1-1-2009

Functional Deconvolution In A Periodic Setting: Uniform Case

Marianna Pensky
University of Central Florida

Theofanis Sapatinas

The Annals of Statistics
2009, Vol. 37, No. 1, 73–104
DOI: 10.1214/07-AOS552

FUNCTIONAL DECONVOLUTION IN A PERIODIC SETTING: UNIFORM CASE

BY MARIANNA PENSKY[1] AND THEOFANIS SAPATINAS

University of Central Florida and University of Cyprus

We extend deconvolution in a periodic setting to deal with functional data. The resulting functional deconvolution model can be viewed as a generalization of a multitude 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. In the case when it is observed at a finite number of distinct points, the proposed functional deconvolution model can also be viewed as a multichannel deconvolution model.

We derive minimax lower bounds for the L^2-risk in the proposed functional deconvolution model when $f(\cdot)$ 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 $f(\cdot)$ that is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls.

In addition, we consider a discretization of the proposed functional deconvolution model and investigate when the availability of continuous data gives advantages over observations at the asymptotically large number of points. As an illustration, we discuss particular examples for both continuous and discrete settings.

1. Introduction. We consider the estimation problem of the unknown response function $f(\cdot)$ based on observations from the following noisy convolutions:

$$Y(u,t) = f * G(u,t) + \frac{\sigma(u)}{\sqrt{n}} z(u,t), \qquad u \in U,\ t \in T, \tag{1.1}$$

where $U = [a,b]$, $-\infty < a \le b < \infty$, and $T = [0,1]$. Here, $z(u,t)$ is assumed to be a two-dimensional Gaussian white noise, that is, a generalized two-dimensional Gaussian field with covariance function

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

Received March 2007; revised September 2007.

[1]Supported in part by NSF Grants DMS-05-05133 and DMS-06-52524.

AMS 2000 subject classifications. Primary 62G05; secondary 62G08, 35J05, 35K05, 35L05.

Key words and phrases. Adaptivity, Besov spaces, block thresholding, deconvolution, Fourier analysis, functional data, Meyer wavelets, minimax estimators, multichannel deconvolution, partial differential equations, wavelet analysis.

where $\delta(\cdot)$ denotes the Dirac δ-function, $\sigma(\cdot)$ is assumed to be a known positive function, and

$$f * G(u,t) = \int_T f(x) G(u, t-x)\, dx, \tag{1.2}$$

with the blurring (or kernel) function $G(\cdot,\cdot)$ in (1.2) also assumed to be known. Note that, since $\sigma(\cdot)$ is assumed to be known, both sides of (1.1) can be divided by $\sigma(\cdot)$ leading to the equation

$$y(u,t) = \int_T f(x) g(u, t-x)\, dx + \frac{1}{\sqrt{n}} z(u,t), \qquad u \in U,\ t \in T, \tag{1.3}$$

where $y(u,t) = Y(u,t)/\sigma(u)$ and $g(u, t-x) = G(u, t-x)/\sigma(u)$. Consequently, without loss of generality, we consider only the case when $\sigma(\cdot) \equiv 1$ and thus, in what follows, we work with observations from model (1.3).

The model (1.3) can be viewed as a *functional deconvolution* model. If $a = b$, it reduces to the standard deconvolution model which attracted attention of a number of researchers. After a rather rapid progress in this problem in late 1980s to early 1990s, authors turned to wavelet solutions of the problem [see, e.g., Donoho (1995), Abramovich and Silverman (1998), Kalifa and Mallat (2003), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Donoho and Raimondo (2004), Johnstone and Raimondo (2004), Neelamani, Choi and Baraniuk (2004) and Kerkyacharian, Picard and Raimondo (2007)]. The main effort was spent on producing adaptive wavelet estimators that are asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls and under mild conditions on the blurring function. [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).]

On the other hand, the functional deconvolution model (1.3) can be viewed as a generalization of a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations of 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 internal domain were studied in Golubev and Khasminskii (1999); the latter problem was also discussed in Golubev (2004). These and other specific models are discussed in Section 5.

Consider now a discretization of the functional deconvolution model (1.3) when $y(u,t)$ is observed at $n = NM$ points (u_l, t_i), $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, that is,

$$y(u_l, t_i) = \int_T f(x) g(u_l, t_i - x)\, dx + \varepsilon_{li}, \qquad u_l \in U,\ t_i = i/N, \tag{1.4}$$

where ε_{li} are standard Gaussian random variables, independent for different l and i. In this case, the functional deconvolution model (1.3) can also be viewed as a multichannel deconvolution problem considered in, for example, Casey and Walnut (1994) and De Canditiis and Pensky (2004, 2006); this model is also discussed in Section 5.

Note that using the same n in (1.3) (continuous model) and (1.4) (discrete model) is not accidental. Under the assumptions (3.3) and (4.1), the optimal (in the minimax sense) convergence rates in the discrete model are determined by the total number of observations, n, and coincide with the optimal convergence rates in the continuous model.

In this paper, we consider functional deconvolution in a periodic setting, that is, we assume that, for fixed $u \in U$, $f(\cdot)$ and $g(u, \cdot)$ are periodic functions with period on the unit interval T. Note that the periodicity assumption appears naturally in the above mentioned special models which (1.3) and (1.4) generalize, and allows one to explore ideas considered in the above cited papers to the proposed functional deconvolution framework. Moreover, not only for theoretical reasons but also for practical convenience [see Johnstone, Kerkyacharian, Picard and Raimondo (2004), Sections 2.3, 3.1–3.2], we use band-limited wavelet bases, and in particular the periodized Meyer wavelet basis for which fast algorithms exist [see Kolaczyk (1994) and Donoho and Raimondo (2004)].

In what follows, we derive minimax lower bounds for the L^2-risk in models (1.3) and (1.4) when $f(\cdot)$ is assumed to belong to a Besov ball and $g(\cdot, \cdot)$ is assumed to possess some smoothness properties, including both regular-smooth and super-smooth convolutions. Furthermore, we propose an adaptive wavelet estimator of $f(\cdot)$ and show that this estimator is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls. We also compare models (1.3) and (1.4), and investigate when the availability of continuous data gives advantages over observations at the asymptotically large number of points.

The paper is organized as follows. In Section 2, we describe the construction of a wavelet estimator of $f(\cdot)$ for both the continuous model (1.3) and the discrete model (1.4). In Section 3, we derive minimax lower bounds for the L^2-risk, based on observations from either the continuous model (1.3) or the discrete model (1.4), when $f(\cdot)$ is assumed to belong to a Besov ball and $g(\cdot, \cdot)$ is assumed to possess some smoothness properties, including both regular-smooth and super-smooth convolutions. In Section 4, we demonstrate that the wavelet estimator derived in Section 2 is adaptive and asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls. In Section 5, we discuss particular examples for both continuous and discrete settings. We conclude in Section 6 with a discussion on the interplay between continuous and discrete models. Finally, in Section 7, we provide some auxiliary statements as well as the proofs of the theoretical results obtained in the earlier sections.

2. Construction of a wavelet estimator. Let $\varphi^*(\cdot)$ and $\psi^*(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively [see, e.g., Meyer (1992) or Mallat (1999)]. As usual,

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

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

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

In what follows, $\langle \cdot, \cdot \rangle$ denotes the inner product in the Hilbert space $L^2(T)$ (the space of squared-integrable functions defined on the unit interval T), that is, $\langle f, g \rangle = \int_T f(t)\overline{g(t)}\, dt$ for $f, g \in L^2(T)$. [Here, and in what follows, $\overline{h(\cdot)}$ (resp. $\bar{h}$) denotes the conjugate of the complex function $h(\cdot)$ (resp. complex number h); $h(\cdot)$ (resp. h) is real if and only if $\overline{h(\cdot)} = h(\cdot)$ (resp. $\bar{h} = h$).]

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

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

be the Fourier coefficients of $\varphi_{jk}(\cdot)$, $\psi_{jk}(\cdot)$ and $f(\cdot)$, respectively. Denote

$$h(u,t) = \int_T f(x) g(u, t-x)\, dx, \qquad u \in U, \ t \in T. \tag{2.1}$$

For each $u \in U$, denote the functional Fourier coefficients by

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

If we have the continuous model (1.3), then, by using properties of the Fourier transform, for each $u \in U$, we have $h_m(u) = g_m(u) f_m$ and

$$y_m(u) = g_m(u) f_m + \frac{1}{\sqrt{n}} z_m(u), \tag{2.2}$$

where $z_m(u)$ are generalized one-dimensional Gaussian processes such that

$$\mathbb{E}[z_{m_1}(u_1) z_{m_2}(u_2)] = \delta_{m_1,m_2}\delta(u_1 - u_2), \tag{2.3}$$

where δ_{m_1,m_2} is Kronecker's delta. In order to find the functional Fourier coefficients f_m of $f(\cdot)$, we multiply both sides of (2.2) by $\overline{g_m(u)}$ and integrate over $u \in U$. The latter yields the following estimators of f_m:

$$\widehat{f_m} = \left(\int_a^b \overline{g_m(u)} y_m(u)\, du\right) \Big/ \left(\int_a^b |g_m(u)|^2\, du\right). \tag{2.4}$$

[Here, we adopt the convention that when $a = b$ the estimator $\widehat{f_m}$ takes the form $\widehat{f_m} = \overline{g_m(a)} y_m(a)/|g_m(a)|^2$.]

If we have the discrete model (1.4), then, by using properties of the discrete Fourier transform, for each $l = 1, 2, \dots, M$, (2.2) takes the form

$$y_m(u_l) = g_m(u_l) f_m + \frac{1}{\sqrt{N}} z_{ml}, \tag{2.5}$$

where z_{ml} are standard Gaussian random variables, independent for different m and l. Similarly to the continuous case, we multiply both sides of (2.5) by $\overline{g_m(u_l)}$ and add them together to obtain the following estimators of f_m:

$$\widehat{f_m} = \left(\sum_{l=1}^{M} \overline{g_m(u_l)} y_m(u_l)\right) \Big/ \left(\sum_{l=1}^{M} |g_m(u_l)|^2\right). \tag{2.6}$$

[Here, and in what follows, we abuse notation and f_m refers to both functional Fourier coefficients and their discrete counterparts. Note also that $y_m(u_l)$, $g_m(u_l)$ and z_{ml} are, respectively, the discrete versions of the functional Fourier coefficients $y_m(u)$, $g_m(u)$ and $z_m(u)$.]

Note that, using the periodized Meyer wavelet basis described above and for any $j_0 \ge 0$, any (periodic) $f(\cdot) \in L^2(T)$ can be expanded as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0k} \varphi_{j_0k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} b_{jk} \psi_{jk}(t). \tag{2.7}$$

Furthermore, by Plancherel's formula, the scaling coefficients, $a_{j_0k} = \langle f, \varphi_{j_0k} \rangle$, and the wavelet coefficients, $b_{jk} = \langle f, \psi_{jk} \rangle$, of $f(\cdot)$ can be represented as

$$a_{j_0k} = \sum_{m \in C_{j_0}} f_m \overline{\varphi_{mj_0k}}, \qquad b_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}}, \tag{2.8}$$

where $C_{j_0} = \{m : \varphi_{mj_0k} \ne 0\}$ and, for any $j \ge j_0$, $C_j = \{m : \psi_{mjk} \ne 0\}$, both subsets of $2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]$, due to the fact that Meyer wavelets are band-limited [see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004), Section 3.1]. We naturally estimate a_{j_0k} and b_{jk} by substituting f_m in (2.8) with (2.4) or (2.6), that is,

$$\widehat{a}_{j_0k} = \sum_{m \in C_{j_0}} \widehat{f_m} \overline{\varphi_{mj_0k}}, \qquad \widehat{b}_{jk} = \sum_{m \in C_j} \widehat{f_m} \overline{\psi_{mjk}}. \tag{2.9}$$

We now construct a block thresholding wavelet estimator of $f(\cdot)$. For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length $\ln n$. Let A_j and U_{jr} be the following sets of indices:

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

Denote

$$B_{jr} = \sum_{k \in U_{jr}} b_{jk}^2, \qquad \widehat{B}_{jr} = \sum_{k \in U_{jr}} \widehat{b}_{jk}^2. \tag{2.10}$$

Finally, for any $j_0 \geq 0$, we reconstruct $f(\cdot)$ as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0 k}\varphi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \widehat{b}_{jk}\mathbb{I}(|\widehat{B}_{jr}| \geq \lambda_j)\psi_{jk}(t), \tag{2.11}$$

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 4.

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

3. Minimax lower bounds for the L^2-risk over Besov balls. Among the various characterizations of Besov spaces for periodic functions defined on $L^p(T)$ in terms of wavelet bases, we recall that for an r-regular multiresolution analysis with $0 < s < r$ and for a Besov ball $B_{p,q}^s(A)$ of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has that, with $s' = s + 1/2 - 1/p$,

$$\begin{aligned} B_{p,q}^s(A) = \Bigg\{ f(\cdot) \in L^p(T) : &\left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0 k}|^p \right)^{1/p} \\ &+ \left(\sum_{j=j_0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \Bigg\}, \end{aligned} \tag{3.1}$$

with respective sum(s) replaced by maximum if $p = \infty$ or $q = \infty$ [see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004), Section 2.4]. (Note that, for the Meyer wavelet basis, considered in Section 2, $r = \infty$.)

We construct below minimax lower bounds for the L^2-risk, for both the continuous model (1.3) and the discrete model (1.4). For this purpose, we define the minimax L^2-risk over the set Ω as

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

where $\|g\|$ is the L^2-norm of a function $g(\cdot)$ and the infimum is taken over all possible estimators $\tilde{f}_n(\cdot)$ (measurable functions taking their values in a set containing Ω) of $f(\cdot)$, based on observations from either the continuous model (1.3) or the discrete model (1.4). [Here, and in what follows, the expectation is taken under the true $f(\cdot)$, and it is assumed that the function class Ω contains $f(\cdot)$.]

In what follows, we shall evaluate a lower bound for $R_n(B_{p,q}^s(A))$. Denote

$$s^* = s + 1/2 - 1/p', \qquad p' = \min(p, 2),$$

and, for $\kappa = 1, 2$, define

$$\tau_\kappa(m) = \begin{cases} \displaystyle\int_a^b |g_m(u)|^{2\kappa}\, du, & \text{in the continuous case,} \\ \displaystyle\frac{1}{M}\sum_{l=1}^{M} |g_m(u_l)|^{2\kappa}, & \text{in the discrete case.} \end{cases} \tag{3.2}$$

[Here, we adopt the convention that when $a = b$, $\tau_\kappa(m)$ takes the form $\tau_\kappa(m) = |g_m(a)|^{2\kappa}$, $\kappa = 1, 2$.] Assume that for some constants $\nu \in \mathbb{R}$, $\alpha \geq 0$, $\beta > 0$ and $K_1 > 0$, independent of m, the choice of M and the selection points u_l, $l = 1, 2, \dots, M$,

$$\tau_1(m) \leq K_1|m|^{-2\nu} \exp(-\alpha|m|^\beta), \qquad \nu > 0 \text{ if } \alpha = 0. \tag{3.3}$$

[Following Fan (1991), we say that the function $g(\cdot, \cdot)$ is *regular-smooth* if $\alpha = 0$ and is *super-smooth* if $\alpha > 0$.]

The following statement provides the minimax lower bounds for the L^2-risk.

THEOREM 1. *Let* $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ *be the periodic Meyer wavelet basis discussed in Section* 2. *Let* $s > \max(0, 1/p - 1/2)$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ *and* $A > 0$. *Then*, *under the assumption* (3.3), *as* $n \to \infty$,

$$R_n(B_{p,q}^s(A)) \geq \begin{cases} Cn^{-2s/(2s+2\nu+1)}, & \text{if } \alpha = 0,\ \nu(2-p) < ps^*, \\ C\left(\dfrac{\ln n}{n}\right)^{2s^*/(2s^*+2\nu)}, & \text{if } \alpha = 0,\ \nu(2-p) \geq ps^*, \\ C(\ln n)^{-2s^*/\beta}, & \text{if } \alpha > 0. \end{cases} \tag{3.4}$$

REMARK 1. The two different lower bounds for $\alpha = 0$ in (3.4) refer to the dense case [$\nu(2 - p) < ps^*$] 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 [$\nu(2 - p) \geq ps^*$] when the worst functions $f(\cdot)$ have only one nonvanishing wavelet coefficient. Note also that the restriction $s > \max(0, 1/p - 1/2)$, $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$ ensures that the corresponding Besov spaces are embedded in $L^2(T)$.

4. Minimax upper bounds for the L^2-risk over Besov balls. Recall (3.3) from Section 3, and assume further that for the constants $\nu \in \mathbb{R}$, $\alpha \geq 0$ and $\beta > 0$, and for a constant $K_2 > 0$, independent of m, the choice of M and the selection points u_l, $l = 1, 2, \dots, M$, with $K_2 \leq K_1$,

$$\tau_1(m) \geq K_2|m|^{-2\nu} \exp(-\alpha|m|^\beta), \qquad \nu > 0 \text{ if } \alpha = 0. \tag{4.1}$$

For any $j \geq j_0$, let $|C_j|$ be the cardinality of the set C_j; note that, for Meyer wavelets, $|C_j| = 4\pi 2^j$ [see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004), page 565]. Let also

$$\Delta_\kappa(j) = \frac{1}{|C_j|}\sum_{m \in C_j} \tau_\kappa(m)[\tau_1(m)]^{-2\kappa}, \qquad \kappa = 1, 2. \tag{4.2}$$

Then, direct calculations yield that

$$\Delta_1(j) \le \begin{cases} c_1 2^{2\nu j}, & \text{if } \alpha = 0, \\ c_2 2^{2\nu j} \exp\left(\alpha \left(\frac{8\pi}{3}\right)^{\beta} 2^{j\beta}\right), & \text{if } \alpha > 0. \end{cases} \tag{4.3}$$

[Note that since the functional Fourier coefficients $g_m(\cdot)$ are known, the positive constants c_1 and c_2 in (4.3) can be evaluated explicitly.]

Consider now the two cases $\alpha = 0$ (regular-smooth) and $\alpha > 0$ (super-smooth) separately. Choose j_0 and J such that

$$2^{j_0} = \ln n, \qquad 2^J = n^{1/(2\nu+1)} \qquad \text{if } \alpha = 0. \tag{4.4}$$

$$2^{j_0} = \frac{3}{8\pi}\left(\frac{\ln n}{2\alpha}\right)^{1/\beta}, \qquad 2^J = 2^{j_0}, \qquad \text{if } \alpha > 0. \tag{4.5}$$

[Since $j_0 > J - 1$ when $\alpha > 0$, the wavelet estimator (2.11) only consists of the first (linear) part and, hence, λ_j does not need to be selected in this case.] Set, for some positive constant d,

$$\lambda_j = d n^{-1} 2^{2\nu j} \ln n \qquad \text{if } \alpha = 0. \tag{4.6}$$

Note that the choices of j_0, J and λ_j are independent of the parameters, s, p, q and A (that are usually unknown in practical situations) of the Besov ball $B^s_{p,q}(A)$; hence, the wavelet estimator (2.11) is adaptive with respect to these parameters.

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

LEMMA 1. *Let the assumption* (4.1) *be valid, and 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 the formula* (2.9) *with* $\widehat{f_m}$ *defined by* (2.4) *in the continuous model and by* (2.6) *in the discrete model. Then, for* $\kappa = 1, 2$, *and for all* $j \ge j_0$,

$$\mathbb{E}|\widehat{a}_{j_0k} - a_{j_0k}|^2 \le C n^{-1} \Delta_1(j_0), \tag{4.7}$$

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^{2\kappa} \le C n^{-\kappa} \Delta_\kappa(j). \tag{4.8}$$

Moreover, under the assumptions (3.3) *and* (4.1) *with* $\alpha = 0$, *for all* $j \ge j_0$,

$$\Delta_2(j) \le C 2^{4(2\nu - \nu_1) j}, \tag{4.9}$$

for any $0 < \nu_1 \le \nu$.

LEMMA 2. *Let the estimators* $\widehat{b}_{jk}$ *of the wavelet coefficients* b_{jk} *be given by the formula* (2.9) *with* $\widehat{f_m}$ *defined by* (2.4) *in the continuous model and by* (2.6) *in the discrete model. If* μ *is a positive constant large enough and* $\alpha = 0$ *in the assumption* (4.1), *then, for all* $j \ge j_0$,

$$\mathbb{P}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \ge 0.25 \mu^2 n^{-1} 2^{2\nu j} \ln n\right) \le n^{-(8\nu - 4\nu_1 + 2)/(2\nu+1)}, \tag{4.10}$$

for any $0 < \nu_1 \le \nu$.

Lemmas 1 and 2 allow to state the following minimax upper bounds for the L^2-risk of the wavelet estimator $\hat{f}_n(\cdot)$ defined by (2.11), with j_0 and J given by (4.4) (if $\alpha = 0$) or (4.5) (if $\alpha > 0$). Set $(x)_+ = \max(0, x)$, and define

$$\varrho_1 = \begin{cases} \dfrac{(2\nu+1)(2-p)_+}{p(2s+2\nu+1)}, & \text{if } \nu(2-p) < ps^*, \\ \dfrac{(q-p)_+}{q}, & \text{if } \nu(2-p) = ps^*, \\ 0, & \text{if } \nu(2-p) > ps^*. \end{cases} \tag{4.11}$$

THEOREM 2. *Let $\hat{f}_n(\cdot)$ be the wavelet estimator defined by* (2.11), *with j_0 and J given by* (4.4) (*if $\alpha = 0$*) *or* (4.5) (*if $\alpha > 0$*). *Let $s > 1/p'$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Then, under the assumption* (4.1), *as $n \to \infty$,*

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \le \begin{cases} Cn^{-2s/(2s+2\nu+1)}(\ln n)^{\varrho_1}, \\ \qquad \textit{if } \alpha = 0, \nu(2-p) < ps^*, \\ C\left(\dfrac{\ln n}{n}\right)^{2s^*/(2s^*+2\nu)}(\ln n)^{\varrho_1}, \\ \qquad \textit{if } \alpha = 0, \nu(2-p) \ge ps^*, \\ C(\ln n)^{-2s^*/\beta}, \\ \qquad \textit{if } \alpha > 0. \end{cases} \tag{4.12}$$

REMARK 2. In the discrete model, assumptions (3.3) and (4.1) require the value of $\tau_1(m)$ to be independent of the choice of M and the selection of points u_l, $l = 1, 2, \dots, M$. If assumptions (3.3) and (4.1) hold, then the minimax convergence rates in discrete and continuous models coincide and are independent of the configuration of the points u_l, $l = 1, 2, \dots, M$. Moreover, the wavelet estimator (2.11) is asymptotically optimal (in the minimax sense) no matter what the value of M is. It is quite possible, however, that in the discrete model, conditions (3.3) and (4.1) both hold but with different values of ν, α and β. In this case, the upper bounds for the risk in the discrete model may not coincide with the lower bounds and with the minimax convergence rates in the continuous model. Proposition 1 in Section 6 provides sufficient conditions for the minimax convergence rates in discrete and continuous models to coincide and to be independent of M and the configuration of the points u_l, $l = 1, 2, \dots, M$. These conditions also guarantee asymptotical optimality of the wavelet estimator (2.11), and can be viewed as some kind of uniformity conditions. If conditions of Proposition 1 are violated, then the rates of convergence in the discrete model depend on the choice of M and u_l, $l = 1, 2, \dots, M$, and some recommendations on their selection should be given. Furthermore, optimality issues become much more complex when $\tau_1(m)$ is not uniformly bounded from above and below (see the discussion in Section 6).

REMARK 3. Theorems 1 and 2 imply that, for the L^2-risk, the wavelet estimator $\hat{f}_n(\cdot)$ defined by (2.11) is asymptotically 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 $1 \le q \le \infty$. In particular, in the cases when (1) $\alpha > 0$, (2) $\alpha = 0$, $\nu(2-p) < ps^*$ and $2 \le p \le \infty$, (3) $\alpha = 0$, $\nu(2-p) > ps^*$, and (4) $\alpha = 0$, $\nu(2-p) = ps^*$ and $1 \le q \le p$, the estimator (2.11) is asymptotically optimal (lower and upper bounds coincide up to a multiplicative constant), that is,

$$R_n(B^s_{p,q}(A)) \asymp \begin{cases} n^{-2s/(2s+2\nu+1)}, & \text{if } \alpha = 0,\ \nu(2-p) < ps^*,\ 2 \le p \le \infty, \\ \left(\dfrac{\ln n}{n}\right)^{2s^*/(2s^*+2\nu)}, & \text{if } \alpha = 0,\ \nu(2-p) > ps^*, \\ & \text{or } \alpha = 0,\ \nu(2-p) = ps^*,\ 1 \le q \le p, \\ (\ln n)^{-2s^*/\beta}, & \text{if } \alpha > 0. \end{cases}$$

On the other hand, in the case when $\alpha = 0$, $\nu(2-p) < ps^*$ and $1 \le p < 2$ or $\alpha = 0$, $\nu(2-p) = ps^*$ and $1 \le p < q$, the wavelet estimator $\hat{f}_n(\cdot)$ defined by (2.11) is asymptotically near-optimal within a logarithmic factor, that is,

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \le \begin{cases} Cn^{-2s/(2s+2\nu+1)}(\ln n)^{(2\nu+1)(2-p)/(p(2s+2\nu+1))}, \\ \qquad \text{if } \alpha = 0,\ \nu(2-p) < ps^*, \\ \qquad 1 \le p < 2, \\ C\left(\dfrac{\ln n}{n}\right)^{2s^*/(2s^*+2\nu)} (\ln n)^{(1-p/q)}, \\ \qquad \text{if } \alpha = 0,\ \nu(2-p) = ps^*, \\ \qquad 1 \le p < q. \end{cases}$$

[Here, and in what follows, $g_1(n) \asymp g_2(n)$ denotes $0 < \liminf(g_1(n)/g_2(n)) \le \limsup(g_1(n)/g_2(n)) < \infty$ as $n \to \infty$.]

REMARK 4. For the L^2-risk, the upper bounds (4.12) are tighter than those obtained by Chesneau (2008) for the regular-smooth case [i.e., $\alpha = 0$ in (3.3) and (4.1)] in the case of the standard deconvolution model [i.e., when $a = b$ in (1.3)], although the difference is only in the logarithmic factors. More specifically, the following minimax upper bounds obtained in Chesneau (2008) for the L^2-risk, as $n \to \infty$:

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \le \begin{cases} Cn^{-2s/(2s+2\nu+1)}(\ln n)^{\varrho_2}, \\ \qquad \text{if } \alpha = 0,\ \nu(2-p) < ps', \\ C\left(\dfrac{\ln n}{n}\right)^{2s'/(2s'+2\nu)} (\ln n)^{\varrho_2}, \\ \qquad \text{if } \alpha = 0,\ \nu(2-p) \ge ps', \end{cases} \tag{4.13}$$

where

$$\varrho_2 = \begin{cases} \dfrac{2s\mathbb{I}(1 \le p < 2)}{2s+2\nu+1}, & \text{if } \nu(2-p) < ps', \\ \dfrac{(2q-p)_+}{q}, & \text{if } \nu(2-p) = ps', \\ 0, & \text{if } \nu(2-p) > ps'. \end{cases} \tag{4.14}$$

[Here, and in what follows, $\mathbb{I}(A)$ is the indicator function of the set A.] Note that when $2 \le p \le \infty$, $s^* = s \le s'$, and only the dense case appears; hence, in this case, the dense cases and the corresponding convergence rates in the minimax upper bounds given by (4.11)–(4.12) and (4.13)–(4.14) coincide since $\nu(2-p) < ps^* = ps < ps'$. On the other hand, when $1 \le p < 2$, $s^* = s'$, both the dense and sparse cases appear; hence, in this case, both the dense and sparse cases and the corresponding convergence rates in the minimax upper bounds given by (4.11)–(4.12) and (4.13)–(4.14) coincide. Looking now at (4.11) and (4.14), we see that $\varrho_2 = \varrho_1$ only when $\nu(2-p) > ps'$. On the other hand, $\varrho_2 > \varrho_1$ when $1 \le p < 2$ and $\nu(2-p) < ps'$ since $(2/p-1)(2\nu+1) - 2s = 2(2\nu - p\nu - ps')/p < 0$, and it is obvious that $\varrho_2 > \varrho_1$ when $\nu(2-p) = ps'$. However, we believe that the slight superiority in the minimax convergence rates for the L^2-risk obtained in Theorems 1 and 2 is due not to a different construction of the wavelet estimator but to a somewhat different way of evaluating the minimax upper bounds.

REMARK 5. Unlike Chesneau (2008) who only considered minimax upper bounds for the regular-smooth case [i.e., $\alpha = 0$ in (3.3) and (4.1)] in the standard deconvolution model [i.e., when $a = b$ in (1.3)], Theorems 1 and 2 provide minimax lower and upper bounds (in the L^2-risk) for both regular-smooth and super-smooth convolutions [i.e., $\alpha > 0$ in (3.3) and (4.1)], not only for the standard deconvolution model but also for its discrete counterpart [i.e., when $M = 1$ in (1.4)].

REMARK 6. The wavelet estimator $\hat{f}_n(\cdot)$ defined by (2.11) is adaptive with respect to the unknown parameters s, p, q and A of the Besov ball $B^s_{p,q}(A)$ but is not adaptive with respect to the parameters α, β and ν in (3.3) and (4.1). It seems that it is impossible to achieve adaptivity with respect to β in the super-smooth case ($\alpha > 0$) because of the very fast exponential growth of the variance. However, in the regular-smooth case ($\alpha = 0$), one can construct a wavelet estimator which is adaptive with respect to the unknown parameter ν. Choose j_0 and J such that $2^{j_0} = \ln n$ and $2^J = n$, and set $\lambda_j = d^* n^{-1} \ln n \Delta_1(j)$, where d^* is large enough. Note that $\Delta_1(j)$ can be calculated whenever the functional Fourier coefficients $g_m(\cdot)$ are available. Also, $K_1^* 2^{2\nu j} \le \Delta_1(j) \le K_2^* 2^{2\nu j}$ for some positive constants K_1^* and K_2^* which depend on the particular values of the constants in the conditions (3.3) and (4.1). Therefore, in this situation, by repeating the proof of Theorem 2 with these new values of the parameters involved, one can easily verify that the optimal convergence rates in Theorem 2 still hold as long as d^* is large enough. How large should be "large enough"? Direct calculations show that d^* should be such that $(0.5 d^* \sqrt{K_2^*/K_1^*} - 1)^2 \ge 8\nu + 2$. Since K_1^*, K_2^* and ν are unknown, it is impossible to evaluate the lower bound for d^*. However, one can replace d^* by a slow-growing function of n, say $\ln \ln n$, leading to, at most, an extra $\ln \ln n$ factor in the obtained maximal L^2-risk.

REMARK 7. We finally note that, although we have only considered L^2-risks in our analysis, the results obtained in Theorems 1 and 2 can be extended to the case of L^π-risks ($1 \le \pi < \infty$). Analogous statements to the ones given in Theorems 1 and 2 but for a wider variety of risk functions can be obtained using the unconditionality and Temlyakov properties of Meyer wavelets [see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004), Appendices A and B]. The details in the derivation of these statements should, however, be carefully addressed.

5. Examples in continuous and discrete models. The functional deconvolution model (1.3) can be viewed as a generalization of a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations of 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 internal domain were studied in Golubev and Khasminskii (1999). More specifically, by studying separately the heat conductivity equation or the Laplace equation on the unit circle, and assuming that the unknown initial or boundary condition belongs to a Sobolev ball, Golubev and Khasminskii (1999) obtained some *linear* and *nonadaptive* solutions to the particular problem at hand; see also Golubev (2004) for a linear adaptive estimator for the Laplace equation on the circle based on the principle of minimization of penalized empirical risk. We also note that, unlike Golubev and Khasminskii (1999) and Golubev (2004) who considered sharp asymptotics, we focus our study on rate optimality results. [Note that the estimation of the unknown initial condition for the heat conductivity equation, allowing also for missing data, has been recently considered by Hesse (2007); however, this latter paper deals with the density deconvolution model and the approach given therein varies from the approach of Golubev and Khasminskii (1999) and Golubev (2004), and it seems to be having a different agenda.]

In view of the general framework developed in this paper, however, the inverse problems mentioned above can all be expressed as a functional deconvolution problem, so that all techniques studied in Sections 2–4 can be directly applied, to obtain *linear/nonlinear* and *adaptive* solutions over a wide range of Besov balls. Such solutions are provided in Examples 1–4 below which discuss some of the most common inverse problems in mathematical physics which have already been studied as well as some other problems which, to the best of our knowledge, have not yet been addressed.

On the other hand, in the case when the functional deconvolution model (1.3) is observed at a finite number of distinct points [see (1.4)], it can also be viewed as a multichannel deconvolution model studied in De Canditiis and Pensky (2004, 2006). Example 5 below deals with this model, providing the minimax convergence rates (in the L^2-risk) for regular-smooth [i.e., $\alpha = 0$ in (3.3) and (4.1)] and super-smooth [i.e., $\alpha > 0$ in (3.3) and (4.1)] convolutions, and also discussing the case when M can increase together with N; both of these aspects were lacking from the theoretical analysis described in De Canditiis and Pensky (2006).

EXAMPLE 1 (*Estimation of the initial condition in the heat conductivity equation*). Let $h(t,x)$ be a solution of the heat conductivity equation

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

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

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

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 initial condition $f(\cdot)$ on the basis of observations $y(t,x)$. This problem was considered by Lattes and Lions (1967) and Golubev and Khasminskii (1999).

It is well known [see, e.g., Strauss (1992), page 48] that, in a periodic setting, the solution $h(t,x)$ can be written as

$$h(t,x) = (4\pi t)^{-1/2} \int_0^1 \sum_{k \in \mathbb{Z}} \exp\left\{-\frac{(x+k-z)^2}{4t}\right\} f(z)\,dz. \tag{5.1}$$

It is easy to see that (5.1) coincides with (2.1) with t and x replaced by u and t, respectively, and that

$$g(u,t) = (4\pi u)^{-1/2} \sum_{k \in \mathbb{Z}} \exp\left\{-\frac{(t+k)^2}{4u}\right\}.$$

Applying the theory developed in Sections 2–4, we obtain functional Fourier coefficients $g_m(\cdot)$ satisfying $g_m(u) = \exp(-4\pi^2 m^2 u)$, and

$$\tau_1(m) = \int_a^b |g_m(u)|^2\,du = C m^{-2} \exp(-8\pi^2 m^2 a)\bigl(1 + o(1)\bigr), \qquad |m| \to \infty,$$

so that $\nu = 1$, $\alpha = 8\pi^2 a$ and $\beta = 2$ in both (3.3) and (4.1).

Hence, one can construct an adaptive wavelet estimator of the form (2.11), with j_0 and J given by (4.5), which achieves minimax (in the L^2-risk) convergence rate of order $(\ln n)^{-s^*}$ over Besov balls $B^s_{p,q}(A)$ of radius $A > 0$ with $s > \max(1/p, 1/2)$, $1 \le p \le \infty$ and $1 \le q \le \infty$.

EXAMPLE 2 (*Estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle*). Let $h(x,w)$ be a solution of the Dirichlet problem of the Laplacian on a region D on the plane

$$\frac{\partial^2 h(x,w)}{\partial x^2} + \frac{\partial^2 h(x,w)}{\partial w^2} = 0, \qquad (x,w) \in D \subseteq \mathbb{R}^2, \tag{5.2}$$

with a boundary ∂D and boundary condition

$$h(x,w)|_{\partial D} = F(x,w). \tag{5.3}$$

Consider the situation when D is the unit circle. Then, it is advantageous to rewrite the function $h(\cdot,\cdot)$ in polar coordinates as $h(x,w) = h(u,t)$, where $u \in [0,1]$ is the polar radius and $t \in [0, 2\pi]$ is the polar angle. Then, the boundary condition in (5.3) can be presented as $h(1,t) = f(t)$, and $h(u,\cdot)$ and $f(\cdot)$ are periodic functions of t with period 2π.

Suppose that only a noisy version $y(u,t) = h(u,t) + n^{-1/2} z(u,t)$ is observed, where $z(u,t)$ is as in Example 1, and that observations are available only on the interior of the unit circle with $u \in [0, r_0]$, $r_0 < 1$, that is, $a = 0$, $b = r_0 < 1$. The goal is to recover the boundary condition $f(\cdot)$ on the basis of observations $y(u,t)$. This problem was investigated in Golubev and Khasminskii (1999) and Golubev (2004).

It is well known [see, e.g., Strauss (1992), page 161] that the solution $h(u,t)$ can be written as

$$h(u,t) = \frac{(1-u^2)}{2\pi} \int_0^{2\pi} \frac{f(x)}{1 - 2u\cos(t-x) + u^2}\, dx.$$

Applying the theory developed in Sections 2–4 with $e_m(t) = e^{imt}$ and

$$g(u,t) = \frac{1-u^2}{1 - 2u\cos(t) + u^2},$$

we obtain functional Fourier coefficients $g_m(\cdot)$ satisfying $g_m(u) = Cu^m$, and

$$\tau_1(m) = \int_0^{r_0} |g_m(u)|^2\, du = C\exp\{-2\ln(1/r_0)|m|\},$$

so that $\nu = 0$, $\alpha = 2\ln(1/r_0)$ and $\beta = 1$ in both (3.3) and (4.1).

Hence, one can construct an adaptive wavelet estimator of the form (2.11), with j_0 and J given by (4.5), which achieves minimax (in the L^2-risk) convergence rate of order $(\ln n)^{-2s^*}$ over Besov balls $B^s_{p,q}(A)$ of radius $A > 0$ with $s > \max(1/p, 1/2)$, $1 \le p \le \infty$ and $1 \le q \le \infty$.

EXAMPLE 3 (*Estimation of the boundary condition for the Dirichlet problem of the Laplacian on a rectangle*). Consider the problem (5.2)–(5.3) in Example 2 above, with the region D being now a rectangle, that is, $(x,w) \in [0,1] \times [a,b]$, $a > 0$, $b < \infty$, and periodic boundary conditions

$$h(x,0) = f(x), \qquad h(0,w) = h(1,w).$$

Again, suppose that only a noisy version $y(x,w) = h(x,w) + n^{-1/2} z(x,w)$ is observed, where $z(x,w)$ is as in Example 1, for $x \in [0,1]$, $w \in [a,b]$, and the goal is to recover the boundary condition $f(\cdot)$ on the basis of observations $y(x,w)$.

It is well known [see, e.g., Strauss (1992), pages 188, 407] that, in a periodic setting, the solution $h(x, w)$ can be written as

$$h(x,w) = \pi^{-1} \int_0^1 \sum_{k\in\mathbb{Z}} \frac{w}{w^2 + (x+k-z)^2} f(z)\,dz. \tag{5.4}$$

It is easy to see that (5.4) coincides with (2.1) with x and w replaced by t and u, respectively, and that

$$g(u,t) = \pi^{-1} \sum_{k\in\mathbb{Z}} \frac{u}{u^2 + (t+k)^2}.$$

Applying the theory developed in Sections 2–4, we obtain functional Fourier coefficients $g_m(\cdot)$ satisfying $g_m(u) = \exp(-2\pi m u)$, and

$$\tau_1(m) = \int_a^b |g_m(u)|^2\,du = C|m|^{-1} \exp(-4\pi|m|a)\big(1 + o(1)\big), \qquad |m| \to \infty,$$

so that $\nu = 1/2$, $\alpha = 4\pi a$ and $\beta = 1$ in both (3.3) and (4.1).

Hence, one can construct an adaptive wavelet estimator of the form (2.11), with j_0 and J given by (4.5), which achieves minimax (in the L^2-risk) convergence rate of order $(\ln n)^{-2s^*}$ over Besov balls $B^s_{p,q}(A)$ of radius $A > 0$ with $s > \max(1/p, 1/2)$, $1 \le p \le \infty$ and $1 \le q \le \infty$.

EXAMPLE 4 (*Estimation of the speed of a wave on a finite interval*). Let $h(t, x)$ be a solution of the wave equation

$$\frac{\partial^2 h(t,x)}{\partial t^2} = \frac{\partial^2 h(t,x)}{\partial x^2}$$

with initial–boundary conditions

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

Here, $f(\cdot)$ is a function defined on the unit interval $[0, 1]$, and the objective is to recover $f(\cdot)$ on the basis of observing a noisy solution $y(t, x) = h(t, x) + n^{-1/2} z(t, x)$, where $z(t, x)$ is as in Example 1, with $t \in [a, b]$, $a > 0$, $b < 1$.

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

$$h(t,x) = \tfrac{1}{2} \int_0^1 \mathbb{I}(|x-z| < t) f(z)\,dz, \tag{5.5}$$

so that (5.5) is of the form (2.1) with $g(u, x) = 0.5\mathbb{I}(|x| < u)$ (a boxcar-like kernel for each fixed u), where u in (2.1) is replaced by t in (5.5). Applying the theory developed in Sections 2–4, with t and x replaced by u and t, respectively, we obtain functional Fourier coefficients $g_m(\cdot)$ satisfying $g_m(u) = \sin(2\pi m u)/(2\pi m)$, and

$$\begin{aligned} \tau_1(m) &= \int_a^b |g_m(u)|^2\,du \\ &= \frac{1}{4\pi^2 m^2}\left(\frac{b-a}{2} + \frac{\sin(4\pi m a) - \sin(4\pi m b)}{8\pi m}\right). \end{aligned} \tag{5.6}$$

Observe that the integral in (5.6) is always positive, bounded from above by Cm^{-2} and from below by $Cm^{-2}[(b-a) - (2\pi m)^{-1}]$, so that $\nu = 1$ and $\alpha = 0$ in both (3.3) and (4.1).

Hence, one can construct an adaptive block thresholding wavelet estimator of the form (2.11), with j_0 and J given by (4.4), which achieves the following minimax upper bounds (in the L^2-risk):

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \le \begin{cases} Cn^{-2s/(2s+3)}(\ln n)^{\varrho_1}, & \text{if } s > 3(1/p - 1/2), \\ C\left(\dfrac{\ln n}{n}\right)^{s'/(s'+1)}(\ln n)^{\varrho_1}, & \text{if } s \le 3(1/p - 1/2), \end{cases}$$

over Besov balls $B^s_{p,q}(A)$ of radius $A > 0$ with $s > 1/p'$, $1 \le p \le \infty$ and $1 \le q \le \infty$, where $\varrho_1 = 3(2/p - 1)_+/(2s+3)$ if $s > 3(1/p - 1/2)$, $\varrho_1 = (1 - p/q)_+$ if $s = 3(1/p - 1/2)$ and $\varrho_1 = 0$ if $s < 3(1/p - 1/2)$. [The minimax lower bounds (in the L^2-risk) have the same form with $\varrho_1 = 0$.]

EXAMPLE 5 (*Estimation in the multichannel deconvolution problem*). Consider the problem of recovering $f(\cdot) \in L^2(T)$ on the basis of observing the following noisy convolutions with known blurring functions $g_l(\cdot)$

$$Y_l(dt) = f * g_l(t)\,dt + \frac{\sigma_l}{\sqrt{n}} W_l(dt), \qquad t \in T, l = 1, 2, \ldots, M. \tag{5.7}$$

Here, σ_l are known positive constants and $W_l(t)$ are independent standard Wiener processes.

The problem of considering systems of convolution equations was first considered by Casey and Walnut (1994) in order to evade the ill-posedness of the standard deconvolution problem, and was adapted for statistical use (in the density deconvolution model) by Pensky and Zayed (2002). Wavelet solutions to the problem (5.7) were investigated by De Canditiis and Pensky (2004, 2006).

Note that deconvolution is the common problem in many areas of signal and image processing which include, for instance, LIDAR (Light Detection and Ranging) remote sensing and reconstruction of blurred images. LIDAR is a laser device which emits pulses, reflections of which are gathered by a telescope aligned with the laser [see, e.g., Park, Dho and Kong (1997) and Harsdorf and Reuter (2000)]. The return signal is used to determine distance and the position of the reflecting material. However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. If M ($M \geq 2$) LIDAR devices are used to recover a signal, then we talk about a multichannel deconvolution problem. Note that a discretization of (5.7) (with $\sigma_l = 1$ for $l = 1, 2, \dots, M$) leads to the discrete setup (1.4).

Adaptive term by term wavelet thresholding estimators for the model (5.7) were constructed in De Canditiis and Pensky (2006) for regular-smooth convolutions [i.e., $\alpha = 0$ in (3.3) and (4.1)]. However, minimax lower and upper bounds were not obtained by these authors who concentrated instead on upper bounds (in the L^π-risk, $1 < \pi < \infty$) for the error, for a fixed response function. Moreover, the case of super-smooth convolutions [i.e., $\alpha > 0$ in (3.3) and (4.1)] and the case when $M \to \infty$ have not been treated in De Canditiis and Pensky (2006).

Let us now discuss the regular-smooth convolution case treated in De Canditiis and Pensky (2006), that is, the case when (in our notation) $|g_m(u_l)| \sim C_l |m|^{-\nu_l}$ with $0 < C_l < \infty$, $l = 1, 2, \dots, M$. If M is fixed, then

$$C_* M^{-1} m^{-2\nu_{\min}} \leq \tau_1(m) \leq C^* m^{-2\nu_{\min}},$$

where $\nu_{\min} = \min\{\nu_1, \nu_2, \dots, \nu_M\}$ and $0 < C_* \leq C_l \leq C^* < \infty$, $l = 1, 2, \dots, M$. Hence, the minimax rates of convergence (in the L^2-risk) are determined by $\nu_{\min}$ only, meaning that one can just rely on the best possible channel and disregard all the others. However, the latter is no longer true if $M \to \infty$. In this case, the minimax rates of convergence (in the L^2-risk) are determined by $\tau_1(m)$ which may not be a function of $\nu_{\min}$ only.

Consider now the adaptive block thresholding wavelet estimator $\hat{f}_n(\cdot)$ defined by (2.11) for the model (5.7) $\sigma_l = 1$ for $l = 1, 2, \dots, M$ or its discrete counterpart (1.4). Then, for the L^2-risk, under the assumption (3.3), the corresponding minimax lower bounds are given by Theorem 1, while, under the assumption (4.1), the corresponding minimax upper bounds are given by Theorem 2. Thus, the proposed functional deconvolution methodology significantly expands on the theoretical findings in De Canditiis and Pensky (2006).

6. Discussion: the interplay between continuous and discrete models. The minimax convergence rates (in the L^2-risk) in the discrete model depend on two aspects: the total number of observations $n = NM$ and the behavior of $\tau_1(m)$ defined in (3.2). In the continuous model, the values of $\tau_1(m)$ are fixed; however, in the discrete model they may depend on the choice of M and the selection of points u_l, $l = 1, 2, \dots, M$. Let us now explore when and how this can happen.

Assume that there exist points $u_*, u^* \in [a, b]$, $-\infty < a \leq b < \infty$ (with $a < b$ in the continuous model while $a = b$ is possible in the discrete model), such that $u_* = \arg\min_u g_m(u)$ and $u^* = \arg\max_u g_m(u)$. (Obviously, this is true if the functional Fourier coefficients $g_m(\cdot)$ are continuous functions on the compact interval $[a, b]$.) In this case, we have $\tau_1(m) \geq L_* |g_m(u_*)|^2$ and $\tau_1(m) \leq L^* |g_m(u^*)|^2$, where $L_* = L^* = b - a$ in the continuous model and $L_* = L^* = 1$ in the discrete model. Assume also that we can observe $y(u, t)$ at the points u_* and u^*. The following statement presents the case when the minimax convergence rates cannot be influenced by the choice of M and the selection of points u_l, $l = 1, 2, \dots, M$.

PROPOSITION 1. *Let there exist constants $\nu_1 \in \mathbb{R}$, $\nu_2 \in \mathbb{R}$, $\alpha_1 \ge 0$, $\alpha_2 \ge 0$, $\beta_1 > 0$, $\beta_2 > 0$, $L_1 > 0$ and $L_2 > 0$, independent of m, such that*

$$|g_m(u_*)|^2 \ge L_1 |m|^{-2\nu_1} \exp(-\alpha_1 |m|^{\beta_1}), \qquad \nu_1 > 0 \text{ if } \alpha_1 = 0, \tag{6.1}$$

$$|g_m(u^*)|^2 \le L_2 |m|^{-2\nu_2} \exp(-\alpha_2 |m|^{\beta_2}), \qquad \nu_2 > 0 \text{ if } \alpha_2 = 0, \tag{6.2}$$

where either $\alpha_1\alpha_2 > 0$ and $\beta_1 = \beta_2$ or $\alpha_1 = \alpha_2 = 0$ and $\nu_1 = \nu_2$. Then, the minimax convergence rates obtained in Theorems 1 and 2 in the discrete model are independent of the choice of M and the selection of points u_l, $l = 1, 2, \dots, M$, and, hence, coincide with the minimax convergence rates obtained in Theorems 1 and 2 in the continuous model.

The validity of Proposition 1 follows trivially from the lower and upper bounds obtained in Theorems 1 and 2. Proposition 1 simply states that asymptotically (up to a constant factor) it makes absolutely no difference whether one samples (1.4) n times at one point, say, u_1 or, say, $\sqrt{n}$ times at $M = \sqrt{n}$ points u_l. In other words, asymptotically (up to a constant factor) each sample value $y(u_l, t_i)$, $l = 1, 2, \dots, M$, $i = 1, 2, \dots, N$, gives the same amount of information and the minimax convergence rates are not sensitive to the choice of M and the selection of points u_l, $l = 1, 2, \dots, M$. The constants in Theorem 2 will, of course, reflect the difference and will be the smallest if one samples (1.4) n times at u^*.

However, conditions (6.1)–(6.2) are not always true. Consider, for example, the case when $g(u, x) = (2u)^{-1}\,\mathbb{I}(|x| \le u)$, that is, the case of a boxcar-type convolution for each $u \in [a, b]$, $0 < a < b < \infty$. Then, $g_m(u) = \sin(2\pi m u)/(2\pi m u)$ and $|g_m(u_*)|^2 = 0$; indeed, for rational points $u = l_1/l_2 \in [a, b]$, the functional Fourier coefficients $g_m(u)$ vanish for any integer m multiple of l_2. This is an example where a careful choice of u_l, $l = 1, 2, \dots, M$, can make a difference. For example, in the multichannel boxcar deconvolution problem (see also Example 5), De Canditiis and Pensky (2006) showed that if M is finite, $M \ge 2$, one of the u_l's is a "badly approximable" (BA) irrational number, and $u_1, u_2, \dots, u_M$ is a BA irrational tuple, then $\Delta_1(j) \le C j\, 2^{j(2+1/M)}$ [for the definitions of the BA irrational number and the BA irrational tuple, see, e.g., Schmidt (1980)]. This implies that, in this case, (the degree of ill-posedness is) $\nu = 1 + 1/(2M)$. [The case $M = 1$, corresponding to the standard boxcar deconvolution problem, was considered by Johnstone, Kerkyacharian, Picard and Raimondo (2004) who showed that $\nu = 3/2$ when u_1 is a BA irrational number.] Furthermore, De Canditiis and Pensky (2006) obtained asymptotical upper bounds (in the L^π, $1 < \pi < \infty$) for the error, for a wavelet estimator, for a fixed response function. They also showed that these bounds depend on M and the larger the M, is the higher the asymptotical convergence rates will be. Hence, in the multichannel boxcar deconvolution problem, it is advantageous to take $M \to \infty$ and to choose $u_1, u_2, \dots, u_M$ to be a BA tuple.

However, the theoretical results obtained in Theorems 1 and 2 cannot be blindly applied to accommodate the blurring scenario represented by the case of boxcar-type convolution for each fixed u, that is, the case when $g(u,x) = (2u)^{-1}\,\mathbb{I}(|x| \le u)$, $u \in [a,b]$, $0 < a < b < \infty$. A careful treatment of this problem is necessary, since it requires nontrivial results in number theory. This is currently under investigation by the authors and the results of the analysis will be published elsewhere.

7. Proofs. In what follows, for simplicity, we use the notation g instead of $g(\cdot)$, for any arbitrary function $g(\cdot)$. Also, ψ_{jk} refer to the periodized Meyer wavelets defined in Section 2.

7.1. *Lower bounds*.

PROOF OF THEOREM 1. The proof of the lower bounds falls into two parts. First, we consider the lower bounds obtained when the worst functions f (i.e., the hardest functions to estimate) are represented by only one term in a wavelet expansion (sparse case), and then when the worst functions f are uniformly spread over the unit interval T (dense case).

Sparse case. Consider the continuous model (1.3). Let the functions f_{jk} be of the form $f_{jk} = \gamma_j \psi_{jk}$ and let $f_0 \equiv 0$. Note that by (3.1), in order $f_{jk} \in B^s_{p,q}(A)$, we need $\gamma_j \le A 2^{-js'}$. Set $\gamma_j = c 2^{-js'}$, where c is a positive constant such that $c < A$, and apply the following classical lemma on lower bounds:

LEMMA 3 [Härdle, Kerkyacharian, Picard and Tsybakov (1998), Lemma 10.1]. *Let V be a functional space, and let $d(\cdot,\cdot)$ be a distance on V. For $f, g \in V$, denote by $\Lambda_n(f,g)$ the likelihood ratio $\Lambda_n(f,g) = d\mathbb{P}_{X_n^{(f)}}/d\mathbb{P}_{X_n^{(g)}}$, where $d\mathbb{P}_{X_n^{(h)}}$ is the probability distribution of the process X_n when h is true. Let V contains the functions $f_0, f_1, \dots, f_\aleph$ such that*:

(a) $d(f_k, f_{k'}) \ge \delta > 0$ *for* $k = 0, 1, \dots, \aleph$, $k \ne k'$,

(b) $\aleph \ge \exp(\lambda_n)$ *for some* $\lambda_n > 0$,

(c) $\ln \Lambda_n(f_0, f_k) = u_{nk} - v_{nk}$, *where v_{nk} are constants and u_{nk} is a random variable such that there exists $\pi_0 > 0$ with* $\mathbb{P}_{f_k}(u_{nk} > 0) \ge \pi_0$,

(d) $\sup_k v_{nk} \le \lambda_n$.

Then, for an arbitrary estimator $\hat{f}_n$,

$$\sup_{f \in V} \mathbb{P}_{X_n^{(f)}}\big(d(\hat{f}_n, f) \ge \delta/2\big) \ge \pi_0/2.$$

Let now $V = \{f_{jk} : 0 \le k \le 2^j - 1\}$ so that $\aleph = 2^j$. Choose $d(f, g) = \|f - g\|$, where $\|\cdot\|$ is the L^2-norm on the unit interval T. Then, $d(f_{jk}, f_{jk'}) = \gamma_j = \delta$. Let $\upsilon_{nk} = \lambda_n = j \ln 2$ and $u_{nk} = \ln \Lambda_n(f_0, f_{jk}) + j \ln 2$. Now, to apply Lemma 3, we need to show that for some $\pi_0 > 0$, uniformly for all f_{jk}, we have

$$\mathbb{P}_{f_{jk}}(u_{nk} > 0) = \mathbb{P}_{f_{jk}}\big(\ln \Lambda_n(f_0, f_{jk}) > -j \ln 2\big) \ge \pi_0 > 0.$$

Since, by Chebyshev's inequality,

$$\mathbb{P}_{f_{jk}}\big(\ln \Lambda_n(f_0, f_{jk}) > -j \ln 2\big) \ge 1 - \frac{\mathbb{E}_{f_{jk}} |\ln \Lambda_n(f_0, f_{jk})|}{j \ln 2},$$

we need to find a uniform upper bound for $\mathbb{E}_{f_{jk}} |\ln \Lambda_n(f_0, f_{jk})|$.

Let $W(u,t)$ and $\widetilde{W}(u,t)$ be Wiener sheets on $U \times T$. Let $\tilde{z}(u,t) = \sqrt{n}\,(g * f_{jk})(u,t) + z(u,t)$, where $z(u,t) = \dot{W}(u,t)$ and $\tilde{z}(u,t) = \dot{\widetilde{W}}(u,t)$ [i.e., $W(u,t)$ and $\widetilde{W}(u,t)$ are the primitives of $z(u,t)$ and $\tilde{z}(u,t)$, resp.]. Then, assuming that $\int_T \int_U n(g * f_{jk})^2(u,t)\,du\,dt < \infty$, by the multiparameter Girsanov formula [see, e.g., Dozzi (1989), page 89], we get

$$-\ln \Lambda_n(f_0, f_{jk}) = \sqrt{n} \int_T \int_U (g * f_{jk})(u,t)\,dW(u,t) - \frac{n}{2} \int_T \int_U (g * f_{jk})^2(u,t)\,du\,dt. \tag{7.1}$$

Hence,

$$\mathbb{E}_{f_{jk}} |\ln \Lambda_n(f_0, f_{jk})| \le A_n + B_n,$$

where

$$A_n = \sqrt{n}\gamma_j \mathbb{E}\left| \int_T \int_U (\psi_{jk} * g)(u,t)\,dW(u,t) \right|,$$
$$B_n = 0.5 n \gamma_j^2 \int_T \int_U (\psi_{jk} * g)^2(u,t)\,du\,dt.$$

Since, by Jensen's inequality, $A_n \le \sqrt{2B_n}$, we only need to construct an upper bound for B_n. For this purpose, we denote the Fourier coefficients of $\psi(\cdot)$ by $\psi_m = \langle e_m, \psi \rangle$, and observe that in the case of Meyer wavelets, $|\psi_{mjk}| \le 2^{-j/2}$ [see, e.g., Johnstone, Kerkyacharian, Picard and Raimondo (2004), page 565]. Therefore, by properties of the Fourier transform, we get

$$B_n = O\left(2^{-j} n \gamma_j^2 \sum_{m \in C_j} \int_U |g_m(u)|^2\,du \right). \tag{7.2}$$

Let $j = j_n$ be such that

$$\frac{B_n + \sqrt{2B_n}}{j \ln 2} \le \frac{1}{2}. \tag{7.3}$$

Then, by applying Lemma 3 and Chebyshev's inequality, we obtain

$$\begin{aligned}\inf_{\tilde f_n}\sup_{f\in B^s_{p,q}(A)}\mathbb{E}\|\tilde f_n-f\|^2&\ge\inf_{\tilde f_n}\sup_{f\in V}\tfrac14\gamma_j^2\mathbb{P}(\|\tilde f_n-f\|\ge\gamma_j/2)\\&\ge\tfrac14\gamma_j^2\pi_0.\end{aligned}\tag{7.4}$$

Thus, we just need to choose the smallest possible $j=j_n$ satisfying (7.3), to calculate $\gamma_j=c2^{-js'}$, and to plug it into (7.4). By direct calculations, we derive, under condition (3.3), that

$$\sum_{m\in C_j}\int_U|g_m(u)|^2\le\begin{cases}C2^{-j(2\nu-1)}, & \text{if } \alpha=0,\\ C2^{-j(2\nu+\beta-1)}\exp(-\alpha(2\pi/3)^\beta 2^{j\beta}), & \text{if } \alpha>0,\end{cases}\tag{7.5}$$

so that (7.3) yields $2^{j_n}=C(n/\ln n)^{1/(2s'+2\nu)}$ if $\alpha=0$ and $2^{j_n}=C(\ln n)^{1/\beta}$ if $\alpha>0$. Hence, (7.4) yields

$$\inf_{\tilde f_n}\sup_{f\in B^s_{p,q}}\mathbb{E}\|\tilde f_n-f\|^2\ge\begin{cases}C(\ln n/n)^{2s'/(2s'+2\nu)}, & \text{if } \alpha=0,\\ C(\ln n)^{-2s'/\beta}, & \text{if } \alpha>0.\end{cases}\tag{7.6}$$

The proof in the discrete case is almost identical to that in the continuous case with the only difference that [compare with (7.1)]

$$\begin{aligned}-\ln\Lambda_n(f_0,f_{jk})&=0.5\sum_{i=1}^N\sum_{l=1}^M\{[y(u_l,t_i)-\gamma_j(\psi_{jk}*g)]^2(u_l,t_i)-y^2(u_l,t_i)\}\\&=-v_{jk}-u_{jk},\end{aligned}$$

where

$$\begin{aligned}u_{jk}&=\gamma_j\sum_{i=1}^N\sum_{l=1}^M(\psi_{jk}*g)(u_l,t_i)\varepsilon_{li},\\ v_{jk}&=0.5\gamma_j^2\sum_{i=1}^N\sum_{l=1}^M(\psi_{jk}*g)^2(u_l,t_i).\end{aligned}$$

Note that, due to $\mathbb{P}(\varepsilon_{li}>0)=\mathbb{P}(\varepsilon_{li}\le 0)=0.5$, we have $\mathbb{P}(u_{jk}>0)=0.5$. Also, by properties of the discrete Fourier transform, we get

$$v_{jk}\le 0.5n2^{-j}\gamma_j^2\sum_{m\in C_j}M^{-1}\sum_{l=1}^M|g_m(u_l)|^2.$$

By replacing B_n and $B_n+\sqrt{B_n}$ with v_{jk} in the proof for the continuous case, and using (3.3), we arrive at (7.6).

Dense case. Consider the continuous model (1.3). Let η be the vector with components $\eta_k = \pm 1$, $k = 0, 1, \ldots, 2^j - 1$, denote by Ξ the set of all possible vectors η, and let $f_{j\eta} = \gamma_j \sum_{k=0}^{2^j-1} \eta_k \psi_{jk}$. Let also η^i be the vector with components $\eta_k^i = (-1)^{\mathbb{I}(i=k)} \eta_k$ for $i, k = 0, 1, \ldots, 2^j - 1$. Note that by (3.1), in order $f_{j\eta} \in B_{p,q}^s(A)$, we need $\gamma_j \le A 2^{-j(s+1/2)}$. Set $\gamma_j = c_\star 2^{-j(s+1/2)}$, where $c_\star$ is a positive constant such that $c_\star < A$, and apply the following lemma on lower bounds:

LEMMA 4 [Willer (2005), Lemma 2]. *Let $\Lambda_n(f, g)$ be defined as in Lemma 3, and let η and $f_{j\eta}$ be as described above. Suppose that, for some positive constants λ and π_0, we have*

$$\mathbb{P}_{f_{j\eta}}\big(-\ln \Lambda_n(f_{j\eta^i}, f_{j\eta}) \le \lambda\big) \ge \pi_0,$$

uniformly for all $f_{j\eta}$ and all $i = 0, \ldots, 2^j - 1$. Then, for any arbitrary estimator $\tilde{f}_n$ and for some positive constant C,

$$\max_{\eta \in \Xi} \mathbb{E}_{f_{j\eta}} \|\tilde{f}_n - f_{j\eta}\| \ge C \pi_0 e^{-\lambda} 2^{j/2} \gamma_j.$$

Hence, similarly to the sparse case, to obtain the lower bounds it is sufficient to show that

$$\mathbb{E}_{f_{j\eta}} |\ln \Lambda_n(f_{j\eta^i}, f_{j\eta})| \le \lambda_1,$$

for a sufficiently small positive constant λ_1. Then, by the multiparameter Girsanov formula [see, e.g., Dozzi (1989), page 89], we get

$$\begin{aligned}
\ln \Lambda_n(f_{j\eta^i}, f_{j\eta}) = \sqrt{n} \int_T \int_U \big(g * (f_{j\eta^i} - f_{j\eta})\big)(u, t)\, dW(u, t) \\
- \frac{n}{2} \int_T \int_U \big(g * (f_{j\eta^i} - f_{j\eta})\big)^2(u, t)\, du\, dt,
\end{aligned}$$

and recall that $|f_{j\eta^i} - f_{j\eta})| = 2|\psi_{ji}|$. Then,

$$\mathbb{E}_{f_{j\eta}} |\ln \Lambda_n(f_{j\eta^i}, f_{j\eta})| \le A_n + B_n,$$

where

$$\begin{aligned}
A_n &= 2\sqrt{n} \gamma_j \mathbb{E} \left| \int_T \int_U (\psi_{ji} * g)(u, t)\, dW(u, t) \right|, \\
B_n &= 2n\gamma_j^2 \int_T \int_U (\psi_{ji} * g)^2(u, t)\, du\, dt.
\end{aligned}$$

Hence, similarly to the sparse case, $A_n \le \sqrt{2B_n}$ and (7.2) is valid. According to Lemma 4, we choose $j = j_n$ that satisfies the condition $B_n + \sqrt{2B_n} \le \lambda_1$. Using

(7.5), we derive that $2^{j_n} = Cn^{1/(2s+2\nu+1)}$ if $\alpha = 0$ and $2^{j_n} = C(\ln n)^{1/\beta}$ if $\alpha > 0$. Therefore, Lemma 4 and Jensen's inequality yield

$$\inf_{\tilde{f}_n} \sup_{f \in B^s_{p,q}} \mathbb{E}\|\tilde{f}_n - f\|^2 \ge \begin{cases} Cn^{-2s/(2s+2\nu+1)}, & \text{if } \alpha = 0, \\ C(\ln n)^{-2s/\beta}, & \text{if } \alpha > 0. \end{cases} \tag{7.7}$$

The proof can be now extended to the discrete case in exactly the same manner as in the sparse case. Now, to complete the proof one just needs to note that $s^* = \min(s, s')$, and that

$$2s/(2s + 2\nu + 1) \le 2s^*/(2s^* + 2\nu) \qquad \text{if } \nu(2 - p) \le ps^*, \tag{7.8}$$

with the equalities taken place simultaneously, and then to choose the highest of the lower bounds (7.6) and (7.7). This completes the proof of Theorem 1. □

7.2. *Upper bounds*.

PROOF OF LEMMA 1. In what follows, we shall only construct the proof for b_{jk} [i.e., the proof of (4.8)] since the proof for a_{j_0k} [i.e., the proof of (4.7)] is very similar. First, consider the continuous model (1.3). Note that, by (2.9),

$$\widehat{b}_{jk} - b_{jk} = \sum_{m \in C_j} (\widehat{f}_m - f_m)\overline{\psi_{mjk}},$$

where

$$\widehat{f}_m - f_m = n^{-1/2}\left(\int_a^b \overline{g_m(u)} z_m(u)\, du\right) \Big/ \left(\int_a^b |g_m(u)|^2\, du\right), \tag{7.9}$$

due to (2.2) and (2.4). Recall that $z_m(u)$ are Gaussian processes with zero mean and covariance function satisfying (2.3). Hence, it is easy to check that

$$\mathbb{E}[(\widehat{f}_{m_1} - f_{m_1})\overline{(\widehat{f}_{m_2} - f_{m_2})}] = n^{-1}[\tau_1(m_1)]^{-1}\delta(m_1 - m_2),$$

implying that

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = n^{-1} \sum_{m \in C_j} |\psi_{mjk}|^2 [\tau_1(m)]^{-1},$$

where $\tau_1(m)$ is defined in (3.2) (the continuous case). To complete the proof of (4.8) in the case of $\kappa = 1$, just recall that $|C_j| = 4\pi 2^j$ and $|\psi_{mjk}|^2 \le 2^{-j}$. If $\kappa = 2$, then

$$\begin{aligned} \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^4\right) + O\left(\left[\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^2\right]^2\right) \\ &= O\left(n^{-2} \sum_{m \in C_j} |\psi_{mjk}|^4 \tau_2(m)[\tau_1(m)]^{-4}\right) \end{aligned}$$

$$+ O\left(n^{-2}\left[|C_j|^{-1} \sum_{m \in C_j} [\tau_1(m)]^{-1}\right]^2\right)$$
$$= O(n^{-2} 2^{-j} \Delta_2(j)) + O(n^{-2}\Delta_1^2(j)) = O(n^{-2}\Delta_2(j)),$$

since, by the Cauchy–Schwarz inequality, $\Delta_1^2(j) \le \Delta_2(j)$. This completes the proof of (4.8) in the continuous case.

In the discrete case, formula (7.9) takes the form [see (2.6)]

$$\widehat{f_m} - f_m = N^{-1/2}\left(\sum_{l=1}^{M} \overline{g_m(u_l)}\, z_{ml}\right) \Big/ \left(\sum_{l=1}^{M} |g_m(u_l)|^2\right), \tag{7.10}$$

where z_{ml} are standard Gaussian random variables, independent for different m and l. Therefore, similarly to the continuous case,

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = N^{-1} \sum_{m \in C_j} |\psi_{mjk}|^2 \left[\sum_{l=1}^{M} |g_m(u_l)|^2\right]^{-1} = O(n^{-1}\Delta_1(j)).$$

In the case of $\kappa = 2$, note that

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 = O\big(2^{-j} N^{-2} M^{-3} \Delta_2(j) + N^{-2} M^{-2} \Delta_1^2(j)\big) = O(n^{-2}\Delta_2(j)),$$

by applying again the Cauchy–Schwarz inequality. This completes the proof of (4.8) in the discrete case.

The last part of the lemma follows easily from (4.2) with $\kappa = 2$, using the assumption (3.3) and the Cauchy–Schwarz inequality, thus completing the proof of Lemma 1. □

PROOF OF LEMMA 2. Consider the set of vectors

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

and the centered Gaussian process defined by

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

The proof of the lemma is based on the following inequality:

LEMMA 5 [Cirelson, Ibragimov and Sudakov (1976)]. *Let D be a subset of $\mathbb{R} = (-\infty, \infty)$, and let $(\xi_t)_{t \in D}$ be a centered Gaussian process. If $\mathbb{E}(\sup_{t \in D} \xi_t) \le B_1$ and $\sup_{t \in D} \mathrm{Var}(\xi_t) \le B_2$, then, for all $x > 0$, we have*

$$\mathbb{P}\left(\sup_{t \in D} \xi_t \ge x + B_1\right) \le \exp(-x^2/(2B_2)). \tag{7.11}$$

To apply Lemma 5, we need to find B_1 and B_2. Note that, by Jensen's inequality, we get

$$\mathbb{E}\Big[\sup_{v\in\Omega_{jr}} Z_{jr}(v)\Big] = \mathbb{E}\Big[\sum_{k\in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2\Big]^{1/2} \le \Big[\sum_{k\in U_{jr}} \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2\Big]^{1/2} \le \sqrt{c_1} n^{-1/2} 2^{\nu j}\sqrt{\ln n}.$$

[Here, c_1 is the same positive constant as in (4.3) with $\alpha = 0$.] Also, by (2.3) and (7.9) or (7.10), we have

$$\mathbb{E}[(\widehat{b}_{jk} - b_{jk})(\widehat{b}_{jk'} - b_{jk'})] = n^{-1}\sum_{m\in C_j} \psi_{mjk}\overline{\psi_{mjk'}}[\tau_1(m)]^{-1},$$

where $\tau_1(m)$ is defined in (3.2). Hence,

$$\sup_{v\in\Omega_{jr}} \operatorname{Var}(Z_{jr}(v)) = n^{-1}\sup_{v\in\Omega_{jr}}\sum_{k\in U_{jr}}\sum_{k'\in U_{jr}} v_k v_{k'}\sum_{m\in C_j}\psi_{mjk}\overline{\psi_{mjk'}}[\tau_1(m)]^{-1} \le c_1 n^{-1}2^{2\nu j}\sum_{k\in U_{jr}} v_k^2 \le c_1 n^{-1} 2^{2\nu j},$$

by using $\sum_{m\in C_j}\psi_{mjk}\overline{\psi_{mjk'}} = \mathbb{I}(k=k')$ and (4.3) for $\alpha = 0$. Therefore, by applying Lemma 5 with $B_1 = \sqrt{c_1} n^{-1/2}2^{\nu j}\sqrt{\ln n}$, $B_2 = c_1 n^{-1}2^{2\nu j}$ and $x = (0.5\mu - \sqrt{c_1})n^{-1/2}2^{\nu j}\sqrt{\ln n}$, we get

$$\begin{aligned}\mathbb{P}\Big(\sum_{k\in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \ge 0.25\mu^2 n^{-1}2^{2\nu j}\ln n\Big) &= \mathbb{P}\Big(\Big[\sum_{k\in U_{jr}}|\widehat{b}_{jk} - b_{jk}|^2\Big]^{1/2} \ge \sqrt{c_1}n^{-1/2}2^{\nu j}\sqrt{\ln n} + x\Big)\\ &\le \exp\big(-(2c_1)^{-1}(0.5\mu - \sqrt{c_1})^2\ln n\big) \le n^{-\theta},\end{aligned}$$

where $\theta = (8\nu - 4\nu_1 + 2)/(2\nu + 1)$, provided that $\mu \ge 2\sqrt{c_1}(1 + \sqrt{2\theta})$. This completes the proof of Lemma 2. $\square$

PROOF OF THEOREM 2. First, note that in the case of $\alpha > 0$, we have

$$\mathbb{E}\|\hat{f}_n - f\|^2 = R_1 + R_2,$$

where

$$R_1 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}^2, \qquad R_2 = \sum_{k=0}^{2^{j_0}-1} \mathbb{E}(\widehat{a}_{j_0k} - a_{j_0k})^2, \tag{7.12}$$

since $j_0 = J$. It is well known [see, e.g., Johnstone (2002), Lemma 19.1] that if $f \in B_{p,q}^s(A)$, then for some positive constant $c^\star$, dependent on p, q, s and A only, we have

$$\sum_{k=0}^{2^j-1} b_{jk}^2 \le c^\star 2^{-2js^*}, \tag{7.13}$$

thus, $R_1 = O(2^{-2Js^*}) = O((\ln n)^{-2s^*/\beta})$. Also, using (4.3) and (4.7), we derive

$$R_2 = O(n^{-1}2^{j_0}\Delta_1(j_0)) = O(n^{-1/2}(\ln n)^{2\nu/\beta}) = o((\ln n)^{-2s^*/\beta}),$$

thus completing the proof for $\alpha > 0$.

Now, consider the case of $\alpha = 0$. Due to the orthonormality of the wavelet basis, we get

$$\mathbb{E}\|\hat{f}_n - f\|^2 = R_1 + R_2 + R_3 + R_4, \tag{7.14}$$

where R_1 and R_2 are defined in (7.12), and

$$R_3 = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}(\widehat{B}_{jr} \ge dn^{-1}2^{2\nu j}\ln n)],$$

$$R_4 = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}[b_{jk}^2 \mathbb{I}(\widehat{B}_{jr} < dn^{-1}2^{2\nu j}\ln n)],$$

where $\widehat{B}_{jr}$ and d are given by (2.10) and (4.6), respectively.

Let us now examine each term in (7.14) separately. Similarly to the case of $\alpha > 0$, we obtain $R_1 = O(2^{-2Js^*}) = O(n^{-2s^*/(2\nu+1)})$. By direct calculations, one can check that $2s^*/(2\nu+1) > 2s/(2s+2\nu+1)$, if $\nu(2-p) < ps^*$, and $2s^*/(2\nu+1) \ge 2s^*/(2s^*+2\nu)$, if $\nu(2-p) \ge ps^*$. Hence,

$$R_1 = O\big(n^{-2s/(2s+2\nu+1)}\big) \qquad \text{if } \nu(2-p) < ps^*, \tag{7.15}$$

$$R_1 = O\big(n^{-2s^*/(2s^*+2\nu)}\big) \qquad \text{if } \nu(2-p) \ge ps^*. \tag{7.16}$$

Also, by (4.7) and (4.3), we get

$$\begin{aligned} R_2 &= O\big(n^{-1}2^{(2\nu+1)j_0}\big) = O(n^{-1}(\ln n)^{2\nu+1}) \\ &= o\big(n^{-2s/(2s+2\nu+1)}\big) = o\big(n^{-2s^*/(2s^*+2\nu)}\big). \end{aligned} \tag{7.17}$$

To construct the upper bounds for R_3 and R_4, note that simple algebra gets

$$R_3 \le (R_{31} + R_{32}), \qquad R_4 \le (R_{41} + R_{42}), \tag{7.18}$$

where

$$R_{31} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq 0.25 d n^{-1} 2^{2\nu j} \ln n \right) \right],$$

$$R_{32} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}(B_{jr} > 0.25 d n^{-1} 2^{2\nu j} \ln n)],$$

$$R_{41} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[b_{jk}^2 \mathbb{I}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq 0.25 d n^{-1} 2^{2\nu j} \ln n \right) \right],$$

$$R_{42} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}[b_{jk}^2 \mathbb{I}(B_{jr} < 2.5 d n^{-1} 2^{2\nu j} \ln n)],$$

since $\widehat{b}_{jk}^2 \leq 2(\widehat{b}_{jk} - b_{jk})^2 + 2b_{jk}^2$. Then, by (7.13), Lemmas 1 and 2, and the Cauchy–Schwarz inequality, we derive

$$\begin{aligned}
R_{31} + R_{41} &= \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\Bigg(((\widehat{b}_{jk} - b_{jk})^2 + b_{jk}^2) \\
&\qquad \times \mathbb{I}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq 0.25 d n 2^{2\nu j} \ln n \right) \Bigg) \\
&= O\Bigg(\sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \left(\sqrt{\mathbb{E}(\widehat{b}_{jk} - b_{jk})^4} + b_{jk}^2 \right) \\
&\qquad \times \sqrt{\mathbb{P}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq 0.25 d n 2^{2\nu j} \ln n \right)} \Bigg) \\
&= O\left(\sum_{j=j_0}^{J-1} [2^j n^{-1} 2^{2(2\nu - \nu_1) j} + 2^{-2js^*}] n^{-(4\nu - 2\nu_1 + 1)/(2\nu + 1)} \right) \\
&= O(n^{-1}),
\end{aligned}$$

provided $d \geq c_1(1 + \sqrt{2\theta})^2$, where $\theta = (8\nu - 4\nu_1 + 2)/(2\nu + 1)$ and c_1 is the same positive constant as in (4.3) with $\alpha = 0$. Hence,

$$\Delta_1 = R_{31} + R_{41} = O(n^{-1}). \tag{7.19}$$

Now, consider

$$\Delta_2 = R_{32} + R_{42}. \tag{7.20}$$

Let j_1 be such that

$$2^{j_1} = n^{1/(2s+2\nu+1)}(\ln n)^{\varrho_1/(2\nu+1)}, \tag{7.21}$$

where ϱ_1 is defined in (4.11).

First, let us study the dense case, that is, when $\nu(2-p) < ps^*$. Then, Δ_2 can be partitioned as $\Delta_2 = \Delta_{21} + \Delta_{22}$, where the first component is calculated over the set of indices $j_0 \le j \le j_1$ and the second component over $j_1 + 1 \le j \le J - 1$. Hence, using (2.10) and Lemma 1, and taking into account that the cardinality of A_j is $|A_j| = 2^j / \ln n$, we obtain

$$\begin{aligned}\Delta_{21} &= O\left(\sum_{j=j_0}^{j_1}\left[\sum_{k=0}^{2^j-1}\mathbb{E}(\widehat{b}_{jk} - b_{jk})^2\right.\right.\\ &\qquad\left.\left. + \sum_{r\in A_j} B_{jr}\mathbb{I}(B_{jr} \le 2.5dn^{-1}2^{2\nu j}\ln n)\right]\right)\\ &= O\left(\sum_{j=j_0}^{j_1}\left[n^{-1}2^{(2\nu+1)j} + \sum_{r\in A_j} n^{-1}2^{2\nu j}\ln n\right]\right)\\ &= O\big(n^{-1}2^{(2\nu+1)j_1}\big) = O\big(n^{-2s/(2s+2\nu+1)}(\ln n)^{\varrho_1}\big).\end{aligned}\tag{7.22}$$

To obtain an expression for Δ_{22}, note that, by (7.13), and for $p \ge 2$, we have

$$\begin{aligned}\Delta_{22} &= O\left(\sum_{j=j_1+1}^{J-1}\sum_{r\in A_j}[n^{-1}2^{2\nu j}\ln n\,\mathbb{I}(B_{jr} \ge 0.25dn^{-1}2^{2\nu j}\ln n) + B_{jr}]\right)\\ &= O\left(\sum_{j=j_1+1}^{J-1}\sum_{r\in A_j} B_{jr}\right) = O\left(\sum_{j=j_1+1}^{J-1} 2^{-2js}\right)\\ &= O\big(n^{-2s/(2s+2\nu+1)}\big).\end{aligned}\tag{7.23}$$

If $1 \le p < 2$, then

$$B_{jr}^{p/2} = \left(\sum_{k\in U_{jr}} b_{jk}^2\right)^{p/2} \le \sum_{k\in U_{jr}} |b_{jk}|^p,$$

so that by Lemma 1, and since $\nu(2-p) < ps^*$, we obtain

$$\begin{aligned}\Delta_{22} = O\bigg(\sum_{j=j_1+1}^{J-1}\sum_{r\in A_j}&[n^{-1}2^{2\nu j}\ln n\,\mathbb{I}(B_{jr} \ge 0.25dn^{-1}2^{2\nu j}\ln n)\\ &+ B_{jr}\mathbb{I}(B_{jr} \le 2.5dn^{-1}2^{2\nu j}\ln n)]\bigg)\end{aligned}\tag{7.24}$$

$$
\begin{aligned}
&= O\Bigg(\sum_{j=j_1+1}^{J-1}\sum_{r\in A_j}[(n^{-1}2^{2\nu j}\ln n)^{1-p/2}B_{jr}^{p/2} \\
&\qquad\qquad + B_{jr}^{p/2}(n^{-1}2^{2\nu j}\ln n)^{1-p/2}]\Bigg) \\
&= O\Bigg(\sum_{j=j_1+1}^{J-1}(n^{-1}2^{2\nu j}\ln n)^{1-p/2}\sum_{r\in A_j}\sum_{k\in U_{jr}}|b_{jk}|^p\Bigg) \\
&= O\Bigg(\sum_{j=j_1+1}^{J-1}(n^{-1}2^{2\nu j}\ln n)^{1-p/2}2^{-pjs^*}\Bigg) \\
&= O\Bigg(\sum_{j=j_1+1}^{J-1}(n^{-1}\ln n)^{1-p/2}2^{(2\nu-p\nu-ps^*)j}\Bigg) \\
&= O\big((n^{-1}\ln n)^{1-p/2}2^{(2\nu-p\nu-ps^*)j_1}\big) \\
&= O\big(n^{-2s/(2s+2\nu+1)}(\ln n)^{\varrho_1}\big).
\end{aligned}
$$

Let us now study the sparse case, that is, when $\nu(2-p) > ps^*$. Let j_1 be defined by (7.21) with $\varrho_1 = 0$. Hence, if $B_{jr} \geq 0.25dn^{-1}2^{2\nu j}\ln n$, then $\sum_{k=0}^{2^j-1} b_{jk}^2 \geq 0.25dn^{-1}2^{(2\nu+1)j}$, implying that j cannot exceed j_2 such that $2^{j_2} = (4c^*n/(d\ln n))^{1/(2s^*+2\nu)}$, where c^* is the same constant as in (7.13). Again, partition $\Delta_2 = \Delta_{21} + \Delta_{22}$, where the first component is calculated over $j_0 \leq j \leq j_2$ and the second component over $j_2+1 \leq j \leq J-1$. Then, using arguments similar to those in (7.24), and taking into account that $\nu(2-p) > ps^*$, we derive

$$
\begin{aligned}
\Delta_{21} &= O\Bigg(\sum_{j=j_0}^{j_2}(n^{-1}2^{2\nu j}\ln n)^{1-p/2}\sum_{r\in A_j}\sum_{k\in U_{jr}}|b_{jk}|^p\Bigg) \\
&= O\Bigg(\sum_{j=j_0}^{j_2}(n^{-1}2^{2\nu j}\ln n)^{1-p/2}2^{-pjs^*}\Bigg) \\
&= O\Bigg(\sum_{j=j_0}^{j_2}(n^{-1}\ln n)^{1-p/2}2^{(2\nu-p\nu-ps^*)j}\Bigg) \qquad (7.25) \\
&= O\big((n^{-1}\ln n)^{1-p/2}2^{(2\nu-p\nu-ps^*)j_2}\big) \\
&= O\big((\ln n/n)^{2s^*/(2s^*+2\nu)}\big).
\end{aligned}
$$

To obtain an upper bound for Δ_{22}, recall (7.20) and keep in mind that the portion of R_{32} corresponding to $j_2+1 \le j \le J-1$ is just zero. Hence, by (7.13), we get

$$\begin{aligned}\Delta_{22} &= O\left(\sum_{j=j_2+1}^{J-1}\sum_{k=0}^{2^j-1} b_{jk}^2\right) = O\left(\sum_{j=j_2+1}^{J-1} 2^{-2js^*}\right)\\ &= O\big((\ln n/n)^{2s^*/(2s^*+2\nu)}\big).\end{aligned}\tag{7.26}$$

Now, in order to complete the proof, we just need to study the case when $\nu(2-p) = ps^*$. In this situation, we have $2s/(2s+2\nu+1) = 2s^*/(2s^*+2\nu) = 1-p/2$ and $2\nu j(1-p/2) = pjs^*$. Recalling (3.1) and noting that $s^* \le s'$, we get

$$\sum_{j=j_0}^{J-1}\left(2^{pjs^*}\sum_{k=0}^{2^j-1}|b_{jk}|^p\right)^{q/p} \le A^q.$$

Then, we repeat the calculations in (7.25) for all indices $j_0 \le j \le J-1$. If $1 \le p < q$, then, by Hölder's inequality, we get

$$\begin{aligned}\Delta_2 &= O\left(\sum_{j=j_0}^{J-1}(\ln n/n)^{1-p/2}2^{pjs^*}\sum_{k=0}^{2^j-1}|b_{jk}|^p\right)\\ &= O\left((\ln n/n)^{1-p/2}(\ln n)^{1-p/q}\left[\sum_{j=j_0}^{J-1}\left(2^{pjs^*}\sum_{k=0}^{2^j-1}|b_{jk}|^p\right)^{q/p}\right]^{p/q}\right)\\ &= O\big((\ln n/n)^{2s^*/(2s^*+2\nu)}(\ln n)^{1-p/q}\big).\end{aligned}\tag{7.27}$$

If $1 \le q \le p$, then, by the inclusion $B_{p,q}^s(A) \subset B_{p,p}^s(A)$, we get

$$\begin{aligned}\Delta_2 &= O\left(\sum_{j=j_0}^{J-1}(\ln n/n)^{1-p/2}2^{pjs^*}\sum_{k=0}^{2^j-1}|b_{jk}|^p\right)\\ &= O\left((\ln n/n)^{1-p/2}\left[\sum_{j=j_0}^{J-1}2^{pjs^*}\sum_{k=0}^{2^j-1}|b_{jk}|^p\right]\right)\\ &= O\big((\ln n/n)^{2s^*/(2s^*+2\nu)}\big).\end{aligned}\tag{7.28}$$

By combining (7.15)–(7.17), (7.19), (7.22)–(7.28), we complete the proof of Theorem 2. □

Acknowledgments. Marianna Pensky is grateful for the hospitality and financial support of the Department of Mathematics and Statistics at the University of Cyprus, Cyprus, and Theofanis Sapatinas is grateful for the hospitality of the Department of Mathematics at the University of Central Florida, USA, where parts of the work of this paper were carried out. The authors would like to thank Thomas Willer for useful discussions. Finally, we would like to thank an Associate Editor and two anonymous referees for their suggestions on improvements to this paper.

REFERENCES

ABRAMOVICH, F. and SILVERMAN, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. MR1627226

CASEY, S. D. and WALNUT, D. F. (1994). Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms. *SIAM Rev.* **36** 537–577. MR1306923

CHESNEAU, C. (2008). Wavelet estimation via block thresholding: A minimax study under L^p-risk. *Statist. Sinica* **18** 1007–1024.

CIRELSON, B. S., IBRAGIMOV, I. A. and 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 Math.* **550** 20–41. Springer, Berlin. MR0458556

DE CANDITIIS, D. and 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. and PENSKY, M. (2006). Simultaneous wavelet deconvolution in periodic setting. *Scand. J. Statist.* **33** 293–306. MR2279644

DONOHO, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Computat. Harmon. Anal.* **2** 101–126. MR1325535

DONOHO, D. L. and RAIMONDO, M. (2004). Translation invariant deconvolution in a periodic setting. *Internat. J. Wavelets, Multiresolution and Information Processing* **14** 415–432. MR2104873

DOZZI, M. (1989). *Stochastic Processes with a Multidimensional Parameter*. Longman, New York. MR0991563

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19** 1257–1272. MR1126324

FAN, J. and KOO, J. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory* **48** 734–747. MR1889978

GOLUBEV, G. (2004). The principle of penalized empirical risk in severely ill-posed problems. *Probab. Theory Related Filelds* **130** 18–38. MR2092871

GOLUBEV, G. K. and KHASMINSKII, R. Z. (1999). A statistical approach to some inverse problems for partial differential equations. *Problems Inform. Transmission* **35** 136–149. MR1728907

HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statist.* **129**. Springer, New York. MR1618204

HARSDORF, S. and REUTER, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

HESSE, C. H. (2007). The heat equation with initial data corrupted by measurement error and missing data. *Statist. Inference Stochastic Processes* **10** 75–95. MR2269605

JOHNSTONE, I. M. (2002). Function estimation in Gaussian noise: sequence models. Unpublished Monograph. Available at http://www-stat.stanford.edu/~imj/.

JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting (with discussion). *J. Roy. Statist. Soc. Ser. B* **66** 547–573. MR2088290

JOHNSTONE, I. M. and RAIMONDO, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Ann. Statist.* **32** 1781–1804. MR2102493

KALIFA, J. and MALLAT, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.* **31** 58–109. MR1962500

KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statist. Sinica* **7** 317–340. MR2352512

KOLACZYK, E. D. (1994). Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

LATTES, R. and LIONS, J. L. (1967). *Methode de Quasi-Reversibilite et Applications*. *Travoux et Recherche Mathematiques* **15**. Dunod, Paris. MR0232549

MALLAT, S. G. (1999). *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, San Diego. MR1614527

MEYER, Y. (1992). *Wavelets and Operators*. Cambridge Univ. Press. MR1228209

NEELAMANI, R., CHOI, H. and BARANIUK, R. (2004). Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Trans. Signal Processing* **52** 418–433. MR2044455

PARK, Y. J., DHO, S. W. and KONG, H. J. (1997). Deconvolution of long-pulse lidar signals with matrix formulation. *Applied Optics* **36** 5158–5161.

PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627

PENSKY, M. and ZAYED, A. I. (2002). Density deconvolution of different conditional densities. *Ann. Instit. Statist. Math.* **54** 701–712. MR1932413

SCHMIDT, W. (1980). *Diophantine Approximation*. *Lecture Notes in Math.* **785**. Springer, Berlin. MR0568710

STRAUSS, W. A. (1992). *Partial Differential Equations: An Introduction*. Wiley, New York. MR1159712

WALTER, G. and SHEN, X. (1999). Deconvolution using Meyer wavelets. *J. Integral Equations and Applications* **11** 515–534. MR1738280

WILLER, T. (2005). Deconvolution in white noise with a random blurring function. Preprint. arXiv:math/0505142.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FLORIDA 32816-1353
USA
E-MAIL: mpensky@pegasus.cc.ucf.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF CYPRUS
P.O. BOX 20537
NICOSIA CY 1678
CYPRUS
E-MAIL: T.Sapatinas@ucy.ac.cy