Statistics and Probability Letters 79 (2009) 1568–1576

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

# Minimax convergence rates under the $L^p$-risk in the functional deconvolution model

Athanasia Petsa, Theofanis Sapatinas *

*Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus*

ARTICLE INFO

*Article history:*
Received 4 October 2008
Received in revised form 24 March 2009
Accepted 25 March 2009
Available online 1 April 2009

ABSTRACT

We derive minimax results in the functional deconvolution model under the $L^p$-risk, $1 \leq p < \infty$. Lower bounds are given when the unknown response function is assumed to belong to a Besov ball and under appropriate smoothness assumptions on the blurring function, including both regular-smooth and super-smooth convolutions. Furthermore, we investigate the asymptotic minimax properties of an adaptive wavelet estimator over a wide range of Besov balls. The new findings extend recently obtained results under the $L^2$-risk. As an illustration, we discuss particular examples for both continuous and discrete settings.



## 1. Introduction

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

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

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

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

* Corresponding author.
*E-mail addresses:* map6pa2@ucy.ac.cy (A. Petsa), T.Sapatinas@ucy.ac.cy (T. Sapatinas).

0167-7152/$ – see front matter 
doi:10.1016/j.spl.2009.03.028

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

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

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

$$y(u_l,t_i) = \int_T f(x)g(u_l,t_i-x)\mathrm{d}x + \epsilon_{li}, \quad t_i = \frac{i}{N} \in T = [0,1],\ u_l \in U = [a,b], \tag{2}$$

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

Pensky and Sapatinas (2009) obtained minimax lower bounds and proposed an adaptive (linear or block thresholding) wavelet estimator, for both the functional deconvolution model (1) and its discrete version (2), that is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, under the $L^2$-risk over a wide range of Besov balls.

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

In what follows, as in Pensky and Sapatinas (2009), we assume that for a fixed $u \in [a,b]$, both $f(\cdot)$ and $g(u,\cdot)$ are periodic functions with period on the unit interval $T = [0,1]$; this assumption appears naturally in the above-mentioned special models which (1) and (2) generalize.

## 2. Meyer wavelets and Besov balls

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

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

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

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

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

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

It is well known that the Meyer wavelet basis satisfies the following three properties (see, e.g., Johnstone et al. (2004)):

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

$$\left\| \sum_{k=0}^{2^j-1} u_{j,k}h_{j,k} \right\|_p^p \le c2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |u_{j,k}|^p. \tag{3}$$

   (Here, and in what follows, $\|g\|_p$ refers to the $L^p$-norm of a function $g(\cdot)$.)
2. **Property of unconditionality.** Let $p \in (1,\infty)$. Let us set $\psi_{\tau-1,k} = \phi_{\tau,k}$. For any sequence $u = (u_{jk})_{j,k}$, we have

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

   (Here, and in what follows, the notation $a \asymp b$ means there exist two positive constants $c_1$ and $c_2$ such that $c_1b \le a \le c_2b$.)

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

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

**Remark 2.1.** The property of concentration is used in the proof of Theorem 4.2, in the case of super-smooth convolutions. The property of unconditionality and Temlyakov property are indirectly used in the proof of Theorem 4.2, since they are used in the proofs of some auxiliary results (i.e., Theorems 5.4.1 and 5.4.2 in Chesneau (2006)).

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

$$\left( \sum_{j=\tau-1}^{\infty} \left( 2^{j(s+1/2-1/\rho)} \left( \sum_{k=0}^{2^j-1} |\beta_{jk}|^\rho \right)^{1/\rho} \right)^r \right)^{1/r} \le M,$$

with respective sum(s) replaced by maximum when $\rho = \infty$ and/or $r = \infty$.

## 3. Construction of the wavelet estimator

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

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

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

$$h(u,t) = \int f(x) g(u, t-x) \mathrm{d}x, \quad t \in T = [0,1], \qquad u \in U = [a,b], \tag{4}$$

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

$$h_m(u) = \langle e_m, h(u,\cdot) \rangle, \quad y_m(u) = \langle e_m, y(u,\cdot) \rangle, \qquad g_m(u) = \langle e_m, g(u,\cdot) \rangle, \qquad z_m(u) = \langle e_m, z(u,\cdot) \rangle.$$

Using the properties of the Fourier transform, then for each $u \in U$, for the continuous model (1), we have

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

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

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

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

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

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

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

A natural estimator of $f_m$ is given by

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

(Here, and in what follows, $\overline{h}$ denotes the conjugate of a complex number or a complex function $h$; $h$ is real if and only if $\overline{h} = h$.)

Consider also the following assumptions on the blurring function $g(\cdot,\cdot)$. Define

$$\tau_1(m) = \begin{cases} \int_a^b |g_m(u)|^2 \mathrm{d}u, & \text{in the continuous case,} \\ \frac{1}{M}\sum_{l=1}^{M} |g_m(u_l)|^2, & \text{in the discrete case,} \end{cases}$$

and suppose that, for some constants $\nu \in \mathbb{R}$, $\alpha \geq 0$ (with $\nu > 0$ if $\alpha = 0$), $\beta > 0$ and some constants $K_1$ and $K_2$, independent of $m$, the choice of $M$ and the selection of points $u_l$, $l = 1, 2, \ldots, M$, with $0 < K_1 \leq K_2$,

$$\tau_1(m) \leq K_2 |m|^{-2\nu} \exp(-\alpha |m|^{\beta}), \tag{5}$$

and

$$\tau_1(m) \geq K_1 |m|^{-2\nu} \exp(-\alpha |m|^{\beta}). \tag{6}$$

Following standard terminology, $\alpha = 0$ corresponds to *regular-smooth* and $\alpha > 0$ corresponds to *super-smooth* blurring functions $g(\cdot,\cdot)$. Define also

$$\begin{aligned} &2^{j_0} = 2^J = \frac{3}{8\pi}\left(\frac{\log(n)}{2\alpha}\right)^{1/\beta}, \quad \alpha > 0, \\ &2^{j_0} = [\log(n)^{\max(p/2,1)}], \quad 2^J = n^{\delta}, \alpha = 0, \end{aligned} \tag{7}$$

where $\delta \in (0, (2\nu+1)^{-1}]$. (Here, and in what follows, $[x]$ denotes the integer part of $x$.)

By Plancherel's formula, the scaling coefficients, $\alpha_{j_0k}$, and the wavelet coefficients, $\beta_{jk}$, can be represented as

$$\alpha_{j_0k} = \sum_{m \in C^*_{j_0}} f_m \overline{\phi_{mj_0k}}, \qquad \beta_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}},$$

where $C^*_{j_0} = \{m : \phi_{mj_0k} \neq 0\}$ and, for all $j \geq j_0$, $C_j = \{m : \psi_{mjk} \neq 0\}$, both subsets of $2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]$, due to the fact that Meyer wavelets are band limited (see, e.g., Johnstone et al. (2004), Section 3.1). Hence, $\alpha_{j_0k}$ and $\beta_{jk}$, are naturally estimated by

$$\hat{\alpha}_{j_0k} = \sum_{m \in C^*_{j_0}} \hat{f}_m \overline{\phi_{mj_0k}} \qquad \hat{\beta}_{jk} = \sum_{m \in C_j} \hat{f}_m \overline{\psi_{mjk}}. \tag{8}$$

We now construct a wavelet (linear or block thresholding) estimator of $f(\cdot)$. For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length $l_j$. More specifically, let the following set of indices

$$A_j = \{1, 2, \ldots, 2^j/l_j\}, \qquad U_{jt} = \{k = 0, 1, \ldots, 2^j - 1 \mid (t-1)l_j \leq k \leq tl_j - 1\}$$

and let

$$l_j \asymp [\log(n)^{\max(p/2,1)}], \qquad \hat{B}_{jt} = \Big(\sum_{k \in U_{jt}} |\hat{\beta}_{jk}|^p / l_j\Big)^{1/p}.$$

For any $j_0 \geq 0$, we finally reconstruct $f(\cdot)$ as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{J-1} \sum_{t \in A_j} \sum_{k \in U_{jt}} \hat{\beta}_{jk} \mathbb{I}\big(|\hat{B}_{jt}| \geq d2^{j\nu} n^{-1/2}\big)\psi_{jk}(t), \tag{9}$$

where $\mathbb{I}(A)$ is the indicator function of the set $A$. (Since $j_0 > J - 1$ when $\alpha > 0$, the estimator (9) only consists of the first (linear) part and, hence, a threshold parameter does not need to be selected in this case.)

Note that since the choices of $j_0$, $J$ and the threshold value are independent of the parameters $s$, $\rho$, $r$ and $M$ (that are usually unknown in practical situations) of the Besov ball $B^s_{\rho,r}(M)$, the wavelet estimator (9) is *adaptive* with respect to these parameters.

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

## 4. Main results

We construct below minimax lower bounds for the $L^p$-risk, $1 \le p < \infty$, both for the continuous model (1) and the discrete model (2). For this purpose, we define the minimax $L^p$-risk, $1 \le p < \infty$, over the set $\Omega$ as

$$R_n(\Omega) = \inf_{\tilde{f}_n} \sup_{f \in \Omega} \mathbb{E}\|\tilde{f}_n - f\|_p^p,$$

where $\|g\|_p$ is the $L^p$-norm, $1 \le p < \infty$, of a function $g(\cdot)$ and the infimum is taken over all possible estimators $\tilde{f}_n(\cdot)$ (measurable functions) of $f(\cdot)$, based on observations either from the continuous model (1) or the discrete model (2).

The following theorem provides the minimax lower bounds for the $L^p$-risk, $1 \le p < \infty$, under assumption (5).

**Theorem 4.1.** *Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 2. Let $s > \max(0, 1/\rho - 1/2)$, $1 \le \rho \le \infty$, $1 \le r \le \infty$ and $M > 0$. Then, under the assumption (5), as $n \to \infty$, there exists some constant $C > 0$ such that,*

$$R_n(B^s_{\rho,r}(M)) \ge \begin{cases} C(\log n)^{-ps^*/\beta}, & \text{if } \alpha > 0, \\ Cn^{-\alpha_1 p}, & \text{if } \alpha = 0,\ \epsilon > 0, \\ C\left(\dfrac{\log n}{n}\right)^{\alpha_2 p}, & \text{if } \alpha = 0,\ \epsilon \le 0, \end{cases} \tag{10}$$

*where*

$$\alpha_1 = \frac{s}{2(s+\nu)+1}, \qquad \alpha_2 = \frac{s - 1/\rho + 1/p}{2(s - 1/\rho + \nu) + 1}, \qquad \epsilon = s\rho + \frac{2\nu+1}{2}(\rho - p)$$
*and* $s^* = s + 1/p - 1/\min(p, \rho)$.

**Remark 4.1.** The two different lower bounds for $\alpha = 0$ in (10) refer to the dense case ($\epsilon > 0$) when the worst functions $f(\cdot)$ (i.e., the hardest functions to estimate) are spread uniformly over the unit interval $T$, and the sparse case ($\epsilon \le 0$) when the worst functions $f(\cdot)$ have only one non-vanishing wavelet coefficient. Also, the restriction $s > \max(0, 1/\rho - 1/2)$, $1 \le \rho \le \infty$, $1 \le r \le \infty$, that appears in the statement of Theorem 4.1, ensures that the corresponding Besov spaces are embedded in $L^2(T)$.

The next theorem provides the minimax upper bounds for the adaptive (with respect to the Besov parameters) wavelet estimator given by (9), under the assumption (6).

**Theorem 4.2.** *Let $\hat{f}_n(\cdot)$ be the adaptive wavelet estimator defined by (9), with $j_0$, $J$ and $\delta$ given by (7). Let $s > 1/\rho - 1/2 + 1/(2\delta) - \nu$ if $\alpha = 0$ and $s > 1/\rho$ if $\alpha > 0$, $1 \le \rho \le \infty$, $1 \le r \le \infty$ and $M > 0$. Then, under assumption (6), as $n \to \infty$, there exists some constant $C > 0$ such that,*

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}(\|\hat{f}_n - f\|_p^p) \le \begin{cases} C(\log n)^{-ps^*/\beta}, & \text{if } \alpha > 0, \\ Cn^{-\alpha_1 p}(\log n)^{\alpha_1 p \mathbb{I}_{\{p > \rho\}}}, & \text{if } \alpha = 0,\ \epsilon > 0, \\ C\left(\dfrac{\log n}{n}\right)^{\alpha_2 p} (\log n)^{\max(0, p - \rho/r)\mathbb{I}_{\{\epsilon = 0\}}}, & \text{if } \alpha = 0,\ \epsilon \le 0, \end{cases} \tag{11}$$

*where $\alpha_1$, $\alpha_2$, $\epsilon$ and $s^*$ as in Theorem 4.1.*

**Remark 4.2.** Theorems 4.1 and 4.2 imply that, for the $L^p$-risk, $1 \le p < \infty$, the estimator $\hat{f}_n(\cdot)$ defined by (9) is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, over a wide range of Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho - 1/2 + 1/(2\delta) - \nu$ if $\alpha = 0$ and $s > 1/\rho$ if $\alpha > 0$, $1 \le \rho \le \infty$ and $1 \le r \le \infty$. In particular, the estimator (9) is asymptotically optimal, except for the cases (i) $\alpha = 0, \epsilon = 0, p > \rho/r$ and (ii) $\alpha = 0, \epsilon > 0, p > \rho$; in these latter cases, the estimator $\hat{f}_n(\cdot)$ defined by (9) is asymptotically near-optimal within a logarithmic factor, i.e.,

$$R_n(B^s_{\rho,r}(M)) \asymp \begin{cases} (\ln n)^{-ps^*/\beta}, & \text{if } \alpha > 0, \\ n^{-\alpha_1 p}, & \text{if } \alpha = 0,\ \epsilon > 0,\ p \le \rho, \\ \left(\dfrac{\ln n}{n}\right)^{\alpha_2 p}, & \text{if } \alpha = 0,\ \epsilon < 0, \\ & \text{or } \alpha = 0,\ \epsilon = 0,\ p \le \rho/r \end{cases}$$

and

$$\sup_{f \in B^s_{\rho,r}(M)} \mathbb{E}\|\hat{f}_n - f\|_p^p \le \begin{cases} Cn^{-\alpha_1 p}(\log n)^{\alpha_1 p}, & \text{if } \alpha = 0,\ \epsilon > 0,\ p > \rho, \\ C\left(\dfrac{\ln n}{n}\right)^{\alpha_2 p} (\log n)^{(p - \frac{\rho}{r})}, & \text{if } \alpha = 0,\ \epsilon = 0,\ p > \rho/r. \end{cases}$$

(Here, and in similar expressions, we abuse notation, and $g_1(n) \asymp g_2(n)$ denotes $0 < \liminf(g_1(n)/g_2(n)) \leq \limsup(g_1(n)/g_2(n)) < \infty$ as $n \to \infty$.) Note that since the constant $C$ in Theorems 4.1 and 4.2 is different, it means that the estimator $\hat{f}_n(\cdot)$ defined by (9) is *rate optimal*.

## 5. Examples

In this section, we briefly present inverse problems discussed in Section 1 which can be seen as applications of the functional deconvolution model (1) or its discrete version (2). The optimality (in the minimax sense), or near-optimality within a logarithmic factor, for the $L^2$-risk over a wide range of Besov balls in the Examples 1–3 below have been discussed in Pensky and Sapatinas (2009) (see their Examples 4, 1, 5, respectively); here, we use the methodology presented in Sections 3 and 4 to check that the corresponding estimators are also optimal or near-optimal under the $L^p$-risk ($1 \leq p < \infty$).

**Example 1** (*Estimation of the Speed of a Wave on a Finite Interval*). Let $h(t, x)$ be a solution of the initial boundary value problem for the wave equation

$$\frac{\partial^2 h(t,x)}{\partial t^2} = \frac{\partial^2 h(t,x)}{\partial x^2} \quad \text{with } h(0,x) = 0,$$
$$\left.\frac{\partial h(t,x)}{\partial t}\right|_{t=0} = f(x), \quad h(t,0) = h(t,1) = 0. \tag{12}$$

Here, $f(\cdot)$ is a function defined on the unit interval $[0, 1]$ and $t \in [a, b]$, $a > 0$, $b < 1$. We assume that a noisy solution $y(t, x) = h(t, x) + n^{-1/2}z(t, x)$ is observed, where $z(t, x)$ is a generalized two-dimensional Gaussian field with covariance function $\mathbb{E}[z(t_1, x_1)z(t_2, x_2)] = \delta(t_1 - t_2)\delta(x_1 - x_2)$, and the goal is to recover the unknown speed of a wave $f(\cdot)$ on the basis of observations $y(t, x)$.

Extending $f(\cdot)$ periodically over the real line, it is well known (see, e.g., Strauss (1992), p. 61) that the solution $h(t, x)$ can then be recovered as

$$h(t,x) = \frac{1}{2}\int_0^1 \mathbb{I}(|x-z| < t) f(z) \mathrm{d}z, \tag{13}$$

so that (13) takes the form (4) with $g(u, x) = 0.5\,\mathbb{I}(|x| < u)$, where $u$ in (4) is replaced by $t$ in (13). It is easily seen that the functional Fourier coefficients $g_m(\cdot)$ satisfy (5) and (6) with $\nu = 1$ and $\alpha = 0$.

Hence, according to Theorems 4.1 and 4.2, the adaptive block thresholding wavelet estimator given by (9) achieves the following minimax upper bounds (in the $L^p$-risk, $1 \leq p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \leq \begin{cases} n^{-\frac{sp}{2s+3}} \, (\ln n)^{\frac{sp}{2s+3}}, & \text{if } s > \frac{3}{2}(1 - p/\rho), \\ \left(\frac{\ln n}{n}\right)^{\frac{p(s-1/\rho+1/p)}{2s-2/\rho+3}} (\ln n)^{\max(0, p-\rho/r)\mathbb{I}(\epsilon=0)}, & \text{if } s \leq \frac{3}{2}(1 - p/\rho), \end{cases}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho - 1/2 - 1/(2\delta) + \nu$, $1 \leq \rho \leq \infty$ and $1 \leq r \leq \infty$. (The minimax lower bounds (in the $L^p$-risk, $1 \leq p < \infty$) have the same form without the extra logarithmic factor.)

**Example 2** (*Estimation of the Initial Condition in the Heat Conductivity Equation*). Let $h(t, x)$ be a solution of the heat conductivity equation

$$\frac{\partial h(t,x)}{\partial t} = \frac{\partial^2 h(t,x)}{\partial x^2}, \quad x \in [0,1], \ t \in [a,b], \ a > 0, \ b < \infty,$$

with initial condition $h(0, x) = f(x)$ and periodic boundary conditions

$$h(t,0) = h(t,1), \qquad \left.\frac{\partial h(t,x)}{\partial x}\right|_{x=0} = \left.\frac{\partial h(t,x)}{\partial x}\right|_{x=1}.$$

Again, suppose that a noisy solution $y(t, x) = h(t, x) + n^{-1/2}z(t, x)$ is observed, where $z(t, x)$ is as in Example 1, and the goal is to recover the unknown initial condition $f(\cdot)$ on the basis of observations $y(t, x)$.

It is well known (see, e.g., Strauss (1992), p. 48) that, under the assumption of periodicity, the solution $h(t, x)$ is given by

$$h(t,x) = (4\rho t)^{-1/2} \int_0^1 \sum_{k\in\mathbb{Z}} \exp\left\{-\frac{(x+k-z)^2}{4t}\right\} f(z) \mathrm{d}z, \tag{14}$$

which coincides with (4) when $t$ and $x$ are replaced by $u$ and $t$, respectively. It is easily seen that the functional Fourier coefficients $g_m(\cdot)$ satisfy (5) and (6) with $\nu = 1$, $\alpha = 8\pi^2 a$ and $\beta = 2$.

Hence, according to Theorems 4.1 and 4.2, the adaptive wavelet estimator given by (9) achieves the following minimax convergence rates (in the $L^p$-risk, $1 \le p < \infty$)

$$R_n(B^s_{\rho,r}(M)) \asymp (\ln n)^{-\frac{p}{2}(s+1/p-1/\min(p,\rho))}$$

over Besov balls $B^s_{\rho,r}(M)$ of radius $M > 0$ with $s > 1/\rho$, $1 \le \rho \le \infty$ and $1 \le r \le \infty$.

**Example 3** (*Estimation in the Multichannel Ddeconvolution Problem*). Consider the problem of estimating $f(\cdot) \in L^p(T)$ on the basis of the following model

$$\mathrm{d}Y_l(t) = f * g_l(t)\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W_l(t), \quad t \in T = [0, 1],\ l = 1, 2, \ldots, M, \tag{15}$$

where $g_l(\cdot)$ are known blurring functions and $W_l(t)$ are independent standard Wiener processes.

Adaptive term-by-term wavelet thresholding estimators for the model (15) were constructed in De Canditiis and Pensky (2006) for regular-smooth convolutions (i.e., $\alpha = 0$ in (5) and (6)), over a wide range of Besov balls. However, minimax lower and upper bounds were not obtained by these authors who concentrate instead on upper bounds (in the $L^p$-risk, $1 < p < \infty$) for the error, for a *fixed* target function (using the *maxiset* approach). Moreover, the case of super-smooth convolutions (i.e., $\alpha > 0$ in (5) and (6)) and the case when $M$ can increase together with $N$ have not been treated in De Canditiis and Pensky (2006).

Consider now the adaptive wavelet estimator $\hat{f}_n(\cdot)$ defined by (9) for the continuous model (1) or the discrete model (2). Then, under the assumption (6), the corresponding minimax lower bounds are given by Theorem 4.1, while, under the assumption (5), the corresponding minimax upper bounds are given by Theorem 4.2. Thus, the proposed functional deconvolution methodology significantly expands on the theoretical findings in De Canditiis and Pensky (2006).

## Acknowledgements

The authors wish to express their thanks to Professors Christophe Chesneau and Marianna Pensky for useful discussions. Helpful comments of a reviewer on improvements to this paper are also gratefully acknowledged.

## Appendix. Proofs

**Proof of Theorem 4.1.** The proof follows by working along the lines of the proof of Theorem 1 in Pensky and Sapatinas (2009), with necessary modifications, hence we omit the details. (For the details, we refer to Petsa (2009).)

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

**Proof of Theorem 4.2.** For the proof of Theorem 4.2, we are going to use Theorems 5.4.1 and 5.4.2 in Chesneau (2006). The important assumptions in these theorems are stated below:

(F1) Let us set $\hat{\beta}_{j_0-1,k} = \hat{\alpha}_{j_0k}$. There exist some constant $c > 0$ such that, for all $j \in \{j_0 - 1, j_0, \ldots, J\}$, $k \in \{0, 1, \ldots, 2^j - 1\}$ and $n$ sufficiently large,

$$\mathbb{E}(|\hat{\beta}_{jk} - \beta_{jk}|^{2p}) \le c2^{2jp\nu}n^{-p}.$$

(F2) There exist two constants $d > 0$ and $c > 0$ such that, for $j \in \{j_0, j_0 + 1, \ldots, J\}$, $t \in A_j$ and $n$ sufficiently large,

$$\mathbb{P}\left(\frac{1}{l_j}\sum_{k\in U_{jt}} |\hat{\beta}_{jk} - \beta_{jk}| \ge 2^{-1}d2^{j\nu}n^{-1/2}\right) \le cn^{-p}.$$

We show below that Assumptions (F1) and (F2) hold in order to apply Theorems 4.5.1 and 4.5.2 in Chesneau (2006), for the case $a = 0$.

**Assumption (F1).** Using the theory of generalized random fields, it is easy to check that $\hat{\alpha}_{j_0k} - \alpha_{j_0k}$ is a centered Gaussian random variable, with

$$\mathrm{Var}(\hat{\alpha}_{j_0k} - \alpha_{j_0k}) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2\left(\int_a^b |g_m(u)|^2\mathrm{d}u\right)^{-1}, & \text{for the continuous model,} \\[2ex] \dfrac{1}{NM}\displaystyle\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2\left(\frac{\sum_{l=1}^M |g_m(u_l)|^2}{M}\right)^{-1}, & \text{for the discrete model.} \end{cases}$$

Under assumption (6), it is easy to see that

$$\mathrm{Var}(\hat{\alpha}_{j_0k}-\alpha_{j_0k}) \le 2^{-j_0}\sum_{m\in C^*_{j_0}} |\phi_{mj_0k}|^2 \frac{\tau_1^{-1}(m)}{n} \le 2^{-j_0}\frac{K_1}{n}\sum_{m\in C^*_{j_0}} |m|^{2\nu} \le \frac{c2^{2j_0\nu}}{n},$$

for both the continuous model (1) and the discrete model (2). Using the same arguments, it is easy to see that, for each $j \ge j_0$, $\hat{\beta}_{jk} - b_{jk}$ are also centered Gaussian random variables with variance

$$\mathrm{Var}\big(\hat{\beta}_{jk}-\beta_{jk}\big) \le \frac{c2^{2j\nu}}{n}.$$

Therefore, the following inequalities hold

$$\mathbb{E}(|\hat{\alpha}_{j_0k}-\alpha_{j_0k}|^p) \le c_p\big(\mathrm{Var}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\big)^{p/2} \le \frac{c2^{j_0\nu p}}{n^{p/2}},$$
$$\mathbb{E}(|\hat{\beta}_{jk}-\beta_{jk}|^{2p}) \le c_p(\mathrm{Var}(\hat{\beta}_{jk}-\beta_{jk}))^p \le \frac{c2^{2p\nu j}}{n^p}.$$

**Assumption (F2).** We first show that Assumption (F2) holds for $p \ge 2$. It is sufficient to show that

$$\mathbb{P}\left(\left(\frac{1}{l_j}\sum_{k\in U_{j,t}} |\hat{\beta}_{jk}-\beta_{jk}|^p\right)^{1/p} \ge \frac{\mathrm{d}2^{j\nu}n^{-1/2}}{2}\right) \le cn^{-p}. \tag{16}$$

Consider the centered Gaussian process

$$Z_{jt} = \sum_{k\in U_{jt}} v_k(\hat{\beta}_{jk}-\beta_{jk}),$$

where $v_k \in \Omega_q = \big\{v_k : k \in U_{jt} \text{ and } \sum_{k\in U_{jt}} |v_k|^q \le 1\big\}$ and $\frac{1}{p}+\frac{1}{q} = 1$. By the duality argument used by Chesneau (2008), we get

$$\sup_{v\in\Omega_q} Z_{jt}(v) = \left(\sum_{k\in U_{jt}} |\hat{\beta}_{jk}-\beta_{jk}|^p\right)^{1/p}.$$

Thus, Jensen's inequality and Assumption (F1) lead to

$$\mathbb{E}(\sup_{v\in\Omega_q} Z_{jt}(v)) = \mathbb{E}\left(\sum_{k\in U_{jt}} |\hat{\beta}_{jk}-\beta_{jk}|^p\right)^{1/p} \le \left(\sum_{k\in U_{jt}} \mathbb{E}\left(|\hat{\beta}_{jk}-\beta_{jk}|^p\right)\right)^{1/p} \le cl_j^{1/p}n^{-1/2}2^{\nu j} := V_1.$$

Under assumption (6), it is easy to see that

$$\mathbb{E}\left((\hat{\beta}_{jk}-\beta_{jk})\overline{(\hat{\beta}_{jk}-\beta_{jk})}\right) = \sum_{m\in C_j} \psi_{mjk}\overline{\psi}_{mjk'}\frac{\tau_1^{-1}(m)}{n}. \tag{17}$$

Hence, using (17), $\psi_{mjk}\psi_{mjk'} = 0$ for $k \ne k'$ and $\sum_{m\in C_j} |\psi_{mjk}|^2 = 1$, we arrive at

$$\begin{aligned}\sup_{v\in\Omega_q} \mathrm{Var}\Big(Z_{jt}(v)\Big) &= \sup_{v\in\Omega_q}\sum_{k\in U_{jt}}\sum_{k'\in U_{jt}} v_k\overline{v}_{k'}\mathbb{E}(\hat{\beta}_{jk}-\beta_{jk})\overline{(\hat{\beta}_{jk'}-\beta_{jk'})}\\ &\le \frac{K_1 2^{2j\nu}}{n}\sup_{v\in\Omega_q}\sum_{k\in U_{jt}} |v_k|^2 \le K_1\frac{2^{2\nu j}}{n} := V_2.\end{aligned} \tag{18}$$

To continue the proof of Theorem 4.2, we are going to use Lemmas 2 and 5 in Pensky and Sapatinas (2009). Applying Lemma 5 in Pensky and Sapatinas (2009) with $x = \frac{\mathrm{d}n^{-1/2}l_j^{1/p}2^{\nu j}}{4}$, $V_1 = cl_j^{1/p}n^{-1/2}2^{\nu j}$, $V_2 = K_1\frac{2^{2\nu j}}{n}$ and $d$ sufficiently large, we have

$$\begin{aligned}\mathbb{P}\left(\left(\frac{1}{l_j}\sum_{k\in U_{jt}} |\hat{\beta}_{jk}-\beta_{jk}|^p\right)^{1/p} \ge 2^{-1}2^{\nu j}\mathrm{d}n^{-1/2}\right) &= \mathbb{P}\left(\sup_{v\in\Omega_q} Z(v) \ge l_j^{1/p}2^{-1}2^{j\delta}dn^{-1/2}\right)\\ &\le \mathbb{P}\left(\sup_{v\in\Omega_q} Z(v) \ge x+V_1\right)\\ &\le \exp\left(-\frac{x^2}{2V_2}\right) \le \exp(-cd^2\log n) \le n^{-p}.\end{aligned}$$

Now, we show that Assumption (F2) holds for $1 \le p < 2$. It is easy to see that the following inequality holds

$$\mathbb{P}\left(\left(\frac{1}{l_j}\sum_{k\in U_{jt}}|\hat{\beta}_{jk}-\beta_{jk}|^p\right)^{1/p} \ge 0.5d2^{j\nu}n^{-1/2}\right) \le \mathbb{P}\left(\left(\frac{1}{l_j}\sum_{k\in U_{jt}}|\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{1/2} \ge 0.5d2^{j\nu}n^{-1/2}\right). \tag{19}$$

In order to complete the proof of Theorem 4.2, we now apply Lemma 2 in Pensky and Sapatinas (2009) which, combining with (19), gives (16). Hence, we have shown that (F1) and (F2) are satisfied for all $1 \le p < \infty$. Applying Theorem 5.4.1 and 5.4.2 in Chesneau (2008), we obtain the required upper bounds.

For the case $\alpha > 0$, the estimator is given by $\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1}\hat{\alpha}_{j_0k}\phi_{j_0k}(t)$. Minkowski's inequality leads to

$$\mathbb{E}(\|\hat{f}_n - f\|_p^p) \le 2^{p-1}\mathbb{E}\left(\left\|\sum_{k=0}^{2^{j_0}-1}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\phi_{j_0k}\right\|_p^p\right) + 2^{p-1}\left\|\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_p^p. \tag{20}$$

Additionally, using the property of concentration (3) and the definition of $j_0$, we have

$$\mathbb{E}\left(\left\|\sum_{k=0}^{2^{j_0}-1}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\phi_{j_0k}\right\|_p^p\right) \le c(\log n)^{p(\nu+1/2)/\beta}n^{-p/4} = o\big((\log n)^{-ps^*/\beta}\big), \tag{21}$$

and

$$\left\|\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_p^p \le \left(\sum_{j=j_0}^{\infty}c2^{-j(s+1/p-1/\min(p,\rho))}\right)^p = c(\log n)^{-\frac{p}{\beta}(s+1/p-1/\min(\rho,p))}. \tag{22}$$

Inequalities (20)–(22) lead to the optimal rate of convergence for the case $\alpha > 0$.

## References


Abramovich, F., Silverman, B.W., 1998. Wavelet decomposition approaches to statistical inverse problems. Biometrika 85, 115–129.

Casey, S.D., Walnut, D.F., 1994. Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms. SIAM Review 36, 537–577.

Chesneau, C., 2006. Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes. Ph.D. Thesis, Université Pierre et Marie Curie - Paris VI, France.

Chesneau, C., 2008. Wavelet estimation via block thresholding: A minimax study under the $L_p$-risk. Statistica Sinica 18, 1007–1024.

De Canditiis, D., Pensky, M., 2004. Discussion on the meeting on statistical approaches to inverse problems. Journal of the Royal Statistical Society, Series B 66, 638–640.

De Canditiis, D., Pensky, M., 2006. Simultaneous wavelet deconvolution in periodic setting. Scandinavian Journal of Statistics 33, 293–306.

Donoho, D.L., 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Applied and Computational Harmonic Analysis 2, 101–126.

Golubev, G.K., Khasminskii, R.Z., 1999. A statistical approach to some inverse problems for partial differential equations. Problems of Information Transmission 35, 136–149.

Johnstone, I.M., Kerkyacharian, G., Picard, D., Raimondo, M., 2004. Wavelet deconvolution in a periodic setting. Journal of the Royal Statistical Society 66, 547–573.

Lattes, R., Lions, J.L., 1967. Methode de Quasi-Reversibilite et Applications. In: Travoux et Recherche Mathematiques, vol. 15. Dunod, Paris.

Meyer, Y., 1992. Wavelets and Operators. Cambridge University Press, Cambridge.

Park, Y.J., Dho, S.W., Kong, H.J., 1997. Deconvolution of long-pulse lidar signals with matrix formulation. Applied Optics 36, 5158–5161.

Pensky, A., Sapatinas, T., 2009. Functional deconvolution in a periodic setting: Uniform case. Annals of Statistics 37, 73–104.

Petsa, A., 2009. Contributions to Wavelet Methods in Nonparametric Statistics. Ph.D. Thesis, Department of Mathematics and Statistics, University of Cyprus, Cyprus.

Strauss, W.A., 1992. Partial Differential Equations: An Introduction. John Wiley & Sons, New York.