Journal of Statistical Planning and Inference 148 (2014) 1–19

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Multichannel deconvolution with long-range dependence: A minimax study

Rida Benhaddou [a], Rafal Kulik [b], Marianna Pensky [a,*], Theofanis Sapatinas [c]

[a] Department of Mathematics, University of Central Florida, USA
[b] Department of Mathematics and Statistics, University of Ottawa, Canada
[c] Department of Mathematics and Statistics, University of Cyprus, Cyprus

ARTICLE INFO

Article history:
Received 30 April 2013
Received in revised form
14 December 2013
Accepted 16 December 2013
Available online 22 December 2013

Keywords:
Adaptivity
Besov spaces
Block thresholding
Deconvolution
Fourier analysis
Functional data
Long-range dependence
Meyer wavelets
Minimax estimators
Multichannel deconvolution
Partial differential equations
Stationary sequences
Sub-Gaussianity
Wavelet analysis

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 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 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

$$y(u_l, t_i) = \int_T g(u_l, t_i - x) f(x)\, dx + \xi_{li}, \quad l = 1, 2, \ldots, M, \;\; i = 1, 2, \ldots, N, \tag{1.1}$$

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

0378-3758/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.12.008

where g is known, $u_l \in U=[a,b]$, $0<a\le b<\infty$, $T=[0,1]$, $t_i=i/N$, and the errors ξ_{li} 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 $l=1,2,\dots,M$, let $\boldsymbol{\xi}^{(l)}$ be a zero mean vector with components ξ_{li}, $i=1,2,\dots,N$, and let $\boldsymbol{\Sigma}^{(l)}:=\mathrm{Cov}(\boldsymbol{\xi}^{(l)}):=\mathbb{E}[\boldsymbol{\xi}^{(l)}(\boldsymbol{\xi}^{(l)})^T]$ be its covariance matrix. Hence errors ξ_{li} are independent for different l's, but dependent for different i's. Let $\boldsymbol{G}^{(l)}$ be a matrix such that $\boldsymbol{G}^{(l)}(\boldsymbol{G}^{(l)})^T=\boldsymbol{\Sigma}^{(l)}$. Then a vector $\boldsymbol{\eta}^{(l)}=(\boldsymbol{G}^{(l)})^{-1}\boldsymbol{\xi}^{(l)}$ has the covariance matrix $\mathbf{I}_N$, the identity matrix of size N.

In order to formulate our main assumption, recall that a random variable ζ is sub-Gaussian if

$$\|\zeta\|_{\psi_2}:=\sup_{p\ge 1} p^{-1/2}(\mathbb{E}[|\zeta|^p])^{1/p}<\infty.$$

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 (*A0G*). Vectors $\boldsymbol{\xi}^{(l)}$ are of the forms

$$\boldsymbol{\xi}^{(l)}=\boldsymbol{G}^{(l)}\boldsymbol{\eta}^{(l)} \tag{1.2}$$

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

(In what follows, we consider the cases when one knows that $\boldsymbol{\eta}^{(l)}$ are Gaussian vectors and refer to this stronger version of Assumption A0 as Assumption A0G.)

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

Assumption A1. For each $l=1,2,\dots,M$, $\boldsymbol{\Sigma}^{(l)}$ satisfies the following condition: there exist constants K_1 and K_2 $(0<K_1\le K_2<\infty)$, independent of l and N, such that, for each $l=1,2,\dots,M$,

$$K_1 N^{2d_l}\le \lambda_{\min}(\boldsymbol{\Sigma}^{(l)})\le\lambda_{\max}(\boldsymbol{\Sigma}^{(l)})\le K_2N^{2d_l},\quad 0\le d_l<1/2, \tag{1.3}$$

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

Assumption A1 is valid when, for each $l=1,2,\dots,M$, $\boldsymbol{\xi}^{(l)}$ is a second-order stationary Gaussian sequence with spectral density satisfying certain assumptions. We shall elaborate on this issue in Section 2. Note that, in the case of independent errors, for each $l=1,2,\dots,M$, $\boldsymbol{\Sigma}^{(l)}$ 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 $d_l=0$, $l=1,2,\dots,M$, 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 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 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 device 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 $\sigma(x_i, \xi_i)$. In Csörgo 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 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 (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 $f(\cdot)$), 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 $\boldsymbol{\xi}^{(l)}$). 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 $\gamma_\xi(k):=\gamma(k)$, $k=0,\pm 1,\pm 2,\ldots$. The spectral density is defined as

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

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

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

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

Let $\boldsymbol{\Sigma}=[\gamma(j-k)]_{j,k=1}^N$ 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 $\boldsymbol{\Sigma}$ is Hermitian, one has

$$\lambda_{\min}(\boldsymbol{\Sigma})=\inf_{\mathbf{x}\in\mathcal{X}}(\mathbf{x}^*\boldsymbol{\Sigma}\mathbf{x}) \quad \text{and} \quad \lambda_{\max}(\boldsymbol{\Sigma})=\sup_{\mathbf{x}\in\mathcal{X}}(\mathbf{x}^*\boldsymbol{\Sigma}\mathbf{x}). \tag{2.1}$$

With the definitions introduced above,

$$\mathbf{x}^*\boldsymbol{\Sigma}\mathbf{x} = \sum_{j,k=1}^{N} \mathbf{x}^*\gamma(j-k)\mathbf{x} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^{N} x_j e^{-ij\lambda} \right|^2 a(\lambda)\, d\lambda. \tag{2.2}$$

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

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

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

$$\mathcal{F}_d = \{a : a(\lambda) = |\lambda|^{-2d} a_*(\lambda),\ 0 < C_{\min} \le |a_*(\lambda)| \le C_{\max} < \infty, \lambda \in [-\pi, \pi]\}. \tag{2.3}$$

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 η_j are uncorrelated, zero-mean, random variables, $\sigma_\eta^2 := \mathrm{Var}(\eta_j) < \infty$, and

$$a_m = (-1)^m \binom{-d}{m} = (-1)^m \frac{\Gamma(1-d)}{\Gamma(m+1)\Gamma(1-d-m)}$$

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

$$a(\lambda) = \frac{\sigma_\eta^2}{2\pi} |A(e^{-i\lambda})|^2 = \frac{\sigma_\eta^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} = \frac{\sigma_\eta^2}{2\pi} |2(1 - \cos\lambda)|^{-d} \sim \frac{\sigma_\eta^2}{2\pi} |\lambda|^{-2d} \quad (\lambda \to 0).$$

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,

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

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 $a(\lambda) \sim |\lambda|^{-2d}$ $(\lambda \to 0)$. 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 $\boldsymbol{\Sigma}$ are of asymptotic order N^{2d} $(N \to \infty)$. 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} $(0 < K_{1d} \le K_{2d} < \infty)$ *that depend on d only,*

$$K_{1d} N^{2d} \le \lambda_{\min}(\boldsymbol{\Sigma}) \le \lambda_{\max}(\boldsymbol{\Sigma}) \le K_{2d} N^{2d}.$$

Remark 1. If $d = 0$, then $\mathcal{F}_d$ is the class of spectral densities $a(\lambda)$ that are bounded away from 0 and ∞ 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* $\boldsymbol{\xi}^{(l)}$ *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^* $(0 \le d^* < 1/2)$ *such that, for each* $l = 1, 2, \ldots, M$,

$$0 \le d_l \le d^* < 1/2. \tag{2.4}$$

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 $e_m(t_i) = e^{-i2\pi m t_i}$, $t_i = i/N$, $i = 1, 2, \ldots, N$, $m = 0, \pm 1, \pm 2, \ldots$, and the complex-conjugate of the matrix $\mathbf{A}$ by $\mathbf{A}^*$.

Recall the multichannel deconvolution model (1.1). Denote

$$h(u_l, t_i) = \int_T g(u_l, t_i - x) f(x)\, dx, \quad l = 1, 2, \ldots, M, \; i = 1, 2, \ldots, N.$$

Then, Eq. (1.1) can be rewritten as

$$y(u_l, t_i) = h(u_l, t_i) + \xi_{li}, \quad l = 1, 2, \ldots, M, \; i = 1, 2, \ldots, N. \tag{3.1}$$

For each $l = 1, 2, \ldots, M$, let $h_m(u_l) = \langle e_m, h(u_l, \cdot) \rangle$, $y_m(u_l) = \langle e_m, y(u_l, \cdot) \rangle$, $z_{lm} = \langle e_m, \xi^{(l)} \rangle$, $g_m(u_l) = \langle e_m, g(u_l, \cdot) \rangle$ 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 (3.1), one obtains, for any $u_l \in U$, $l = 1, 2, \ldots, M$,

$$h_m(u_l) = g_m(u_l) f_m \tag{3.2}$$

and

$$y_m(u_l) = g_m(u_l) f_m + N^{-1/2} z_{lm}. \tag{3.3}$$

Multiplying both sides of (3.3) by $N^{-2d_l}\overline{g_m(u_l)}$, and adding them together, we obtain the following estimator of f_m:

$$\widehat{f}_m = \left(\sum_{l=1}^{M} N^{-2d_l} \overline{g_m(u_l)} y_m(u_l) \right) \Bigg/ \left(\sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 \right). \tag{3.4}$$

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 $j \ge 0$ 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 $j_0 \ge 0$, expand $f(\cdot) \in L^2(T)$ as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0 k} \varphi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j - 1} b_{jk} \psi_{jk}(t), \quad t \in T. \tag{3.5}$$

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

$$a_{j_0 k} = \sum_{m \in C_{j_0}} f_m \overline{\varphi_{mj_0 k}}, \quad b_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}}, \tag{3.6}$$

where $\varphi_{mj_0k} = \langle e_m, \varphi_{j_0 k} \rangle$, $C_{j_0} = \{m : \varphi_{mj_0k} \ne 0\}$, $\psi_{mjk} = \langle e_m, \psi_{jk} \rangle$ and, for any $j \ge j_0$,

$$C_j = \{m : \psi_{mjk} \ne 0\} \subseteq 2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}].$$

(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_{jk} are readily obtained by substituting f_m in (3.6) with (3.4), i.e.,

$$\widehat{a}_{j_0 k} = \sum_{m \in C_{j_0}} \widehat{f}_m \overline{\varphi_{mj_0 k}}, \quad \widehat{b}_{jk} = \sum_{m \in C_j} \widehat{f}_m \overline{\psi_{mjk}}. \tag{3.7}$$

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_j and U_{jr} be the following sets of indices:

$$A_j = \{r \mid r = 1, 2, \ldots, [2^j / \ln n]\},$$

$$U_{jr} = \{k \mid k = 0, 1, \ldots, 2^j - 1; (r-1)\ln n \le k \le r \ln n - 1\}.$$

Denote

$$B_{jr} = \sum_{k \in U_{jr}} b_{jk}^2, \quad \widehat{B}_{jr} = \sum_{k \in U_{jr}} \widehat{b}_{jk}^2. \tag{3.8}$$

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

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \widehat{b}_{jk}\mathbb{I}(|\widehat{B}_{jr}| \geq \lambda_j)\psi_{jk}(t), \quad t \in T, \tag{3.9}$$

where $\mathbb{I}(A)$ 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.

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

$$s' = s + 1/2 - 1/p, \quad s^* = s + 1/2 - 1/p', \quad p' = \min\{p, 2\}. \tag{4.1}$$

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_0k} and b_{jk} defined in (3.6) satisfy the following relation:

$$B^s_{p,q}(A) = \left\{ f \in L^2(U) : \left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0k}|^p \right)^{1/p} + \left(\sum_{j=j_0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \right\}. \tag{4.2}$$

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

$$R_n(V) = \inf_{\tilde{f}} \sup_{f \in V} \mathbb{E}\|\tilde{f} - f\|^2,$$

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 $f(\cdot)$, based on observations from model (1.1)).

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

$$\tau_\kappa(m,n) = M^{-1} \sum_{l=1}^{M} N^{-2\kappa d_l} |g_m(u_l)|^{2\kappa}, \quad \kappa = 1 \text{ or } 2 \text{ or } 4, \tag{4.3}$$

and

$$\Delta_\kappa(j,n) = |C_j|^{-1} \sum_{m \in C_j} \tau_\kappa(m,n)[\tau_1(m,n)]^{-2\kappa}, \quad \kappa = 1 \text{ or } 2. \tag{4.4}$$

The expression $\tau_1(m,n)$ 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 $\nu_1, \nu_2, \vartheta_1, \vartheta_2 \in \mathbb{R}$, $\alpha_1, \alpha_2 \geq 0$ ($\vartheta_1, \vartheta_2 > 0$ 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

$$K_3\varepsilon_n|m|^{-2\nu_1}(\ln|m|)^{-\vartheta_1}e^{-\alpha_1|m|^\beta} \leq \tau_1(m,n) \leq K_4\varepsilon_n|m|^{-2\nu_2}(\ln|m|)^{-\vartheta_2}e^{-\alpha_2|m|^\beta}, \tag{4.5}$$

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 (4.5) is such that

$$n^* = n\varepsilon_n \to \infty \quad (n \to \infty). \tag{4.6}$$

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 $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section3. Let $s > \max(0, 1/p - 1/2)$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Then, as $n \to \infty$,*

$$R_n\left(B^s_{p,q}(A)\right) \geq \begin{cases} C(n^*)^{-2s/(2s+2\nu+1)}(\ln n^*)^{2s\vartheta_2/(2s+2\nu+1)} & \text{if } \alpha_1 = \alpha_2 = 0, \ \nu(2-p) < ps^*, \\ C\left(\dfrac{\ln n^*}{n^*}\right)^{2s^*/(2s^*+2\nu)} (\ln n^*)^{2s^*\vartheta_2/(2s^*+2\nu)} & \text{if } \alpha_1 = \alpha_2 = 0, \ \nu(2-p) \geq ps^*, \\ C(\ln n^*)^{-2s^*/\beta} & \text{if } \alpha_1\alpha_2 \neq 0. \end{cases} \tag{4.7}$$

5. Minimax upper bounds for the L^2–risk: Gaussian case

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

Let $\hat{f}_n(\cdot)$ be the (block thresholding) wavelet estimator defined by (3.9). Choose now j_0 and J such that

$$2^{j_0}=\ln n^*, \quad 2^J=(n^*)^{1/(2\nu+1)} \text{ if } \alpha_1=\alpha_2=0, \tag{5.1}$$

$$2^{j_0}=\frac{3}{8\pi}\left(\frac{\ln n^*}{2\alpha}\right)^{1/\beta}, \quad 2^J=2^{j_0} \text{ if } \alpha_1\alpha_2>0. \tag{5.2}$$

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

$$\lambda_j=\mu^2\,(n^*)^{-1}\ln n^* 2^{2\nu j}j^{\vartheta_1} \quad \text{if } \alpha_1=\alpha_2=0. \tag{5.3}$$

(Since $j_0>J-1$ when $\alpha_1\alpha_2>0$, the estimator (3.9) only consists of the first (linear) part and, hence, λ_j does not need to be selected in this case.) 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 (3.9) is adaptive with respect to these parameters.

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

$$\varrho=\begin{cases}\dfrac{(2\nu+1)(2-p)_+}{p(2s+2\nu+1)} & \text{if } \nu(2-p)<ps^*,\\ \dfrac{(q-p)_+}{q} & \text{if } \nu(2-p)=ps^*,\\ 0 & \text{if } \nu(2-p)>ps^*.\end{cases} \tag{5.4}$$

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

$$-h_1\ln n\le \ln\varepsilon_n\le h_2\ln n \tag{5.5}$$

for some constants $h_1,h_2\in(0,1)$. Observe that condition (5.5) implies (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 $0<C_1v(n)\le u(n)\le C_2v(n)<\infty$ 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

$$\Delta_1(j,n)\le\begin{cases}c_1\ \varepsilon_n^{-1}\ 2^{2\nu j}j^{\vartheta_1} & \text{if } \alpha_1=\alpha_2=0,\\ c_2\ \varepsilon_n^{-1}\ 2^{2\nu_1 j}j^{\vartheta_1}\exp\left\{\alpha_1\left(\dfrac{8\pi}{3}\right)^{\beta}2^{j\beta}\right\} & \text{if } \alpha_1\alpha_2>0.\end{cases} \tag{5.6}$$

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 $\widehat{a}_{j_0k}$ and $\widehat{b}_{jk}$ of the scaling and wavelet coefficients a_{j_0k} and b_{jk}, respectively, be given by (3.6) with $\hat{f}_m$ defined by (3.4). Then, for all $j\ge j_0$,*

$$\mathbb{E}|\widehat{a}_{j_0k}-a_{j_0k}|^2\le Cn^{-1}\Delta_1(j_0,n) \quad \text{and} \quad \mathbb{E}|\widehat{b}_{jk}-b_{jk}|^2\le Cn^{-1}\Delta_1(j,n). \tag{5.7}$$

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

$$\mathbb{E}|\widehat{b}_{jk}-b_{jk}|^4\le Cn^3(\ln n)^{3\vartheta_1}(n^*)^{-3/(2\nu+1)}. \tag{5.8}$$

Lemma 3. *Let Assumptions A0G, A1, A2 and (5.5) hold. Let the estimators $\widehat{b}_{jk}$ of the wavelet coefficients b_{jk} be given by (3.6) with $\hat{f}_m$ defined by (3.4). Let*

$$\mu\ge\frac{2}{\sqrt{1-h_1}}\left[\sqrt{c_1}+\frac{\sqrt{8\pi\kappa}}{\sqrt{K_3}}(\ln 2)^{\vartheta_1/2}\left(\frac{2\pi}{3}\right)^{\nu}\right], \tag{5.9}$$

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

$$\mathbb{P}\left(\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2\ge 0.25\mu^2\,(n^*)^{-1}2^{2\nu j}j^{\vartheta_1}\ln n^*\right)\le n^{-\kappa}. \tag{5.10}$$

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 $\hat{f}_n(\cdot)$ 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 μ satisfying (5.9) with $\kappa=5$. Let $s>1/p'$, $1\le p\le\infty$, $1\le q\le\infty$ and $A>0$. Then, under*

(4.6) *if* $\alpha_1\alpha_2>0$ *or* (5.5) *if* $\alpha_1=\alpha_2=0$, *as* $n\to\infty$,

$$\sup_{f\in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n-f\|^2 \le \begin{cases} C(n^*)^{-2s/(2s+2\nu+1)}(\ln n)^{\varrho+2s\vartheta_1/(2s+2\nu+1)} & \text{if } \alpha_1=\alpha_2=0,\ \nu(2-p)<ps^*,\\ C\left(\dfrac{\ln n}{n^*}\right)^{2s^*/(2s^*+2\nu)}(\ln n)^{\varrho+2s^*\vartheta_1/(2s^*+2\nu)} & \text{if } \alpha_1=\alpha_2=0,\ \nu(2-p)\ge ps^*,\\ C(\ln n^*)^{-2s^*/\beta} & \text{if } \alpha_1\alpha_2>0. \end{cases} \tag{5.11}$$

Remark 2. Theorems 1 and 2 imply that, for the L^2-risk, the wavelet estimator $\hat{f}_n(\cdot)$ 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 $B^s_{p,q}(A)$ of radius $A>0$ with $s>\max(1/p,1/2)$, $1\le p\le\infty$ and $1\le q\le\infty$. The convergence rates depend on the balance between the smoothness parameter s (of the response function $f(\cdot)$), the kernel parameters ν,β,ϑ_1 and ϑ_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

$$\varepsilon_n=M^{-1}\sum_{l=1}^{M}N^{-2d_l}, \tag{5.12}$$

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 7 below.

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

In this section, we shall assume that random variables η_{li}, 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 ϱ be defined in (5.4). Assume that, in the case of $\alpha_1=\alpha_2=0$, sequence ε_n satisfies condition (5.5). For some constant $\mu>0$, large enough, choose

$$\lambda_j=4c_1(1+\mu^2\ln n)(n^*)^{-1}\ln(n)2^{2\nu j}j^{\vartheta_1}\quad \text{if } \alpha_1=\alpha_2=0, \tag{6.1}$$

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 $\boldsymbol{G}$, let $\|\boldsymbol{G}\|_{sp}$ and $\|\boldsymbol{G}\|_2$ be, respectively, the spectral and the Frobenius norms.

Lemma 4 (*The matrix version of the Hanson–Wright inequality, Rudelson and Vershynin, 2013*). *Let* $\mathbf{X}=(X_1,\ldots,X_n)$ *be a random vector with independent components such that* $E[X_i]=0$, $\|X_i\|_{\psi_2}\le K$. *Then, for any matrix* $\mathbf{B}$, *and some absolute constant* $c_0>0$, *one has*

$$P\left(\left|\mathbf{X}^T\mathbf{B}\mathbf{X}-\mathbb{E}\left[\mathbf{X}^T\mathbf{B}\mathbf{X}\right]\right|>t\right)\le 2\exp\left(-c_0\min\left\{\frac{t^2}{K^4\|\mathbf{B}\|_2^2},\frac{t}{K^2\|\mathbf{B}\|_{sp}}\right\}\right). \tag{6.2}$$

Lemma 5. *Let Assumptions A0, A1, A2 and (5.5) hold. Let the estimators* $\widehat{b}_{jk}$ *of the wavelet coefficients* b_{jk} *be given by (3.6) with* $\widehat{f}_m$ *defined by (3.4). Then, for all* $j\ge j_0$, *(5.7) holds. Moreover, if* $\alpha_1=\alpha_2=0$ *and (5.5) holds, then, for any* $j\ge j_0$,

$$\mathbb{E}|\widehat{b}_{jk}-b_{jk}|^4\le Cn^3(n^*)^{-2}. \tag{6.3}$$

In addition, for all $j\ge j_0$ *and any* $\kappa>0$,

$$P\left(\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2>c_1(1+\mu^2\ln n)(n^*)^{-1}\ln n\,2^{2\nu j}j^{\vartheta_1}\right)\le 2\,n^{-\kappa}, \tag{6.4}$$

provided

$$\mu\ge K\sqrt{c_0\kappa}, \tag{6.5}$$

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* $\hat{f}_n(\cdot)$ *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* μ *satisfying (6.5) with* $\kappa=5$. *Let* $s>1/p'$, $1\le p\le\infty$, $1\le q\le\infty$ *and* $A>0$. *Then, under*

(4.6) *if* $\alpha_1\alpha_2>0$ *or* (5.5) *if* $\alpha_1=\alpha_2=0$, *as* $n\to\infty$,

$$\sup_{f\in B_{p,q}^s(A)}\mathbb{E}\|\hat{f}_n-f\|^2\le\begin{cases}C(n^*)^{-2s/(2s+2\nu+1)}(\ln n)^{1+\varrho+2s\vartheta_1/(2s+2\nu+1)} & \text{if } \alpha_1=\alpha_2=0,\ \nu(2-p)<ps^*,\\ C\left(\dfrac{\ln n}{n^*}\right)^{2s^*/(2s^*+2\nu)}(\ln n)^{1+\varrho+2s^*\vartheta_1/(2s^*+2\nu)} & \text{if } \alpha_1=\alpha_2=0,\ \nu(2-p)\ge ps^*,\\ C(\ln n^*)^{-2s^*/\beta} & \text{if } \alpha_1\alpha_2>0.\end{cases}\tag{6.6}$$

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_5, K_6, θ_1 and θ_2, such that $M=M_n$ satisfies

$$K_5n^{\theta_1}\le M\le K_6n^{\theta_2},\quad 0\le\theta_1\le\theta_2<1,\quad 0<K_5\le K_6<\infty.\tag{7.1}$$

(Note that, under (7.1), $K_5n^{1-\theta_2}\le N\le K_6n^{1-\theta_1}$.)

Example 1. Consider the case when $g_m(\cdot)$, $m=0,\pm1,\pm2,\dots$, is of the form

$$g_m(u)=C_g\exp(-K|m|^\beta q(u)),\quad u\in U,\tag{7.2}$$

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

$$0<q_1\le q(u)\le q_2<\infty,\quad u\in U.\tag{7.3}$$

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, $g_m(u)=\exp(-4\pi^2m^2u)$, $u\in U$, 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 (7.2) and (7.3), for $\tau_1(m,n)$ given in (4.3),

$$\tau_1(m,n)\le C_g\varepsilon_n\exp(-2Kq_1|m|^\beta)\quad\text{and}\quad\tau_1(m,n)\ge C_g\varepsilon_n\exp(-2Kq_2|m|^\beta),$$

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

$$K_5n^{1-2d^*(1-\theta_1)}\le n^*\le n,$$

so that $\ln n\asymp\ln n^*$. Therefore, according to Theorems 1 and 2,

$$R_n(B_{p,q}^s(A))\asymp(\ln n)^{-2s^*/\beta}.\tag{7.4}$$

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 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 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.,

$$g(u,t)=0.5q(u)\mathbb{I}(|t|<u),\quad u\in U,\ t\in T,\tag{7.5}$$

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

$$g_0(u)=1\quad\text{and}\quad g_m(u)=(2\pi m)^{-1}\gamma(u)\sin(2\pi mu),\ m\in\mathbb{Z}\setminus\{0\},\ u\in U.\tag{7.6}$$

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) $u_l=l/M$, $l=1,2,\dots,M$, such that

$$d_l=a_1u_l+a_2,\quad 0\le a_2\le d^*<1/2,\ 0\le a_1+a_2\le d^*<1/2,\tag{7.7}$$

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

$$\tau_1(m,n)\asymp M^{-1}N^{-2a_2}(4\pi^2m^2)^{-1}\sum_{l=1}^M\sin^2(2\pi ml/M),$$

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 $2^{j_0} > (\ln n)^\delta$ for some $\delta > 0$ and $M \ge (32\pi/3)n^{1/3}$, then, for n and $|m|$ large enough,

$$\tau_1(m,n) \asymp N^{-2a_2} m^{-2}.$$

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 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_l, $l = 1,2,\dots,M$, be given by (7.7) with $a_1 \ge 0$. Assume (without loss of generality) that $U = [0,1]$, and consider $u_l = l/M$, $l = 1,2,\dots,M$. Let $M = M_n$ satisfy (7.1) 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,*

$$\tau_1(m,n) \asymp N^{-2a_2} m^{-2} (\log n)^{-1}. \tag{7.8}$$

It follows immediately from Lemma 6 that, if

$$M = M_n \asymp n^\theta, \quad 0 < \theta < 1,$$

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 $\vartheta_1 = \vartheta_2 = 0$. Note that ε_n satisfies conditions (4.6) and (5.5), so that $\ln n \asymp \ln n^*$. Therefore, according to Theorems 1 and 2,

$$R_n\left(B^s_{p,q}(A)\right) \ge \begin{cases} C(n^*)^{-2s/(2s+5)} & \text{if } 4-2p < ps^*, \\ C\left(\dfrac{\ln n^*}{n^*}\right)^{s^*/(s^*+2)} & \text{if } 4-2p \ge ps^*, \end{cases} \tag{7.9}$$

and

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \le \begin{cases} C(n^*)^{-2s/(2s+5)}(\ln n)^\varrho & \text{if } 4-2p < ps^*, \\ C\left(\dfrac{\ln n}{n^*}\right)^{s^*/(s^*+2)}(\ln n)^\varrho & \text{if } 4-2p \ge ps^*, \end{cases} \tag{7.10}$$

where

$$n^* = n^{1-2a_2(1-\theta)}(\ln n)^{-1}$$

and

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

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 ($M \ge 2$) 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) 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 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 Khasminskii (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 $f(\cdot)$), 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).

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) $u_l=l/M$, $l=1,2,\ldots,M$, $n^{\theta_1}\leq M=M_n\leq n^{\theta_2}$, for some $0\leq\theta_1\leq\theta_2<1$ when $d_l=a_1u_l+a_2$, $l=1,2,\ldots,M$, $0\leq a_2<1/2$, $0\leq a_1+a_2<1/2$.

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 $\alpha\geq1/2$), 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

Marianna 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 $\mathcal{F}_d$,

$$\lambda_{\max}(\boldsymbol{\Sigma})\leq C_{\max}\sup_{h\in\mathcal{H}_N}\int_{-\pi}^{\pi}h(\lambda)|\lambda|^{-2d}\,d\lambda=2C_{\max}\sup_{h\in\mathcal{H}_N}\int_{0}^{\pi}h(\lambda)|\lambda|^{-2d}\,d\lambda.$$

Now, we split $\int_0^\pi=\int_0^{\pi/N}+\int_{\pi/N}^{\pi}$. Since $d<1/2$, for the first integral, we have

$$\int_0^{\pi/N}h(\lambda)|\lambda|^{-2d}\,d\lambda\leq N\int_0^{\pi/N}\lambda^{-2d}\,d\lambda=N\frac{1}{1-2d}\left(\frac{\pi}{N}\right)^{-2d+1}=\frac{\pi^{-2d+1}}{1-2d}N^{2d}.$$

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. □

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 Fanno 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 Θ be a set of functions of cardinality* $\text{card}(\Theta) \geq 2$, *such that*

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

Then, for some absolute constant $C > 0$, one has

$$\inf_{T_n} \sup_{f \in \Theta} \mathbb{E}_f \|T_n - f\|^2 \geq C\delta^2,$$

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

The dense case: Let $\boldsymbol{\omega}$ be the 2^j-dimensional vector with components $\omega_k = \{0, 1\}$. Denote the set of all possible vectors $\boldsymbol{\omega}$ by Ω: $\Omega = \{(0,1)^{2^j}\}$, the set of binary sequences of length 2^j. Note that the vector $\boldsymbol{\omega}$ has $\aleph = 2^j$ entries and, hence, $\text{card}(\Omega) = 2^{\aleph}$. Let $H(\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}) = \sum_{k=0}^{2^j-1} \mathbb{I}(\tilde{\omega}_k \neq \omega_k)$ 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 $\boldsymbol{\omega}, \tilde{\boldsymbol{\omega}} \in \Omega_1$.

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

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

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

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

Hence, we get $4\delta^2 = 2^j\rho_j^2/8$ in condition (i) of Lemma 7.

In order to apply Lemma 7, one needs to also verify condition (ii). For $f_{\boldsymbol{\omega}}$ with $\boldsymbol{\omega} \in \Omega$, denote by $\mathbf{h}_{l,\boldsymbol{\omega}}$ and $\mathbf{h}_{l,\tilde{\boldsymbol{\omega}}}$, the vectors with components, respectively,

$$\begin{aligned} h_{\boldsymbol{\omega}}(u_l, t_i) &= g(u_l, t_i - \cdot) * f_{\boldsymbol{\omega}}(\cdot), \quad i = 1, 2, \ldots, N, \\ h_{\tilde{\boldsymbol{\omega}}}(u_l, t_i) &= g(u_l, t_i - \cdot) * f_{\tilde{\boldsymbol{\omega}}}(\cdot), \quad i = 1, 2, \ldots, N. \end{aligned}$$

Then,

$$\begin{aligned} K(P_{f_{\boldsymbol{\omega}}}, P_{\tilde{f}_{\tilde{\boldsymbol{\omega}}}}) &= 0.5 \sum_{l=1}^{M} (\mathbf{h}_{l,\boldsymbol{\omega}} - \mathbf{h}_{l,\tilde{\boldsymbol{\omega}}})^T (\boldsymbol{\Sigma}^{(l)})^{-1} (\mathbf{h}_{l,\boldsymbol{\omega}} - \mathbf{h}_{l,\tilde{\boldsymbol{\omega}}}) \\ &\leq 0.5 \sum_{l=1}^{M} \lambda_{\max}((\boldsymbol{\Sigma}^{(l)})^{-1}) \|\mathbf{h}_{l,\boldsymbol{\omega}} - \mathbf{h}_{l,\tilde{\boldsymbol{\omega}}}\|^2. \end{aligned}$$

Now, since $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ 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,

$$\begin{aligned} K\left(P_{f_{\boldsymbol{\omega}}}, P_{\tilde{f}_{\tilde{\boldsymbol{\omega}}}}\right) &\leq 0.5 N M \rho_j^2 \sum_{m \in C_j} \frac{1}{M} \sum_{l=1}^{M} |g_m(u_l)|^2 K_1^{-1} N^{-2d_l} \\ &\leq 2\pi K_1^{-1} n 2^j \rho_j^2\, \Delta_1(j,n) \leq 2\pi A^2 K_1^{-1} n 2^{-2js}\, \Delta_1(j,n), \end{aligned}$$

where $\Delta_1(j,n)$ is defined by (4.4).

Direct calculations yield that, under Assumptions A1, A2 and (4.5), for some constants $c_3>0$ and $c_4>0$, independent of n,

$$\Delta_1(j,n)\le\begin{cases} c_3\ \varepsilon_n^{-1}\ 2^{2\nu j}j^{\vartheta_2} & \text{if } \alpha_1=\alpha_2=0,\\ c_4\ \varepsilon_n^{-1}\ 2^{2\nu_1 j}j^{\vartheta_2}\ \exp\left\{\alpha_1\left(\dfrac{8\pi}{3}\right)^{\beta}2^{j\beta}\right\} & \text{if } \alpha_1\alpha_2>0.\end{cases}\tag{A.1}$$

Apply now Lemma 7 with j such that

$$2\pi A^2K_1^{-1}n2^{-2js}\ \Delta_1(j,n)\le 2^j\ \ln 2/16,$$

i.e.,

$$2^j\asymp\begin{cases}[n^*(\ln n^*)^{-\vartheta_2}]^{1/(2s+2\nu+1)} & \text{if } \beta=0,\\ (\ln n^*)^{1/\beta} & \text{if } \beta>0,\end{cases}$$

to obtain

$$\delta^2=\begin{cases}[n^*(\ln n^*)^{-\vartheta_2}]^{-2s/(2s+2\nu+1)} & \text{if } \beta=0,\\ (\ln n^*)^{-2s/\beta} & \text{if } \beta>0.\end{cases}\tag{A.2}$$

The sparse case: Consider the functions $f_k(\cdot)$ of the form $f_k(t)=\rho_j\psi_{jk}(t)$, $t\in T$, $k=0,1,\ldots,2^j-1$, and denote

$$\Theta=\{f_k(t)=\rho_j\psi_{jk}(t):k=0,1,\ldots,2^j-1,\ f_0=0\}.$$

Thus, card$(\Theta)=2^j$. Choose now $\rho_j=A2^{-js'}$, so that $f_k\in B^s_{p,q}(A)$. It is easy to check that, in this case, one has $4\delta^2=\rho_j^2$ in Lemma 7, and that

$$K(P_{f_k},P_{f_{\hat k}})\le 2\pi A^2K_1^{-1}n2^{-2js'}\ \Delta_1(j,n).$$

With

$$2^j\asymp\begin{cases}[n^*(\ln n^*)^{-\vartheta_2-1}]^{1/(2s'+2\nu)} & \text{if } \beta=0,\\ (\ln n^*)^{1/\beta} & \text{if } \beta>0,\end{cases}$$

we then obtain that $K(P_{f_k},P_{f_{\hat k}})\le 2\pi A^2K_1^{-1}n2^{-2js'}\Delta_1(j,n)$ and

$$\delta^2=\begin{cases}\left[\dfrac{n^*}{(\ln n^*)^{\vartheta_2+1}}\right]^{-2s'/(2s'+2\nu)} & \text{if } \beta=0,\\ (\ln n^*)^{-2s'/\beta} & \text{if } \beta>0.\end{cases}\tag{A.3}$$

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

$$2s/(2s+2\nu+1)\le 2s^*/(2s^*+2\nu)\quad \text{if } \nu(2-p)\le ps^*,\tag{A.4}$$

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

$$\widehat{a}_{j_0k}-a_{j_0k}=\sum_{m\in C_{j_0}}(\widehat{f}_m-f_m)\overline{\varphi_{mj_0k}},\quad \widehat{b}_{jk}-b_{jk}=\sum_{m\in C_j}(\widehat{f}_m-f_m)\overline{\psi_{mjk}},$$

where

$$\widehat{f}_m-f_m=\frac{1}{\sqrt{N}}\left(\sum_{l=1}^{M}N^{-2d_l}\overline{g_m(u_l)}z_{lm}\right)\Big/\left(\sum_{l=1}^{M}N^{-2d_l}|g_m(u_l)|^2\right).\tag{A.5}$$

Define

$$v_m=\sum_{l=1}^{M}N^{-2d_l}|g_m(u_l)|^2=M\tau_1(m,n).\tag{A.6}$$

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

$$V_m^{(l)}=N^{-2d_l}\psi_{mjk}g_m(u_l)\left[\sum_{j=1}^{M}N^{-2d_j}|g_m(u_j)|^2\right]^{-1}=N^{-2d_l}\psi_{mjk}g_m(u_l)v_m^{-1}.\tag{A.7}$$

It is easy to see that, due to $|\psi_{mjk}| \le 2^{-j/2}$ and the definition of C_j,

$$\|\mathbf{V}^{(l)}\|^2 = N^{-4d_l} \sum_{m \in C_j} |\psi_{mjk}|^2 |g_m(u_l)|^2 \left[\sum_{t=1}^{M} N^{-2d_t} |g_m(u_t)|^2 \right]^{-2}$$
$$\le 4\pi |C_j|^{-1}\, N^{-4d_l} \sum_{m \in C_j} |g_m(u_l)|^2 \left[\sum_{t=1}^{M} N^{-2d_t} |g_m(u_t)|^2 \right]^{-2}.$$

Hence,

$$\|\mathbf{V}^{(l)}\|^2 \le 4\pi |C_j|^{-1} N^{-2d_l} N^{-2d_l} \sum_{m \in C_j} |g_m(u_l)|^2 v_m^{-2}.$$

Using Assumption A1, since z_{lm} are independent for different l's, we obtain

$$\begin{aligned}
\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 &= \frac{1}{N} \sum_{m_1, m_2 \in C_j} \overline{\psi}_{m_1 jk} \psi_{m_2 jk} \sum_{l=1}^{M} N^{-4d_l} v_{m_1}^{-1} v_{m_2}^{-1} \overline{g_{m_1}(u_l)} g_{m_2}(u_l) \mathrm{Cov}\left(z_{lm_1}, \overline{z}_{lm_2}\right) \\
&= \frac{1}{N} \sum_{l=1}^{M} \overline{\mathbf{V}^{(l)}}^T \mathbf{\Sigma}^{(l)} \mathbf{V}^{(l)} \\
&\le \frac{1}{N} \sum_{l=1}^{M} \lambda_{\max}\left(\mathbf{\Sigma}^{(l)}\right) \|\mathbf{V}^{(l)}\|^2 \\
&\le 4\pi K_2 |C_j|^{-1} N^{-1} \sum_{l=1}^{M} N^{-2d_l} \sum_{m \in C_j} |g_m(u_l)|^2 v_m^{-2} \\
&= 4\pi K_2 |C_j|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-2} \sum_{l=1}^{M} N^{-2d_l} |g_m(u_l)|^2 = 4\pi K_2 |C_j|^{-1} N^{-1} \sum_{m \in C_j} v_m^{-1},
\end{aligned}$$

so that

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 \le C n^{-1} |C_j|^{-1} \sum_{m \in C_j} [\tau_1(m,n)]^{-1} := C n^{-1} \Delta_1(j,n).$$

(One can obtain an upper bound for $\mathbb{E}|\widehat{a}_{j_0 k} - a_{j_0 k}|^2$ by the following similar arguments.)

In order to prove (5.8), define

$$B_{lm} = N^{-2d_l} \left[\sum_{j=1}^{M} N^{-2d_j} |g_m(u_j)|^2 \right]^{-1} = N^{-2d_l} v_m^{-1}.$$

Note that

$$\mathbb{E}(z_{lm_1} z_{lm_2} z_{lm_3} z_{lm_4}) \le \left[\prod_{i=1}^{4} \mathbb{E}|z_{m_i l}|^4 \right]^{1/4}.$$

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

$$\begin{aligned}
\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(N^{-2} \sum_{l=1}^{M} B_{lm}^4 \left[\sum_{m \in C_j} |\psi_{mjk}| |g_{m_2}(u_l)| (\mathbb{E}|z_{lm}|^4)^{1/4} \right]^4 \right) \\
&\quad + O\left(\left[N^{-1} \sum_{l=1}^{M} B_{lm}^2 \sum_{m_1, m_2 \in C_j} \overline{\psi}_{m_1 jk} \psi_{m_2 jk} \overline{g_{m_1}(u_l)} g_{m_2}(u_l) \mathrm{Cov}(z_{lm_1}, \overline{z}_{lm_2}) \right]^2 \right) \\
&= O\left(N^{-2} \sum_{l=1}^{M} B_{lm}^4 \left[\sum_{m \in C_j} |\psi_{mjk}|^2 |g_m(u_l)|^2 \sum_{m \in C_j} \mathbb{E}|z_{ml}|^2 \right]^2 \right) \\
&\quad + O\left(\left[n^{-1} |C_j|^{-1} \sum_{m \in C_j} [\tau_1(m,n)]^{-1} \right]^2 \right).
\end{aligned}$$

Since $\sum_{m \in C_j} \mathbb{E}|z_{lm}|^2 = O(|C_j|)$, one derives

$$\begin{aligned}
\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(|C_j|^{-1} \sum_{m \in C_j} \left[\frac{1}{M^3} \frac{\tau_2(m,n)}{[\tau_1(m,n)]^4} \right] + \frac{\Delta_1^2(j,n)}{n^2} \right) \\
&= O(M^{-3} \Delta_2(j,n) + n^{-2} \Delta_1^2(j,n)).
\end{aligned} \tag{A.8}$$

It is straightforward to show that, when $\alpha_1 = \alpha_2 = 0$, one has

$$\Delta_2(j,n) = O(2^{6j\nu} j^{3\vartheta_1} \varepsilon_n^{-3}).$$

Thus, using (A.1) and the fact that $2^j \le 2^{J-1} < (n^*)^{1/(2\nu+1)}$, (A.8) can be rewritten as

$$\begin{aligned}\mathbb{E}|\widehat{b}_{jk}-b_{jk}|^4 &= O(2^{6\nu j} j^{3\vartheta_1}\varepsilon_n^{-3}M^{-3}+2^{4j\nu}j^{2\vartheta_1}\varepsilon_n^{-2}n^{-2})\\ &= O(n^3(\ln n)^{3\vartheta_1}(n^*)^{-3/(2\nu+1)}).\end{aligned}$$

Hence, (5.8) follows. This completes the proof of the lemma. □

Proof of Lemma 3. Consider a set of vectors

$$\Omega_{jr}=\left\{v_k, k\in U_{jr} : \sum_{k\in U_{jr}}|v_k|^2\le 1\right\}$$

and a centered Gaussian process

$$Z_{jr}=\sum_{k\in U_{jr}} v_k(\widehat{b}_{jk}-b_{jk}).$$

Note that, by Jensen's inequality,

$$\sup_v Z_{jr}(v)=\sqrt{\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2}.$$

We shall apply below a lemma of Cirelson et al. (1976) which states that, for any $x>0$,

$$\Pr\left(\sqrt{\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2}\ge x+B_1\right)\le \exp\left(-\frac{x^2}{2B_2}\right), \tag{A.9}$$

where

$$\mathbb{E}\left[\sqrt{\sum_{k\in U_{jr}}|\widehat{b}_{jk}-b_{jk}|^2}\right]\le \frac{\sqrt{c_1}2^{j\nu}j^{\vartheta_1/2}\sqrt{\ln n}}{\sqrt{n^*}}:=B_1$$

with c_1 defined in (5.6), and B_2 is an upper bound for

$$\sup_{v\in\Omega_{jr}}\mathrm{Var}(Z_{jr}(v))=\sup_{v\in\Omega_{jr}}\mathbb{E}\left|\sum_{k\in U_{jr}} v_k(\widehat{b}_{jk}-b_{jk})\right|^2.$$

Denote

$$w_{jm}=\sum_{k\in U_{jr}} v_k\psi_{mjk}\left[\sum_{l=1}^{M}N^{-2d_l}|g_m(u_l)|^2\right]^{-1}, \quad m\in C_j.$$

Then, under Assumption A2 with $\alpha_1=\alpha_2=0$, using argument similar to the proof of (5.7), one obtains

$$\begin{aligned}\sup_{v\in\Omega_{jr}}\mathrm{Var}(Z_{jr}(v)) &= \sup_{v\in\Omega_{jr}}\left\{N^{-1}\sum_{m_1,m_2\in C_j}\overline{w_{jm_1}}w_{jm_2}\mathbb{E}\left[\sum_{l=1}^{M}N^{-4d_l}\overline{g_{m_1}(u_l)}g_{m_2}(u_l)z_{lm_1}\overline{z}_{lm_2}\right]\right\}\\ &\le \sup_{v\in\Omega_{jr}} N^{-1}\sum_{l=1}^{M}N^{-4d_l}\lambda_{\max}(\boldsymbol{\Sigma}^{(l)})\sum_{m\in C_j}|w_{jm}g_m(u_l)|^2\\ &\le K_3 n^{-1}\sup_{v\in\Omega_{jr}}\left\{\sum_{m\in C_j}|w_{jm}|^2[\tau_1(m,n)]^{-1}\right\}\le 4\pi C_3^* 2^{2j\nu}j^{\vartheta_1}(n^*)^{-1}:=B_2,\end{aligned}$$

where $C_3^*=(K_3)^{-1}(\ln 2)^{\vartheta_1}(2\pi/3)^{2\nu}$. 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 $\boldsymbol{\xi}$. We need to establish upper bounds for the corresponding fourth moment and large deviation inequality.

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

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

$$\mathbf{Z}^{(l)} = \mathbf{\Phi}\boldsymbol{\xi}^{(l)} = \mathbf{\Phi}\boldsymbol{G}^{(l)}\boldsymbol{\eta}^{(l)}, \quad l = 1, \dots, M.$$

Let $\mathbf{V}^{(l)}$ be a vector in $\mathbb{R}^n$ with entries $V_m^{(l)}$, $l=1,\dots,M$, $m \in C_j$, defined in (A.7) and let ν_m be defined in (A.6). Define further vectors $\overrightarrow{\mathbf{Z}}$ and $\overrightarrow{\mathbf{V}}$ obtained by stacking vectors $\mathbf{Z}^{(l)}$ and $\mathbf{V}^{(l)}$, respectively, into one long vector:

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

Define block diagonal matrices

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

Since

$$|\widehat{b}_{jk} - b_{jk}|^2 = \sum_{l_1,l_2=1}^{M} \sum_{m_1,m_2}^{N} N^{-2d_{l_1}-2d_{l_2}} \psi_{m_1jk}\overline{\psi_{m_2jk}}\nu_{m_1}\nu_{m_2}g_{m_1}(u_{l_1})\overline{g_{m_2}(u_{l_2})}z_{lm_1}z_{lm_2},$$

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

$$\begin{aligned} |\widehat{b}_{jk} - b_{jk}|^2 &= \frac{1}{N}\left[\sum_{l=1}^{M}\sum_{m \in C_j} \tilde{\nu}_{lm} z_{lm}\right]^2 = \frac{1}{N}(\overrightarrow{\mathbf{V}}^T \overrightarrow{\mathbf{Z}})^2 = \frac{1}{N}(\overrightarrow{\mathbf{V}}^T \tilde{\mathbf{\Phi}}\boldsymbol{G}\boldsymbol{\eta})^2 \\ &= \frac{1}{N}\boldsymbol{\eta}^T \underbrace{\boldsymbol{G}^T\tilde{\mathbf{\Phi}}^T\overrightarrow{\mathbf{V}}\overrightarrow{\mathbf{V}}^T\tilde{\mathbf{\Phi}}\boldsymbol{G}}_{\mathcal{A}(k)}\boldsymbol{\eta} = \frac{1}{N}\boldsymbol{\eta}^T\mathcal{A}(k)\boldsymbol{\eta}. \end{aligned}$$

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

$$\mathcal{A} = \sum_{k \in U_{jr}} \mathcal{A}(k) = \boldsymbol{G}^T\tilde{\mathbf{\Phi}}^T\left(\sum_{k \in U_{jr}} \overrightarrow{\mathbf{V}}(k)[\overrightarrow{\mathbf{V}}(k)]^T\right)\tilde{\mathbf{\Phi}}\boldsymbol{G} \tag{A.10}$$

and

$$\hat{B}_{jr} = \sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 = \frac{1}{N}\boldsymbol{\eta}^T\mathcal{A}\boldsymbol{\eta}.$$

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

$$\begin{aligned} N\mathbb{E}(\hat{B}_{jr}) &= \mathbb{E}(\boldsymbol{\eta}^T\mathcal{A}\boldsymbol{\eta}) = \mathbb{E}[\mathrm{Tr}(\mathcal{A}\boldsymbol{\eta}\boldsymbol{\eta}^T)] = \mathrm{Tr}(\mathcal{A}) = \mathrm{Tr}\left(\sum_{k \in U_{jr}} \mathcal{A}(k)\right) \\ &= \sum_{k \in U_{jr}} \mathrm{Tr}(\boldsymbol{G}^T\tilde{\mathbf{\Phi}}^T\overrightarrow{\mathbf{V}}(k)[\overrightarrow{\mathbf{V}}(k)]^T\tilde{\mathbf{\Phi}}\boldsymbol{G}) = \sum_{k \in U_{jr}} \mathrm{Tr}\left(\tilde{\mathbf{\Phi}}^T\overrightarrow{\mathbf{V}}(k)[\overrightarrow{\mathbf{V}}(k)]^T\tilde{\mathbf{\Phi}}\underbrace{\boldsymbol{G}\boldsymbol{G}^T}_{\mathbf{\Sigma}}\right) \\ &= \sum_{k \in U_{jr}} [\overrightarrow{\mathbf{V}}(k)]^T\tilde{\mathbf{\Phi}}\mathbf{\Sigma}\tilde{\mathbf{\Phi}}^T\overrightarrow{\mathbf{V}}(k) = \sum_{k \in U_{jr}}\sum_{l=1}^{M} [\mathbf{V}^{(l)}(k)]^T\mathbf{\Phi}\mathbf{\Sigma}^{(l)}\mathbf{\Phi}^T\mathbf{V}^{(l)}(k). \end{aligned}$$

For any matrix $\boldsymbol{G}$, recall that $\|\boldsymbol{G}\|_{\mathrm{sp}}$ and $\|\boldsymbol{G}\|_2$ denote, respectively, the spectral and the Frobenius norms. Denote

$$D_{jn} = (n^*)^{-1}c_1 2^{2j\nu}j^{\vartheta_1}\ln n. \tag{A.11}$$

Then, by Assumption A1, $\|\mathbf{\Phi}\mathbf{\Sigma}^{(l)}\mathbf{\Phi}^T\|_{\mathrm{sp}} = \|\mathbf{\Phi}\|_{\mathrm{sp}}\|\mathbf{\Sigma}^{(l)}\|_{\mathrm{sp}}\|\tilde{\mathbf{\Phi}}^T\|_{\mathrm{sp}} \le K_2N^{2d_l}$. Hence,

$$\begin{aligned} \mathbb{E}(\hat{B}_{jr}) &\le \frac{K_2}{N}\sum_{k \in U_{jr}}\sum_{l=1}^{M} [\mathbf{V}^{(l)}(k)]^T\mathbf{V}^{(l)}(k)N^{2d_l} \\ &= \frac{1}{N}\sum_{k \in U_{jr}}\sum_{l=1}^{M}\sum_{m \in C_j} |\psi_{mjk}|^2|g_m(u_l)|^2\nu_m^{-2}N^{-2d_l} \le D_{jn}. \end{aligned} \tag{A.12}$$

Next, using the definition of $\mathcal{A}$, obtain

$$\|\mathcal{A}\|_{\mathrm{sp}} \le \sum_{k \in U_{jr}} \|\mathcal{A}_k\|_{\mathrm{sp}}$$

where $\mathcal{A}_k = \mathbf{u}(k)[\mathbf{u}(k)]^T$ with $\mathbf{u}_k = \mathbf{G}^T \tilde{\mathbf{\Phi}}^T \overrightarrow{\mathbf{v}}(k)$. Hence, $\mathcal{A}_k$ is of rank 1 and, consequently, $\lambda_{\max}(\mathcal{A}_k) = \|\mathcal{A}_k\|_{sp} = [\overrightarrow{\mathbf{v}}(k)]^T \tilde{\mathbf{\Phi}} \mathbf{\Sigma} (\tilde{\mathbf{\Phi}})^T \overrightarrow{\mathbf{v}}(k)$. The latter implies that $\|\mathcal{A}\|_{sp} \le \sum_{k \in U_{jr}} [\overrightarrow{\mathbf{v}}(k)]^T \tilde{\mathbf{\Phi}} \mathbf{\Sigma} (\tilde{\mathbf{\Phi}})^T \overrightarrow{\mathbf{v}}(k)$ and, hence,

$$N^{-1} \|\mathcal{A}\|_{sp} \le \mathbb{E}(\hat{B}_{jr}) \le D_{jn}. \tag{A.13}$$

Moreover, since $\mathcal{A}_k$ are matrices of rank 1, one has

$$\|\mathcal{A}\|_2 \le \sum_{k \in U_{jr}} \|\mathcal{A}_k\|_2 = \sum_{k \in U_{jr}} \|\mathcal{A}_k\|_{sp}. \tag{A.14}$$

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

$$\mathbb{E}[|\widehat{b}_{jk} - b_{jk}|^4] = \mathbb{E}\left[\left(\frac{1}{N} \boldsymbol{\eta}^T \mathcal{A}(k) \boldsymbol{\eta}\right)^2\right] \le \frac{1}{N^2} \|\mathcal{A}\|_{sp} (\boldsymbol{\eta}^T \boldsymbol{\eta})^2 \le \frac{c_1^2 2^{4j\nu} j^{2\vartheta_1} (\ln n)^2}{(n^*)^2} (\boldsymbol{\eta}^T \boldsymbol{\eta})^2.$$

Since

$$\mathbb{E}(\boldsymbol{\eta}^T \boldsymbol{\eta})^2 = \mathbb{E}\left(\sum_{l=1}^{M} \sum_{m=1}^{N} \eta_{lm}^2\right)^2 \le C n^2,$$

we derive

$$\mathbb{E}\left[|\widehat{b}_{jk} - b_{jk}|^4\right] \le \frac{C n^3}{(n^*)^2},$$

which implies (6.3).

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

$$P(|\hat{B}_{jr}| > D_{jn}(1 + \mu^2 \log n)) \le 2 n^{-\kappa}$$

provided $\mu \ge K \sqrt{c_0 \kappa}$.

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

$$\tau_1(m,n) = M^{-1} (4\pi^2 m^2)^{-1} N^{-2a_2} \sum_{l=1}^{M} q^2(l/M) \sin^2(2\pi m l M^{-1}) N^{-2a_1 l/M}.$$

Therefore,

$$(4\pi^2 m^2)^{-1} q_1^2 N^{-2a_2} S(m,n) \le \tau_1(m,n) \le (4\pi^2 m^2)^{-1} q_2^2 N^{-2a_2} S(m,n), \tag{A.15}$$

where

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

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

$$\begin{aligned} p &= \exp(-2a_1 M^{-1} \ln N) \\ &= 1 - 2a_1 M^{-1} \ln N + 2a_1^2 M^{-2} \ln^2 N + o(M^{-2} \ln^2 N), \end{aligned} \tag{A.16}$$

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 and Ryzhik (1980), we obtain

$$S(m,n) = \frac{1}{M}\left[\frac{1 - p^M}{1 - p} - \frac{1 - p \cos x - p^M \cos(Mx) + p^{M+1} \cos((M-1)x)}{1 - 2p \cos x + p^2}\right].$$

Since m is an integer and $x = 4\pi m M^{-1}$,

$$\cos(Mx) = 1, \quad \sin(Mx) = 0, \quad \cos((M-1)x) = \cos x.$$

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

$$\frac{(1-p^M)}{M(1-p)} \approx \frac{1-N^{-2a_1}}{4a_1 \ln N(1-a_1 M^{-1} \ln N)}, \tag{A.17}$$

so that, if N is large enough, due to $p<1$, one obtains an upper bound for $S(m,n)$:

$$S(m,n) = \frac{(1-p^M)}{M(1-p)} \left[\frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \right]^{-1} \le \frac{1}{2a_1 \ln N}. \tag{A.18}$$

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: $x \ge \pi/3$ and $x < \pi/3$. If $x \ge \pi/3$, then $\cos x \le 1/2$ and

$$F(p,x) = \frac{(1-p)^2}{p(p+1)(1-\cos x)} + \frac{2}{p+1} \le 2,$$

If $x < \pi/3$, we can use the fact that $1-\cos x = 2\sin^2(x/2) \ge 3x^2/8$, so that

$$F(p,x) \le \frac{4}{3}\left[1+\frac{8(1-p)^2}{3x^2}\right] \le \frac{4}{3}\left[1+\frac{2a_1^2 \ln^2 N}{3\pi^2 m^2}\right]$$

for N large enough. □

Since $|m| = C_m 2^j > C_m C_0 \ln n$ for some $C_0 > 0$ and $\ln n \ge (1-\theta_1)^{-1} \ln N(1+o(1))$ (as $n \to \infty$) due to assumption (7.1), one has $m^2 \ge C_m^2 C_0^2 (1-\theta_1)^{-2} \ln^2 N$ and

$$S(m,n) \ge C(\ln N)^{-1}. \tag{A.19}$$

Observe now that $\ln N \asymp \ln n$. This completes the proof of the theorem. □

References

Abramovich, F., Silverman, B.W., 1998. Wavelet decomposition approaches to statistical inverse problems. Biometrika 85, 115–129.
Beran, J., 1992. Statistical methods for data with long-range dependence. Statist. Sci. 4, 404–416.
Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, New York.
Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes: Probabilistic Properties and Statistical Methods. Springer-Verlag, New York.
Bunea, F., Tsybakov, A., Wegkamp, M.H., 2007. Aggregation for Gaussian regression. Ann. Statist. 35, 1674–1697.
Box, G.E.P., Jenkins, G.M., 1970. Times Series Analysis. Forecasting and Control. Holden-Day, San Francisco, CA, London, Amsterdam.
Cheng, B., Robinson, P.M., 1994. Semiparametric estimation from time series with long-range dependence. J. Econom. 64, 335–353.
Cirelson, B.S., Ibragimov, I.A., Sudakov, V.N., 1976. Norm of Gaussian sample function. In: Proceedings of the 3rd Japan-U.S.S.R. Symposium on Probability Theory. Lecture Notes in Mathematics, vol. 550. Springer-Verlag, Berlin, pp. 20–41.
Comte, F., Dedecker, J., Taupin, M.L., 2008. Adaptive density deconvolution with dependent inputs. Math. Methods Statist. 17, 87–112.
Csörgo, S., Mielniczuk, J., 2000. The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. Statist. Sinica 10, 771–787.
De Canditiis, D., Pensky, M., 2004. Discussion on the meeting on "Statistical Approaches to Inverse Problems". J. Roy. Statist. Soc. Ser. B 66, 638–640.
De Canditiis, D., Pensky, M., 2006. Simultaneous wavelet deconvolution in periodic setting. Scand. J. Statist. 33, 293–306.
Donoho, D.L., 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl. Comput. Harmon. Anal. 2, 101–126.
Doukhan, P., Oppenhein, G., Taqqu, M.S., 2003. Theory and Applications of Long-Range Dependence. Birkhaüser, Boston.
Fan, J., Koo, J., 2002. Wavelet deconvolution. IEEE Trans. Inform. Theory 48, 734–747.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Ser. Anal. 4, 221–238.
Golubev, G., 2004. The principle of penalized empirical risk in severely ill-posed problems. Probab. Theory Related Fields 130, 18–38.
Golubev, G.K., Khasminskii, R.Z., 1999. A statistical approach to some inverse problems for partial differential equations. Problems Inform. Transmission 35, 136–149.
Grenander, U., Szegö, G., 1958. Toeplitz Forms and Their Applications. California Monographs in Mathematical Sciences. University of California Press, Berkeley, Los Angeles.
Gradshtein, I.S., Ryzhik, I.M., 1980. Tables of Integrals, Series, and Products. Academic Press, New York.
Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A., 1998. Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics, vol. 129. Springer-Verlag, New York.
Harsdörf, S., Reuter, R., 2000. Stable deconvolution of noisy lidar signals. In: Proceedings of EARSeL-SIG-Workshop LIDAR, Dresden/FRG, June 16–17.
Johnstone, I.M., Kerkyacharian, G., Picard, D., Raimondo, M., 2004. Wavelet deconvolution in a periodic setting. Journal of the Royal Statistical Society, Series B 66, 547–573. (with discussion, 627–657).
Johnstone, I.M., Raimondo, M., 2004. Periodic boxcar deconvolution. Diophantine approximation. Ann. Stat. 32, 1781–1804.
Kalifa, J., Mallat, S., 2003. Thresholding estimators for linear inverse problems and deconvolutions. Ann. Stat. 31, 58–109.
Kerkyacharian, G., Picard, D., Raimondo, M., 2007. Adaptive boxcar deconvolution on full Lebesgue measure sets. Statistica Sinica 7, 317–340.
Kulik, R., 2008. Nonparametric deconvolution problem for dependent sequences. Electron. J. Statist. 2, 722–740.
Kulik, R., Raimondo, M., 2009. L^p wavelet regression with correlated errors. Statist. Sinica 19, 1479–1489.
Lattes, R., Lions, J.L., 1967. Methode de Quasi-Reversibilite et Applications. Travoux et Recherche Mathematiques, vol. 15. Dunod, Paris.
Mallat, S., 1999. A Wavelet Tour of Signal Processing, 2nd edition Academic Press, San Diego.
Meyer, Y., 1992. Wavelets, Operators, Cambridge University Press, Cambridge.
Mielniczuk, J., Wu, W.B., 2004. On random-design model with dependent errors. Statist. Sinica 14, 1105–1126.

Navarro, F., Chesneau, C., Fadili, J., Sassi, T., 2013. Block thresholding for wavelet-based estimation of function derivatives from a heteroscedastic multichannel convolution model. Electron. J. Statist. 7, 428–453.
Park, Y.J., Dho, S.W., Kong, H.J., 1997. Deconvolution of long-pulse lidar signals with matrix formulation. Appl. Opt. 36, 5158–5161.
Pensky, M., Sapatinas, T., 2009. Functional deconvolution in a periodic setting: uniform case. Ann. Statist. 37, 73–104.
Pensky, M., Sapatinas, T., 2010. On convergence rates equivalency and sampling strategies in a functional deconvolution model. Ann. Statist. 38, 1793–1844.
Pensky, M., Sapatinas, T., 2011. Multichannel boxcar deconvolution with growing number of channels. Electron. J. Statist. 5, 53–82.
Pensky, M., Vidakovic, B., 1999. Adaptive wavelet estimator for nonparametric density deconvolution. Ann. Statist. 27, 2033–2053.
Rudelson, M., Vershynin, R., 2013. Hanson-Wright inequality and sub-Gaussian concentration. arXiv:1306.2872.
Tsybakov, A.B., 2008. Introduction to Nonparametric Estimation. Springer-Verlag, New York.
Van Zanten, H., Zareba, P., 2008. A note on wavelet density deconvolution for weakly dependent data. Statist. Inference Stochastic Process. 11, 207–219.
Vershynin, R., 2011. Introduction to the non-asymptotic analysis of random matrices. arXiv:1011.3027v7.
Walter, G., Shen, X., 1999. Deconvolution using Meyer wavelets. J. Integral Equations Appl. 11, 515–534.
Wang, Y., 1997. Minimax estimation via wavelets for indirect long-memory data. J. Statist. Plann. Inference 64, 45–55.
Wishart, J.M., 2013. Wavelet deconvolution in a periodic setting with long-range dependent errors. J. Statist. Plann. Inference 143, 867–881.