \cdots, r\}, \quad \mathcal{K}(\mathbf{j'}) = \{\mathbf{k'} : 0 \le k_l' \le j_l' - 1,\ l = 1, \cdots, r\}. \tag{4.6.2}$$

If $\boldsymbol{\infty}$ is the r-dimensional vector with all components being ∞, one can expand $f(\mathbf{u},t)$ into wavelet series as

$$f(\mathbf{u},t) = \sum_{j=m_0-1}^{\infty} \sum_{k=0}^{2^j-1} \sum_{\mathbf{j'}\in\Upsilon(\mathbf{m'},\boldsymbol{\infty})} \sum_{\mathbf{k'}\in\mathcal{K}(\mathbf{j'})} \beta_{jk,\mathbf{j'},\mathbf{k'}} \psi_{jk}(t) \prod_{l=1}^{r} \eta_{j'_l,k'_l}(u_l), \tag{4.6.3}$$

where coefficients $\beta_{jk,\mathbf{j'},\mathbf{k'}}$ are of the form

$$\beta_{jk,\mathbf{j'},\mathbf{k'}} = \sum_{m\in W_j} \overline{\psi_{j,k,m}} \int_{[0,1]^d} \frac{h_m(\mathbf{u})}{g_m(\mathbf{u})} \prod_{l=1}^{r} [\eta_{j'_l,k'_l}(u_l)]\, d\mathbf{u}, \tag{4.6.4}$$

the set W_j is defined by formula (4.2.5) and $h_m(\mathbf{u}) = \langle (f*g)(\cdot,\mathbf{u}), e_m(\cdot)\rangle$. Similarly to the two-dimensional case, we estimate $f(\mathbf{u},t)$ by

$$\widehat{f}(\mathbf{u},t) = \sum_{j=m_0-1}^{J-1} \sum_{k=0}^{2^j-1} \sum_{\mathbf{j'}\in\Upsilon(\mathbf{m'},\mathbf{J'})} \sum_{\mathbf{k'}\in\mathcal{K}(\mathbf{j'})} \widehat{\beta}_{j,k,\mathbf{j'},\mathbf{k'}}\, \psi_{jk}(t) \prod_{l=1}^{r} \eta_{j'_l,k'_l}(u_l) \tag{4.6.5}$$

with

$$\widehat{\beta}_{j,k,\mathbf{j'},\mathbf{k'}} = \widetilde{\beta}_{j,k,\mathbf{j'},\mathbf{k'}} \mathbf{1}\left(\left|\widetilde{\beta}_{j,k,\mathbf{j'},\mathbf{k'}}\right| > \lambda_{j,\varepsilon}\right). \tag{4.6.6}$$

Here

$$\widetilde{\beta}_{j,k,\mathbf{j'},\mathbf{k'}} = \sum_{m\in W_j} \overline{\psi_{j,k,m}} \int \frac{y_m(\mathbf{u})}{g_m(\mathbf{u})} \prod_{l=1}^{r} [\eta_{j'_l,k'_l}(u_l)] d\mathbf{u} \tag{4.6.7}$$

are the unbiased estimators of $\beta_{jk,\mathbf{j'},\mathbf{k'}}$, J is defined in (4.4.8), J'_l are such that $2^{J'_l} = \varepsilon^{-2}$, $l = 1,\cdots,r$, and $\lambda_{j,\varepsilon}$ is given by formula (4.4.7). Assume, as before, that functional Fourier coefficients $g_m(\mathbf{u})$ of function $g(\mathbf{u},t)$ are uniformly bounded from above and below

$$C_1 |m|^{-2\nu} \le |g_m(\mathbf{u})|^2 \le C_2 |m|^{-2\nu} \tag{4.6.8}$$

and that function $f(\mathbf{u},t)$ belongs to an $(r+1)$-dimensional Besov ball. As described in section 4.3.1 to define these Besov balls, we introduce the vector $\mathbf{s_2} = (s_{21},\cdots,s_{2r})$ and denote by $\mathbf{s'_2}$ and $\mathbf{s^*_2}$ vectors with components $s'_{2l} = s_{2l} + 1/2 - 1/p'$ and $s^*_{2l} = s_{2l} + 1/2 - 1/p$, $l = 1,\cdots,r$, respectively, where $p' = \min\{p,2\}$. If $s_0 \ge \max_l s_{2l}$, then the $(r+1)$-dimensional Besov ball of radius A is characterized by its wavelet coefficients

$\beta_{j,k,\mathbf{j}',\mathbf{k}'}$ as follows (see, e.g. Heping (2004))

$$B_{p,q}^{s_1,\mathbf{s_2}}(A) = \left\{ f \in L^2([0,1]^{r+1}) : \; \left(\sum_{j,\mathbf{j}'} 2^{[js_1^* + \mathbf{j}'^T \mathbf{s_2^*}]q} \left(\sum_{k,\mathbf{k}'} |\beta_{jk,\mathbf{j}',\mathbf{k}'}|^p \right)^{\frac{q}{p}} \right)^{1/q} \le A \right\}. \tag{4.6.9}$$

It is easy to show that, with the above assumptions, similarly to the two-dimensional case, as $\varepsilon \to 0$, one has

$$\mathrm{Var}\left(\widetilde{\beta}_{j,k,\mathbf{j}',\mathbf{k}'}\right) \asymp \varepsilon^2 2^{2j\nu}, \quad \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} |\beta_{j,k,\mathbf{j}',\mathbf{k}'}|^2 \le A^2 2^{-2(js_1' + \mathbf{j}'^T \mathbf{s_2^*})}, \tag{4.6.10}$$

$$\Pr\left(\left|\widetilde{\beta}_{j,k,\mathbf{j}',\mathbf{k}'} - \beta_{j,k,\mathbf{j}',\mathbf{k}'}\right| > \alpha \lambda_{j\varepsilon}\right) = O\left(\varepsilon^{\frac{\alpha^2 C_\beta^2}{2\sigma_0^2}} [\ln(1/\varepsilon)]^{-\frac{1}{2}}\right). \tag{4.6.11}$$

The upper and the lower bounds for the risk are expressed via

$$s_{2,0} = \min_{l=1,\cdots,r} s_{2,l} = s_{2,l_0}, \tag{4.6.12}$$

where $l_0 = \arg\min s_{2,l}$. In particular, the following statements hold.

Theorem 8. Let $\min\{s_1, s_{2,l_0}\} \ge \max\{1/p, 1/2\}$ with $1 \le p, q \le \infty$. Then, under assumption (4.6.8), as $\varepsilon \to 0$,

$$R_\varepsilon(B_{p,q}^{s_1,\mathbf{s_2}}(A)) \ge CA^2 \left(\frac{\varepsilon^2}{A^2}\right)^D \tag{4.6.13}$$

where

$$D = \min\left(\frac{2s_{2,0}}{2s_{2,0}+1}, \frac{2s_1}{2s_1+2\nu+1}, \frac{2s_1'}{2s_1'+2\nu}\right). \tag{4.6.14}$$

or,

$$D = \begin{cases} \frac{2s_{2,0}}{2s_{2,0}+1}, & \text{if } \; s_1 > s_{2,0}(2\nu+1), \\ \frac{2s_1}{2s_1+2\nu+1}, & \text{if } \; (\frac{1}{p} - \frac{1}{2})(2\nu+1) \le s_1 \le s_{2,0}(2\nu+1), \\ \frac{2s_1'}{2s_1'+2\nu}, & \text{if } \; s_1 < (\frac{1}{p} - \frac{1}{2})(2\nu+1). \end{cases} \tag{4.6.15}$$

Proof of Theorem 8. Repeating the proof of Theorem 5 with j' and k' replaced by $\mathbf{j}'$ and $\mathbf{k}'$, respectively, and $s_2 j'$ replaced by $\mathbf{j}'^T \mathbf{s_2'}$, we again arrive at two cases. Denote the r-dimensional vector with all unit components by $\mathbf{e}$.

In the dense-dense case, we use $(r+1)$-dimensional array w, so that $N = 2^{j+\mathbf{e}^T\mathbf{j}'}$. Choose $\gamma_{j,\mathbf{j}'}^2 =$

$A^2 2^{-j(2s_1+1)-\mathbf{j}'^T(\mathbf{2s_2}+\mathbf{e})}$ and observe that $K(f,\tilde{f}) \le C\varepsilon^{-2}\gamma^2_{jj'}2^{j+\mathbf{e}^T\mathbf{j}'}2^{-2\nu j}$. Now, applying Lemma 14 with

$$\delta^2 = \gamma^2_{jj'}2^{j+\mathbf{e}^T\mathbf{j}'}/32 = A^2 2^{-2s_1 j - 2\mathbf{j}'^T\mathbf{s_2}}/32 \tag{4.6.16}$$

one arrives at the following optimization problem

$$2js_1 + 2j's_2 \Rightarrow \min \; j(2s_1+2\nu+1) + \sum_{l=1}^{r}(2s_{2,l}+1)j'_l \ge \tau_\varepsilon, \quad j, j'_l \ge 0, \tag{4.6.17}$$

where τ_ε is defined in formula (4.3.15). Setting $j = \tau_\varepsilon/(2s_1+2\nu+1) - \sum_{l=1}^{r}(2s_l+1)/(2s_1+2\nu+1)$, arrive at optimization problem

$$\frac{2s_1\tau_\varepsilon}{2s_1+2\nu+1} + \sum_{l=1}^{r}\frac{2j'_l[s_{2,l}(2\nu+1)-s_1]}{2s_1+2\nu+1} \Rightarrow \min, \quad j'_l \ge 0, \; l = 1,\cdots,r. \tag{4.6.18}$$

If $s_{2,l_0}(2\nu+1) \ge s_1$, then each j'_l is multiplied by a nonnegative number and minimum is attained when $j'_l = 0$, $l = 1,\cdots,r$. Then, $j = \tau_\varepsilon/(2s_1+2\nu+1)$. On the other hand, if $s_{2,l_0}(2\nu+1) < s_1$, then j_{l_0} is multiplied by the smallest factor which is negative. Therefore, minimum in (4.6.18) is attained if $j = 0$, $j'_l = 0$, $l \ne l_0$ and $j_{l_0} = \tau_\varepsilon/(2s_{2,l_0}+1)$. Plugging those values into (4.6.16), obtain

$$\delta^2 = \begin{cases} CA^2\,(\varepsilon^2/A^2)^{\frac{2s_{2,0}}{2s_{2,0}+1}}, & \text{if } \; s_1 > s_{2,0}(2\nu+1), \\ CA^2\,(\varepsilon^2/A^2)^{\frac{2s_1}{2s_1+2\nu+1}}, & \text{if } \; s_1 \le s_{2,0}(2\nu+1). \end{cases} \tag{4.6.19}$$

In the sparse-dense case, we use r-dimensional array w, so that $N = 2^{\mathbf{e}^T\mathbf{j}'}$. Choose $\gamma^2_{j,\mathbf{j}'}$ as,
$\gamma^2_{j,\mathbf{j}'} = A^2 2^{-2js_1^* - \mathbf{j}'^T(\mathbf{2s_2}+\mathbf{e})}$ and observe that $K(f,\tilde{f}) \le C\varepsilon^{-2}\gamma^2_{jj'}2^{j+\mathbf{e}^T\mathbf{j}'}2^{-2\nu j}$.

Now, applying Lemma 14 with

$$\delta^2 = A^2 2^{-2s_1^* j - 2\mathbf{j}'^T\mathbf{s_2}}/32 \tag{4.6.20}$$

one arrives at the following optimization problem

$$2js_1 + 2j's_2 \Rightarrow \min \; j(2s_1^*+2\nu+1) + \sum_{l=1}^{r}(2s_{2,l}+1)j'_l \ge \tau_\varepsilon, \quad j, j'_l \ge 0, \tag{4.6.21}$$

Again, setting $j = \tau_\varepsilon/(2s_1^*+2\nu) - \sum_{l=1}^{r}(2s_l+1)/(2s_1^*+2\nu)$, arrive at optimization problem

$$\frac{2s_1^*\tau_\varepsilon}{2s_1^*+2\nu} + \sum_{l=1}^{r}\frac{2j'_l[2s_{2,l}\nu - s_1^*]}{2s_1^*+2\nu} \Rightarrow \min, \quad j'_l \ge 0, \; l = 1,\cdots,r. \tag{4.6.22}$$

Repeating the reasoning applied in the dense-dense case, we obtain $j = 0$, $j'_l = 0$, $l \neq l_0$ and $j_{l_0} = \tau_\varepsilon/(2s_{2,l_0}+1)$ if $2s_{2,l_0}\nu < s_1^*$, and $j = \tau_\varepsilon/(2s_1+2\nu+1)$, $j'_l = 0$, $l = 1, \cdots, r$, if $2s_{2,l_0}\nu > s_1^*$. Plugging those values into (4.6.20), obtain

$$\delta^2 = \begin{cases} CA^2\,(\varepsilon^2/A^2)^{\frac{2s_{2,0}}{2s_{2,0}+1}}, & \text{if } \ 2\nu s_{2,0} \le s_1^*, \\ CA^2\,(\varepsilon^2/A^2)^{\frac{2s_1^*}{2s_1^*+2\nu}}, & \text{if } \ 2\nu s_{2,0} > s_1^*. \end{cases} \tag{4.6.23}$$

In order to complete the proof, combine (4.6.19) and (4.6.23) and note that $s_1^* = s'_1$ if $p \le 2$.

Theorem 9. Let $\widehat{f}(.,.)$ be the wavelet estimator defined in (4.6.5), with J defined in (4.4.8), J'_l such that $2^{J'_l} = (\varepsilon^2)^{-1}$, $l = 1, \cdots, r$, and $\lambda_{j,\varepsilon}$ given by formula (4.4.7). Let condition (4.3.6) hold and $\min\{s_1, s_{2,0}\} \ge \max\{1/p, 1/2\}$, with $1 \le p, q \le \infty$. If C_β in (4.4.7) satisfies condition (4.4.45), then, as $\varepsilon \to 0$,

$$\sup_{f \in B_{p,q}^{s_1,\mathbf{s_2}}(A))} \mathbf{E}\left|\left|\widehat{f} - f\right|\right|^2 \le CA^2\ \left(A^{-2}\,\varepsilon^2 \ln(1/\varepsilon)\right)^D\ \ln(1/\varepsilon)^{D_1} \tag{4.6.24}$$

where D is defined in (4.6.14) and

$$D_1 = \mathbf{1}(s_1 = s_{2,0}(2\nu+1)) + \mathbf{1}(s_1 = (2\nu+1)(1/p - 1/2)) + \sum_{l \neq l_0} \mathbf{1}(s_{2,l} = s_{2,0}). \tag{4.6.25}$$

Proof of Theorem 9. Repeat the proof of Theorem 6 with j' and k' replaced by $\mathbf{j'}$ and $\mathbf{k'}$, respectively, $s_2 j'$ replaced by $\mathbf{j'}^T\mathbf{s'_2}$ and

$$2^{j_0} = (\chi_{\varepsilon,A})^{-\frac{d}{2s'_1}}, \quad 2^{j'_{0,l}} = (\chi_{\varepsilon,A})^{-\frac{d}{2s'_{2,l}}}, \quad l = 1, \cdots, r. \tag{4.6.26}$$

Then, formulae (4.4.19)–(4.4.32) are valid. One can also partition Δ in (4.4.33) into Δ_1, Δ_2 and Δ_3 given by expressions similar to (4.4.34), (4.4.35) and (4.4.36) with $r+1$ sums in (4.4.34) instead of two, $\sum_{j'=m'_0}^{j'_0}$ replaced by r respective sums and $\mathbf{1}\left(2^{j(2\nu+1)+j'} > \chi_{\varepsilon,A}^{d-1}\right)$ replaced by $\mathbf{1}\left(2^{j(2\nu+1)+\mathbf{e}^T\mathbf{j'}} > \chi_{\varepsilon,A}^{d-1}\right)$. Then, upper bounds (4.4.37) and (4.4.38) hold. In order to construct upper bounds for Δ_3, we again need to consider three different cases.

In Case 1, $s_1 \ge s_{2,0}(2\nu+1)$, replace $\sum_{j'=m'_0}^{j'_0}$ by $\sum_{j'_{l_0}=m'_{l_0}}^{j'_{0,l_0}}$ and $\sum_{j=m_0}^{j_0}$ by the sum over $j, j'_1, \cdots, j'_{l_0-1}, j'_{l_0+1}, \cdots, j'_r$. Repeating calculations for this case, keeping in mind that $s'_{2,l} \ge s'_{2,0}$ for any l and noting that, whenever $s'_{2,l} = s'_{2,0}$, we gain an extra logarithmic factor, we arrive at

$$\Delta_3 = O\left(A^2\, \chi^d_{\varepsilon,A}\, [\ln(1/\varepsilon)]^{\mathbf{1}(s_1 = s_2(2\nu+1)) + \sum_{l \neq l_0} \mathbf{1}(s_{2,l} = s_{2,0})}\right). \tag{4.6.27}$$

In Case 2, $(1/p - 1/2)(2\nu+1) < s_1 < s_{2,0}(2\nu+1)$, replace $\sum_{j'=m'_0}^{j'_0}$ by $\sum_{\mathbf{j}' \in \Upsilon(\mathbf{m}', \mathbf{j}'_0)}$ where $\mathbf{j}'_0 = (j'_{0,1}, \cdots, j'_{0,r})$ and arrive at (4.4.42). In Case 3, $s_1 \le (\frac{1}{p} - \frac{1}{2})(2\nu+1)$, since the sum over $\mathbf{j}'$ is uniformly bounded, calculations for the two-dimensional case hold and (4.4.44) is valid. Combination of (4.6.27), (4.4.42) and (4.4.44) completes the proof.

Remark 5. Observe that convergence rates in Theorems 8 and 9 depend on s_1, p, ν and $\min_l s_{2l}$ but not on the dimension r.

It could be also natural to ask what would the corresponding results be if s_1 itself was multidimensional, that is, if one considers the case of convolution in more than one direction where

$$h(\mathbf{u}, \mathbf{t}) = \int_{[0;1]^d} g(\mathbf{u}, \mathbf{t} - \mathbf{x}) f(\mathbf{u}, x) d\mathbf{x}, \quad \mathbf{t} \in [0;1]^d;\ \mathbf{u} \in [0;1]^r. \tag{4.6.28}$$

Although this is beyond the scope of this discussion, let us just mention that, as soon as one establishes upper bounds for the variances of the wavelet coefficients like (4.6.10) as well as concentration inequalities for the wavelet coefficients estimators like in (4.6.11), one expects to obtain convergence rates similar to Theorems 8 and 9 with s_1 replaced with $\min_k s_{1k}$.

CHAPTER 5: FUNCTIONAL DECONVOLUTION MODEL WITH LONG-RANGE DEPENDENT ERRORS

5.1 Formulation of the Problem

We consider the estimation problem of the unknown response function $f(\cdot) \in L^2(T)$ from observations $y(u_l, t_i)$ driven by

$$y(u_l, t_i) = \int_T g(u_l, t_i - x) f(x) dx + \xi_{li}, \quad l = 1, 2, \ldots, M, \; i = 1, 2, \ldots, N, \tag{5.1.1}$$

where $u_l \in U = [a, b]$, $0 < a \leq b < \infty$, $T = [0, 1]$, $t_i = i/N$, and the errors ξ_{li} are Gaussian random variables, independent for different l's, but dependent for different i's.

Denote the total number of observations $n = NM$ and assume, without loss of generality, that $N = 2^J$ for some integer $J > 0$. For each $l = 1, 2, \ldots, M$, let $\boldsymbol{\xi}^{(l)}$ be a Gaussian vector with components ξ_{li}, $i = 1, 2, \ldots, N$, and let $\boldsymbol{\Sigma}^{(l)} := \text{Cov}(\boldsymbol{\xi}^{(l)}) := \mathbf{E}[\boldsymbol{\xi}^{(l)}(\boldsymbol{\xi}^{(l)})^T]$ be its covariance matrix.

Assumption A1: For each $l = 1, 2, \ldots, M$, $\boldsymbol{\Sigma}^{(l)}$ satisfies the following condition: there exist constants K_1 and K_2 ($0 < K_1 \leq K_2 < \infty$), independent of l and N, such that, for each $l = 1, 2, \ldots, M$,

$$K_1 N^{2d_l} \leq \lambda_{\min}(\boldsymbol{\Sigma}^{(l)}) \leq \lambda_{\max}(\boldsymbol{\Sigma}^{(l)}) \leq K_2 N^{2d_l}, \quad 0 \leq d_l < 1/2, \tag{5.1.2}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and the largest eigenvalues of (the Toeplitz matrix) $\boldsymbol{\Sigma}^{(l)}$. (Here, and in what follows, "^{T}" denotes the transpose of a vector or a matrix.)

Assumption A1 is valid when, for each $l = 1, 2, \ldots, M$, $\boldsymbol{\xi}^{(l)}$ is a second-order stationary Gaussian sequence with spectral density satisfying certain assumptions. We shall elaborate on this issue in Section 5.2. Note that, in the case of independent errors, for each $l = 1, 2, \ldots, M$, $\boldsymbol{\Sigma}^{(l)}$ is proportional to the identity matrix and that $d_l = 0$. In this case, the multichannel deconvolution model (5.1.1) reduces to the one with independent and identically distributed Gaussian errors. In a view of (5.1.1), the limit situation $d_l = 0$, $l = 1, 2, \ldots, M$, can be thought of as the standard multichannel deconvolution model described in Pensky and Sapatinas (2009, 2010).

In the multichannel deconvolution model studied by Pensky and Sapatinas (2009, 2010), as well as in the very current extension of their results to derivative estimation by Navarro *et al.* (2013), it is assumed that errors are independent and identically distributed Gaussian random variables. However, empirical evidence has shown that even at large lags, the correlation structure in the errors can decay at a hyperbolic rate, rather than an exponential rate. To account for this, a great deal of papers on long-range dependence (LRD) have been developed. The study of LRD (also called long memory) has a number of applications, as it can be reflected by the very large number of articles having LRD or long memory in their titles, in areas such as climate study, DNA sequencing, econometrics, finance, hydrology, internet modeling, signal and image processing, physics and even linguistics. Other applications can be found in Beran (1992, 1994), Beran *et al.* (2013) and Doukhan *et al.* (2003).

Although quite a few LRD models have been considered in the regression estimation framework, very little has been done in the standard deconvolution model. The density deconvolution set up has also witnessed some shift towards analyzing the problem for dependent processes. The argument behind that was that a number of statistical models, such as non-linear GARCH and continuous-time stochastic volatility models, can be looked at as density deconvolution models if we apply a simple logarithmic transformation, and thus there is need to account for dependence in the data. This started by Van Zanten *et al.* (2008) who investigated wavelet based density deconvolution studied by Pensky and Vidakovic (1999) with a relaxation to weakly dependent processes. Comte *et al.* (2008) analyzed another adaptive estimator that was proposed earlier but under the assumption that the sequence is strictly stationary but not necessarily independent. However, it was Kulik (2008), who considered the density deconvolution for LRD and short-range dependent (SRD) processes. However, Kulik (2008) did not considered nonlinear wavelet estimators but dealt instead with linear kernel estimators.

In nonparametric regression estimation, ARIMA-type models for the errors were analyzed in Cheng and Robinson (1994), with error terms of the form $\sigma(x_i, \xi_i)$. In Csörgo and Mielniczuk (2000), the error terms were modeled as infinite order moving averages processes. Mielniczuk and Wu (2004) investigated another form of LRD, with the assumption that x_i and ξ_i are not necessarily independent for the same i. ARIMA-type error models were also considered in Kulik and Raimondo (2009). In the standard deconvolution model, and using a maxiset approach, Wishart (2012) applied a fractional Brownian motion to model the presence of LRD, while Wang (2012) used a minimax approach to study the problem of recovering a function f from a more general noisy linear transformation where the noise is also a fractional Brownian motion.

The objective of the discussion in this chapter is to study the multichannel deconvolution model from a minimax point of view, with the relaxation that errors exhibit LRD. We do not limit our consideration to a specific type of LRD: the only restriction is that the errors should satisfy Assumption A1. In particular, we derive minimax lower bounds for the L^2-risk in model (5.1.1) under Assumption A1 when $f(\cdot)$ is assumed to belong to a Besov ball and $g(\cdot,\cdot)$ has smoothness properties similar to those in Pensky and Sapatinas (2009, 2010), including both regular-smooth and super-smooth convolutions. In addition, we propose an adaptive wavelet estimator for $f(\cdot)$ and show that such estimator is asymptotically optimal (or near-optimal within a logarithmic factor) in the minimax sense, in a wide range of Besov balls. We prove that the convergence rates of the resulting estimators depend on the balance between the smoothness parameter (of the response function $f(\cdot)$), the kernel parameters (of the blurring function $g(\cdot,\cdot)$), and the long memory parameters d_l, $l = 1, 2 \dots, M$ (of the error sequence $\boldsymbol{\xi}^{(l)}$). Since the parameters d_l depend on the values of l, the convergence rates have more complex expressions than the ones obtained in Kulik and Raimondo (2009) when studying nonparametric regression estimation with ARIMA-type error models. The convergence rates we derive are more similar in nature to those in Pensky and Sapatinas (2009, 2010). In particular, the convergence rates depend on how the total number $n = NM$ of observations is distributed among the total number M of channels. As we illustrate in two examples, convergence rates are not affected by long range dependence in case of super-smooth convolutions, however, the situation changes in regular cases.

5.2 Stationary Sequences with Long-Range Dependence

In this section, for simplicity of exposition, we consider one sequence of errors $\{\xi_j : j = 1, 2, \dots\}$. Assume that $\{\xi_j : j = 1, 2, \dots\}$ is a second-order stationary sequence with covariance function $\gamma_\xi(k) := \gamma(k)$, $k = 0, \pm 1, \pm 2, \dots$. The spectral density is defined as

$$a_\xi(\lambda) := a(\lambda) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\lambda), \qquad \lambda \in [-\pi, \pi]. \tag{5.2.1}$$

On the other hand, the inverse transform which recovers $\gamma(k)$, $k = 0, \pm 1, \pm 2, \ldots$, from $a(\lambda)$, $\lambda \in [-\pi, \pi]$, is given by

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} a(\lambda) d\lambda, \quad k = 0, \pm 1, \pm 2, \ldots, \tag{5.2.2}$$

under the assumption that the spectral density $a(\lambda)$, $\lambda \in [-\pi, \pi]$, is squared-integrable.

Let $\mathbf{\Sigma} = [\gamma(j-k)]_{j,k=1}^{N}$ be the covariance matrix of $(\xi_1, \ldots, \xi_N)$. Define $\mathcal{X} = \{\mathbf{x} \in \mathbb{C}^N : \mathbf{x}^* \mathbf{x} = 1\}$, where $\mathbf{x}^*$ is the complex-conjugate of $\mathbf{x}$. Since $\mathbf{\Sigma}$ is Hermitian, one has

$$\lambda_{\min}(\mathbf{\Sigma}) = \inf_{\mathbf{x} \in \mathcal{X}} (\mathbf{x}^* \mathbf{\Sigma} \mathbf{x}) \qquad \text{and} \qquad \lambda_{\max}(\mathbf{\Sigma}) = \sup_{\mathbf{x} \in \mathcal{X}} (\mathbf{x}^* \mathbf{\Sigma} \mathbf{x}). \tag{5.2.3}$$

With the definitions introduced above,

$$\mathbf{x}^* \mathbf{\Sigma} \mathbf{x} = \sum_{j,k=1}^{N} \mathbf{x}^* \gamma(j-k) \mathbf{x} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^{N} x_j e^{-ij\lambda} \right|^2 a(\lambda) d\lambda. \tag{5.2.4}$$

Note that, by the Parseval identity, the function $h(\lambda) = \left| \sum_{j=1}^{N} x_j e^{-ij\lambda} \right|^2$, $\lambda \in [-\pi, \pi]$, belongs to the set

$$\mathcal{H}_N = \left\{ h : \, h \;\text{ symmetric, } |h|_\infty \le N, \int_{-\pi}^{\pi} h(\lambda) d\lambda = 2\pi \right\}. \tag{5.2.5}$$

Let $d \in [0, 1/2)$. Consider the following class of spectral densities

$$\mathcal{F}_d = \left\{ a : \; a(\lambda) = |\lambda|^{-2d} a_*(\lambda), 0 < C_{\min} \le |a_*(\lambda)| \le C_{\max} < \infty, \, \lambda \in [-\pi, \pi] \right\}. \tag{5.2.6}$$

Below we provide two examples of second-order stationary sequences such that their spectral densities $a(\lambda)$, $\lambda \in [-\pi, \pi]$, belong to the class $\mathcal{F}_d$ described in (5.2.6).

Fractional ARIMA(0, d, 0). Let $\{\xi_j : j = 1, 2, \ldots\}$ be the second-order stationary sequence

$$\xi_j = \sum_{m=0}^{\infty} a_m \eta_{j-m}, \tag{5.2.7}$$

where η_j are uncorrelated, zero-mean, random variables, $\sigma_\eta^2 := \mathrm{Var}(\eta_j) < \infty$, and

$$a_m = (-1)^m \binom{-d}{m} = (-1)^m \frac{\Gamma(1-d)}{\Gamma(m+1)\Gamma(1-d-m)} \tag{5.2.8}$$

with $d \in [0, 1/2)$. Then, a_m, $m = 0, 1, \ldots$, are the coefficients in the power-series representation

$$A(z) := (1-z)^{-d} := \sum_{m=0}^{\infty} a_m z^m. \tag{5.2.9}$$

Therefore, the spectral density $a(\lambda)$, $\lambda \in [-\pi, \pi]$, of $\{\xi_j : j = 1, 2, \ldots\}$, is given by

$$a(\lambda) = \frac{\sigma_\eta^2}{2\pi} \left|A(e^{-i\lambda})\right|^2 = \frac{\sigma_\eta^2}{2\pi} \left|1 - e^{-i\lambda}\right|^{-2d} = \frac{\sigma_\eta^2}{2\pi} \left|2(1-\cos\lambda)\right|^{-d} \sim \frac{\sigma_\eta^2}{2\pi} |\lambda|^{-2d} \;\; (\lambda \to 0). \tag{5.2.10}$$

Hence, the sequence $\{\xi_j : j = 1, 2, \ldots\}$ has spectral density $a(\lambda)$, $\lambda \in [-\pi, \pi]$, that belongs to the class $\mathcal{F}_d$ described in (5.2.6). (The sequence $\{\xi_j : j = 1, 2, \ldots\}$ is called the fractional ARIMA(0,d,0) time series.)

Fractional Gaussian Noise. Assume that $B_H(u)$, $u \in [0, \infty]$, is a fractional Brownian motion with the Hurst parameter $H \in [1/2, 1)$. Define the second-order stationary sequence $\xi_j = B_H(j) - B_H(j-1)$, $j = 1, 2, \ldots$. Its spectral density $a(\lambda)$, $\lambda \in [-\pi, \pi]$, is given by (see, e.g., [26], p. 222)

$$a(\lambda) = \sigma^2 (2\pi)^{-2H-2} \Gamma(2H+1) \sin(\pi H) 4 \sin^2(\lambda/2) \times \sum_{k=-\infty}^{\infty} |k + (\lambda/2\pi)|^{-2H-1}, \tag{5.2.11}$$

and, hence,

$$a(\lambda) = \frac{2\sigma^2}{\pi} \Gamma(2H+1) \sin(\pi H) \lambda^{1-2H} \qquad (\lambda \to 0). \tag{5.2.12}$$

Hence, the sequence $\{\xi_j : j = 1, 2, \ldots\}$ has spectral density $a(\lambda)$, $\lambda \in [-\pi, \pi]$, that belongs to class $\mathcal{F}_d$ with $d = H - 1/2$. (The sequence $\{\xi_j : j = 1, 2, \ldots\}$ is called the fractional Gaussian noise.)

It follows from (5.2.6) that, for $a \in \mathcal{F}_d$, one has $a(\lambda) \sim |\lambda|^{-2d}$ $(\lambda \to 0)$. It also turns out that the condition $a \in \mathcal{F}_d$, $d \in [0, 1/2)$, implies that all eigenvalues of the covariance matrix $\mathbf{\Sigma}$ are of asymptotic order N^{2d} $(N \to \infty)$. In particular, the following lemma is true.

Lemma 20. Assume that $\{\xi_j : j = 1, 2, \ldots\}$ is a second-order stationary sequence with spectral density $a \in \mathcal{F}_d$, $d \in [0, 1/2)$. Then, for some constants K_{1d} and K_{2d} $(0 < K_{1d} \leq K_{2d} < \infty)$, that depend on d only,

$$K_{1d} N^{2d} \leq \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) \leq K_{2d} N^{2d}. \tag{5.2.13}$$

Proof of Lemma 20. We prove the upper bound only since the proof of the lower bound is similar. By (5.2.3)-(5.2.4), and the definitions of $\mathcal{H}_N$ and $\mathcal{F}_d$,

$$\lambda_{\max}(\mathbf{\Sigma}) \leq C_{\max} \sup_{h \in \mathcal{H}_N} \int_{-\pi}^{\pi} h(\lambda)|\lambda|^{-2d} d\lambda = 2C_{\max} \sup_{h \in \mathcal{H}_N} \int_{0}^{\pi} h(\lambda)|\lambda|^{-2d} d\lambda. \tag{5.2.14}$$

Now, we split $\int_0^{\pi} = \int_0^{\pi/N} + \int_{\pi/N}^{\pi}$. Since $d < 1/2$, for the first integral, we have

$$\int_0^{\pi/N} h(\lambda)|\lambda|^{-2d} d\lambda \leq N \int_0^{\pi/N} \lambda^{-2d} d\lambda = N \frac{1}{1-2d} \left(\frac{\pi}{N}\right)^{-2d+1} = \frac{1}{1-2d} N^{2d}. \tag{5.2.15}$$

For the second integral, since $d \geq 0$, we have

$$\int_{\pi/N}^{\pi} h(\lambda)|\lambda|^{-2d} d\lambda \leq \left(\frac{\pi}{N}\right)^{-2d} \int_{\pi/N}^{\pi} h(\lambda) d\lambda \leq \left(\frac{\pi}{N}\right)^{-2d} \int_0^{\pi} h(\lambda) d\lambda \leq \pi (2\pi)^{-2d} N^{2d}. \tag{5.2.16}$$

This completes the proof of the lemma. □

Remark 6. If $d = 0$, then $\mathcal{F}_d$ is the class of spectral densities $a(\lambda)$ that are bounded away from 0 and ∞ for all $\lambda \in [-\pi, \pi]$. In particular, the corresponding second-order stationary sequences $\{\xi_j : j = 1, 2, \ldots\}$ are weakly dependent. Then, the statement of Lemma 20 reduces to a result in Grenander and Szegö [32], Section 5.2.

It follows immediately from Lemma 20 that if, for each $l = 1, 2, \ldots, M$, $\boldsymbol{\xi}^{(l)}$ is a second-order stationary Gaussian sequence with spectral density $a_l \in \mathcal{F}_{d_l}$, $d_l \in [0, 1/2)$, that $\boldsymbol{\xi}^{(l)}$ are independent for different l's, and that d_l's are uniformly bounded, then Assumption A1 holds.

Corollary 3. For each $l = 1, 2, \ldots, M$, let $\boldsymbol{\xi}^{(l)}$ be a second-order stationary Gaussian sequence with spectral density $a_l \in \mathcal{F}_{d_l}$, $d_l \in [0, 1/2)$. We assume that $\boldsymbol{\xi}^{(l)}$ are independent for different l's. Let d_l, $l = 1, 2, \ldots, M$, be uniformly bounded, i.e., there exists d^* $(0 \leq d^* < 1/2)$ such that, for each $l = 1, 2, \ldots, M$,

$$0 \leq d_l \leq d^* < 1/2. \tag{5.2.17}$$

Then, Assumption A1 holds.

5.3 The Estimation Algorithm

In what follows, $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^N$. We also denote the complex-conjugate of $a \in \mathbb{C}$ by $\bar{a}$, the discrete Fourier basis on the interval T by $e_m(t_i) = e^{-i2\pi m t_i}$, $t_i = i/N$, $i = 1, 2, \ldots, N$, $m = 0, \pm 1, \pm 2, \ldots$, and the complex-conjugate of the matrix $\mathbf{A}$ by $\mathbf{A}^*$.

Recall the multichannel deconvolution model (5.1.1). Denote

$$h(u_l, t_i) = \int_T g(u_l, t_i - x) f(x) dx, \quad l = 1, 2, \ldots, M, \quad i = 1, 2, \ldots, N. \tag{5.3.1}$$

Then, equation (5.1.1) can be rewritten as

$$y(u_l, t_i) = h(u_l, t_i) + \xi_{li}, \quad l = 1, 2, \ldots, M, \quad i = 1, 2, \ldots, N. \tag{5.3.2}$$

For each $l = 1, 2, \ldots, M$, let $h_m(u_l) = \langle e_m, h(u_l, \cdot) \rangle$, $y_m(u_l) = \langle e_m, y(u_l, \cdot) \rangle$, $z_{lm} = \langle e_m, \boldsymbol{\xi}^{(l)} \rangle$, $g_m(u_l) = \langle e_m, g(u_l, \cdot) \rangle$ and $f_m = \langle e_m, f \rangle$ be the discrete Fourier coefficients of the $\mathbb{R}^N$ vectors $h(u_l, t_i)$, $y(u_l, t_i)$, ξ_{li}, $g(u_l, t_i)$ and $f(t_i)$, $i = 1, 2, \ldots, N$, respectively. Then, applying the discrete Fourier transform to (5.3.2), one obtains, for any $u_l \in U$, $l = 1, 2, \ldots, M$,

$$y_m(u_l) = g_m(u_l) f_m + N^{-1/2} z_{lm} \tag{5.3.3}$$

and

$$h_m(u_l) = g_m(u_l) f_m. \tag{5.3.4}$$

Multiplying both sides of (5.3.3) by $N^{-2d_l}\overline{g_m(u_l)}$, and adding them together, we obtain the following estimator of f_m

$$\widehat{f}_m = \left(\sum_{l=1}^{M} N^{-2d_l}\, \overline{g_m(u_l)} y_m(u_l) \right) / \left(\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right). \tag{5.3.5}$$

Let $\varphi^*(\cdot)$ and $\psi^*(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively, defined on the real line (see, e.g., Meyer (1992)), and obtain a periodized version of Meyer wavelet basis as in Johnstone *et al.* (2004), i.e., for $j \geq 0$ and $k = 0, 1, \ldots, 2^j - 1$,

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

Following Pensky and Sapatinas (2009, 2010), using the periodized Meyer wavelet basis described above, for some $j_0 \geq 0$, expand $f(\cdot) \in L^2(T)$ as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0 k}\varphi_{j_0 k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(t), \quad t \in T. \tag{5.3.6}$$

Furthermore, by Plancherel's formula, the scaling coefficients, $a_{j_0 k} = \langle f, \varphi_{j_0 k}\rangle$, and the wavelet coefficients, $b_{jk} = \langle f, \psi_{jk}\rangle$, of $f(\cdot)$ can be represented as

$$a_{j_0 k} = \sum_{m \in C_{j_0}} f_m \overline{\varphi_{m j_0 k}}, \quad b_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}}, \tag{5.3.7}$$

where $C_{j_0} = \{m : \varphi_{m j_0 k} \neq 0\}$ and, for any $j \geq j_0$,

$$C_j = \{m : \psi_{mjk} \neq 0\} \subseteq 2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]. \tag{5.3.8}$$

(Note that the cardinality $|C_j|$ of the set C_j is $|C_j| = 4\pi 2^j$, see, e.g., Johnstone *et al.* (2004).) Estimates of $a_{j_0 k}$ and b_{jk} are readily obtained by substituting f_m in (5.3.7) with (5.3.5), i.e.,

$$\widehat{a}_{j_0 k} = \sum_{m \in C_{j_0}} \widehat{f}_m \overline{\varphi_{m j_0 k}}, \quad \widehat{b}_{jk} = \sum_{m \in C_j} \widehat{f}_m \overline{\psi_{mjk}}. \tag{5.3.9}$$

We now construct a (block thresholding) wavelet estimator of $f(\cdot)$, suggested by Pensky & Sapatinas (2009, 2010). For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length $\ln n$. Let A_j and U_{jr} be the following sets of indices

$$A_j = \left\{r \mid r = 1, 2, \ldots, 2^j/\ln n\right\}, \tag{5.3.10}$$

$$U_{jr} = \left\{k \mid k = 0, 1, \ldots, 2^j - 1;\ (r-1)\ln n \leq k \leq r\ln n - 1\right\}. \tag{5.3.11}$$

Denote

$$B_{jr} = \sum_{k \in U_{jr}} b_{jk}^2, \quad \widehat{B}_{jr} = \sum_{k \in U_{jr}} \widehat{b}_{jk}^2. \tag{5.3.12}$$

Finally, for any $j_0 \geq 0$, the (block thresholding) wavelet estimator $\hat{f}_n(\cdot)$ of $f(\cdot)$ is constructed as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0 k} \varphi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \widehat{b}_{jk} \mathbf{1}(|\widehat{B}_{jr}| \geq \lambda_j)\, \psi_{jk}(t), \quad t \in T, \tag{5.3.13}$$

where $\mathbf{1}(A)$ is the indicator function of the set A, and the resolution levels j_0 and J and the thresholds λ_j will be defined in Section 5.5.

In what follows, the symbol C is used for a generic positive constant, independent of n, while the symbol K is used for a generic positive constant, independent of m, n, M and $u_1, u_2, \ldots, u_M$. Either of C or K may take different values at different places.

5.4 Minimax Lower Bounds for the L^2-Risk

Recall that

$$s^* = s + 1/2 - 1/p, \quad s' = s + 1/2 - 1/p', \quad p' = \min\{p, 2\}. \tag{5.4.1}$$

Assume that the unknown response function $f(\cdot)$ belongs to a Besov ball $B^s_{p,q}(A)$ of radius $A > 0$, so that the wavelet coefficients $a_{j_0 k}$ and b_{jk} defined in (5.3.7) satisfy the following relation

$$B^s_{p,q}(A) = \left\{ f \in L^2(U) : \left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0 k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{\infty} 2^{j s^* q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{\frac{q}{p}} \right)^{1/q} \leq A \right\}. \tag{5.4.2}$$

Below, we construct minimax lower bounds for the (quadratic) L^2-risk. For this purpose, we define the minimax L^2-risk over the set $V \subseteq L^2(T)$ as in (4.3.5).

For $M = M_n$ and $N = n/M_n$, denote

$$\tau_\kappa(m, n) = M^{-1} \sum_{l=1}^{M} N^{-2\kappa d_l}\, |g_m(u_l)|^{2\kappa}, \quad \kappa = 1 \text{ or } 2 \text{ or } 4, \tag{5.4.3}$$

and

$$\Delta_\kappa(j, n) = |C_j|^{-1} \sum_{m \in C_j} \tau_\kappa(m, n)\ [\tau_1(m, n)]^{-2\kappa}, \quad \kappa = 1 \text{ or } 2. \tag{5.4.4}$$

The expression $\tau_1(m, n)$ appears in both the lower and the upper bounds for the L^2-risk. Hence, we impose the following assumption:

Assumption A2: For some constants $\nu_1, \nu_2, \lambda_1, \lambda_2 \in \mathbb{R}$, $\alpha_1, \alpha_2 \geq 0$ ($\lambda_1, \lambda_2 > 0$ if $\alpha_1 = \alpha_2 = 0$, $\nu_1 = \nu_2 = 0$) and $K_3, K_4, \beta > 0$, independent of m and n, and for some sequence $\varepsilon_n > 0$, independent of m, one has

$$K_3 \varepsilon_n \; |m|^{-2\nu_1} (\ln |m|)^{-\lambda_1} e^{-\alpha_1 |m|^\beta} \leq \tau_1(m,n) \leq K_4 \varepsilon_n \; |m|^{-2\nu_2} (\ln |m|)^{-\lambda_2} e^{-\alpha_2 |m|^\beta}, \tag{5.4.5}$$

where either $\alpha_1 \alpha_2 \neq 0$ or $\alpha_1 = \alpha_2 = 0$ and $\nu_1 = \nu_2 = \nu > 0$. The sequence ε_n in (5.4.5) is such that

$$n^* = n \varepsilon_n \to \infty \quad (n \to \infty). \tag{5.4.6}$$

In order to construct minimax lower bounds for the L^2-risk, we consider two cases: the dense case and the sparse case, when the hardest functions to estimate are, respectively, uniformly spread over the unit interval T and are represented by only one term in a wavelet expansion. Here also, we apply Lemma (14). Under Assumptions A1 and A2, the following statement is true.

The dense case. Let $\boldsymbol{\omega}$ be the vector with components $\omega_k = \{0, 1\}$. Denote the set of all possible vectors $\boldsymbol{\omega}$ by $\Omega = \{(0,1)^{2^j}\}$. Note that the vector $\boldsymbol{\omega}$ has $\aleph = 2^j$ entries and, hence, $\text{card}(\Omega) = 2^{\aleph}$. Let $H(\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}) = \sum_{k=0}^{2^j-1} \mathbf{1}\left(\tilde{\omega}_k \neq \omega_k\right)$ be the Hamming distance between the binary sequences $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$. Then, the Varshamov-Gilbert Lemma (see, e.g., Tsybakov (2008), p. 104) states that one can choose a subset Ω_1 of Ω, of cardinality at least $2^{\aleph/8}$, such that $H(\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}) \geq \aleph/8$ for any $\boldsymbol{\omega}, \tilde{\boldsymbol{\omega}} \in \Omega_1$.

Consider two arbitrary sequences $\boldsymbol{\omega}, \tilde{\boldsymbol{\omega}} \in \Omega_1$ and the functions f_j and $\tilde{f}_j$ given by

$$f_j(t) = \rho_j \sum_{k=0}^{2^j-1} \omega_k \psi_{jk}(t) \quad \text{and} \quad \tilde{f}_j(t) = \rho_j \sum_{k=0}^{2^j-1} \tilde{\omega}_k \psi_{jk}(t), \quad t \in T. \tag{5.4.7}$$

Choose $\rho_j = A 2^{-j(s+1/2)}$, so that $f_j, \tilde{f}_j \in B^s_{p,q}(A)$. Then, calculating the L^2-norm difference of f_j and $\tilde{f}_j$, we obtain

$$\left|\left|\tilde{f}_j - f_j\right|\right|^2 = \rho_j^2 \left|\left|\sum_{k=0}^{2^j-1} (\tilde{\omega}_k - \omega_k) \psi_{jk}\right|\right|^2 = \rho_j^2 H(\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}) \geq 2^j \rho_j^2 / 8. \tag{5.4.8}$$

Hence, we get $4\delta^2 = 2^j \rho_j^2/8$ in condition (i) of Lemma 14.

Direct calculations yield that, under Assumptions A1, A2 and (5.4.6), for some constants $c_3 > 0$ and

$c_4 > 0$, independent of n,

$$\Delta_1(j,n) \le \begin{cases} c_3\ \varepsilon_n^{-1}\ 2^{2\nu j} j^{\lambda_2}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \\ \\ c_4\ \varepsilon_n^{-1}\ 2^{2\nu_1 j} j^{\lambda_2}\ \ \exp\left\{\alpha_1 \left(\frac{8\pi}{3}\right)^{\beta} 2^{j\beta}\right\}, & \text{if} \quad \alpha_1\alpha > 0. \end{cases} \tag{5.4.9}$$

Apply now Lemma 14 with j such that

$$2\pi A^2 K_1^{-1} n 2^{-2js}\ \Delta_1(j,n) \le 2^j\ \ln 2/16, \tag{5.4.10}$$

i.e.,

$$2^j \asymp \begin{cases} \left[n^*(\ln n^*)^{-\lambda_2}\right]^{\frac{1}{2s+2\nu+1}}, & \text{if} \ \ \beta = 0, \\ (\ln n^*)^{1/\beta}, & \text{if} \ \ \beta > 0, \end{cases} \tag{5.4.11}$$

to obtain

$$\delta^2 = \begin{cases} \left[n^*(\ln n^*)^{-\lambda_2}\right]^{-\frac{2s}{2s+2\nu+1}}, & \text{if} \quad \beta = 0, \\ (\ln n^*)^{-2s/\beta}, & \text{if} \ \ \beta > 0. \end{cases} \tag{5.4.12}$$

The sparse case. Let the functions f_j be of the form $f_j(t) = \rho_j \psi_{jk}(t)$, $t \in T$, and denote

$$\Omega = \{f_j(t) = \rho_j \psi_{jk}(t) : k = 0, 1, \ldots, 2^j - 1, f_0 = 0\}. \tag{5.4.13}$$

Thus, $\text{card}(\Omega) = 2^j$. Choose now $\rho_j = A 2^{-js^*}$, so that $f_j \in B^s_{p,q}(A)$. It is easy to check that, in this case, one has $4\delta^2 = \rho_j^2$ in Lemma 14, and that

$$K(P_\omega, P_{\tilde{\omega}}) \quad \le \quad 2\pi A^2 K_1^{-1} n 2^{-2js'}\ \Delta_1(j,n). \tag{5.4.14}$$

With

$$2^j \asymp \begin{cases} \left[n^*(\ln n^*)^{-\lambda_2 - 1}\right]^{\frac{1}{2s'+2\nu}}, & \text{if} \ \ \beta = 0, \\ (\ln n^*)^{1/\beta}, & \text{if} \ \ \beta > 0, \end{cases} \tag{5.4.15}$$

we then obtain that $K(P_\omega, P_{\tilde{\omega}}) \le 2\pi A^2 K_1^{-1} n 2^{-2js'}\ \Delta_1(j,n)$ and

$$\delta^2 = \begin{cases} \left[\frac{n^*}{(\ln n^*)^{\lambda_2+1}}\right]^{-\frac{2s'}{2s'+2\nu}}, & \text{if} \ \ \beta = 0, \\ (\ln n^*)^{-2s'/\beta}, & \text{if}\ \beta > 0. \end{cases} \tag{5.4.16}$$

Recall that $s' = \min\{s, s^*\}$. By noting that

$$2s/(2s + 2\nu + 1) \le 2s'/(2s' + 2\nu), \quad \text{if} \quad \nu(2 - p) \le ps', \tag{5.4.17}$$

we then choose the highest of the lower bounds in (5.4.12) and (5.4.16). The results can be summarized in the next theorem. □

Theorem 10. Let Assumptions A1 and A2 hold. Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 5.3. Let $s > \max(0, 1/p - 1/2)$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Then, as $n \to \infty$,

$$R_n(B^s_{p,q}(A)) \ge \begin{cases} C(n^*)^{-\frac{2s}{2s+2\nu+1}} \; (\ln n^*)^{\frac{2s\lambda_2}{2s+2\nu+1}}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \; \nu(2-p) < ps', \\ C\Big(\frac{\ln n^*}{n^*}\Big)^{\frac{2s'}{2s'+2\nu}} (\ln n^*)^{\frac{2s'\lambda_2}{2s'+2\nu}}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \; \nu(2-p) \ge ps', \\ C(\ln n^*)^{-\frac{2s'}{\beta}}, & \text{if} \quad \alpha_1\alpha_2 \neq 0. \end{cases} \tag{5.4.18}$$

5.5 Minimax Upper Bounds for the L^2-Risk

Let $\hat{f}_n(\cdot)$ be the (block thresholding) wavelet estimator defined by (5.3.13). Choose now j_0 and J such that

$$2^{j_0} = \ln n^*, \;\; 2^J = (n^*)^{\frac{1}{2\nu+1}}, \quad \text{if} \quad \alpha_1 = \alpha_2 = 0, \tag{5.5.1}$$

$$2^{j_0} = \frac{3}{8\pi}\left(\frac{\ln n^*}{2\alpha}\right)^{\frac{1}{\beta}}, \;\; 2^J = 2^{j_0}, \qquad \text{if} \quad \alpha_1\alpha > 0. \tag{5.5.2}$$

(Since $j_0 > J - 1$ when $\alpha_1\alpha > 0$, the estimator (5.3.13) only consists of the first (linear) part and, hence, λ_j does not need to be selected in this case.) Set, for some constant $\mu > 0$, large enough,

$$\lambda_j = \mu^2 (n^*)^{-1} \ln(n^*)\, 2^{2\nu j} j^{\lambda_1}, \quad \text{if} \;\; \alpha_1 = \alpha_2 = 0. \tag{5.5.3}$$

Note that the choices of j_0, J and λ_j are independent of the parameters, s, p, q and A of the Besov ball $B^s_{p,q}(A)$; hence, the estimator (5.3.13) is adaptive with respect to these parameters.

Denote $(x)_+ = \max(0, x)$,

$$\varrho = \begin{cases} \frac{(2\nu+1)(2-p)_+}{p(2s+2\nu+1)}, & \text{if} \quad \nu(2-p) < ps', \\ \frac{(q-p)_+}{q}, & \text{if} \quad \nu(2-p) = ps', \\ 0, & \text{if} \quad \nu(2-p) > ps'. \end{cases} \tag{5.5.4}$$

Assume that, in the case of $\alpha_1 = \alpha_2 = 0$, the sequence ε_n is such that

$$-h_1 \ln n \le \ln(\varepsilon_n) \le h_2 \ln n \tag{5.5.5}$$

for some constants $h_1, h_2 \in (0,1)$. Observe that condition (5.5.5) implies (5.4.6) and that $\ln n^* \asymp \ln n$ ($n \to \infty$). (Here, and in what follows, $u(n) \asymp v(n)$ means that there exist constants C_1, C_2 ($0 < C_1 \le C_2 < \infty$), independent of n, such that $0 < C_1 v(n) \le u(n) \le C_2 v(n) < \infty$ for n large enough.)

The proof of the minimax upper bounds for the L^2-risk is based on the following two lemmas.

Lemma 21. Let Assumptions A1 and A2 hold. Let the estimators $\widehat{a}_{j_0k}$ and $\widehat{b}_{jk}$ of the scaling and wavelet coefficients a_{j_0k} and b_{jk}, respectively, be given by (5.3.7) with $\widehat{f}_m$ defined by (5.3.5). Then, for all $j \ge j_0$,

$$\mathbf{E}|\widehat{a}_{j_0k} - a_{j_0k}|^2 \le Cn^{-1}\Delta_1(j_0, n) \quad \text{and} \quad \mathbf{E}|\widehat{b}_{jk} - b_{jk}|^2 \le Cn^{-1}\Delta_1(j, n). \tag{5.5.6}$$

If $\alpha_1 = \alpha_2 = 0$ and (5.5.5) holds, then, for any $j \ge j_0$,

$$\mathbf{E}|\widehat{b}_{jk} - b_{jk}|^4 \le Cn^3 (\ln n)^{3\lambda_1} (n^*)^{-\frac{3}{2\nu+1}}. \tag{5.5.7}$$

Proof of Lemma 21. First, consider model (5.1.1). Then, using (5.3.3), (5.3.5), (5.3.7) and (5.3.9), one has

$$\widehat{a}_{j_0k} - a_{j_0k} = \sum_{m \in C_{j_0}} \left(\widehat{f}_m - f_m\right) \overline{\varphi_{mj_0k}}, \quad \widehat{b}_{jk} - b_{jk} = \sum_{m \in C_j} \left(\widehat{f}_m - f_m\right) \overline{\psi_{mjk}}, \tag{5.5.8}$$

where

$$\widehat{f}_m - f_m = \frac{1}{\sqrt{N}} \left(\sum_{l=1}^{M} N^{-2d_l} \overline{g_m(u_l)} z_{lm}\right) / \left(\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2\right). \tag{5.5.9}$$

Consider vector $\mathbf{V}^{(l)}$ with components

$$V_m^{(l)} = N^{-2d_l} \psi_{mjk} g_m(u_l) \left[\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right]^{-1}. \tag{5.5.10}$$

It is easy to see that, due to $|\psi_{mjk}| \le 2^{-j/2}$ and the definition of C_j,

$$\begin{aligned} \left|\left| \mathbf{V}^{(l)} \right|\right|^2 &= N^{-4d_l} \sum_{m \in C_j} |\psi_{mjk}|^2 |g_m(u_l)|^2 \left[\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right]^{-2} \\ &\le 4\pi |C_j|^{-1}\ N^{-4d_l} \sum_{m \in C_j} |g_m(u_l)|^2 \left[\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right]^{-2}. \end{aligned} \tag{5.5.11}$$

Define

$$v_m = \sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 = M \tau_1(m, n). \tag{5.5.12}$$

Hence,

$$\left|\left| \mathbf{V}^{(l)} \right|\right|^2 \le 4\pi |C_j|^{-1}\ N^{-2d_l} N^{-2d_l} \sum_{m \in C_j} |g_m(u_l)|^2 v_m^{-2}. \tag{5.5.13}$$

Using Assumption A1, since z_{lm} are independent for different l's, we obtain

$$\begin{aligned} \mathbf{E}|\widehat{b}_{jk} - b_{jk}|^2 &= \frac{1}{N} \sum_{m_1, m_2 \in C_j} \overline{\psi}_{m_1 jk} \psi_{m_2 jk} \sum_{l=1}^{M} N^{-4d_l} v_{m_1}^{-1} v_{m_2}^{-1} \overline{g_{m_1}(u_l)} g_{m_2}(u_l) \mathrm{Cov}\left(z_{lm_1}, \overline{z}_{lm_2}\right) \\ &= \frac{1}{N} \sum_{l=1}^{M} \overline{\mathbf{V}^{(l)}}^T \mathbf{\Sigma}^{(l)} \mathbf{V}^{(l)} \\ &\le \frac{1}{N} \sum_{l=1}^{M} \lambda_{\max}(\mathbf{\Sigma}^{(l)}) \left|\left| \mathbf{V}^{(l)} \right|\right|^2 \\ &\le 4\pi K_2\, |C_j|^{-1} N^{-1} \sum_{l=1}^{M} N^{-2d_l} \sum_{m \in C_j} |g_m(u_l)|^2 v_m^{-2} \\ &= 4\pi K_2\, |C_j|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-2} \sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 = 4\pi K_2\, |C_j|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-1}, \end{aligned} \tag{5.5.14}$$

so that

$$\mathbf{E}|\widehat{b}_{jk} - b_{jk}|^2 \le C n^{-1} |C_j|^{-1} \sum_{m \in C_j} \left[\tau_1(m, n)\right]^{-1}. \tag{5.5.15}$$

(One can obtain an upper bound for $\mathbf{E}|\widehat{a}_{j_0k} - a_{j_0k}|^2$ by following similar arguments.)

In order to prove (5.5.7), define

$$B_l = N^{-2d_l}\left[\sum_{l=1}^{M} N^{-2d_l}|g_m(u_l)|^2\right]^{-1}. \tag{5.5.16}$$

Note that

$$\mathbf{E}\left(z_{lm_1}z_{lm_2}z_{lm_3}z_{lm_4}\right) \quad \le \quad \left[\Pi_{i=1}^4 \mathbf{E}|z_{m_il}|^4\right]^{1/4}. \tag{5.5.17}$$

Consequently, using Assumption A1, the fact that z_{lm} are independent for different l's, and that $\mathbf{E}|z_{ml}|^4 = 3\left[\mathbf{E}|z_{ml}|^2\right]^2$ for standard (complex-valued) Gaussian random variables z_{ml}, one obtains

$$\begin{aligned}
\mathbf{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(N^{-2}\sum_{l=1}^{M} B_l^4\left[\sum_{m\in C_j}|\psi_{mjk}||g_{m_2}(u_l)|\left(\mathbf{E}|z_{ml}|^4\right)^{1/4}\right]^4\right)\\
&+ O\left(\left[N^{-1}\sum_{l=1}^{M} B_l^2 \sum_{m_1,m_2\in C_j}\overline{\psi}_{m_1jk}\psi_{m_2jk}\overline{g_{m_1}(u_l)}g_m(u_l)\mathrm{Cov}\left(z_{lm_1},\overline{z}_{lm_2}\right)\right]^2\right)\\
&= O\left(N^{-2}\sum_{l=1}^{M} B_l^4\left[\sum_{m\in C_j}|\psi_{mjk}|^2|g_m(u_l)|^2\sum_{m\in C_j}\mathbf{E}|z_{ml}|^2\right]^2\right)\\
&+ O\left(\left[n^{-1}|C_j|^{-1}\sum_{m\in C_j}[\tau_1(m,n)]^{-1}\right]^2\right)
\end{aligned} \tag{5.5.18}$$

Since $\sum_{m\in C_j}\mathbf{E}|z_{ml}|^2 = O(|C_j|)$, one derives

$$\begin{aligned}
\mathbf{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(|C_j|^{-1}\sum_{m\in C_j}\left[\frac{1}{M^3}\frac{\tau_2(m,n)}{[\tau_1(m,n)]^4}\right] + \frac{\Delta_1^2(j,n)}{n^2}\right)\\
&= O\left(M^{-3}\Delta_2(j,n) + n^{-2}\Delta_1^2(j,n)\right).
\end{aligned} \tag{5.5.19}$$

To calculate the asymptotic order of $\Delta_2(j,n)$ when $\alpha_1 = \alpha_2 = 0$, recall that $|g_m(u_l)|^2 \le ||g||_\infty$. Then,

$$\Delta_2(j,n) = O\left(2^{6j\nu}j^{3\lambda_1}\varepsilon_n^{-3}\right). \tag{5.5.20}$$

$$
\begin{aligned}
\Delta_2(j,n) &= O\left(|C_j|^{-1}\sum_{m\in C_j}\tau_2(m,n)\;[\tau_1(m,n)]^{-4}\right)\\
&= O\left(|C_j|^{-1}\sum_{m\in C_j}\frac{M^{-1}\sum_{l=1}^{M}N^{-4d_l}\,|g_m(u_l)|^4}{\left[M^{-1}\sum_{l=1}^{M}N^{-2d_l}\,|g_m(u_l)|^2\right]^4}\right)\\
&= O\left(|C_j|^{-1}\sum_{m\in C_j}\frac{M^{-1}\sum_{l=1}^{M}N^{-2d_l}\,|g_m(u_l)|^2}{\left[M^{-1}\sum_{l=1}^{M}N^{-2d_l}\,|g_m(u_l)|^2\right]^4}\right)\\
&= O\left(|C_j|^{-1}\sum_{m\in C_j}[\tau_1(m,n)]^{-3}\right)\\
&= O\left(2^{6j\nu}j^{3\lambda_1}\varepsilon_n^{-3}\right) \qquad (5.5.21)
\end{aligned}
$$

Thus, using (5.4.9) and the fact that $2^j \le 2^{J-1} < (n^*)^{1/(2\nu+1)}$, (5.5.19) can be rewritten as

$$
\begin{aligned}
\mathbf{E}|\widehat{b}_{jk}-b_{jk}|^4 &= O\left(2^{6\nu j}j^{3\lambda_1}\varepsilon_n^{-3}M^{-3}+2^{4j\nu}j^{2\lambda_1}\varepsilon_n^{-2}n^{-2}\right)\\
&= O\left(n^3\,(\ln n)^{3\lambda_1}\,(n^*)^{-3/(2\nu+1)}\right). \qquad (5.5.22)
\end{aligned}
$$

Hence, (5.5.7) follows. This completes the proof of the lemma. □

Lemma 22. Let Assumptions A1, A2 and (5.5.5) hold. Let the estimators $\widehat{b}_{jk}$ of the wavelet coefficients b_{jk} be given by (5.3.7) with $\widehat{f}_m$ defined by (5.3.5). Let

$$
\mu \ge \sqrt{\frac{2}{1-h_1}}\;\left[\sqrt{c_1}+\frac{\sqrt{8\pi\kappa}}{\sqrt{K_3}}(\ln 2)^{\lambda_1/2}\left(\frac{2\pi}{3}\right)^{\nu}\right], \qquad (5.5.23)
$$

where c_1, K_3 and h_1 are defined in (5.5.36), (5.4.5) and (5.5.5), respectively. Then, for all $j \ge j_0$ and any $\kappa > 0$,

$$
\mathbb{P}\left(\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2 \ge (4n^*)^{-1}\;\mu^2\,2^{2\nu j}\,j^{\lambda}\;\ln n^*\right) \le n^{-\kappa}. \qquad (5.5.24)
$$

Proof of Lemma 22. Consider a set of vectors

$$
\Omega_{jr} = \left\{v_k, k\in U_{jr} : \sum_{k\in U_{jr}}|v_k|^2 \le 1\right\} \qquad (5.5.25)
$$

and a centered Gaussian process

$$Z_{jr} = \sum_{k \in U_{jr}} v_k (\widehat{b}_{jk} - b_{jk}). \tag{5.5.26}$$

Note that

$$\sup_v Z_{jr}(v) = \sqrt{\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2}. \tag{5.5.27}$$

We shall apply below a lemma of Cirelson, Ibragimov and Sudakov (1976) which states that, for any $x > 0$,

$$\Pr\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq (x + B_1) \right) \leq \exp\left(-\frac{x^2}{2B_2} \right), \tag{5.5.28}$$

where,

$$B_1 \quad = \quad \mathbf{E}\left[\sqrt{\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2} \right] \leq \frac{\sqrt{c_1} 2^{j\nu} j^{\lambda_1/2} \sqrt{\ln n}}{\sqrt{n^*}} \tag{5.5.29}$$

with c_1 defined in (5.4.9), and

$$B_2 = \sup_{v \in \Omega_{jr}} \mathrm{Var}(Z_{jr}(v)) = \sup_{v \in \Omega_{jr}} \mathbf{E}|\sum_{k \in U_{jr}} v_k (\widehat{b}_{jk} - b_{jk})|^2. \tag{5.5.30}$$

Denote

$$w_{jm} = \sum_{k \in U_{jr}} v_k \psi_{mjk} \left[\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right]^{-1}, \quad m \in C_j. \tag{5.5.31}$$

Then, under Assumption A2 with $\alpha_1 = \alpha_2 = 0$, using argument similar to the proof of (5.5.6), one obtains

$$\begin{aligned} B_2 \quad &= \quad \sup_{v \in \Omega_{jr}} \left\{ N^{-1} \sum_{m_1, m_2 \in C_j} \overline{w_{jm_1}} w_{jm_2} \ \mathbf{E}\left[\sum_{l=1}^{M} N^{-4d_l} \overline{g_{m_1}(u_l)} g_{m_2}(u_l) z_{lm_1} \overline{z}_{lm_2} \right] \right\} \\ &\leq \quad \sup_{v \in \Omega_{jr}} N^{-1} \sum_{l=1}^{M} N^{-4d_l} \lambda_{\max}(\mathbf{\Sigma}^{(l)}) \sum_{m \in C_j} |w_{jm} g_m(u_l)|^2 \\ &\leq \quad K_3 n^{-1} \sup_{v \in \Omega_{jr}} \left\{ \sum_{m \in C_j} |w_{jm}|^2 \left[\tau_1(m,n) \right]^{-1} \right\} \leq 4\pi C_3^* 2^{2j\nu} j^{\lambda_1} \, (n^*)^{-1}, \end{aligned} \tag{5.5.32}$$

where $C_3^* = (K_3)^{-1} (\ln 2)^{\lambda_1} (2\pi/3)^{2\nu}$.

Apply now inequality (5.5.28) with x such that $x^2 = 2B_2\kappa \ln n$, and note that

$$(x + B_1)^2 = (n^*)^{-1} 2^{2j\nu} j^{\lambda_1} \ln n \left(\sqrt{c_1} + \sqrt{8\pi\kappa K_3^{-1} (\ln 2)^{\lambda_1} (2\pi/3)^{2\nu}} \right)^2 \tag{5.5.33}$$

and

$$\mu^2 \geq 4(1 - h_1)^{-1} \left(\sqrt{c_1} + \sqrt{8\pi\kappa K_3^{-1} (\ln 2)^{\lambda_1} (2\pi/3)^{2\nu}} \right)^2 , \tag{5.5.34}$$

which guarantees (5.5.28). This completes the proof of the lemma. □

Under Assumptions A1 and A2, and using Lemmas 21 and 22, the following statement is true.

Theorem 11. Let Assumptions A1 and A2 hold. Let $\hat{f}_n(\cdot)$ be the wavelet estimator defined by (5.3.13), with j_0 and J given by (5.5.1) (if $\alpha_1 = \alpha_2 = 0$) or (5.5.2) (if $\alpha_1\alpha_2 > 0$) and μ satisfying (5.5.23) with $\kappa = 5$. Let $s > 1/p'$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Then, under (5.4.6) if $\alpha_1\alpha_2 > 0$ or (5.5.5) if $\alpha_1 = \alpha_2 = 0$, as $n \to \infty$,

$$\sup_{f \in B^s_{p,q}(A)} \mathbf{E} \left|\left| \widehat{f}_n - f \right|\right|^2 \leq \begin{cases} C(n^*)^{-\frac{2s}{2s+2\nu+1}} \, (\ln n)^{\varrho + \frac{2s\lambda_1}{2s+2\nu+1}}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \ \nu(2-p) < ps', \\ C\left(\frac{\ln n}{n^*}\right)^{\frac{2s'}{2s'+2\nu}} \, (\ln n)^{\varrho + \frac{2s'\lambda_1}{2s'+2\nu}}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \ \nu(2-p) \geq ps', \\ C(\ln n^*)^{-\frac{2s'}{\beta}}, & \text{if} \quad \alpha_1\alpha_2 > 0. \end{cases} \tag{5.5.35}$$

Proof of Theorem 11. Direct calculations yield that under Assumptions A1, A2 and (5.5.5), for some constants $c_1 > 0$ and $c_2 > 0$, independent of n, one has

$$\Delta_1(j,n) \leq \begin{cases} c_1 \, \varepsilon_n^{-1} \, 2^{2\nu j} j^{\lambda_1}, & \text{if} \quad \alpha_1 = \alpha_2 = 0, \\ c_2 \, \varepsilon_n^{-1} \, 2^{2\nu_1 j} j^{\lambda_1} \, \exp\left\{ \alpha_1 \left(\frac{8\pi}{3}\right)^\beta 2^{j\beta} \right\}, & \text{if} \quad \alpha_1\alpha > 0. \end{cases} \tag{5.5.36}$$

Using 5.5.36, the proof of this theorem is now almost identical to the proof of Theorem 9 of Section 4.6, when $r = 1$. □

Remark 7. Theorems 10 and 11 imply that, for the L^2-risk, the wavelet estimator $\widehat{f}_n(\cdot)$ defined by (5.3.13) is asymptotical optimal (in the minimax sense), or near optimal within a logarithmic factor, over a wide range of Besov balls $B^s_{p,q}(A)$ of radius $A > 0$ with $s > \max(1/p, 1/2)$, $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$. The convergence rates depend on the balance between the smoothness parameter s (of the response function

$f(\cdot)$), the kernel parameters ν, β, λ_1 and λ_2 (of the blurring function $g(\cdot,\cdot)$), the long memory parameters d_l, $l = 1, 2 \ldots, M$ (of the error sequence $\boldsymbol{\xi}^{(l)}$), and how the total number of observations n is distributed among the total number of channels M. In particular, M and d_l, $l = 1, 2, \ldots, M$, jointly determine the value of ε_n which, in turn, defines the "essential" convergence rate $n^* = n\varepsilon_n$ which may differ considerably from n. For example, if $M = M_n = n^\theta$, $0 \le \theta < 1$ and $|g_m(u_l)|^2 \asymp |m|^{-2\nu}$ for every $l = 1, 2 \ldots, M$, then

$$\varepsilon_n = M^{-1} \sum_{l=1}^{M} N^{-2d_l}, \tag{5.5.37}$$

and, therefore, $n^{1-2d^*(1-\theta)} \le n^* \le n$, where $d^* = \max_{1 \le l \le M} d_l$, so that, n^* can take any value between $n^{1-2d^*(1-\theta)}$ and n. This is further illustrated in Section 5.6 below.

5.6 Illustrative Examples

In this section, we consider some illustrative examples of application of the theory developed in the previous sections. They are particular examples of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation.

We assume that condition (5.2.17) holds true and that there exist θ_1 and θ_2, such that $M = M_n$ satisfies

$$n^{\theta_1} \le M \le n^{\theta_2}, \quad 0 \le \theta_1 \le \theta_2 < 1. \tag{5.6.1}$$

(Note that, under (5.6.1), $n^{1-\theta_2} \le N \le n^{1-\theta_1}$.)

Example 14. Consider the case when $g_m(\cdot)$, $m = 0, \pm 1, \pm 2, \ldots$, is of the form

$$g_m(u) = C_g \exp\left(-K|m|^\beta q(u)\right), \quad u \in U, \tag{5.6.2}$$

where $q(\cdot)$ in (5.6.2) is such that, for some q_1 and q_2,

$$0 < q_1 \le q(u) \le q_2 < \infty, \quad u \in U. \tag{5.6.3}$$

This set up takes place in the estimation of the initial condition in the heat conductivity equation or the estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle (see Examples 1 and 2 of Pensky and Sapatinas (2009, 2010)). In the former case, $g_m(u) = \exp(-4\pi^2 m^2 u)$, $u \in U$, so that $K = 4\pi^2$, $\beta = 2$, $q(u) = u$, $q_1 = a$ and $q_2 = b$. In the latter case, $g_m(u) = Cu^{|m|} = C\exp(-|m|\ln(1/u))$, $0 < r_1 \le u \le r_2 < 1$, so that $K = 1$, $\beta = 1$, $q(u) = \ln(1/u)$, $q_1 = \ln(1/r_2)$ and $q_2 = \ln(1/r_1)$.

It is easy to see that, under conditions (5.6.2) and (5.6.3), for $\tau_1(m,n)$ given in (5.4.3),

$$\tau_1(m,n) \le C_g\,\varepsilon_n\, \exp\left(-2Kq_1|m|^\beta\right) \quad \text{and} \quad \tau_1(m,n) \ge C_g\,\varepsilon_n\, \exp\left(-2Kq_2|m|^\beta\right), \tag{5.6.4}$$

where ε_n is of the form (5.5.37). Assumptions (5.2.17) and (5.6.1) lead to the following bounds for n^*:

$$n^{1-2d^*(1-\theta_1)} \le n^* \le n, \tag{5.6.5}$$

so that $\ln n \asymp \ln n^*$. Therefore, according to Theorems 10 and 11,

$$R_n(B^s_{p,q}(A)) \asymp (\ln n)^{-\frac{2s^*}{\beta}}. \tag{5.6.6}$$

Note that, in this case, the value of d^* has absolutely no bearing on the convergence rates of the linear wavelet estimators: the convergence rates are determined entirely by the properties of the smoothness parameter s^* (of the response function $f(\cdot)$) and the kernel parameter β (of the blurring function $g(\cdot,\cdot)$).

In other words, in case of super-smooth convolutions, LRD does not influence the convergence rates of the suugested wavelet estimator. A similar effect is observed in the case of kernel smoothing, see Section 2.2 in Kulik (2008).

Example 15. Suppose that the blurring function $g(\cdot,\cdot)$ is of a box-car like kernel, i.e.,

$$g(u,t) = 0.5\,q(u)\,\mathbf{1}(|t| < u), \quad u \in U, \quad t \in T, \tag{5.6.7}$$

where $q(\cdot)$ is some positive function which satisfies conditions (5.6.3). In this case, the functional Fourier coefficients $g_m(\cdot)$ are of the form

$$g_0(u) = 1 \quad \text{and} \quad g_m(u) = (2\pi m)^{-1}\,\gamma(u)\sin(2\pi m u), \quad m \in \mathbb{Z}\setminus\{0\}, \qquad u \in U. \tag{5.6.8}$$

It is easy to see that estimation of the initial speed of a wave on a finite interval (see Example 4 of Pensky and Sapatinas (2009) or Example 3 of Pensky and Sapatinas (2010)) leads to $g_m(\cdot)$ of the form (5.6.8) with $q(u) = 1$. Assume, without loss of generality, that $u \in [0, 1]$, so that $a = 0$, $b = 1$, and consider (equispaced channels) $u_l = l/M$, $l = 1, 2, \ldots, M$, such that

$$d_l = a_1 u_l + a_2, \quad 0 \le a_2 \le d^* < 1/2,\ 0 \le a_1 + a_2 \le d^* < 1/2, \tag{5.6.9}$$

i.e., condition (5.2.17) holds. Note that if $a_1 = 0$, then

$$\tau_1(m, n) \asymp M^{-1} N^{-2a_2} (4\pi^2 m^2)^{-1} \sum_{l=1}^{M} \sin^2(2\pi m l/M), \tag{5.6.10}$$

which is similar to the expression for $\tau_1(m, n)$ studied in Section 6 of Pensky and Sapatinas (2010). Following their calculations, one obtains that, if j_0 in (5.3.13) is such that $2^{j_0} > (\ln n)^\delta$ for some $\delta > 0$ and $M \ge (32\pi/3) n^{1/3}$, then, for n and $|m|$ large enough,

$$\tau_1(m, n) \asymp N^{-2a_2} m^{-2}. \tag{5.6.11}$$

Assume now, without loss of generality, that $a_1 \ge 0$. (Note that the case of $a_1 \le 0$ can be handled similarly by changing u to $1 - u$.) Below, we shall show that, in this case, a similar result can be obtained under less stringent conditions on $M = M_n$. Indeed, the following statement is true.

Lemma 23. Let $g(\cdot, \cdot)$ be of the form (5.6.7), where $q(\cdot)$ is some positive function which satisfies (5.6.3), and let d_l, $l = 1, 2, \ldots, M$, be given by (5.6.9) with $a_1 \ge 0$. Assume (without loss of generality) that $U = [0, 1]$, and consider $u_l = l/M$, $l = 1, 2, \ldots, M$. Let $M = M_n$ satisfy (5.6.1) with $\theta_1 > 0$ if $a_1 > 0$ and $M \ge (32\pi/3) n^{1/3}$ if $a_1 = 0$. If $m \in A_j$, where $|A_j| = C_m 2^j$, for some absolute constant $C_m > 0$, with $j \ge j_0 > 0$, where j_0 is such that $2^{j_0} \ge C_0 \ln n$ for some $C_0 > 0$, then, for n and $|m|$ large enough,

$$\tau_1(m, n) \asymp N^{-2a_2} m^{-2} (\log n)^{-1}. \tag{5.6.12}$$

Proof of Lemma 23. Below we consider only the case of $a_1 > 0$. Validity of the satement for $a_1 = 0$ follows from Pensky and Sapatinas (2010).

By direct calculations, one obtains that

$$\tau_1(m,n) = M^{-1}(4\pi^2 m^2)^{-1} N^{-2a_2} \sum_{l=1}^{M} q^2(l/M)\, \sin^2(2\pi m l M^{-1}) N^{-2a_1 l/M}. \tag{5.6.13}$$

Therefore,

$$(4\pi^2 m^2)^{-1} q_1^2\, N^{-2a_2} S(m,n) \le \tau_1(m,n) \le (4\pi^2 m^2)^{-1} q_2^2\, N^{-2a_2} S(m,n), \tag{5.6.14}$$

where

$$S(m,n) = M^{-1} \sum_{l=1}^{M} \sin^2(2\pi m l M^{-1}) N^{-2a_1 l/M}. \tag{5.6.15}$$

Denote $p = N^{-2a_1/M}$, $x = 4\pi m M^{-1}$ and note that, as $n \to \infty$,

$$p^M = N^{-2a_1} \to 0 \tag{5.6.16}$$

and

$$\begin{aligned} p &= \exp\left(-2a_1 M^{-1} \ln N\right) \\ &= 1 - 2a_1 M^{-1} \ln N + 2a_1^2 M^{-2} \ln^2 N + o(M^{-2} \ln^2 N), \end{aligned} \tag{5.6.17}$$

since $M^{-1} \ln N \to 0$ as $n \to \infty$.

Using the fact that $\sin^2(x/2) = (1 - \cos x)/2$ and formula 1.353.3 of Gradshtein & Ryzhik (1980), we obtain

$$S(m,n) = \frac{1}{M} \left[\frac{1 - p^M}{1 - p} - \frac{1 - p\cos x - p^M \cos(Mx) + p^{M+1} \cos((M-1)x)}{1 - 2p\cos x + p^2} \right]. \tag{5.6.18}$$

Since m is an integer and $x = 4\pi m M^{-1}$,

$$\cos(Mx) = 1, \quad \sin(Mx) = 0, \quad \cos((M-1)x) = \cos x. \tag{5.6.19}$$

Therefore, simple algebraic transformations yield

$$S(m,n) = \frac{p(p+1)(1-p^M)(1-\cos x)}{M(1-p)[(1-p)^2 + 2p(1-\cos x)]} \tag{5.6.20}$$

The asymptotic expansion (5.6.17) for p as $n \to \infty$, leads to

$$\frac{(1-p^M)}{M(1-p)} \approx \frac{1-N^{-2a_1}}{4a_1 \ln N (1 - a_1 M^{-1} \ln N)}, \tag{5.6.21}$$

so that, if N is large enough, due to $p < 1$, one obtains an upper bound for $S(m,n)$:

$$S(m,n) = \frac{(1-p^M)}{M(1-p)} \left[\frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \right]^{-1} \leq \frac{1}{2a_1 \ln N}. \tag{5.6.22}$$

In order to obtain a lower bound for $S(m,n)$, we note that for N large enough, one has $1/2 < p < 1$. Consider the following two cases: $x \geq \pi/3$ and $x < \pi/3$. If $x \geq \pi/3$, then $\cos x \leq 1/2$ and

$$F(p,x) = \frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \leq 2, \tag{5.6.23}$$

If $x < \pi/3$, we can use the fact that $1 - \cos x = 2\sin^2(x/2) \geq 3x^2/8$, so that

$$F(p,x) \leq \frac{4}{3} \left[1 + \frac{8(1-p)^2}{3x^2} \right] \leq \frac{4}{3} \left[1 + \frac{2a_1^2 \ln^2 N}{3\pi^2 m^2} \right] \tag{5.6.24}$$

for N large enough.

Since $|m| = C_m 2^j > C_m C_0 \ln n$ for some $\delta > 0$ and $\ln n \geq (1-\theta_1)^{-1} \ln N$ due to assumption (5.6.1), one has $m^2 \geq C_m C_0 (1-\theta_1)^{-1} \ln^2 N$ and

$$S(m,n) \geq C(\ln N)^{-1}. \tag{5.6.25}$$

Observe now that $\ln N \asymp \ln n$. This completes the proof of the theorem. $\square$

It follows immediately from Lemma 23 that, if

$$M = M_n = n^\theta, \quad 0 < \theta < 1, \tag{5.6.26}$$

then Assumption A2 holds with $\alpha_1 = \alpha_2 = 0$, $\nu_1 = \nu_2 = \nu = 2$, $\varepsilon_n = n^{-2a_2(1-\theta)} (\ln n)^{-1}$ and $\lambda_1 = \lambda_2 = 0$. Note that ε_n satisfies conditions (5.4.6) and (5.5.5), so that $\ln n \asymp \ln n^*$. Therefore, according to Theorems 10 and 11,

$$R_n(B_{p,q}^s(A)) \geq \begin{cases} C(n^*)^{-\frac{2s}{2s+5}}, & \text{if} \quad 4 - 2p < ps^*, \\ C\left(\frac{\ln n^*}{n^*}\right)^{\frac{s^*}{s^*+2}}, & \text{if} \quad 4 - 2p \geq ps^*, \end{cases} \tag{5.6.27}$$

and

$$\sup_{f\in B^s_{p,q}(A)} \mathbf{E}\left|\left|\widehat{f}_n - f\right|\right|^2 \le \begin{cases} C(n^*)^{-\frac{2s}{2s+5}}\ (\ln n)^{\varrho}, & \text{if} \quad 4-2p < ps^*, \\ C\left(\frac{\ln n}{n^*}\right)^{\frac{s^*}{s^*+2}}\ (\ln n)^{\varrho}, & \text{if} \quad 4-2p \ge ps^*, \end{cases} \tag{5.6.28}$$

where

$$n^* = n^{1-2a_2(1-\theta)}\,(\ln n)^{-1} \tag{5.6.29}$$

and

$$\varrho = \begin{cases} \frac{(5(2-p)_+}{p(2s+5)}, & \text{if} \quad 4-2p < ps^*, \\ \frac{(q-p)_+}{q}, & \text{if} \quad 4-2p = ps^*, \\ 0, & \text{if} \quad 4-2p > ps^*. \end{cases} \tag{5.6.30}$$

Note that LRD affects the convergence rates in this case via the parameter a_2 that appears in the definition (5.6.9).

CHAPTER 6: DISCUSSION

In this dissertation, we have discussed two different nonparametric models using the minimax approach; empirical Bayes model and functional deconvolution model.

In the case of the nonparametric empirical Bayes estimation, we derived lower bounds for the risk of the nonparametric empirical Bayes estimators. In order to attain this convergence rate, we suggested an adaptive wavelet-based method of EB estimation. The method is based on approximating Bayes estimator $t(y)$ corresponding to observation y as a whole using finitely supported wavelet family. The wavelet estimator is used in a rather non-ortodox way: $t(y)$ is estimated locally using only a linear scaling part of the expansion at the resolution level m where coefficients are recovered by solving a system of linear equations.

The advantage of the method lies in its flexibility. The technique works for a variery of families of conditional distributions. Computationally, it leads to solution of a finite system of linear equations which, due to decorrelation property of wavelets, is sparse and well conditioned. The size of the system depends on the size and regularity of the wavelet which is used for representation of the EB estimator $t(y)$.

A non-adaptive version of the method was introduced in Pensky and Alotaibi (2005). However, since no mechanism for choosing the resolution level m of the expansion was suggested, the Pensky and Alotaibi (2005) paper remained of a theoretical interest only. In this dissertation, we use Lepski method for choosing an optimal resolution level m and show that the resulting EB estimator remains nearly asymptotically optimal (within a logarithmic factor of the number of observations n).

Finally, we should comment that, although the choice of a wavelet basis for representation of $t(y)$ is convenient, it is not unique. Indeed, one can use a local polynomial or a kernel estimator for representation of $t(y)$. In this case, the challenge of finding support of the estimator for the local polynomials or bandwidth for a kernel estimator can be addressed by Lepski method in a similar manner. However, the disadvantage of abandoning wavelets will be that the system of equations will cease to be sparse and well-posed.

Another model investigated in the paper is the functional deconvolution model introduced by Pensky and Sapatinas (2009, 2010, 2011). Our study of this model expanded results of Pensky and Sapatinas (2009, 2010, 2011) to the case of estimating an $(r+1)$-dimensional function or the situation of dependent errors. In both cases, we derived minimax lower bounds for the integrated square risk over a wide range of Besov balls and constructed adaptive wavelet estimators that attain those optimal convergence rates.

In particular, in the case of estimating a periodic $(r+1)$-dimensional function, we constructed functional deconvolution estimators based on the hyperbolic wavelet thresholding procedure. We derived the lower and the upper bounds for the minimax convergence rates which confirm that estimators derived in here are adaptive and asymptotically near-optimal, within a logarithmic factor, in a wide range of Besov balls of mixed regularity.

Although results of Kerkyacharian, Lepski and Picard (2001, 2008) have been obtained in a slightly different framework (no convolution), they can nevertheless be compared with the results obtained in the present dissertation. Set $\nu = 0$ to account for the absence of convolution, $p_i = p$ and $d = r+1$. Then, convergence rates in the latter can be identified as rates of a one-dimensional setting with a regularity parameter which is equal to the harmonic mean

$$\bar{s} = \left(\frac{1}{s_1} + \cdots + \frac{1}{s_d} \right)^{-1} < \min_{i=1,\cdots,d} s_i. \tag{6.0.1}$$

In our case, the rates can also be identified as the rates in the one-dimensional setting with a regularity parameter $\min_i s_i$ which is always larger than $\bar{s}$. Moreover, if $s_i = s$, one obtains $\bar{s} = sd > s = \min s_i$, showing that estimators of Kerkyacharian, Lepski and Picard (2001, 2008) in the Nikolski spaces are affected by "the curse of dimensionality" while the estimators in the anisotropic Besov spaces of mixed regularity considered above are free of "the curse of dimensionality" and, therefore, have higher convergence rates.

The problem of deconvolution of a two-dimensional function is related to seismic inversion which can be reduced to solution of noisy convolution equations which deliver underground layer structures along the chosen profiles. The common practice in seismology is to recover layer structures separately for each profile and then to combine them together. This, however, usually is not the best strategy and leads to estimators which are inferior to the ones obtained as two-dimensional functional deconvolutions. Indeed, as it is shown above, unless function f is very smooth in the direction of the profiles, very spatially inhomogeneous along another dimension and the number of profiles is very limited, functional deconvolution solution has precision superior to combination of M solutions of separate convolution equations. The precise condition when separate recoveries are preferable to the two-dimensional one is given by formula (4.5.22) which, essentially, is very reasonable. In fact, if the number M of profiles is small, there is no reason to treat f as a two-dimensional function. Small value of s_2 indicates that f is very spatially inhomogeneous and, therefore, the links between its values on different profiles are very weak. Finally, if s_1 is large, deconvolutions are quite precise, so that combination of various profiles cannot improve the precision.

Finally, we considered a multichannel deconvolution model with long-range dependent (LRD) Gaussian errors. Deconvolution is the common problem in many areas of signal and image processing which include, for instance, LIDAR (Light Detection and Ranging) remote sensing and reconstruction of blurred images. LIDAR is a laser device which emits pulses, reflections of which are gathered by a telescope aligned with the laser (see, e.g., Park, Dho & Kong (1997) and Harsdorf & Reuter (2000)). The return signal is used to determine distance and the position of the reflecting material. However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. If M $(M \geq 2)$ LIDAR devices are used to recover a signal, then we talk about a multichannel deconvolution problem. This leads to the discrete model (5.1.1) considered in this work.

The multichannel deconvolution model (5.1.1) can also be thought of as the discrete version of a model referred to as the functional deconvolution model by Pensky and Sapatinas (2009, 2010). The functional deconvolution model has a multitude of applications. In particular, it can be used in a number of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation. Lattes & Lions (1967) initiated research in the problem of recovering the initial condition for parabolic equations based on observations in a fixed-time strip. This problem and the problem of recovering the boundary condition for elliptic equations based on observations in an internal domain were studied in Golubev & Khasminskii (1999); the latter problem was also discussed in Golubev (2004). Some of these specific models were considered in Section 5.6.

The multichannel deconvolution model (5.1.1) and its continuous version, the functional deconvolution model, were studied by Pensky and Sapatinas (2009, 2010), under the assumption that errors are independent and identically distributed Gaussian random variables. The objective of this discussion was to study the multichannel deconvolution model (5.1.1) from a minimax point of view, with the relaxation that errors exhibit LRD. We were not limited in our consideration to a specific type of LRD: the only restriction made was that the errors should satisfy a general assumption in terms of the smallest and larger eigenvalues of their covariance matrices. In particular, minimax lower bounds for the L^2-risk in model (5.1.1) under such assumption were derived when $f(\cdot)$ is assumed to belong to a Besov ball and $g(\cdot,\cdot)$ has smoothness properties similar to those in Pensky and Sapatinas (2009, 2010), including both regular-smooth and super-smooth convolutions.

In addition, an adaptive wavelet estimator of $f(\cdot)$ was constructed and it was shown that such estimator is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls. The convergence rates of the resulting estimators depend on the balance between the smoothness parameter (of the response function $f(\cdot)$), the kernel parameters (of the blurring function $g(\cdot,\cdot)$), and the long memory parameters d_l, $l = 1, 2 \ldots, M$ (of the error sequence $\boldsymbol{\xi}^{(l)}$), and how the total number of observations is distributed among the total number of channels. Note that SRD is implicitly included in our results by selecting $d_l = 0$, $l = 1, 2, \ldots, M$. In this case, the convergence rates we obtained coincide with the convergence rates obtained under the assumption of independent and identically distributed Gaussian errors by Pensky and Sapatinas (2009, 2010).

If the errors are independent and identically distributed Gaussian random variables, for box-car kernels, it is known that, when the number of channels in the multichannel deconvolution model (5.1.1) is finite, the precision of reconstruction of the response function increases as the number of channels M grow (even when the total number of observations n for all channels M remains constant) and this requires the channels to form a Badly Approximable (BA) M-tuple (see De Canditiis and Pensky (2004, 2007)). Under the same assumption for the errors, Pensky and Sapatinas (2009, 2010) showed that the construction of a BA M-tuple for the channels is not needed and a uniform sampling strategy for the channels with the number of channels increasing at a polynomial rate (i.e., $u_l = l/M$, $l = 1, 2, \ldots, M$, for $M = M_n \geq (32\pi/3)n^{1/3}$) suffices to construct an adaptive wavelet estimator that is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls, when the blurring function $g(\cdot,\cdot)$ is of box-car like kernel (including both the standard box-car kernel and the kernel that appears the estimation of the initial speed of a wave on a finite interval). Example 15 showed that a similar result is still possible under long-range dependence with (equispaced channels) $u_l = l/M$, $l = 1, 2, \ldots, M$, $n^{\theta_1} \leq M = M_n \leq n^{\theta_2}$, for some $0 \leq \theta_1 \leq \theta_2 < 1$ when $d_l = a_1 u_l + a_2$, $l = 1, 2, \ldots, M$, $0 \leq a_2 < 1/2$, $0 \leq a_1 + a_2 < 1/2$.

However, in real-life situations, the number of channels $M = M_n$ usually refers to the number of physical devices and, consequently, may grow to infinity only at a slow rate as $n \to \infty$. When $M = M_n$ grows slowly as n increases, (i.e., $M = M_n = o((\ln n)^\alpha)$ for some $\alpha \geq 1/2$), in the multichannel deconvolution model with independent and identically distributed Gaussian errors, Pensky and Sapatinas (2011) developed a procedure for the construction of a BA M-tuple on a specified interval, of a non-asymptotic length, together with a lower bound associated with this M-tuple, which explicitly shows its dependence on M as M grows. This result was further used for the derivation of upper bounds for the L^2-risk of the suggested adaptive wavelet thresholding estimator of the unknown response function and, furthermore, for the choice of the optimal number of channels M which minimizes the L^2-risk. It would be of interest to see whether or not similar upper bounds are possible under long-range dependence. Another avenue of possible research is to consider an analogous minimax study for the functional deconvolution model (i.e., the continuous version of the multichannel deconvolution model (5.1.1)) under long range-dependence (e.g., modeling the errors as a fractional Brownian motion) and examine the effect of the convergence rates between the two models, similar to the convergence rate study of Pensky and Sapatinas (2010) when the errors were considered to be independent and identically distributed Gaussian random variables.

For future work, in the case of the empirical Bayes Model, we plan to consider the compound estimation problem, i.e., the case when n values $x_1, \cdots, x_n$ are observed where x_i's are conditionally independent and are distributed according to the pdfs $q(\cdot \mid \theta_i)$, $i = 1, \cdots, n$, and θ_i are independent with the common prior pdf $g(\theta)$. In this scheme, the form of conditional pdf $q(x \mid \theta)$ is known, $g(\theta)$ is unknown and the goal is to estimate the collection of unknown θ's, $\theta_1, \cdots, \theta_n$. This is an important problem which has a variety of applications (see, e.g. Brown and Greenshtein (2009), Brown *et al.* (2005) and Raykar and Zhao (2010)).

REFERENCES

[1] Abramovich, F. and Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115-129.

[2] Alvarez-Lacalle, E., Dorow, B., Eckmann, J. P., Moses, E. (2006). Hierarchical Structures Induce Long-Range Dynamical Correlation in Written Texts. *PNAS.* **103**, 7956–7961.

[3] Benhaddou, R., Kulik, R., Pensky, M., Sapatinas, T. (2013). Multichannel Deconvolution with Long-Range Dependence: A Minimax Study. *Journal of Statistical Planning and Inference.* ArXiv1305-1.0274.pdf

[4] Benhaddou, R., Pensky, M. (2012). Adaptive Nonparametric Empirical Bayes Estimation Via Wavelet Series. *Journal of Statistical Planning and Inference.*(accepted) ArXiv:1205.0990v1

[5] Benhaddou, R., Pensky, M., Picard, D. (2013). Anisotropic Denoising in Functional Deconvolution Model with Dimension-free Convergence Rates. *Electronic Journal of Statistics.*(accepted) ArXiv1211.7114v2.pdf

[6] Beran, J. (1992). Statistical Methods for Data with Long-Range Dependence. *Statistical Science.* **4**, 404-416.

[7] Beran, J. (1994). Statistics for long-memory processes. *Monographs on Statistics and Applied Probability,* Chapman and Hall, New York.

[8] Berkhout, A.J. (1986) The seismic method in the search for oil and gas: Current techniques and future developments. *Proc. IEEE*, **74**, 1133-1159.

[9] Brown, L. D., Greenshtein, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Ann. Statist.*, **37**, no. 4, 1685-1704,

[10] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., Zhao, L. (2005). Statistical analysis of a telephone call center: a queueing-science perspective. *J. Amer. Statist. Assoc.*, **100**, 36-50,

[11] Bunea, F., Tsybakov, A. and Wegkamp, M.H. (2007). Aggregation for Gaussian regression. *Ann. Statist.*, **35**, 1674-1697.

[12] Carlin, B.P., Louis, T.A.(2000) *Bayes and empirical Bayes methods for data analysis*, Second ed., Chapman& Hall/CRC, Boca Raton.

[13] Casella, G. (1985) An introduction to empirical Bayes data analysis. *The American Statistician*, **39**, 83-87.

[14] Casella, G. and Berger, R. L. (2002). *Statistical Inference* (2nd ed.), Duxbury Advanced Series, Pacific Grove.

[15] Cheng, B., Robinson, P. M. (1994) Semiparametric Estimation from time series with long-range dependence. *J. Econometrics*, **64**, 335-353.

[16] Comte, F., Dedecker, J., Taupin, M. L. (2008) Adaptive density deconvolution with dependent inputs. *Math. Methods in Statistics*, **17**, 87-112.

[17] Csorgo, S., Mielniczuk, J. (2000) The Smoothing Dichotomy in Random-Design Regression with Long-Memory Errors Based on Moving Averages. *Statist. Sinica*, **10**, 771-787.

[18] Datta, S. (1991) Nonparametric empirical Bayes estimation with $O(n^{-1/2})$ rate of a truncation parameter. *Statist. Decisions*, **9**, 45-61.

[19] Datta, S. (2000) Empirical Bayes estimation with non-identical components. *J. Nonparametr. Statist.*, **12**, 709-725.

[20] Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, **2**, 101-126.

[21] Donoho, D.L. and Raimondo, M. (2004). Translation invariant deconvolution in a periodic setting. *International Journal of Wavelets, Multiresolution and Information Processing*, **14**, 415-432.

[22] Doukhan, P., Oppenhein, G., Taqqu, M. S. (2003). Theory and applications of long-range dependence. *Birkhauser Boston Inc.*, Boston, MA.

[23] Durrett, R. (1996). *Probability: Theory and Examples* (2nd ed.), Duxbury Press, Belmont.

[24] Efron, B., Morris, C. N. (1977) Stein's paradox in statistics. *Scientific American*, **236**, 119-127.

[25] Fan, J. and Koo, J. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, **48**, 734-747.

[26] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.

[27] Ghosh, M., Lahiri, P. (1987) Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.*, **82**, 1153-1162.

[28] Ghosh, M., Meeden, G. (1986) Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.*, **81**, 1058-1062.

[29] Golubev, G. (2004) The principle of penalized empirical risk in severely ill-posed problems. *Probability Theory and Related Filelds*, **130**, 18–38.

[30] Golubev, G.K. and Khasminskii, R.Z. (1999). A statistical approach to some inverse problems for partial differential equations. *Problems of Information Transmission*, **35**, 136–149.

[31] Gradshtein, I.S. and Ryzhik, I.M. (1980). *Tables of Integrals, Series, and Products.* New York: Academic Press.

[32] Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and their Applications.* California Monographs in Mathematical Sciences, Berkeley-Los Angeles: University of California Press.

[33] Harsdörf, S. and Reuter, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

[34] Heimer, A., Cohen, I. (2008) Multichannel blind seismic deconvolution using dynamic programming. *Signal Processing*, **88**, 1839-1851.

[35] Heping, W. (2004) Representation and approximation of multivariate functions with mixed smoothness by hyperbolic wavelets *J. Math. Anal. Appl.*, **291**, 698-715.

[36] Holdai, V., and Korostelev, A. (2008) Image Reconstruction in Multi-Channel Model under Gaussian Noise. **Math. Meth. Statist.**, **17**, 198-208.

[37] Huang, S.Y. (1997) Wavelet based empirical Bayes estimation for the uniform distribution. *Statist. Probab. Lett.*, **32**, 141-146.

[38] Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004) Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, Series B, **66**, 547-573 (with discussion, 627-657).

[39] Johnstone, I.M. and Raimondo, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Annals of Statistics*, **32**, 1781-1804.

[40] Jiang, W., Zhang, C.-H. (2009) General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.*, **37**, 1647-1684.

[41] Kalifa, J. and Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, **31**, 58-109.

[42] Karagiannis, T., Molle, M., Faloutsos, M. (2004) Long-Range Dependence, ten years of internet traffic modeling. *IEEE Internet Computing*, **8**, 57-64.

[43] Karmeshu, D., Krishnamashari, A. (2004) Sequence Variability and Long-Range Dependence in DNA. *IEEE Internet Computing*, **8**, 1354-1361, Berlin: Springer, 3316 of Lecture Notes in Computer Science.

[44] Karunamuni, R. J., Singh, R. S., Zhang, S. (2002) On empirical Bayes estimation in the location family. *J. Nonparam. Stat.* , **14**, 435-448.

[45] Karunamuni, R.J.; Zhang, S. (2003) Optimal linear Bayes and empirical Bayes estimation and prediction of the finite population mean. *J. Statist. Plan. Inference*, **113**, 505-525.

[46] Kerkyacharian, G., O. Lepski, O., Picard, D. (2001) Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields* , **121**, 137-170.

[47] Kerkyacharian, G., O. Lepski, O., Picard, D. (2008) Nonlinear Estimation in Anisotropic Multi-Index Denoising. Sparse Case. *Theory Probab. Appl.*, **52**, 58-77.

[48] Kerkyacharian, G., Picard, D. and Raimondo, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statistica Sinica*, **7**, 317-340.

[49] Kulik, R. (2008) Nonparametric deconvolution problem for dependent sequences. *Electronic Journal of Statistics*, **2**, 722-740.

[50] Kulik, R., Raimondo, M. (2009a) L^p wavelet regression with correlated errors. *Statistica Sinica*, **19**, 1479-1489.

[51] Kulik, R., Raimondo, M. (2009b) Wavelet Regression in Random Design with Heteroscedastic Dependent Errors. *Annals of Statistics*, **37**, 3396-3430.

[52] Lattes, R. and Lions, J.L. (1967). *Methode de Quasi-Reversibilite et Applications*. Travoux et Recherche Mathematiques, **15**. Paris: Dunod.

[53] Lepski, O.V. (1991). Asymptotic mimimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 654-659.

[54] Lepski, O. V., Mammen, E., Spokony, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimators with variable bandwidth selectors. *Ann. Statist*, **25**, 929-947.

[55] Lo, A. (2001) Fat tails, long memory and the stock market since the 1960s. *Economic Notes*, Banca Monte dei Paschi di Siena.

[56] Louis, T. A. (1984) Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc*, **79**, 393-398.

[57] Ma, Y., Balakrishnan, N. (2000) Empirical Bayes estimation for truncation parameters. *J. Statist. Plan. Inference*, **84**, 111-120.

[58] Mallat, S. *a Wavelet Tour of Signal Processing*. 2nd Edition, Academic Press, San Diego.

[59] Maritz, J. S., Lwin, T. (1989). *Empirical Bayes Methods with Applications*. Chapman & Hall, London.

[60] Mielniczuk, J. Wu, W. B. (2004) On Random-Design model with Dependent Errors. *Statist. Sinica*, **14**, 1105-1126 .

[61] Morettin, P. S. *Wavelets in Statistics*. Text for Tutorial in The Third International Conference on Statistical Data Analysis Based on the L^1-Norm and Related Methods. Neuchatel, Switzerland, August 11-15, 1997.

[62] Morris, C. N. (1983) Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.*, **78**, 47-65.

[63] Navarro, F., Chesneau, C., Fadili, J., and Sassi, T. (2013) Block thresholding for wavelet-based estimation of function derivatives from a heteroscedastic multichannel convolution model. *Electron. J. Statist.*, **7**, 428-453.

[64] Neelamani, R., Choi, H. and Baraniuk, R. (2004). Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, **52**, 418-433.

[65] Nikolski S.M. (1975) *Approximation of functions of several variables and imbedding theorems* (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975.

[66] Nogami, Y. (1988) Convergence rates for empirical Bayes estimation in the uniform $U(0, \theta)$ distribution. *Ann. Stat.*, **16**, 1335-1341.

[67] Painter, S. (1998) Long-Range Dependence in subsurfaces: Empirical Evidence and Simulation Methods. *Invited paper at the American Geophysical Union*, Fall Meeting, 1998.

[68] 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.

[69] Pensky, M. (1997). Empirical Bayes estimation of a location parameter. *Statist. Decisions*, **15**, 1-16.

[70] Pensky, M. (1998). Empirical Bayes estimation based on wavelets. *Sankhya*, **A60**, 214-231.

[71] Pensky, M. (2000). Adaptive wavelet empirical Bayes estimation of a location or a scale parameter. *J. Statist. Plann. Inference*, **90**,275-292.

[72] Pensky, M. (2002). Locally adaptive wavelet Bayes estimation of location parameter. *Ann. Inst. Statist. Math*, **54**, 83-99.

[73] Pensky, M., Allotaibi, M. (2005). Generalization of linear empirical Bayes estimation via wavelet series. *Statistics and Decisions*, **23**, 181-198.

[74] Pensky, M., Ni, P. (2000) Extended linear empirical Bayes estimation. *Communications in Statistics - Theory and Methods*, **29**, 579-592.

[75] Pensky, M., Sapatinas, T. (2009) Functional Deconvolution in a Periodic Setting: Uniform Case. *Annals of Statistics*, **37**, 73-104.

[76] Pensky, M., Sapatinas, T. (2010) On Convergence Rates Equivalency and Sampling Strategies in a Functional Deconvolution Model. *Annals of Statistics*, **38**, 1793-1844.

[77] Pensky, M., Sapatinas, T. (2011) Multichannel Boxcar Deconvolution with Growing Number of Channels. *Electronic Journal of Statistics*, **5**, 53-82. .

[78] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, **27**, 2033–2053.

[79] Raykar, V., Zhao, L. (2010). Empirical Bayesian thresholding for sparse signals using mixture loss functions to appear in Statistica Sinica.

[80] Robbins, H. (1955). An Empirical Bayes approach to statistics. In *proceedings of the third Berkeley symposium of mathematical statistics and probability.* University of California Press, Berkeley: 157-163.

[81] Robbins, H. (1983). Some thoughts on Empirical Bayes Estimation. *Ann. Statist.*, **11**, 713-723.

[82] Robinson, E.A. (1999) *Seismic Inversion and Deconvolution: Part B: Dual-Sensor Technology.* Elsevier, Oxford.

[83] Robinson, E.A., Durrani, T.S., Peardon, L.G. (1986) *Geophysical Signal Processing.* Prentice-Hall, London.

[84] Singh, R. S. (1976) Empirical Bayes estimation with convergence rates in noncontinuous Lebesgue exponential families. *Ann. Statist.*, **4**, 431-439.

[85] Singh, R. S. (1979). Empirical Bayes Estimation in Lebesgue-exponential families with rates near the best possible rate. *Ann. Statist*, **7**, 890-902.

[86] Tsybakov, A.B. (2008) *Introduction to Nonparametric Estimation*, Springer-Verlag, New York.

[87] Van Zanten, H., Zareba, P. (2008) A note on wavelet density deconvolution for weakly dependent data. *Statistical Inference for Stochastic Processes*, **11**, 207-219.

[88] Varotsos, C., Kirk-Davidoff, D. (2006) Long-memory processes in global ozone and temperature variations at the region 60S-60N. *Atmospheric Chemistry and Physics*, **6**, 4093-4100.

[89] Walter, G. and Shen, X. (1999). Deconvolution using Meyer wavelets. *Journal of Integral Equations and Applications* **11** 515-534.

[90] Walter, G. G., Shen, X. (2001) Wavelets and Other Orthogonal Systems. Chapman and Hall/ CRC, Boca Raton.

[91] Wang, Y. (2012) Minimax Estimation via Wavelets for Indirect Long-Memory Data. *The Journal of Statistical Planning and Inference* . To appear.

[92] Wason, C.B., Black, J.L., King, G.A. (1984) Seismic modeling and inversion. *Proc. IEEE*, **72**, 1385-1393.

[93] Wishart, J. M. (2012) Wavelet Deconvolution in a Periodic Setting with Long-Range Dependent Errors. *The Journal of Statistical Planning and Inference*, **5**, 867-881.

[94] Yang, Y. (2001) Nonparametric Regression with Dependent Errors. *Bernoulli*, **7**, 633-655. Diversity Press, 1989.