Multichannel deconvolution with long-range dependence: A minimax study

Rida Benhaddou\textsuperscript{a}, Rafal Kulik\textsuperscript{b}, Marianna Pensky\textsuperscript{a,}\textsuperscript{*}, Theofanis Sapatinas\textsuperscript{c}

\textsuperscript{a} Department of Mathematics, University of Central Florida, USA
\textsuperscript{b} Department of Mathematics and Statistics, University of Ottawa, Canada
\textsuperscript{c} Department of Mathematics and Statistics, University of Cyprus, Cyprus

Abstract

We consider the problem of estimating the unknown response function in the multichannel deconvolution model with long-range dependent Gaussian or sub-Gaussian errors. We do not limit our consideration to a specific type of long-range dependence rather we assume that the errors should satisfy a general assumption in terms of the smallest and largest eigenvalues of their covariance matrices. We derive minimax lower bounds for the quadratic 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, for both Gaussian and sub-Gaussian errors. It is shown that the optimal convergence rates depend on 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 or sub-Gaussian errors.

1. Introduction

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

$$y(u_l, t_i) = \int_T g(u_l, t_i - x) f(x) \, dx + \xi_{li}, \quad l = 1, 2, \ldots, M, \quad i = 1, 2, \ldots, N,$$

(1.1)

\textsuperscript{*} Corresponding author.

E-mail address: Marianna.pensky@ucf.edu (M. Pensky).

0378-3758/$ - see front matter \copyright\ 2014 Elsevier B.V. All rights reserved.

http://dx.doi.org/10.1016/j.jspi.2013.12.008
where \( g \) is known, \( u_i \in U = [a, b], 0 < a \leq b < \infty, T = \{0, 1\}, \) and the errors \( \xi_{il} \) are Gaussian or sub-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 \( \xi^b = \{\xi_{il} \}_{i=1}^{N} \) be a zero mean vector with components \( \xi_{il}, l = 1, 2, \ldots, N \), and let \( \Sigma^b = \text{Cov}(\xi^b) = \mathbb{E}[\xi^b(\xi^b)^T] \) be its covariance matrix. Hence errors \( \xi_{il} \) are independent for different \( l \)'s, but dependent for different \( i \)'s. Let \( G^0(b) \) be a matrix such that \( G^0(b)^T \Sigma^b G^0(b) = \Sigma^0 \). Then a vector \( \eta(b) = (G^0(b)^T - G^0(b) \) has the covariance matrix \( I_N^b \), the identity matrix of size \( N \).

In order to formulate our main assumption, recall that a random variable \( \zeta \) is sub-Gaussian if

\[
\|\zeta\|_{\psi_2} = \sup_{p > 1} \left( \mathbb{E}[(\zeta^p)^2] \right)^{1/p} < \infty.
\]

Examples of sub-Gaussian random variables include Gaussian, Bernoulli or any bounded random variable. See Section 5.2.3 of Vershynin (2011) for more details. We consider the following assumption on the errors:

**Assumption A0 (AOG).** Vectors \( \xi^b \) are of the forms

\[
\xi^b = G^0(b) \eta^b
\]

where \( \eta^b \) are independent vectors with independent sub-Gaussian (or Gaussian) components \( \eta_b \) for every \( l = 1, 2, \ldots, M \), and \( i = 1, 2, \ldots, N \), such that \( \|\eta^b\|_{\psi_2} < K, 0 < K < \infty. \)

(In what follows, we consider the cases when one knows that \( \eta^b \) are Gaussian vectors and refer to this stronger version of Assumption A0 as Assumption AOG.)

Furthermore, we impose the following condition on the dependence structure.

**Assumption A1.** For each \( l = 1, 2, \ldots, M \), \( \Sigma^0 \) 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 \( l = 1, 2, \ldots, M \),

\[
K_1N^{2d_l} \leq \lambda_{\min}(\Sigma^0) \leq \lambda_{\max}(\Sigma^0) \leq K_2N^{2d_l}, \quad 0 \leq d_l < 1/2,
\]

where \( \lambda_{\min}(\Sigma^0) \) and \( \lambda_{\max}(\Sigma^0) \) are the smallest and largest eigenvalues of (the Toeplitz matrix) \( \Sigma^0 \).

Assumption A1 is valid when, for each \( l = 1, 2, \ldots, M \), \( \xi^b \) is a second-order stationary Gaussian sequence with spectral density satisfying certain assumptions. We shall elaborate on this issue in Section 2. Note that, in the case of independent errors, for each \( l = 1, 2, \ldots, M \), \( \Sigma^0 \) is proportional to the identity matrix and that \( d_l = 0 \). In this case, the multichannel deconvolution model (1.1) reduces to the one with independent and identically distributed Gaussian errors. In a view of (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).

Model (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. For instance, the problem of recovering the initial condition for parabolic equations based on observations in a fixed-time trip was first investigated in Lattices 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 Khasminskii (1999) and Golubev (2004).

In the case when \( a = b \), the functional deconvolution model reduces to the standard deconvolution model. This model has been the subject of a great array of research papers since late 1980s, but the most significant contribution was that of Donoho (1995) who was the first to devise a wavelet solution to the problem. This has attracted the attention of a good deal of researchers, see, e.g., Abramovich and Silverman (1998), Kalifa and Mallat (2003), Donoho and Raimondo (2004), Johnstone and Raimondo (2004), Johnstone et al. (2004), Kerkyacharian et al. (2007). (For related results on the density deconvolution problem, we refer to, e.g., Pensky and Vidakovic, 1999; Walter and Shen, 1999; Fan and Koo, 2002.)

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) has 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, e.g., 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 setup 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 and Zareba (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 consider 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 (\(\sigma(x, \xi_i)\)). In Cśorgo and Mielniczuk (2000), the error terms were modeled as infinite order
moving average processes. Mielniczuk and Wu (2004) investigated another form of LRD, with the assumption that \(x_i\) and \(\xi_i\)
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 (2013) applied a fractional Brownian
motion to model the presence of LRD, while Wang (1997) 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. For further
reference on nonparametric regression with long range dependent errors we refer to Sections 7.4 and 7.5 in Beran et al.
(2013).

The objective of this paper is to study the multichannel deconvolution model from a minimax point of view, with the
relaxation that errors may be sub-Gaussian and 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 (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 when the errors are Gaussian, and
near-optimal (within a logarithmic factor) when the errors are sub-Gaussian. Moreover, the estimator adapts to sub-
Gaussianity of errors since its form does not depend on the nature of errors.

We prove that the convergence rates of the resulting estimators depend on the balance between the smoothness
parameter (of the response function \(G(\cdot)\)) and 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 \(\xi(0)\)). 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 LRD in the case of super-smooth convolutions, however, the situation changes in the case of regular-smooth
convolutions.

The paper is organized as follows. Section 2 discusses stationary sequences with LRD errors, justifies Assumption A1 and
provides illustrative examples of stationary sequences satisfying this assumption. Section 3 describes the construction of the
suggested wavelet estimator of \(f(\cdot)\). Section 4 derives minimax lower bounds for the \(L^2\)-risk for observations from model (1.1). Section 5 proves that the suggested wavelet estimator is adaptive and asymptotically optimal (in the minimax sense) or near-optimal (within a logarithmic factor), in a wide range of Besov balls. The Gaussian and sub-Gaussian cases are
treated separately. Section 7 presents 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. Section 8 concludes with a brief discussion. Appendix A contains the proofs of
the theoretical results obtained in earlier sections.

2. Stationary sequences with long-range dependence

In this section, for simplicity of exposition, we consider one sequence of errors \(\{\xi_j : j = 1, 2, \ldots\}\). Assume that \(\{\xi_j : j = 1, 2, \ldots\}\) is a second-order stationary sequence with covariance function \(\gamma(l) = \gamma(k), k = 0, \pm 1, \pm 2, \ldots\). The spectral density is defined as

\[
a(\omega) := a(\lambda) := \frac{1}{2\pi} \sum_{k = -\infty}^{\infty} \gamma(k) e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi].
\]

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

\[
\gamma(k) = \frac{1}{\pi} \int_{-\pi}^{\pi} a(\lambda) e^{ik\lambda} d\lambda, \quad k = 0, \pm 1, \pm 2, \ldots,
\]

under the assumption that the spectral density \(a(\lambda), \lambda \in [-\pi, \pi]\), is squared-integrable.

Let \(\Sigma = [\gamma(j-k)]_{j,k=1}^{\infty}\) be the covariance matrix of \(\{\xi_1, \ldots, \xi_N\}\). Define \(X = (x \in C^N : x^*x = 1)\), where \(x^*\) is the complex-conjugate of \(x\). Since \(\Sigma\) is Hermitian, one has

\[
\lambda_{\min}(\Sigma) = \inf_{x \in X} (x^* \Sigma x) \quad \text{and} \quad \lambda_{\max}(\Sigma) = \sup_{x \in X} (x^* \Sigma x). \quad (2.1)
\]
With the definitions introduced above,

\[
x^* \Sigma x = \sum_{j,k=1}^{N} x^*_j (j-k) x_k = \int_{-\pi}^{\pi} \left| \sum_{j=1}^{N} x_j e^{-ij \lambda} \right|^2 a(\lambda) \, d\lambda.
\]  
(2.2)

Note that, by the Parseval identity, the function \( h(\lambda) = |\sum_{j=1}^{N} x_j e^{-ij \lambda}|^2, \lambda \in [-\pi, \pi], \) belongs to the set

\[ \mathcal{H}_N = \left\{ h : h \text{ symmetric}, \ |h|_\infty \leq N, \int_{-\pi}^{\pi} h(\lambda) \, d\lambda = 2\pi \right\}. \]

Let \( d \in [0,1/2] \). Consider the following class of spectral densities:

\[ \mathcal{F}_d = \{a : a(\lambda) = |\lambda|^{-2d} a_0(\lambda), \ 0 < C_{\min} \leq |a_0(\lambda)| \leq C_{\max} < \infty, \lambda \in [-\pi, \pi]\}. \]

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 (2.3).

**Fractional ARIMA**(0, d, 0). Let \( \xi_j : j = 1,2,\ldots \) be the second-order stationary sequence

\[ \xi_j = \sum_{m=0}^{\infty} a_m \eta_{j-m}, \]

where \( \eta_j \) are uncorrelated, zero-mean, random variables, \( \sigma^2 = \text{Var}(\eta_j) < \infty \), and

\[
a_m = (-1)^m \left( -\frac{d}{m} \right) = (-1)^m \frac{\Gamma(1-d)}{\Gamma(m+1) \Gamma(1-d-m)},
\]

with \( d \in [0,1/2] \). Then, \( a_m, m = 0,1,\ldots \), are the coefficients in the power-series representation

\[ A(z) = (1-z)^{-d} = \sum_{m=0}^{\infty} a_m z^m. \]

Therefore, the spectral density \( a(\lambda), \lambda \in [-\pi, \pi], \) of \( \{\xi_j : j = 1,2,\ldots\} \), is given by

\[
a(\lambda) = \frac{a_0}{2\pi} |A(e^{-i\lambda})|^2 = \frac{\sigma^2}{2\pi} \left| 1 - e^{-i\lambda} \right|^{-2d} = \frac{\sigma^2}{2\pi} |2(1 - \cos \lambda)|^{-d} \sim \frac{\sigma^2}{2\pi} |\lambda|^{-2d} \quad (\lambda \to 0). \]

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 (2.3). The sequence \( \{\xi_j : j = 1,2,\ldots\} \) is called the fractional ARIMA(0, d, 0) time series. Such models were introduced in Box and Jenkins (1970) and studied extensively since then. We refer to Section 2.1.1.4 of Beran et al. (2013) for summary of its properties.

**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., Geweke and Porter-Hudak, 1983, p. 222)

\[
a(\lambda) = \sigma^2 (2\pi)^{-2H-2} \Gamma(2H+1) \sin(\pi H)4 \sin^2(\lambda/2) \times \sum_{k = -\infty}^{\infty} |k+(\lambda/2\pi)|^{-2H-1},
\]

and, hence,

\[
a(\lambda) = \frac{2\sigma^2}{\pi} \Gamma(2H+1) \sin(\pi H) \lambda^{1-2H}(1+o(1)) \quad (\lambda \to 0). \]

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. We refer to Section 1.3.5 in Beran et al. (2013) for its further properties.

It follows from (2.3) that, for \( a \in \mathcal{F}_d, \) one has \( a(\lambda) \sim |\lambda|^{-2d} \quad (\lambda \to 0). \) 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 \( \Sigma \) are of asymptotic order \( N^{2d} \quad (N \to \infty). \) In particular, the following lemma is true.

**Lemma 1.** 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,

\[ K_{1d} N^{2d} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq K_{2d} N^{2d}. \]

**Remark 1.** If \( d = 0 \), then \( \mathcal{F}_d \) is the class of spectral densities \( a(\lambda) \) that are bounded away from 0 and \( \infty \) 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 1 reduces to a result in Grenander and Szegö (1958, Section 5.2).

**Corollary 1.** For each \( l = 1,2,\ldots,M \), let \( \xi_l^{(0)} \) be a second-order stationary Gaussian sequence with spectral density \( a_l \in \mathcal{F}_d, \) \( dl \in [0,1/2) \). We assume that \( \xi_l^{(0)} \) 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 \( l = 1, 2, \ldots, M, \)
\[ 0 \leq d_l \leq d^* < 1/2. \]
Then, Assumption A1 holds.

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 \( \overline{a}, \) the discrete Fourier basis on the interval \( T \) by \( e_m(t_i) = e^{-2\pi i m t_i}, t_i = i/N, i = 1, 2, \ldots, N, \) \( m = 0, \pm 1, \pm 2, \ldots, \) and the complex-conjugate of the matrix \( A \) by \( A^\dagger. \)

Recall the multichannel deconvolution model (1.1). Denote
\[ h(u_l, t_i) = \int_T g(t_i - x)f(x) \, dx, \quad l = 1, 2, \ldots, M, \ t_i = 1, 2, \ldots, N. \]
Then, Eq. (1.1) can be rewritten as
\[ y_m(u_l) = h_m(u_l) + \epsilon_m, \quad l = 1, 2, \ldots, M, \ i = 1, 2, \ldots, N. \]  

For each \( l = 1, 2, \ldots, M, \) let \( h_m(u_l) = \langle \epsilon_m, h(u_l, \cdot) \rangle, \ y_m(u_l) = \langle \epsilon_m, y(u_l, \cdot) \rangle, \ z_m = \langle \epsilon_m, \psi_i \rangle, \ g_m(u_l) = \langle \epsilon_m, g(u_l, \cdot) \rangle \) and \( f_m = \langle \epsilon_m, f \rangle \) be the discrete Fourier coefficients of the \( \mathbb{R}^N \) vectors \( h(u_l, t_i), y(u_l, t_i), \psi_i \) and \( f(t_i), i = 1, 2, \ldots, N, \) respectively. Then, applying the discrete Fourier transform to (3.1), one obtains, for any \( u_l \in U_l, l = 1, 2, \ldots, M, \)
\[ h_m(u_l) = g_m(u_l)f_m \]  
and
\[ y_m(u_l) = g_m(u_l)f_m + N^{-1/2}z_m. \]

Multiplying both sides of (3.3) by \( N^{-2d}g_m(u_l), \) and adding them together, we obtain the following estimator of \( f_m: \)
\[ \hat{f}_m = \left( \frac{M}{l=1} N^{-2d} \left( g_m(u_l) y_m(u_l) \right) \right) / \left( \frac{M}{l=1} N^{-2d} |g_m(u_l)|^2 \right). \]  

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 or Mallat, 1999) and obtain a periodized version of Meyer wavelet basis 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
\[ f(t) = \sum_{k=0}^{2^{j_0} - 1} a_{j_0 k} \varphi_{j_0 k}(t) + \sum_{j = j_0}^{\infty} \sum_{k=0}^{2^j - 1} b_{j k} \psi_{j k}(t), \ t \in T. \]  

Furthermore, by Plancherel’s formula, the scaling coefficients, \( a_{j_0 k} = \langle f, \varphi_{j_0 k} \rangle, \) and the wavelet coefficients, \( b_{j k} = \langle f, \psi_{j k} \rangle, \) of \( f(\cdot) \) can be represented as
\[ a_{j_0 k} = \sum_{m \in C_{j_0}} f_m \overline{\varphi_{j_0 k}}, \quad b_{j k} = \sum_{m \in C_j} f_m \overline{\psi_{j k}}, \]  
where \( \varphi_{j_0 k} = \langle \epsilon_m, \varphi_{j_0 k} \rangle, \ C_{j_0} = \{ m : \varphi_{j_0 k} \neq 0 \}, \) \( \psi_{j k} = \langle \epsilon_m, \psi_{j k} \rangle \) and, for any \( j \geq j_0, \)
\[ C_j = \{ m : \psi_{j k} \neq 0 \} \subseteq 2\pi/3 [-2^j + 2, -2^j] \cup [2^j, 2^j + 2]. \]  
(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 \( b_{j k} \) are readily obtained by substituting \( f_m \) in (3.6) with (3.4), i.e.,
\[ a_{j_0 k} = \sum_{m \in C_{j_0}} \hat{f}_m \overline{\varphi_{j_0 k}}, \quad b_{j k} = \sum_{m \in C_j} \hat{f}_m \overline{\psi_{j k}}. \]

We now construct a (block thresholding) wavelet estimator of \( f(\cdot), \) suggested by Pensky and Sapatinas (2009, 2010). For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length \( \ln n. \) Let \( A_r \) and \( U_r \) be the following sets of indices:
\[ A_r = \{ r | r = 1, 2, \ldots, [2^r / \ln n] \}, \]
\[ U_r = \{ k | k = 0, 1, \ldots, 2^r - 1 ; (r-1) \ln n \leq k \leq r \ln n - 1 \}. \]

Denote
\[ B_{jr} = \sum_{k \in U_r} b_{jk}^2, \quad \hat{B}_{jr} = \sum_{k \in U_r} \hat{b}_{jk}^2. \]  

(3.8)
Finally, for any $j_0 \geq 0$, the (block thresholding) wavelet estimator $\hat{f}_n(\cdot)$ of $f(\cdot)$ is constructed as

$$
\hat{f}_n(t) = \sum_{k=0}^{2^n-1} \hat{a}_{jk} \psi_{jk}(t) + \sum_{j=j_0}^{j} \sum_{k \in U_j} \hat{b}_{jk} I(|\hat{b}_{jk}| \geq \lambda_j) \nu_j(t), \quad t \in T,
$$

(3.9)

where $I(A)$ is the indicator function of the set $A$, and the resolution levels $j_0$ and $j$ and the thresholds $\lambda_j$ will be defined in Section 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.

4. Minimax lower bounds for the $L^2$-risk

Denote

$$
s' = s + 1/2 - 1/p, \quad s^* = s + 1/2 - 1/p', \quad p' = \min\{p, 2\}.
$$

(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 $a_{jk}$ and $b_{jk}$ defined in (3.6) satisfy the following relation:

$$
B^s_{p,q}(A) = \left\{ f \in L^2(U) : \left( \sum_{k=0}^{2^n-1} |a_{jk}|^p \right)^{1/p} + \left( \sum_{j=0}^{\infty} 2^{j(q_p - q)} \left( \sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \right\}.
$$

(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

$$
R_0(V) = \inf_{\hat{f}} \sup_{f \in V} \|\hat{f} - f\|^2,
$$

where $\|g\|$ is the $L^2$-norm of a function $g(\cdot)$ and the infimum is taken over all possible estimators $\hat{f}(\cdot)$ (measurable functions taking their values in a set containing $V$) of $f(\cdot)$, based on observations from model (1.1)).

For $M = M_n$ and $N = n/M_n$, denote

$$
\tau_\kappa(m, n) = M^{-1} \sum_{l=1}^{M} N^{-2n/|g_m(u_l)|^{2s}}, \quad \kappa = 1 \text{ or } 2 \text{ or } 4,
$$

(4.3)

and

$$
\Delta_\kappa(j, n) = |C_j|^{-1} \sum_{m \in C_j} \tau_\kappa(m, n)[r_1(m, n)]^{-2s}, \quad \kappa = 1 \text{ or } 2.
$$

(4.4)

The expression $r_1(m, n)$ appears in both the lower and the upper bounds for the $L^2$-risk and contains the dependence parameters $d_l$, $l = 1, 2, \ldots, M$. Hence, we impose the following assumption:

**Assumption A2.** For some constants $\nu_1, \nu_2, \theta_1, \theta_2 \in \mathbb{R}$, $\alpha_1, \alpha_2 \geq 0$ (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

$$
K_3 \varepsilon_n |m|^{-2s}\ln |m| - \theta_1 e^{-\alpha_1 |m|^\beta} \leq \tau_1(m, n) \leq K_4 \varepsilon_n |m|^{-2s}\ln |m| - \theta_2 e^{-\alpha_2 |m|^\beta},
$$

(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 $\varepsilon_n$ in (4.5) is such that

$$
n^* = n \varepsilon_n \rightarrow \infty \quad (n \rightarrow \infty).
$$

(4.6)

Since we expect estimator (3.9) to adapt to the case of sub-Gaussian errors and since Gaussian random variables is a particular case of sub-Gaussian ones, it is sufficient to derive lower bounds in the Gaussian case.

**Theorem 1.** Let Assumptions A0G, A1 and A2 hold. Let $(\psi_{j_k}, \psi_{j_k})$ be the periodic Meyer wavelet basis discussed in Section 3. Let $s > \max(0, 1/p - 1/2)$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Then, as $n \rightarrow \infty$,

$$
R_0(B^s_{p,q}(A)) \geq \begin{cases} 
C(n^*)^{-2s/(2s + 2p + 1)}(\ln n^*)^{-2s/(2s + 2p + 1)} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) < ps^*, \\
C(\ln n^*)^{-2s/(2s + 2p)}(\ln n^*)^{-2s/(2s + 2p)} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) \geq ps^*,
\end{cases}
$$

(4.7)

$$
C(\ln n^*)^{-2s/\beta} & \text{if } \alpha_1 \alpha_2 \neq 0.
\end{cases}
$$
5. Minimax upper bounds for the $L^2$-risk: Gaussian case

In this section, we shall assume that random variables $\eta_l$, for every $l = 1, 2, ..., M$, and $i = 1, 2, ..., N$, in (1.2) are Gaussian, that is, Assumption A0G holds.

Let $f_n(\cdot)$ be the (block thresholding) wavelet estimator defined by (3.9). Choose now $j_0$ and $J$ such that

$$2^k = \ln n^*, \quad 2^j = (n^*)^{1/(2\alpha + 1)} \text{ if } \alpha_1 = \alpha_2 = 0,$$

$$2^k = \frac{3}{8\pi} \left( \frac{\ln n^*}{2\alpha} \right)^{1/\beta}, \quad 2^j = 2^k \text{ if } \alpha_1 \alpha_2 > 0. \tag{5.1}$$

Set, for some constant $\mu > 0$, large enough,

$$\lambda_j = \mu^2 (n^*)^{-1} \ln n^* 2^{2k^* j^*} \text{ if } \alpha_1 = \alpha_2 = 0. \tag{5.3}$$

Since $j_0 > J - 1$ when $\alpha_1 \alpha_2 > 0$, the estimator (3.9) only consists of the first (linear) part and, hence, $\lambda_j$ does not need to be selected in this case.) Note that the choices of $j_0$, $J$ and $\lambda_j$ are independent of the parameters, $s$, $p$, $q$ and $A$ of the Besov ball $B_{p,q}^S(A)$; hence, the estimator (3.9) is adaptive with respect to these parameters.

Denote $(\chi)_+ = \max(0, \chi)$,

$$q = \begin{cases} (2\epsilon + 1)(2 - p) \overline{p}(25 + 2\epsilon + 1) & \text{if } \epsilon(2 - p) < ps^*, \\ (q - P)_+ \overline{q} & \text{if } \epsilon(2 - p) = ps^*, \\ 0 & \text{if } \epsilon(2 - p) > ps^*. \end{cases} \tag{5.4}$$

Assume that, in the case of $\alpha_1 = \alpha_2 = 0$, the sequence $e_n$ is such that

$$-h_1 \ln n \leq \ln e_n \leq h_2 \ln n$$

for some constants $h_1, h_2 \in (0, 1)$. Observe that condition (5.5) implies (4.6) and that $\ln n^* \approx \ln n$ ($n \to \infty$). (Here, and in what follows, $u(v) = o(v)$ means that there exist constants $C_s, C_2$ ($0 < C_1 \leq C_2 < \infty$), independent of $n$, such that $0 < C_1 v(n) \leq u(n) \leq C_2 v(n) < \infty$ for $n$ large enough.)

Direct calculations yield that under Assumptions A1, A2 and (5.5), for some constants $c_1 > 0$ and $c_2 > 0$, independent of $n$, for $\Delta_1(j, n)$ defined in (4.4), one has

$$\Delta_1(j, n) \leq \begin{cases} c_1 e_n^{-1} 2^{2k^* j^*} & \text{if } \alpha_1 = \alpha_2 = 0, \\ c_2 e_n^{-1} 2^{2k^* j^*} \exp \left( \alpha_1 \left( \frac{8\pi}{3} \right)^{1/\beta} \right) & \text{if } \alpha_1 \alpha_2 > 0. \end{cases} \tag{5.6}$$

The proof of the minimax upper bounds for the $L^2$-risk is based on the following two lemmas.

**Lemma 2.** Let Assumptions A0G, A1 and A2 hold. Let the estimators $\hat{a}_{j,k}$ and $\hat{b}_{j,k}$ of the scaling and wavelet coefficients $a_{j,k}$ and $b_{j,k}$, respectively, be given by (3.6) with $f_m$ defined by (3.4). Then, for all $j \geq j_0$,

$$\mathbb{E} |\hat{a}_{j,k} - a_{j,k}|^2 \leq C_n^{-1} \Delta_1(j_0, n) \quad \text{and} \quad \mathbb{E} |\hat{b}_{j,k} - b_{j,k}|^2 \leq C_n^{-1} \Delta_1(j, n). \tag{5.7}$$

If $\alpha_1 = \alpha_2 = 0$ and (5.5) holds, then, for any $j \geq j_0$,

$$\mathbb{E} |\hat{b}_{j,k} - b_{j,k}|^2 \leq C_n^{-1} (\ln n)^{3\eta^*(n^*)^{3/(2\alpha + 1)}}. \tag{5.8}$$

**Lemma 3.** Let Assumptions A0G, A1 and A2 hold. Let the estimators $\hat{b}_{j,k}$ of the wavelet coefficients $b_{j,k}$ be given by (3.6) with $f_m$ defined by (3.4). Let

$$\mu \geq \frac{2}{\sqrt{1 - h_1^2}} \left[ \sqrt{c_1 + \frac{8\pi}{\sqrt{K_3} \ln 2}} \left( \frac{2\pi}{3} \right)^{1/\beta} \right], \tag{5.9}$$

where $c_1, K_3$ and $h_1$ are defined in (5.6), (4.5) and (5.5), respectively. Then, for all $j \geq j_0$ and any $\kappa > 0$,

$$\mathbb{P} \left( \sum_{j \in U_p} |\hat{b}_{j,k} - b_{j,k}|^2 \geq 0.25 \mu^2 (n^*)^{-1} 2^{2k^* j^*} \ln n^* \right) \leq n^{-\kappa}. \tag{5.10}$$

Under Assumptions A0G, A1 and A2, and using Lemmas 2 and 3, the following statement is true.

**Theorem 2.** Let Assumptions A0G, A1 and A2 hold. Let $f_{n,\cdot}(\cdot)$ be the wavelet estimator defined by (3.9), with $j_0$ and $J$ given by (5.1) (if $\alpha_1 = \alpha_2 = 0$) or (5.2) (if $\alpha_1 \alpha_2 > 0$) and $\mu$ satisfying (5.9) with $\kappa = 5$. Let $s > 1/p$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Then, under
(4.6) if \(\alpha_1\alpha_2 > 0\) or (5.5) if \(\alpha_1 = \alpha_2 = 0\), as \(n \to \infty\),

\[
\sup_{f \in B_{p,q}(A)} \| f_n - f \|^2 \leq \left\{ \begin{array}{ll}
C(n^*)^{-2/(2s + 2 + 1)}(\ln n)^{\beta_1/(2s + 2 + 1)} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) < ps^*, \\
\frac{(\ln n)^{\beta_1/(2s + 2 + 2s^*)}}{C(n^*)} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) \geq ps^*, \\
\frac{(\ln n^*)^{-2\beta}}{C(n^*)} & \text{if } \alpha_1 \alpha_2 > 0.
\end{array} \right.
\] (5.11)

**Remark 2.** Theorems 1 and 2 imply that, for the \(L^2\)-risk, the wavelet estimator \(\hat{f}_m(\cdot)\) defined by (3.9) is asymptotical optimal (in the minimax sense), or near optimal within a logarithmic factor, over a wide range of Besov balls \(B_{p,q}(A)\) 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 \(\nu, \beta_1, \beta_2\) (of the blurring function \(g(\cdot, \cdot)\)), the long memory parameters \(d_i\), \(l = 1, 2, \ldots, M\) (of the error sequence \(\xi(i)\)), and how the total number of observations \(n\) is distributed among the total number of channels \(M\). In particular, \(M\) and \(d_i\), \(l = 1, 2, \ldots, M\), jointly determine the value of \(\epsilon_n\), which, in turn, defines the “essential” convergence rate \(n^* = n c_2\) which may differ considerably from \(n\). For example, if \(M = M_n = n^0, 0 \leq \theta < 1\) and \(|g_m(u)|^2 \approx |m|^{-2\theta}\) for every \(l = 1, 2, \ldots, M\), then

\[
\epsilon_n = M^{-1} \sum_{l=1}^{M} N^{-2d_i}.
\] (5.12)

and, therefore, \(n^1 - 2d^*(1 - \theta) \leq n^* \leq n\), where \(d^* = \max_{1 \leq l \leq M} d_i\), so that \(n^*\) can take any value between \(n^1 - 2d^*(1 - \theta)\) and \(n\). This is further illustrated in Section 7 below.

### 6. Minimax upper bounds for the \(L^2\)–risk: sub-Gaussian case

In this section, we shall assume that random variables \(\eta_{l,n}\), for every \(l = 1, 2, \ldots, M\) and \(i = 1, 2, \ldots, n\), in (1.2) are sub-Gaussian, that is, the general version of Assumption A0 holds. Indeed, by slightly modifying the threshold, one can adapt the estimator (3.9) to the case of sub-Gaussian noise.

Let \(J\) and \(J_0\) be defined in (5.1) or (5.2) and \(\varrho\) be defined in (5.4). Assume that, in the case of \(\alpha_1 = \alpha_2 = 0\), sequence \(\epsilon_n\) satisfies condition (5.5). For some constant \(\mu > 0\), large enough, choose

\[
\lambda_j = 4c_1(1 + \mu^2 \ln n)(n^*)^{-1} \ln(n)2^{2j}J^\varrho
\] (6.1)

where \(c_1\) is defined in (5.6). Note that, similar to the case of Gaussian errors, estimator (3.9) is adaptive with respect to parameters of the Besov space where it belongs as well to sub-Gaussian noise without the knowledge of its exact distribution.

The proof of the minimax upper bounds for the \(L^2\)-risk in sub-Gaussian case is based on the following two lemmas. To state it, for any matrix \(G\), let \(\|G\|_{\infty}\) and \(\|G\|_2\) be, respectively, the spectral and the Frobenius norms.

**Lemma 4.** (The matrix version of the Hanson–Wright inequality, Rudelson and Vershynin, 2013). Let \(X = (X_1, \ldots, X_n)\) be a random vector with independent components such that \(E[X_i] = 0, \|X_i\|_2 \leq K\). Then, for any matrix \(B\), and some absolute constant \(c_0 > 0\), one has

\[
P\left(\left\|X^\top BX - E[X^\top BX]\right\| > t\right) \leq 2 \exp\left(-c_0 \min\left\{ \frac{t^2}{K^2\|B\|_2^2}, \frac{t}{K^2\|B\|_{\infty}} \right\}\right).
\] (6.2)

**Lemma 5.** Let Assumptions A0, A1, A2 and (5.5) hold. Let the estimators \(\tilde{b}_j\) of the wavelet coefficients \(b_j\) be given by (3.6) with \(\hat{f}_m\) defined by (3.4). Then, for all \(j \geq J_0\), (5.7) holds. Moreover, if \(\alpha_1 = \alpha_2 = 0\) and (5.5) holds, then, for any \(j \geq J_0\),

\[
E[|\tilde{b}_j - b_j|^4] \leq Cn^3(n^*)^{-2}.
\] (6.3)

In addition, for all \(j \geq J_0\) and any \(\kappa > 0\),

\[
P\left(\sum_{k \in U_j} |\tilde{b}_j - b_j|^2 > c_1(1 + \mu^2 \ln n)(n^*)^{-1} \ln n 2^{2j}J^\varrho\right) \leq 2 n^{-\kappa}.
\] (6.4)

provided

\[
\mu \geq K \sqrt{c_0 \kappa},
\] (6.5)

where \(c_0\) and \(c_1\) are defined in (6.2) and (5.6), respectively.

**Lemma 5** implies the following version of the upper bounds for quadratic risk in the case of sub-Gaussian errors.

**Theorem 3.** Let Assumptions A0, A1 and A2 hold. Let \(\hat{f}_m(\cdot)\) be the wavelet estimator defined by (3.9), with \(j_0\) and \(f\) given by (5.1) (if \(\alpha_1 = \alpha_2 = 0\)) or (5.2) (if \(\alpha_1 \alpha_2 > 0\)) and \(\mu\) satisfying (6.5) with \(\kappa = 5\). Let \(s > 1/p', 1 \leq p \leq \infty, 1 \leq q \leq \infty\) and \(A > 0\). Then, under
(4.6) if \( \alpha_1 \alpha_2 > 0 \) or (5.5) if \( \alpha_1 = \alpha_2 = 0 \), as \( n \to \infty \),

\[
\sup_{f \in B^s_p(A)} \mathbb{E} \| f - \hat{f} \|^2 \leq \begin{cases} C (n^s)^{2s/(2s+2s+1)} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) < ps^*, \\ C \left( \frac{\ln n}{n^s} \right)^{2s/(2s+2s)} \left( \ln n \right)^{1+\frac{2s\theta_1}{2s+2s}} & \text{if } \alpha_1 = \alpha_2 = 0, \nu(2-p) \geq ps^*, \\ C (\ln n^s)^{-2s/\beta} & \text{if } \alpha_1 \alpha_2 > 0.
\end{cases}
\] (6.6)

7. 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 (2.4) holds true and that there exist \( K_6, \theta_1 \) and \( \theta_2 \), such that \( M = M_n \) satisfies

\[
K_5 n^{\theta_1} \leq M \leq K_6 n^{\theta_2}, \quad 0 \leq \theta_1 < \theta_2 < 1, \quad 0 < K_5 \leq K_6 < \infty.
\] (7.1)

(Not that, under (7.1), \( K_5 n^{1-\theta_2} \leq N \leq K_6 n^{1-\theta_1} \).)

Example 1. Consider the case when \( g_m(\cdot), m = 0, \pm 1, \pm 2, \ldots \), is of the form

\[
g_m(u) = C g \exp(-K|m|^\beta q(u)), \quad u \in U,
\] (7.2)

where \( q(\cdot) \) in (7.2) is such that, for some \( q_1 \) and \( q_2 \),

\[
0 < q_1 \leq q(u) \leq q_2 < \infty, \quad u \in U.
\] (7.3)

This setup 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 Pensky and Sapatinas, 2009, 2010, Examples 1 and 2). In the former case, \( g_m(u) = \exp(-4\pi^2 m^2 u), u \in U \), so that \( K = 4\pi^2, \quad \beta = 2 \), \( q(u) = u, \quad q_1 = 0 \) and \( q_2 = 1 \). In the latter case, \( g_m(u) = C u^{m_1} = C \exp(-|m|\ln(1/u)), \quad 0 < r_1 \leq u \leq r_2 < 1 \), so that \( K = 1, \quad \beta = 1, \quad q(u) = \ln(1/u), \quad q_1 = \ln(1/r_2) \) and \( q_2 = \ln(1/r_1) \).

It is easy to see that, under conditions (7.2) and (7.3), for \( r_1(m,n) \) given in (4.3),

\[
r_1(m,n) \leq C \epsilon_n \exp(-2Kq_1|m|^\beta) \quad \text{and} \quad r_1(m,n) \geq C \epsilon_n \exp(-2Kq_2|m|^\beta),
\]

where \( \epsilon_n \) is of the form (5.12). Assumptions (2.4) and (7.1) lead to the following bounds for \( n^* \):

\[
K_5 n^{1-2d(1-\theta)} \leq n^* \leq n,
\]

so that \( \ln n^* \leq \ln n^* \). Therefore, according to Theorems 1 and 2,

\[
R_n(B_{\rho_3}(A)) = (\ln n)^{-2s/\beta}.
\] (7.4)

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 \( \beta \) (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 suggested wavelet estimator. A similar effect is observed in the case of kernel smoothing, see Section 2.2 in Kulik (2008).

Example 2. Suppose that the blurring function \( g(\cdot, \cdot) \) is of a box-car like kernel, i.e.,

\[
g(u, t) = 0.5q(u)|(|t| < u), \quad u \in U, \quad t \in T,
\] (7.5)

where \( q(\cdot) \) is some positive function which satisfies conditions (7.3). In this case, the functional Fourier coefficients \( g_m(\cdot) \) are of the form

\[
g_0(u) = 1 \quad \text{and} \quad g_m(u) = (2\pi m)^{-1} \gamma(u) \sin(2\pi mu), \quad m \in \mathbb{Z}, \quad u \in U.
\] (7.6)

It is easy to see that estimation of the initial speed of a wave on a finite interval (see Pensky and Sapatinas, 2009, Example 4 or Pensky and Sapatinas, 2010, Example 3) leads to \( g_m(\cdot) \) of the form (7.6) 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

\[
d_l = a_l u_l + a_2, \quad 0 \leq a_2 \leq d^* < 1/2, \quad 0 \leq a_1 + a_2 \leq d^* < 1/2,
\] (7.7)

i.e., condition (2.4) holds. Note that if \( a_1 = 0 \), then

\[
r_1(m,n) \approx M^{-1} N^{-2a_2(4\pi^2 m^2)} -1 \sum_{l=1}^{M} \sin^2(2\pi m l/M).
\]
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 (3.9) is such that $2^b > (\ln n)^\theta$ for some $\delta > 0$ and $M \geq (32\pi/3)n^{1/3}$, then, for $n$ and $|m|$ large enough,

$$\tau_1(m,n) \approx N^{-2a_1} m^{-2}.$$ 

Assume now, without loss of generality, that $a_1 \geq 0$. (Note that the case of $a_1 \leq 0$ 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 6.** Let $g(\cdot, \cdot)$ be of the form (7.5), where $q(\cdot)$ is some positive function which satisfies (7.3), and let $d_1, l = 1, 2, \ldots, M$, be given by (7.7) with $a_1 \geq 0$. Assume (without loss of generality) that $U = [0, 1]$, and consider $u_1 = l/M, l = 1, 2, \ldots, M$. Let $M = M_n$ satisfy (7.1) with $\theta_1 > 0$ if $a_1 > 0$ and $M \geq (32\pi/3)n^{1/3}$ if $a_1 = 0$. If $m \in A$, where $|A| = C_m 2^l$, for some absolute constant $C_m > 0$, if $j \geq j_0 > 0$, where $j_0$ is such that $2^b \geq C_0 \ln n$ for some $C_0 > 0$, then, for $n$ and $|m|$ large enough,

$$\tau_1(m,n) \approx N^{-2a_1} m^{-2}(\log n)^{-1}. \quad (7.8)$$

It follows immediately from Lemma 6 that, if $M = M_n \approx n^\theta, \quad 0 < \theta < 1,$

then Assumption A2 holds with $a_1 = a_2 = 0, \nu_1 = \nu_2 = \nu = 2, \epsilon_n = n^{-2a_1(1-\theta)(\ln n)^{-1}}$ and $\theta_1 = \theta_2 = 0$. Note that $\epsilon_n$ satisfies conditions (4.6) and (5.5), so that $\ln n s \ln n^s$. Therefore, according to Theorems 1 and 2,

$$R_n \left( B_{p,q}^{\theta}(A) \right) \begin{cases} C(n^* - 2^{\nu}(2s + 5)) & \text{if } 4 - 2p < s^*, \\ C\left( \frac{\ln n^*}{n^*} \right)^{(p/s + 2)} & \text{if } 4 - 2p \geq s^*, \end{cases} \quad (7.9)$$

and

$$\sup_{f \in \mathcal{A}} \frac{\|f_n - f\|^2}{\|f\|^2} \leq \begin{cases} C(n^* - 2^{\nu}(2s + 5)(\ln n)^\theta) & \text{if } 4 - 2p < s^*, \\ C\left( \frac{\ln n}{n^*} \right)^{(p/s + 2)} (\ln n)^\theta & \text{if } 4 - 2p \geq s^*, \end{cases} \quad (7.10)$$

where $n^* = n^{1 - 2a_1(1-\theta)(\ln n)^{-1}}$ and

$$q = \begin{cases} \frac{5(2-p)\nu}{p(2s+5)} & \text{if } 4 - 2p < s^*, \\ \frac{(q-p)\nu}{q} & \text{if } 4 - 2p = s^*, \\ 0 & \text{if } 4 - 2p > s^*. \end{cases}$$

Note that LRD affects the convergence rates in this case via the parameter $a_2$ that appears in the definition (7.7).

8. Discussion

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 et al., 1997; Harsdörf and 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 (1.1) considered in this work.

The multichannel deconvolution model (1.1) or 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 and 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 interval domain were studied in Golubev and Khamsinskii (1999); the latter problem was also discussed in Golubev (2004). Some of these specific models were considered in Section 7.

The multichannel deconvolution model (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 work was to study the multichannel deconvolution model (1.1) from a minimax point of view, with the relaxation that errors exhibit LRD, covering also both Gaussian and sub-Gaussian cases. 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 largest eigenvalues of their covariance matrices. In particular, minimax lower bounds for the $L^2$-risk in model (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 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, for both Gaussian and sub-Gaussian errors. The convergence rates of the resulting estimators depend on the balance between the smoothness parameter (of the response function or of the blurring function $g(\cdot, \cdot)$), and the long memory parameters $\theta_1, \theta_2, \ldots, \theta_r$ (of the error sequence $u_\ell$), 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 $d_r=0, l=1, 2, \ldots, M$. 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).

Under the assumption that 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 (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, 2006). 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 in the estimation of the initial speed of a wave on a finite interval). Example 2 showed that a similar result is still possible under long-range dependence with (equispaced channels) $u_l = l/M, l = 1, 2, \ldots, M$, $n^\alpha \leq M = M_n \leq n^{\alpha+1}$, for some $0 \leq \alpha \leq 1/2$ such that $d_1 = d_2 = \ldots = d_\alpha, l = 1, 2, \ldots, M$, $0 \leq \alpha_2 < 1/2$, $0 \leq \alpha_1 + \alpha_2 < 1/2$.

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., $M = M_n = o(n^\alpha)$ 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 $M$ is growing. 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 (1.1)) under long range-dependence (e.g., modeling the errors as fractional Brownian motions) 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.

Acknowledgments

Mariani Pensky was supported in part by National Science Foundation (NSF), Grant DMS-1106564. The authors would like to thank the anonymous reviewer for useful comments and suggestions on improvements to this paper.

Appendix A. Proofs

A.1. Proofs of the statements in Section 2

Proof of Lemma 1. We prove the upper bound only since the proof of the lower bound is similar. By (2.1) and (2.2), and the definitions of $\mathcal{H}_N$ and $\mathcal{F}_d$, we have

$$\lambda_{\max}(\Sigma) \leq C_{\max} \sup_{h \in \mathcal{H}_N} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda = 2C_{\max} \sup_{h \in \mathcal{F}_d} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda.$$ 

Now, we split $\int_{-\pi}^{\pi} = \int_{-\pi}^{\pi/N} + \int_{\pi/N}^{\pi}$. Since $d < 1/2$, for the first integral, we have

$$\int_{-\pi}^{\pi/N} |h(\lambda)|^{-2d} \, d\lambda \leq N \int_{-\pi}^{\pi/N} \lambda^{-2d} \, d\lambda = N \frac{1}{1-2d} \left(\frac{\pi}{N}\right)^{-2d+1} = \frac{\pi^{-2d+1}}{1-2d} N^{2d}.$$ 

Proof of Lemma 2. We prove the upper bound only since the proof of the lower bound is similar. By (2.1) and (2.2), and the definitions of $\mathcal{H}_N$ and $\mathcal{F}_d$, we have

$$\lambda_{\max}(\Sigma) \leq C_{\max} \sup_{h \in \mathcal{H}_N} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda = 2C_{\max} \sup_{h \in \mathcal{F}_d} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda.$$ 

Now, we split $\int_{-\pi}^{\pi} = \int_{-\pi}^{\pi/N} + \int_{\pi/N}^{\pi}$. Since $d < 1/2$, for the first integral, we have

$$\int_{-\pi}^{\pi/N} |h(\lambda)|^{-2d} \, d\lambda \leq N \int_{-\pi}^{\pi/N} \lambda^{-2d} \, d\lambda = N \frac{1}{1-2d} \left(\frac{\pi}{N}\right)^{-2d+1} = \frac{\pi^{-2d+1}}{1-2d} N^{2d}.$$ 

Proof of Lemma 3. We prove the upper bound only since the proof of the lower bound is similar. By (2.1) and (2.2), and the definitions of $\mathcal{H}_N$ and $\mathcal{F}_d$, we have

$$\lambda_{\max}(\Sigma) \leq C_{\max} \sup_{h \in \mathcal{H}_N} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda = 2C_{\max} \sup_{h \in \mathcal{F}_d} \int_{-\pi}^{\pi} |h(\lambda)|^{-2d} \, d\lambda.$$ 

Now, we split $\int_{-\pi}^{\pi} = \int_{-\pi}^{\pi/N} + \int_{\pi/N}^{\pi}$. Since $d < 1/2$, for the first integral, we have

$$\int_{-\pi}^{\pi/N} |h(\lambda)|^{-2d} \, d\lambda \leq N \int_{-\pi}^{\pi/N} \lambda^{-2d} \, d\lambda = N \frac{1}{1-2d} \left(\frac{\pi}{N}\right)^{-2d+1} = \frac{\pi^{-2d+1}}{1-2d} N^{2d}.$$
For the second integral, since \( d \geq 0 \), we have

\[
\int_{\pi/N}^{\pi} h(\lambda)|\lambda|^{-2d} \, d\lambda \leq \left( \frac{\pi}{N} \right)^{-2d} \int_{\pi/N}^{\pi} h(\lambda) \, d\lambda \leq \left( \frac{\pi}{N} \right)^{-2d} \int_{0}^{\pi} h(\lambda) \, d\lambda \leq \pi(2\pi)^{-2d}N^{2d}.
\]

This completes the proof of the lemma. \( \square \)

### A.2. Proof of the minimax lower bounds for the \( L_2 \)-risk

In order to prove Theorem 1, 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.

The proof of Theorem 1 is based on Lemma A.1 of Bunea et al. (2007), an easy corollary of the Fano lemma, which we reformulate here for completeness for the case of the \( L_2 \)-risk. [Note that the proof of the corresponding lower bound in Pensky and Sapatinas, 2009, 2010, in the case of independent and identically distributed Gaussian errors, uses a different but similar lemma (see Härdle et al., 1998, Lemma 10.1).]

**Lemma 7 (Bunea et al., 2007, Lemma A.1).** Let \( \theta \) be a set of functions of cardinality \( \text{card}(\theta) \geq 2 \), such that

1. \( \|f - g\|^2 \geq 4\sigma^2 > 0 \) for \( f, g \in \Theta, f \neq g \).
2. the Kullback divergences \( K(P_f, P_g) \) between the measures \( P_f \) and \( P_g \) satisfy the inequality \( K(P_f, P_g) \leq \log(\text{card}(\Theta))/16 \) for \( f, g \in \Theta \).

Then, for some absolute constant \( C > 0 \), one has

\[
\inf_{f_n} \sup_{f \in \Theta} E_f \|T_n - f\|^2 \geq C\sigma^2,
\]

where \( \inf_{T_n} \) denotes the infimum over all estimators.

The dense case: Let \( \omega \) be the \( 2^j \)-dimensional vector with components \( \omega_k = 0 \) or 1. Denote the set of all possible vectors \( \omega \) by \( \Omega : \Omega = \{0, 1\}^j \), the set of binary sequences of length \( 2^j \). Note that the vector \( \omega \) has \( N = 2^j \) entries and, hence, \( \text{card}(\Omega) = 2^N \).

Let \( H(\omega, \tilde{\omega}) = \sum_{k=1}^{2^j} (\omega_k \neq \tilde{\omega}_k) \) be the Hamming distance between the binary sequences \( \omega \) and \( \tilde{\omega} \). Then, the Varshamov–Gilbert Lemma (see, e.g., Tsybakov, 2008, p. 104) states that one can choose a subset \( \Omega_1 \) of \( \Omega \), of cardinality at least \( 2^{N/8} \), such that \( H(\omega, \tilde{\omega}) \geq N/8 \) for any \( \omega, \tilde{\omega} \in \Omega_1 \).

Let \( \Theta = \{ \omega : \omega \in \Omega_1 \} \). Consider two arbitrary sequences \( \omega, \tilde{\omega} \in \Omega_1 \) and the functions \( f_\omega \) and \( \tilde{f}_\tilde{\omega} \) given by

\[
f_\omega(t) = \rho_j \sum_{k=0}^{2^j-1} \omega_k \psi_{jk}(t), \quad \tilde{f}_\tilde{\omega}(t) = \rho_j \sum_{k=0}^{2^j-1} \tilde{\omega}_k \psi_{jk}(t), \quad t \in T.
\]

Choose \( \rho_j = A2^{-j^2 s + 1/2} \), so that \( f_\omega, \tilde{f}_\tilde{\omega} \in B_{2q}(A) \). Then, calculating the \( L_2 \)-norm difference of \( f_\omega \) and \( \tilde{f}_\tilde{\omega} \), we obtain

\[
\|f_\omega - \tilde{f}_\tilde{\omega}\|^2 = \rho_j^2 \left( \sum_{k=0}^{2^j-1} (\omega_k - \tilde{\omega}_k) \psi_{jk}(t) \right)^2 = \rho_j^2 H(\omega, \tilde{\omega}) \geq 21/8.
\]

Hence, we get \( 4\sigma^2 = 2 J_2 / 8 \) in condition (i) of Lemma 7.

In order to apply Lemma 7, one needs to also verify condition (ii). For \( f_\omega \) with \( \omega \in \Omega_1 \), denote by \( h_{\omega} \) and \( h_{\tilde{\omega}} \), the vectors with components, respectively,

\[
h_\omega(u_i, t_i) = g(u_i, t_i) \psi_\omega(\cdot), \quad i = 1, 2, \ldots, N, \\
h_{\tilde{\omega}}(u_i, t_i) = g(u_i, t_i) \psi_{\tilde{\omega}}(\cdot), \quad i = 1, 2, \ldots, N.
\]

Then,

\[
K(P_{f_\omega}, P_{f_{\tilde{\omega}}}) = 0.5 \sum_{m=1}^{M} (h_{\omega} - h_{\tilde{\omega}})^T (\Sigma^{1/2})^{-1} (h_{\omega} - h_{\tilde{\omega}}) \\
\leq 0.5 \sum_{m=1}^{M} \lambda_{\text{max}}((\Sigma^{1/2})^{-1}) \|h_{\omega} - h_{\tilde{\omega}}\|^2.
\]

Now, since \( \omega \) and \( \tilde{\omega} \) are binary vectors, using Plancherel’s formula and the fact that \( |\psi_{jk,m}| \leq 2^{-j/2} \), we derive that, under Assumptions A1 and A2,

\[
K \left( P_{f_\omega}, P_{f_{\tilde{\omega}}} \right) \leq 0.5NM\rho_j^2 \sum_{m \in C} \frac{1}{M} \sum_{i=1}^{M} \|g_m(u_i)\|^2 K_1^{-1}N^{-2d} \\
\leq 2\pi K_1^{-1}N2^{1/2} \Delta_1(j, n) \leq 2\pi K_1^{-1}N2^{-1/2} \Delta_1(j, n),
\]

where \( \Delta_1(j, n) \) is defined by (4.4).
Direct calculations yield that, under Assumptions A1, A2 and (4.5), for some constants \(c_1 > 0\) and \(c_4 > 0\), independent of \(n\),

\[
\Delta_1(j,n) \leq \begin{cases} 
  c_3 \varepsilon_n^{-1} 2^{2 \varepsilon_j} & \text{if } \alpha_1 = \alpha_2 = 0, \\
  c_4 \varepsilon_n^{-1} 2^{2 \epsilon_j} \exp \left\{ \alpha_1 \left( \frac{8 \pi}{3} \right)^{2 \beta} \right\} & \text{if } \alpha_1 \alpha_2 > 0.
\end{cases}
\]  

(A.1)

Apply now Lemma 7 with \(j\) such that

\[ 2 \pi A^2 K_1^{-1} n 2^{-2p} \Delta_1(j,n) \leq 2^l \ln 2 / 16, \]

i.e.,

\[ 2^l \leq \begin{cases} 
  [n^*(\ln n^*)]^{1/(2s + 2v + 1)} & \text{if } \beta = 0, \\
  (\ln n^*)^{1/\beta} & \text{if } \beta > 0,
\end{cases} \]

(A.2)

to obtain

\[ \delta^2 = \begin{cases} 
  [n^*(\ln n^*)]^{1/2s + 1} & \text{if } \beta = 0, \\
  (\ln n^*)^{1-2s/\beta} & \text{if } \beta > 0.
\end{cases} \]

The sparse case: Consider the functions \(f_k()\) of the form \(f_k(t) = \rho_j \psi_k(t)\), \(t \in T\), \(k = 0, 1, ..., 2^l - 1\), and denote

\[ \Theta = \{f_k(t) = \rho_j \psi_k(t) : k = 0, 1, ..., 2^l - 1, f_0 = 0\}. \]

Thus, \(\text{card}(\Theta) = 2^l\). Choose now \(\rho_j = A_j^{-\beta_j}\), so that \(f_k \in B_{\rho_j}(A)\). It is easy to check that, in this case, one has \(4 \delta^2 = \rho_j^2\) in Lemma 7, and that

\[ K(P_{1'}, P_{1'}) \leq 2 \pi A^2 K_1^{-1} n 2^{-2p'} \Delta_1(j,n). \]

With

\[ 2^l \leq \begin{cases} 
  [n^*(\ln n^*)]^{1/2s + 1} & \text{if } \beta = 0, \\
  (\ln n^*)^{1-2s/\beta} & \text{if } \beta > 0,
\end{cases} \]

we then obtain that \(K(P_{1'}, P_{1'}) \leq 2 \pi A^2 K_1^{-1} n 2^{-2p'} \Delta_1(j,n)\) and

\[ \delta^2 = \begin{cases} 
  [n^*(\ln n^*)]^{1/2s + 1} & \text{if } \beta = 0, \\
  (\ln n^*)^{1-2s/\beta} & \text{if } \beta > 0.
\end{cases} \]

(A.3)

Recall that \(s^* = \min\{s, s'\}\). By noting that

\[ 2s/(2s + 2v + 1) \leq 2s^*/(2s^* + 2v) \quad \text{if } s(2 - p) \leq ps^*, \]

we then choose the highest of the lower bounds in (9.2) and (9.3). This completes the proof of the theorem.

A.3. Proof of the minimax upper bounds for the \(L^2\)-risk: Gaussian case

We start with proofs of Lemmas 2 and 3.

Proof of Lemma 2. First, consider model (1.1). Then, using (3.3), (3.4), (3.6) and (3.7), one has

\[ a_{jk} - a_{jk} = \sum_{m \in C_i} (\tilde{f}_m - f_m) \psi_{mk}, \quad b_{jk} - b_{jk} = \sum_{m \in C_j} (\tilde{f}_m - f_m) \psi_{jk}. \]

where

\[ \tilde{f}_m - f_m = \frac{1}{\sqrt{N}} \left( \sum_{i=1}^M N^{-2d} g_m(u_i) z_{im} \right) = \sqrt{N} \left( \sum_{i=1}^M N^{-2d} |g_m(u_i)|^2 \right). \]

(A.5)

Define

\[ v_m = \sum_{i=1}^M N^{-2d} |g_m(u_i)|^2 = M \tau_1(m, n). \]

(A.6)

For \(l = 1, 2, ..., M\), consider vector \(V^{(l)}\) with components

\[ V^{(l)} = N^{-2d} \psi_{mj} \sum_{i=1}^M N^{-2d} |g_m(u_i)|^2 \]

\[ = N^{-2d} \psi_{mj} g_m(u_i) v_m^{-1}. \]

(A.7)
It is easy to see that, due to $|\psi_{mk}| \leq 2^{-1/2}$ and the definition of $C_j$
\[
\|V_0^0\|^2 = N^{-4d_1} \sum_{m \in C_j} |\psi_{mk}|^2 |g_{m}(u_i)|^2 \left[ \sum_{l=1}^{M} N^{-2d_1} |g_{m}(u_l)|^2 \right]^{-2} \\
\leq 4\pi |C_j|^{-1} N^{-2d_1} \sum_{m \in C_j} |g_{m}(u_i)|^2 v_m^{-2}.
\]

Hence,
\[
\|V_0^0\|^2 \leq 4\pi |C_j|^{-1} N^{-2d_1} \sum_{m \in C_j} |g_{m}(u_i)|^2 v_m^{-2}.
\]

Using Assumption A1, since $z_{lm}$ are independent for different $l$'s, we obtain
\[
\mathbb{E}[\tilde{C}_j - b_j]^2 = \frac{1}{N} \sum_{m \in C_j} \mathbb{E} \left[ \sum_{m \in C_j} |\psi_{mk}|^2 |g_{m}(u_i)|^2 \right]^{-2} \\
\leq \frac{1}{N} \sum_{m \in C_j} \mathbb{E} \left[ \sum_{m \in C_j} |g_{m}(u_i)|^2 v_m^{-2} \right] \\
\leq 4\pi K \|C_j\|^{-1} N^{-1} \sum_{m \in C_j} |g_{m}(u_i)|^2 v_m^{-2} \\
= 4\pi K \|C_j\|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-2} = 4\pi K \|C_j\|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-1}.
\]

so that
\[
\mathbb{E}[\tilde{C}_j - b_j]^2 \leq Cn^{-1} |C_j|^{-1} \sum_{m \in C_j} [r_1(m, n)]^{-1} = Cn^{-1} \Delta_1(j, n).
\]

(One can obtain an upper bound for $\mathbb{E}[\tilde{C}_j - a_j]^2$ by the following similar arguments.)

In order to prove (5.8), define
\[
B_{lm} = N^{-2d_1} \left[ \sum_{m \in C_j} N^{-2d_1} |g_{m}(u_i)|^2 \right]^{-1} = N^{-2d_1} v_m^{-1}.
\]

Note that
\[
\mathbb{E}[z_{lm} z_{lm}, z_{lm}, z_{lm}] \leq \left\{ \prod_{l=1}^{4} \mathbb{E}[z_{lm}]^4 \right\}^{1/4}.
\]

Consequently, using Assumption A1, the fact that $z_{lm}$ are independent for different $l$'s, and that $\mathbb{E}[z_{lm}]^4 = 3[\mathbb{E}[z_{lm}]^2]^2$ for standard (complex-valued) Gaussian random variables $z_{lm}$, one obtains
\[
\mathbb{E}[\tilde{C}_j - b_j]^4 = O \left( N^{-2} \sum_{l=1}^{M} B_{lm}^2 \left[ \sum_{m \in C_j} |\psi_{mk}|^2 |g_{m}(u_i)|^2 \mathbb{E}[z_{lm}]^4 \right]^{1/2} \right) \\
+ O \left( N^{-1} \sum_{l=1}^{M} B_{lm}^2 \left[ \sum_{m \in C_j} |\psi_{mk}|^2 |g_{m}(u_i)|^2 \mathbb{E}[z_{lm}]^2 \right] \right) \\
= O \left( N^{-2} \sum_{l=1}^{M} B_{lm}^2 \left[ \sum_{m \in C_j} |\psi_{mk}|^2 |g_{m}(u_i)|^2 \mathbb{E}[z_{lm}]^2 \right] \right) \\
+ O \left( \frac{1}{\Delta_1(j, n)} \right)^2.
\]

Since $\sum_{m \in C_j} \mathbb{E}[z_{lm}]^2 = O(|C_j|)$, one derives
\[
\mathbb{E}[\tilde{C}_j - b_j]^4 = O \left( \sum_{m \in C_j} \left[ \frac{1}{M^2} \tau_2(m, n) \right] + \Delta_2(j, n) \right) \\
= O(M^{-3} \Delta_2(j, n) + n^{-2} \Delta_2(j, n)).
\]

It is straightforward to show that, when $\alpha_1 = \alpha_2 = 0$, one has
\[
\Delta_2(j, n) = O(2^{6b-13} e_n^{-3}).
\]
Thus, using (A.1) and the fact that $2^j \leq 2^{j-1} < (n^*)^{1/(2\nu+1)}$, (A.8) can be rewritten as
\[
\mathbb{E}[\hat{b}_{jk} - b_{jk}]^4 = O(2^{6j}2^{3j_1}2^{3j_2}M^{-3} + 2^{4j}2^{3j_1}2^{3j_2}2^{-2}) = O(n^3(\ln n)^{3j_1}(n^*)^{-3/(2\nu+1)}).
\]
Hence, (5.8) follows. This completes the proof of the lemma. □

**Proof of Lemma 3.** Consider a set of vectors
\[
\Omega_y = \left\{ v_k, k \in U_y : \sum_{k \in U_y} |v_k|^2 \leq 1 \right\}
\]
and a centered Gaussian process
\[
Z_{jk} = \sum_{k \in U_y} v_k (\hat{b}_{jk} - b_{jk}).
\]
Note that, by Jensen’s inequality,
\[
\sup_v \mathbb{E}[Z_{jk}(v)] = \sqrt{\sum_{k \in U_y} |\hat{b}_{jk} - b_{jk}|^2}.
\]
We shall apply below a lemma of Cirelson et al. (1976) which states that, for any $x > 0$,
\[
\Pr \left( \sum_{k \in U_y} |\hat{b}_{jk} - b_{jk}|^2 \geq x + B_1 \right) \leq \exp \left( -\frac{x^2}{2B_2^2} \right),
\]
where
\[
\frac{\mathbb{E}}{\sqrt{\sum_{k \in U_y} |\hat{b}_{jk} - b_{jk}|^2}} \leq \frac{\sqrt{c_1}2^j\ln \sqrt{n^*}}{\sqrt{n^*}} \leq B_1
\]
with $c_1$ defined in (5.6), and $B_2$ is an upper bound for
\[
\sup_{v \in U_y} \text{Var}(Z_{jk}(v)) = \sup_{v \in U_y} \mathbb{E} \left( \sum_{k \in U_y} v_k (\hat{b}_{jk} - b_{jk})^2 \right).
\]
Denote
\[
W_{jm} = \sum_{k \in U_y} v_k w_{jm} \left( \sum_{l=1}^M N^{-2d_l}|g_{ml}(U_l)|^2 \right)^{-1}, \quad m \in \mathcal{C}_j.
\]
Then, under Assumption A2 with $\alpha_1 = \alpha_2 = 0$, using argument similar to the proof of (5.7), one obtains
\[
\sup_{v \in U_y} \text{Var}(Z_{jk}(v)) = \sup_{v \in U_y} \left\{ N^{-1} \sum_{m_1, m_2 \in \mathcal{C}_j} W_{jm_1} W_{jm_2} \mathbb{E} \left[ \sum_{l=1}^M N^{-4d_l}\frac{g_{ml}(U_l)g_{ml}(U_l)}{g_{m_1}(U_l)g_{m_1}(U_l)g_{m_2}(U_l)g_{m_2}(U_l)} Z_{lm_1} Z_{lm_2} \right] \right\}
\leq \sup_{v \in U_y} N^{-1} \sum_{l=1}^M N^{-4d_l} \sum_{m \in \mathcal{C}_j} |W_{jm}|^2 |g_{ml}(U_l)|^2
\leq K_j n^{-1} \sup_{v \in U_y} \left\{ \sum_{m \in \mathcal{C}_j} |W_{jm}|^2 [r_1(m, n)]^{-1} \right\} \leq 4\pi C_3^{2j} 2^j \frac{\ln(n^*)}{n^*} \leq B_2,
\]
where $C_3 = (K_j)^{-1}(\ln 2)^2 (2\pi)^{2\nu}$. Apply now inequality (A.9) with $x = B_1((\mu/\sqrt{1-H_1})/2\sqrt{c_1} - 1)$, in order to obtain large deviation inequality (5.10) provided that (5.9) holds. This completes the proof of the lemma. □

**Proof of Theorem 2.** With (5.6), the proof of this theorem is now almost identical to the proof of Theorem 2 in Pensky and Sapatinas (2010). □

**A.4. Proof of the minimax upper bounds for the $L^2$–risk: sub-Gaussian case**

In this section, we prove Lemma 5. Once the lemma is proved, Theorem 3 will follow from the same arguments that are used in the proof of Theorem 2.

We note first that conclusion (5.7) of Lemma 2 holds, since its proof relies only on the correlation structure of the vector $\xi$. We need to establish upper bounds for the corresponding fourth moment and large deviation inequality.

Recall that for each $l = 1, 2, \ldots, M$, $\xi^{(l)}$ is a vector with components $\xi^{(l)}_i$, $i = 1, 2, \ldots, N$, given by (1.2) with $G^{(l)}G^{(l)^T} = \Sigma^{(l)}$. Then a vector $\eta^{(l)} = (G^{(l)})^{-1} \xi^{(l)}$ has the covariance matrix $I_N$, the identity matrix of size $N$. 

Let $\Phi$ be a matrix of Fourier transforms. Then we define

$$Z^{(l)} = \Phi x^{(l)} = \Phi G^{(l)} \eta^{(l)}, \quad l = 1, \ldots, M.$$  

Let $V^{(l)}$ be a vector in $\mathbb{R}^n$ with entries $v^{(l)}_{m}, l = 1, \ldots, M, m \in \mathcal{C}_l$, defined in (A.7) and let $\nu_m$ be defined in (A.6). Define further vectors $\tilde{Z}$ and $\tilde{V}$ obtained by stacking vectors $Z^{(l)}$ and $V^{(l)}$, respectively, into one long vector:

$$\tilde{Z} = \begin{bmatrix} Z^{(1)} \\ \vdots \\ Z^{(M)} \end{bmatrix}, \quad \tilde{V} = \begin{bmatrix} V^{(1)} \\ \vdots \\ V^{(M)} \end{bmatrix}.$$  

Define block diagonal matrices

$$G = \begin{bmatrix} G^{(1)} & 0 & \ldots & 0 & 0 \\ 0 & G^{(2)} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & G^{(M)} & 0 \end{bmatrix}, \quad \tilde{\Phi} = \begin{bmatrix} \Phi & 0 & \ldots & 0 \\ 0 & \Phi & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Phi \end{bmatrix}.$$  

Since

$$|\tilde{b}_{jk} - b_{jk}|^2 = \frac{1}{N} \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} \tilde{v}_{m_c l} \eta \sum_{m_c} \tilde{v}_{m_c l} \sum_{m_c} \tilde{v}_{m_c l},$$

and using the above notation, we can write $\tilde{Z} = \Phi G \eta$, and calculate

$$|\tilde{b}_{jk} - b_{jk}|^2 = \frac{1}{N} \left[ \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} \tilde{v}_{m_c l} \eta \sum_{m_c} \tilde{v}_{m_c l} \right]^2 = \frac{1}{N} \left( \tilde{v}^T \Phi \eta \right)^2 \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} \tilde{v}_{m_c l} \eta \sum_{m_c} \tilde{v}_{m_c l}.$$  

Note that the quantities in the last line depend on $j$ and $k$, in particular, $\tilde{V} = \tilde{V}(k)$. Define further

$$A = \sum_{k \in U_p} A(k) = \Phi^T \Phi \left( \sum_{k \in U_p} \tilde{V}(k) \tilde{V}(k)^T \right) \Phi G$$  

and

$$\tilde{\beta}_p = \sum_{k \in U_p} |\tilde{b}_{jk} - b_{jk}|^2 = \frac{1}{N} \eta^T A \eta.$$  

Using this notation and bearing in mind block-diagonal structure of the matrices, we can evaluate

$$N \mathbb{E}(\tilde{\beta}_p) = \mathbb{E}(\eta^T A \eta) = \mathbb{E}[\text{Tr}(A \eta \eta^T)] = \text{Tr}(A) = \text{Tr} \left( \sum_{k \in U_p} A(k) \right) = \sum_{k \in U_p} \text{Tr}(G^T \Phi^T \Phi G) = \sum_{k \in U_p} \text{Tr} \left( \Phi^T \Phi \left( \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} \tilde{v}_{m_c l} \eta \sum_{m_c} \tilde{v}_{m_c l} \right) \Phi G \right) = \sum_{k \in U_p} \left\{ \tilde{V}(k)^T \Phi \Sigma \Phi^T \tilde{V}(k) \right\} = \sum_{k \in U_p} \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} \tilde{v}_{m_c l} \eta \sum_{m_c} \tilde{v}_{m_c l}.$$  

For any matrix $G$, recall that $\|G\|_2$ and $\|G\|_F$ denote, respectively, the spectral and the Frobenius norms. Denote

$$D_m = (n^*)^{-1} c_1 2^{2j} n^*, \quad \ln n.$$  

Then, by Assumption A1, $\|\Phi \Sigma^{(l)} \Phi^T \|_2 = \|\Phi \|_2 \Sigma^{(l)} \|_2 \|\Phi^T \|_2 \leq K_2 N^{2d_l}$. Hence,

$$\mathbb{E}(\tilde{\beta}_p) \leq \frac{K^2}{N} \sum_{k \in U_p} \sum_{l=1}^{M} \sum_{m_c} N^{-2d_l} \sum_{l=1}^{N} \sum_{m_c} |\tilde{v}_{m_c l}|^2 \|g_m(u_l)\|^2 \nu_m^2 N^{-2d_l} \leq D_m.$$  

Next, using the definition of $A$, obtain

$$\|A\|_2 \leq \sum_{k \in U_p} \|A_k\|_2$$  

(A.11)

(A.12)
where $A_k = u_k(u_k)\,^T$ with $u_k = G^T \Phi^T \tilde{v}(k)$. Hence, $A_k$ is of rank 1 and, consequently, $\lambda_{\text{max}}(A_k) = \|A_k\|_{sp} = \|\tilde{v}(k)\| \Phi \tilde{v}(k)$. The latter implies that $\|A\|_{sp} \leq \sum_{\ell} u_{\ell} \|\tilde{v}(k)^T \Phi \tilde{v}(k)\|$ and, hence,

$$N^{-1} \|A\|_{sp} \leq E(\tilde{B}_{jr}) \leq D_{jr}. \quad (A.13)$$

Moreover, since $A_k$ are matrices of rank 1, one has

$$\|A\|_2 \leq \sum_{\ell} \|A_k\|_2 = \sum_{\ell} \|A_k\|_{sp}. \quad (A.14)$$

Now, in order to prove (6.3), we compute

$$E[|b_{jk} - \tilde{b}_{jk}|^4] = \left[\frac{1}{N} \sum_{N=1}^{N} A(k(\eta)) \right]^2 \leq \frac{1}{N^2} \|A\|_{sp}^2 \eta^2 \leq \frac{C_2^2 N \eta^2}{(N^2)^2}.$$

Since

$$E(\eta\eta)^2 = E\left( \sum_{l=1}^{M} \sum_{m=1}^{N} \eta_{lm}^2 \right)^2 \leq Cn^2,$$

we derive

$$E[|b_{jk} - \tilde{b}_{jk}|^4] \leq \frac{Cn^2}{(N^2)^2},$$

which implies (6.3).

In order to prove the large deviation inequality (6.4), we use Lemma 4. We apply inequality (6.2) with $t = \mu^2 D_{jr} \log n$, $X = \eta$ and $B = N^{-1}A$ where $A$ and $D_{jr}$ are defined in (9.10) and (9.11), respectively. Taking into account (9.13) and (9.14), we obtain

$$P(|\tilde{B}_{jr}| > D_{jr}(1 + \mu^2 \log n)) \leq 2n^{-k}$$

provided $\mu \geq K \sqrt{c_0}$. A.5. Proofs of the statement in Section 7

**Proof of Lemma 6.** Below we consider only the case of $a_1 > 0$. Validity of the statement for $a_1 = 0$ follows from Pensky and Sapatinas (2010).

By direct calculations, one obtains that

$$\tau_1(m, n) = M^{-1}(4\pi^2 m^2)^{-1} N^{-2a_1} \sum_{l=1}^{M} q^2(l/M) \sin^2(2\pi ml^{-1}) N^{-2a_1 l/M}. \quad (A.15)$$

Therefore,

$$(4\pi^2 m^2)^{-1} q_1^2 N^{-2a_1} S(m, n) \leq \tau_1(m, n) \leq (4\pi^2 m^2)^{-1} q_2^2 N^{-2a_1} S(m, n),$$

where

$$S(m, n) = M^{-1} \sum_{l=1}^{M} \sin^2(2\pi ml^{-1}) N^{-2a_1 l/M}. \quad (A.16)$$

Denote $p = N^{-2a_1/M}$, $x = 4\pi m M^{-1}$ and note that, as $n \to \infty$,

$$p^M = N^{-2a_1} \to 0$$

and

$$p = \exp(-2a_1 M^{-1} \ln N) = 1 - 2a_1 M^{-1} \ln N + 2a_1^2 M^{-2} \ln^2 N + o(M^{-2} \ln^2 N),$$

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 Gradshteyn and Ryzhik (1980), we obtain

$$S(m, n) = \frac{1}{M} \left[ \frac{1 - p}{1 - p} - \frac{1 - p \cos x - p^M \cos (Mx) + p^{M+1} \cos ((M-1)x)}{1 - 2p \cos x + p^2} \right] \cos (Mx) = 1, \quad \sin (Mx) = 0, \quad \cos ((M-1)x) = \cos x.$$
Therefore, simple algebraic transformations yield
\[
S(m, n) = \frac{p(p+1)(1-p^m)(1-\cos x)}{M(1-p)(1-p)^2 + 2p(1-\cos x)}
\]
The asymptotic expansion (A.16) for \( p \) as \( n \to \infty \) leads to
\[
\frac{(1-p^m)}{M(1-p)} \approx \frac{1-N^{-2a}}{4a_1 \ln N(1-a_1 M^{-1} \ln N)},
\]
so that, if \( N \) is large enough, due to \( p < 1 \), one obtains an upper bound for \( S(m, n) \):
\[
S(m, n) \leq \left( \frac{1}{M(1-p)} \right)^{-1} \left( \frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \right) \leq \frac{1}{2a_1 \ln N}.
\]
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
\[
F(p, x) = \frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \leq 2.
\]
If \( x < \pi/3 \), we can use the fact that \( 1-\cos x = 2 \sin^2(x/2) \geq 3x^2/8 \), so that
\[
F(p, x) \leq \frac{4}{3} \left( 1 + \frac{8(1-p)^2}{3x^2} \right) \leq \frac{4}{3} \left( 1 + \frac{2a_1^2 \ln^2 N}{3x^2 m^2} \right)
\]
for \( N \) large enough. \( \Box \)

Since \( |m| = C_m n^2 > C_m C_0 \ln n \) for some \( C_0 > 0 \) and \( \ln n \geq (1-\theta_1)^{-1} \ln N(1+o(1)) \) (as \( n \to \infty \)) due to assumption (7.1), one has \( m^2 \geq C_m C_0^2 (1-\theta_1)^{-2} \ln^2 N \) and
\[
S(m, n) \geq C(\ln N)^{-1}.
\]
Observe now that \( \ln N \approx \ln n \). This completes the proof of the theorem. \( \Box \)

References


